UNIVERSITY OF SOUTHAMPTON

A REVISION OF *Vicia* SUBGENUS *Vicia* USING DATABASE TECHNIQUES

A Thesis submitted for the Degree
of Doctor of Philosophy in the
University of Southampton

by

Nigel Maxted
Department of Biology
September, 1990

"What we cannot speak about we must passover in silence"

Wittgenstein (1974)

"The revolution , if such it is to be, will arise from the application of computers to systematics, both in data-processing and in the more restricted field of production of classifications and keys."

Heywood (1974)

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SCIENCE

BIOLOGY

Doctor of Philosophy

A REVISION OF *VICIA* SUBGENUS *VICIA* USING DATABASE TECHNIQUES

by

Nigel Maxted

The thesis investigates the applications of contemporary database technology to enhance the process and production of a taxonomic revision. The current uses of databases and their specific applications to plant taxonomy are discussed. The stages of a revision are outlined and as a result, the revision process is clarified.

In order to experiment with the uses of databases during a revision, an exemplar revision of *Vicia* L. subgenus *Vicia* was undertaken. A detailed taxonomic history of the subgenus is provided and the phenetic approach to data analysis is introduced. Phenetic analysis of 1539 specimens, representing the variation within subgenus *Vicia*, was undertaken for 174 characters, using cluster analysis and ordination techniques. The results of the phenetic analysis are considered in conjunction with data abstracted from the literature and a novel classification of the subgenus is proposed. This comprises nine sections, nine series, thirty eight species and thirty six subspecific taxa. The classification is compared with previous conceptions of taxa relationships. Other revision products are incorporated into a conspectus for the accepted species which includes: a checklist of accepted taxa; comments on type material; synonymised lists; descriptions; phenology; chromosome numbers; geographical distributions; ecological data; taxonomic notes and specimen citations.

All data collected during the revision were held in the revision database. The application of database techniques within the exemplar revision were discussed critically. It is found that use of the techniques facilitated the revision process. This is most apparent in the production of the conspectus, a large proportion of which was produced directly from the revision database. Taxon descriptions were produced directly (via a dBASE program) from the descriptive data for representative specimens of each taxon. After discussing the advantages and disadvantages of applying database techniques to the exemplar revision, a novel conception of the revision paradigm, which incorporates database techniques, is proposed. Comments are made on the need for further developments of taxonomic database management systems to facilitate revisions and broader taxonomic processes in the future.

CONTENTS

VOLUME 2

ACKNOWLEDGEMENTS

I am particularly grateful to the following:

Dr. F.A. Bisby, Dr. R.J. White and Dr. R.J. Pankhurst - for supervision, encouragement and stimulating discussion.

Dr. R. Allkin - for discussion and advice on the database techniques and numerical methods.

Prof. J.T. Williams and Mr. R. Reid - of the International Board for Plant Genetic Resources, who provided advice and financial support for the germplasm collecting trips to Syria, Turkey, Yugoslavia and the Soviet Union. These field trips enabled me to gain a more detailed understanding of Vicia taxa inter-relationships and gave me a chance to obtain germplasm of the rare and novel species included in my study group.

Dr. R.M. Polhill, Mr. G. Lewis and Mrs. B. Mackinder - for encouragement, advice and discussion of the project.

Mr. R. Crust, Ms. D. Richardson, Mr. J. Williams and Mr. P. Winfield - for help in writing the dBASE II and C programs used during this project.

Mr. J. Hepworth, Mr. A. Cotton, Dr. P. Hancock and Ms. F. Baker - for trying to make the CLUSTAN analysis of a large data set work and Dr. J. Henderson of CLUSTAN Ltd. for, at least, partially succeeding.

Dr. S. Hollis and Mr. R. Cornick - for assistance in drawing the figures.

The Staff of the Department of the Biology, Southampton - for help in numerous ways throughout the project.

My thanks are also due to the curators of the Herbaria listed in appendix 1 from whom the specimens used in the project were borrowed.

Science and Engineering Research Council - for the earmarked studentship, which provided the core funds for this research project.

Finally, I would like to thank Barbara and Rebecca for their love and their healthy disbelief in my intellectual ability.

VOLUME 1

SECTION

ONE

INTRODUCTION

CHAPTER ONE
GENERAL INTRODUCTION

It is difficult to imagine a world in which humankind does not characterise, describe and order their surroundings, an activity which is as basic to them as the use of language. This process is central to taxonomy which, in the biological context, is concerned with living organisms. This project involved the efficient cataloguing and ordering of biological information which could then be used, via the taxonomic process, to catalogue, describe and order living organisms.

1.1 Problem Elucidation

Taxonomy, more than any other branch of biology, requires a high investment in data capture, storage and retrieval. As taxonomic procedures themselves evolve, they will become increasingly impractical without a means of easy access, management and utilisation of their vast data resources. Heywood (1974) states that one of the basic problems of contemporary taxonomy is that it is "largely chaotically organised", but that "electronic data-processing and the establishment of databanks will be powerful tools" in resolving this problem. It is therefore imperative to introduce and apply the ordering abilities of database technology to organise the taxonomic data.

When undertaking a taxonomic revision, a taxonomist gathers data relating to a particular group of organisms, synthesises these data and then produces an indexed catalogue or classification based on the synthesised data. Traditionally, when presenting a classification, taxonomists do not usually present the full synthesised data set. They present their conclusions summarised in the form of a classification. Generally this will be accompanied by segments of the synthesised data set, which substantiate their conclusions (see Taylor, 1971). This process wastes expensively gathered data. Cutbill (1971), Heywood (1973),

Morse (1975) and Raven (1977) all point out that much taxonomic information is discarded or at least not recorded in the highly stylised published form of revision products. There is in fact a significant counter pressure on taxonomists to present even simpler and more succinct products.

As well as wasting data, conventional methods of preparing a revision do not require the production of comparable datasets. If comparable datasets were produced they could be made available to other researchers working in similar fields. Watson (1971) underlines the point that taxonomic description are rarely comparable. He suggests that adopting a computerised approach to tackling taxonomic problems forces taxonomists to organise their data in a more logical manner, which in itself makes the available data more comprehensive and subsequently useful.

The scientific justification for this project therefore was to clarify the traditional paradigm used during a revision, then to use this to help formulate a novel paradigm which incorporated, where appropriate, contemporary computer technology. This clarified paradigm will attempt to catalogue, describe and order the biological data in a more logical fashion, the data may then be more efficiently used to catalogue, describe and order the biological diversity inherent in nature. The central innovation implicit in this research is, therefore, the application of database technology to the revision. This innovation will enhance data gathering and data presentation throughout the revision process. From the start all systematic data will be kept in an individual database. This will be extended to include the resultant taxonomic treatment and used as a source of data for other computer-aided activities including numerical taxonomy, description generation and the production of identification aids. While individual parts of the taxonomic revision have frequently been assisted by computer before (see Allkin and Bisby, 1984) there is no known study where computer assistance

has been used at all appropriate stages.

Present applications of database, numerical taxonomy and identification software tend to occur at an advanced stage of a revision by which time bibliographic sources, nomenclatural information, location of types, loans of materials etc. are well organised. However database software can be of use in this preliminary organisation and in moving from the collection of 'primary' data to the synthesised structure of descriptions, classification, nomenclature and identification aids of the final product.

To assess the validity of using computer assistance throughout the revision process, one particular exemplar taxon will be revised, Vicia L. subg. Vicia. This taxon was selected because it contains economically important species (broad bean, Hungarian vetch, field vetch and narrow-leaf vetch); it is relatively small (32 species, Kupicha 1976) and is therefore manageable in the time available and lastly, because Southampton University taxonomy laboratory is an established centre for Vicieae taxonomic research, thus making available facilities and expertise particularly suitable for a revision of this group.

1.2 What is a Taxonomic Database System ?

Computers have shaped the Twentieth Century and in the last twenty years database technology has been the most rapidly growing area in computer and information science (Date, 1981). The ever increasing desire by society for information of all kinds has necessitated the development of fast, efficient methods of data storage and retrieval. The most efficient means of tackling this set of problems is the application of a database system. Date defines a database system thus:

> "it is a nothing more than a computer-based recordkeeping system whose overall purpose is to record and maintain information."

This is the definition provided in the standard text, but it has led to confusion due to its generality. A more restrictive definition of a database is provided by Deen (1977), a database is:

> "a generalised integrated collection of data which is structured on natural data relationships, so that all necessary access paths to each unit of data are ordered to fulfil the differing needs of all users."

The essential characteristic of a database system is that the data are organised so that they can be shared by several programs. This means that the data are the central resource and exist independently of the programs that use or change them (Barron, 1984). If the above definitions are reduced to their constituent elements, they compose: the computer, the recordkeeping system, storage and retrieval, and the information. If these elements are applied to taxonomy they produce the following working definition of a taxonomic database system:

> a taxonomic database system is a computer based, taxonomic information storage and retrieval system.

1.3 Aims

As outlined in the thesis section 1.1 the general aim of this project is to investigate the applications of contemporary database technology as a tool to enhance the process and production of a taxonomic revision.

This general aim will be achieved by the realisation of the following specific aims:

A. To undertake a particular revision of Vicia L. subgenus Vicia (Vicieae, Leguminosae) and to experiment with and assess the use of contemporary database techniques on this particular exemplar.

B. To assess whether using database technology would shorten the time taken to carry out a revision.

C. To assess whether it is possible to produce descriptions for taxa based directly on and produced automatically (via program) from the primary data scored from specimens.

D. To investigate the possibility of making the exemplar database available to the taxonomic community, along with publications drawn from the revision once the revision is completed.

E. To investigate whether using a database encourages better and more detailed data presentation of the data in the revision products.

F. To assess at which stages of the revision process the use of database technology enhances or detracts from the traditional revision paradigm and so propose a revised revision paradigm that incorporates database technology where appropriate.

1.4 Thesis Plan

This thesis is divided into three sections and eleven chapters.

The first section, INTRODUCTION contains three chapters. Chapter 1, General Introduction, outlines the subject of the thesis and the aims of the research. Chapter 2, An Introduction to Taxonomic Databases, provides a background to database technology and detail how databases are currently being used in taxonomy. Chapter 3, The Taxonomic Revision, discusses what research is normally undertaken as part of a taxonomic revision and attempts to clarify the traditional revision paradigm.

The second section, EXEMPLAR REVISION contains six chapters, which detail the process of the exemplar revision. Chapter 4, Taxonomic Introduction, provides an introduction to the exemplar taxon, subgenus Vicia and discusses the aims of the revision. Chapter 5, Taxonomic History, concerns the taxonomic background to subgenus Vicia and discusses prior classifications of the subgenus. Chapter 6, Methods of Data Gathering and Analysis, provides a background to the phenetic methods used in the revision and discusses why particular methods were selected. Chapter 7, Phenetic Studies Analysis Results, describes the results of the phenetic analysis undertaken. Chapter 8, Literature Based Taxonomic Evidence, is a discussion of the taxonomic data available from the literature concerning subgenus Vicia. Chapter 9, Proposed Classification And Taxonomic Discussion, presents the taxonomic primary and secondary products of the revision and compares these products to those previously available.

The third and final section, DATABASE RESULTS AND DISCUSSION discusses the results and implications of the application of database techniques to the exemplar revision. Chapter 10, Application of Databases to a Specific Taxonomic Revision, details how database technology was used during the exemplar revision. Chapter 11, Proposed Taxonomic Database Paradigm, makes more general points about the uses of database management systems during the revision process and discusses how database technology can be used to facilitate the general revision process.

CHAPTER TWO
AN INTRODUCTION TO TAXONOMIC DATABASES

2.1 Introduction to Databases

Taxonomic databases are not intrinsically dependent upon computers (Abbott et al., 1985) but have been used for centuries. In this simple sense a taxonomic database is any store of taxonomic data that is retrievable, e.g. a flora, monograph, card index system or even an illustration. The computer revolution has enabled the taxonomist to save time and effort in the storage, retrieval and processing of taxonomic data. In the following a "database" will be used in the computing sense and will imply the use of a computer in data processing, storage and retrieval.

The central innovation of this research project is the application of database technology to the revision. This chapter introduces database technology. A more comprehensive understanding can be gained from texts such as Cardenas (1985), Date (1986), Howe (1985) and Sundgren (1985). The history of database technology is relatively short. The post-war economic expansion in Western Europe and the USA brought with it the information explosion. It was as a result of attempting to meet the requirements of managing larger datasets that in the mid-1960's the first database systems began to appear (Date, 1986). The use of these systems has expanded exponentially and as a result today all major organizations rely on some form of database system for everyday management.

Bisby (1984b) stresses the importance of distinguishing between textual and codified databases. A textual database contains text that can be printed or displayed on request. It is in effect no more than a sophisticated word processor. Textual databases may be searched for specific strings, but the searches are slow and imprecise if the original text permits a broad vocabulary. A codified database, on the other

hand, contains codes which represent text rather than text itself. The codified database allows multiple uses of the same dataset, which can be searched quickly and precisely as the coding should reduce any problem of text synonyms. Here codified databases will be referred to unless otherwise indicated.

A database system has four components: data, hardware, software and the user (Date, 1981). Data or information is stored within the system in one or more databases, these simply being data stores. Data held in the database is said to be 'integrated' and 'shared'. Integration means that the data item held in each database file is considered to be united with several otherwise distinct data files, which partially or wholly eliminates data duplication. Shared implies that each data item may be shared among several different uses and users, in the sense that each of those users may have access to the same piece of data and may use it for different purposes. This means that different users of the database perceive the database in different ways, allowing multiple means of giving access to partial data sets held in the database.

The hardware is composed of the secondary storage volumes, the disks, the drums, etc., on which the database resides, together with the associated devices and control units which provide the physical element to the system. The software provides an intermediary layer between the actual data and the user. The software that manipulates the data is usually referred to as a database management system or DBMS. All requests by the user for data manipulation are mediated by the DBMS, which acts like a high level computer language shielding the user from the machine level detail.

Date (1981) lists three kinds of database system user. The application programmer, who writes application programs using the database. These application programs manipulate the

data, e.g. retrieve information, create new information, deleting and changing information. The second user category is the end-user, who gain access to the data from the terminal. These users will normally employ a query language provided by the DBMS. The end-user will invoke a particular application program by typing a key word or command, e.g. CREATE or LIST in dBASE (dBASE being the widely used family of DBMS). The third user category is the database administrator who has responsibility for the overall control of the database system. The administrator must consider what information will be included in the database, design the storage structure and access strategy to the data, liaise with end-users, define authorisation checks and validation procedures, define backup and recovery strategies and monitor the performance of the database, as well as responding to the demands of the end-users.

The main advantage of using a database for information storage is related to the centralised control of the data that results from its use. Date (1981) summarises many user's experiences in his list of the advantages of centralised data control, thus:

1. "Redundancy can be reduced."

   Data duplication is prevalent among manual systems of data storage leading to a wastage of storage space and the time and resources required to enter the same data item more than once. Within a database redundancy should be kept to a minimum. Each case of data duplication should result from a positive decision to store the data item more than once and not from poor design of overall data storage.

2. "Inconsistency can be avoided."

This follows from the previous point: if a data item is only stored once then it cannot be internally inconsistent. If a data item is stored more than once then it is possible that one replicate of the item may be updated while another is left at the original value. This would result in internal inconsistency, e.g a recent specimen determination being updated in one part of the database, but not being automatically updated in other parts of the database would result in the same specimen having different identifications within the database.

3. "The data can be shared."

This means that it is possible to view the same data set in more than one way. A database may be indexed on different kinds of items or subsets of the database could be produced. It is possible to use boolean searches to abstract particular subsets of data.

4. "Standards can be enforced."

The centralised control implicit in a database system means that, prior to inclusion in the database, data must be standardised to fit the particular database rules or structure. This is of particular advantage if data are to be inter-changed with other systems.

5. "Security restrictions can be applied."

As database administrators have jurisdiction over the data, they can enforce centralised control over data input and output and thus guard sensitive data.

6. "Integrity can be maintained."

   If a data item is stored only once it is easier for the database administrator to retain data integrity and validate that item. This will ensure that the data items are logically precise e.g. that a specimen is not unintentionally scored twice for the same analysis.

7. "Conflicting requirements can be balanced."

   Knowing the overall requirements of the end-users the database administrator can design a system which provides the optimal overall service to all the end-users.

Barron (1984) makes the additional point that using a database makes it possible for different programs to manipulate the data and so the data are independent of the physical characteristics of the storage medium. This would makes it possible, for example, to change the data storage medium from a floppy disc to a hard disc without any changes to the programs that accesses the database. Bisby (1984b) adds four further advantages of database use: that database information storage allows continuous development, recurrent editing and updating of the database and allows remote access to the database by electronic means.

It is usual for the user to be presented with a particular view of the data matching the users' requirements, in which details of the data storage structures are deliberately omitted. The manner in which the information is held in the database will, however, dictate the design of the data manipulation language and have an effect on data item relationships. The traditional view of data consists of files of records which are made up of fields, e.g. a taxonomist undertaking a revision might maintain files of herbaria from

which loans were obtained, taxa that were loaned and the actual loans received, with the structure shown in Table 2.1.

It should be noted that the data are held in the form of a table of text. Each database user is faced with a choice: how should the data be stored and which basic database design will be employed? The three most widely applied approaches to database architecture are the hierarchical, network and relational approaches. A concise summary of these three approaches is provided by Barron (1984) and the following examples are adapted from his examples for fish supply. All three approaches incorporate the use of codes, which as stated above save computer space and speed up searching operations. The codes used in the examples below are as follows:

Herbarium codes	Herbarium names
BM	Natural History Museum
E	Royal Botanic Garden
HUJ	Hebrew University

Taxon codes	Taxon names
06	V. bithynica
16	V. galilaea
46	V. oroboides
52	V. sativa

The Hierarchical Approach

Data are represented in the form of a simple tree structure. The database administrator decides on the orientation of the hierarchy, e.g. if taxa were placed at the top of the hierarchy the structure shown in Table 2.2 would be obtained. The person querying the database is presented with

Table 2.1. Example of curatorial data files held by a taxonomist.

Herbaria file

Herbarium name	Address
Natural History Museum	London
Royal Botanic Garden	Edinburgh
Hebrew University	Jerusalem

Taxa file

Taxon name	Number of specimens
V. bithynica	40
V. galilaea	12
V. oroboides	5
V. sativa	130

Loans file

Loan origin	Taxon name	Number of specimens
Natural History Museum	V. bithynica	30
Natural History Museum	V. oroboides	2
Natural History Museum	V. sativa	65
Royal Botanic Garden	V. bithynica	10
Royal Botanic Garden	V. galilaea	5
Royal Botanic Garden	V. oroboides	3
Royal Botanic Garden	V. sativa	55
Hebrew University	V. galilaea	7
Hebrew University	V. sativa	10

Table 2.2. Example of curatorial data files held using the hierarchical approach.

06	V. bithynica	40
BM	Natural History Museum, London	30
E	Royal Botanic Garden, Edinburgh	10
16	V. galilaea	12
E	Royal Botanic Garden, Edinburgh	5
HUJ	Hebrew University, Jerusalem	7
46	V. oroboides	5
BM	Natural History Museum, London	2
E	Royal Botanic Garden, Edinburgh	3
52	V. sativa	130
BM	Natural History Museum, London	65
E	Royal Botanic Garden, Edinburgh	55
HUJ	Hebrew University, Jerusalem	10

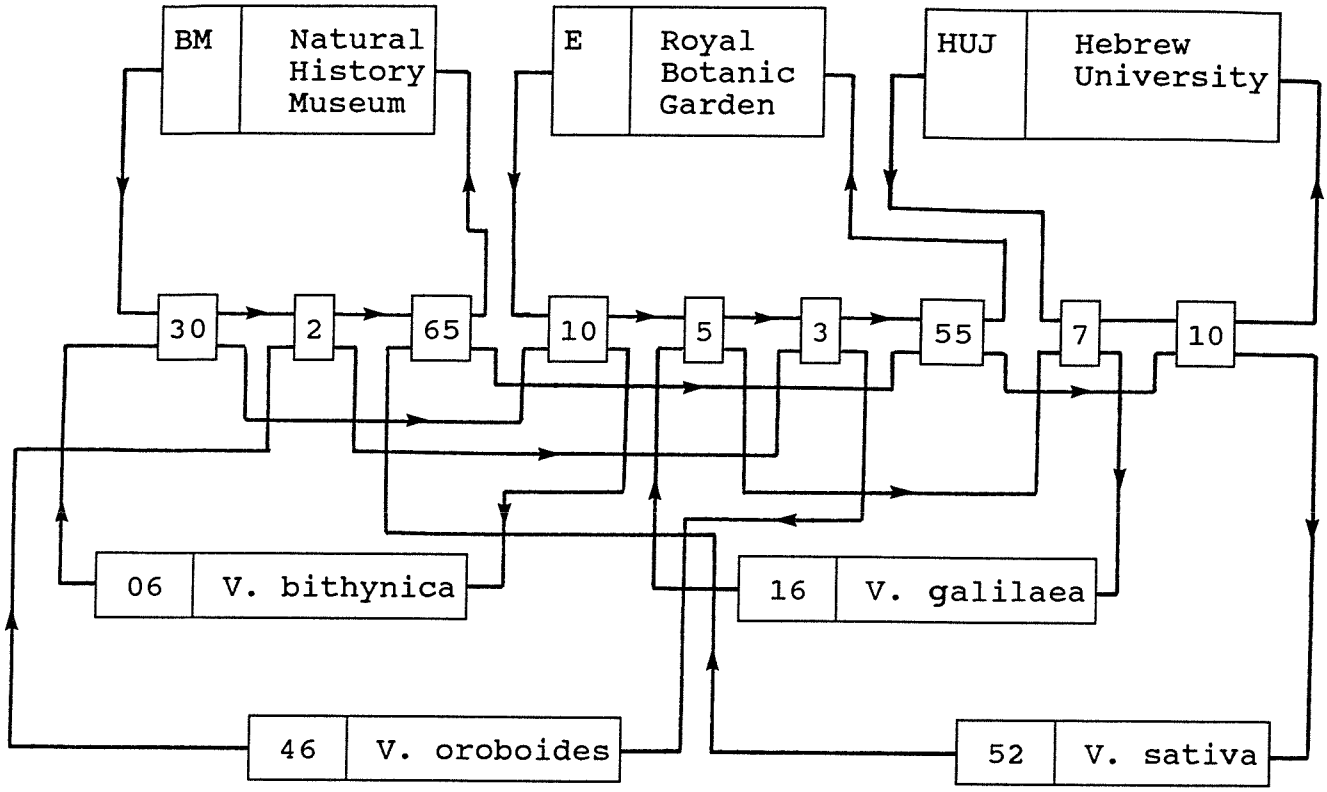
a number of trees, one for each taxon. The dependent records for each taxon represent loans of herbarium specimens with the number of loaned specimens added. Querying this structure can be difficult, e.g. "from which herbaria are V. sativa specimens loaned?". This query requires listing the tree under the taxon V. sativa, but if the loans made from a particular herbarium are queried, the search is much more complicated and involve searching each taxon tree to see if a loan has been obtained for that taxon from that herbarium. This asymmetry in the way the data are retrieved, does not therefore suit data that involves many to many relationships. Here the term relationship infers a table in which multiple data items are inter-connected (Howe, 1985).

The Network Approach

The data are stored in the form of records and links. A general network is one in which any particular occurrence of a record can have any number of links to both superior records and inferior records. The taxonomic database example is illustrated in a network form in Table 2.3.

Both taxa and herbaria records occur, but there is also a new type of record called a connector which links taxa to herbaria. A loan associates a taxon with a herbarium and the connector record stipulates the quantity and involves two fields for chains. One is used to link together on a circular chain all loans from a particular herbarium, the second chain links the connectors for all loans of a particular taxon. The chain from this taxon will take us through a sequence of connectors and from each connector we can follow the other chain which will lead to a herbarium record. The major disadvantage of this approach is that it requires a program to make any enquiry. If the programmer has foreseen the queries, the enquirer will pose these queries which may be answered directly. If the programmer has not foreseen the query, however, then the enquirer must wait for the programmer to

Table 2.3. Example of curatorial data files held using the network approach.



write a program which will produce the answer when applied to the database.

The Relational Approach

Information in a relational database is stored in the form of rectangular tables, in which the rows correspond to records and the columns to fields of the records. The rows of the table link together related values and individual tables can be linked by the use of common items, the coded identifiers. The taxonomic database example is illustrated in a relational form in Table 2.4. Each of these tables is actually a special case of the construct known in mathematics as a relation, where a row is known as a tuple and a column is known as a domain. These files may then be manipulated using three basic operators (Barron, 1984)

> "SELECT extract a horizontal subset from the relation (i.e. select rows).
>
> PROJECT extract a vertical subset from the relation (i.e. delete certain columns) and remove duplicate rows from the subset so obtained.
>
> JOIN join tables where rows have common values in a given domain."

Using these operators the data set may be queried, e.g. "what V. bithynica specimens are on loan from Edinburgh" or "SELECT loan where T # = 06 and H # = E". Barron concludes by under scoring the "beauty" of the relational model, the way in which the users specifies what information they require, not how to extract that information from the database.

Table 2.4. Example of curatorial data files held using the relational approach.

Herbaria relation

H #	name	address
BM	Natural History Museum	London
E	Royal Botanic Garden	Edinburgh
HUJ	Hebrew University	Jerusalem

Taxa relation

T #	name	number of specimens
06	V. bithynica	40
16	V. galilaea	12
46	V. oroboides	5
52	V. sativa	130

Loans relation

L #	T #	number of specimens
BM	06	30
BM	46	2
BM	52	65
E	06	10
E	16	5
E	46	3
E	52	55
HUJ	16	7
HUJ	52	10

2.2 Taxonomic Databases

Taxonomy is data intensive: large quantities of data are related in a complex manner to one another. Data are collated, stored and produced during taxonomic operations (Heywood, 1984). As the quantities of taxonomic data increase, so the need for an efficient means of data management and manipulation becomes imperative. The application of database systems to taxonomic processes has and will increasingly fulfil this requirement (Bisby, 1984b).

Bisby (1984b) distinguishes various types of taxonomic database, based on the four basic types of taxonomic data identified by Leenhouts (1968). The four data types are curatorial, descriptive, bibliographic and nomenclatural. These data types are distinguished by their function and the biological meaning of their contents. It is useful to discuss the content of taxonomic databases using these four data types although, in practice, contemporary taxonomic databases contain complex data structures that are often not restricted to a single data type.

Each database type is illustrated by reference to examples; the list of examples cited is not exhaustive. Further examples are discussed in Sarasan and Neuner (1983) and Allkin and Bisby (1984).

1. Curatorial

Data are stored for the items found in various kinds of collections, e.g museums, herbaria, botanic gardens, germplasm collections and zoos. These databases are now used extensively, both on a large scale for cataloguing institutional collections and also on a smaller scale, where they may be tailored to suit the needs of an individual worker. The information can be queried online or hard copies of the whole, part or an indexed version of the data produced. One of the earliest large curatorial databases was that at the

Botanical Research Institute, Pretoria, South Africa (Morris, 1974). PRECIS (Pretoria Computerised Information System) contains herbarium label information for specimens from the Flora of Southern Africa area. Novel database management software was developed in Pretoria by Morris and co-workers. In 1984 it contained data on 620,000 specimens, with about 20,000 further accessions being added annually (Gibbs Russell and Gonsalves, 1984). The Flora Veracruz Project (Gómez-Pompa et al., 1984) uses a large relational database, which contains data on more than 60,000 specimens collected from the state of Veracruz, Mexico. The International Board for Plant Genetic Resources (IBPGR, 1987) has a large database which contains information on holdings of crops and crop relatives accessions conserved in national and international genebanks.

2. Descriptive

A descriptive database contains taxonomic character data, which help in distinguishing specimens or taxa, e.g. data concerning morphology, anatomy, cytology, chemistry or phytogeography. Phytogeographic data may be thought of as a special kind of descriptive data as the data are stored for taxa throughout the whole or part of their ranges. The information held in the database can be presented as flat files with taxa tabulated against the character scores. In the case of phytogeographic data the information may be presented in the form of a flat file or as a map where presence or absence in particular geographic units is indicated.

Two large-scale projects involving the storage of descriptions are those coordinated by Watson, Dallwitz and co-workers on grass genera (Watson and Dallwitz, 1981; Watson et al., 1986) and Caesalpinioideae Genera (Watson, 1981). Both projects utilise the data coding format DELTA and associated programs; CONFOR for the production of natural language descriptions (Dallwitz, 1980; Dallwitz & Paine, 1986), KEY

which produces keys automatically (Dallwitz, 1984) and TYPESET which is used to produce book quality typeset output of descriptions (Dallwitz, 1984). The DELTA system will be discussed in detail in section 10.2.2 of this thesis. The United Kingdom law requires that all commercial Pisum sativum cultivars are registered to ensure their distinct nature from other cultivars. The descriptive information on approximately 800 UK registered cultivars is held in an EXIR (Brill and Estabrook, 1984) maintained database at East Craigs, Scotland (Winfield and Green, 1984). EXIR is used in conjunction with special purpose software to produce various output products: partial classifications and list of registered cultivars.

3. Bibliographic

A bibliographic database contains details of publications on taxa being studied. Examples of bibliographic databases are: BIOSIS PREVIEWS, an internationally available online database; B-I-T-S (Schultz, 1983) is a subscription search service, which supplies subscribers periodically with floppy discs containing selected references tailored to meet their research interests. There are numerous other large online bibliographic databases, such as CABI Abstracts from the CAB International and AGRICOOLA from the United States Department of Agriculture. Another large scale bibliographic project is the storage of the approximately two million species names and associated bibliographic details held in the Index Kewensis (Lucas, 1984).

4. Nomenclatural

A nomenclatural database includes nomenclatural details, such as authorities, publication and typification details, and an indication of whether a name is accepted, which may be helpful in cross-indexing accepted names to their synonyms. The program ALICE (Winfield et al., 1987), a partial implementation of the BAOBAB project (Allkin & White, 1988),

is currently being used to build a taxon-based nomenclatural (and other data type) database for the International Legume Database and Information Service. The data is stored using the relational database model and involves a highly structured approach (see Hollis, 1988).

The Vicieae Database Project based at Southampton can provide examples of all four types of database files. Due to the location of the research described in this thesis in the Vicieae Database Group at Southampton and the exemplar taxon used being part of the Vicieae, the Vicieae Database Project will be described in more detail than the above examples of taxonomic databases.

The Vicieae Database Project was an experimental monographic database project (see Bisby, 1981; Bisby et al., 1983; Bisby, 1984c; Adey et al., 1984), containing nomenclatural, morphological, chemical and geographical data. The aims of the project were fourfold: to construct a monographic database for the Legume tribe Vicieae; to test whether it could replace the standard printed monograph; to establish an information centre for the Vicieae and to explore the needs and reactions of users to the taxonomic products produced by the Vicieae workers. A Vicieae Project "users group" was established to assess the project products. The user group consisted of a wide group of scientists with a special interest in the Vicieae.

The Vicieae (vetches and peas) were chosen because they were a relatively small group (approx. 327 species), were well known in the temperate region and contained economically important food, forage and ornamental species. The project was divided into two phases: the first involved the construction of the database and during the second phase the database was used for on-line data query, production of pamphlets containing subsets of data and production of tailor-made products for members of the database user group. The

setting up of the Vicieae information centre involved the establishment of several comprehensive facilities whose establishment was stimulated by the Vicieae database, such as the Vicieae herbarium, Vicieae reprint collection and the Vicieae genebank.

The project used largely existing programs (EXIR, CONFOR, GRAPH and PANKEY). This allowed existing programs to be assessed to see whether they were suitable for this type of monographic project or whether a novel taxonomic database management system should be developed. The five database files contain nomenclatural, morphological, geographical, curatorial and chemical information. EXIR holds the main database for direct retrieval and for passing files to other programs. However it was necessary to write two programs (EXIRPOST and SYNONYMS) especially for the project which bridged gaps between the existing programs. EXIRPOST is used to tabulate the products of retrieval or to prepare them for input into external application programs, such as CONFOR for the production of taxon descriptions or SYNONYMS for the production of nomenclatural lists.

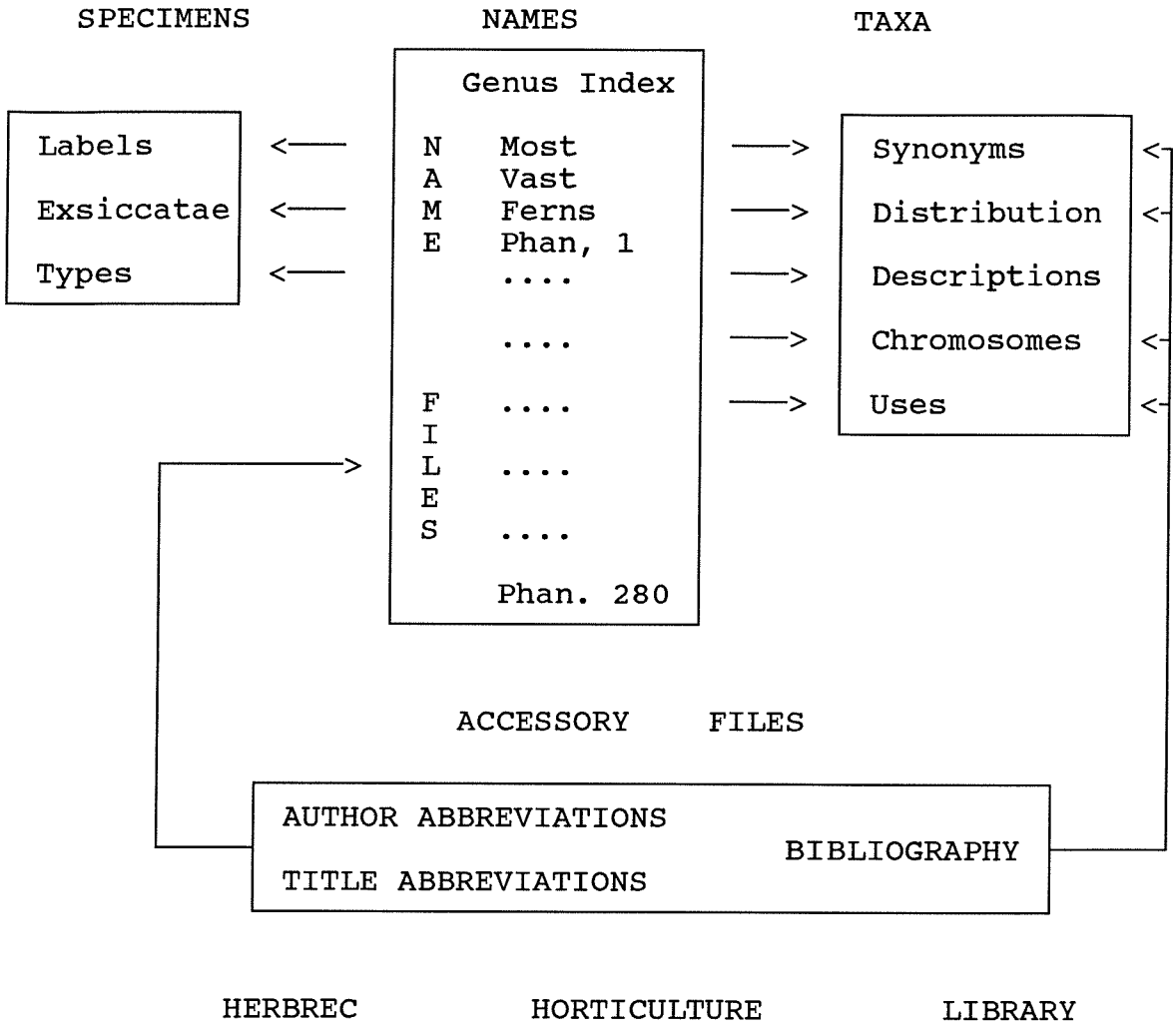
The project products included: a list of accepted Vicieae species names (Allkin et al., 1983a); a synonymised list, with each accepted name associated with its common synonyms (Allkin et al., 1983b, revised for Allkin et al., 1986); distribution of chemicals among species in the Vicieae (Adey et al., 1983a and 1983b); the geographical distribution of Vicia species (Allkin et al., 1983c) Lathyrus species (Allkin et al., 1985) and descriptions of the taxa in Vicia section Faba (Bisby et al., 1987). Many of the requests for information came from outside the study group. Enquiries came from taxonomists, phytochemists, agriculturalists, archaeologists, horticulturalists, physicians and from other disciplines. Tailor-made products were also produced, e.g. cluster analysis of Vicieae species on the basis of chemical data,

morphological descriptions of Portuguese species of Vicieae in Portuguese and keys to subsets of Vicieae species.

The second example of a taxonomic database system, which combines all four types of data, is the TROPICOS system (Crosby & Magill, 1988), based at Missouri Botanical Garden, St. Louis, Missouri, U.S.A. The system's primary function is similar to that of the Vicieae Project, i.e. to provide a tool for taxonomic research, but in this case the bias is towards floristic studies in the Americas rather than a particular monographic coverage. There is another important difference between the Vicieae Project and TROPICOS: the Vicieae Project was taxon based and TROPICOS is name based. A taxon based system is one where the information held in the system is tied to an accepted taxon, where as in a name based system the information is tied to a name that may be either accepted or a synonym.

TROPICOS runs on an ADDS Mentor 5300, using an IBM XT implementation of the PICK operating system. A PC version is also available (called PC-TROPICOS) which runs under revelation. All the TROPICOS software has been specifically developed for TROPICOS by Magill. Figure 2.1 is taken from Magill (1987), where the interrelationships of the various files are fully explained, but here it can be used to indicate the kinds of data held in the system. Crosby & Magill (1988) state that TROPICOS currently maintains hundreds of files, but they can be split into four file types: NAMES; SPECIMEN; TAXA and ACCESSORY FILES. This division of file type is similar to that used in the current project, as discussed in Chapter 10. The content of the first three file types is self-evident from their titles; synonyms, specimen holdings and accepted taxa respectively. The data held in ACCESSORY FILES relates to bibliographic and general herbarium and botanical garden management information.

Figure 2.1 Interrelationships within the TROPICOS database system (Taken from Magill, 1987)



A third example of a database project that combines data of several of the types listed above is the European Documentation System Project, which was a development from Flora Europaea (Heywood and Derrick, 1984). The system was basically an attempt to place the data found in the five volumes of Flora Europaea in a database, so that it could be queried and updated more easily. This information was then to be made more easily accessible to the general public, via online presentation of the PRESTEL type. The system would be menu driven and provide information on taxonomic hierarchy, geography, cytology, conservation status, phytochemistry, bibliography and taxon illustrations.

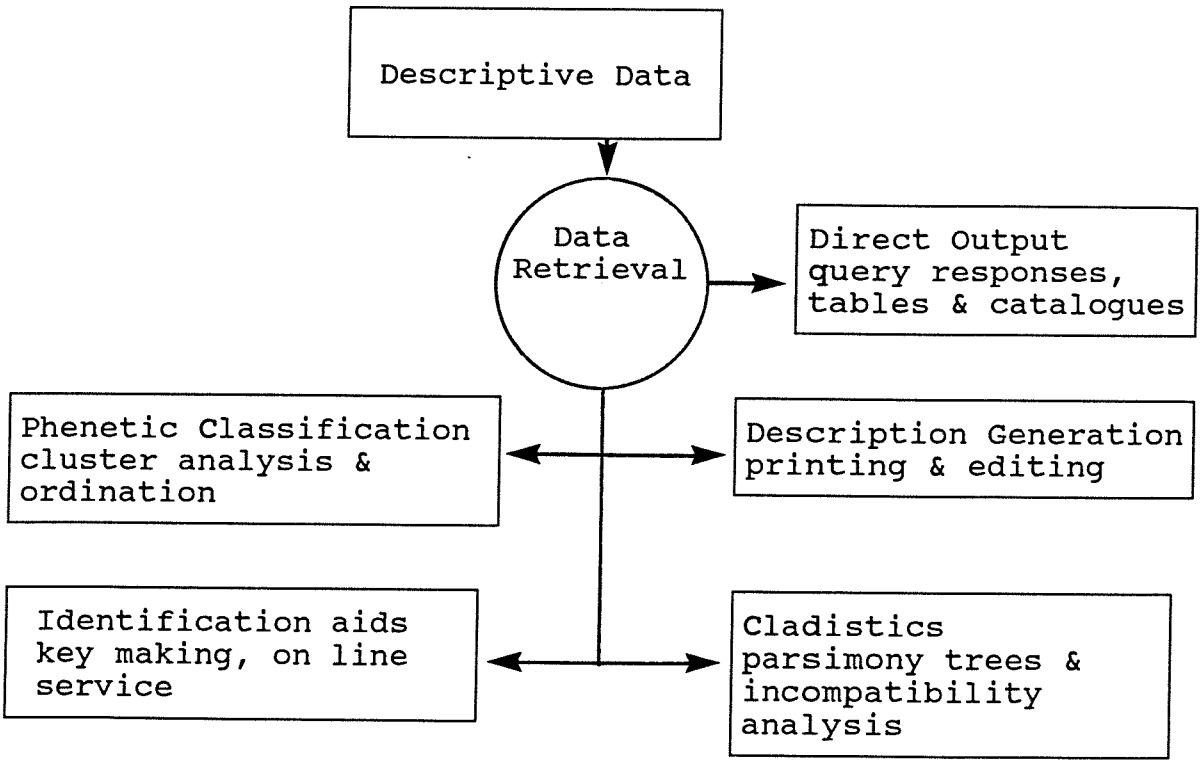
Most of the projects described above are large scale activities involving mainframe computers. This underscores the fact that the initial application of database techniques to taxonomy was at the institutional level via large scale projects e.g. Flora North America, PRECIS and the other examples detailed above. This was largely due to the cost of obtaining the database packages and the cost of the mainframe computers these packages ran on. The advent of relatively cheap micro or personal computers and the widespread availability of suitable inexpensive database packages (notably dBASE) means that taxonomic database usage should now be considered at two levels, the institutional and the personal. Effectively the data stored still divide into the four basic types detailed above, the difference between the institutional and personal databases being one of scale. Institutional taxonomic databases are likely to remain reference sources, whereas the personal taxonomic database will be a more dynamic working tool to aid the individual taxonomist.

2.3 Taxonomic Database Management Systems

The central feature of Bisby's (1984a) "Dream" was the production of an automatic taxonomic information system, where

> "a central computerised taxonomic database can be connected to programs which carry out a range of taxonomic activities such as data retrieval, description writing, identification, phenetic classification and cladistic analysis."

Bisby also provides the following idealised flow diagram to help clarify his vision of how descriptive data would be processed in a taxonomic information system.



Essentially the system has a central codified database, which is linked to an assortment of existing analysis and product generation packages. The use of a codified (as opposed to a textural) database enables the same basic data set to be used for several quite different purposes, as is shown in the flow diagram above. This approach allows greater taxonomic flexibility: easy semi-automatic analysis or product generation of partial subsets of data. An example of the

application of such a system may be a botanical expedition going to collect Vicia in Iran. Basically they wish to know all there is to know about how to identify and where to look for Iranian Vicia. A checklist, keys, description and ecogeographic data could all be generated for Iranian Vicia by making a simple request of the taxonomic database system. Without the use of such a system, tailor-made products could not be produced without extensive work in a good herbarium and taxonomic library.

A background to the development of such taxonomic data management systems is provided by Allkin (1988). He points out that techniques have evolved through four stages from the pre-computer era when traditional taxonomy used paper and card index systems, to the use of sophisticated taxonomic software. In the first phase there was one data set and one use. A data set was generated from scoring specimen attributes and this was then analysed using a single application program (e.g. cluster analysis). The second phase followed in which one data set was used for multiple purposes, such as the production of identification aids, descriptions and the production of similarity coefficients which could then be used in analysis packages. This was followed by the third phase, the application of database management programs to catalogue and organise taxonomic data. Such commercial packages as dBASE or Revelation were used. This phase used databases in the 'raw'. A detailed knowledge of the database management system was required and as will be shown below, little account was taken of the specific problems that were inherent in taxonomic data. The final phase was, in effect, a marriage of phases two and three, in which a taxonomically intelligent database program was used to manage the taxonomic data and link it to specific application packages. The program would enable the multiple use of data sets without the necessity of learning how to run a conventional DBMS and would have a built in understanding of the complex relationships between taxonomic data items.

There have been numerous pieces of software developed by taxonomic database projects to aid the processing of taxonomic data. The majority of the programs produced, however, have been developed as "one-off jobs" that bridge specific gaps between existing software packages and usually sit on top of existing programs. Allkin (1988) points out that the development of reliable, taxonomically useful, software is labour intensive, which possibly explains why comparatively little software has been written to specifically manage taxonomic datasets (Allkin and White, 1988).

One of the first taxonomic database management systems to attempt to meet these goals was TAXIR, developed at the University of Colorado (Estabrook and Brill, 1969; Brill, 1971). Initially seen as a TAXonomic Information storage / Retrieval system, additions were made to the version named EXIR (Abbott, 1976), which was used in the Vicieae Database Project. Further developments to the system which make it more flexible have been discussed by Brill (1983) and Brill and Estabrook (1984).

TAXIR was written in the late 1960's for large mainframe computers but has subsequently undergone many changes and will now run on a variety of machines and operating systems. The system has been used extensively in taxonomic database project worldwide (Brill and Estabrook, 1984). Data are entered into the database in the form of a table of items (flat file) by item descriptors, such as the following example taken from Estabrook and Brill, (1984):

	Family	Genus	Species	Habitat
Plant 1	25	392	106	11
Plant 2	31	112	82	4
Plant 3	98	45	16	6
Plant 4	16	309	30	8

The database can then be queried via use of a list of descriptors and a boolean expression. A simple query would have the form, "desc relop ds", where: desc is a descriptor name; relop is one of the comparison operators =, =, >, >=, <, =<; and ds is a descriptor-state name of descriptor, desc. Some examples of queries might be: GENUS = 309; SPECIES < 82; NOT PLANT = 4 and (FAMILY < 31 or HABITAT > 4). The query would then be answered by printing out a list of the requested descriptors of the items which satisfied these conditions. TAXIR/EXIR is limited by the necessity to use flat files but it is this flat file structure that the algorithm uses so efficiently to extract subsets of data quickly. The structure, however, is not well suited to data that cannot be fitted into flat files. Brill and Estabrook (1984) point out that database files are often not flat and thus they suggest various ways in which non-flat data can be flattened. The TAXIR/EXIR based database management systems are now largely being replaced by the more generally available DBMS. The reason for this is the greater flexibility of the DBMS in use, even if they become less easily manageable as dataset size increases. These general DBMS'S, as will be discussed below, will in turn be replaced by specialised taxonomic DBMS's (such as BAOBAB).

One of the first attempts to build a more comprehensive (in the taxonomic and computing sense) DBMS was attempted by the Flora North America Program (Shetler, 1975). They endeavoured to build a generalized descriptive database with associated software to produce various tailored taxonomic products, classification, descriptions and identification

aids. Unfortunately, one of the project's central conclusions (Shetler, 1974) was that the building of such a system, was much more time consuming and expensive than had initially been envisaged.

Pankhurst (1983), when discussing the potential of using DBMS's to construct floristic databases, suggests that they could contain all normal forms of taxonomic data. The database could then be queried to provide: a) specimen identifications via diagnostic keys, b) comparisons of included taxa and c) retrieval of more general information about taxa. He points out that the problem with this is the way in which the data are structured, stored and retrieved for diverse taxa. Traditionally, taxonomy has not placed emphasis on using comparable character sets. A brief examination of any standard flora will show that characters used in descriptions are not consistent even for closely related species. Pankhurst outlines five basic difficulties which may be encountered when trying to produce a kingdom-wide floristic database:

1. How many characters are required? A complete set would be impossible to collect and too few would not rigorously distinguish the species.

2. Which terminology to use to describe characters? Lawrence (1951) is the standard text but within this there are inconsistencies, incompleteness and incomparabilities.

3. How much redundant information to include? He provides the example of leaf shape where some redundancy may be desirable in describing a complex shape which is very difficult to describe accurately in words.

4. The characters must be homologous, similar organs must be described in the same way.

5. The descriptions must allow for intra-taxon polymorphism, where there is variation between individual specimens of the same taxon.

He does not provide solutions to these problems but does detail a pilot project, an attempt to construct a floristic database for the British Flora. He took one taxon from each of the fifty British plant orders and recorded eleven characters for each, the data being included in a database. He concluded from this experiment that it was possible to create general character sets for use in floristic databases of this kind. Although creating a character set of this kind is problematic, once established, the addition of new taxa, characters and states and the extraction of special subset character sets would be straightforward. The problems of building such a database should not be under-estimated. Pankhurst is optimistic and it seems likely that a new set of problems would become apparent if a more detailed project were undertaken.

In a later paper Pankhurst (1988d) discusses the design of a "completely" relational taxonomic database structure. He illustrates his design with the example of Rubus (Rosaceae), which he was writing for the Flora Meso-Americana project. Using dBASE III, as it was then the most widely available DBMS, he constructs a database with twelve linked files in a fully relational structure. The database is associated with curatorial, bibliographic, determination and nomenclatural data capture programs, but as yet no morphological data capture program. The database is small (in comparison with the one described in this thesis), containing information relating to 100 specimens, 50 taxa and 40 bibliographic

entities, but does illustrate some of the design intricacies associated with the complex problem of constructing a taxonomic database. He discusses how a database of this kind could be linked to a file containing descriptive data for specimens or taxa using the DELTA based sets of programs (Pankhurst, 1986; Dallwitz & Paine, 1986), but concludes that morphological data is generally too complex to be dealt with in a simple DBMS such as dBASE III. The conclusion of this pilot database project was:

> "that even if databases are a convenient means to begin a Flora project, and the analysis of DELTA files an effective way to end it, there is still a gap between them which needs to be bridged."

Similar conclusions became evident during the Vicieae Database Project. Firstly, that existing methods of managing taxonomic data and the application of taxonomic database management systems were inadequate for general taxonomic use. Secondly that, "there is a need for an improved system which takes into account the needs of taxonomy" (Adey et al., 1984). This type of comment had been underlined and repeated since the 1950's, see Just, 1954; Estabrook and Brill, 1969; Crovello and MacDonald, 1970; Watson, 1971; Kedler and Crovello, 1973, and Brenan et al., 1975.

Stimulated by the challenge of developing such a system, Allkin & White (1988) are designing and attempting to implement Baobab, a novel taxonomic database management system that will attempt to meet this need. The basic design features of Baobab were discussed by Allkin and White (1982). Later Allkin (1988) lists three specific requirements of taxonomic data management software:

> "1. provision of communication links between application programs;
>
> 2. assistance in data file management;
>
> 3. provision of a biological interface to make data capture, revision, and retrieval easier and less error prone."

Allkin and White (1988) expand on these points, reviewing what is required by the taxonomic community and what Baobab will be able to provide:

> "It will enable the professional taxonomist to score, record, retrieve, edit, reorganise and summarise their data and select subsets for subsequent use in classificatory analysis, in the construction of keys or to respond to enquiries. The Baobab design consists of a suite of interlocking data structures following the relational approach to database design, algorithms to manipulate these structures and a taxonomic interface allowing the user to deal in familiar taxonomic concepts without knowledge of the internal structure."

When fully implemented it will provide a taxonomic tool of immense value in aiding taxonomists to collate, manage and develop products from the complex data sets they deal with. Baobab will in effect be a generalised version of the specific taxonomic database management system presented in this thesis.

Baobab will comprise three basic components:

> "(i) structures to support a general descriptive taxonomic database
>
> (ii) algorithms to interpret those structures and model taxonomic practices
>
> (iii) an interface to shield taxonomists from the complexity of the internal structures and provide a familiar environment."

Allkin (1988) adds a further point, that the taxonomic database management program must replace the interpretative role of taxonomists using conventional data tables. This means the program must be taxonomically intelligent and must appreciate as many of the complexities of taxonomic data as is possible.

Allkin and White (1988), discussing their ideas for the design and implementation of Baobab, consider some of the

problems of taxonomic description storage, such as character dependence, character variability and the need to incorporate the taxonomic hierarchy.

The storage of descriptive taxonomic data is a much more complex problem than the storage of other types of taxonomic data due to, for example, dependencies among characters and hierarchical relationships between taxa (see Watson, 1970; Pankhurst, 1975c; Allkin, 1980; Allkin 1984; Bisby 1984a; Allkin and White 1988). This is an important point and will be illustrated by an example taken from Allkin (1988); see Table 2.5.

Table 2.5. An example descriptive taxonomic data set. (Taken from Allkin, 1988) ? = Unknown or missing observation, X = Logically inapplicable observation.

Taxon	Flower colour	Corolla length (mm)	Leaf	Leaflet pubescence
V. cracca	blue	8-12	present	?
V. sativa				
subsp. *sativa*	purple	18-30	present	present
subsp. *nigra*	purple	10-19	present	present
L. aphaca	yellow	6-18	absent	X

Allkin points out that once *Lathyrus aphaca* is described as "having no leaflets", it becomes illogical to ask the question, "are hairs present on the leaflet?" Only realisation of this logical relationship between the two characters, leaflet presence and leaflet pubescence, stops data being erroneously entered on leaflet pubescence for *L. aphaca*. Similarly if we ask the question, "what is the corolla length of *Vicia sativa*?", the answer is 10-30mm. To answer this question the taxonomic hierarchy must be understood, so this question involves both the data for *Vicia sativa* subsp. *sativa* and *V. sativa* subsp. *nigra*.

Another problem of structuring data of this type is the effect on dependent characters if the character they depend upon is deleted. For instance, in the example character set discussed above, if leaf presence / absence is deleted, how will the character leaflet pubescence be affected? or if the character is multistate, what are the logical implications if the character states are redefined? All these complexities of taxonomic data present very complex problems for the programmer. There are ways of overcoming these problems such as not permitting dependent characters, but clearly these are serious problems and the long-term consequences of their solutions must be thought through thoroughly before implementation.

One of the other central questions which must be addressed to those setting up a database is whether to use a textual or codified system. Lightowlers (1988) recently described the Antarctic Plant Database system, which contains curatorial and descriptive information on about 40,500 specimens. This database has been converted from a codified to a text based system using STATUS, a sophisticated text retrieval system. The reasons for this conversion were not made explicit, but it does pose the question of how they have overcome the problems associated with the text based database; such as indexing on text strings, location of data substrings or data size problems that might arise. Allkin & Bisby (1988) comment that:

> "Free text databases offer few advantages over conventional texts stored on word processor, while structured database design ensures greater consistency, permits automatic checks of data integrity, and provides for flexible indexing and reliable retrieval."

Some software does allow free-text searching, but they consider this to be unreliable for the following reasons:

"1) two phrases with exactly the same meaning may have quite a different syntax. Compare, for example, our search pattern "leaves lanceolate" with "lanceolate leaves" or "leaves long and lanceolate",

2) taxonomic variability results in descriptions with more complex syntax. The description "leaves elliptical or lanceolate" or "elliptical to lanceolate", for example, would not necessarily match our simple search pattern despite indicating that at least some plants have lanceolate leaves;

3) authors may use alternative synonymous terms within different descriptions; compare, for example, "lanceolate leaves" with "lancifolius"."

Even though there are some sophisticated free text search programs, it would still appear that a codified database provides a more flexible solution to the problem of data storage.

It is envisaged that Baobab will be a codified database with a user interface which allows the taxonomist to employ familiar (taxonomic) terminology, to refer to taxa by name or obtain data for a particular taxonomic level (Allkin and White, 1988). This interface will mean the user is shielded from the algorithm and is presented with a familiar user-friendly package. Allkin (1988) makes the point that as well as data files the well designed database must also contain data about data or 'meta-data', which means files will also exist that are largely composed of codes which enable the correct interpretation of the data. He goes on to add that these meta-data files can far outnumber the raw data files in a complex relational system.

The software will be written, at least partially, in the high level language C, with calls to a general database management package. C was chosen because of the requirement for programming flexibility. ALICE (Winfield et al., 1987) a program that is intended to aid biologists building checklists or species diversity databases has been developed and different file structures explored for the management of nomenclatural, bibliographic, geographic distribution and

simple factual data. ALICE is taxonomically shrewd. It does not allow logical inconsistencies such as species recorded from India being absent in Asia. ALICE is taxon based and so data can be retrieved from the system using a synonym. It runs on IBM PC-compatible microcomputers (such as XT/AT, etc.), under the PC-DOS or MS-DOS operating systems with a hard disc and 512K RAM. It can be used to produce taxon checklists, to store ethnobotanical data and is going to be used by IUCN for conservation catalogues prepared at remote sites.

ALICE is currently being used for phase one of the ILDIS (International Legume Database and Information Service) project (Bisby, 1987, 1989a), which involves the building of a worldwide checklist for legume species. Once completed it is envisaged that the checklist will be distributed to the scientific and non-scientific public via on-line services, floppy discs, enquiries to the ILDIS coordinating centre and via publications. Phase two of the project will mean the creation and distribution of more specialised legume data sets, with information concerning legume biochemistry, legume germplasm holdings, detailed descriptions of legume taxa, etcetera. For phase one the information is uniformly structured using the relational database model and contains data on:

1)	Accepted names	*
2)	Principal synonyms	*
3)	Vernacular names	*
4)	Legume tribe membership	
5)	Habitat within continent	*
6)	Geographical distribution	*
7)	Life form	*
8)	Conservation status	*
9)	Importance to man	*
10)	Notes	*
11)	Reference to a description	
12)	Reference to an illustration	
13)	Reference to maps	

Fields marked with asterisk can be linked to references.

Currently data for all legume species in Africa has been entered to the system (Lock, 1989) and data for the Indian sub-continent is being entered. Information on legumes from other geographical regions has been stored using other storage programs and this data, for North America (TROPICOS) and Europe (ESFEDS) has been converted to ALICE format and merged with the data already held in the ILDIS phase 1 database. ALICE is also being used to produce a checklist of orchids of Vanuatu (B. Lewis, P. Cribb & R. Allkin, pers. comm.), CITES world Cactaceae database (Kew/CITES), Checklist of Araracuara, Colombia (Scientific Reserve Amazonia) and Zoological invertebrates (Smithsonian Institute, Washington, U.S.A.).

ALICE is, however, not Baobab. ALICE aims to satisfy the requirements of one type of systematic data. Baobab requires other 'limbs' such as a specimen management program (Allkin, 1988), a descriptive data management program and a data synthesis module, which can take information about specimens or species and draw conclusions about species or genera before the 'Baobab tree' reaches maturity.

CHAPTER THREE

THE TAXONOMIC REVISION

> "By a monograph we understand a complete account of any one family, tribe, or genus, nothing being neglected which is necessary for a perfect knowledge of it."
>
> A.P. de Candolle

3.1 What is a Taxonomic Revision ?

De Candolle (1821) in the above definition of a monograph, illustrates both the breadth and the lack of clarity of taxon-based taxonomic research. He fails to differentiate between a revision and a monograph. It would seem logical to expect that something as central to taxonomy as the revision would be clearly defined and be accompanied by an explicit and established methodology, however this is not the case.

Attempts to describe an explicit revision methodology are almost exclusively limited to general taxonomic textbooks, where definitions and objectives are required for taxonomic work to be discussed in any detail. It is clear from the literature (Lawrence, 1951; Benson, 1962; Davis and Heywood, 1963; Leenhouts, 1968; Radford, 1986; Jones and Luchsinger, 1987; Stuessy, 1990) that taxon-based herbarium research can be discussed under two headings, monographs and revisions. The difference between these activities is essentially one of scale and taxonomic focus; a monograph is a synthesis of all available taxonomic information relating to a particular plant group, whereas a revision is more limited and is largely confined to a synthesis of morphological and geographic data for a plant group from a particular area. The precise difference between a monograph and a revision is of limited semantic importance and in practice the one grades into the other.

Current fiscal limits mean that time allowed for taxon-based work is increasingly restricted and this, combined with

the exponential expansion of possible sources of taxonomic data, means that very few taxonomists are allowed the luxury of undertaking a monograph. As the revision grades into the monograph, the following discussion of the revision paradigm is equally applicable for a monograph.

A formal definition of a revision is provided by Deborah Qualls in Radford (1986),

> "A revision is a treatment of selected taxa throughout at least a major portion of their range, including a study of nomenclature and classification along with descriptions based on several types of evidence."

This definition is inelegant and vague. It also poses the question: what is a 'treatment'? This is though the sole attempt at a clear, succinct definition of a revision.

Davis and Heywood (1963) define the objective of a revision,

> "is to delimit the taxa clearly (particularly species), to group them in a natural manner and provide a means of identification."

The essential products of a such a revision would be: a classification of the circumscribed taxa, complete specimen citation (including types), taxon descriptions, a key, synonymised nomenclatural lists, distribution maps, ecological descriptions and critical notes to the circumscribed taxa.

The revision that is undertaken is influenced by the scope and purpose of the work and so may be referred to as complete or partial. A complete revision is one where the taxa studied are investigated throughout their range, whereas a partial revision is associated with work for a Flora of a political or geographic area and is thus more restricted. The extent of previous work on the group, the facilities, the material and the time available, will each influence the

manner in which the revision is carried out and affect the final products of the revision.

Although data synthesis for a revision is commonly based on morphological and geographic evidence alone, it may also include data from literature reviews, field and garden studies and cytological, anatomical, phytochemical, palaeobotanical, ecological and palynological studies. The value of the final products of a revision will depend on the amount of original research and the thoroughness of the data compilation and analysis under taken.

Based upon the above discussion the following working definition of a revision is proposed:

> A revision is a novel analysis of the variation pattern within a particular taxon, considered in conjunction with information from the literature which results in the generation of primary and secondary revision products. The primary product is a novel classification of the taxon, which is complemented by a range of secondary products, such as descriptions, keys, synonymised lists, taxon illustrations, critical notes, etc.

3.2 How to Undertake a Taxonomic Revision

Leenhouts (1968) comments on the clarification of a revision methodology thus:

> "Whereas it is simple to give generally applicable rules for the documentation of literature and specimens, it is difficult if not impossible to compose them for the work itself."

Though it is true that trying to establish a comprehensive revision methodology is difficult, taxonomists should persevere, as it will result in greater clarity of purpose, i.e. note the continuing attempts at a species definition.

Davis and Heywood (1963) make the first detailed, systematic attempt to break down the steps involved in the revision process, listing the following nine empirical steps:

"1. Assess the scope of the taxonomic problem by consulting the literature. We need to know if a previous revision exists, and to have some idea of the number of species involved and the distributional range of the genus. For monographic work, a card index of all species listed in the Index Kewensis is usually compiled at the beginning of the investigation. Herbarium material must be assembled, preferably from as many herbaria as possible.

2. Study the specimens and select characters which show discontinuities within the whole group. A picture of the variation is more easily obtained if the character expressions are tabulated for individual plants or herbarium sheets; these may be referred to at later stages in the revision and in the preparation of keys, descriptions and graphical methods of presentation.

3. Group individuals together according to the highest correlation (maximum co-variation) of characters which have been selected and which show more or less continuous variation between these individuals. These groups will represent the major "species", and often possess a characteristic facies. Graphic methods to portray variation can be used at this stage (or earlier) to check taxonomic conclusions.

4. Seek discontinuities which may separate the resultant groups from one another; that is, the diagnostic characters are established.

5. Next study variation in those differential characters that show incomplete separation between the groups, and any other patterns of character variation that occur within the groups. This may disclose further discontinuities within the groups first established under (3).

6. Map the geographical distribution of the groups, at the same time seeking ecological regularities of variation within the groups; these may reveal infra-specific patterns that deserve subspecific or varietal recognition. The taxonomist should decide on his specific and infra-specific criteria within the group and apply them as consistently as he can.

7. The species that have been recognised by the above processes can then be grouped on the basis of their similarities, so that natural series, sections, etc. can be built up. A natural classification of the genus should thus be achieved. In some genera it may prove impossible to recognise satisfactory sections, the best that can be achieved being groups of closely allied species which may constitute species aggregates or series. In many cases it is possible to recognise major groupings at an early stage of the revision (corresponding to sections, subsections, etc.) which can be broken down into species at a later stage.

8. For the purpose of communication, names must be attached to the groups in accordance with the International Code of Botanical Nomenclature, synonyms listed, keys and descriptions drawn up, etc. Naming involves the scrutiny of original descriptions and, often, examination of type material. Additional methods of presentation (scatter diagrams, maps, etc.) may be used to describe the patterns of variation.

9. The taxonomic characters can be arranged (or considered) in series. The interpretation of these in terms of evolutionary trends may be attempted with some degree of probability, but very seldom with certainty. The same applies, but with even greater force, to the phylogeny of the group. This can only be discussed in terms of possibilities."

To these nine points they add a further nine, they suggest should be born in mind when undertaking a revision.

A. The revision should avoid an over-reliance on herbarium specimens, material should also be grown under uniform garden conditions and field observation made. This is especially important if infra-specific taxa are to be recognised, as they are often distinguished on the basis of characters which are difficult to detect in dried material. They warn against the recognition of single character variants that have no population basis.

B. The assumption should not be made that the specimens held in herbaria are necessarily representative of taxon variability throughout the range of a particular taxon.

C. The growth of material will provide extra characters, but these characters will not necessarily reflect the entire taxon variation pattern.

D. It is important to collect geographic and ecological data from the herbarium labels during the course of the revision.

E. Care should be taken when scoring herbarium material to compare like with like. Meaningful taxonomic information cannot be obtained from the leaf dimensions of lush woodland plants when compared to plants growing on the edge of the desert.

F. Specimens which show intermediacy between established taxa can provide important data about the established taxa. The distribution pattern of the intermediate material can indicate whether there is a cline between the two taxa or whether the two taxa hybridise where their distributions abut.

G. Taxonomic validity can only be given to the identification of taxa, if the original description and type specimens are consulted. It is also a general rule that the more material seen during the revision, the more variation will be recorded and the more accurately taxa will be described.

H. Care should be taken in attributing unidentified material to a new name because a name may already exist that encompasses the material:

   i. the plant may already be described and reduced to a synonym and so require reinstatement,

   ii. the plant may be described as a generic synonym,

   iii. the plant may have been described recently and not yet included in Index Kewensis,

iv. the material may be of a taxon from another country where its presence has not yet been recorded,

v. the material may represent an extreme genetic variant of an existing taxon,

vi. the material may have been growing under environmental stress, which may have resulted in the development of aberrant forms,

vii. the material may be of hybrid origin, from two previously existing taxa.

I. A revision should take into account all existing names associated with a taxon, should make decisions on the names taxonomic status or should say if their status remains uncertain.

These points made by Davis and Heywood (1963) form the clearest attempt yet to clarify the methodology of a taxonomic revision. Although these points do summarise the revision process, they fail to acknowledge the routine part played in contemporary revisions by various computer aided techniques, e.g. numerical phenetic or phylogenetic analysis and automatic product generation programs (as discussed in later thesis sections).

Subsequent authors (Leenhouts, 1968; Qualls in Radford, 1986) have attempted to summarise the operations of a revision, but their attempts proved less precise and succinct. Leenhouts provides a detailed guide to the practice of herbarium taxonomy, which is comprehensive, but lacks any clear, concise summary of how a revision is undertaken. Interestingly, both Davis and Heywood and most markedly Leenhouts suggest the extensive use of index cards to catalogue taxonomic data. Index cards in general have been superseded by the application of computer database technology, which in its simplest form is composed of multi-indexed data tables.

3.3 A Clarification of the Revision Paradigm

The efficient application of a DBMS to the revision process requires that the revision paradigm is fully understood. The lack of such an explicit paradigm necessitated the clarification of the revision process. Once clarified an attempt could then be made to model the paradigm using the DBMS. Bisby (1984a) makes the point that:

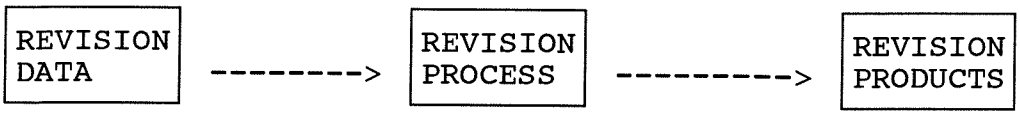
> "it is widely acknowledged in industry that a DBMS works for an organisation only as well as it succeeds in modelling that organisation's flow of information."

While Allkin (1988) also stresses a similar point, that:

> "taxonomists subconsciously use their knowledge of the nature of taxonomic data and the rules governing taxonomic procedures to interpret and manage their data effectively. If taxonomic programs are to be equally effective that knowledge must be made explicit and incorporated within them."

For these reasons it is important that a precise understanding of the taxonomic revision process is obtained. In practice this is difficult as each revision is unique; unique in this case meaning that the published results of each revision contain a subset of the total potential revision products. However, it is possible to produce a general summary of the research process and the products that commonly accrue from a revision.

The primary objective of a revision is to take an established taxonomic hierarchy and based upon novel data analysis synthesise a novel ("revised") taxonomic hierarchy. Thus during the course of a revision a fundamental alteration of the taxonomic hierarchy occurs between the initial and the novel hierarchy. The novel hierarchy should be a closer approximation of the intrinsic "natural" hierarchy or classification of the plant group. The actual process of undertaking a revision can be summarised in the following model:

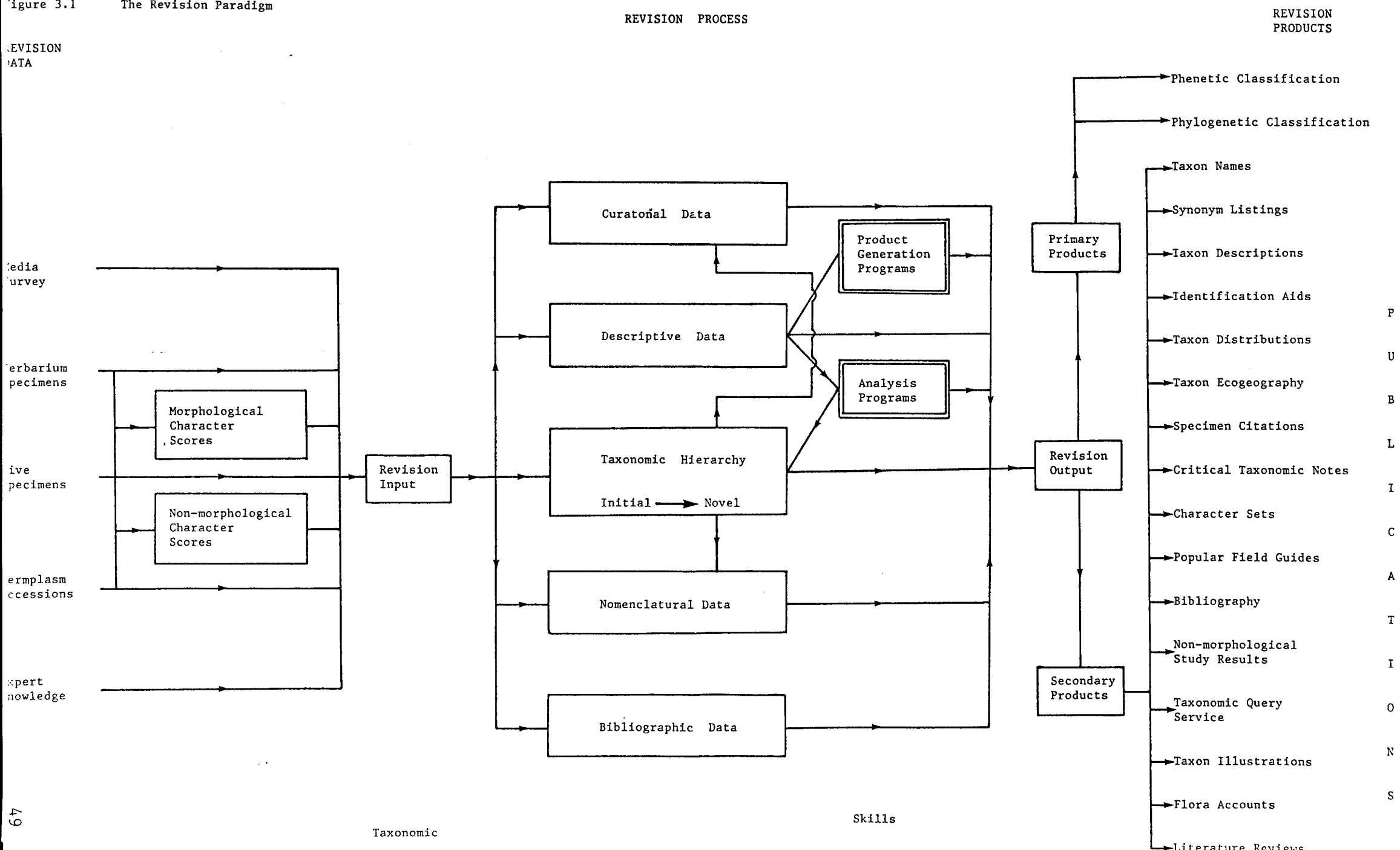


This simplistic model can be seen as a variation on the "diffusion model" proposed by Bisby (1984b, p. 20) for the taxonomic profession as a whole. With the expansion of this simple model the explicit model for the revision paradigm is produced, as shown in Figure 3.1. The three components of the simple model are evident, but the internal structures of the components have been elaborated to help clarify the model.

The data that the revision will work upon enters the revision from five basic sources. A media survey is an essential element of the revision, as it provides a taxonomic history for the taxon and thus provides the starting point from which the revision can progress; the starting taxonomic hierarchy. The media survey will also provide revision data from sources not included within the revision study. An example of the latter is a large proportion of the chromosome count data included in Chapter Eight, which was simply abstracted from the literature for subg. Vicia. Another example is the nomenclatural information held in the Vicieae Project Database and published by Allkin et al. (1986), which forms the basis for the TAXNOME1 database file (discussed in Chapter Ten).

An important point from the latter example is that the review of preceding knowledge can be seen as more than a literature review where the data is solely obtained from books, pamphlets or papers. In fact, the search for existing taxonomic knowledge should be seen as a "media review", where taxonomic databases, television programs and other forms of contemporary media are searched as well as the more traditional sources, such as taxonomic journals (e.g. Kew Bulletin, Taxon), name indices (e.g. Index Kewensis, Gray

Figure 3.1 The Revision Paradigm



Index) or books (e.g. Advances In Legume Systematics, Polhill and Raven, 1981).

It is important to stress that representative herbarium specimens should be used in the revision: representative in two senses, firstly representative of the inherent variation pattern for that taxon and secondly, representative in the scope of geographical coverage for that taxon. The larger the sample of specimens the better will be the estimate of the variation pattern for the taxon. Geographical coverage can be secured by obtaining specimens from a broad selection of international herbaria.

Individual herbarium label data should be recorded from the specimens for use in producing generalising information about taxon geography and ecology. Scoring morphological characters from herbarium specimens limits the character set that can be used and it is difficult to score important characters such as flower colour from herbarium material. For this reason, where specimen cultivation or field study is possible it is advisable to score characters that are obscured on dried herbarium material. Field study will also enable the recording of more detailed ecogeographic data, such as pollination vectors, soil preferences or biotic interactions which may be difficult to obtain from herbarium label data.

In practice, a contemporary revision will often be based on a study of morphological characteristics, but where possible other sources of data should be used to augment the morphological data. These other sources of data (e.g. anatomy, cytology, phytochemistry, etc) usually require vouchered germplasm to be used either directly as experimental material or for growing plants which are in turn used as experimental material.

The final source of input to the revision is expert knowledge; this may be difficult to define, but it cannot be

ignored. Before commencing a revision, taxonomists will discuss the project and problems associated with the taxon with colleagues, project supervisors or international experts in the particular group of plants to be revised.

Having established the sources of revision data or input, the revision process can essentially be seen as the analysis and synthesis of the revision input. This data processing will result in the conversion of the initial or starting hierarchy to the novel taxonomic hierarchy. As pointed out by Leenhouts (1968) the data that is gathered for a revision can be divided into four types: curatorial, descriptive, nomenclatural and bibliographic. These four data types can be defined as follows:

Curatorial - information that can be used to manage and identify particular accessions,

Descriptive - information which can be used to distinguish taxa,

Nomenclatural - information concerning accepted or synonymised taxon names,

Bibliographic - information about the location of literature references to taxonomic data.

Essentially "active" data (information that can be used to relate taxa) is confined to the descriptive data. It is the analysis of these data, via either phenetic or phylogenetic analysis, which will be used to identify character correlation between Operational Taxonomic Unit's (e.g. specimens, species, genera). Based on these character correlation OTU's will be clustered by the analysis program. The number and type of correlations will help to ascribe taxonomic rank to the

clusters of OTU's. The results of the descriptive data analysis are used in the conversion of the taxonomic hierarchy from the initial to the novel form. The hierarchy taken from the literature at the commencement of the revision is superseded by the hierarchy that is drawn from the revision results.

The novel form of the taxonomic hierarchy is the primary product of the revision. The novel classification is central to the revision, because once established it will directly affect the production of the secondary products, such as taxon descriptions, identification aids, accepted nomenclature, etc. Having established the hierarchy, descriptive data from the representatives of taxa can be used to produce some of the secondary revision products via product generation programs, such as PANKEY (Pankhurst, 1986), CONFOR (Dallwitz, 1979) and EXPERT KEY (Atkinson and Gammerman, 1987).

The primary and secondary revision products are listed on the right side of the model. The primary product, the classification, will be produced by each revision, but the secondary products listed are those potentially available. As mentioned above, the secondary products of individual revisions will usually be a subset of the potential products. The products are used by two groups of users; internal products, for use within taxonomy by other taxonomists and external products, for use by general biologists and amateurs. To make these products available to both user communities the taxonomist must publish the products of the revision.

Finally, it must be stressed that the whole revision paradigm is mediated by taxonomic skill. The taxonomist intercedes making skilled taxonomic judgements throughout the various stages of the revision, e.g. deciding which data are to be collected, interpreting the intrinsic patterns in the data analysis and supervising the synthesis of the various revision products.

SECTION

TWO

EXEMPLAR REVISION

CHAPTER FOUR
TAXONOMIC INTRODUCTION

4.1 Taxonomic Aims

The general aim of the taxonomic element of this project was to undertake a revision of Vicia L. subgenus Vicia (Vicieae, Leguminosae). To achieve this aim, both novel morphological data and non-morphological evidence available from the literature were used. This information was collated, analysed and the subsequent patterns of relatedness were used to produce a revised classification of subgenus Vicia.

More precisely the revision aimed at answering questions concerned with the sectional, specific and supra-specific relatedness and status of the included taxa.

These questions are made explicit in the following:

1. Sectional - to test the sectional classification of subgenus Vicia proposed by Kupicha (1976), to learn whether the species circumscribed as being members of her subgenus Vicia do form a natural grouping and whether the taxa included group naturally into five sections as proposed by Kupicha. Several authors (Medikus, 1787; Fedtschenko, 1948; Stankevich, 1982) believe that V. faba should be split into a separate higher level taxon from Vicia, either subgeneric or generic level. The revision should resolve this question.

2. Species - to examine the relationships between the species included in the sections of the subgenus.

    a. To assess the relationship between V. faba, the four V. narbonensis complex species and V. bithynica. These six species are all included in section Faba (sensu Kupicha, 1976), but, as mentioned above,

there are doubts about their inclusion in the same subgenus, or even in the same section.

b. To investigate the relationship between V. sepium, V. oroboides and the other two species from sect. Atossa. Geographically V. sepium is distributed throughout Europe and Western Asia, while the other species are endemic to the North-east Mediterranean Basin. V. oroboides is also an 'oroboid' species and, as such, it is linked with the other 'oroboid' species of subgenus Vicilla. This allegiance would distance it from the other three species, which are not considered close 'oroboid' relatives.

c. To investigate the relationship between the perennial V. pyrenaica and the five other annual species of section Vicia.

d. To assess the validity of using lens position as the diagnostic feature of section Hypechusa. Historically this has been considered an important character (Alefeld, 1860a) for defining this group, but is the grouping of these twelve species of section Hypechusa (sensu Kupicha, 1976) substantiated by other correlated characters?

e. To resolve the problem concerning the natural placement of V. mollis. Kupicha places this species in sect. Peregrinae, but the presence of a short peduncle and lens placed opposite the hilum indicates a closer relationship than she suggests with sect. Hypechusa species.

f. To investigate the relationship between the five species not known to Kupicha (1976) and the 32 species she included in Vicia subg. Vicia.

3. Supra-specific - to examine the relationships between the supra-specific taxa included in the species of the subgenus.

   a. To investigate the supra-specific variation found in the pan-temperate V. sativa. This species has a broad range of variation being both phenotypically and genotypically plastic. This high level of plasticity has led to the growth of a copious synonymy which in turn has resulted in no supra-specific classifications being generally accepted.

   b. To investigate the acceptability of the supra-specific taxonomic status used by Allkin et al. (1983b).

   c. To investigate the relationship between the two supra-specific taxa (V. sepium var. eriocalyx and V. sativa subsp. devia) unknown to Allkin et al. (1983b) and the species, they are suggested as belonging to, within Vicia subg. Vicia.

4.2 Delimitation of the Exemplar Taxon

Kupicha (1976) comments that historically Vicia species have been grouped into three or four major species clusters: 'Cracca', 'Vicia', 'Ervum' and sometimes 'Faba'. Kupicha (1974, 1976) argues convincingly that this division of the species is artificial. She believes that correlated characters divide the species into two natural subgenera, Vicia and Vicilla. These two major clusters within Vicia are distinguished by the presence of a stipule nectary spot, the number of flowers per inflorescence and the relative lengths of the peduncle and subtending leaf. The validity of Kupicha's hypothesis will be discussed further in the following chapters but, in terms of this revision, it is accepted that Vicia subg. Vicia (sensu Kupicha) is a

distinctive, natural grouping and as such forms the taxon which was revised.

Kupicha divides subgenus Vicia into five sections and 32 species. She does not include infra-specific taxa in her classification and the subspecies and varieties used at the beginning of the revision are those listed by Allkin et al. (1983b), which are in turn taken from various works (principally: Ball, 1968; Davis & Plitmann, 1970 and Fedtschenko, 1948). The 65 taxa circumscribed are listed in Table 4.1. Each taxon is listed with its authority and the beginning taxon code, the code given to each taxon name at the beginning of the revision. Beginning taxon codes all start with a B, which indicates that the taxon code refers to a taxon at the start of the revision. The two central numbers are arbitrary numbers, which consistently represent a particular taxon e.g. 06 = V. bithynica or 52 = V. sativa. The final alpha-numeric character represents the taxonomic status of that taxon name, 0 is always taken to mean that the taxon name is accepted at the beginning of the revision, while any number from 1-9 or letter from A-Z indicates that the taxon name is a synonym of an accepted taxon. At the start of the revision all 65 taxa were accepted. Taxa accepted at the conclusion of the revision have end taxon codes which have a similar code structure to the beginning code but the initial letter is a T rather than the B. So, for example, the code B060 indicates a beginning code for the accepted species V. bithynica.

During the course of the revision other Vicia subg. Vicia taxa not included in Vicia subg. Vicia by either Kupicha (1976) or Allkin et al. (1983b) were discovered. These five species and two subspecies, unknown to the above authors (pers. comm.) are distinct entities and so were added to the list of accepted taxa for the revision. Of these five taxa two were species new to science, V. kalakhensis Khattab, Maxted and Bisby (1988) and V. eristalioides Maxted (1989).

V. kalakhensis was collected by Maxted, Ehrman, Khattab and Bisby in Syria during 1986 and V. eristalioides was collected by Maxted, Kitiki and Allkin in Turkey in 1987.

As with all taxonomic investigations, the scope of the revision was limited by time, facilities and material availability. Seventy-two taxa were finally included within the scope of the revision. Due to the restricted nature of a Ph.D. project and the lack of seed of several of the taxa, the project concentrated on scoring morphological characters from dried herbarium specimens. Seventy seven accessions were, however, grown under uniform garden conditions in the University of Southampton greenhouses from seed held in the Vicieae Database Project Genebank. Collecting missions were also undertaken and 592 Vicia subg. Vicia taxon accessions were collected from Britain, France, Spain, Syria, Turkey, the Soviet Union and Yugoslavia. As characters were scored directly into the revision database, it was found easier to score all specimens as pressed specimens, so the specimens could be scored near the personal computer where the data were entered into the database. However, all Maxted collections were scored as freshly pressed material and so 'fresh specimen' characters such as flower colour were still easily recorded.

As well as scoring specimens personally collected, specimens were also borrowed from other herbaria. Davis and Heywood (1973) stress the importance of obtaining herbarium material from as many herbaria as possible, so that a true estimate of within-taxon variation can be made. Thus herbarium specimens were borrowed from the major international herbaria listed in Appendix 1.

TABLE 4.1 Circumscribed taxa of Vicia subgenus Vicia included in the exemplar revision

Sectional placement of taxa follows Kupicha (1976)

Taxon begin code	Section	Taxon name	Authority
B010	Peregrinae	V. aintabensis	Boiss. et Hausskn. ex Boiss.
B020	Hypechusa	V. anatolica	Turrill
B030	Hypechusa	V. assyriaca	Boiss.
B040	Atossa	V. balansae	Boiss.
B050	Vicia	V. barbazitae	Ten. et Guss.
B060	Faba	V. bithynica	(L.) L.
B070	Hypechusa	V. ciliatula	Lipsky
B080	Vicia	V. cuspidata	Boiss.
B090	Hypechusa	V. esdraelonensis	Warb. et Eig
B100	Faba	V. faba	L.
B110	Faba	V. faba subsp. faba var. faba	
B120	Faba	V. faba subsp. faba var. equina	Pers.
B130	Faba	V. faba subsp. faba var. minor	Beck
B140	Faba	V. faba subsp. paucijuga	(Alef.) Murat.
B150	Hypechusa	V. galeata	Boiss.
B160	Faba	V. galilaea	Plitm. et Zoh.
B170	Faba	V. galilaea subsp. faboidea	Plitm. et Zoh.
B180	Faba	V. galilaea subsp. galilaea	Plitm. et Zoh.
B190	Vicia	V. grandiflora	Scop.
B200	Faba	V. hyaeniscyamus	Mout.
B210	Hypechusa	V. hybrida	L.
B220	Hypechusa	V. hyrcanica	Fischer et C. Meyer
B230	Faba	V. johannis	Tamamschjan in Karyagin
B240	Faba	V. johannis var. ecirrhosa	(Popov) Schäfer
B250	Faba	V. johannis var. johannis	
B260	Faba	V. johannis var. procumbens	Schäfer
B270	Vicia	V. lathyroides	L.
B280	Hypechusa	V. lutea	L.
B290	Hypechusa	V. lutea subsp. lutea var. lutea	

Taxon Section begin code	Taxon name	Authority
B300 Hypechusa	V. lutea subsp. lutea var. laevigata	(Smith, J.) Boiss.
B310 Hypechusa	V. lutea subsp. vestita	(Boiss.) Rouy
B320 Hypechusa	V. melanops	Sibth. et Smith
B330 Hypechusa	V. melanops var. melanops	
B340 Hypechusa	V. melanops var. loiseaui	Alleiz.
B350 Peregrinae	V. michauxii	Sprengel
B360 Peregrinae	V. mollis	Boiss. et Haussk. ex Boiss.
B370 Faba	V. narbonensis	L.
B380 Faba	V. narbonensis var. aegyptiaca	Korn. ex Asch. et Schweinf.
B390 Faba	V. narbonensis var. affinis	Korn. ex Asch. et Schweinf.
B400 Faba	V. narbonensis var. jordanica	Schäfer
B410 Faba	V. narbonensis var. narbonensis	
B420 Faba	V. narbonensis var. salmonea	(Mout.) Schäfer
B430 Hypechusa	V. noeana	Boiss. et Reut. ex Boiss.
B440 Hypechusa	V. noeana subsp. noeana	
B450 Hypechusa	V. noeana subsp. megalodonta	Rech.
B460 Atossa	V. oroboides	Wulfen in Jacq.
B470 Hypechusa	V. pannonica	Crantz
B480 Hypechusa	V. pannonica subsp. pannonica	
B490 Hypechusa	V. pannonica subsp. striata	(M. Bieb.) Nymen
B500 Peregrinae	V. peregrina	L.
B510 Vicia	V. pyrenaica	Pourret
B520 Vicia	V. sativa	L.
B530 Vicia	V. sativa subsp. amphicarpa	(Dorthes) Asch. et Graebn.
B540 Vicia	V. sativa subsp. cordata	(Wulfen ex Hoppe) Asch. et Graebn.
B550 Vicia	V. sativa subsp. incisa	(M. Bieb.) Arcang.
B560 Vicia	V. sativa subsp. macrocarpa	(Moris) Arcang.

Taxon Section begin code	Taxon name	Authority
B570 Vicia	V. sativa subsp. nigra	(L.) Ehrh.
B580 Vicia	V. sativa subsp. nigra var. segatalis	(Thuill.) Ser.
B590 Vicia	V. sativa subsp. sativa	
B600 Atossa	V. sepium	L.
B610 Atossa	V. sepium subsp. sepium	
B620 Atossa	V. sepium subsp. montana	(Koch) Hamet-Ahti
B630 Hypechusa	V. sericocarpa	Fenzl
B640 Faba	V. serratifolia	Jacq.
B650 Atossa	V. truncatula	Fischer ex M. Bieb.

Additional taxa added during the revision

B660	V. kalakhensis	Khattab, Maxted et Bisby
B670	V. dionysiensis	Mout.
B680	V. qatmensis	Gomb.
B690	V. tigridis	Mout.
B700	V. sepium subsp. eriocalyx	Celak
B710	V. eristalioides	Maxted
B720	V. sativa subsp. devia	J.G. de Costa

4.3 Taxonomic Revision Study Plan

A generalised methodology or paradigm for undertaking a taxonomic revision has been discussed extensively in the previous chapter and so this discussion will not be reiterated here. However, it may assist the understanding of the exemplar revision if a specific revision study plan is provided which emphasises the logical sequence followed during the revision. A summary of the revision study plan is provided in Figure 4.1.

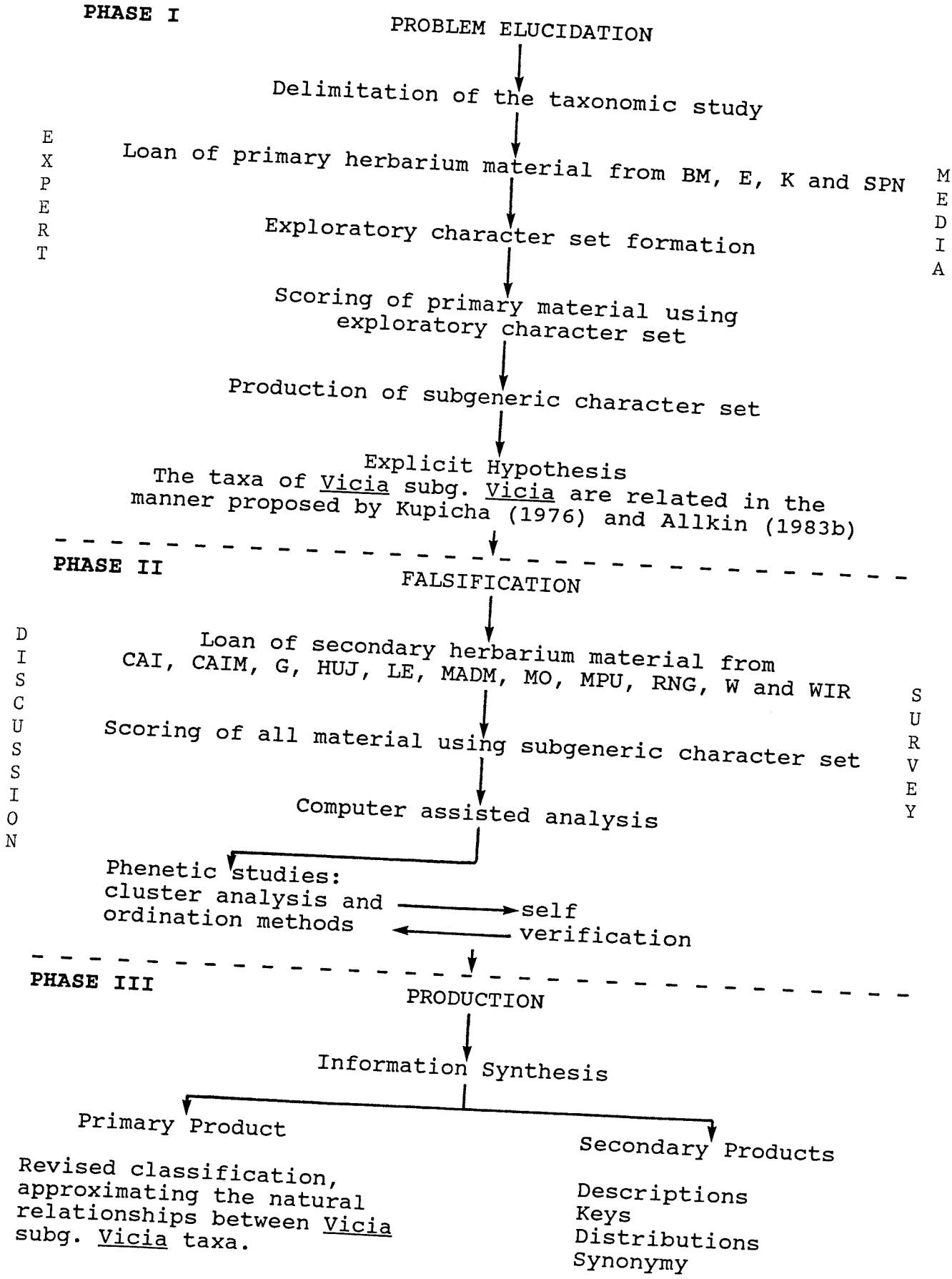
The revision divides naturally into three phases. The first phase, 'problem elucidation', involved the clarification of the taxonomic problems associated with Vicia subg. Vicia. An initial subgeneric character set was formulated from the literature and from a personal study of the taxa. This character set was tested by scoring the primary herbarium material, using a few specimens per taxon, as a representative sample of the variation included within Vicia subg. Vicia. The exploratory scoring of the primary herbarium material resulted in the formulation of the subgeneric character set that was to be used in the second phase. Phase one concluded with the adoption of a general hypothesis of taxon relatedness. This was derived from discussion with Vicieae experts, initial plant observations and comparison of the different published classifications. The hypothesis selected for falsification was that the species of Vicia subg. Vicia were related in the manner proposed by Kupicha (1976), for the sectional and specific level, and Allkin et al. (1983b), for the taxa below the species level.

The second phase, 'falsification', commences with the loan of the secondary herbarium material from major international herbaria. This material was then scored, along with the living specimens, using the subgeneric character set. The data set produced was then analysed using both cluster analysis and ordination techniques. This analysis was undertaken using several different methods, so the results of

each method would verify the results of the other methods used. In this phase the general hypothesis that the taxa of Vicia subg. Vicia were related in the manner suggested was falsified.

The final section, 'production', was where the results and the information gathered concerning the taxa throughout the revision were synthesised to provide the primary and secondary revision products. The primary product was the classification and the secondary products were a range of commodities, such as keys, descriptions, distribution maps, etc.

Figure 4.I. Taxonomic study plan



CHAPTER FIVE
TAXONOMIC HISTORY

5.1 Introduction

An understanding of prior treatments is an essential pre-requisite of any effective systematic study. The five genera of the tribe Vicieae have an extensive taxonomic history. This may be due to the presence of both major and minor crop species, and to the fact that the five genera are centred on the Eastern Mediterranean, where the flora has been extensively studied. Thus this chapter is not an attempt at a comprehensive history of Vicieae taxonomy, but is an attempt to abstract *a posteriori* major taxonomic advances as they have affected *Vicia* subgenus *Vicia*.

The tribe Vicieae (Adans.) DC. (1825a) belongs to the family Leguminosae, subfamily Papilionoideae (Fabaceae, Faboideae, Papilionaceae). The Leguminosae are morphologically diverse, ranging from trees to aquatics to xerophytes, and only rank second to the Gramineae in economic importance to man (Heywood, 1978). The family contains approximately 650 genera and 18,000 species (Polhill, Raven and Stirton, 1981). It is normally divided into three sub-families, of which the largest is the Papilionoideae. The Papilionoideae are distributed through the temperate, sub-tropical and tropical regions. They are mostly herbs with some trees and shrubs among the 400-500 genera and 10,000 species (Polhill, Raven and Stirton, 1981). The leaves are usually pinnate; the flowers are irregular with lateral petals enclosed by the standard in the bud; there are usually 10 stamens, commonly diadelphous but sometimes monadelphous or free.

The Papilionoideae have been traditionally divided into 10 or 11 tribes on habit, vegetative and floral characters. More recently legume taxonomists have tended to increase the number of tribes; Gillett, Polhill and Verdcourt (1971) detail

17 tribes and Polhill (in Lackey 1977) suggests a complete break with prior tribal delimitations to form 31 tribes, while Polhill and Raven (1981) divide the sub-family into 32 tribes. The tribe Vicieae is considered relatively advanced and is placed twenty-first in Polhill and Raven's classification.

Traditionally the tribe Vicieae has included numerous genera with relatively small numbers of species in each; Abacosa Alef., Aphaca Miller, Arachus Medic., Arbus L., Atossa Alef., Bona Medic., Cicer L., Cicercula Medik, Clymenum Miller, Cracca (Riv) Medik, Cujunia Alef., Endusia Benth. & Hook., Ervilia Link., Ervum L., Faba L., Graphiosa Alef., Hypechusa Alef., Lastila Alef., Lathyrus L., Lens Mill., Navidura Alef., Nissolia L. non Jacq., Orobus L., Parallosa Alef., Pisum L., Sallunia Alef., Swantia Alef., Tuamina Alef., Vicia L. Vicilla Schur and Wiggersia Alef. In recent years the number of genera has decreased and the number of species per genus has increased. The Vicieae generic classification has stabilised into a generally accepted grouping of five genera, Kupicha (1981) lists these as follows: Vicia L.; Lathyrus L.; Lens Mill.; Pisum L. and Vavilovia A. Fedorov. She comments that the Vicieae, narrowly defined, excluding Arbus Adans. and Cicer L.,

> "form a small, distinct group with several specialised features: tendrilous leaves; unusual stem vasculature; precise and elaborate floral details."

She then goes on to provide the following key to the five genera:

"1 Style dorsally compressed, folded longitudinally, with margins meeting adaxially, and pubescent on adaxial (inner) face:
2 Annuals; leaves tendrilous, usually with more than one pair of leaflets; stipules large, foliaceous; leaflets conduplicate in bud **Pisum**
2 Perennials; leaves mucronate to shortly tendrilous, unijugate; stipules small; leaflets supervolute in bud **Vavilovia**
1 Style not as above (i.e. not folded longitudinally):
3 Style dorsally compressed, pubescent only on adaxial (inner) face:
4 Leaflet ptyxis supervolute **Lathyrus**
4 Leaflet ptyxis conduplicate:
5 Seeds lenticular **Lens**
5 Seeds +/- spherical **Vicia**
p.p. (V. koeieana + V. ervilia)
3 Style not as above; if dorsally compressed then pubescent all round or only on abaxial face **Vicia**"

Kupicha (1981) considers Vicia L. to comprise approximately 140 species, chiefly located in Europe, Asia and North America, extending to temperate South America and tropical East Africa, but the genus is primarily centred on the Mediterranean and Irano-Turanian regions. However, Allkin et al. (1986) have increased the number of accepted species to 166 and this estimate errs on the conservative side (Goyder pers. comm.). Kupicha (1976) provides the following description:

Vicia L., Sp. Pl. 734 (1753).

"Perennial and annual herbs with erect or more usually climbing or sprawling habit; plants never tuberous. Stems angled but not winged, usually with complete replacement of cortical vascular bundles at the nodes, occasionally with partial replacement. Leaves hypostomatic to epistomatic, paripinnate and tendrilous or mucronate or very rarely imparipinnate,usually with several to many pairs of leaflets, very rarely unijugate; stipules semisagittate or simple, sometimes toothed or laciniate, occasionally dimorphic, sometimes with a nectary on abaxial side; vernation of leaflets conduplicate (supervolute in V. biennis); venation pinnate, brochidodrome. Inflorescence racemose, 1 - many-flowered, occasionally branched. Calyx usually with oblique mouth and teeth of unequal length ('irregular'), sometimes actinomorphic ('regular'). Vexillum oblong, stenonychioid or platonychioid, very rarely bossed or pouched at the

fold, rarely pubescent on inner-face. Alae usually with 'pleat' in upper edge of limb. Staminal tube oblique at apex. Style linear, not contorted, dorsally or laterally compressed or occasionally terete, always hairy; distribution of pubescence various but style never hairy on adaxial side only (except in some specimens of V. ervilia). Legume compressed or occasionally subtorulose, often stipitate, sometimes hairy but hairs rarely tuberculate; pod sometimes occasionally 'woolly' parenchymatous tissue between the seeds. Seeds with short to long hilum; testa smooth or very rarely rough; lens near hilum or occasionally on opposite side of seed; free amino acid canavanine sometimes present.

Type: V. sativa L."

5.2 Taxonomy History of Subgenus Vicia

Each revision must have a taxonomic starting point from which it progresses and from which the existing taxonomy is tested. For this revision the starting point selected was the classification of Vicia L. described in Kupicha (1976). In this classification she divided the genus into two subgenera, Vicilla (Schur) Rouy in Rouy & Fouc. and Vicia. The current revision is restricted to subgenus Vicia and the following taxonomic history provides a concise history of Vicia, with specific emphasis given to Vicia subg. Vicia (sensu Kupicha, 1976) species.

As Kupicha (1976) points out, the group she defined as subgenus Vicia has in fact existed in the classification of numerous prior authors. She makes clear the taxonomic history of Vicia is really a history of three or four groups of species; 'Cracca', 'Vicia', 'Ervum' and sometimes 'Faba' (perhaps she might also have included 'Orobus'). Thus the taxonomic history that follows will focus on the Vicia and Faba groups of species, that she was later to unite in her subg. Vicia.

Another factor to be considered when preparing a taxonomic history for Vicia is that very few authors actually studied the genus worldwide. Revisions were commonly undertaken during the preparation of regional floras, thus, by definition, they deal with a

restricted numbers of species from a particular region. This indicates the need for a worldwide revision of Vicia, not just a revision of Vicia subg. Vicia. A detailed taxonomic history of the tribe Vicieae and the genus Vicia, as a whole, are provided in Kupicha (1974 and 1976).

The discussion of each classification will follow the same general format: a summary of the classification in tabular form; a discussion of the innovations of that classification; a discussion of the authors' conception in relation to that of previous authors; a discussion of the characters used to distinguish the author's taxonomic groups. Each classification will be discussed in chronological order, except Kupicha (1976). Kupicha's classification will be discussed first as it forms the starting point for this revision. If her conception is understood it will assist in the interpretation of other authors' conceptions.

5.2.1 Vicia Subgenus Vicia Sensu Kupicha (1976)

Kupicha's classification of Vicia was produced for her Ph.D. thesis in 1974 and subsequently published as Kupicha (1976). It was a major step in Vicia taxonomy as it was the first comprehensive worldwide account of Vicia since Seringe (1825).

Kupicha (1976) considers that the division of Vicia into four subgroups (Cracca, Ervum, Vicia and Faba) does not reflect a balanced natural sub-division of Vicia species and argues that a division into two broader groups reflects the natural relationships within the genus. Her classification is radical, in that no previous author had identified within Vicia (in this sense containing all four specific groupings) two major subgeneric groupings. One of these groups contains Ervum and Ervilia and the other contains Vicia and Faba. The history of Vicia since Linnaeus can be read as a history of these four taxonomic groups together with the different rankings given to them by various authors.

Kupicha comments that the number of species in subg. Vicia is smaller and is more internally coherent than subg. Vicilla. The former is therefore easier to circumscribe. Her subgenera can be distinguished using the following characters:

Character	Subg. Vicilla	Subg. Vicia
Stipule nectary	absent	present on abaxial stipule face
Peduncle length	equal to or longer than leaf	short or absent
Style type	terete, dorsally compressed pubescent or tufted, laterally compressed	dorsally compressed, tufted
Keel shape	edge curved round above style	apical part encircled style and anthers
Legume	stipitate, sub-torulose	never stipitate
Canavanine	present	absent

She provided a conspectus and key for subg. Vicia and its sections (see Table 5.1). The characters she used to distinguish her groupings are those used throughout the taxonomic history of Vicia to segregate the Vicia and Faba groups of species from other Vicia.

As with any major, new classification there are a few small problems with it and with the characters she used to define the sub-generic groups. Firstly, the set of species circumscribed is incomplete, as was discussed in thesis section 4.1. Secondly, V. bithynica and V. sativa subsp. devia often have peduncles longer than the subtending leaf, but are clearly members of her subg. Vicia (the latter is a rare endemic of Madeira and was unknown to Kupicha). Kupicha understood the problem of the varying peduncle length of V. bithynica; she comments on the anomalous position of the species, but is unable to propose a solution. Thirdly, purple flower

colour is a strange character to use to distinguish sect. Peregrinae (in her key), as three of the four included species have yellowish coloured flowers. This point is corrected in the description of the section, where flower colour is given as "whitish, pale yellow or dark violet".

Table 5.1 Vicia subgenus Vicia (Kupicha, 1976)

All classifications listed include the authors of the names where they were listed by the author of the classification.

I Subgenus Vicilla (Schur) Rouy in Rouy & Fouc.
(Stipules without nectariferous spot; inflorescence usually equalling or exceeding the subtending leaf, usually many-flowered.)

Seventeen sections

II Subgenus Vicia
(Stipules with nectariferous spot on abaxial surface; inflorescence much shorter than the subtending leaf, usually 1-few-flowered.)

18. Section Atossa (Alef.) Aschers. & Graebner
(Perennial; calyx irregular; inflorescence several-flowered, vexillum oblong; suture of legume not parallel; lens of seed close to hilum)

V. oroboides Wulfen
V. sepium L.
V. balansae Boiss.
V. truncatula Fisch. ex Bieb.

19. Section Vicia
(Leaves usually with more than three pairs of leaflets [if fewer, then leaflets less than 1 cm long], lateral veins of leaflets prominent and straight; calyx subregular; sutures of legume parallel)

V. pyrenaica Pourret
V. sativa L. agg.
V. grandiflora Scop.
V. barbazitae Ten. & Guss.
V. lathyroides L.
V. cuspidata Boiss.

20. Section Faba (Miller) Ledeb.
(Leaves with 1-3 pairs of leaflets which are more than 2 cm long, lateral veins of leaflets not prominent, curved toward apex; calyx subregular, sutures of legume parallel)

V. faba L.
V. narbonensis L.
V. galilaea Plitm. & Zoh.
V. hyaeniscyamus Mouterde
V. johannis Tamamschjan
V. bithynica L.

Table 5.1. Continued

21. Section Hypechusa (Alef.) Aschers. & Graebner
(Annual; Calyx irregular; inflorescence 1-many-flowered, vexillum oblong or stenonychioid [i.e. banner wider than the claw]; suture of legume not parallel; lens of seed opposite hilum)

V. anatolica Turrill
V. assyriaca Boiss.
V. ciliatula Lipsky
V. esdraelonensis O. Warb.
V. galeata Boiss.
V. hybrida L.
V. hyrcanica Fischer & C. A. Meyer
V. lutea L.
V. melanops Sibth. & Smith
V. noeana Reuter ex Boiss.
V. pannonica Crantz
V. sericocarpa Fenzl

22. Section Peregrinae Kupicha
(Calyx irregular; inflorescence 1-2-flowered, vexillum stenonychioid, flowers purplish; suture of legume not parallel; lens of seed close to hilum)

V. aintabensis Boiss. & Hausskn.
V. michauxii Sprengel
V. mollis Boiss. & Hausskn. ex Boiss.
V. peregrina L.

5.2.2 A Concise History of Subgenus Vicia

Early classifications

Linnaeus (1753) based his conception of the genera later to be included in the Vicieae, on the work of Tournefort (1694). Tournefort used the names Faba, Lupinus, Orobus, Pisum, Lathyrus, Clymenum, Ochrus, Vicia, Ervum, Galega, Astragaloides, and Aphaca, but does not attempt a comprehensive list of taxa. Linnaeus listed taxa and included 23 species, which would now be regarded as members of Vicia (see Table 5.2). He included 17 Vicia species, one of which (V. nissoliana) is currently regarded as a synonym of V. cassubica (Allkin et al., 1986), and six further species that were later to be transferred to Vicia from related genera (Orobus, Lathyrus and Ervum). Ten of these 23 species would be regarded by Kupicha as being members of Vicia subg. Vicia. Interestingly, even as early as the Species Plantarum, Linnaeus recognised that Vicia could be subdivided into two groups on the basis of peduncle length. This character was to remain important in subdividing the genus throughout its taxonomic history.

One of the problems that recurs in the history of subgenus Vicia is the placement of the fababean: should it be included in Vicia or separated into the monospecific genus Faba? Linnaeus (1753) included the fababean in Vicia as V. faba. Linnaeus (1737) commented that he examined hundreds of flowers of Vicia and Faba and found no important distinguishing characters. He notes that only the seeds are different, but does not consider this important enough to warrant generic distinction.

Adanson (1763) splits his family XLIII, the Leguminosae into six "sections", the sixth of which is the Vicieae. Within this section he includes 12 genera, three of which, Faba, Vicia and Ervum, are referable to contemporary Vicia. Interestingly, Adanson concurs with Tournefort (1694) and considers V. faba sufficiently distinct from Vicia to warrant

Table 5.2. Linnaean species of Vicia

In this and subsequent classifications current members of subg. Vicia are marked with an asterisk

Linnaeus 1753 Binomial	Later transfers Or Synonyms
Orobus lathyroides	V. oroboides Wulfen *
Lathyrus bithynicus	V. bithynica (L.) L. *
Vicia	
a. (Pedunculis elongatis.)	
V. pisiformis	
V. dumetorum	
V. sylvatica	
V. cassubica	
V. cracca	
V. onobrychioides	
V. nissoliana	
V. biennis	V. cassubica L.
V. benghalensis	
b. (Floribus axillaribus, subsessilibus.)	
V. sativa *	
V. lathyroides *	
V. lutea *	
V. hybrida *	
V. peregrina *	
V. sepium *	
V. narbonensis *	
V. faba *	

generic rank. Adanson's treatment is followed by Bronn (1822), who gives the Leguminosae order rank, and makes the Vicieae a subtribe of the tribe Curvembryae Diadelphae, while at the same time retaining the three genera Vicia, Ervum and Faba. De Candolle (1825b) comments that the Vicieae have been well covered by Adanson and Bronn and he basically concurs with their conclusions. In the Prodromus, Seringe (1825) raises the Vicieae to tribal rank. He maintains the same three genera, with 13 species that would now be considered to be accepted members of subg. Vicia (see Table 5.3). Seringe included two subgroups within Vicia, the distinction between the two subgroups is made on the basis of peduncle length. Current subg. Vicia species are restricted to the monospecific genus Faba and Vicia subgroup B. The latter contains 40 species, 12 of which are currently accepted, the other 28 being synonyms of the 12 species (Allkin et al., 1986).

Koch (1836) moves away from the previous classifications and unites Vicia, Faba and part of Ervum into the one genus Vicia. His tribe Vicieae contains six genera: Cicer, Vicia, Ervum, Pisum, Lathyrus and Orobus. All the species, that are now considered subg. Vicia members, are included in his circumscription of Vicia. His revision is presented in a regional flora (Florae Germanicae Et Helveticae) and thus is not species comprehensive. However, he does provide the first detailed subdivision of Vicia (shown in Table 5.4), largely on the basis of style pubescence, peduncle length, flower number / inflorescence, leaf and seed shape characters. The species that would be currently be considered members of subg. Vicia are marked with an asterisk and form a distinct group at one extreme of his Vicia (the last three subgroups in his section Viciae). This conception returned V. faba to within Vicia.

In a regional flora for Dalmatia, Visiani (1842) places all the subg. Vicia taxa in his Vicia. He splits the genus into two sections on the basis of the relative length of the peduncle. The current subg. Vicia taxa are placed in his

Table 5.3. Abstracted classification of Vicia - Seringe (1825)

Seringe 1825 Binomial	Synonyms
Faba	
Faba Vulgaris Moench	V. faba *
Vicia section A. - Floribus pedunculatis.	
50 spp. not included in subg. Vicia sensu Kupicha (1976)	
Vicia section B. - Floribus subsessilibus.	
V. pimpinelloides Mauri	
V. sativa L. *	V. sativa *
V. intermedia Vivian.	
V. incisa Bieb.	V. sativa *
V. cornigera Chaub.	V. sativa *
V. repens D'Urvill.	
V. canadensis Zuccagni	
V. globosa Retz.	V. sativa *
V. cordata Wulf. ex Hoppe in Sturm.	V. sativa *
V. pilosa Bieb.	V. sativa *
V. peregrina L. *	V. sativa *
V. michauxii Spreng. *	
V. amphicarpa Dorth.	
V. pyrenaica Pourr. *	V. sativa *
V. lathyroides L. *	
V. laevigata Smith	
V. dubia Schult.	V. lutea *
V. lutea L. *	V. articulata
V. hybrida L. *	
V. spuria Rafin.	
V. pusilla Muhlenb.	V. hybrida *
V. grandiflora Scop. *	
V. sordida Waldst. et Kit.	
V. biebersteinii Besser	V. grandiflora *
V. tricolor Sebast. et Mauri	V. grandiflora *
V. sepium L. *	V. melanops *
V. lineata Bieb.	
V. pannonica Jacq. *	V. pannonica *
V. bipartita Moench	
V. truncatula Fisch. *	
V. narbonensis L. *	
V. platycarpos Roth.	
V. monodelpha Roth.	V. narbonensis *
V. bacla Moench	V. narbonensis *
V. ciliaris Sibth. et Smith	
V. melanops Sibth. et Smith *	
V. bactra Zuccagni	
V. bicolor Willd.	
V. triflora Tenor	V. melanops *
V. leptophylla Rafin.	V. calcarata
	V. peregrina *

11 Vicia spp. Not subg. Vicia sensu Kupicha (1976) and that Seringe considers to have dubious Vicia status.

Table 5.4. Classification of Vicia - Koch (1836)

Sect. 1. Erviliae (Stylus superne undique aequaliter pilosus.)

Subgroup 1. (Pedunculi elongati pauciflori.)
- V. hirsuta
- V. tetrasperma
- V. gracilis Loiseleur
- V. monantha
- V. ervilia Willd.

Subgroup 2. (Pedunculi elongati multiflori.)
- V. pisiformis L.
- V. sylvatica L.
- V. cassubica L.
- V. orobus DC.

Sect. 2. Viciae (Stylus in latere inferiore apicis barbatus, ceterum glaber vel simul apice undique villosus.)

Subgroup 1. (Pedunculi elongati multiflori, folio fulcrante longiores brevioresve.)
- V. dumetorum L.
- V. cracca L.
- V. tenuifolia Roth.
- V. villosa Roth.
- V. onobrychioides L.

Subgroup 2. (Pedunculi elongati uni, pauciflori.)
- V. bithynica L. *

Subgroup 3. (Pedunculi breviter racemosi 4 6 flori, flore breviores vel eos vix superantes. Sem. oblonga, hilum in angustiore latere gerentia.)
- V. faba L. *

Subgroup 4. (Pedunculi uni-biflori, vel 4 6 flori et breviter racemosi, pedunculis racemisve flore brevioribus. Sem. globosa vel oblonga, hilum non in latere angustiore gerentia.)
- V. narbonensis L. *
- V. oroboides Wulf. *
- V. sepium L. *
- V. pannonica Jacq. *
- V. hybrida L. *
- V. lutea L. *
- V. grandiflora Scop. *
- V. sativa L. *
- V. angustifolia Roth.
- V. cordata Wulfen ex Hopp. syn. V. sativa *
- V. peregrina L. * syn. V. sativa *
- V. lathyroides L. *

section 1. Euvicia, except for V. bithynica (Table 5.5). This placement is logical due to the variable length peduncle (4-45mm long) of V. bithynica. The placement of this species will be discussed in detail in Chapter 9. Visiani fails to include the taxonomically distinct V. oroboides in his classification, although it grew in the area covered by his flora.

A second important regional flora was published in 1842 by Ledebour, which covered Poland and much of what is today the West and Southern republics of the Soviet Union. He divides Vicia into two sections (Table 5.6), V. faba being the monospecific member of section Faba. All the remaining subg. Vicia sensu Kupicha (1976) species are placed in the first subsection of section Vicia except for V. bithynica. Like Visiani, Ledebour uses relative peduncle length as the major distinguishing character between the two major infra-generic clusters and thus V. bithynica is separated. This conception of the subgenus Vicia species was subsequently followed by Dietrich (1847). Though he places V. faba remote from V. narbonensis, allying it with V. sativa, unlike previous or subsequent authors; he does, however, return V. bithynica to its more common placement with other subg. Vicia species.

Godron (1849) took a narrower view of Vicia and placed the 'Cracca' group of species in the separate genus, Cracca. The ten subg. Vicia (sensu Kupicha, 1976) species are included in his Vicia on the basis of calyx and style shape, and the distribution of hairs on the style. Godron attempts some detailed distinction of supra-generic groups, using sections, subsections groupings and sub-subsections (Table 5.7). The subg. Vicia species are included in the first four supra-generic groupings in the classification. This groups the species into sections Vicia, Peregrinae with Hypechusa, Faba and Atossa sensu Kupicha.

Table 5.5. Classification of Vicia - Visiani (1842)

Section 1. Euvicia Vis. (Flores solitarii, aut racemosi, in axilla subsessiles.)

V. faba L. *
V. narbonensis L. *
V. hybrida L. *
V. lutea L. *
V. pannonica Crantz *
V. melanops Sibth. et Smith *
V. grandiflora Scop. *
V. sativa L. *
V. cordata Wulf. ex Hoppe in Sturm. syn. V. sativa *
V. lathyroides L. *
V. peregrina L. *
V. sepium L. *

Section 2. Cracca (Flores solitarii aut racemosi, in axilla longe pedunculati.)

V. bithynica L. *
V. monantha Desf.
V. ervilia Willd.
V. gracilis Loisel.
V. hirsuta Koch
V. leucantha Bivon.
V. cassubica L.
V. dumetorum L.
V. gerardi DC.
V. dasycarpa Ten.
V. cracca L.
V. tenuifolia Roth.
V. ochroleuca Ten.
V. altissima Desf.
V. onobrychioides L.
V. atropurpurea Desf.

Table 5.6. Classification of Vicia - Ledebour (1842)

Ledebour 1842 Binomial	Synonyms
Section 1. Faba Tournef. Legumen coriaceum, subturgidum, torulosum. Semen elliptica, compressa: umbilico terminali, margini breviori imposito. - Caulis erectus. Cirrhi simplices, subnulli. Peduncli breviter racemosi.	
V. faba L. *	
Section 2. Vicia Tournef. Legumen oblongum. Semina subglobosa v. lenticularia v. oblonga et tunc umbilico laterali, margini longiori imposito. - Herbae saepius scandentes. Folia cirrhosa: cirrho plerumque ramoso.	
1. Flores subsessiles, v. breviter pedunculati.	
V. narbonensis L. *	
V. sativa L. *	
V. angustifolia Roth.	V. sativa L. *
V. cordata Wulf. ex Hoppe	V. sativa L. *
V. incisa Bieb.	V. sativa L. *
V. pilosa Bieb.	V. sativa L. *
V. lathyroides L. *	
V. peregrina L. *	
V. lutea L. *	
V. hybrida L. *	
V. hyrcanica Fisch. et Meyer *	
V. grandiflora Scop. *	
V. pannonica Jacq. *	
V. lineata Bieb.	V. pannonica Crantz *
V. sepium L. *	
V. truncatula Fisch. *	
2. Floribus pedunculatis: pedunculis elongatis.	
V. pisiformis L.	
V. dumetorum L.	
V. pseudo-orobus Fisch. et Meyer.	
V. amoena Fisch. ex DC.	
V. cassubica L.	
V. abbreviata Fisch. ex Spreng.	
V. monosperma C. Koch	
V. pallida Turcz.	
V. sitchensis Bong.	
V. megalotropis Ledeb.	
V. cracca L.	
V. branchitropis Karel et Kiril.	
V. lilacina Ledeb.	
V. tenuifolia Roth.	
V. villosa Roth.	
V. sylvatica L.	
V. costata Ledeb.	
V. biennis L.	
V. picta Fisch. et Meyer	
V. variegata Willd.	
V. purpurea Steven	
V. alpestris Steven	
V. multicaulis Ledeb.	
V. cinerea Bieb.	
V. bithynica L. *	

Table 5.7. Classification of Vicia - Godron (1849)

Godron 1849 Binomial	Synonyms
Section 1. Fleurs axillaires, solitaires ou géminées, non portées par un pédoncule cummun.	
a. Gousse sessile: calice régulier.	
V. sativa L. *	
V. cordata Wulf. ex Hoppe	V. sativa L. *
V. angustifolia Roth.	V. sativa L. *
V. cuneata Guss.	V. sativa L. *
V. lathyroides L. *	
V. pyrenaica Pourr. *	
V. amphicarpa Dorth.	V. sativa L. *
b. Gousse stipitée; calice irrégulier.	
V. peregrina L. *	
V. lutea L. *	
V. hybrida L. *	
Section 2. Fleurs en grappes plus ou moins pédonculées,	
a. Gousse sessile.	
V. faba L. *	
V. narbonensis L. *	
V. bithynica L. *	
b. Gousse stipitée	
1. Style épais, barbu sous le stigmate.	
V. sepium L. *	
V. pannonica Jacq. *	
V. syrtica Dub.	
V. argentea Lapey.	
V. onobrychioides L.	
V. altissima Desf.	
V. dumetorum L.	
2. Style fin, velu tout autour et non barbu au sommet.	
V. pisiformis L.	
V. sylvatica L.	
V. orobus DC.	
V. cassubica L.	

Although Godron produces a perceptive classification from which embryonic Kupicha sections can be identified, he does not accurately record the distinguishing characters used for identifying the major groups in his classification. He, for instance, splits the two Vicia sections on the basis of number of flowers per inflorescence and the presence of a peduncle. His section 1 species should not possess a peduncle, but this is true for one of the species he includes, V. peregrina.

A distinct difference of emphasis is seen between the classifications of Visiani and Godron. Visiani follows Seringe (1825), whereas Godron presents a more radical classification using the characters shown to be useful by Koch (1836). Visiani's classification does not attempt to reflect the detailed supra-generic grouping of species known at that time. Godron presents a more precise breakdown of the species relationships, but his inaccurate scoring of characters results in unusual alliances, such as the linking of V. pannonica to V. orobrychioides and V. dumetorum. V.pannonica has throughout been regarded as a more natural ally of V. hybrida.

The first detailed conceptions

Alefeld (1859, 1860a,b,c, 1861a,b & 1866), in a series of publications, completely deconstructed and refashioned contemporary concepts of the Vicieae and can be considered a watershed in Vicieae research. He raised the rank of the tribe to a sub-family (Alefeld, 1859) with three component tribes. His tribe Viciidae is equivalent to what is now regarded as Vicia which was divided into two subtribes, see Table 5.8. Alefeld's Viciosae is a development of Euvicia sensu Visiani (1842), although V. bithynica was placed with its current allies. There is a striking similarity between Viciosae and subg. Vicia Kupicha (1976). Alefeld is restrictive in his generic concept and thus produces, in

Table 5.8. Classification of sub-tribe Viciosae - Alefeld (1859, 1860, 1861)

Alefeld Binomial	Synonyms
Subtribe Viciosae Alef. (Stipular nectaries, inflorescence few flowered, sessile or sub-sessile)	
Hypechusa Alef. subg. 1 Masarunia Alef.	
H. hybrida Alef.	Vicia hybrida L. *
Hypechusa Alef. subg. 2 Euhypechusa	
H. purpurascens Alef.	V. pannonica Crantz *
H. pannonica Alef.	V. pannonica Crantz *
H. hircania Alef.	V. hyrcanica Fischer & C. Meyer *
H. lutea Alef.	V. lutea L. *
H. sericocarpa Alef.	V. sericocarpa Fenzl *
H. tricolor Alef.	V. melanops Sibth. et Smith *
Wiggersia minima Alef.	V. lathyroides L. *
W. cuspidata Alef.	V. cuspidata Boiss. *
Vicia subg. 1 Megalusa Alef.	
V. truncatula Fisch. *	
V. trichomera Alef.	V. truncatula *
Vicia subg. 2 Euvicia	
V. pyrenaica Pourr. *	
V. sativa L. *	
V. angustifolia Roth.	V. sativa L. *
V. canariensis Alef.	V. sativa L. *
V. abyssinica Alef.	V. sativa L. *
V. cordata Wulf. ex Hoppe	V. sativa L. *
V. amphicarpa Dorthes	V. sativa L. *
Vicia subg. 3 Alangula Alef.	
V. peregrina L. *	
Vicia subg. 4 Taenifila Alef.	
V. bithynica L. *	
V. narbonensis L. *	
V. serratifolia Jacq. *	
Atossa sepium Alef.	V. sepium L. *
A. clusii Alef.	V. oroboides Wulfen *
Cujunia grandiflora Alef.	V. grandiflora Scop. *
Faba vulgaris Moench	V. faba L. *
Tuamina michauxii Alef.	V. michauxii Spr. *
Subtribe Ervosae Alef. (No stipular nectaries, peduncle elongated)	
No subgenus Vicia taxa.	

total, thirteen new generic names for Vicieae (sensu lato) taxa. However, his understanding of the detailed relationships of the species is acute. He also combines established 'good' characters, such as relative peduncle length and numbers of flowers per inflorescence, with others such as presence of stipular nectaries, used to distinguish his subtribe Viciosae and the importance of lens position in defining the group he refers to as the genus Hypechusa. Out of the 32 species included in subg. Vicia by Kupicha (1976), 20 are included in sub-tribe Viciosae.

Bentham (1865) included six genera in his tribe Vicieae, with Vicia returned to its pre-Alefeld conception, including the Faba, Vicia, Ervum and Cracca groups of species. He does not discuss intra-generic classification in detail, but does state that he supports the division of Vicia into two sub-groups on the basis of relative peduncle length and the numbers of flowers per inflorescence. Likewise, Schur (1866), in a regional flora, uses relative peduncle length and numbers of flowers per inflorescence to define sub-generic groupings. He also uses the generic name Vicilla to include the V. ervilia groups of species, the name later used to include the Ervum and Cracca groups by Boissier (1872), Rouy (1899) and Kupicha (1976).

Boissier (1872) follows the Visiani classification, but presents a very detailed supra-generic classification of Vicia. He includes 25 of the 26 species from subg. Vicia sensu Kupicha (1976) in section Euvicia, only excluding V. bithynica in the second section Cracca (Table 5.9). V. bithynica is excluded from the subg. Vicia grouping due to its extended peduncle, as in Visiani's (1842) classification, discussed above. The characters used for the separation of intra-generic groups are the established 'good' characters of relative peduncle length and the numbers of flowers per inflorescence. Boissier added to these style pubescence characters used previously by Koch (1836) and Godron (1849).

Table 5.9. Classification of Vicia - Boissier (1872)

Vicia

Section 1. Euvicia Vis. (Flores 1-2, rarius racemosi, in axillis subsessiles. Stylus ad latus inferius apicus barbatum.)

*. Perennes, flores racemosi.
V. oroboides Wulf. in Jacq. *
V. sepium L. *
V. truncatula M.B. *
V. balansae Boiss. *

**. Annuae, flores saepius solitarii vel bini, legumen ad suturas non denticulatum. Plantae caulibus debilibus scandentibus.

+ Hilus ad seminis latus radiculae oppostum situs.

a. Vexillum hirsutum.
V. pannonica Jacq. *
V. hybrida L. *

b. Vexillum glabrum.

X Legumen hirsutum.
V. lutea L. *
V. sericocarpa Fenzl *

XX Legumen glabrum.
V. melanops Sibth. et Smith *
V. hyrcanica Fischer et Meyer *
V. galeata Boiss. *
V. assyriaca Boiss. *
V. noeana Reut. in Boiss. *

++ Hilus radiculae proximus vel contiguus.

a. Calyx regularis.

X Semina laevia.
V. grandiflora Scop. *
V. barbazitae Ten. et Guss. *
V. sativa L. *
V. angustifolia Roth.
(syn. V. sativa L. *)

XX Semina tuberculata.
V. lathyroides L. *
V. cuspidata Boiss. *

b. Calyx irregularis.
V. mollis Boiss. et Haussk. *
V. peregrina L. *
V. michauxii Spreng. *
V. aintabensis Boiss. et Haussk. *

***. Annuae, flores 1-2 axillares, legumen ad suturas denticulatum. Plantae crasso erecto.
V. narbonensis L. *
V. serratifolia Jacq. *

Section 2. Cracca Riv.

Series A. - Perennes, Legumen apice oblique truncatum acutum
22 species from subg. Vicilla sensu Kupicha

Series B. - Species monocarpicae

*. Stylus a dorso compressus apice subtus barbatus. Legumen apice oblique truncatum acutum.

V. bithynica L. *

**. - 5 species from subg. Vicilla sensu Kupicha
***. - 7 species from subg. Vicilla sensu Kupicha
****. - 12 species from subg. Vicilla sensu Kupicha

Faba

Faba vulgaris Moench (syn V. faba L.*)

Boissier was the first author to use life form as a distinguishing character. He separated the four perennial species in his section Euvicia. These four species were later to form section Atossa (Alef.) Aschers. & Graebn. in Kupicha (1976) classification. Boissier's circumscription of Euvicia closely resembles Kupicha's, as well as her sect. Atossa, her complete section Peregrinae Kupicha is found.

Taubert (1894) presents a classification which develops on from Boissier (1872). Section 1. Euvicia is complete, but section 2. Cracca is split into three sections, Cracca, Ervum and Ervilia. Taubert includes the subg. Vicia (sensu Kupicha) species in section 1. Euvicia (Table 5.10). His classification does not include all the known species, but does provide a detailed intra-sectional review of species groupings. He makes some radical changes to proposed species relationships of previous authors. He places V. pyrenaica with V. sepium and V. oroboides due to their shared perennial habit, even though the former bears the general facies of a V. sativa form. Taubert also places V. grandiflora with V. pannonica and V. melanops, because all three species have yellow flowers. V. grandiflora is, however, distinct from the latter two species, because of its different calyx mouth shape and seed lens position, as shown by Alefeld. These unusual allegiances are explained by his use of characters not previously considered significant (e.g. life form, flower colour) to distinguish major groups and his failure to included established "good" characters used by previous authors (e.g. relative peduncle length, number of flowers per inflorescence and presence of stipular nectaries).

Rouy (1899) includes all the subg. Vicia sensu Kupicha species in subg. Euvicia sensu Rouy (Table 5.11). Although Rouy's Floral account covers a restricted area away from the centre of diversity of the genus it attempts a comprehensive intra-genus classification. His classification of subg. Vicia has evolved from Godron (1849), using very similar character

Table 5.10. Classification of section Euvicia - Taubert (1894)

Section 1. Euvicia Vis.

A. Perennes. Taubert (Pfl. perennierend; Bl. in Trauben.)
6 species
V. sepium L. *
V. oroboides Wulf. in Jacq. *
V. pyrenaica Pourr. *

B. Annuae. Taubert (Pfl. jährig; Bl. einzeln oder zu 2 in den Blattachseln.)

a. Vicieinae Taubert (Pfl. mit dünnen, niederliegenden oder kletternden Stengeln.)

+ Platycarpae Taubert (Hülse kurz und briet.)
8 species
V. lutea L. *
V. hirta Balb. (syn. V. lutea L. *)
V. hybrida L. *

++ Niphocarpae Taubert (Hülse verlängert, schwertförmig.)
30 species

*** Ochroleucae Taubert** (Bl. +/- gelblich, bisweilen purpurn werdend)
V. pannonica Crantz *
V. grandiflora Scop. *
V. melanops Sibth. et Smith *

**** Purpurascentes Taubert** (Bl. von Anfang an +/- purpurn.)
V. sativa L. *
V. angustifolia Benth. (syn. V. sativa L. *)
V. lathyroides L. *

b. Fabinae Taubert (Pfl.+/- steif aufrecht, fleischig.)
V. faba L. *
V. narbonensis L. *

Table 5.11. Classification of sub-genus Euvicia - Rouy (1899)

Rouy Binomial	Synonyms
Sub-genus 1. Euvicia Rouy (Style barbu sous le stigmate à la face inférieure. Légume comprimé plus rarement subcylindrique, sessile ou stipité, rostré.)	
Section I. Subsessiles Nob. (Fleurs axillaires, 1-2, très rarement 3-4, subsessiles ou brièvement pédicellées.)	
* (Calice régulier; légume sessile.)	
1. (Plantes annuelles ou bisannuelles.)	
V. communis Rouy	V. sativa L. *
V. sativa L. *	
V. cordata Wulf. ex Hoppe	V. sativa L. *
V. maculata Presl	V. sativa L. *
V. heterophylla Presl	V. sativa L. *
V. angustifolia Reichdt.	V. sativa L. *
V. amphicarpa Dorthes	V. sativa L. *
V. barbazitae Ten. et Guss. *	
V. lathyroides L. *	
V. olbiensis Reut. et Sh.	V. lathyroides L. *
2. (Plante vivace.)	
V. pyrenaica Pourr. *	
** (Calice irrégulier; légume stipité.)	
V. peregrina L. *	
V. lutea L. *	
V. linnaei Rouy	V. hybrida L. *
Section 2. Pedunculatae Nob. (Fleurs en grappes ou, plus rarement, 1-5 au sommet d'un pédoncule axillaire.)	
* (Légume sessile; feuilles à 1-3 paires de folioles; plantes annuelles.)	
V. narbonensis L. *	
V. faba L. *	
V. bithynica L. *	
** (Légume stipité; feuilles à plusieurs paires de folioles.)	
1. (Grappes 2-5 flores; pédoncule beaucoup plus court que la feuille.)	
V. melanops Sibth. et Smith *	
V. pannonica Jacq. *	
V. sepium L. *	
2. (Grappes multiflores; pédoncules égalant ou dépassant la feuille; plantes vivaces.)	
V. argentea Lapeyr.	
V. altissima Desf.	
V. onobrychioides L.	
V. dumetorum L.	
Sub-genus II. Vicilla Nob.	No subgenus Vicia taxa.
Sub-genus III. Pseudervoidea Rouy	No subgenus Vicia taxa.
Sub-genus IV. Ervoidea Nob.	No subgenus Vicia taxa.
Sub-genus V. Ervum Rouy	No subgenus Vicia taxa.
Sub-genus VI. Ervilia Rouy	No subgenus Vicia taxa.

combinations to define intra subgeneric groupings. Rouy's classification approaches modern conceptions, he takes Godron's subg. Vicia species and places them in two sections, Subsessiles and Pedunculatae of his subg. Euvicia. Rouy uses a narrower conception of the V. sativa grouping, placing V. peregrina, V. lutea and V. hybrida in a separate taxon. He also reinstates the alliance between V. pyrenaica and the three included species of the V. sativa group. The inclusion of four species in sect. 2, which Kupicha considered members of subg. Vicilla, seems erroneous. V. sepium and V. dumetorum have a least a superficial similarity, but V. melanops and V. pannonica have little in common with V. argentea, V. altissima and V. onobrychioides.

Twentieth Century Classifications

Ascherson and Graebner (1909), in their classification of Vicia, present very detailed subspecific classifications for the species in their "Synopsis der Mittel-europäischen Flora" (Table 5.12). As noted by Kupicha (1974) these authors had a very confused concept of the taxonomic hierarchy above the species level. They included sections within sections and leave taxonomic groupings without a defined taxonomic ranking. For this reason the synopsis of their classification, presented in Janchen (1957), is discussed. They divide Vicia into four "sections", Ervum, Cracca, Euvicia and Faba. The subg. Vicia sensu Kupicha (1976) species are included in the later two "sections". The delimitation of Euvicia follows Visiani (1842) and is further subdivided into two "sections", Atossa and Hypechusa. These two groups are distinguished on relative hilum length and lens position; in Atossa the hilum encircled more than half the seed and the lens is near the hilum, while in Hypechusa the hilum is less than half the seed circumference and the lens is opposite the hilum. Having made this dichotomy, they then include species which do not conform to this character combination in their sections, e.g. V. lathyroides and V. sativa, V. peregrina in their "section" Hypechusa. Their fourth "section" is monotypic, containing V.

Table 5.12. Classification of sections Euvicia and Faba - Ascherson and Graebner (1909)

Section I Cracca (Tourn.) L. No subgenus Vicia taxa.

Section II Ervum (Tourn.) L. No subgenus Vicia taxa.

Section III Euvicia Vis. (Blüthen in kurzgestielten Trauben oder einzeln oder zu zweien in den Blattachseln. Griffel auf der Achse zugewandten Seite bärtig. Frucht mit angedeuteten Scheidewänden, mit lederartigen Klappen, mehrsamig. Blätter in der Knospenlage gefaltet.)

a. Atossa Alef.

1. Annuae Nyman

V. grandiflora Scop. *

2. Perennes Nyman

V. sepium L. *

V. oroboides Wulf. in Jacq. *

b. Hypechusa Alef.

1. Perennes Nyman

V. truncatula Fisch. *

V. pyrenaica Pourr. *

2. Annuae Nyman

a. (Nebenblatter klein oder ziemlich klein, ganzrandig oder gezahnt.)

V. lathyroides L. *

V. melanops Sibth. et Smith *

V. noeana Reut. in Boiss. *

V. sativa L. *

V. peregrina L. *

V. lutea L. *

V. hibrida (syn. V. hybrida L. *)

V. pannonica Crantz *

b. (Nebenblatter gross, fast stets gezahnt. Blatter nur mit 1-3 Paaren von Blattchen.)

V. bithynica L. *

V. narbonensis L. *

V. serratifolia Jacq. *

Section IV Faba Tourn.

(Früchte gedunsen mit schwammigen Querscheidewänden. Samen länglich mit endständigem Nabel. Griffel auf der Achse abgewandten Seite bärtig. - Stengel steif aufrecht. Blätter ohne Wickelranke. - Einjahrig.)

V. faba L. *

faba. This section they distinguished by the presence of the hilum at the end of the oblong seed and the legume containing spongy parenchyma between the seeds.

Following on from this work, Gams (1924) produced a classification for the illustrated "Flora von Mittel Europa". In this he returned to the pre-Ascherson and Graebner (1909) concept, using an intra-generic grouping containing all the subg. Vicia sensu Kupicha (1976) species. Thus he reunites Ascherson and Graebner's Euvicia and Faba in his subgenus Euvicia. He distinguished Euvicia from his other two subgenera on the basis of relative flower size, relative length of the peduncle and presence of stipular nectaries. However, he does not provide an intra-subgeneric classification of the subgenus or included a comprehensive species listing, so his full conception can not be appreciated.

In an innovative treatment of Vicia for the "Flore du Liban et De La Syrie", Boulomov (1930) anticipates Kupicha's classification within subg. Vicia. He is retrogressive, in that, he placed the Ervum group in a separate genus, while keeping the Vicia, Faba and Cracca groups in his Vicia. He does not give sub-generic ranking to his subgroups within Vicia but from the key (Table 5.13), it is clear that his conception of subg. Vicia is very similar to Kupicha's. As the Flora covers a relatively restricted area he included a limited number of species, but four of the five subg. Vicia (sensu Kupicha, 1976) sections are present. The fifth (sect. Atossa) is absent, but this is not surprising as the four species included in this section do not grow in the area covered by his Flora. Boulomov distinguishes his sub-generic groupings with the same characters used by previous authors; style pubescence, relative numbers of flowers per inflorescence and length of the peduncle.

Fedtschenko (1948) presents his classification of Vicia in the "Flora Of The U.S.S.R.". He divides Vicia into three

Table 5.13. Key To Vicia - Adapted From Boulomov (1930).

	Species	Kupicha Sub-generic Ranking
1. Style barbu sur sa face infér. au-dessous du sommet; fl. solitaires ou géminées, brièvement pédicellées; très rarement 2-4 en grappes axillaires.....	2	Subg. Vicia
Style portant un anneau complet de poils vers son sommet; fl. pédicellées, en grappes axillaires, très rarement solitaires ou géminées aux aisselles.....	3	Subg. Vicilla
2. Gousse de 3-5 centim. de long sur 1 de large, glabre, à suture pourvues de tubercules dentiformes, pilifères; tiges épaisses, rameuses; pétioles des feuilles supér. terminées en vrille; stipules grandes, orbiculaires; fl. grandes, d'un violet pourpre; graines globuleuses.	V. narbonensis L.* V. serratifolia Jacq.*	(Sect. Faba sensu Kupicha)
Sutures de la gousse non tuberculeuses...	4	
4. Hile opposé à la radicule.....	V. hybrida L.* V. lutea L.* V. sericocarpa Fenzl* V. galeata Boiss.* V. noeana Reut. in Boiss.*	(Sect. Hypechusa sensu Kupicha)
Hile rapproché de la radicule ou même lui étant contigu.....	5	
5. Cal. régulier; gousse ordinairement glabre.....	V. sativa L.* V. angustifolia Roth.* V. lathyroides L.* V. cuspidata Boiss.*	(Sect. Vicia sensu Kupicha)
Cal. irrégulier gousse poilue ou pubescente.....	V. peregrina L.* V. mollis Boiss. et Haussk.* V. michauxii Spreng.* V. aintabensis Boiss. et Haussk.*	(Sect. Peregrinae sensu Kupicha)

subgenera, Ervilia, Craccoidea and Faba, using primarily legume characters to distinguish these major groups. He then further subdivides subgenus Craccoidea into four sections, Lenticula, Ervum, Cracca and Euvicia, using other pod characters, relative peduncle length, number of flowers per inflorescence, flower size and leaf characters. This was a radical departure both from the preceding classifications and from the traditional use of characters to distinguish infra-generic groups. He is the first author to give prominence to legume characters, at the expense of those traditionally used. In this classification, the Faba and Vicia groups of species are found in sect. Euvicia and subg. Faba (see Table 5.14). The latter is monotypic and contains only V. faba, which is a clear statement by Fedtschenko that he believed the fababean to be quite distinct from other Vicia species.

Within section Euvicia, Fedtschenko divides the seventeen Soviet Vicia species into eleven series. He produces an obvious clustering of closely related species, but when using as many supra-specific taxa as species, it is not surprising that a natural picture is presented. The question must be asked: does the user of the classification gain sufficient extra information from such a "fine" classification to warrant its use? Kupicha (1976, p.288) makes the point that if a good classification is required, the taxonomist must balance two opposing desires:

> "variation in a single part of the plant is used as a basis for sectional grouping, then the groups are conveniently large but certainly unnatural. If, however, only convincingly natural assemblages of species (i.e. groups that share several traits) are accepted as sections, a too finely divided system is produced."

It appears that Fedtschenko adopted the latter approach.

Following a detailed morphological study of 18 Vicia species, selected to reflect the extremes of variation pattern in the genus, Kiffman (1952) divides the genus into five subgenera (Ervilia, Ervum, Cracca, Euvicia and Faba). Kiffman clearly identifies the major specific groupings in Vicia and

Table 5.14. Classification of Vicia and Faba groups - Fedtschenko (1948)

Note: several taxa below are attributed to Fedtschenko, but he did not actually publish these names, they were later legitimately published by Radzhi (1971b).

Subgenus I Ervilia Link. (Pods linear, slightly compressed, moniliformly attenuate. Rachis not terminating in cirrus.)
No subg. Vicia species

Subgenus II Craccoidea B. Fedtsch. (Pods conspicuously compressed laterally, without internal septa. Rachis usually terminating in tendril, usually branching. Flowers axillary or in racemes at ends of peduncles.)
Section 1 Lenticula B. Fedtsch. (Annuals. Leaves terminating in branching tendril. Leaflets narrow, linear. Flowers small, long pedunculate, solitary or few. Pods short, usually rhombic; seeds 2-3.)
No subg. Vicia species
Section 2 Ervum (Annuals. Leaves ending in branched tendrils; leaflets narrow, linear, sometimes elliptic; flowers small, few, on long peduncles; pods linear, 4-6-seeded)
No subg. Vicia species
Section 3 Cracca (Perennials, rarely annuals, with flowers larger and [excluding V. calcarata] more numerous than in sections Lenticula and Ervum, borne on tip of peduncles.)
No subg. Vicia species
Section 4 Euvicia Vis. (Annuals, rarely perennials. Flowers solitary or in pairs, subsessile in axils of leaves, rarely 3-5 on very short peduncles.)
Series 1 Sepium Buchenau (Perennials; flowers in dense raceme on short peduncles; hilum two-thirds to three-fourths the circumference of the seed.)
V. sepium L. *
Series 2 Truncatulae B. Fedtsch. (Perennials; leaves terminating in tendril or awn; peduncles very short; flowers yellow in racemes; hilum up to one half the circumference of seed.)
V. truncatula Fisch. *
V. balansae Boiss. *
Series 3 Lathyroides Buchenau (Strongly branched annuals; flowers small; seed verrucose.)
V. lathyroides L. *
V. olbiensis Reuter (syn. V. lathyroides L. *)
V. saxatilis (Vent.) Tropea (syn. Lathyrus saxatilis)

Series 4 Grandiflorae B. Fedtsch. (Annuals or biennials. Flowers axillary, 1(2-3), yellow. Radix dorsal, its end turned away from the placenta. Hilum two-thirds to four-fifths of the seed circumference.)
V. grandiflora Scop. *

Series 5 Sativae B. Fedtsch. (Annuals. Flowers axillary, lilac or pink. calyx teeth subequal. Hilum one-eighth to one-third of seed circumference.)
V. sativa L. *
V. incisa M.B. (syn. V. sativa L. *)
V. cordata Wulf. in Sturm (syn. V. sativa L. *)
V. angustifolia L. (syn. V. sativa L. *)
V. pilosa M.B. (syn. V. sativa L. *)
V. amphicarpa Dorthes (syn. V. sativa L. *)

Series 6 Hyrcanicae B. Fedtsch. (Annuals. Flowers yellowish, whitish. Calyx teeth unequal. Hilum up to three-eighths the circumference of seed.)
V. hyrcanica Fischer et Meyer *

Series 7 Peregrinae B. Fedtsch. (Annuals. Rachis ending in a tendril or awn. Flowers purple-violet, whitish or yellowish; calyx oblique, lower teeth longer than the upper.)
V. peregrina L. *
V. gracilior M. Pop. (syn. V. peregrina L. *)
V. megalosperma M.B. (syn. V. peregrina L. *)
V. michauxii Spreng. *

Series 8 Luteae B.Fedtsch. (Annuals. Flowers yellow; calyx obliquely truncate. Pods hirsute, hairs tubercled at base.)
V. lutea L. *

Series 9 Hybridae B. Fedtsch. (Annuals. Lower calyx teeth longer than upper; standard outwardly pubescent; pods glabrous or pubescent but hairs not tubercled.)
V. hybrida L. *
V. hajastana Grossh. (syn. V. anatolica Tur. *)
V. pannonica Crantz *
V. ciliatula Lipsky *

Series 10 Bithynicae B. Fedtsch. (Stipules wide, usually large-toothed; leaflets 1-3-paired.)
V. bithynica L. *

Series 11 Narbonenses B. Fedtsch. (Annuals, and biennials, with firm erect stems; leaflets large, wide, 1-3-paired. Flowers axillary. Pods scabrous-hairy, hairs often on tubercles.)
V. narbonensis L. *
V. serratifolia Jacq. *

Subgenus III Faba (Adans.) Gray (Annuals. Flowers axillary; pods slightly inflated, nearly cylindrical, with incised septa between seeds. Seeds oblong, hilum apical. Leaves always atendrilous.)
V. faba L. *

splits the subg. Vicia sensu Kupicha (1976) species that he investigated into subg. Euvicia and subg. Faba:

Euvicia	Faba
V. sepium	V. narbonensis
V. angustifolia	V. faba
V. pannonica	
V. peregrina	
V. lutea	
V. grandiflora	

It is difficult to draw extensive conclusions from Kiffman's classification, because of the limited number of species known at that time and included in his survey. However, it is interesting to note his grouping of V. narbonensis with V. faba away from Euvicia, which marks a return to the nineteenth century conceptions of Visiani (1842), Godron (1849) and Taubert (1894).

The annual Vicia species of the "Middle East" are extensively reviewed by Plitmann (1967). He divides Vicia into six sections, Faba, Vicia, Ervum, Trigonellopsis, Anatropostylia and Cracca. The Vicia and Faba groups of species are restricted, as one might expect from their names, to the first two sections (Table 5.15). As with Fedtschenko's classification, Plitmann proposes a very detailed classification, using ten supra-specific taxa to cluster 30 species. There are two problems in his classification: the inclusion of V. montbretii, which is clearly not a natural member of the subg. Vicia and the exclusion of V. esdraelonensis from sect. Vicia. Referring to the latter, he commented, that the authors of this species allied it to either the V. peregrina or the V. hyrcanica groups. Plitmann, however, believed it was a natural ally of V. cretica in sect. Cracca. Although he later altered his opinion (Davis and Plitmann, 1970), returning the species to the V. hyrcanica group. The characters used to distinguish groups within his sections Faba and Vicia are those that had proved useful through out the history of the group.

Table 5.15. Classification of Vicia and Faba Groups - Plitmann (1967)

(With the exception of V. montbretii all species are members of Subg. Vicia sensu Kupicha 1976)

Section I Faba Aschers. et Graebn. (1-3 pairs, large leaflets. Flowers and seed large. Pods broad pubescent, ciliate or glandular. Calyx regular. Non-climbers.)

Series 1. Fabae (1-3 pairs of leaflets, flowers large, calyx regular. Style compressed, apex with tuft hairs.Pods large, straight edged, mostly pubescent, ciliate or glandular)

1a. Bithynica
V. bithynica L.

1b. Narbonensis
V. narbonensis L.
V. galilaea Plitm. et Zoh.
V. hyaeniscyamus Mout.

1c. Faba
V. faba L.

Section II Vicia (Leaves =or> 3 pairs leaflets, medium or small. Flowers medium, 1.5-2.0 cm. Pods < 1.3 cm. Calyx regular or irregular. Mostly climbers.)

Series 2. Hyrcanicae (Annuals with peduncles. Calyx irregular. Corolla medium to large. Pods glabrous, edges non-parallel.)
V. hyrcanica Fischer et Meyer
V. assyriaca Boiss.
V. galeata Boiss.
V. melanops Sibth. et Smith
V. noeana Reut. in Boiss.

Series 3. Peregrinae (Annuals, narrow leaflets. Peduncle absent. Calyx irregular, pubescent. Corolla medium, purple to yellow. Pods pubescent or not, edges non-parallel.)
V. peregrina L.
V. aintabensis Boiss. et Haussk.
V. michauxii Spreng.

Series 4. Luteae (Annuals. Calyx gibbous, glabrous. Corolla large, yellow or purple. Pods pubescent, edges non-parallel.)
V. lutea L.

Series 5. Sericocarpae (Annual. Calyx gibbous, densely pubescent. Peduncle short. Corolla yellow. Pods dense adpressed pubescent, edges non-parallel.)
V. sericocarpa Fenzl
V. bombycina Stapf ex Post (syn. V. montbretii Fischer & C. Meyer)
V. mollis Boiss. et Haussk.

V. camptopoda Townsend (syn. V. mollis Boiss. et Haussk.)

Series 6. Hybridae (Annuals. Calyx gibbous, pubescent. Peduncle short. Corolla usually yellow, standard pubescent on upper surface. Pods pubescent, edges non-parallel.

V. hybrida L.
V. anatolica Turrill
V. ciliatula Lipsky
V. hajastana Grossh. (syn. V. anatolica Turrill)
V. pannonica Crantz

Series 7. Sativae (Annuals & perennial. Peduncle mostly absent. Calyx regular. Style compressed or and bearing tuft hairs near apex. Pods large, edges parallel, many seeded.)

7a Lathyroides
V. lathyroides L.
V. cuspidata Boiss.
7b Sativa
V. sativa L.
V. barbazitae Ten. et Guss.
V. grandiflora Scop.
V. pyrenaica Pourr.

Section III Ervum (L.) Gray
No subgenus Vicia taxa.

Section IV Trigonellopsis Rech.
No subgenus Vicia taxa.

Section V Anatropostylia Plitm.
No subgenus Vicia taxa.

At the same time as Plitmann was completing his Ph.D. research, Ball (1968) was compiling his account of Vicia and other legume groups for the "Flora Europaea". Ball does not attempt such a detailed intra-generic ranking of species as Plitmann, but presents similar conclusions. He divides Vicia into four sections; Cracca, Ervum, Vicia and Faba. As with Plitmann's classification, the Vicia and Faba groups of species are restricted to sections Vicia and Faba (see Table 5.16). The characters used to distinguish the two sections are those established by preceding authors. Within his sect. Vicia he does not present a clear pattern of specific relationships, e.g. he places V. pyrenaica with the sect. Atossa sensu Kupicha (1976) species, not with its generally accepted allies, the V. sativa complex. This alliance was previously suggested by Ascherson and Graebner (1909), but not taken up by subsequent authors. Ball places V. pannonica in the centre of the V. sativa complex, rather than in its established natural position with the other sect. Hypechusa sensu Kupicha (1976) species; a view not adopted by previous or subsequent authors.

The treatment of Vicia for the "Flora of Turkey and the East Aegean Islands" (Davis and Plitmann, 1970) follows Plitmann (1967), although perennial and recently described species are added. V. truncatula, V. balansae and V. sepium are added as a perennial unit at the beginning of sect. Vicia. V.esdraelonensis is placed with its natural allies in ser. Hyrcanicae sensu Plitmann (1967) and not in sect. Cracca as suggested by Plitmann (1967).

In the first of a series of Soviet accounts of Vicia, Stankevich (1970) presents a phylogenetic classification of the Vicia species of the Soviet Union. She considers Fedtschenko's (1948) classification to be inaccurate and to contain many specific names which do not represent acceptable taxa. She uses pod and flower shape and dimension characters, with the numbers of pairs of leaflets to split

Table 5.16. Classification of Vicia and Faba groups - Ball (1968)

(All species are members of subg. Vicia sensu Kupicha 1976)

Section 1. Cracca S.F. Gray
No subgenus Vicia taxa.

Section 2. Ervum (L.) S.F. Gray
No subgenus Vicia taxa.

Section 3. Vicia (Leaflets usually more than 3 pairs; flowers solitary, axillary or in few-flowered, sessile or shortly pedunculate racemes; corolla usually large (more than 10 mm); style pubescent on the lower side beneath the stigma.)
V. oroboides Wulf. in Jacq.
V. truncatula Fisch.
V. pyrenaica Pourr.
V. sepium L.
V. grandiflora Scop.
V. barbazitae Ten. et Guss.
V. pannonica Crantz
V. sativa L.
V. lathyroides L.
V. cuspidata Boiss.
V. peregrina L.
V. melanops Sibth. et Smith
V. lutea L.
V. hybrida L.

Section 4. Faba (Miller) S.F. Gray (Leaflets 1-3 pairs; flowers solitary, axillary or in few-flowered, sessile or shortly pedunculate racemes; corolla large (more than 10 mm); style pubescent on the lower side beneath the stigma.)
V. bithynica L.
V. narbonensis L.
V. faba L.

Vicia into three sections (see Table 5.17). She moves away from the contemporary view; removing V. faba to the separate genus, Faba, and reinstates the 'oroboid' species as a distinct sub-group of Vicia. Apart from these points the classification is similar to that of Kupicha (1976). Stankevich comments that her sect. Vicia includes the most evolutionary advanced Vicia species. The character states which she considers to reflect evolutionary advancement, are the presence of reduced numbers of flowers, shorter peduncle and seed hilum, smaller area of lamina per leaf and reduced chromosome number.

Stankevich's sect. Vicia, includes the Vicia and Faba groups of species in one taxon, remote from other Vicia taxa, as does Kupicha. She includes three sub-sections. Her subsect. Laticarpa Stankev. is equivalent to sect. Faba sensu Kupicha (1976) with the exclusion of V. faba itself; her subsect. Vicia is synonymous with Kupicha's conception; her third subsect. Brevicarpa Stankev. contains sections Hypechusa Peregrinae and Atossa sensu Kupicha (1976). Unlike Kupicha, she follows Fedtschenko (1948) and emphasises pod shape and dimension characters to separate her subsections of Vicia. Although she uses the established characters of leaflet size, peduncle length and numbers of flowers per inflorescence, she does not, however, use stipular nectaries as a distinguishing character.

A year later, Radzhi (1971a) suggested a far more traditional classification of the Vicieae in which she reinstated Ervum and Orobus as genera distinct from Vicia. Radzhi (1971b) provided an equally traditional account of the Caucasian species of Vicia. She proposed three subgenera and a detailed hierarchical sub-generic structure (see Table 5.18). She uses and validly publishes a lot of the categories and names suggested, but not validly published by Fedtschenko (1948). She describes her subgenus Vicia species as erect or ascending plants, shorter than 50 cm, with solitary or 2(4)

Table 5.17. Classification of Vicia and Faba groups - Stankevich (1970)

(All species are members of subg. Vicia sensu Kupicha 1976) V. faba L. is excluded from Vicia and is placed in the monotypic genus Faba as F. vulgaris.

Section Vicia (Reduced number of flowers and peduncle and hilum length, smaller lamina area per leaf and lower chromosome base number)

Sub-section 1 Laticarpa Stankev. (Legume wide and linear)
- V. narbonensis L.
- V. serratifolia Jacq.
- V. bithynica L.

Sub-section 2 Vicia (Legume narrow and linear)
- V. sativa L.
- V. incisa M.B. (syn. V. sativa L.)
- V. grandiflora Scop.

Sub-section 3 Brevicarpa Stankev. (Legume short, rhomboid; calyx oblique)
- V. hyrcanica Fischer et Meyer
- V. hybrida L.
- V. pannonica Crantz
- V. hajastana Grossh. (syn. V. anatolica Turrill)
- V. ciliatula Lipsky
- V. peregrina L.
- V. michauxii Spreng.
- V. sepium L.
- V. balansae Boiss.
- V. truncatula Fisch. et Mey.

Section Cracca Vis.
No subg. Vicia taxa

Section Oroboidea Stankev.
No subg. Vicia taxa

Table 5.18. Classification of Vicia and Faba groups - Radzhi (1971b)

(All species are members of subg. Vicia sensu Kupicha 1976)

Subgenus Cracca (Medic.) Gams
No subg. Vicia taxa

Subgenus Ervum (L.) Gams
No subg. Vicia taxa

Subgenus Vicia
Section 1 Sepium Radzhi (Plants perennial, or mesophytic, wooded regions. Stems usually straight. racemes axillary, 4-6-flowered. Stipules small, 3-5 mm long. Legume glabrous. Hilum two-thirds seed circumference.)
Subsection 1 Sepium (Buchenau) Radzhi
V. sepium L.
Subsection 2 Truncatulae Radzhi
V. truncatula Fisch.
Section 2 Vicia
Subsection 1 Vicia Radzhi
Series 1 Grandiflorae Radzhi
V. grandiflora Scop.
Series 2 Sativae (Buchenau) B. Fedtsch.
V. sativa L.
Series 3 Lathyroides (Buchenau) B. Fedtsch.
V. lathyroides L.
Subsection 2 Bithynicae Radzhi
Series 1 Bithynicae Radzhi
V. bithynica L.
Series 2 Paucijugae Radzhi
V. paucijuga (Trautv.) B. Fedtsch. (syn. V. cappadocica Boiss. & Bal.)
Subsection 3 Peregrinae Radzhi
V. peregrina L.
Subsection 4 Hybridae Radzhi
Series 1 Luteae Radzhi
V. lutea L.
Series 2 Hyrcanicae Radzhi
V. hyrcanica Fischer et Meyer
Series 3 Hybridae Radzhi
V. hybrida L.
V. hajastana Grossh. (syn. V. anatolica Turrill)
V. pannonica Crantz
V. ciliatula Lipsky
Section 3 Faba Ledeb.
Subsection 1 Narbonensis Radzhi
V. serratifolia Jacq.
V. narbonensis L.
Subsection 2 Faba Radzhi
V. faba L.

flowers per inflorescence, the flowers are either sessile or auxiliary, the peduncle is rarely longer than the leaf and the stipule has a dark coloured extra-floral nectary. Like Fedtschenko (1948), whose scheme she follows closely, she uses almost as many taxonomic categories as she includes species, which makes Vicia seem unnecessarily disjointed.

Townsend's (1974) account of Vicia for the "Flora of Iraq" divides the genus into six sections, both the Vicia and Faba groups attaining sectional rank. His presentation of subg. Vicia sensu Kupicha essentially follows Ball (1968). He (Townsend, 1967), questions the specific status of V. noeana, a point that will be discussed in Chapter 9. Chronologically the next account of Vicia is that presented by Kupicha (1974) and published as Kupicha (1976), but this has been discussed above in thesis section 5.2.1.

Post Kupicha Classifications

For the "Flora Iranica", Chrtkova-Zertova (1979) excluded V. faba from Vicia and placed it in the monospecific genus Faba as F. vulgaris. However, she retained the other species associated with V. faba, V. narbonensis and V. serratifolia within her section Vicia (Table 5.19), so, in effect, she unified the Vicia and Faba groups (except for V. faba itself). Her conception returns to that suggested by Ascherson & Graebner (1909). She presents a radical classification, allying species not previously considered closely related, but unfortunately she does not justify her position. For example she places the perennial V. truncatula between the relatively closely related annuals V. anatolica and V. hyrcanica with which it shares no obvious close relationship. She also disperses, throughout her classification, the three closely allied species of sect. Peregrinae sensu Kupicha, V. michauxii, V. aintabensis and V. peregrina. On the positive side, she retains the V. sativa complex of species as one

Table 5.19. Classification of Vicia and Faba groups - Chrtkova-Zertova (1979)

(All species are members of subg. Vicia sensu Kupicha 1976)

Vicia L.

Section 1 Ervilia (Link) W.D.J. Koch
No subg. Vicia taxa

Section 2 Ervum L.
No subg. Vicia taxa

Section 3 Cracca L.
No subg. Vicia taxa

Section 4 Anatroppostylia Plitmann
No subg. Vicia taxa

Section 5 Vicia (Annuae, rarius perennes. Legumen inter semina non vel paulo tantum constrictum. Flores solitarii, subsessiles vel in racemes brevipedunculatis paucifloris dispositi. Stipulae glandulosae. Stylus extus infra stigma fasciculo pilorum provisus.)
V. hybrida L.
V. pannonica Crantz
V. anatolica Turrill
V. truncatula Fisch.
V. hyrcanica Fischer et Meyer
V. michauxii Spreng.
V. mollis Boiss. et Haussk.
V. assyriaca Boiss.
V. aintabensis Boiss. et Haussk.
V. sericocarpa Fenzl
V. lutea L.
V. hirta Balbis ex DC. (syn. V. lutea L.)
V. peregrina L.
V. sativa L.
V. angustifolia L. (syn. V. sativa L.)
V. amphicarpa Lam. (syn. V. sativa L.)
V. grandiflora Scop.
V. lathyroides L.
V. cuspidata Boiss.
V. narbonensis L.

Faba Miller
(Plantae annuae. Caules erecti, crassi. Foliorum rhachis cuspide provisa vel deficiente. Foliola (1-)2(-3)-juga, magna. Pedunculi breves. Inflorescentia (1-)4(-5)-flora. Corolla sordide alba; alae carina atro-coeruleo-vel coeruleo-viola ceo-maculatae. Legumen suturis parallelis spongiosis. Semina magna, late oblongo-reniformia.)
F. vulgaris Moench

entity and uses the presence of a stipular gland to define her sect. Vicia, the character Kupicha (1976) had shown to be so useful.

In 1979 Bueno Perez undertook, for her doctoral thesis, a numerical investigation of 34 Vicia species, using morphological and cytological characters. In conclusion she believed it would be premature to present a novel classification of Vicia, but she makes certain suggestions which require further development. She suggests uniting Vicia sensu Kupicha (1976) with Lens (on the basis of shared seed characteristics between V. ervilia and L. culinaris). She suggests splitting Vicia into three subgenera, Faba, Euvicia and Lens. Interestingly, she found that V. faba was quite distinct from other Vicia using both morphological and cytological data. Thus she questions whether V. faba should be split into a separate subgenus or even genus Faba. She subdivides Euvicia into three sections, Ervum, Craccoidea and Vicioidea, the latter is defined as species with either a short peduncle or sessile flowers and with a few flowers per inflorescence. Within Vicioidea, she suggests five "lines":

a) Line sativa - flowers sessile, 2n = 10 or 12, short chromosomes with frequent satellites.
b) Line arvense - flowers solitary or with short peduncle, few flowers per inflorescence, 2n = 10, 12 or 14, long chromosomes frequently in metacentric pairs.
c) Line narbonensis - long and wide leaflets, flowers solitary, 2n = 14.
d) Line bithynica - long and narrow leaflets flowers solitary, 2n = 14.
e) Line grandiflora - ovary with many ovules, 2n = 14.

The impact of Bueno Perez's proposals are limited by the inclusion of a restricted number of Vicia species in her study and by the fact that the species included do not fully represent the extent of variation encountered in Vicia. She does not attempt to list which species she would include in each of her taxa, possibly because of the limited number of taxa she studied. However, her section Vicioidea is distinguished using similar characters to subg. Vicia sensu

Kupicha (1976) and so would contain similar taxa. Interestingly, she, like Fedtschenko (1948), excluded V. faba from her sect. Vicioidea. As she does not detail circumscribed taxa it is difficult to compare her "lines" with other authors' groupings, but the names used and their defining characteristics can be seen to indicate correspondence with sections in Kupicha's classification.

The next in the series of Soviet classifications of Vicia, was presented by Tzvelev (1980) for the Vicia of Soviet Europe (Table 5.20). Tzvelev suggests similar species clusters to previous Soviet taxonomists. His classification is closer to that proposed by Fedtschenko (1948), than that of Radzhi (1971b), especially concerning the appropriate taxonomic rank to give to the different clusters of species. He returns V. faba to Vicia and splits the genus into four subgenera, Vicilla, Ervum, Vicia and Faba, the Vicia and Faba groups of species are placed in the latter two groups. Tzvelev, unlike Fedtschenko, does ally V. faba to V. narbonensis and places both species in his subg. Vicia. He compliments Kupicha (1976) on her classification, but does not believe V. bithynica to be naturally allied to V. faba. Therefore he places V. bithynica and V. faba in the same subgenus, but separates V. bithynica into a new section (Pseudolathyrus Tzvel.). Like the classification presented by Stankevich (1970), Tzvelev does not provide an over "fine" hierarchy and presents groups of species similar in composition to Kupicha's sections Atossa, Hypechusa, Vicia and Faba. Tzvelev's classification is the closest of all the Soviet botanists to Kupicha's.

In a regional classification of Soviet Vicia, Avazneli (1981) divides the Georgian Vicia into five sections (Table 5.21). The Vicia and Faba groups of species are found in two sections, Vicia and Faba. As with the previous Soviet classifications of Vicia, Avazneli suggests a new hierarchy, but within sections Vicia and Faba, the species are aligned in

Table 5.20. Classification of Vicia and Faba groups - Tzvelev (1980)
(All species are members of subg. Vicia sensu Kupicha 1976)

Subgenus 1 Vicilla (Schur) Rouy
No subg. Vicia species
Subgenus 2 Ervum (L.) Rouy
No subg. Vicia species
Subgenus 3 Vicia (Flowers clustered into auxiliary racemes, 2-6 flowers on very short peduncles, up to 8 mm. long, sometimes solitary, inflorescence axis terminating in a flower, calyx mouth regular or irregular, corolla (5)7-30(35) mm., style at right angles to the axis, style round, hairs on abaxial side only, mostly annuals, stipules with extra floral nectaries.)
Section 7 Sepium (Buchenau) Radzhi
V. sepium L.
Section 8 Hypechusa (Alef.) Ascher. et Graebn. ex Kupicha
V. pannonica Crantz
V. striata Bieb. (syn. V. pannonica Crantz)
V. hybrida L.
V. anatolica Turrill
V. lutea L.
V. ciliatula Lipsky
V. peregrina L.
Section 9 Vicia
V. grandiflora Scop.
V. sativa L.
V. incisa Bieb. (syn. V. sativa L.)
V. cordata Wulf. ex Hoppe (syn. V. sativa L.)
V. segetalis Thuill. (syn. V. sativa L.)
V. angustifolia L. ex Reichard (syn. V. sativa L.)
V. pilosa Bieb. (syn. V. sativa L.)
V. amphicarpa Dorthes (syn. V. sativa L.)
Section 10 Lathyroides (Buchenau) Tzvel.
V. olbiensis Reut. ex Timb.-Lagr. (syn. V. lathyroides L.)
V. lathyroides L.

Subgenus 4 Faba (Mill.) Peterm. (Flowers clustered into auxillary racemes, 2-6 flowers on short peduncles, up to 30 mm. long, inflorescence axis terminating in a flower, calyx mouth regular, corolla 13-25(30) mm., style at right angles to the axis, style round, hairs on abaxial side only, annuals, stipules with extra floral nectaries.)
Section 11 Pseudolathyrus Tzvel.
V. bithynica L.
Section 12 Faba (Mill.) Ledeb.
V. narbonensis L.
V. faba L.

Table 5.21. Classification of Vicia and Faba groups - Avazneli (1981)

(All species are members of subg. Vicia sensu Kupicha 1976)

Section 1 Cracca S.F. Gray
No subg. Vicia species

Section 2 Ervilum S.F. Gray
No subg. Vicia species

Section 3 Vicia
V. truncatula Fisch.
V. balansae Boiss.
V. sepium L.
V. ciliatula Lipsky
V. peregrina L.
V. lutea L.
V. pannonica Crantz
V. grandiflora Scop.
V. lathyroides L.
V. sativa L.
V. cordata Wulf. in Sturm (syn. V. sativa L.)
V. angustifolia L. (syn. V. sativa L.)
V. cinerea Bieb. (syn. V. sativa L.)

Section 4 Faba Aschers. et Graebn.
V. bithynica (L.) L.
V. narbonensis L.
V. serratifolia Jacq.
V. faba L.

Section 5 Variegatae Radzhi
No subg. Vicia species

a similar way to that suggested by Tzvelev. Within sect. Faba, however, Avazneli goes further and includes V. bithynica, V. narbonensis, V. serratifolia and V. faba in one section, as did Kupicha (1976). This paper is written in Georgian and no translation is available so that a thorough interpretation of Avazneli's conception is difficult.

In 1982 Stankevich radically revised her 1970 classification of Vicia. She takes the Ervum and Ervilia groups of species and concludes that they deserve generic rank, a divisive idea not used since Alefeld (1861a & b). She argues on the basis of a set of morphological and distributional characters that Vicia, as she previously conceived it, should be split into five genera, Ervum, Ervilia, Bona, Faba and Vicia. She believes Vicia to be characterised as: herbs of open spaces; with a weak prostrate or ascending stem; the leaves are paripinnate with a well developed tendril; the leaflets are small, elongate and or elliptic; the style is filiform with hairs distally. This definition excludes the species of the V. narbonensis complex, so she resurrects Bona to contain these species. She concludes that Bona and Vicia are not a monophyletic grouping, they have arisen from different ancestral forms at different times and that the similarity between these groups is superficial.

Stankevich (1982) does not include any specific reference to Faba vulgaris. Stankevich (1983) clarified the position by including this species in Faba and separating the two species, she recognised, of the V. narbonensis complex in Bona, as B. narbonensis and B. serratifolia. She concluded that although Bona is superficially close to Faba in the structure of the stem, leaves and flowers, the pod of Faba is inflated, fleshy and softly pubescent, whereas in Bona it is leathery and covered in tubercular hairs. She also notes that the seeds and karyotypes of these two genera are quite distinct. In these later papers Stankevich uses a much narrower conception

of the Vicieae genera than "Western" botanists or even other Soviet botanists. She included ten genera within her tribe Fabeae: Faba Mill., Bona Medik., Orobus L., Vavilovia Fed., Pisum L., Vicia L., Lathyrus L., Ervum L., Lens Mill. and Ervilia (L.) Link.

Another Soviet taxonomic system for Vicia is proposed by Nikiforova (1985) for the Siberian region of the USSR. She does not follow the restricted generic view of Stankevich (1982, 1983) and reunites the Ervum, Ervilia and Vicia groups into one genus. Nikiforova comments that she favours the generic conception of Vicia used by Kupicha and this, in general terms, is reflected in her classification. However, only one subg. Vicia sensu Kupicha (1976) species is present in Siberia (V. sepium) and so she makes no attempt to produce a subg. Vicia classification.

During the final editing process of this thesis, Hanelt & Mettin (1989) published a review of the classifications and taxonomy of Vicia. They commented that the most acceptable classification of Vicia is provided by Kupicha (1976). However, they suggest one alteration; that V. bithynica should be separated from sect. Faba sensu Kupicha (1976) into the monotypic section Pseudolathyrus Tzvel.

5.3 Summary of Subgenus Vicia Taxonomic History

As the taxonomic history of subg. Vicia is extensive, there being 20 major classifications discussed above, a short summary of the themes and developments of taxonomic groupings, authors' conceptions and the characters used in distinguishing the subg. Vicia will follow.

Major taxonomic groupings

The similarity between the classifications, discussed above, should be emphasised. Certain species groupings are met throughout the taxonomic history of the genus e.g. the close linking of V. sativa, V. barbazitae, V. pyrenaica, V. grandiflora, V. lathyroides and V. cuspidata (sect. Vicia sensu Kupicha, 1976) which clearly indicates a natural grouping. In more general terms, the taxonomic history has shown how the "Vicia" and "Faba" groups of species distinguished by Kupicha (1976) tend to be linked more closely in the taxonomic hierarchy to each other, than to either of the other two groups she identified, Ervum and Cracca.

Specifically, the "Faba" group is a less cohesive unit than Vicia and as a result "Faba" has had a more controversial history. Opinions about the placement of V. faba have been especially erratic. Some authors believing it should be regarded as a separate genus, (Seringe, 1825; Alefeld, 1861, Stankevich 1970, 1982), a separate subgenus (Fedtschenko, 1948; Bueno Perez 1979), a separate section (Ascherson & Graebner, 1909; Ball, 1968; Chrtkova-Zertova, 1979), a separate subsection (Koch, 1836; Godron, 1849; Taubert, 1894; Radzhi, 1971b), a separate subseries (Plitmann, 1967) or regarded on the same level as certain other Vicia species (Linnaeus, 1753; Visiani, 1842; Kupicha, 1976).

It is perhaps not surprising that a crop plant like V. faba, without any obvious close progenitor should have such a chequered taxonomic history, but another species in the "Faba" group, V. bithynica, has also proved problematic to place.

Most commonly it is allied with the other "Faba" group species but it was originally placed in Lathyrus by Linnaeus (1753) and was placed with the Cracca group (by Visiani, 1842 and Boissier, 1872) due to its extended peduncle. The taxonomic background of V. narbonensis has not proved so problematic for when not linked to V. faba and V. bithynica it has been allied to the perennial species of sect. Atossa sensu Kupicha (1976).

Within the "Vicia" group of species Kupicha (1976) places the perennial species, except for V. pyrenaica (the close ally of V. sativa), in one section, Atossa. Boissier (1872) originally grouped these four species by placing them in one subsection. Similarly other authors link two or three of these species, but as few authors include all four species in their classification, it is difficult to judge the naturalness of their relationship.

Another subgroup often linked within the "Vicia" group is composed of V. peregrina, V. aintabensis, and V. michauxii. These three species are the only ones to lack a peduncle and are obviously closely related. This led Ponert (1973) to regard the latter two taxa as subspecies of V. peregrina. Kupicha like Boissier (1872) and Boulomov (1930) links V. mollis with these species in sect. Peregrinae, but more recently this species has only occasionally been allied to the V. peregrina complex. V. mollis has been linked to sect. Hypechusa, specifically to V. sericocarpa (Plitmann, 1967) and V. assyriaca Chrtkova-Zertova (1979).

The remaining species of the "Vicia" grouping largely come from sect. Hypechusa sensu Kupicha (1976). The taxonomic history shows the species tend to cluster together rather than forming links with the sectional groups described by Kupicha. Rather than the whole of sect. Hypechusa sensu Kupicha being commonly found, subgroups of species are more commonly seen, e.g. V. hybrida, V. pannonica and V. anatolica or V. hyrcanica, V. galeata and V. noeana. There may be a

particular problem in trying to interpret the history of this section as most of the included species are relatively rare and of restricted distribution.

Authors' conceptions

If the authors' conceptions of subg. Vicia are considered, then it can be seen that certain authors adhere more to the general patterns established above than others. It is interesting that the first author to present a subgeneric classification of Vicia, Linnaeus (1753), divides the genus into two subgroups, which match Kupicha's subgenera. This division of Vicia on the basis of relative peduncle length has remained the dominant theme in the subgeneric classification. The next important step was taken by Koch (1836), who further sub-divided the short-peduncled or sessile Vicia (subg. Vicia sensu Kupicha). Koch separated out from subg. Vicia, V. faba and V. bithynica the two species that have remained peripheral within subg. Vicia. Godron's (1849) classification, although producing some radical juxtapositions (i.e. V. sepium and V. pannonica allied to V. argentea and V. onobrychioides respectively), is the first author to unite the sect. Faba sensu Kupicha (1976) species as one distinct unit.

Alefeld's contribution to Vicia taxonomy can not be under-rated. The basic groups of species he identified (Alefeld, 1860b & c) are those still recognised today. Although the taxonomic ranking he uses may currently seem eccentric, his use of 'good' characters is perceptive and it was this that enabled him to identify natural groupings. Boissier's (1872) classification, in essence, takes the groupings distinguished by Alefeld, reinstates the broader concept of Vicia and provides a more natural taxonomic ranking of subgeneric species clusters. Boulomov's classification of Vicia for the Flora of Lebanon and Syria in 1930 is advanced and perceptive both in his use of characters to define

subgroups and in the subgroups defined which mirror those later erected by Kupicha (1976).

Like Alefeld, Fedtschenko and later Soviet classifications are probably natural, but they subdivide Vicia to such an extent, that the classifications becomes less predictive. Kupicha's (1976) overall conception is masterful, because, as has been established in the taxonomic history, it can be read as a culmination of previous classifications while at the same time being species comprehensive.

Characters used to distinguish subg. Vicia groupings

Throughout the taxonomic history of Vicia certain characters have been repeatedly used to define and distinguish subgeneric groupings. These characters may be divided into two kinds, those used to define major divisions (i.e. subgenera or sections) and those used to distinguish subdivisions within the previously established major divisions.

There are a few characters that are commonly used to distinguish the major subgeneric groupings within subg. Vicia. These are used by almost all authors. They are: presence of peduncle and, if present, the length of the peduncle relative to either flower or subtending leaf size, the number of flowers per inflorescence, presence of stipular nectaries and the positioning of hairs around the style.

Once the major groups have been established many more characters are used to produce more restricted subgroups. These are: life span; plant stature; number of flowers per inflorescence; number of leaflets per leaf; size of leaflets; calyx mouth shape; flower colour; vexillum shape; vexillum pubescence; legume pubescence; legume protrusion; relative legume size; relative legume shape; legume suture curvature; legume hairs type; seed shape; relative hilum length; relative position of the seed hilum and lens; and seed surface type.

CHAPTER SIX
METHODS OF DATA GATHERING AND ANALYSIS

6.1 Introduction

Once the taxon to be revised has been delimited, there follows a series of skilful taxonomic decisions associated with the choice of data gathering and analysis methods. There are three basic decisions to be made: the choice of the appropriate level for operational taxonomic units (OTUs), the choice of suitable characters for use in the analysis, and the choice of the methods of data analysis to be used. The purpose of this chapter is to provide each of these three points with, firstly, an introduction to their theoretical background and, secondly, an explanation of the choices made during this study.

The third decision, choice of data analysis method, can be subdivided into two parts, because of the fundamental difference between phenetic and phylogenetic or cladistic analysis. Phenetic techniques use the occurrence of character combinations as they are now perceived to construct taxonomic groups, while phylogenetic techniques add the dimension of time and ancestry to describing organisms (definitions taken from Dunn and Everitt, 1982, and Davis and Heywood, 1963). Having outlined the essential difference between these two methods of analysis, only the phenetic approach was used during this study. Phylogenetic methods do have a vital innovative part to play in contemporary taxonomy, but there was insufficient time to apply both phenetic and phylogenetic methods rigorously, especially as the data analysis of the taxonomic element of the research programme formed only part of the overall project. However, a discussion of possible phylogenetic trends in subg. Vicia is provided in thesis section 9.4 and in the revision conspectus provided in Appendix 7.

This chapter is not intended as a justification of the underlying philosophies of the phenetic techniques to be employed. Surely twenty seven years after the publication of Sokal and Sneath's (1963) 'Principles of Numerical Taxonomy' the case for the use of phenetic methods has been extensively validated.

6.2 Operational Taxonomic Units

6.2.1 Theoretical Foundation

The fundamental taxonomic units employed in computer aided taxonomy are referred to as Operational Taxonomic Units (OTUs). An OTU may be defined as the lowest taxonomically ranked unit to be studied in a particular investigation, i.e. an element in a set that is to be classified. Operational is used instead of fundamental to imply that the taxonomic rank can be of any level: for example OTUs may be individual specimens, exemplars of genera, or average specimens representing species (Sneath and Sokal, 1973).

The choice of the level of taxonomic rank at which the OTUs will be established is important. Jardine and Sibson (1971) advise that the choice should be preceded by stating the purpose of the investigation. This may effectively constrain the choice to a particular taxonomic level. A simple example might be a study of the intra-specific variants in Vicia sativa where the ideal taxonomic unit would be the individual and not averaged varieties which would provide little information about the intra-varietal variation pattern.

Sneath and Sokal (1973) pose two basic questions in the context of OTU level: should numerical taxonomy rely on the validity of prior classifications for its choice of OTUs and should every study be based on individuals in order to be taxonomically consistent and rigorous? The answer is usually "yes" to the first question. All revisions use an existing classification as a starting point. This will usually suggest the most appropriate OTU level. However, if the wrong OTU

level is selected this is likely to become apparent during the course of the revision. The answer to the second question is "no". It would be almost impossible to undertake a revision which included all the specimens of that taxon. This is especially true for a revision of a high level taxon, such as a tribe or family.

In order to answer any query, the question must be formulated within the framework of the existing knowledge base. Thus a revision must take as a starting point a pre-existing classification. Within this chosen classification representative OTUs are sampled in an attempt to falsify the starting classification. Following this falsification, a novel classification is suggested which is closer to the intrinsic natural classification of the taxon being revised.

It is impossible to sample every individual of a taxon, time and facilities are always limiting. Thus OTUs are necessarily sampled. Williams and Lance (1965) discuss the problems involved in choosing a sample of OTUs and conclude that the sample must represent the taxon for the study to be probabilistic. Sneath and Sokal (1973) conclude their discussion of this point by stating that if:

> "we wish to reexamine the relationship of numerous genera in a family we cannot reassert the validity of every genus from a study of its species (and the validity of the species from a study of their individuals)".

Thus OTUs must be sampled.

Variation within a taxon may present a problem when choosing a sample of OTUs. This is particularly true above the species level where within a taxon the majority of characters will vary. This problem may be overcome by including one OTU to represent each varying character combination per original OTU. An alternative is to use a median representative for the polymorphic OTU with the expectation that variance in the polymorphic OTU does not

exceed between taxon variation. A more commonly applied solution is to restrict the choice of characters to those that show little intra-OTU variation.

6.2.2 Choice of Operational Taxonomic Units

The taxon to be revised was Vicia subgenus Vicia, as delimited by Kupicha (1976). The reason for the choice of this taxon and classification is discussed in Chapter Four. She includes thirty two species in the subgenus. To this number, the subspecific taxa, which she does not detail, and the taxa not known to her, which using her circumscription appear to be members of the subgenus, were added. If these extra taxa are included, the number of taxa recognised at the beginning of the revision is increased to seventy two.

Following this selection of the base classification, Jardine and Sibson (1971) suggest that the next step in selecting the appropriate OTU level is to state the purpose of the investigation. The central taxonomic objective of the revision was to attempt to establish the 'natural' subgeneric classification for subg. Vicia: i.e. to establish which taxa were accepted and the relationship between these accepted taxa. The taxonomic study was orientated at a low enough level for each specimen to be given OTU status. This meant that for the analysis concerning subgeneric taxon recognition, 1539 OTUs (specimens) representing the 72 taxa were analysed, using the full subgeneric character set of 171 characters.

Following the analysis of the complete data set, which located taxa (specimen clusters), each specimen was attributed to one of the accepted taxa. The description synthesis programs LICEA and CALIE (discussed in thesis section 10.3.4) were then used to produce taxon descriptions direct from the specimen data for each taxon. These taxon descriptions were subsequently converted to a taxon-based data set which was then analysed. So in this subgeneric analysis the taxon was given OTU status. This second analysis was used to

corroborate and clarify the relationships between taxa suggested by the results of the complete data set analysis. As the taxon data were synthesised from numerous representative specimens for each taxon, the data set had a small proportion of missing data compared to the complete character set and so produced results uncomplicated by missing data problems.

For the specimen-based analysis, unequal number of OTUs (specimens) were scored from each taxon. By far the largest number of specimens was scored for V. sativa because it is the most cosmopolitan and polymorphic Vicia species and because of the controversy surrounding its sub-specific classification (Plitmann, 1967; Hollings and Stace, 1978; Stankevich, 1978a). At the other extreme there are several taxa for which very few specimens have been collected and so only a few specimens could be included in the study, e.g. V. esdraelonensis, V. tigridis, and V. sativa subsp. devia, as well as the species only known from Southampton-based collecting missions (V. kalakhensis and V. eristalioides).

6.3 Character Selection

6.3.1 Theoretical Foundation

Animal perception is mediated by the concept of characters. As a result of the accumulation of sense perception data, animals are able to distinguish the world in which they exist. The concept of characters is no less central to taxonomy and provides the descriptive information about taxa on which classifications are founded. It may at first seem curious that something so fundamental to taxonomy as the character is defined in virtually unique terms in each taxonomic text book. This is, perhaps, because characters fall into the category of concepts that are so fundamental that they defy accurate definition but at the same time they are universally understood.

That is not to say, attempts at definitions should not and have not been made. Davis and Heywood (1973) provide the following general definition:

> "any attribute (or descriptive phase) referring to form, structure or behaviour which the taxonomist separates from the whole organism for a particular purpose such as comparison or interpretation".

It is clear from this definition that differences, similarities and discontinuities between taxa are reflected in their defining characters (Jones and Luchsinger, 1987) and it is these differences, embodied by characters, which distinguish organisms and on which classifications are based.

Within the context of numerical taxonomy the concept of a character is more restricted. In an attempt to make character selection more objective and free character definition from logical inconsistencies (as discussed in thesis section 2.3), Sneath and Sokal (1973) refer more specifically to 'unit characters', which they defined as:

> "a taxonomic character of two or more states, which within the study at hand cannot be further subdivided logically, except for subdivision brought about by changes in the method of coding the state".

The later definition reflects the general philosophy of numerical taxonomy. By defining and utilising a larger number of 'simple' characters it is possible to reflect more accurately the genome in the resultant classification. Here 'simple' is used in terms of lack of logical subdivision.

Davis and Heywood (1963) stress the abstract nature of characters. It is character states that taxonomists actually utilise. An example of this might be the character vexillum adaxial surface pubescence. The presence or absence of hairs on the adaxial surface of the vexillum would represent the two character states which would be used in the analysis.

Davis and Heywood (1963) distinguish four pairs of character categories which relate to their intrinsic nature and must be borne in mind when selecting characters. The first pair of categories consist of phenetic or phylogenetic characters. These are defined by their use in either phenetic or phylogenetic studies. Phylogenetic characters may be used in phenetic studies, but are generally chosen for a specific purpose (to infer phylogeny) and so produce not a general but a special classification based usually on far fewer characters than phenetic studies. The phylogenetic characters related to subg. Vicia, are discussed in Chapter Nine. The second pair of character categories are analytic and synthetic characters. Analytic characters are the diagnostic characters used in identification, characterisation and delimitation. Synthetic characters are used in a more generally descriptive manner. They show a wide occurrence and are constant for a particular taxon. Davis and Heywood's third pair of character categories are qualitative and quantitative characters. The features assessed by size, length, number, etc. are quantitative (e.g. peduncle length) and those assessed by form are qualitative (e.g. calyx tube apex shape). The fourth pair of character categories is the practically important division into 'good' and 'bad' characters. They recognise good characters as being those that:

1. are not subject to wide variation within the samples being considered.

2. do not have a high intrinsic genetic variability.

3. are not easily susceptible to environmental modification.

4. show consistency, i.e. agree with the correlations of characters that exist in a natural system of classification constructed without their use.

Stace (1980) refers to good and bad characters as those with greater or lesser taxonomic value. He continues by stressing the relativity of these terms. What may be a good character in a generic study may be a poor one at the species

level, e.g. in the present study, occurrence of stipular nectaries is a good intra-generic discriminating character (in fact it is the most important character for distinguishing the two subgenera), but within subgenus Vicia it is invariant and so a bad character.

Good characters may be derived from various sources. Sneath and Sokal (1973) group the sources of taxonomic characters into the following 'rough' categories:

> "a) morphological characters (external, internal, microscopic, including cytological and developmental characters).
>
> b) physiological and chemical characters.
>
> c) behavioural characters.
>
> d) ecological and distributional characters (habitats, food, hosts, parasites, population dynamics, geographical distribution)"

While not suggesting that this list is completely comprehensive, it does show the wide variety of character sources available to taxonomists. In recent years there has been some argument over the comparative usefulness of the different categories of characters. The more traditional school, exemplified by Cronquist (1975), argues that due to the relative ease with which category "a" characters can be scored, they will remain the "mainstay" of taxonomy. Turner (1977) disagrees, arguing that Cronquist's approach neglects the fundamental nature of chemical data to the plant genome and subsequently its direct effect on plant classification. Whether this is proven to be so or not, the traditional school is likely to remain dominant, simply because of the relative ease of scoring morphological characters compared to non-morphological characters. Turner's point should be registered, i.e. the broader the range of information used in forming a general classification, the more stable the resultant classification will be.

Once the source of characters is decided, Sneath and Sokal (1973) point out that two important questions relating to characters must be answered:

(a) whether each attribute selected should be equally weighted, and

(b) how many characters should be used in a particular taxonomic study?

They argue against character weighting and believe that all attributes should be weighted equally. This is, in fact, not feasible as the character set must always be a sample from the potential one and the most explicit form of weighting is that of inclusion or exclusion from the analysis (Dunn and Everitt, 1982). In this sense all taxonomic study uses weighted characters.

Weighting of characters may take two forms, a priori and a posteriori, weighting before and in the light of experience respectively. The latter are characters previously considered useful in the group, those characters known to be good diagnostically or those thought to be phylogenetically important. Sneath and Sokal (1973) argue that any weighting is inappropriate as it, "presupposes a knowledge that is not yet available, either about the classification of the organisms or about the presumed significance of their characters". For this reason no weighting of characters was used in this study, other than the implicit weighting via character inclusion or exclusion in the analysis.

Each organism possesses a limitless number of characters. The taxonomist will always be, however, limited temporally and economically, so characters must be selected. Stace (1980) points out that for practical reasons characters are often selected because they are most easily scored or because they seem likely to be reliable and discriminating in taxon delimitation. He proposes the truism that, "it is most convenient if taxa are delimited by obvious features rather

than by cryptic ones". Traditionally these convenient characters have often been morphological, mostly based on flowers because of their "conservatism". A conservative character is defined as one which remains relatively unchanged in fundamental structure over long periods of evolutionary development. Commonly the reproductive structures of closely related taxa are relatively conservative. As pointed out earlier, congruence studies have challenged these classically orientated beliefs and Stace himself stresses the importance of using a wide range of information from diverse sources to produce a more predictive classification.

An increase in the objective nature of character selection has been the least successful aspect in the application of computer technology to taxonomy. Davis and Heywood (1963) comment that, "character selection is the weak link in this whole approach". Unfortunately, this comment remains as pertinent today, perhaps due to the problems involved in defining the process of character selection, which a classical taxonomist undertakes largely intuitively. Bisby (1970), in a study of character evaluation and selection using taximetric procedures, concludes that computer aided character selection will prove very useful to taxonomists. He comments that for:

> "taxonomists working with large amounts of data, the evaluation and selection of characters may be one of the most difficult and time-consuming stages of taxonomic work. Using 'character analysis' to discover the most useful characters can save a large amount of time."

This is true. The character set cannot expand exponentially and taxonomists retain the same level of intuitive accuracy in selecting 'good' characters. Logically there must be a limit to the dataset taxonomists can hold in their memory and still 'understand' the dataset structure. In fact Bisby might also have added that there is a law of diminishing returns in character set selection. As the data set increases in size, the accuracy of intuitively selecting 'good' characters diminishes, because it becomes more difficult for the

taxonomists to understand the increasingly complex correlation patterns within the data.

6.3.2 Reasoned Choice of Characters

Having detailed the problems involved in 'good' character selection, it must reluctantly be admitted that subjective and intuitive character selection was used in this study. However, a strong effort was made throughout to select characters logically and to follow the principles outlined above.

The initial set of characters was limited, as are all taxonomic studies, by temporal, economic and material factors. The relatively large size of the taxon (seventy two taxa) and the fact that they are mostly endemic to the Eastern Mediterranean Region, almost completely restricted the study to scoring morphological characters from herbarium material. This emphasis was underlined by the lack of seed for many of the taxa (at least at the start of the project) and the limited facilities available for cultivating plants in a Mediterranean environment.

The morphological characters chosen were readily scored from herbarium material, although use of dried plants restricts character choice, i.e. many characters (e.g. flower colour, shape of stipuliferous nectary) cannot readily be scored from dried plant fragments. The focus on using herbarium material meant that some of the characters shown to be useful in sect. Faba by Khattab (1987) were not used in this study. These characters relate to attributes that cannot be easily scored from dried specimens, e.g. degree of basal shoot branching, height of main shoot, presence of upper plant branching and stipule extra-floral nectary colour.

Although it is not directly relevant to the scoring of the material, the extensive field work undertaken in the Eastern Mediterranean during the course of the project,

permitted both a better understanding of these more "ephemeral" characters and enabled the extensive collection of seed and fresh herbarium material. Davis and Heywood (1963) stress the advantage of having a good field knowledge of the taxon being revised. The field experience gained, during this project, certainly enabled a more comprehensive understanding of the group relationships to be formulated.

Phase 1 of the taxonomic study plan, as detailed in thesis section 4.3, includes a period for exploratory character scoring of Vicia subgenus Vicia specimens (the primary herbarium material) using an experimental character set of over 350 characters. This gross character set was derived from the literature, discussion with Vicieae specialists and from original observations. The characters derived from the literature were taken from: Fedtschenko, 1948; Zertova, 1962; Plitmann, 1967; Ball, 1968; Davis and Plitmann, 1970; Gunn, 1970; Kupicha, 1974, 1976; Townsend, 1974; Gunn and Kluve, 1976; Lerston, 1982; Maxted, 1984; Perrino et al., 1984 and the Vicieae Project score sheet. The Vicieae systematists consulted were Khattab, Allkin, Goyder and Bisby.

Following experimentation with the gross character set, a subset was chosen intuitively and is henceforth referred to as the 'subgeneric character set'. It contained 171 characters: 39 vegetative; 83 inflorescence; 24 fruit and 25 seed characters. This character set was used for the initial analysis of the complete specimen data set. For the taxon-based analysis three extra characters were added: character 2 - life form; character 3 - plant height and character 67 - calyx hair elevation. The two former characters could not be scored from herbarium material, but are important taxon characteristics. The data for these characters were taken from the literature and incorporated in the taxon data set. The third character, calyx hair elevation, was a very useful character for distinguishing the varieties of V. sepium. It was not recorded for individual specimens, but was considered

sufficiently useful to be included for the taxon based analysis. The subgeneric character set is detailed in Appendix 2.

The first set of analyses discussed in Chapter Seven, which refers to the splitting of the complete data set, used the 171 character set, but subsequent analyses each used a different character set. The use of multiple character sets, all abstracted from the subgeneric character set, was necessitated by the different taxonomic questions being addressed by each analysis. Character set selection was also complicated by the different taxonomic levels of the various studies and the limitations of the computer programs in accepting certain types of data and sizes of data set. Thus individual character selection will be explained prior to the discussion of the analysis results for each study in Chapter 7. The characters selected for each analysis are provided in Appendix 3. In each study, however, it is generally true that characters were chosen which showed variation between OTUs, but were constant for a particular OTU.

The SPSS^{x} Batch System (Norusis, 1988) was used in an attempt to make character selection more objective. SPSS^{x} is a comprehensive program for managing, analysing and displaying data. Norusis lists the capabilities of SPSS^{x}, as allowing:

> "Input from almost any type of data file.
>
> File management, including sorting, splitting, and aggregating files, match-merging multiple file, and saving fully defined system files
>
> Data management, including sampling, selecting and, weighting cases, recoding variables, and creating new variables using extensive numeric and string functions.
>
> Tabulation and statistical analysis - from describing single variables to performing complex multivariate analysis.
>
> Report writing.
>
> Device-independent graphics."

SPSS^x is a suite of integrated procedures within a single program. The central program calls up subroutines recognised by keywords. Each keyword initiates one SPSS^x subprogram so that an SPSS^x run can involve several distinct procedures. SPSS^x was used for two distinct purposes within the analysis, for character selection and for initial group formation in the phenetic analysis; the latter is described in section 6.4.2.

Discriminant analysis was undertaken via the subroutine DISCRIMINANT. Discriminant analysis is a multivariate technique in which linear combinations of variables are used to distinguish between two or more groups of OTUs. The first step of discriminant analysis is to find the linear combination of variables that best discriminates the groups of OTUs, thus an estimate of character discriminating power is obtained. The dataset was pre-processed using CONDESCRIPTIVE, so that all quantitative characters were standardised and, as discriminant analysis is unsuitable for the inclusion of multi-state qualitative characters, all multi-state characters with more than two states were converted to binary characters.

The DISCRIMINANT Statistic 6 prints the univariate F ratios for each variable. These are, in fact, one-way analysis of variance tests for equality of group means on a single discriminating variable. The larger the F ratio value the better the character is at discriminating the taxa. Experimentation using characters of various F ratio levels led to the decision to use characters with value of 10.000 or over in the phenetic analysis.

6.4 Phenetic Analysis

6.4.1 Theoretical Foundations

Of all the three topics covered in this chapter, this is the most problematic to define under the heading 'theoretical foundation'. The reason is the speed with which this field has progressed since the publication of Sokal and Sneath's

"Principles of Numerical Taxonomy" in 1963. This section will, however, attempt to introduce the subject of phenetic analysis (= numerical taxonomy) and will be followed by a discussion of the actual computer programs that were utilised. For a more comprehensive introduction one should turn to Dunn and Everitt (1982) and from there to Sneath and Sokal (1973).

Phenetic techniques analyse the overall affinities between organisms. The presence of consistent character combinations are used to define particular taxa. The larger the number of characters and the broader their base, the clearer the taxon definition (Davis and Heywood, 1963). Sneath and Sokal (1973) define numerical taxonomy as, "the grouping by numerical methods of taxonomic units into taxa on the basis of their character states". Davis and Heywood (1963) point out that the traditional techniques of taxonomists in neurally estimating resemblance on observable features is a form of phenetic analysis. Sneath (1961) criticises these traditional methods for being subjective because they employ imprecise statistical Q-techniques (Q-techniques are defined as a means of associating pairs of OTUs over all characters; Sneath and Sokal, 1973).

Historically, classifications were based on characters that were thought to have inherent importance to the plant. In each case a few morphological characters were chosen. Caesalpinus, who produced the first methodical classification of plants, "De Plantis" (1583), considered nutritional and reproductive characters to be the most important characteristics and so his classification was based on these characters; Tournefort (1700) used flowers and fruits primarily in his classification; Linnaeus (1753) used floral characters; de Jussieu (1789) used the number of cotyledons, nature of perianth and position of ovary; De Candolle (1824-1873) however used a more complex system retaining the use of number of cotyledons but basing his sub-classes on the degree of reduction and fusion of floral parts.

Two botanists, Ray and Adanson, stand out as opposing this system of artificial classifications (classifications based on a few characters). Ray formulated the principle that all part of plants should be used for classifying plants. His classification system presented in 'Methodus Plantarum' (1682) retains the traditional division of plants on habit, but also uses cotyledon number, fruit type, leaf and floral characters. Adanson, after trying to apply Tournefort and Linnaeus' classifications, rejected artificial systems as being hopelessly inadequate and therefore attempted to create a more natural system (Davis and Heywood, 1963). For the 'Familles des Plants' (1763) Adanson produced 65 single character systems in which generic position was very variable. He united these systems so that agreement between the greatest number of systems reflected the natural (phenetic) relationships.

Cain (1959) comments that Adanson's approach was the forerunner of contemporary numerical principles. De Candolle, however, rejected Adanson's work on two counts: (a) it presupposes that we know not only all the organs of plants but also all the points of view in which to consider them, and (b) it did not allow for a priori character weighting. The first point remains as relevant today as then. In the present study lack of legumes and seeds on specimens caused missing data problems during the analysis. The second of De Candolle's points is rejected (see Section 6.3 of this chapter), a priori weighting of characters is considered by pheneticists to be deleterious.

The publication of the 'Origin of Species' by Darwin (1859), in effect, caused little change in the procedure of classification. Taxonomists simply put a greater emphasis on including phylogenetically important characters in their analysis, but they retained the artificiality of using a small number of characters. Then in the 1950's, as the potential sources of taxonomic characters began to mushroom, some

taxonomists began to advocate a more objective method of classification. Michener and Sokal (1957) argued that taxa should be classified on the total degree of difference or similarity using as many characters as possible from different sources. This was an obvious revival of Adansonian principles and, finally, with the development of computer technology, multivariate analysis became a practical option.

The principles of numerical taxonomy, formulated by Sneath and Sokal (1963), are as follows:

> "1. The greater the content of information in the taxa of a classification and the more characters on which it is based, the better a given classification will be.
>
> 2. A priori, every character is of equal weight in creating natural taxa.
>
> 3. Overall similarity between any two entities is a function of their individual similarities in each of their many characters in which they are being compared.
>
> 4. Distinct taxa can be recognised because correlations of characters differ in the groups of organisms under study.
>
> 5. Phylogenetic inferences can be made from the taxonomic structures of a group and from character correlations, given certain assumptions about evolutionary pathways and mechanisms.
>
> 6. Taxonomy is viewed and practised as an empirical science.
>
> 7. Classifications are based on phenetic similarity."

Stace (1980) comments on these principles thus:

> "These seven principles underlie not only Sneath and Sokal's concepts of numerical taxonomy, but the subject as a whole and nowadays essentially the same aims and methods are being applied to all groups of microorganisms, animals and plants".

Stace could have added that the reason these principles are so important to contemporary taxonomy is largely due to the

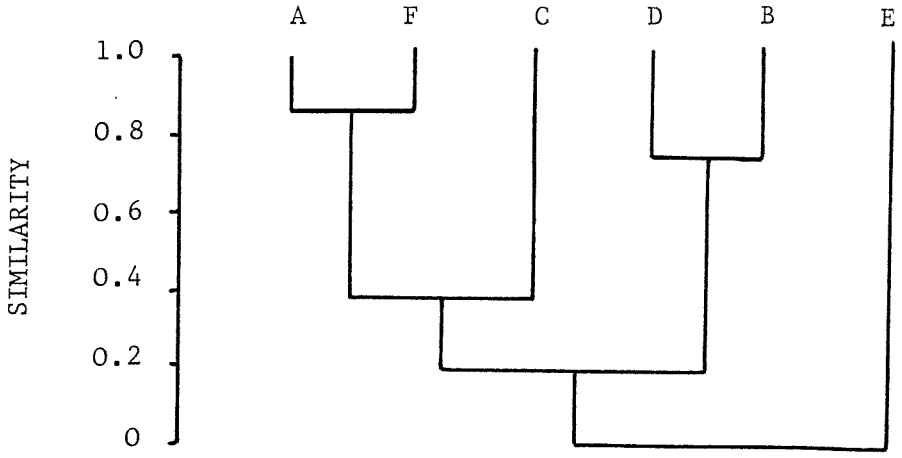
impact of numerical taxonomy on taxonomy as a whole in the past twenty years.

Sneath and Sokal (1973) continue by outlining the practical sequence of numerical taxonomy:

> "organisms and characters are chosen and recorded, the resemblances between organisms are calculated, taxa are based upon these resemblances and last, generalisations are made about the taxa (such as inferences about their phylogeny, choice of discriminating characters, etc.)."

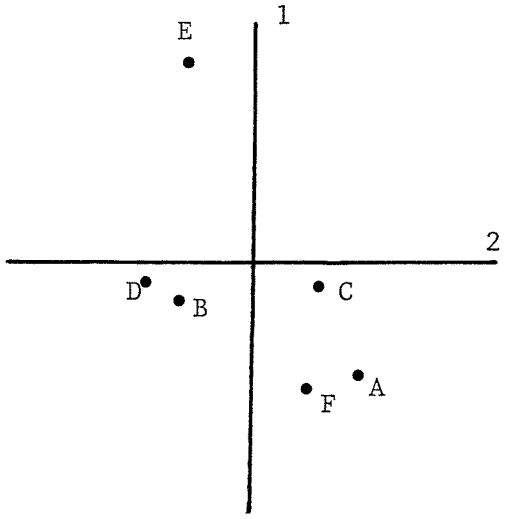
The choice of organisms and characters has been discussed in the preceding sections of this chapter. Following these choices, the resemblance between operational taxonomic units (OTUs) is calculated. This can be divided into two sub-stages. In the first the OTUs are scored and the data are organised into a data matrix table (OTUs x characters or t x n). The data are codified to facilitate entry into the computer. The second stage involves calculation of the triangular similarity or dissimilarity matrix between OTUs (t x t). Numerous similarity and dissimilarity coefficients are available for calculating this matrix (see Dunn and Everitt, 1982: Chapter 3). Sneath and Sokal recommend that the simplest coefficient applicable to the study's data should be chosen as it will ease interpretation of the final results. However, the coefficient must be able to reflect adequately the natural patterns within the data.

The triangular similarity matrix is then analysed using a range of possible techniques which generally fall into two categories, cluster and ordination methods. For both these categories there are numerous variations on the basic techniques (for cluster analysis see Everitt, 1974 and for ordination see Dunn and Everitt, 1982). Both these categories of analysis produce distinctive means of displaying the result visually. Cluster analysis usually yields a hierarchical classification of OTUs, which is commonly depicted in the form of a dendrogram (see below).



OTUs are linked at particular levels of similarity; the higher the level of similarity the more closely allied are the OTUs. The abscissa has no special significance as it simply spaces the OTUs. The ordinate, however, represents the linking of OTUs at a particular level of similarity. Linkage could extend upwards until OTUs exhibit maximum similarity (1.0) and extend downwards as far as is needed to unite all OTUs in the particular study (in the above example 0.2).

The second category of analysis techniques, ordination, usually displays its results in the form of a scatter diagram, in which the most closely allied OTUs are spatially juxtaposed (see below).



Sneath and Sokal (1973) stress the important advance made by numeric taxonomy, in enabling the delimitation of taxonomic groups in a repeatable and objective manner. Repeatability implies that two taxonomists will produce comparable

classifications if presented with the same problem. This has at least been partially invalidated by congruence studies, which have shown that different character sets can produce different classifications of the same sets of OTUs. However, Sneath and Sokal (1973) suggest that in general the larger the number of characters and the broader their origin (from morphology, biochemistry, anatomy, etc.) the more likely the classification is to be predictive and the less likely it is for estimates of similarity to be significantly changed. Objectivity is a relative concept, but use of numerical taxonomy has reduced subjectivity and bias, so enhancing the predictivity of the classification subsequently produced.

In their concluding remarks on the potential value of numerical taxonomy Davis and Heywood (1963) comment that, "It would be rash to predict the reaction of Angiosperm taxonomists in general to numerical procedures, but clearly a cautious welcome should be extended". Thirty seven years later the note of caution can be withdrawn: numerical procedures have proved a major advantage to taxonomists, allowing them to thoroughly tackle problems and to use sources of information that would have been previously beyond their scope.

6.4.2 Reasoned Choice of Phenetic Methods

For the phenetic analysis, three computer programs were utilised: SPSS^{x}, LINKAGE, and CLUSTAN 3, all available on the IBM 3090 mainframe computer at the University of Southampton. These programs were selected to carry out particular forms of analysis on the different data sets. More than one program was used to analyse the data sets because, as pointed out by Duncan and Baum (1981) and Davies (1985), different algorithms bias results in different ways. Duncan and Baum stress that comparison of different analysis methods should be an integral feature of the application of phenetic techniques. So to derive a natural picture of taxa inter-relationships, several methods of analysis were used and then the results were

"pooled" to produce a synthetic but hopefully a natural classification. For this reason three different programs with different forms of analysis were used allowing each to verify the other programs' results. Each program will be discussed together with an introduction to concepts underlying the program and the program's advantages and disadvantages.

(a) SPSS^{x}

The SPSS^{x} Batch System, as well as being used for the character selection (described above), was used for initial group formation in the phenetic analysis. Due to the relatively large size of the gross subgeneric data set (1539 OTUs x 171 characters with missing values) it was difficult to find an analysis program that could be used which was theoretically sound (did not force unacceptable subjective assumptions about the data) and would produce useful output. SPSS^{x} appeared to provide the best option using the subroutine DISCRIMINANT and undertaking discriminant analysis facilitated the initial division of the gross subgeneric data set, which could then be followed by more vigorously using the LINKAGE and CLUSTAN routines.

This use of SPSS^{x} forced a degree of subjectivity into the analysis, which was not considered to be unacceptable. To undertake discriminant analysis a grouping variable must be specified. As discussed in Chapter Ten, the specimen identification at the commencement of the revision is coded into the specimen code, e.g. the specimen 06003 was identified as V. bithynica, the first two numbers "06" representing V. bithynica. Thus within DISCRIMINANT the grouping variable used was the first two figures of the identification code. This might have been unacceptable if closely allied taxa were placed in separate clusters, as specimens may have been wrongly identified and then forced into different clusters but this was not the case. DISCRIMINANT was only used for a crude division of the gross data set and any taxa considered closely allied were contained within one of the subgeneric clusters.

Perhaps the most taxonomically problematic group there was a possibility of misidentification was the V. narbonensis complex but all V. narbonensis complex specimens were included in one cluster. A misplaced specimen would also have become apparent in the later stages of the analysis.

As for the previous use of DISCRIMINANT, the dataset was pre-processed using CONDESCRIPTIVE, so that all quantitative characters were standardised and the multi-state qualitative characters, with more than two states, were converted to binary characters.

The canonical discriminant functions, evaluated by group means (group centroids), were plotted for the first two functions in order to identify major groupings. To aid further clarification of the results, the first ten functions were then used as variables themselves and analysed using average linkage cluster analysis via the program CLUSTAN, described below.

(b) LINKAGE

LINKAGE is a single linkage cluster analysis FORTRAN program, written by Wirth, Estabrook and Rogers (1966) and introduced in a handbook to the program written by Fleming and Appan (1972).

The algorithm used by LINKAGE can be divided into two main stages of operation. Firstly, the similarity matrix is produced by calculating the similarity coefficients between all the pairs of OTUs involved in the study. Secondly, in decreasing similarity order, the OTUs at a specific similarity level are connected to form clusters. The clusters gradually become more inclusive as the similarity level drops. This iterative procedure halts when all OTUs are contained in a single cluster. Measures of connectedness and isolation are calculated at each similarity level for each cluster to aid the taxonomist in interpreting the results. The similarity

coefficient used by LINKAGE is the simple matching coefficient. The similarity S(a,b) attributed to the pair of OTUs a and b will be the number of characters for which the same state has been attributed to OTUs a and b divided by the total number of characters used in the comparison.

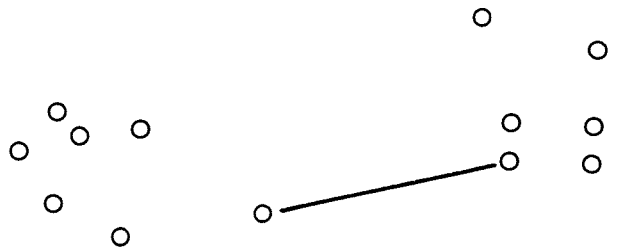
$$S(a,b) = \frac{\text{Number of characters in which states are shared by a and b}}{\text{Number of characters on which a and b have been compared}}$$

An extremely useful modification has been added to the measure of similarity, which enables the taxonomist to specify that two or more character states are partially similar instead of completely different. This allows the character states of logically ordered characters to be given partial relatedness, e.g. if there are ten states to the character leaf shape, the character states are ordered but this facility can be used to infer, for example, that character state 2 is partially allied to character state 5, but not with character state 7 which is quite different.

LINKAGE is also capable of compensating for missing data. As often occurs, the taxonomist may not be able to score data from each OTU for every character. The 'overall similarity' of a pair of OTUs is calculated from those characters where data is present for both OTUs, ignoring those characters that are absent for either OTU. The type, ordering (unordered or ordered) and number of character states per character are specified individually for each character and so the program has the advantage of permitting mixtures of character types to be used.

When the similarity matrix has been calculated for all pairs of OTUs the matrix is analysed using single linkage (nearest neighbour) clustering. A detailed background to the clustering techniques used during this project is provided by Sneath and Sokal (1973). Clustering follows the general rules

given above but, with single linkage, an OTU is admitted to a cluster once it has reached a level of similarity similar to that of any one of the OTUs already present in the cluster. Similarly any pair of clusters will coalesce if the similarity between any pair of OTUs (one representing each cluster) exceeds a threshold level of similarity. Thus fusions are based on single links between particular OTUs, see diagram below (Note: other members of a cluster might be quite dissimilar to the OTU or cluster of OTUs with which the cluster links).



O- OTU

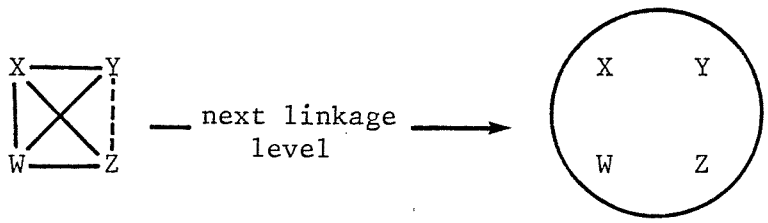
With LINKAGE, prior to actual clustering, the program orders the similarity measures between pairs of OTUs into decreasing similarity order, so that OTU pairs with a similarity level of 1.0 will be presented at the beginning of the print out. Following this, the program instigates a cycle of cluster forming iterations with decreasing similarity so that successive OTUs are incorporated into clusters. Every time a cluster is modified, by the addition of a new OTU or cluster of OTUs, the measure of cluster isolation is printed called 'moat'. It states how long, on the axis of decreasing similarity, it will be before the cluster is further modified, eg:

"MOAT = .07500 NEXT PAIR TO JOIN (15006, 15004) (15007, 15004)"

means that the present cluster will be modified after a further decrease in similarity of .07500 and that the next two OTUs to join the cluster will be 15006 and 15007. These will join by both forming a link with 15004

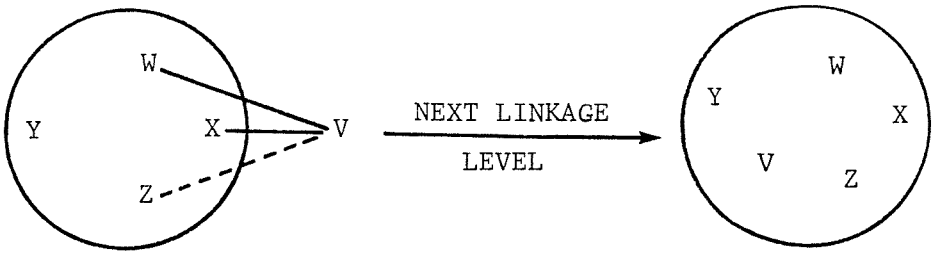
For practical reasons the drawing of linkage diagrams or sub-graphs from LINKAGE output requires a brief introduction. The LINKAGE output details a sub-graph for each level of similarity at which a novel clustering occurs. At a given threshold level of similarity each pair of OTUs will cluster and this is demonstrated in the sub-graph by a line connecting the OTUs. In effect, this connecting line could be of three types: a double line indicating a new clustering of either OTUs or clusters, a single line indicating a linkage established at a higher level of similarity, which at this level of similarity demonstrates internal cluster structure, and a broken line which indicates a new link within a cluster between OTUs at that particular level of similarity.

With large numbers of OTUs, especially at lower levels of similarity, the OTUs become complicatedly inter-connected. To simplify interpretation of the sub-graphs, highly connected OTUs are encircled. The criterion for inclusion into a circle is that each OTU should have at least three connections with other members of the same cluster. As an example, suppose that four OTUs W, X, Y and Z are connected as shown below:



It can be seen that XW, XZ, XY, WZ and WY have established intra-OTU links at that particular linkage level. A new intra-OTU link is formed between YZ indicated by the broken line joining Y to Z. Then within the cluster all four OTUs are joined so at the next linkage level the four linked OTUs will be drawn inside a circle as shown.

If then a fifth OTU, V, forms links with the encircled OTUs W, X, Y and Z, as shown below, then it will also be included in the circle as shown.



This kind of linkage diagram enables LINKAGE to overcome the main criticism of the single linkage algorithm: that clusters consisting of loosely linked chains of outliers and intermediates cannot always be distinguished from clusters of highly interconnected OTUs. LINKAGE records all links between OTUs and these can be seen by drawing the individual subgraphs detailed in the output.

Several workers have concluded that single linkage cluster analysis, used in conjunction with linkage diagrams, provide the most satisfactory taxonomic arrangement of studied taxa (Prance, Rogers and White, 1969; Stearn, 1971). Bisby (1973), after undertaking a comparative study of multivariate data analysis using the legume genus Crotalaria L., comments on this method:

> "It gives an excellent indication of the pattern of similarities within the groups formed, and an indication of the affinities of marginal and isolated species excluded from these groups. Some of these results have been sufficiently useful for improvements to be made to the classification of Crotalaria"

(c) CLUSTAN

CLUSTAN is a suite of FORTRAN IV programs for data analysis designed by David Wishart, which include both cluster analysis and ordination analysis. Generation 3.2 was

published in June, 1987; CLUSTAN 3 is introduced in a manual written by Wishart (1987).

CLUSTAN is similar to SPSS^{x}, in that it has been designed as an integrated suite of separate procedures within a single program. These procedures are activated by the CLUSTAN driver. Wishart has attempted, in developing CLUSTAN, to simplify the program specifications so that routines are easy to use while incorporating special features and non-standard options.

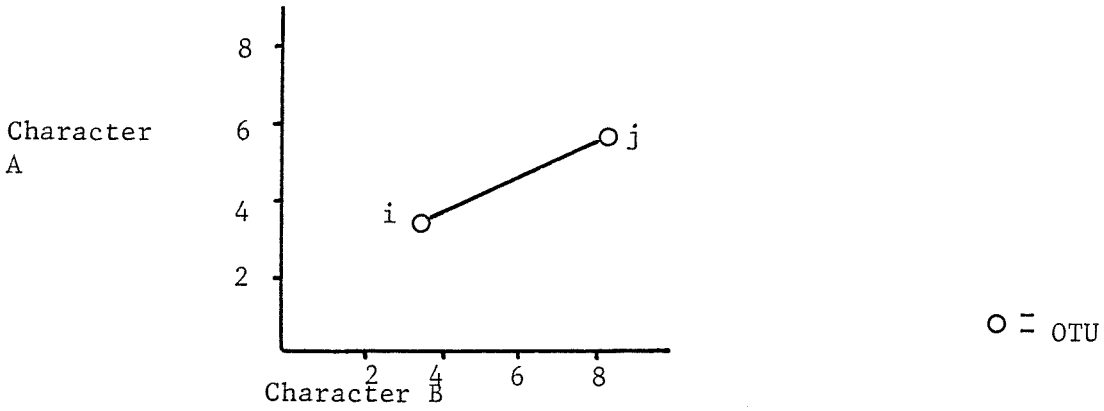
The CLUSTAN driver recognises key words which contain the name of subroutine procedures. Each keyword initiates one CLUSTAN step, thus a single CLUSTAN job can comprise any number of different steps which provides CLUSTAN with a versatility very useful to taxonomists. The program will accept data in the form of continuous measurements, multi-state attributes or as binary data. The program is also able to accept a matrix of similarity or distance values in a triangular table (t x t). Details of the actual combinations of subroutines used in the analysis will be detailed in Chapter Seven together with the results of the analysis.

CLUSTAN was used to carry out cluster analysis, principal component analysis and to find the minimum spanning tree. The theoretical background and advantages of each of these techniques will now be discussed.

Following the keyword CLUSTER, a range of different cluster analysis programs can be selected. Following the use of single linkage in the previous program, LINKAGE, it was decided to use three clustering procedures which undertake cluster analysis in a different manner. Average linkage (unweighed pair-group), centroid and Ward's method (error sum of squares) were selected. CLUSTER checks that the appropriate dissimilarity coefficient is used for the method of cluster analysis selected. The dissimilarity coefficients

selected were squared Euclidean distance for the group average and the centroid methods, and the distance coefficient for Wards method.

The most widely used dissimilarity coefficient used is Euclidean distance (Wishart, 1987), from which both squared Euclidean distance and the distance coefficient are derived. Euclidean distance is easy to conceptualise as the distance between two points as is shown in the simple two character example below:



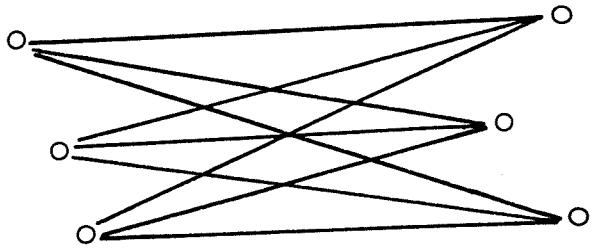
The line drawn between OTU i and j demonstrates the Euclidean distance between the two points. To calculate the Euclidean distance between OTUi and OTUj the following formula is applied:

$$d_{ij} = \left[(X_{ja} - X_{ia})^2 + (X_{jb} - X_{ib})^2 \right]^{\frac{1}{2}}$$

This can be generalised for p characters per OTU thus:

$$d_{ij} = \left[\sum_{k=1}^{p} (X_{ik} - X_{jk})^2 \right]^{\frac{1}{2}}$$

Average linkage cluster analysis was one of the earliest attempts at taking group structure into account when clusters are formed. Dunn and Everitt (1982) describe average linkage as the method that "defines the proximity between two clusters as the average of the proximities between all pairs of OTUs that are made of one OTU from each group". This is demonstrated in the diagram below.



Wishart (1987) comments that this method does "tend to find spherical clusters", which is indeed a serious criticism if spherical clusters are not naturally inherent in the data. It would appear to have the opposite problems of single linkage, which Wishart asserts, "will find 'straggling' clusters but often fails with large populations due to chaining".

The second method of cluster analysis called from CLUSTAN was the centroid method. Dunn and Everitt (1982) describe centroid linkage as the method where "groups once formed are replaced by their mean vectors, and inter-group distance is defined as the distance between these means." This can be illustrated in the following diagram:

The third cluster analysis method employed from the CLUSTAN suite was Ward's method recommended by Wishart (1975) as "possibly the best of the HIERARCHY options". Ward (1963) proposes that at any stage of the analysis, the loss of

information, resulting from the grouping of individuals into clusters, can be measured by the total sum of squared deviations of every point from the mean of the cluster to which it belongs. At each step in the analysis, union of every possible pair of clusters is considered and the two clusters whose fusion results in the minimum increase in the error sum of squares are combined (Everitt, 1974). Wishart (1987) states that, "this method finds minimum-variance spherical clusters" and Dunn and Everitt (1982) conclude that group average clustering and Ward's method are generally the best methods of clustering available. Sneath and Sokal (1973), however, point out that Ward's methods of clustering on the minimum increase in the error sum of squares is questionable and "may correspond to unacceptable partitions". So all four methods of cluster analysis used in this study have their proponents and their detractors. This should make the comparative interpretation of the results interesting in the context of comparing techniques of clustering.

The program, CLUSTAN, was also used to carry out ordination procedures. Using the sub-routine SCATTER, principal components were calculated. Before production of principal components, the data matrix X was standardized (unit variance zero mean) producing a matrix of N X N product-moment correlation coefficients between pairs of characters and was computed using the formula :

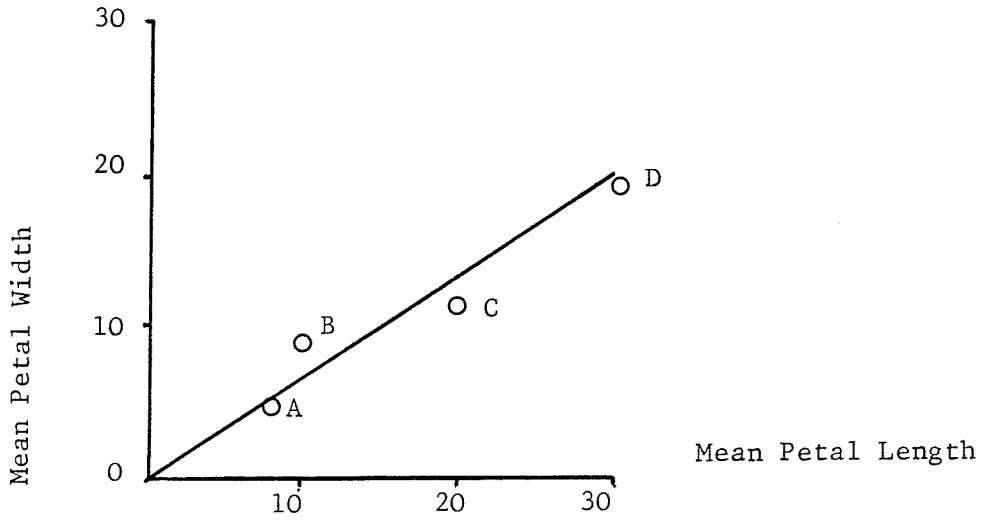
$$R = 1/(t - 1)\ XX'$$

From this matrix R the principal components were computed which involved the calculation of eigenvalues and eigenvectors. The importance of these eigenvectors is that they describe the relationship between OTUs with economy. This means that a large proportion of the dispersion of the OTUs can be accounted for in few dimensions or principal components. The coordinates of these axes are linear combinations of the original variables and they summarise the

major dimensions of variation. An eigenvalue is equal to the variance of the OTUs along its corresponding axis, so that the principal axis corresponding to their largest eigenvalue is the dimension that accounts for the greatest amount of variance from the sample of OTUs. The second principal axis, drawn at right angles to the first, accounts for the second largest amount of variance and so on. Quite often as few as three principal axes will account for the majority of the variation (Sneath and Sokal, 1973). Dunn and Everitt (1982) provide a simple example to demonstrate the transformation involved in calculating principal components. Consider the hypothetical set of data consisting of pairs of measurements relating to the petal dimensions of four buttercups:

OTU	A	B	C	D
mean petal length (mm)	8	10	20	30
mean petal width (mm)	4	9	11	18

This data is plotted below.



To express the variation of these two characters on a single axis, the line Z would be drawn which maximises the variance of the projections of the four points onto it. This will show the maximum discrimination of the four OTUs. This line is equivalent to the first principal axis. Projection of the four points onto this line gives the following first principal component scores for each OTU.

OTU	A	B	C	D
Score	9	13	23	35

The second principal axis drawn at right angles to the first would give second principal component scores of:

OTU	A	B	C	D
Score	1	-1.5	0.5	0

Dunn and Everitt (1982) note, "that the values of the first principal component are all positive and are clearly related to the size of the buttercup flowers. Those of the second component, however, can have either positive or negative values and may give information about the variation in the shape of the flowers".

Rohlf (1968) has pointed out that principal components analysis faithfully represents the distances between major groups or clusters of OTUs but it is unreliable for demonstrating the distances between near neighbours. The method underlying the calculation of principal components is sound but the choice of axis for plotting leads to the distortion of the relationship between near neighbouring OTUs.

Dunn and Everitt (1982) point to two other problems associated with the use of principal components analysis. The first is the tendency to produce principal components plots with a Gaussian distribution if discontinuous data are used. Points tend to lie on a horseshoe-shaped curve rather than reflect the inherent structure of the data. This is especially noticeable with binary data where the points tend to occupy the apices of a hypercube. Although Williamson (1978) has described a method to overcome these problems, the solution is not incorporated in CLUSTAN 3. The present study involved very little binary (presence-absence) data; proportionally continuous data predominates and so it is hoped that this malformation of data structure will have little

effect on the results. Dunn and Everitt also point out, that in principal components plots, the Euclidean distance between OTU points is an approximation of the Euclidean distance between OTU points in p-dimensional character space. This implies that if Euclidean distance is not considered the appropriate metric for analysis of the data using PCA would also be inappropriate. This second point is not considered to be a problem because the data gathered in the present study are compatible with the Euclidean distance metric.

The third problem associated with PCA is that it is scale sensitive, meaning that the most abundant character states will dominate the analysis to the detriment of rarer character states. This problem may be overcome by transforming the original data, either by taking their square or cubed roots.

The sub-routine SCATTER which carries out the PCA can be used secondarily to produce a minimum spanning tree. Dunn and Everitt (1982) state that the usefulness of any ordination technique can be judged by how well it preserves the inherent structure of the data. The use of a minimum spanning tree allows assessment of how well the original proximities of the data are preserved by two-dimensional mapping. Dunn and Everitt define it as follows:

> "Suppose n points are given (possibly in many dimensions), the tree spanning these points, i.e. a spanning tree, is any set of straight-line segments joining pairs of points such that;
>
> (a) no closed loops occur,
>
> (b) each point is visited by at least one line,
>
> (c) the tree is connected, that is, it has paths between any pair of points.
>
> If a weight is assigned to each segment in the tree then its length is defined to be the sum of these weights. The minimum spanning tree of the n points is then defined as the spanning tree of minimum length."

They illustrate the use of a minimum spanning tree in assessing the data distortion of ordination techniques, using the example described by Gower and Ross (1969). In this example, canonical variate analysis is used to analyse skull measurements of white-toothed shrews from the Scilly and Channel Islands. Their two-dimensional representation of the results incorporates 89% of the variance. However, distortion of the data by the ordination method can be clearly seen, e.g. in the two-dimensional representation shrews from Jersey and Sark are closely allied, but when the minimum spanning tree of the original distance matrix is superimposed onto the two-dimensional representation, it indicates that the Jersey and Sark races are closer to the Tresco race than to each other.

Wishart (1987) points out that a minimum spanning tree is analogous to single linkage clustering, where the pair of nearest neighbours at each fusion step define an edge of the graph. It has been used to partition a population with respect to its error sum of squares. For the above reasons, the option of obtaining the first and second principal components with minimum spanning tree overlaid, was chosen.

The sub-routine SCATTER also has the facility of plotting cluster circles onto the scatter diagram of the principal component analysis, the cluster circle being an indicator of cluster variance. Wishart (1987) explains the use of cluster circles as follows:

> "Suppose that XM, YM are the means and VX, VY the variances of the X and Y distributions for the subset of individuals which belong to cluster L, then the circle for cluster L will be drawn with its centre located at (XM, YM) and radius proportional to the square root (VX + VY)."

Cluster circles are drawn for the results of the following PCA plots to provide an extra aid in interpreting cluster conformity.

So to conclude this section it should be noted that each method of phenetic analysis has it's advantages and disadvantages, it's proponents and it's detractors. This is one reason why a combination of analytical methods was used. The results of each method were then used to verify the results of the other methods of analysis. Taken together the results of the different methods used provided an approximation of the actual, 'natural' relationships of the plants. This approximation can then be used in constructing a more natural composite formulation of the plants' relationships.

CHAPTER SEVEN
PHENETIC STUDIES ANALYSIS RESULTS

7.1 Introduction

The purpose of this chapter is to present and discuss the results of the phenetic analysis undertaken in order to clarify the natural relationships between the taxa of subgenus Vicia. Estimation of relatedness between taxa will be based on overall resemblances as indicated by the results of the phenetic analysis. The analysis is divided into three component studies: subgeneric analysis of the specimen data (specimens = OTUs), which attempts to delimit groupings of specimens and suggest taxa; subgeneric analysis of the taxon data (taxa = OTUs), which attempts to clarify the broad intra-subgenus taxon relationships; and detailed analysis of closely related taxa (taxa = OTUs), which form the larger sections. For each of these three studies the results will be discussed under the appropriate sub-heading of the phenetic analysis program utilised.

All three sets of analysis were based on the scores for the subgeneric character set. The matrix of 171-character-scores were taken from the dBASE II file MORPDATA. Due to the limitation in dBASE II of 32 fields per file, MORPDATA was, in practice, split into seven subfiles. An example of the first five records of the DATA subfile DATAVEGA is provided below:

```
1001 2 2  3  4  75 1 2 1 2 2 1 4  38  3  7  6  2  20  24
1002 2 2  0  0   0 1 2 1 1 0 1 0  55  2  6  7  2  35  46
1003 2 2  5  6  83 1 2 1 2 2 1 4  34  5 10 26  2  27  20
1004 2 2  3  2 150 1 2 1 1 2 1 4  60  4 11 12  2  39  37
1005 2 2  4  4 100 1 2 1 1 1 1 4  44  5  7 10  1  29  35
```

Each record of MORPDATA contains the specimen identification code number (1001, 1002, etc.) followed by the character scores, 19 in the above example for DATAVEGA. Prior to use in the phenetic analysis these seven files were merged into one, using a conversion program written by Mr. R. Crust (Dept. of Biology, University of Southampton). This conversion program

merged the seven records from the seven files that held the character scores for each specimen. The program then reordered the character scores, so that each record began with the identification code, followed by the 66 continuous character scores which were in turn followed by the 105 multi-state character scores. Due to differing requirements of the analysis programs, the output was required in two basic formats for use by SPSS^{x} and CLUSTAN, and by LINKAGE. The first two records of the CLUSTAN output file for specimens 1001 and 1002 are shown below:

```
 1001      03 04 075 038 03 07 06 02 020 024 035 12 000 00 06
00 0000 0000 0000 04 03 02 08 0050 00 00 00 00 00 000 000 000
00 00 00 00 000 000 00 00 00 00 000 000 00 00 000 00 00 00 00
018 07 03 0257 0233 02 08 08 06 24 02 02 100 133 0120 2 2 1 2
1 2 2 1 4 4 1 2 3 1 1 2 1 1 1 2 1 1 3 4 1 2 1 1 1 1 3 3 3 2 2
1 2 1 3 4 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 0 1 1 1 4 1 2 1 1 2 3 2 2 4 1 1 1 2 1 2 3 2 2 2 1 2 3
2 1 2 3 4 1 1 1

 1002      00 00 000 055 02 06 07 02 035 046 010 05 008 04 02
24 0200 0200 0033 24 19 15 12 0200 44 22 22 10 09 200 100 045
44 19 25 09 488 076 37 12 25 10 048 120 32 05 640 19 10 08 08
054 11 06 0490 0183 05 10 11 09 29 04 02 090 111 0725 2 2 1 2
1 1 0 1 0 2 2 1 3 1 1 4 5 1 2 2 1 1 2 3 1 1 2 3 2 3 2 2 1 1 1
1 2 1 2 2 1 1 1 4 1 2 2 1 1 1 1 1 1 2 2 2 3 1 1 1 2 1 2 1 1 1
2 4 2 1 1 1 1 1 4 2 2 1 1 3 4 4 1 6 1 2 3 2 1 2 3 1 2 2 1 2 1
1 1 3 3 1 1 1 1
```

LINKAGE requires continuous variables to be converted to alpha-numeric characters, so character scores over "9" were converted to letters, as is shown for specimens 1001 and 1002 below:

```
1001      23592432779C00600004328600000000000000000000000005
          738I2886C22IA32212122144123112111211341211113332212 1
          34200000000000000000000000000001114121123224111212
          322212321234111

1002      000D2432CD35542J242OJFCMMIM99C83QJP9M4VCPA9PQ5NJA85F
          B6FE5AB9F42G8E22121101022131145122112311232322111121
          22111412211111122231112121112421111142211344161232 12
          312212111331111
```

7.2 Subgeneric Analysis of Specimen Data

7.2.1 Splitting the Initial Complete Data Set

The first step in the phenetic analysis was the identification of clusters of specimens ("embryonic taxa") from the complete specimen data set. Unfortunately, as is discussed in detail in Chapter 11, there is no phenetic character selection or analysis program currently available that is ideal for the analysis of such a comparatively large data set. The use of database technology in the revision process will lead to the routine production of such relatively large data sets and contemporary analysis techniques will need appropriate enhancement.

The lack of any ideal analysis program required the adoption of a less-than-optimal solution to the problem of analysing the morphological character set. The complete data set of 1539 specimens by 171 characters was analysed using a combination of iterative applications of SPSS^{x} (Norusis, 1988) and CLUSTAN (Wishart, 1987).

Initially SPSS^{x} was used via procedure DISCRIMINANT to partition the complete data set into major clusters, which could then be analysed individually in a more objective manner. Each specimen included in the data set was subjectively attributed to one of the taxa accepted at the beginning of the study. Discriminant analysis of these OTU clusters was then undertaken to identify OTU cluster groupings. The complete data set was believed <u>a priori</u> to contain 67 taxon clusters and so each OTU was classified as belonging to one of these taxon clusters. Of the 171 characters, the 48 legume and seed characters included a high percentage of missing data. Although DISCRIMINANT can compensate for missing data (via substitution of missing values by means during classification), it was thought preferable, due to the predominance of missing legume and seed character scores, to exclude them from this initial analysis.

An example of an SPSS^{x} - DISCRIMINANT "run file" is provided in Appendix 4, as are all the "run files" for the various forms of analysis used. A "run file" is the file containing the directives required by the analysis program to undertake the analysis of the data set. Throughout the phenetic analysis various combinations of characters were used for each analysis and so, as discussed above, the character sets used in each analysis are detailed in Appendix 3.

The results of the DISCRIMINANT analysis for the subgenus Vicia, using specimen data, are shown in Figures 7.1 and 7.2. Throughout the results discussion, taxa will be identified using their numeric codes. A key to the codes is provided in Table 7.1. Taxa will also be attributed to sections sensu Kupicha (1976) using the following symbols:

Kupicha Section	Symbol
Atossa	Filled square
Vicia	Filled circle
Faba	Filled rhombus
Hypechusa	Filled triangle
Peregrinae	Inverted filled triangle
Unknown to Kupicha	Empty square

Figure 7.1 shows the scatter diagram for the first two discriminant functions for all 1539 specimens of the 67 specified taxon clusters, using vegetative and inflorescence characters and Figure 7.2 shows the first two discriminant functions for the 566 OTUs that contained no missing values, so using all 171 characters. In the latter the exclusion of specimens with missing values reduced the number of expected taxon clusters to 51.

Figure 7.1. Discriminant analysis of specimen data using 123 characters

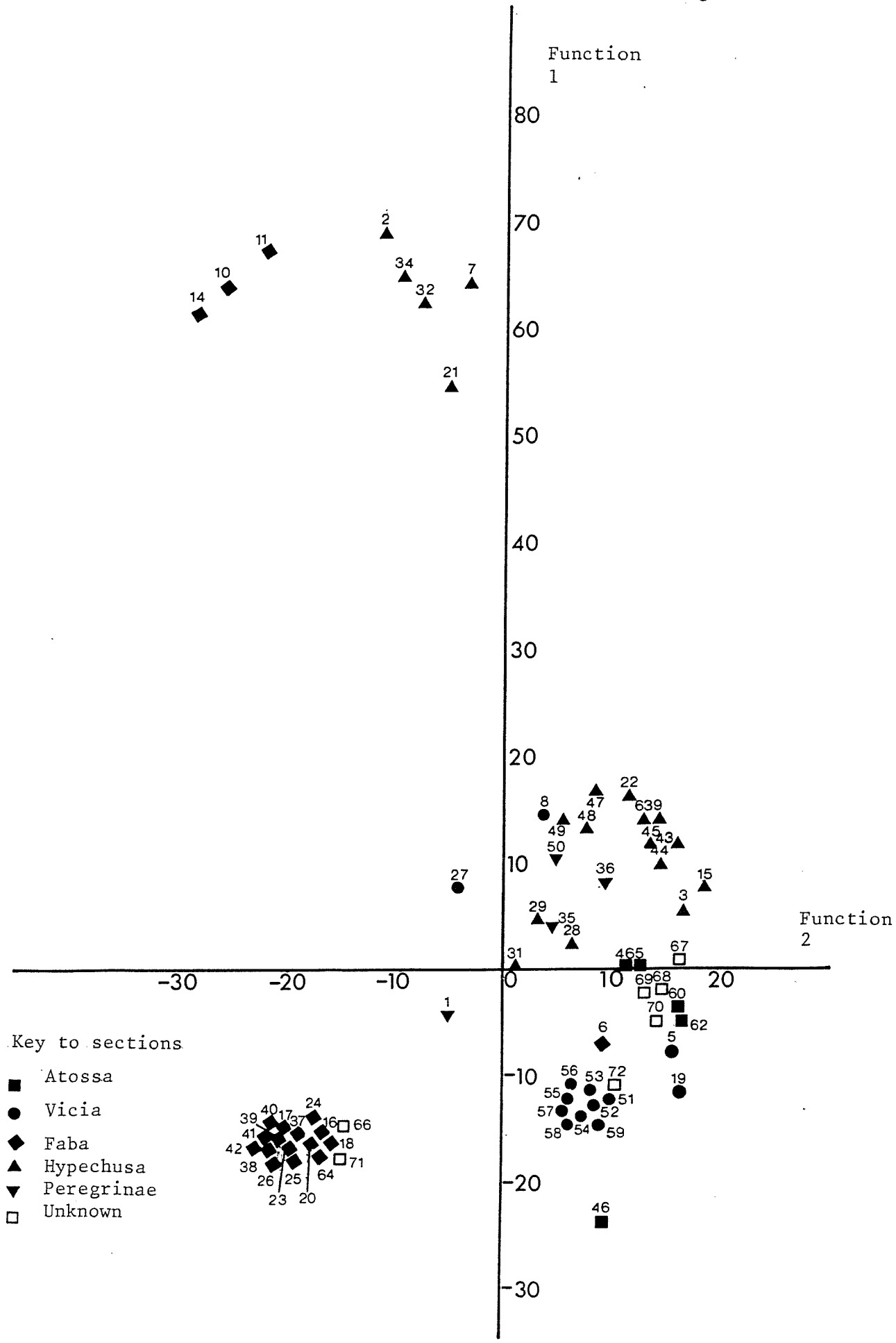


Figure 7.2. Discriminant analysis of specimen data using 171 characters

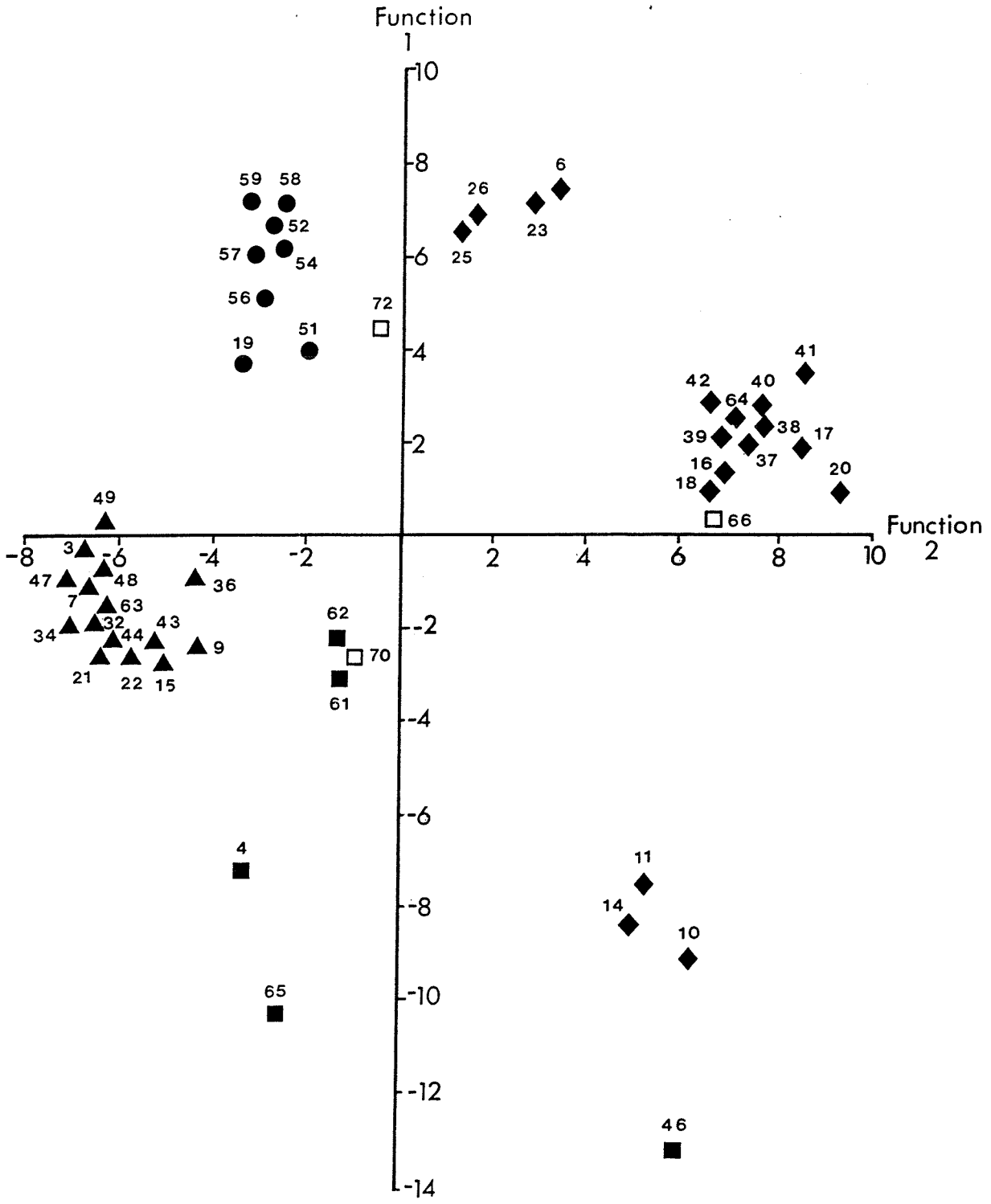


Table 7.1. Key to phenetic results taxon numeric codes and symbols

Taxon Code	Sensu Kupicha Section	Taxon Name
01	Peregrinae	V. aintabensis
02	Hypechusa	V. anatolica
03	Hypechusa	V. assyriaca
04	Atossa	V. balansae
05	Vicia	V. barbazitae
06	Faba	V. bithynica
07	Hypechusa	V. ciliatula
08	Vicia	V. cuspidata
09	Hypechusa	V. esdraelonensis
10	Faba	V. faba
11	Faba	V. faba subsp. faba var. faba
12	Faba	V. faba subsp. faba var. equina
13	Faba	V. faba subsp. faba var. minor
14	Faba	V. faba subsp. paucijuga
15	Hypechusa	V. galeata
16	Faba	V. galilaea
17	Faba	V. galilaea subsp. faboidea
18	Faba	V. galilaea subsp. galilaea
19	Vicia	V. grandiflora
20	Faba	V. hyaeniscyamus
21	Hypechusa	V. hybrida
22	Hypechusa	V. hyrcanica
23	Faba	V. johannis
24	Faba	V. johannis var. ecirrhosa
25	Faba	V. johannis var. johannis
26	Faba	V. johannis var. procumbens
27	Vicia	V. lathyroides
28	Hypechusa	V. lutea
29	Hypechusa	V. lutea subsp. lutea
31	Hypechusa	V. lutea subsp. vestita
32	Hypechusa	V. melanops
33	Hypechusa	V. melanops var. melanops
34	Hypechusa	V. melanops var. loiseaui
35	Peregrinae	V. michauxii
36	Peregrinae	V. mollis
37	Faba	V. narbonensis
38	Faba	V. narbonensis var. aegyptiaca
39	Faba	V. narbonensis var. affinis
40	Faba	V .narbonensis var. jordanica
41	Faba	V. narbonensis var. narbonensis
42	Faba	V. narbonensis var. salmonea
43	Hypechusa	V. noeana
44	Hypechusa	V. noeana subsp. noeana
45	Hypechusa	V. noeana subsp. megalodonta
46	Atossa	V. oroboides
47	Hypechusa	V. pannonica
48	Hypechusa	V. pannonica subsp. pannonica
49	Hypechusa	V. pannonica subsp. striata
50	Peregrinae	V. peregrina

Taxon Code	*Sensu* Kupicha Section	Taxon Name
51	Vicia	V. pyrenaica
52	Vicia	V. sativa
53	Vicia	V. sativa subsp. amphicarpa
54	Vicia	V. sativa subsp. cordata
55	Vicia	V. sativa subsp. incisa
56	Vicia	V. sativa subsp. macrocarpa
57	Vicia	V. sativa subsp. nigra
58	Vicia	V. sativa subsp. nigra var. segetalis
59	Vicia	V. sativa subsp. sativa
60	Atossa	V. sepium
61	Atossa	V. sepium subsp. sepium
62	Atossa	V. sepium subsp. montana
63	Hypechusa	V. sericocarpa
64	Faba	V. serratifolia
65	Atossa	V. truncatula
66	Unknown	V. kalakhensis
67	Unknown	V. dionysiensis
68	Unknown	V. qatmensis
69	Unknown	V. tigridis
70	Atossa	V. sepium subsp. eriocalyx
71	Unknown	V. eristalioides
72	Vicia	V. sativa subsp. devia

Both Figures 7.1 and 7.2 show the taxa forming major clusters of related taxa with a few taxa, such as taxon 46 (V. oroboides), remaining isolated from these clusters. There are distinct differences in cluster membership between the two scatter diagrams. As this initial step of dividing the complete subgeneric data set into smaller units will affect the later analysis steps, it was decided to analyse the DISCRIMINANT output further. Thus the first nine canonical discriminant functions for each taxon were used as variables for average linkage cluster analysis via CLUSTAN procedure CLUSTER.

The two dendrograms showing the results of these average linkage analyses are drawn in Figures 7.3 and 7.4. These relate to the analysis of all 1539 specimens using the 123 vegetative and inflorescence characters and for the 566 OTUs using 171 characters respectively. As with the results of the previous analysis, similar clusters of taxa are found. The fact that these clusters are evident following both analyses, strengthens the hypothesis that the taxa within clusters are more similar to each other than they are to taxa contained in other clusters. However, this analysis still involved the subjective a priori attribution of specimens to taxa and so a further analysis was undertaken to group specimens and taxa into major clusters, which could then be followed by a less subjective analysis of the internal structure of these major clusters.

Theoretically, it should be possible to analyse the complete data set of 1539 specimens by 171 characters using CLUSTAN (see Wishart, 1987). However, analysis of this size data sets is not possible. After many false starts, Dr. J. Henderson of CLUSTAN Ltd managed to produce a "one-off" version of CLUSTAN, which increased the current capabilities of CLUSTAN. Dr. Henderson managed to undertake average linkage cluster analysis using squared Euclidean distance via procedure CLUSTER on two subsets of the complete data set,

Figure 7.3. Average linkage cluster analysis of specimen data using first nine discriminant functions and 123 characters.

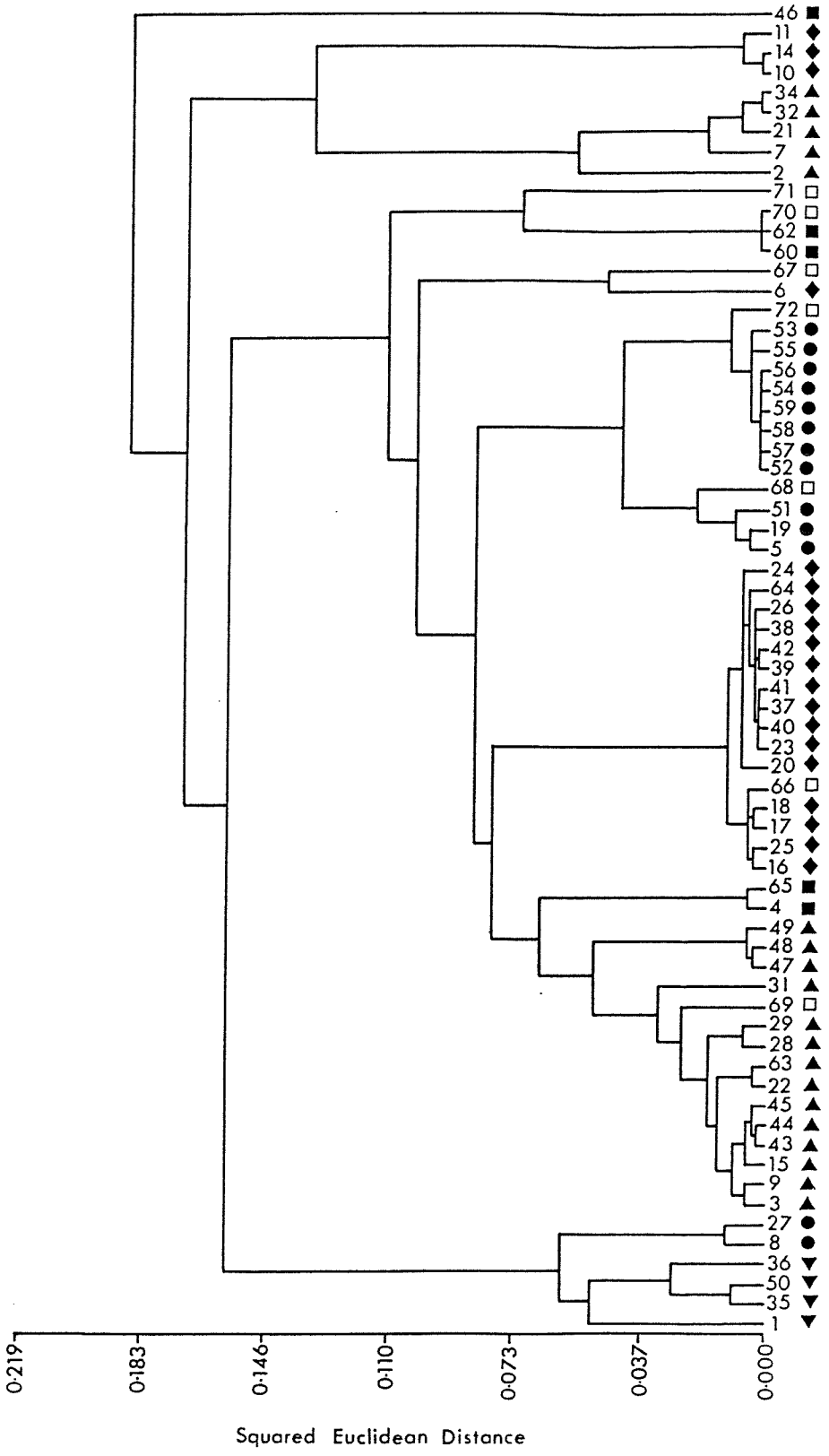
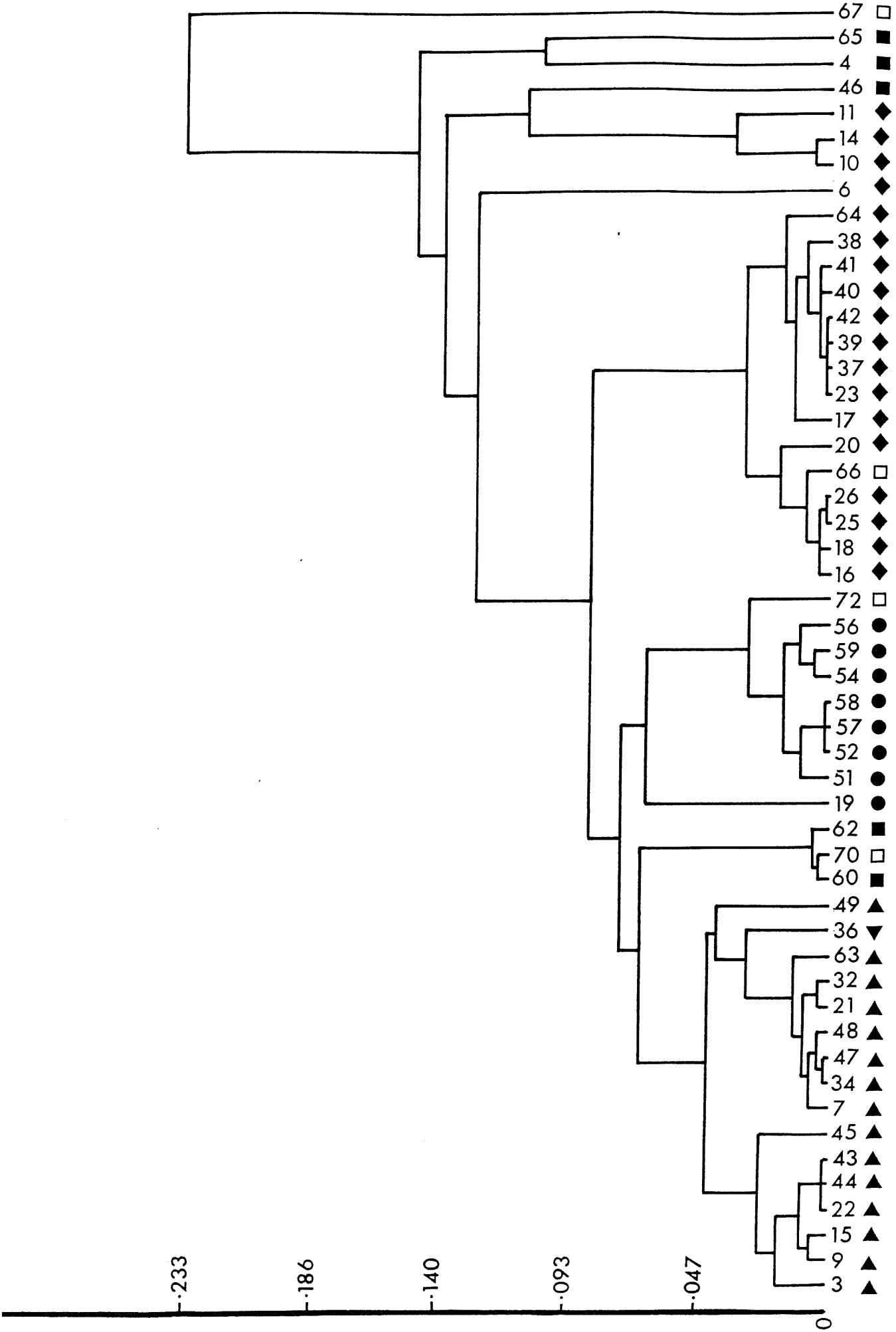


Figure 7.4.
Average linkage cluster analysis of specimen data using first nine discriminant functions and 171 characters.



Squared Euclidean Distance

each comprising 999 OTUs and using the 123 vegetative and inflorescence characters. For the initial average linkage analysis, the first 999 OTUs were used and for the second run, the last 999 OTUs of the complete data set were analysed. Fruit and seed characters, which contained numerous missing values, were excluded. As the complete data set comprised 1539 specimens, it did allow an element of replicated analysis between these two analyses.

The results of the two average linkage cluster analyses are shown in the dendrograms drawn in Figures 7.5 and 7.6, for the first 999 and second 999 OTUs respectively. These dendrograms have been simplified to show clusters of OTUs and not individual OTU links. In identifying clusters of OTUs, a level of dissimilarity was taken which tended to maintain uniform taxon clusters. This threshold dissimilarity level was 0.35 for the first 999 OTUs and 0.31 for the second 999 OTUs, which produced 46 and 42 OTU clusters respectively, as indicated in Figures 7.5 and 7.6. For both figures the clusters are labelled consecutively from A to T'. Ideally it would have been better to indicate cluster membership in the figures, but this would have made their interpretation more difficult. So cluster membership is indicated in Tables 7.2 and 7.3 which refer to Figures 7.5 and 7.6 respectively.

Figure 7.5. Average linkage cluster analysis of specimen data using 123 characters for the first 999 specimens.

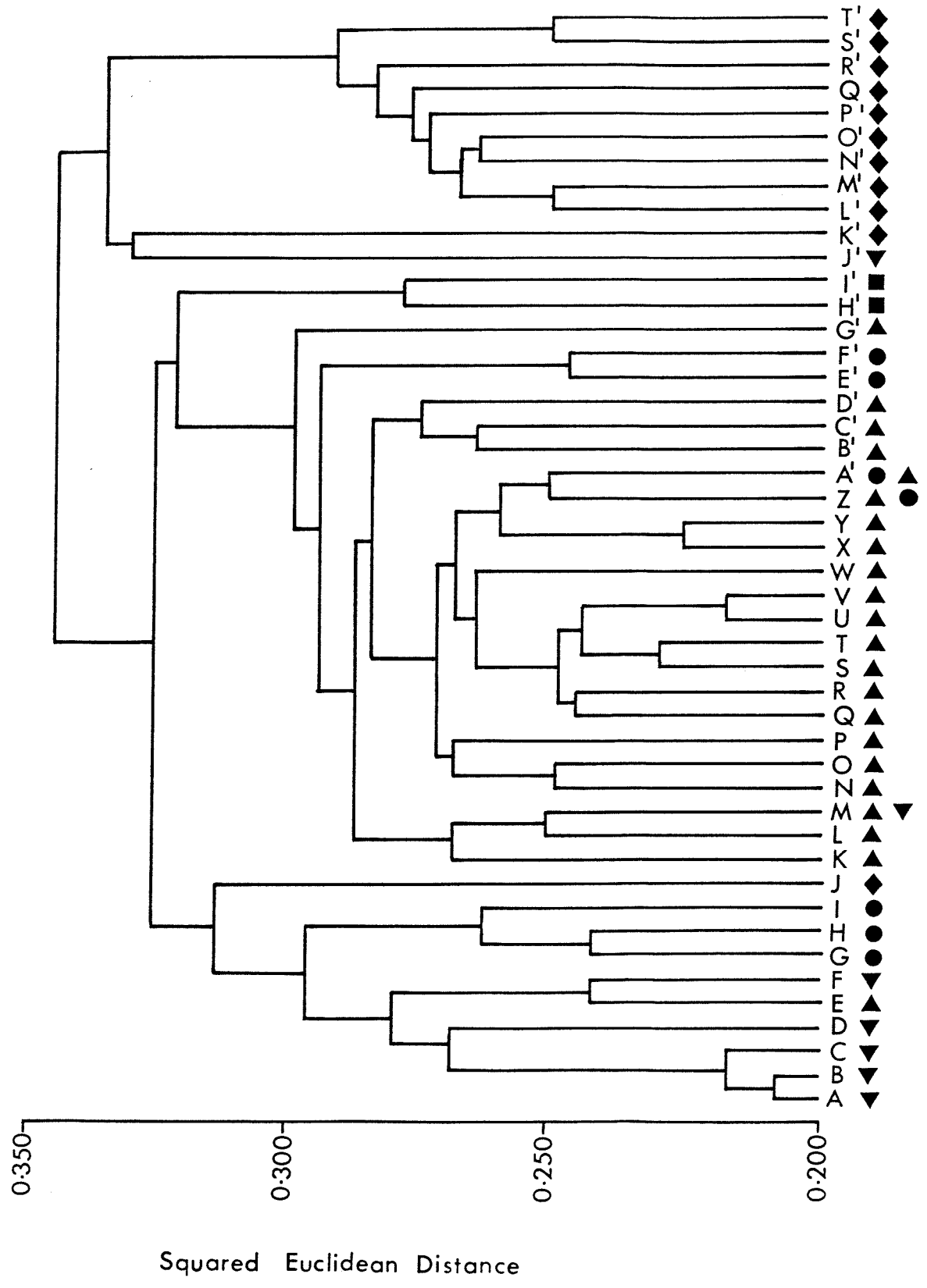


Figure 7.6. Average linkage cluster analysis of specimen data using 123 characters for the last 999 specimens.

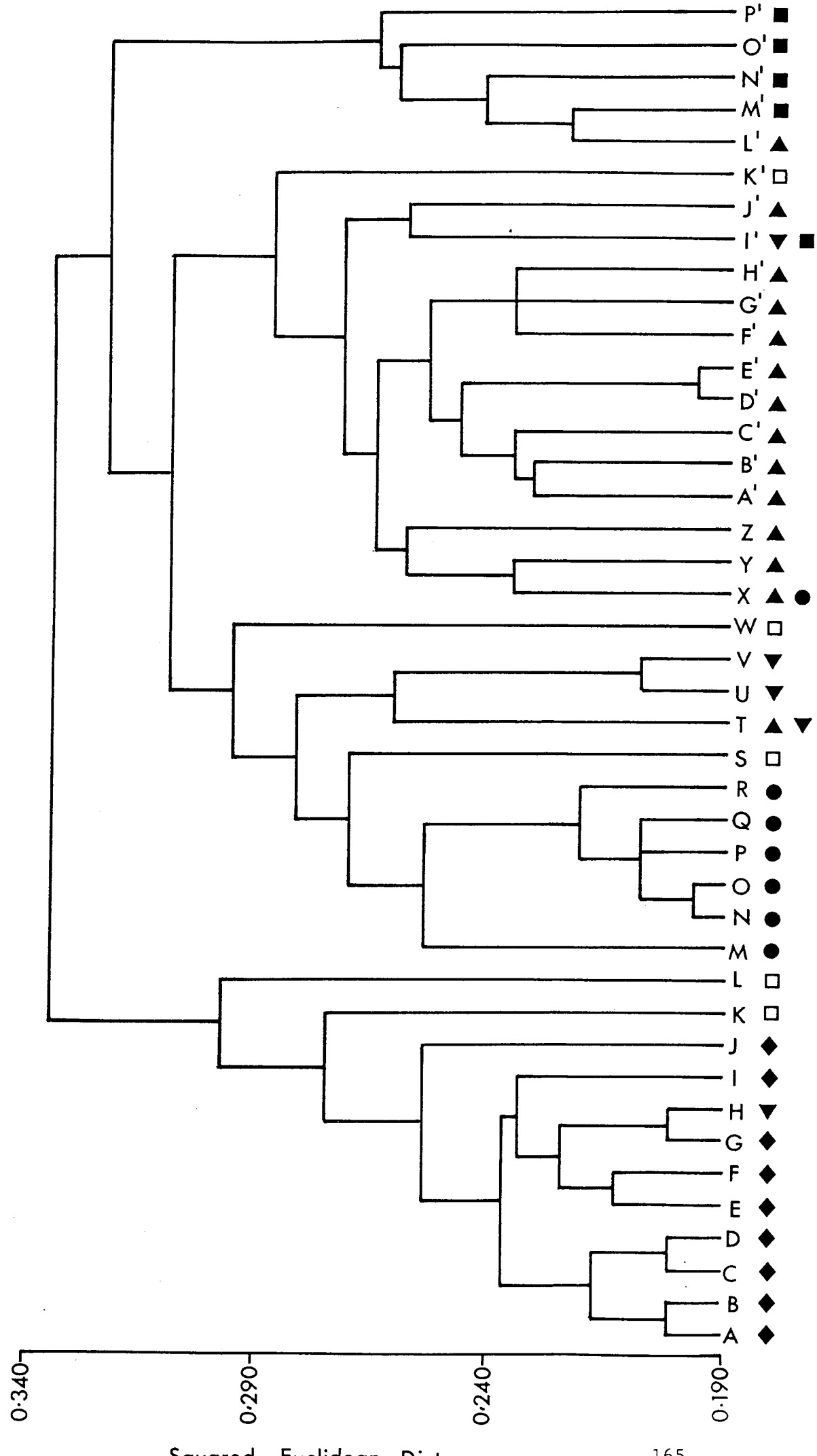


Table 7.2 Cluster membership of first 999 OTUs average linkage cluster analysis (relates to Figure 7.5)

Cluster	Taxa (number of OTUs x species, sectional symbol)		
A	7 x V. aintabensis		
B	1 x V. aintabensis	+	21 x V. michauxii
C	1 x V. aintabensis		
D	1 x V. michauxii		
E	1 x V. lutea		
F	17 x V. mollis		
G	1 x V. barbazitae	+	49 x V. cuspidata
+	2 x V. lathyroides		
H	1 x V. cuspidata	+	57 x V. lathyroides
I	1 x V. lathyroides		
J	86 x V. bithynica		
K	18 x V. anatolica	+	1 x V. hybrida
L	8 x V. ciliatula		
M	1 x V. hybrida	+	1 x V. hyrcanica
+	25 x V. melanops	+	1 x V. michauxii
N	12 x V. assyriaca		
O	7 x V. pannonica		
P	1 x V. assyriaca		
Q	1 x V. assyriaca	+	1 x V. noeana
R	27 x V. noeana		
S	2 x V. esdraelonensis		
T	28 x V. galeata		
U	24 x V. hyrcanica		
V	8 x V. noeana		
W	1 x V. galeata	+	1 x V. noeana
X	1 x V. ciliatula	+	9 x V. pannonica
Y	8 x V. melanops		
Z	1 x V. grandiflora	+	53 x V. hybrida
+	2 x V. mollis		
A'	1 x V. grandiflora	+	1 x V. lutea
B'	3 x V. hybrida	+	1 x V. hyrcanica
+	69 x V. lutea	+	1 x V. pannonica
C'	1 x V. lutea		
D'	1 x V. lutea		
E'	13 x V. barbazitae		
F'	1 x V. barbazitae	+	62 x V. grandiflora
G'	1 x V. galeata		
H'	22 x V. balansae		
I'	23 x V. oroboides		
J'	1 x V. aintabensis		
K'	33 x V. faba		
L'	1 x V. faba		
M'	9 x V. galilaea		
N'	15 x V. galilaea	+	85 x V. johannis
+	135 x V. narbonensis		
O'	11 x V. galilaea		
P'	1 x V. johannis	+	1 x V. narbonensis
Q'	1 x V. johannis		
R'	5 x V. galilaea	+	4 x V. johannis
S'	1 x V. galilaea		
T'	8 x V. hyaeniscyamus		

Table 7.3 Cluster membership of second 999 OTUs average linkage cluster analysis (relates to Figure 7.6)

Cluster	Taxa (number of OTUs x species, sectional symbol)		
A	27 x V. johannis	+	5 x V. narbonensis
B	5 x V. johannis	+	33 x V. narbonensis
C	6 x V. johannis	+	1 x V. serratifolia
D	4 x V. kalakhensis		
E	10 x V. narbonensis	+	72 x V. serratifolia
F	61 x V. narbonensis	+	6 x V. serratifolia
G	24 x V. narbonensis		
H	1 x V. peregrina		
I	4 x V. serratifolia		
J	2 x V. serratifolia		
K	3 x V. eristalioides		
L	2 x V. kalakhensis		
M	59 x V. lathyroides		
N	1 x V. lathyroides	+	222 x V. sativa
O	1 x V. sativa		
P	2 x V. sativa		
Q	21 x V. pyrenaica		
R	4 x V. sativa		
S	5 x V. qatmensis		
T	1 x V. lutea	+	19 x V. mollis
+	1 x V. sericocarpa		
U	21 x V. peregrina		
V	1 x V. michauxii	+	43 x V. peregrina
W	5 x V. dionysiensis		
X	65 x V. lutea	+	1 x V. pannonica
+	1 x V. sativa		
Y	1 x V. lutea		
Z	1 x V. lutea		
A'	1 x V. lutea	+	35 x V. sericocarpa
B'	36 x V. pannonica		
C'	1 x V. pannonica	+	1 x V. sericocarpa
D'	8 x V. melanops	+	1 x V. pannonica
E'	8 x V. pannonica		
F'	25 x V. melanops		
G'	2 x V. noeana		
H'	34 x V. noeana	+	1 x V. pannonica
I'	1 x V. michauxii	+	1 x V. truncatula
J'	3 x V. sericocarpa		
K'	2 x V. tigridis		
L'	1 x V. noeana		
M'	17 x V. truncatula		
N'	21 x V. oroboides		
O'	2 x V. oroboides		
P'	47 x V. sepium		

The first conclusion to be drawn from the results of the average linkage cluster analysis using the two data subsets, is that specimens identified at the start of the project as belonging to certain taxa do form distinct, uniform taxon clusters. Secondly, clusters of taxa do form supra-specific groupings, which approximate the classification of subgenus Vicia sensu Kupicha (1976).

If the information provided by the results of the six analyses, reported above, is synthesised, similar groupings of taxa can be seen to be clustered by each analysis method. Of the 67 taxa included a priori in the analysis, 15 consistent specimen and taxon clusters are indicated, 63 taxa are included in 11 clusters and four taxa and their specimens form individual taxon clusters. A summary of which taxa are included in these 15 clusters is provided in Table 7.4 below:

Table 7.4 Taxon membership of fifteen major clusters

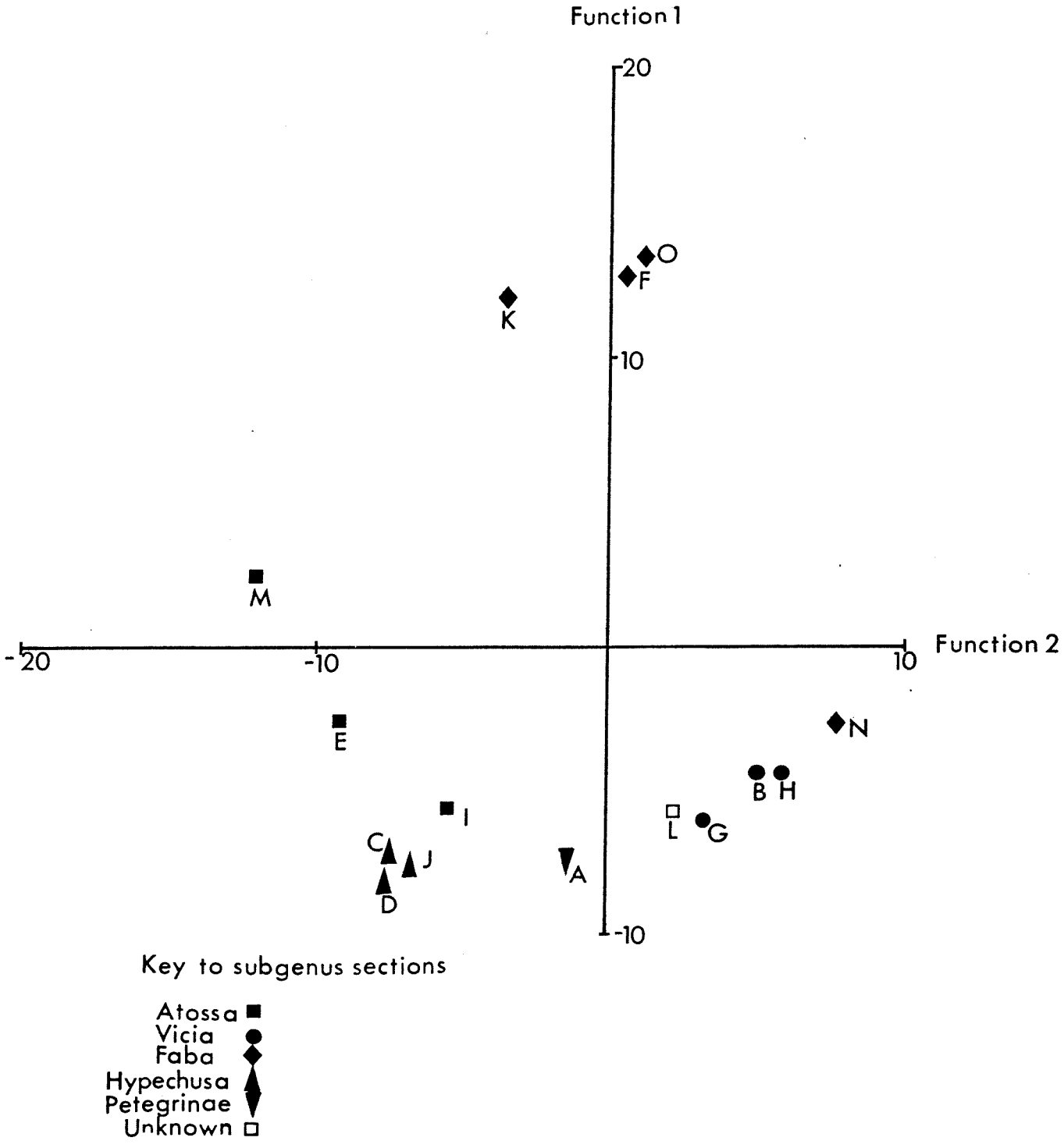
Cluster	Taxa Codes	Sensu Kupicha Section
A	V. aintabensis, V. michauxii, V. mollis, V. peregrina	Peregrinae
B	V. cuspidata, V. lathyroides	Vicia
C	V. assyriaca, V. esdraelonensis, V. galeata, V. hyrcanica, V. lutea (3 taxa), V. noeana (3 taxa), V. sericocarpa, V. tigridis	Hypechusa
D	V. pannonica (3 taxa)	Hypechusa
E	V. balansae, V. truncatula	Atossa
F	V. galilaea (3 taxa), V. hyaeniscyamus, V. johannis (4 taxa), V. narbonensis (6 taxa), V. serratifolia, V. kalakhensis	Faba
G	V. barbazitae, V. grandiflora, V. pyrenaica, V. qatmensis	Vicia
H	V. sativa (9 taxa)	Vicia
I	V. sepium (4 taxa)	Atossa

Cluster	Taxa Codes	*Sensu* Kupicha Section
J	V. anatolica, V. ciliatula, V. hybrida, V. melanops (3 taxa)	Hypechusa
K	V. faba (5 taxa)	Faba
L	V. dionysiensis	-
M	V. oroboides	Atossa
N	V. bithynica	Faba
O	V. eristalioides	Faba

It is interesting to note how these groupings of taxa are closely correlated with Kupicha's (1976) classification of subgenus *Vicia*. None of the groupings indicated by the analyses contains taxa from different sections *sensu* Kupicha. One mono-specific group, *V. dionysiensis* (L), was not known to Kupicha and cannot be placed in her classification of subgenus *Vicia* using the characters she cites. Though using her key to the subgenera, this species is clearly a member of subgenus *Vicia*. The clusters of taxa do however split Kupicha's sections, but this will be discussed below.

To estimate the degree of distinction between the fifteen major clusters and to evaluate the likelihood of specimens being included in an inappropriate cluster, every specimen was reidentified as belonging to one of these fifteen major clusters. The data were then analysed using SPSS^{x} procedure DISCRIMINANT using vegetative and floral characters alone (character set 7 in Appendix 3). The results of this analysis are shown in Figure 7.7. These results show the spatial closeness of major clusters from the same sections *sensu* Kupicha (1976). For instance, the close juxtaposition of the three sect. *Hypechusa* and three sect. *Vicia* clusters should be noted. However, sect. *Atossa*, and more notably, sect. *Faba* are spatially divergent. Particularly within the latter, *V. bithynica* is distant from the other sect. *Faba* taxa.

Figure 7.7. Discriminant analysis of taxon cluster data using 123 characters.



There is a possibility that specimens may have been misidentified during the initial identification and so attributed to the wrong taxon during the above analysis. It is, however, assumed that if a specimen was misidentified, the misplacement will be within one of the fifteen groupings identified. This assumption is likely to be valid, as all closely related taxa, that might be confused, will be included in the same major grouping. If specimens were misplaced between groupings this would become evident during the analysis of the individual groupings. This analysis should be seen as a practical means of splitting the complete data set into major clusters that could then be analysed more rigorously. The problems of analysing such a comparatively large data set were thus circumvented.

7.2.2 Taxon Location

Having split the complete specimen data set into eleven multi-taxon clusters and four single taxon clusters, each multi-taxon grouping was then analysed. This analysis was undertaken using CLUSTAN procedure CLUSTER, using the dissimilarity coefficient, squared Euclidean distance and centroid linkage cluster analysis. Each multi-taxon cluster was analysed using a character subset selected for its importance in that particular group.

Character selection for each multi-taxon group analysis was undertaken using SPSS^{x} procedure DISCRIMINANT (discussed in thesis section 6.3.2). This procedure produces F values for each character, which indicate the discriminating value of each character to distinguish the groups included in the data set. This provides a more objective means of assessing "good discriminating characters" for that particular data set. An F value in excess of 9.99 was considered sufficient to warrant the inclusion of the character in the analysis. This level of F was selected, as it tended to include characters shown to be useful by previous authors and those thought useful within the current study. This method of character selection was only

possible for vegetative and inflorescence characters as legume and seed characters contained too high a proportion of missing data. Legume and seed characters were selected subjectively.

A small proportion of vegetative and inflorescence characters were included even though their F value was lower than 9.99, if they were judged to be taxonomically important. In these cases the characters had proved useful to previous authors and their F values were only slightly lower than 9.99 and so it was considered expedient to include them in the analysis. The inclusion of characters with a lower F value was necessitated when analysing groups of closely related taxa (clusters D, I and K), for which all characters had lower F values than 9.99. The more closely related the taxa the more difficult it was to find good discriminating characters and the lower were the character's F values. This was especially evident for characters being used to distinguish infra-specific taxa. Characters were excluded from the complete (171 character) character set using the program RADISH written by Mr. R. Crust (Dept. Of Biology, The University, Southampton).

The grouping which contains V. aintabensis, V. michauxii, V. mollis and V. peregrina was analysed using 80 characters (character set A in Appendix 3). The dendrogram resulting from the cluster analysis is shown in Figure 7.8. The results indicate nine clusters, of which five are composed of single OTUs (specimens). These five distinct specimens (Haussknecht 24/04/1865, K; Haussknecht 24/04/1865, G; Willdenow s.n., K; Kotscky 1845, K; and Foures 983, MPU) each contained a high level of missing data, which possibly explains their separation by this analysis. The bulk of the specimens are contained in four clusters. The most distinct of these multiple OTU clusters are V. mollis and V. peregrina, each of which contain specimens of a single taxon. One of the other two multiple OTU clusters also contains specimens of one taxon, V. michauxii. The other contains 16 V. aintabensis

Figure 7.8. Centroid linkage cluster analysis of V. aintabensis, V. michauxii, V. mollis and V. peregrina group.

Numbers prior to specific epithets indicate the number of specimens per cluster.

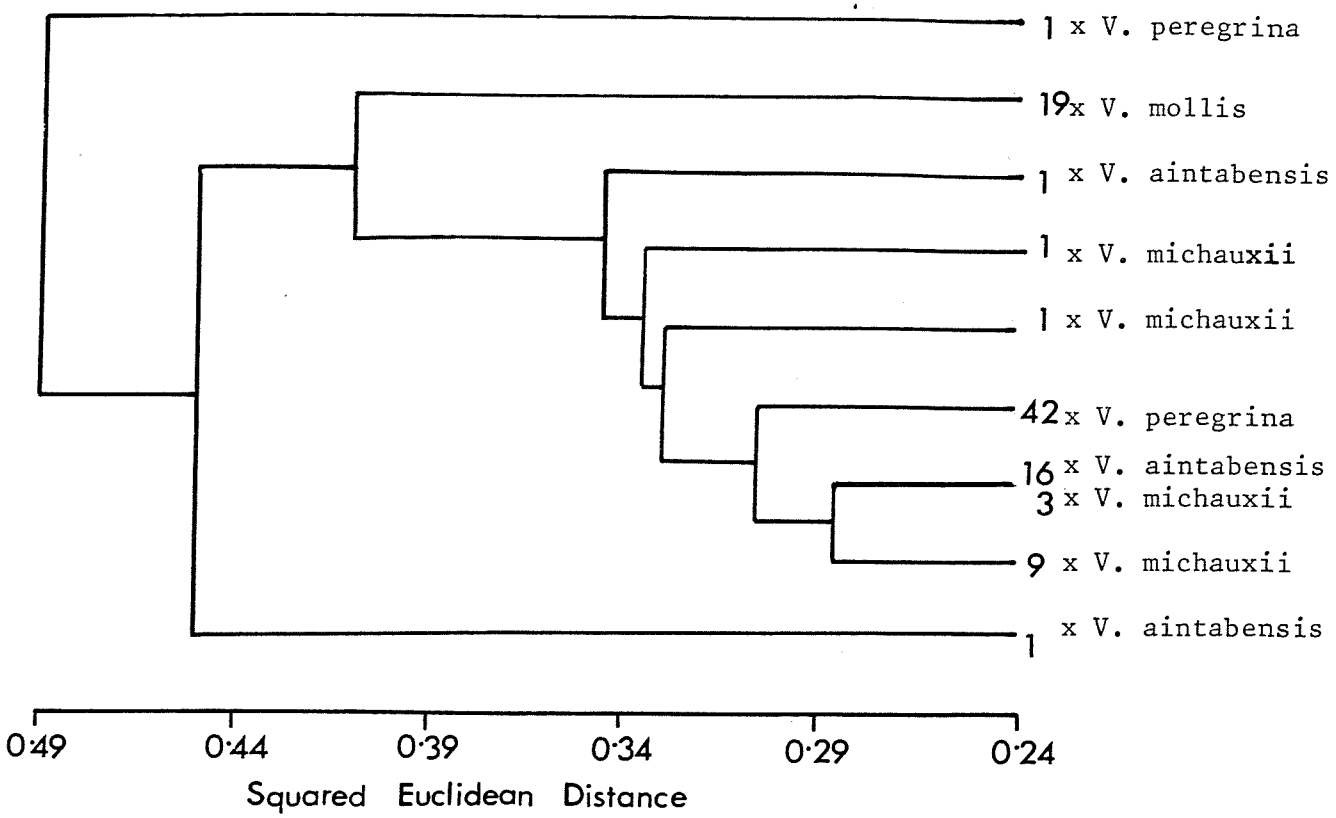
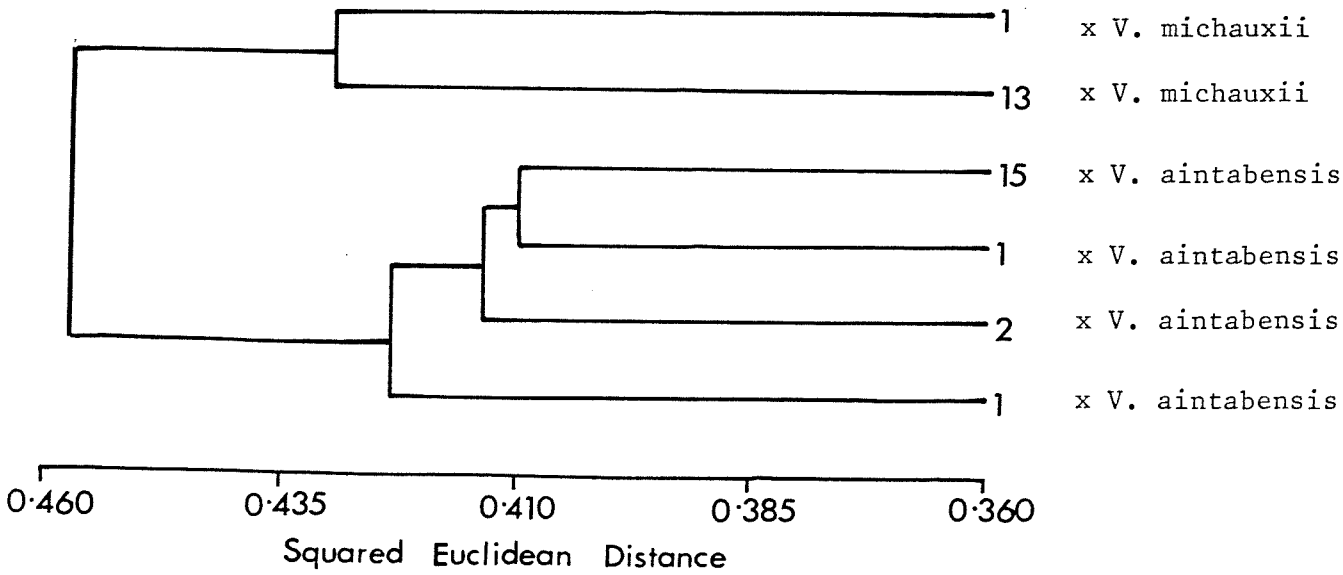


Figure 7.9. Centroid linkage cluster analysis of V. aintabensis and V. michauxii group.



with 3 V. michauxii specimens. The analysis has failed, using this character set, to differentiate between specimens of these two closely related taxa.

V. aintabensis and V. michauxii are indisputably closely related and Ponert (1973) goes so far as to regard them as subspecies of V. peregrina. For this reason the data for these two species were reanalysed using 14 characters (character set R in Appendix 3). The results of this cluster analysis are shown in Figure 7.9. These 14 characters do effectively distinguish the two species. Even the specimens that were distinct in the previous analysis join their specific cluster. The key characteristics for distinguishing these two species are legume and seed, dimension and shape. So this presents identification problems if legume and seed data are absent. If these data are absent the results of analysis would tend to indicate the two taxa are more closely related than they would be if a full data set was scored. For this reason I retain V. aintabensis and V. michauxii as distinct species.

The grouping of V. cuspidata and V. lathyroides was analysed using 72 characters (character set B in Appendix 3). The dendrogram, resulting from the cluster analysis, is shown in Figure 7.10. The results indicate a clear distinction between the two species. One specimen of V. lathyroides (Anon. 27/05/1874, SPN) remains distinct from the two main clusters. This specimen did have some of the key characters missing, which explains its failure to link with the V. lathyroides cluster prior to forming links with the V. cuspidata cluster.

The large grouping, from sect Hypechusa sensu Kupicha (1976), containing V. assyriaca, V. esdraelonensis, V. galeata, V. hyrcanica, V. lutea, V. noeana, V. sericocarpa and V. tigridis, was analysed using 71 characters (character set C in Appendix 3). The results of the centroid linkage cluster

Figure 7.10. Centroid linkage cluster analysis of V. cuspidata and V. lathyroides group.

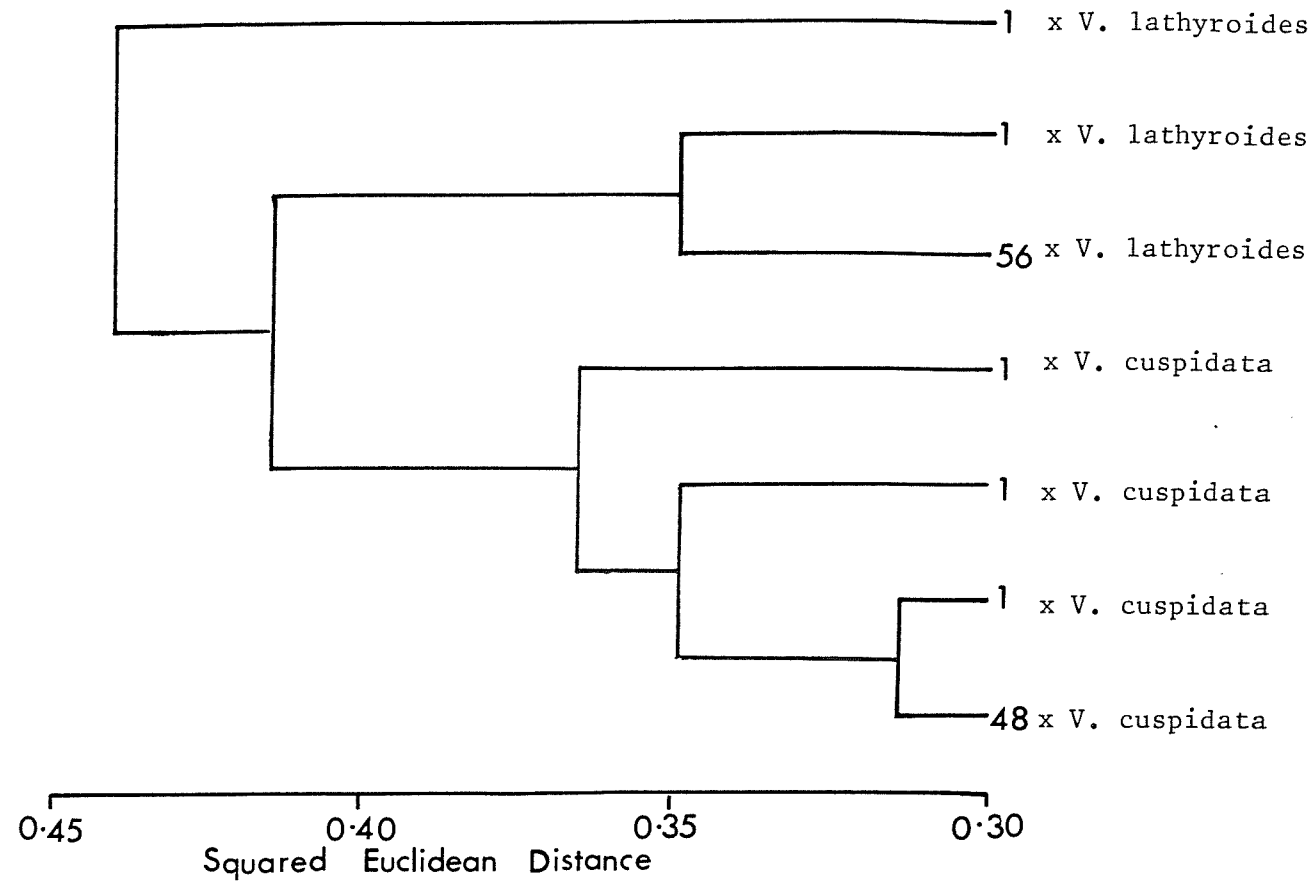
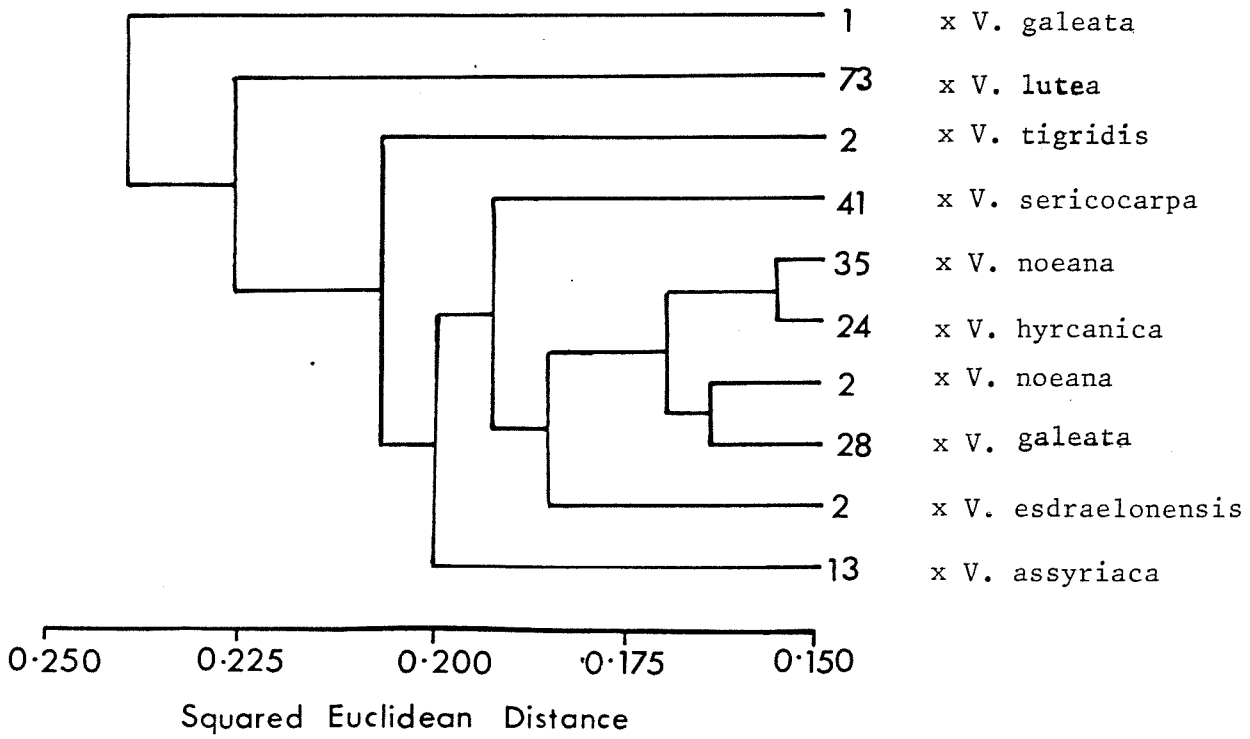


Figure 7.11. Centroid linkage cluster analysis of V. assyriaca, V. esdraelonensis, V. galeata, V. hyrcanica, V. lutea, V. sericocarpa and V. tigridis group.



analysis for this group is shown in Figure 7.11. The first point to note from the dendrogram is the overall formation of specific clusters, containing specimens of one species. However, two species, V. noeana and V. galeata, form two clusters each. The bulk of the V. noeana specimens form one cluster, but two of them do not link with the other 34 specimens, linking instead to the main V. galeata specimen cluster. These two species with V. hyrcanica form a closely related complex and this is reflected in the dendrogram. Two authors, Townsend (1967) and Ponert (1973), question the specific distinction between these three species. So the relationships between these species will be re-examined in the detailed taxon analysis (thesis section 7.4.3) to establish whether the three warrant specific status. The most distinct cluster of the analysis is composed of one V. galeata specimen. As in the previous analyses, specimens with a high proportion of missing data tend to take aberrant positions in the dendrogram.

Of the eight species analysed in the above cluster analysis, two, V. noeana and V. lutea, have been widely accepted as including infra-specific taxa. A priori, V. noeana was considered to be composed of two subspecies (noeana and megalodonta) and V. lutea of two subspecies (lutea and vestita) and two varieties of subsp. lutea (lutea and laevigata). The two subsp. of V. noeana are distinguished on the basis of calyx dimension and colour, and leaflet shape. These characteristics were used in the analysis of the 37 V. noeana specimens. The results of the cluster analysis using 10 characters (character set Q in Appendix 3) are shown in Figure 7.12. A clear distinction between the 34 specimens of subsp. noeana and the 3 specimens of subsp. megalodonta is indicated.

The three subspecific taxa of V. lutea are distinguished on the basis of overall pubescence and presence of tubercle based hairs. Using 19 characters, the 73 specimens of V.

Figure 7.12. Centroid linkage cluster analysis of V. noeana specimens.

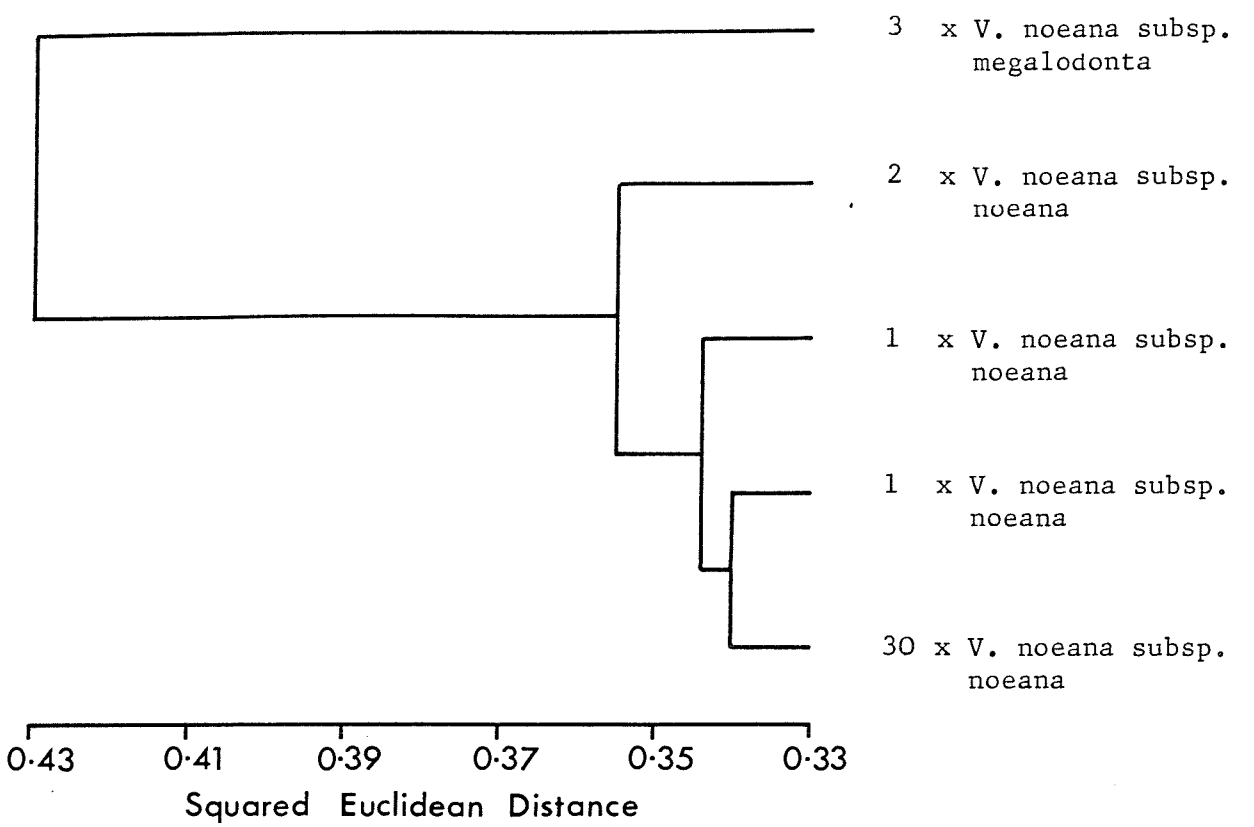
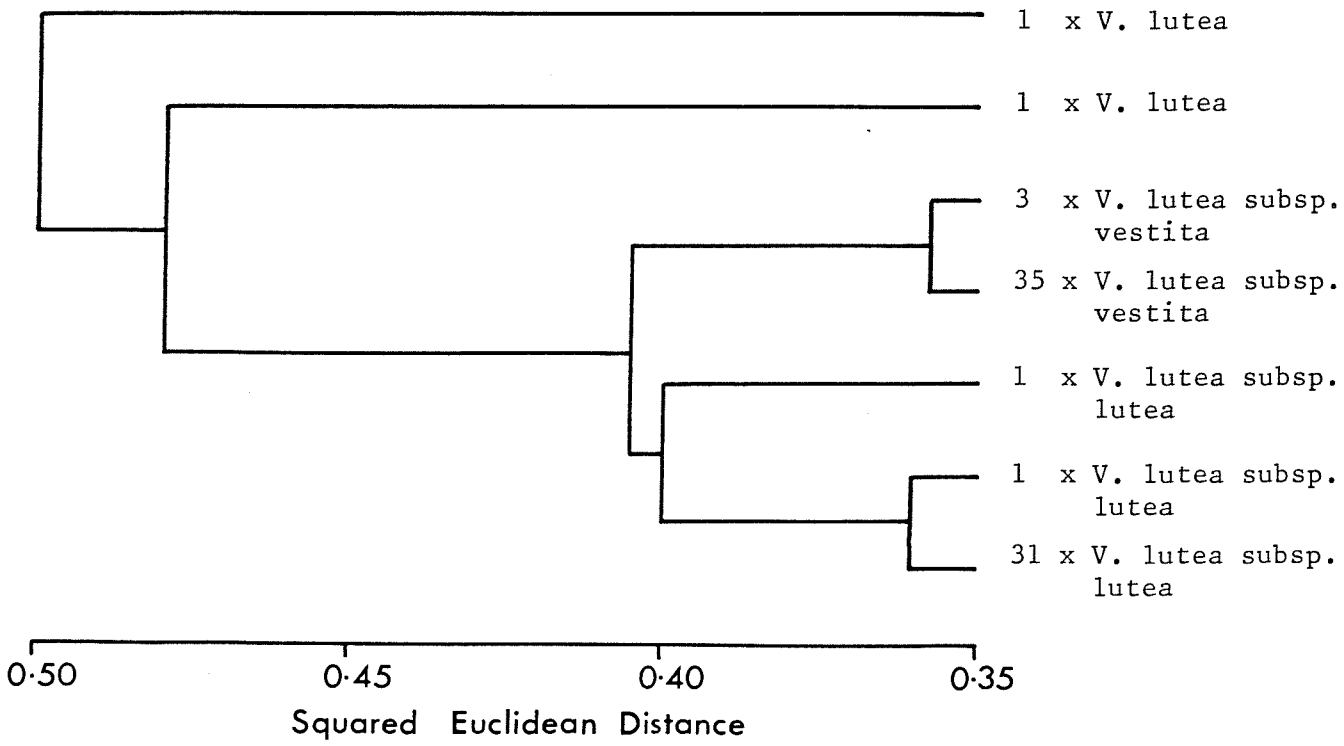


Figure 7.13. Centroid linkage cluster analysis of V. lutea specimens.



lutea were analysed (character set N in Appendix 3) and the resulting dendrogram is shown in Figure 7.13. Seventy one of the seventy three V. lutea specimens examined are included in the two major clusters. These two major clusters each contain specimens of one taxon, subsp. vestita and subsp. lutea specimens respectively. No third cluster containing subsp. lutea var. laevigata specimens was identified. None of the specimens investigated showed the character combination of subsp. lutea var. laevigata as described by Smith (1798) and it is suggested that this taxon is a synonym of subsp. lutea. The two most distinct clusters are comprised of single specimens. Both these specimens lacked legumes and it proved impossible to attribute them to either of the two subspecies.

V. pannonica was identified as a single species group by the initial analysis of the complete specimen data. This species was considered a priori to contain two subspecies and the 49 specimens of this taxon were analysed using 11 corolla colour and dimension characters (character set D in Appendix 3). The results of the cluster analysis are shown in Figure 7.14. The dendrogram indicates three major clusters, one containing subsp. striata, the second containing subsp. pannonica and a third, containing those specimens that could not be attributed to either subspecies due to missing data problems. The flowers of the two subspecies have different colours and in older specimens the colour fades. As specimens age they become impossible to score for this important character and so can only be identified as V. pannonica. One misplaced specimen of subsp. striata was included in the subsp. pannonica cluster. The misplacement of this specimen (Maxted 1179, SPN) can be explained by the deformation of it s flower.

Another group, identified in the initial complete specimen data analysis, is the grouping of V. balansae and V. truncatula. These are closely related and Stankevich (1988) does not consider them sufficiently distinct to warrant

Figure 7.14. Centroid linkage cluster analysis of V. pannonica specimens.

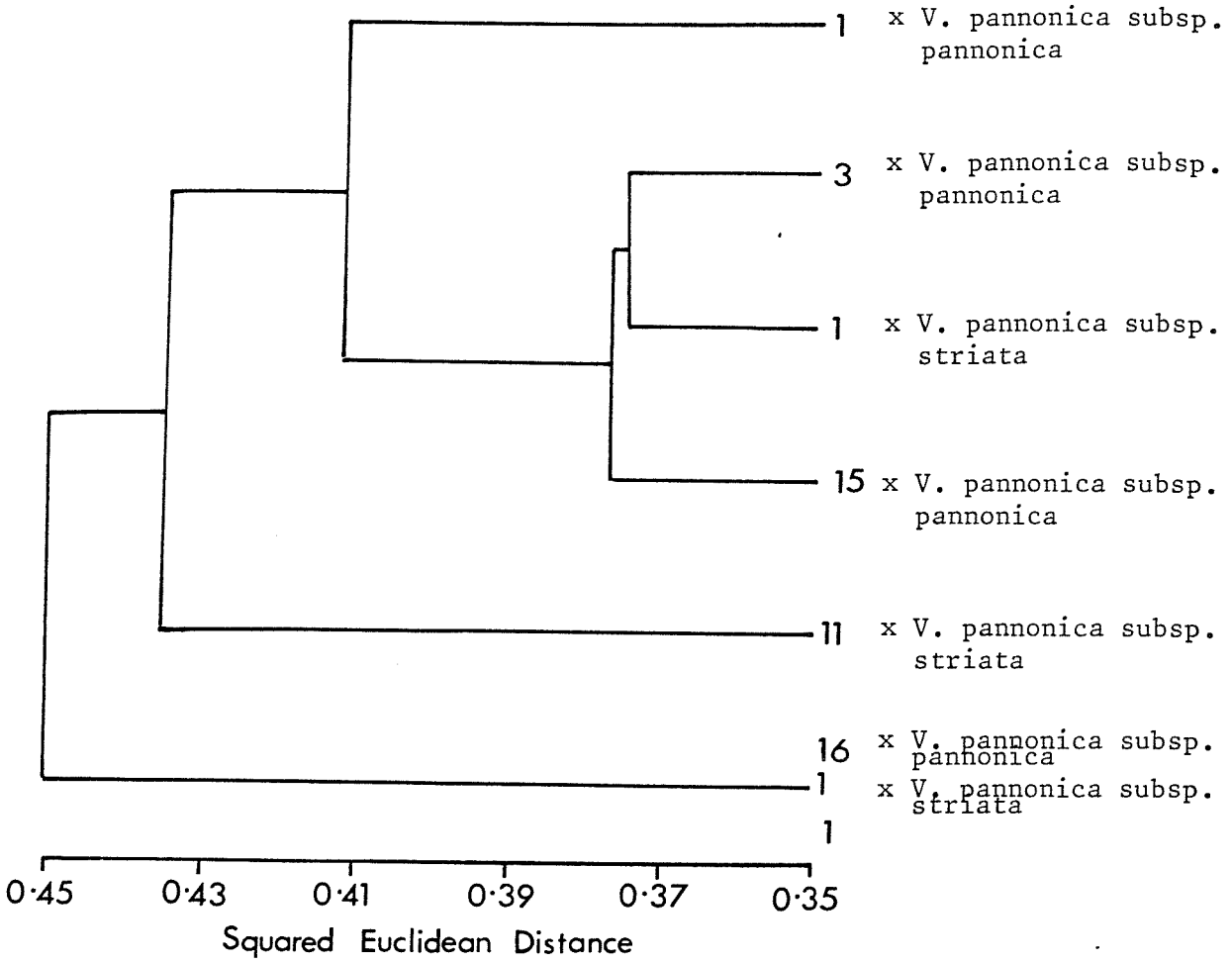
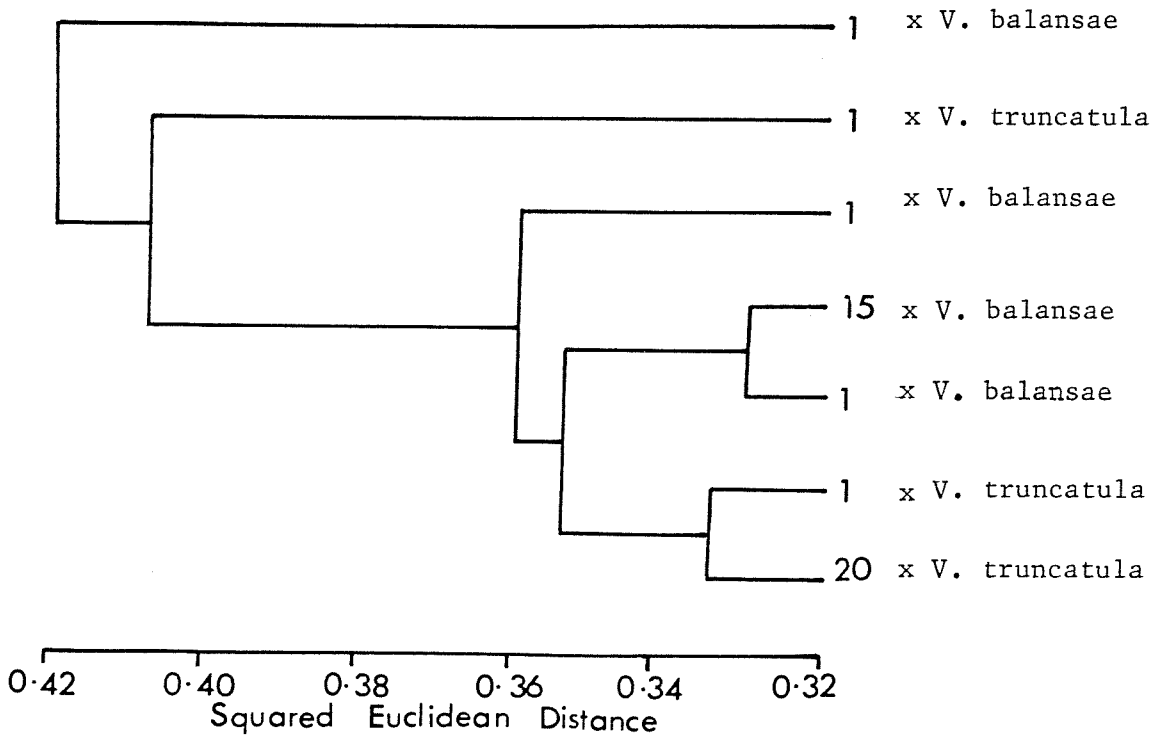


Figure 7.15. Centroid linkage cluster analysis of V. balansae and V. truncatula group.



specific status. The 40 specimens were analysed using 47 characters (character set E in Appendix 3) and the results are shown in Figure 7.15. The majority of the specimens cluster in two distinct clusters, each containing one species. There are also three single specimen clusters, two contain one V. balansae specimen each and the other contains a single V. truncatula specimen. The reason for the comparative isolation of these three specimens into distinct clusters is unclear. The distinction of these two species is discussed in detail in the conspectus (Appendix 5). Stankevich reports the existence of specimens with intermediate characteristics and she has shown these specimens to me. However, the distinction between the two specific clusters of specimens is sufficient, that I consider they should be retained as distinct species.

In the initial analysis of the complete specimen data, specimens representing individual taxa cluster tightly together with, perhaps, a few aberrant specimens. However, for the group which contained V. galilaea, V. hyaeniscyamus, V. johannis, V. narbonensis, V. serratifolia and V. kalakhensis, the specimens tended at least in the initial analysis to form multi-taxon clusters. Several authors (Schäfer, 1973; Khattab, 1987; Khattab et al. 1989 and Greuter, 1989) consider this group of taxa, known as the V. narbonensis complex, to be very closely related or some even regard them as micro-species. When dealing with such a critical group it is essential that the full data set is available for analysis, for the taxonomic structure of the group to be clearly understood. In practice, however, this is rarely possible for herbarium based studies, as herbarium specimens can seldom be scored for the complete character set.

The 357 specimens of the V. narbonensis complex were analysed using 23 characters (character set F in Appendix 3). The results of the analysis are shown in Figure 7.16. As in the results of the initial analysis, multi-taxon clusters are

Figure 7.16. Centroid linkage cluster analysis of V. galilaea, V. hyaeniscyamus, V. johannis, V. kalakhensis, V. narbonensis and V. serratifolia group.

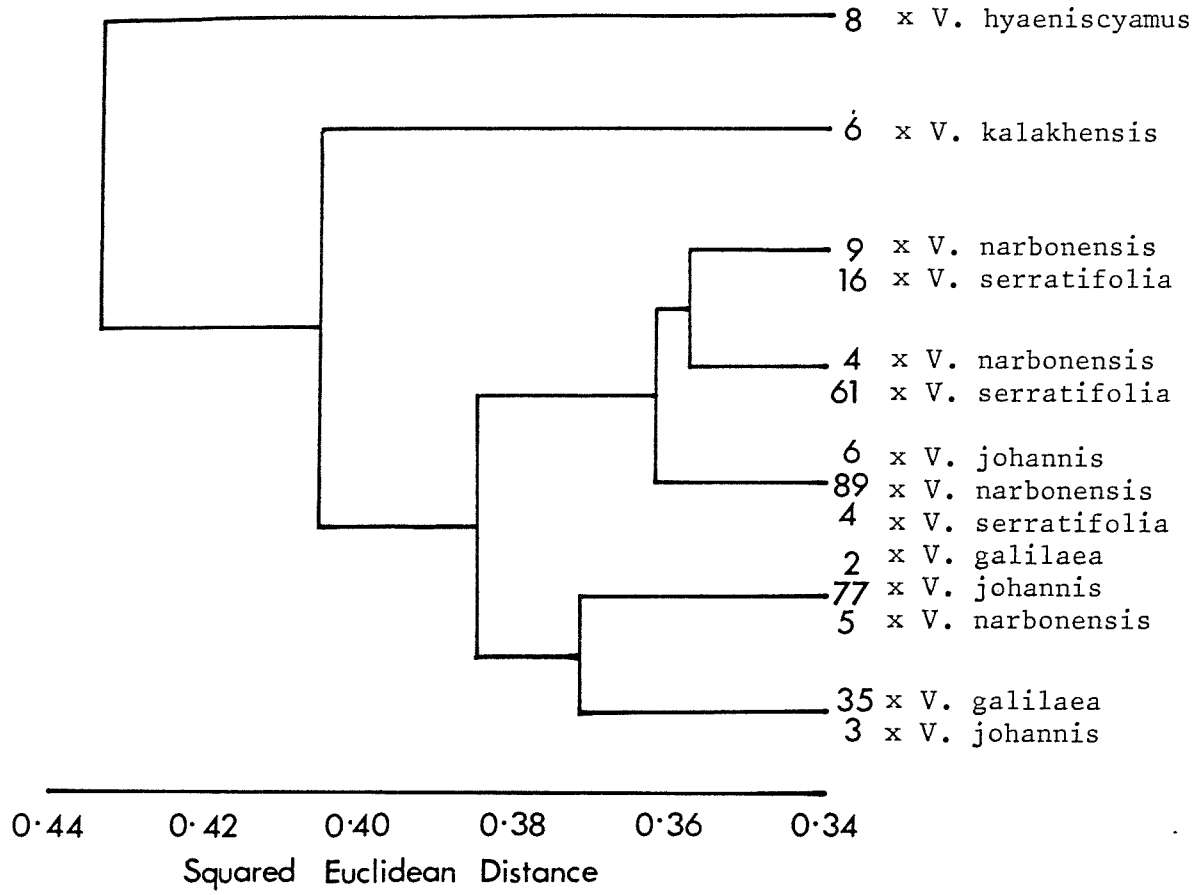
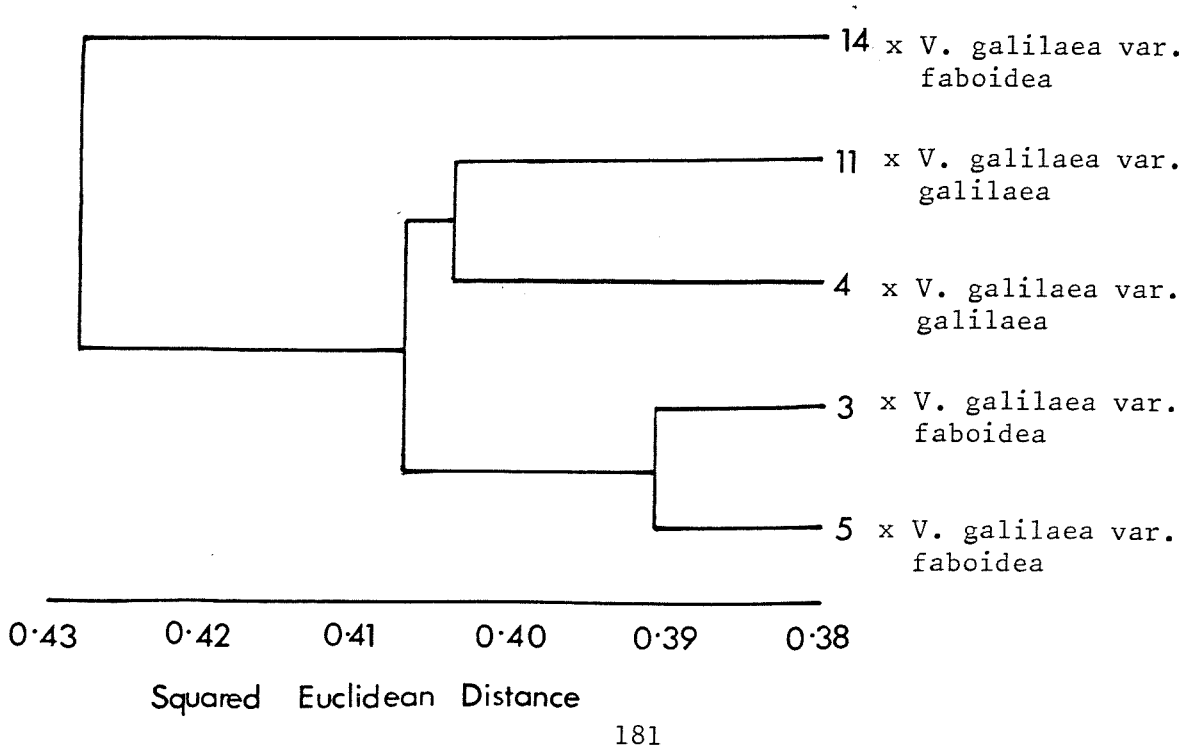


Figure 7.17. Centroid linkage cluster analysis of V. galilaea specimens.



formed. Only V. hyaeniscyamus and V. kalakhensis form discrete clusters. The most useful discriminating characters for this group are those associated with flower colour. As was found with V. pannonica, this subset of characters is almost invariably impossible to score from older herbarium specimens. V. galilaea, V. johannis, V. narbonensis and V. serratifolia are obviously closely related and the difficulty in obtaining a complete data set for these taxa makes their distinction problematic. However, the bulk of each of these four species' specimens are seen to be located in single species clusters. To study this closely related group more accurately, one would need to undertake uniform garden experiments and score key characters from living material, as was undertaken by Khattab (1987).

Within the V. narbonensis complex Schäfer (1973) considers three species (V. galilaea, V. johannis and V. narbonensis) to have subordinate taxa. The two varieties of V. galilaea are distinguished on the basis of leaflet pubescence, the presence of leaflet serrations and the relative size of the flowers and legumes. The 37 V. galilaea specimens were analysed using 23 characters (character set L in Appendix 3) and the results of the centroid linkage cluster analysis are shown in Figure 7.17. The dendrogram shows two main clusters, one containing 14 var. faboidea specimens and the other 23 specimens in two subclusters. The first subcluster contains the 15 var. galilaea specimens and the second 8 further var. faboidea specimens. This appears to indicate that var. galilaea is less variable than var. faboidea, which forms two cluster groups, one of which is more closely related to var. galilaea, than to other var. faboidea specimens. Once again, it should be stressed that the interpretation of the result is seriously impaired by missing data problems.

Several of the key characters used to distinguish the varieties of V. narbonensis are legume and seed

characteristics. This again meant there were problems arising from missing data. Five attempts were made, using various character combinations, to analyse all 148 V. narbonensis specimens. In each case the clusters formed did not relate to their accepted taxonomy. To resolve this problem the number of specimens was decreased to 60 and only those with a nearly complete data set were included in the analysis. These were analysed using 12 characters (character set P in Appendix 3). The results of the centroid linkage cluster analysis are shown in Figure 7.18.

The analysis splits the five varieties into ten clusters, four of which contain single specimens. The bulk of the specimens are contained in the remaining six clusters, one for each variety except for var. salmonea which is split into two. The smaller var. salmonea cluster contains five specimens (Al Eisawi 1311, RNG; Maxted, Ehrman et Khattab 2561 SPN; Simpson 467, K; Dinsmore 13090, K and Adamson s.n., BM). All five of these specimens have upper leaflets with more than six margin serrations and lower leaflets with a crenate margin. Although upper leaflet serrations are common in all V. narbonensis complex species, only V. serratifolia was thought to have more than six margin serrations per leaflet (Khattab, 1987). Crenation of the lower leaflet margin was thought to be diagnostic for V. narbonensis var. salmonea. These five specimens show intermediacy between V. serratifolia and V. narbonensis var. salmonea. On the basis of other characters, these specimens share more in common with V. serratifolia than with V. narbonensis and thus they have been attributed to V. serratifolia. Of the remaining V. narbonensis specimens, clusters of each of the five varieties can be seen, var. affinis and var. salmonea are most closely related, then var. jordanica and var. narbonensis, with var. aegyptiaca being the most remote.

Schäfer (1973) splits V. johannis into three varieties, distinguished by tendril presence, number of leaflets per leaf

Figure 7.18. Centroid linkage cluster analysis of V. narbonensis specimens.

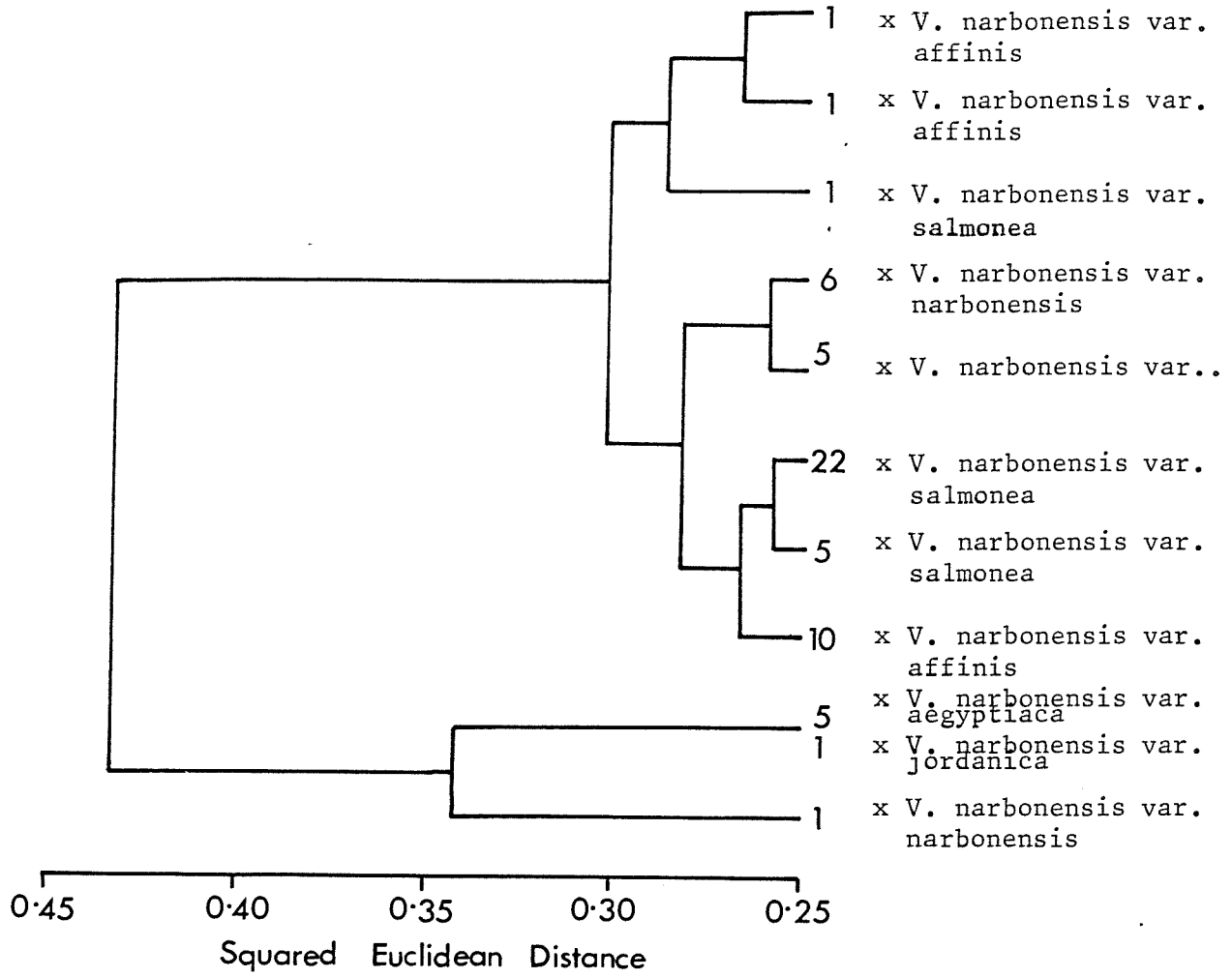
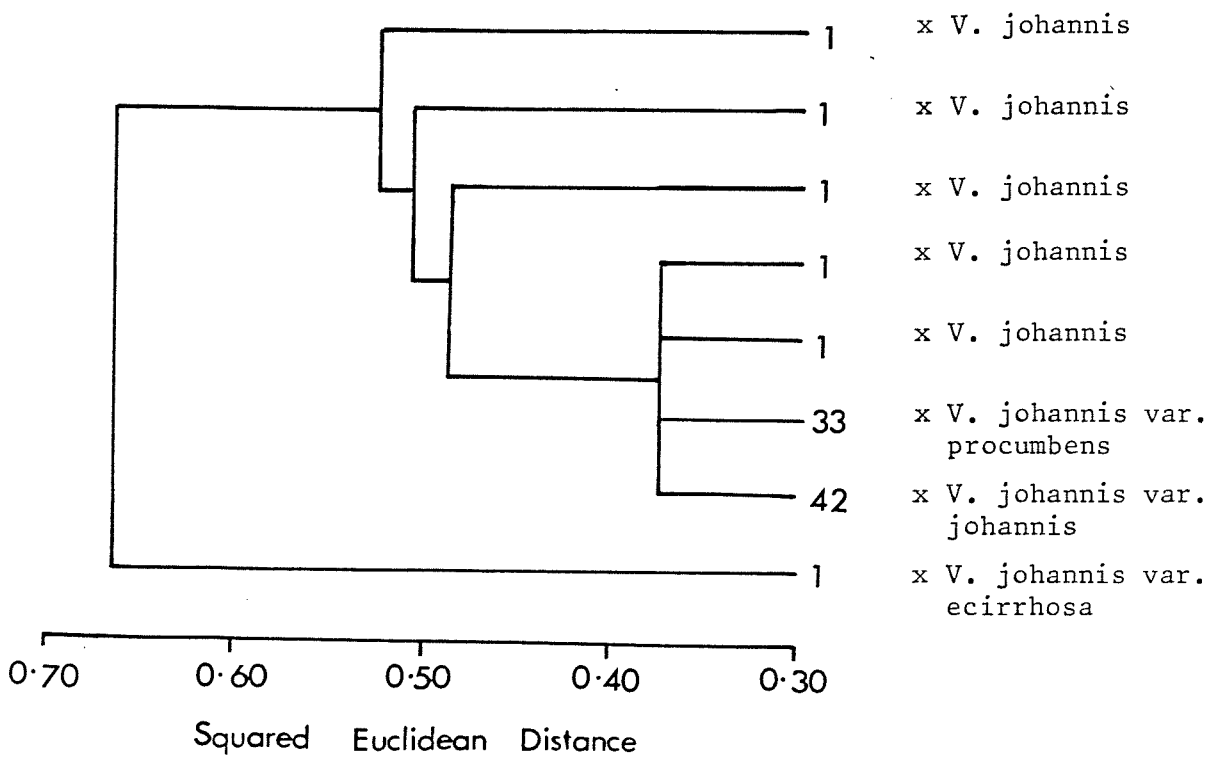


Figure 7.19. Centroid linkage cluster analysis of V. johannis specimens.



and colour of the wing apex spot. In the analysis, the 81 specimens of V. johannis were clustered using these 5 characters (character set M in Appendix 3). The results are shown in Figure 7.19. The dendrogram shows the specimens split into eight clusters. Three clusters can be attributed to the three varieties, var. ecirrhosa, var. johannis and var. procumbens. The other five clusters contain single specimens, all of which possess a tendril and so are either var. johannis or var. procumbens. The wing spot colour for these specimens could not be recorded and so no varietal identification could be made.

One of the groups identified by the initial analysis contains those species closely allied to V. sativa. It includes V. barbazitae, V. grandiflora, V. pyrenaica and V. gatmensis. Representative specimens were analysed using 69 characters (character set G in Appendix 3). The results are shown in Figure 7.20. Four of the groupings shown in the dendrogram are composed largely of one taxon each. The V. grandiflora cluster contains one V. barbazitae specimen (Haussknecht 20/06/1885) and the V. barbazitae cluster contains two small flowered V. grandiflora specimens (Montbret 10/5/1833, W and Krendl and Krendl 17/5/1980, W). The two remaining clusters each contain one specimen of V. grandiflora each and both these specimens (Krendl and Krendl 12/5/1980, W and Hepper 3165, K) were found to have a high proportion of missing data, possibly explaining their separation from the other specimens of their species.

The 236 specimens of V. sativa were analysed using 23 characters (character set H in Appendix 3). The dendrogram resulting from the cluster analysis is shown in Figure 7.21. The separation of infra-specific taxa within V. sativa depends on a small set of characteristics and if some of these characters cannot be scored for a particular specimen, then it is difficult to attribute specimens to a subspecies. This problem was also encountered with specimens of the V.

Figure 7.20. Centroid linkage cluster analysis of V. barbazitae, V. grandiflora, V. pyrenaica and V. qatmensis group.

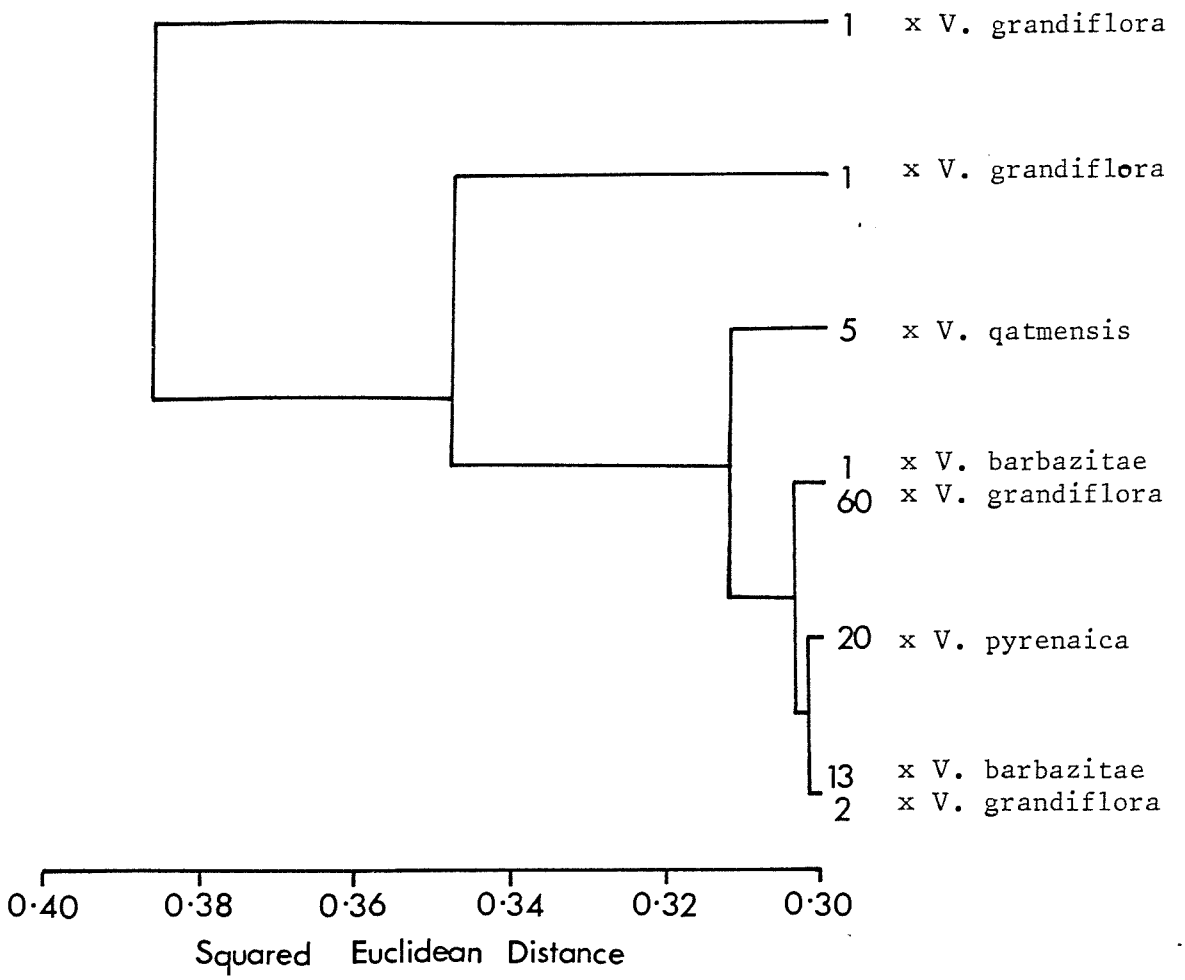
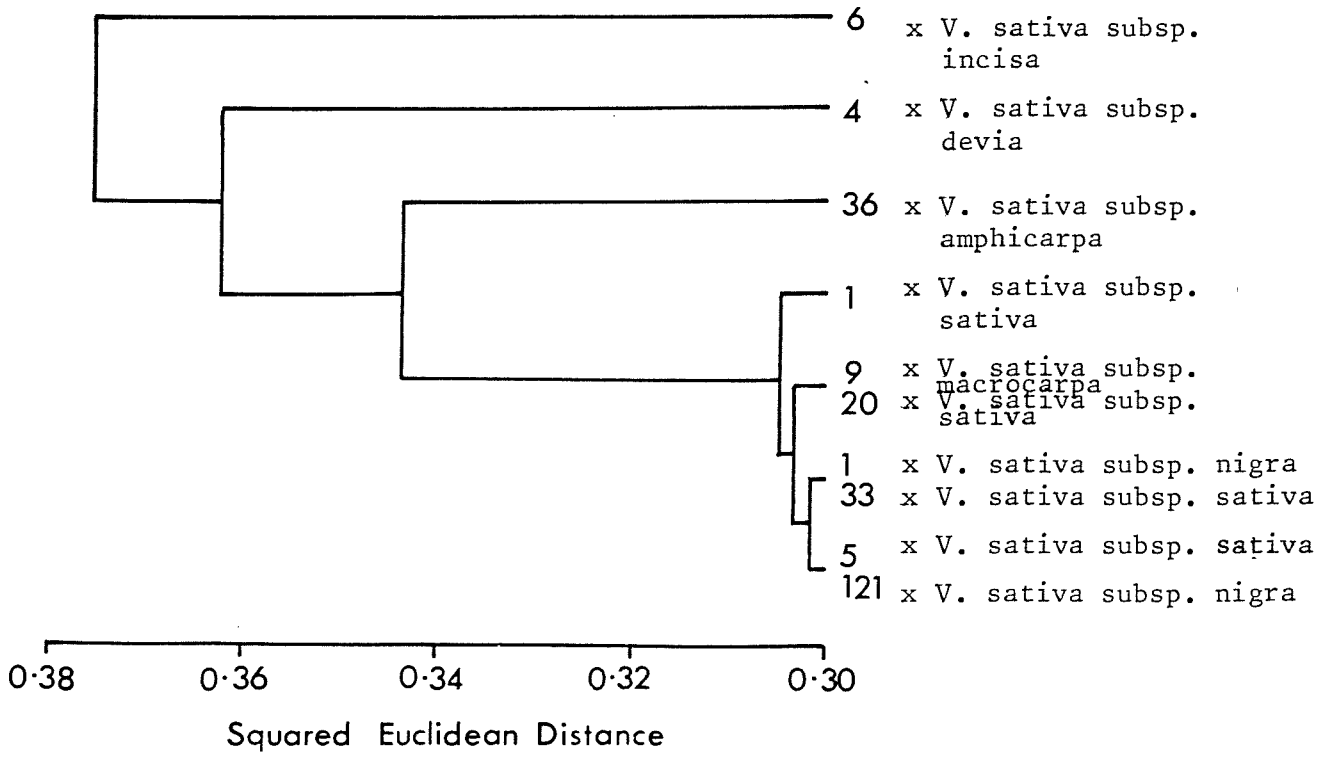


Figure 7.21. Centroid linkage cluster analysis of V. sativa specimens.



narbonensis complex. The interpretation of the results shown in Figure 7.21 is similarly difficult. Seven taxon clusters can be identified and three of these clusters contain more than one taxon. Three subspecies form clearly distinct clusters, subsp. amphicarpa, subsp. devia and subsp. incisa. It was, however, particularly difficult to distinguish between subsp. macrocarpa, subsp. nigra and subsp. sativa, unless legume characters were scored. Lack of legumes from some of the specimens included in this study has resulted in the formation of multi-taxon clusters for these three taxa. Two further taxa, subsp. nigra var. segetalis and subsp. cordata, were included a priori as accepted taxa. However, during scoring and analysis these two taxa were not distinguishable from the other six infra-specific V. sativa taxa.

The initial cluster and ordination analysis indicated that V. sepium was a distinct group. The three varieties accepted at the beginning of the project are distinguished using stipule, leaflet dimension, calyx pubescence and calyx colour characters (Chrtkova-Zertova, 1969; Hämet-Ahti, 1970). Five characters, derived from these differences (character set I in Appendix 3), were used in the analysis of this species. The results of the cluster analysis are shown in Figure 7.22. Seven major clusters can be seen. Three clusters are largely composed of specimens attributable to one of the three varieties each, although the var. sepium and var. montana clusters each contain a few specimens from var. eriocalyx. Observation of the material and the results of the cluster analysis indicate that the three varieties are closely related and there may be a cline between the extremes of the three taxa. Extreme specimens are easily determined, but, more rarely, specimens are found which are hard to attribute to one variety. Even though var. sepium and var. montana are not as distinctive as var. eriocalyx, the existence of three varieties is retained as the majority of the specimens were easily attributed to one of the three varieties.

Figure 7.22. Centroid linkage cluster analysis of V. sepium specimens.

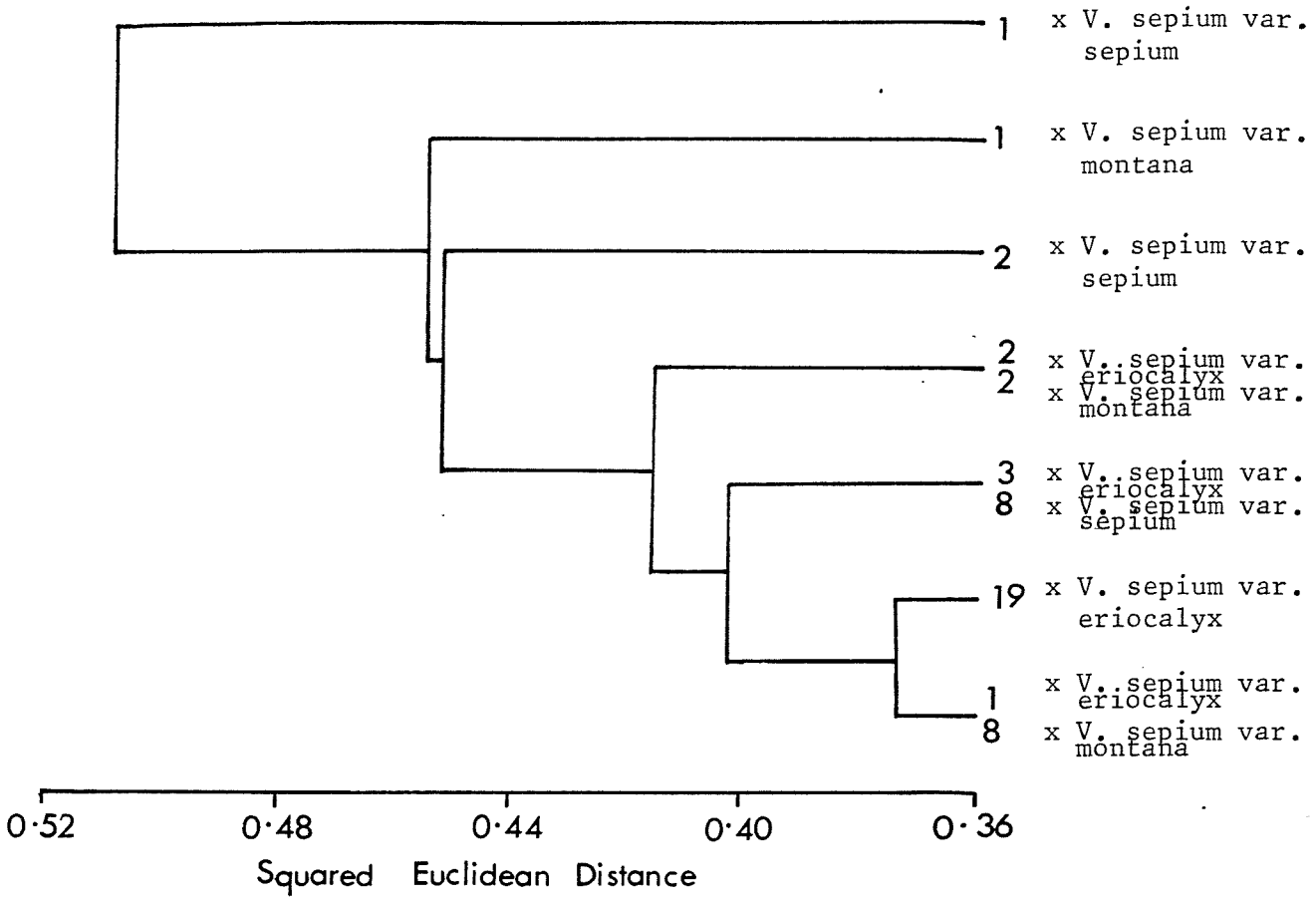
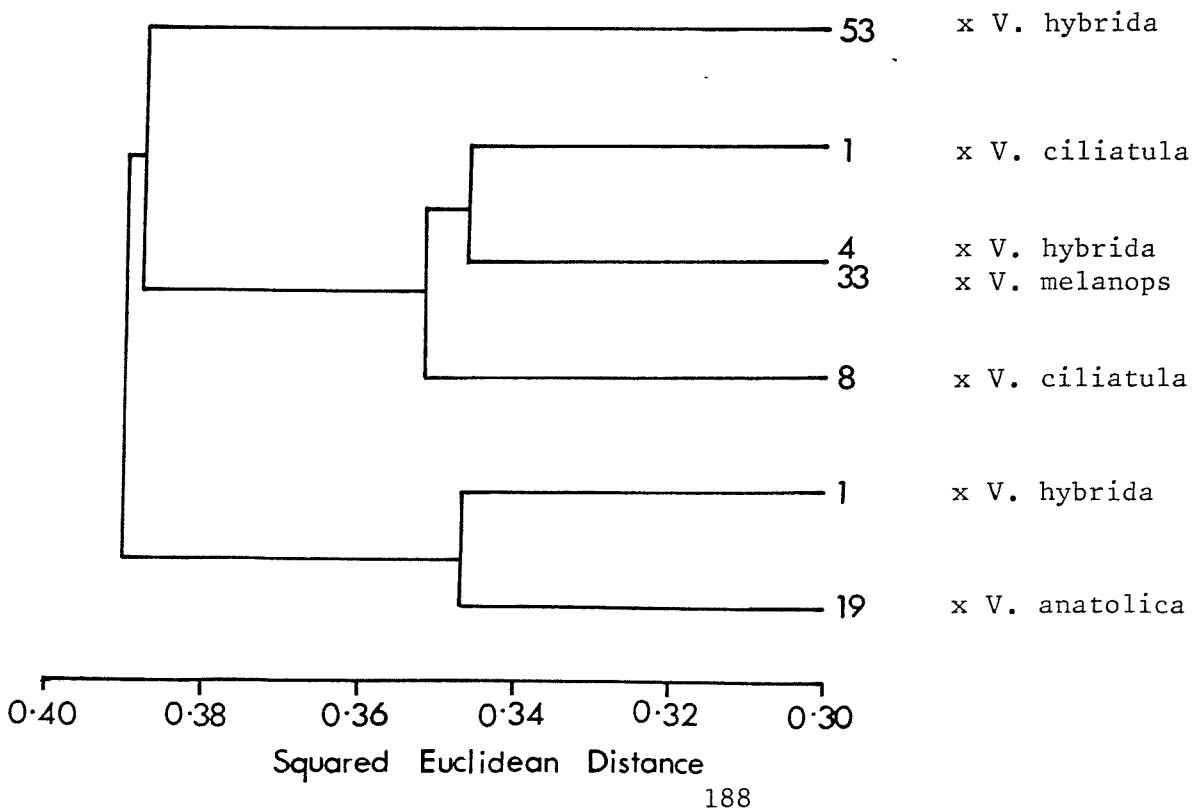


Figure 7.23. Centroid linkage cluster analysis of V. anatolica, V. ciliatula, V. hybrida and V. melanops group.



Another group of species, identified by the initial analysis, was that composed of V. anatolica, V. ciliatula, V. hybrida and V. melanops. Representative specimens of these four species were analysed using 66 characters (character set J in Appendix 3) and the results are shown in Figure 7.23. The bulk of the 119 specimens form species based clusters. However, four V. hybrida specimens join the V. melanops cluster. The reason for the separation of these four specimens from the main V. hybrida cluster cannot easily be explained. The flowers of two of these specimens were slightly immature and this might explain their association with the smaller flowered V. melanops. Another of these aberrant V. hybrida specimens has no flowers and so the analysis is based on vegetative and legume characters alone, which explains its misplacement.

D'Alleizette (1958) distinguishes two varieties of V. melanops on the basis of leaflet, flower dimension and flower colour and using these 13 characters (character set O in Appendix 3), the 33 V. melanops were analysed. The results of the analysis are shown in Figure 7.24. The results indicate a clear separation of the two varieties. Variety loiseaui is rarer and has a more restricted distribution than var. melanops, but the distinction between them is a good one.

The final group, distinguished by the initial analysis, to be further analysed is V. faba. The infra-specific classification of the fababean most commonly used is that suggested by Muratova (1931). She splits the species into two subspecies and three varieties on the basis of leaflet number and seed dimension characters. Representative specimens of these four taxa were analysed using 16 characters (character set K in Appendix 3) and the results are shown in Figure 7.25. Four taxa can be distinguished from the dendrogram. The most distinct cluster is that containing subspecies paucijuga. Within the second major cluster, which contains subspecies

Figure 7.24. Centroid linkage cluster analysis of V. melanops specimens.

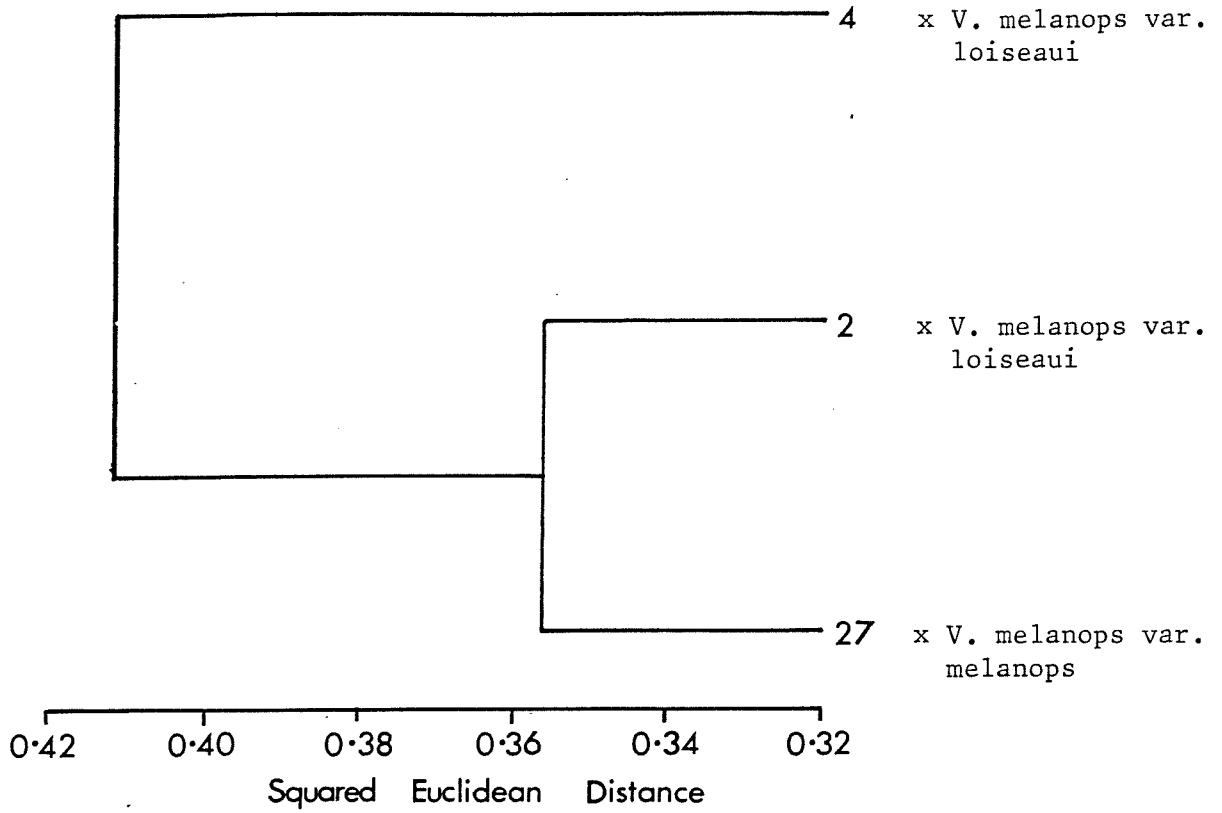
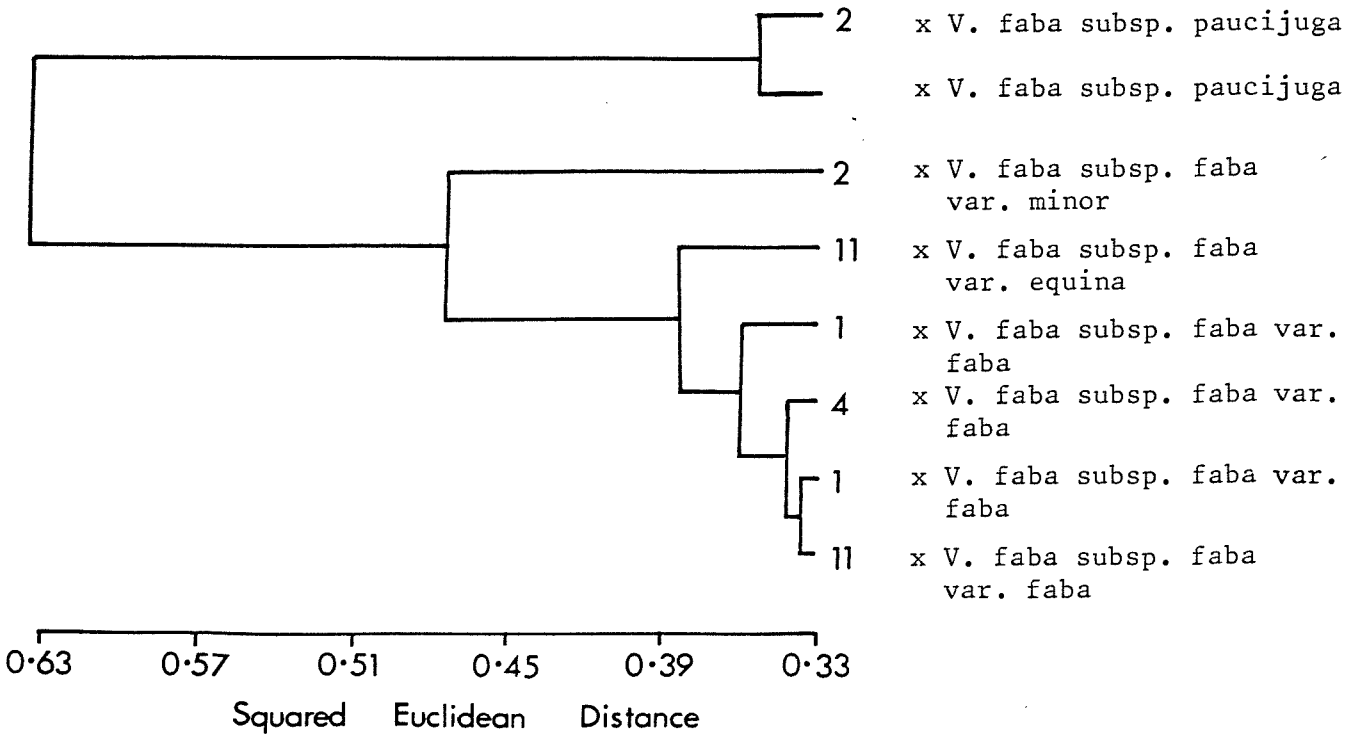


Figure 7.25. Centroid linkage cluster analysis of V. faba specimens.



faba specimens, var. minor is the most distinct taxon, with var. equina and var. faba also distinguishable.

7.3 Subgeneric Analysis of Taxon Data

Once the 74 taxa which make up subgenus Vicia were identified, the relationship between them could be investigated in more detail. As will be discussed in Chapter 10, the specimen data were used directly (via a series of dBASE programs) to produce DELTA descriptions for each taxon. Dallwitz & Paine (1986) include in their suite of programs, DIST. A program that can generate a distance matrix, using the Gower metric, from the DELTA items file. This distance matrix could then be used for multivariate analysis.

DIST was used to produce a distance matrix, which was then read by CLUSTAN and analysed, via the procedure HIERARCHY, using average linkage cluster analysis. Unfortunately, the results of this analysis were misleading. Taxa considered by previous authors to be closely related were shown not to be allied. The reason for this supposed misplacement was explained by the method used to calculate the data values, which formed the base for the calculation of the distance matrix. In the manual Dallwitz & Paine (1986) state that for

> "ordered multistate and numeric characters, a 'central' value X_{ik} is calculated for each masked in item i and character k. For ordered multistate characters, X_{ik} is the mean of all the coded values."

This method may be acceptable if two basic premises are met: firstly, that ordered multistate characters have minimal intra-taxon character variability and secondly, for quantitative characters, individual character scores are distributed normally throughout the character range. As the taxon data set was generated from a large number of individual specimens with high levels of intra-taxon variability, these two premises were not met. Which explains why the results did not reflect the natural relationships of the taxa.

Thus the subgeneric taxon data set for the 74 taxa was generated by hand from the specimen data sets for each taxon. To produce the taxon scores for each taxon the mode was calculated for each continuous character and the most common character state was used for the multistate characters. The mode was calculated by dividing the range into ten equal bands, scoring the number of records that fell in each band and then using the mean figure for the most common band. This does imply a certain characteristic for the data, i.e. for the multistate characters, that only one score is common, that the scores are not evenly distributed between two or more states. This is not true for all the characters, but is true for the majority and so was considered a satisfactory assumption.

Three characters were added to the original character set at this stage: plant life form and plant height (both of which would have been difficult to score from herbarium specimens) and calyx hair elevation. The latter was overlooked in the original analysis, but proved very useful in distinguishing V. sepium var. eriocalyx from the other V. sepium varieties. For the taxon analysis only the lowest level taxa were included in the analysis, in an attempt to avoid problems of scoring variable character scores (see discussion in Chapters Ten and Eleven). For example the two varieties of V. galilaea were included in the analysis, but not V. galilaea itself. This produced a complete taxon data set of 61 taxa x 174 characters. Subsets of this taxon based data set were analysed by the single linkage cluster analysis program LINKAGE, and by various other cluster analysis methods and principal components analysis using CLUSTAN. Multiple methods of analysis were used so that a general view of taxa relatedness could be established and any bias introduced by the use of one particular analysis method could be avoided.

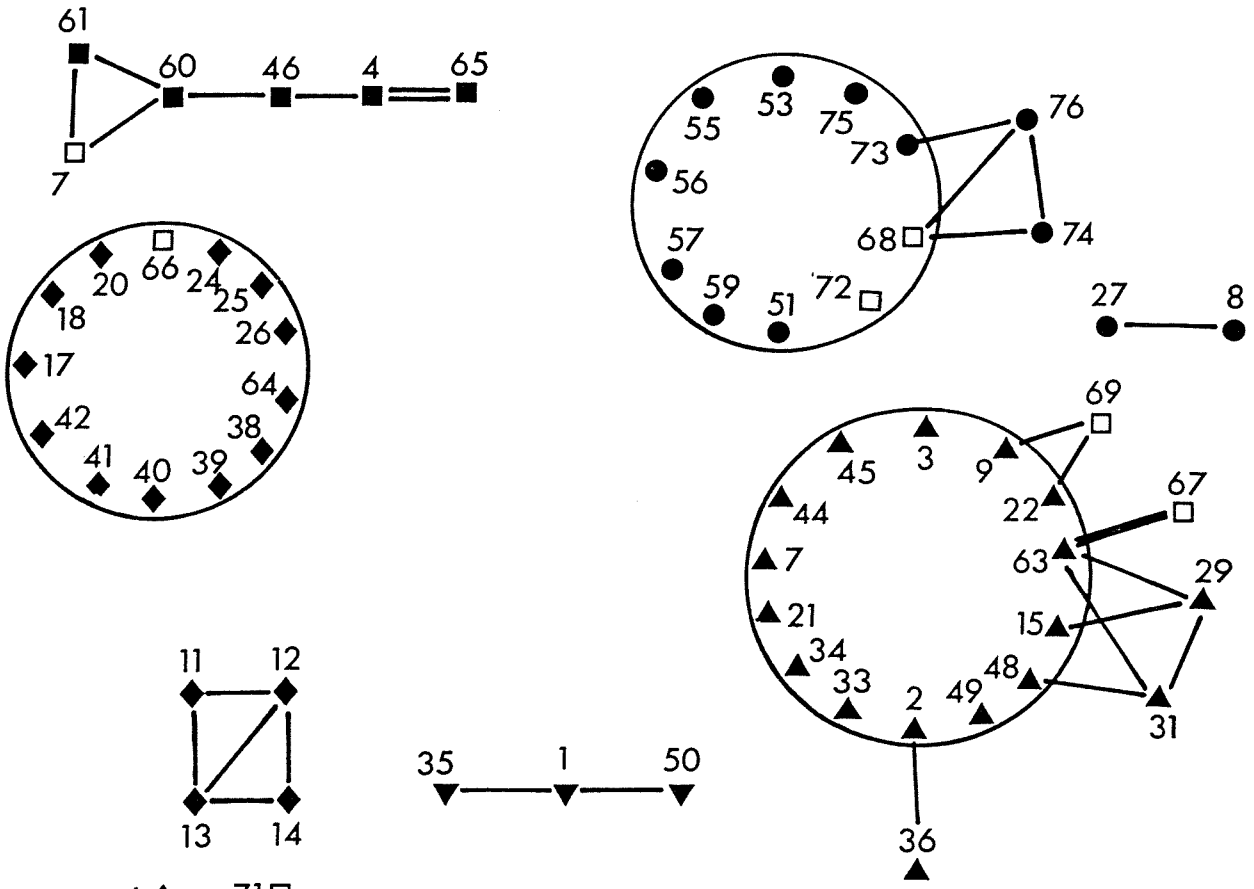
LINKAGE is a FORTRAN program written by Wirth, Estabrook & Rogers (1966), which undertakes single linkage (nearest

neighbour) cluster analysis. The program uses the simple matching coefficient to calculate the similarity matrix, using only those characters for which data are present for both OTUs. For the analysis 122 characters were selected a posteriori from the complete data set. The characters selected were those from the SPSS^{x} procedure DISCRIMINANT analysis with high F values and those which had proved useful during the specimen based analysis (character set 122 in Appendix 3). The results of the single linkage cluster analysis using the program LINKAGE may be displayed in the form of linkage diagrams or sub-graphs: 60 diagrams for the data set analysed. The diagrams are arranged in decreasing similarity, from a level of 0.9144, for the first major inter-taxon link, to 0.5139, when all the OTUs are joined in one cluster. Those diagrams which are most use in clarifying subgeneric relations are shown in Figures 7.26-7.33.

The interpretation of the linkage diagrams requires some discussion. At a given threshold level of similarity some pairs of OTUs will cluster and this is demonstrated in the diagram by a line connecting the OTUs. This connecting line may be of three kinds, indicating three possible kinds of relationships between OTUs; a single line indicating a relationship already established at a higher level of similarity, a double line indicating a new relationship established at that particular similarity level and a broken line which indicates a new internal (within cluster) link at that similarity level. To simplify interpretation of the diagrams highly intra-connected clusters are encircled. The criterion for inclusion in a circle is that each OTU should have at least three links with other members of the same encircled cluster.

The linkage diagram, drawn in Figure 7.26, shows seven clusters of taxa and two independent taxa. At this level of similarity, one cluster is referable to sect. Atossa, two clusters to Vicia, one to Hypechusa, two to Faba and one to

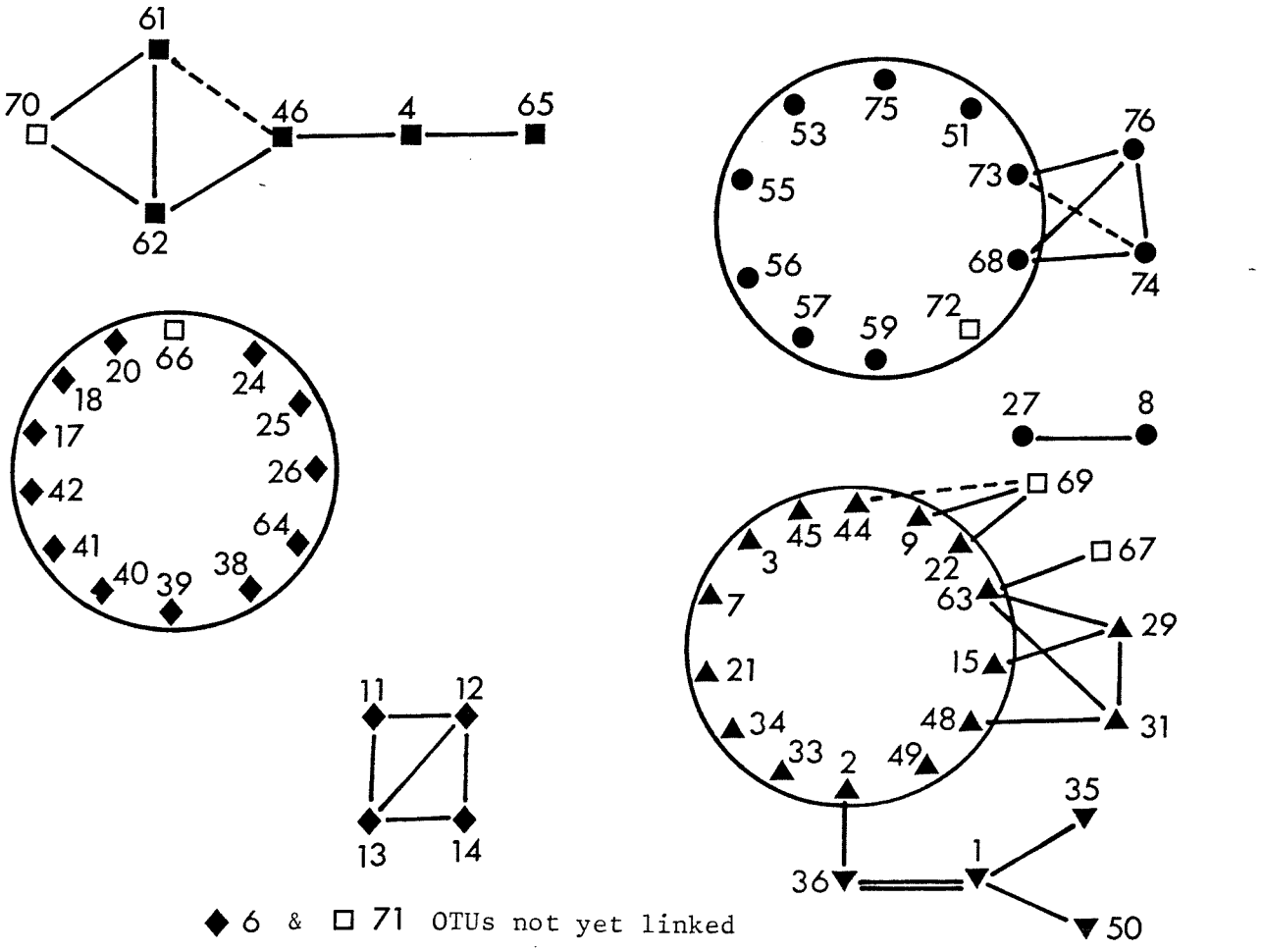
Figure 7.26. Single linkage cluster analysis of taxon data using 122 characters. Level 52.



6 ◆ & 71 □ OTUs not yet linked

Similarity - 0.6171

Figure 7.27. Single linkage cluster analysis of taxon data using 122 characters. Level 53. Similarity - 0.6156.



◆ 6 & □ 71 OTUs not yet linked

Peregrinae sensu Kupicha (1976). Each cluster contains members of one Kupicha section, except for the Hypechusa cluster, which contains one sect. Peregrinae taxon, V. mollis (36). The two sections which are split into two clusters each are: Faba, which has V. faba (11, 12, 13 and 14) separated from the V. narbonensis complex taxa and Vicia, which has the two small flowered forms, V. lathyroides (27) and V. cuspidata (8) separated from the main Vicia cluster. This level also shows V. dionysiensis (67), which in the previous analysis was shown to be isolated, forming links with V. sericocarpa (63) of the sect. Hypechusa cluster.

The next linkage diagram drawn in Figure 7.27, shows the linking of the two sectional clusters containing Hypechusa and Peregrinae taxa. The link between these two clusters was made between V. mollis (36) and V. aintabensis (1). This link might have been expected as Kupicha places V. mollis in sect. Peregrinae. Two taxa V. bithynica (6) and V. eristalioides (71) still remain distinct at this level. Figure 7.28 shows the next inter-sectional clustering, the linking of the perennial Atossa with the Vicia cluster. The link between the clusters is made between V. sepium var. sepium (61) and V. pyrenaica (51), the latter is the one perennial member of sect. Vicia. Also between the previous level shown and this level, V. eristalioides (71) has joined the V. narbonensis complex cluster.

Figure 7.29 shows the sect. Atossa and Vicia cluster linking next with the cluster containing the Hypechusa and Peregrinae taxa. The actual link between these clusters is made by V. qatmensis (68) of sect. Vicia and V. sericocarpa (63) of sect. Hypechusa. The next linkage diagram, drawn in Figure 7.30, shows V. lathyroides (27) and V. cuspidata (8) joining the main cluster, via V. sativa subsp. nigra (57) of sect. Vicia. It is somewhat surprising that these two taxa had not joined the sect. Vicia cluster at a higher similarity level as they are considered members of that section by

Figure 7.28. Single linkage cluster analysis of taxon data using 122 characters. Level 55. Similarity - 0.6009.

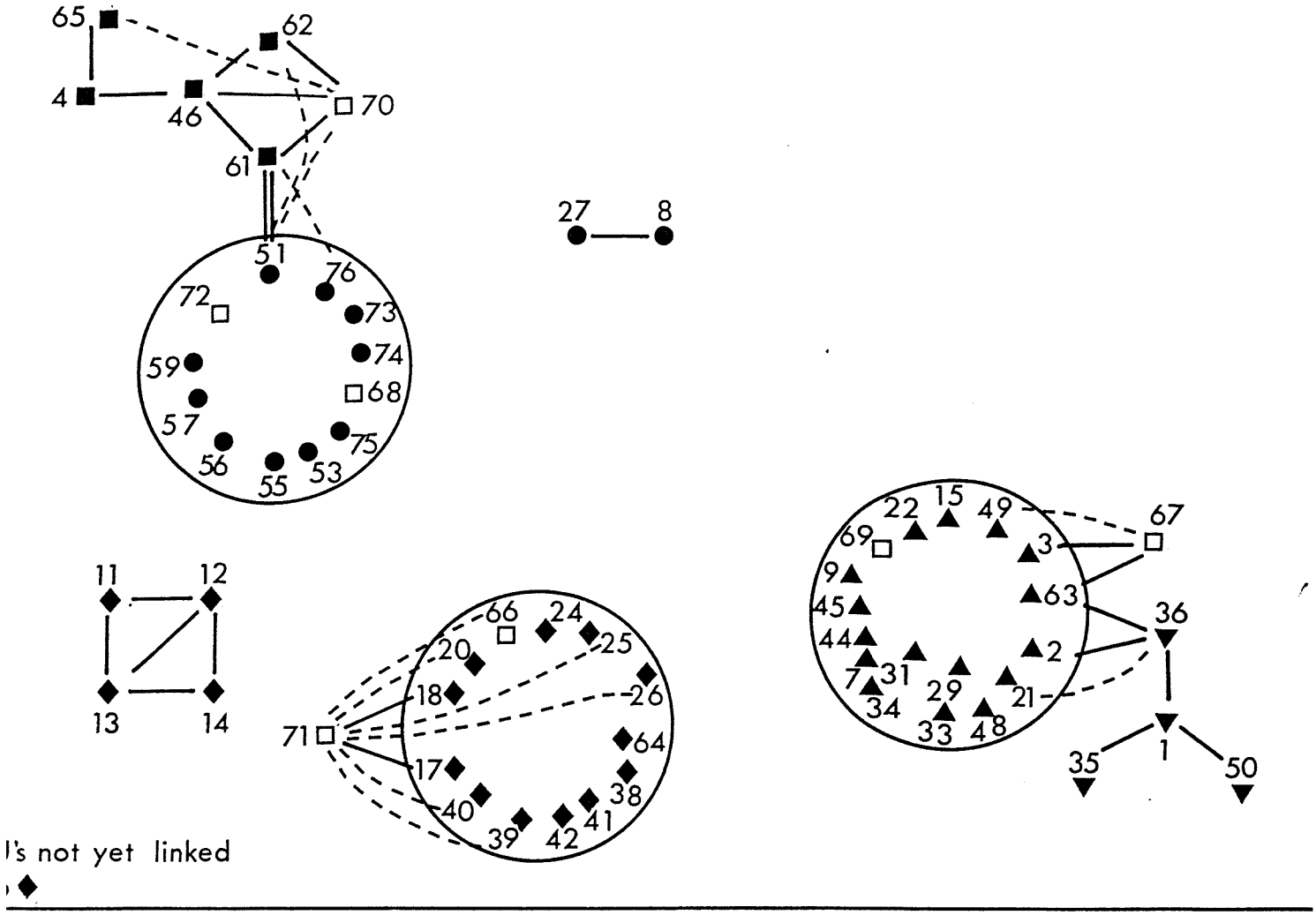


Figure 7.29. Single linkage cluster analysis of taxon data using 122 characters. Level 56. Similarity - 0.5827.

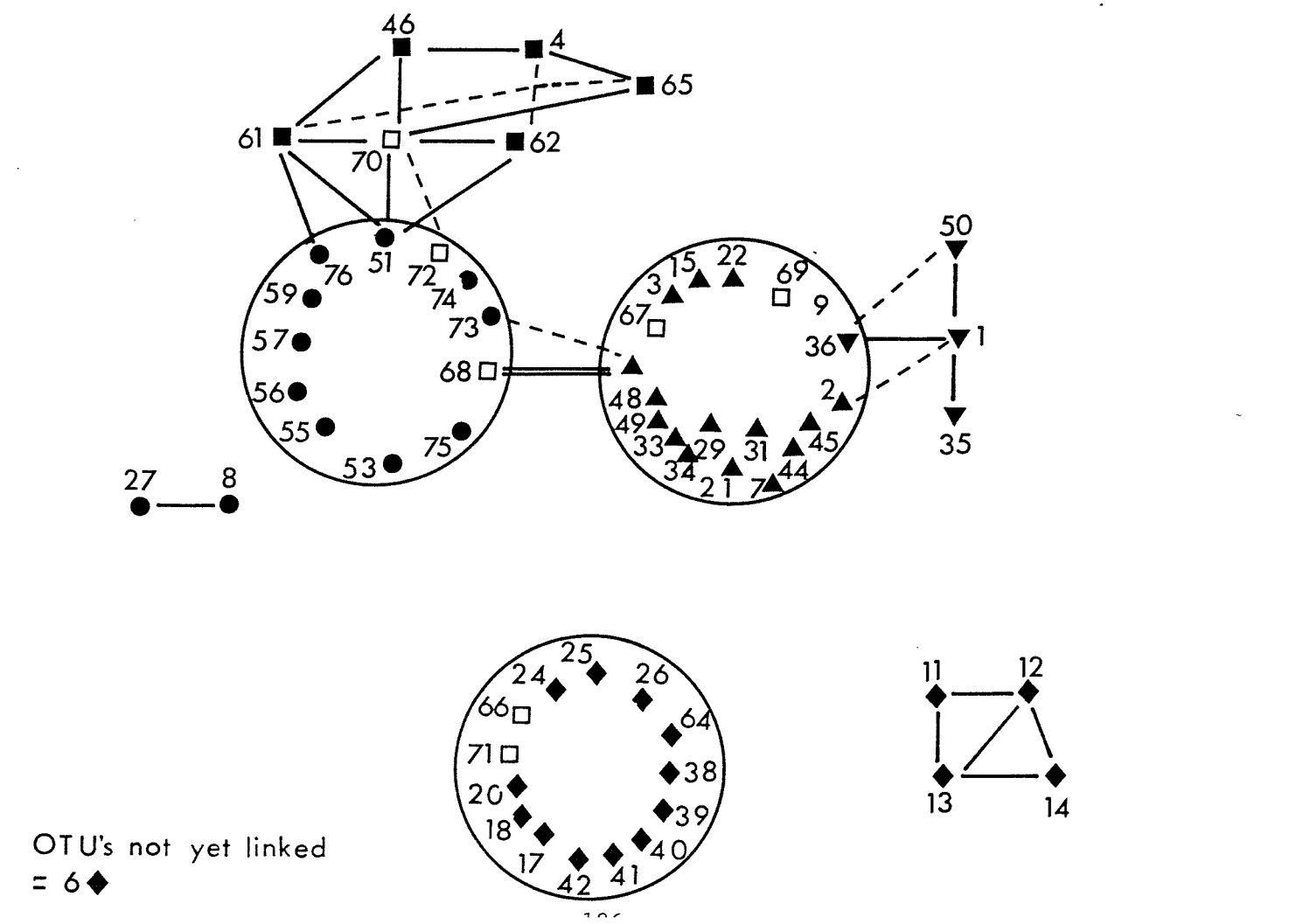
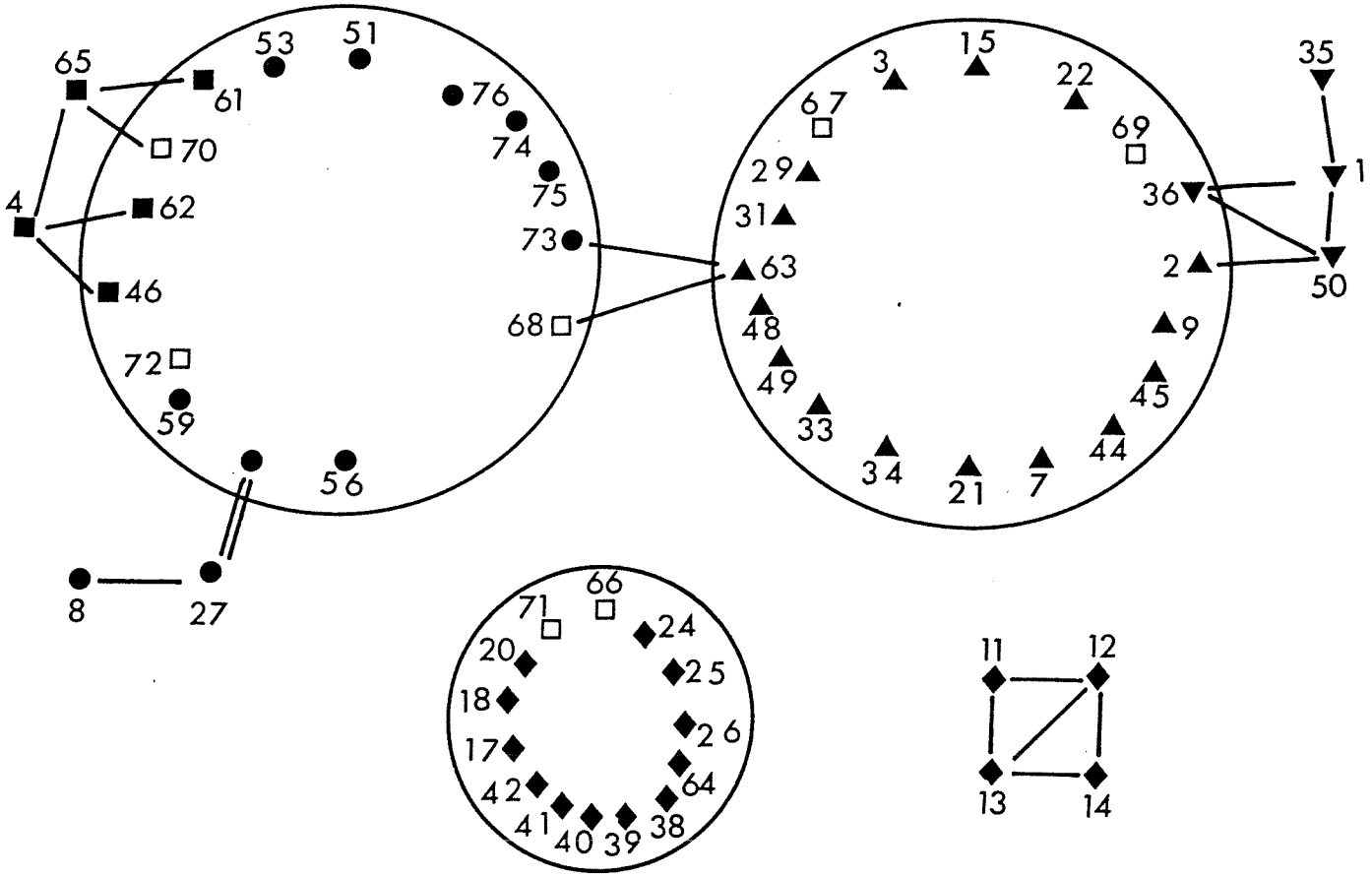
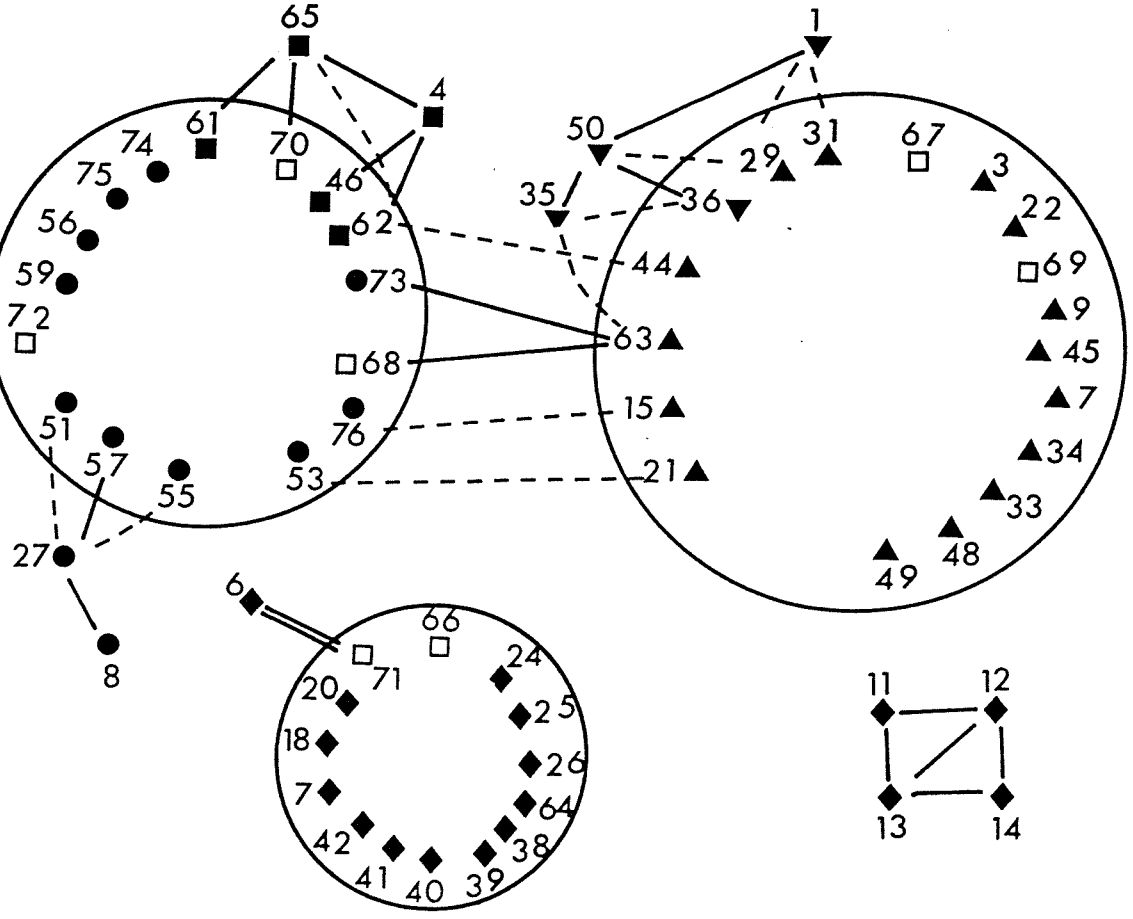


Figure 7.30. Single linkage cluster analysis of taxon data using 122 characters. Level 57. Similarity - 0.5765



OTU's not yet linked
= 6◆

Figure 7.31. Single linkage cluster analysis of taxon data using 122 characters. Level 58. Similarity - 0.5729.



Kupicha (1976). This indicates these species are more isolated from the other sect. Vicia than was previously thought.

At a similarity level of 0.5729, the final remaining independent taxon, V. bithynica (6), joins the V. narbonensis complex cluster. The link is made via V. eristalioides (71), as is shown in Figure 7.31. Within the main cluster of taxa, the two subclusters containing Vicia and Atossa, and Hypechusa and Peregrinae are beginning to form multiple links. The two other clusters contain V. faba, and the V. narbonensis complex with V. bithynica. Figure 7.32 shows the V. narbonensis complex cluster joining the main cluster via the linking of V. narbonensis var. affinis (39) with V. sativa subsp. sativa (59). The final diagram drawn in Figure 7.33, shows the linking of V. faba with the main cluster, V. faba subsp. faba var. equina (12) links to V. narbonensis var. narbonensis (41).

The same data set of 61 taxa by 122 characters was used for the centroid linkage cluster analysis, using CLUSTAN (version 3.1). The results of the analysis are shown in the dendrogram drawn in Figure 7.34. Similar groupings are found as in the single linkage cluster analysis. The following sections were identified: Atossa; Vicia, divided into two subclusters with V. pyrenaica, V. barbazitae, V. sativa, V. grandiflora and V. qatmensis in the first and V. lathyroides and V. cuspidata in the second; Hypechusa; Peregrinae, excluding V. mollis which is included in sect. Hypechusa; and Faba, which is divided into three subclusters with V. faba in the first, V. bithynica in the second and the V. narbonensis complex taxa in the third. The dendrogram indicates that the clusters containing Vicia sensu stricto and Hypechusa are the most closely related sections, followed by sect. Atossa, the V. narbonensis complex cluster, V. dionysiensis and sect. Peregrinae, then V. bithynica, V. faba and finally V. lathyroides and V. cuspidata.

Figure 7.32. Single linkage cluster analysis of taxon data using 122 characters. Level 59. Similarity - 0.5496.

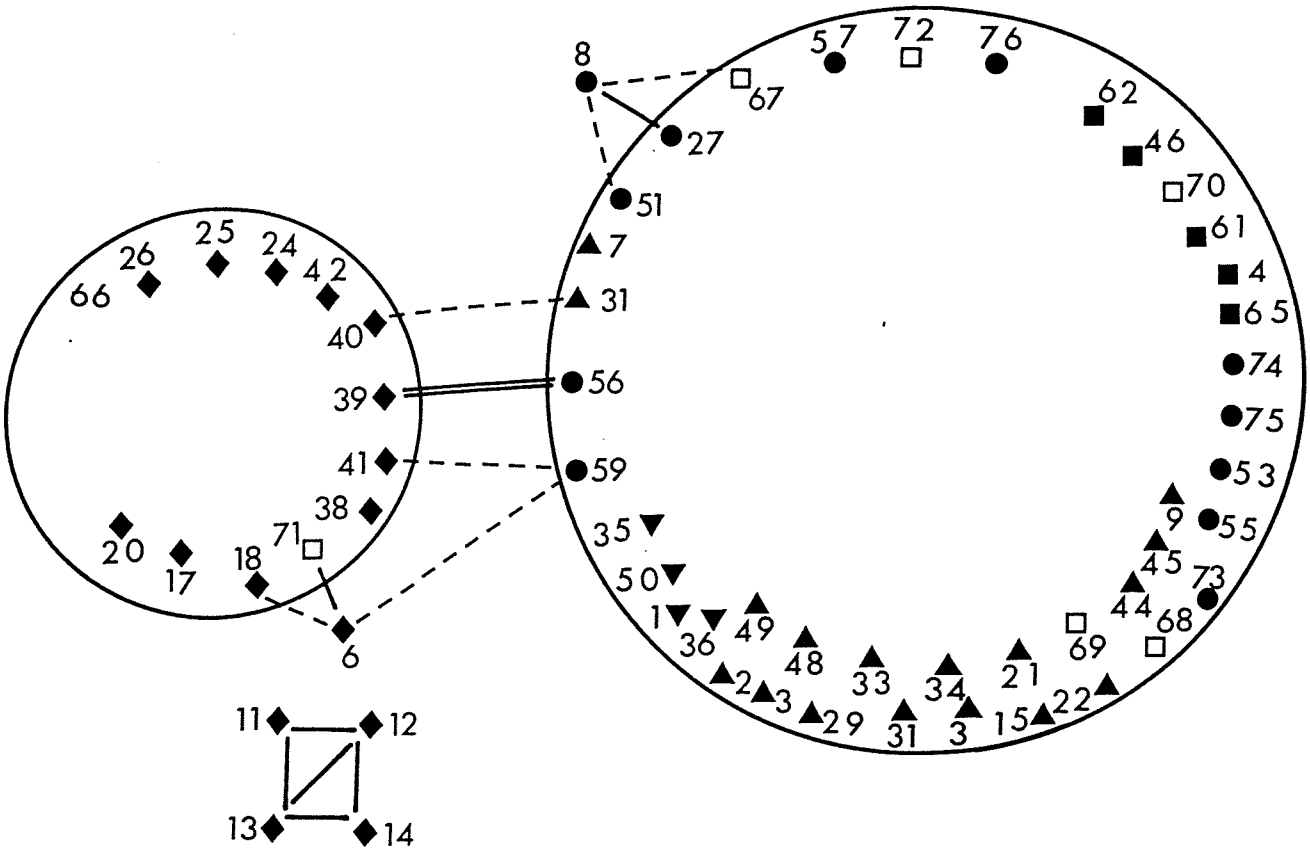


Figure 7.33. Single linkage cluster analysis of taxon data using 122 characters. Level 60. Similarity - 0.5139.

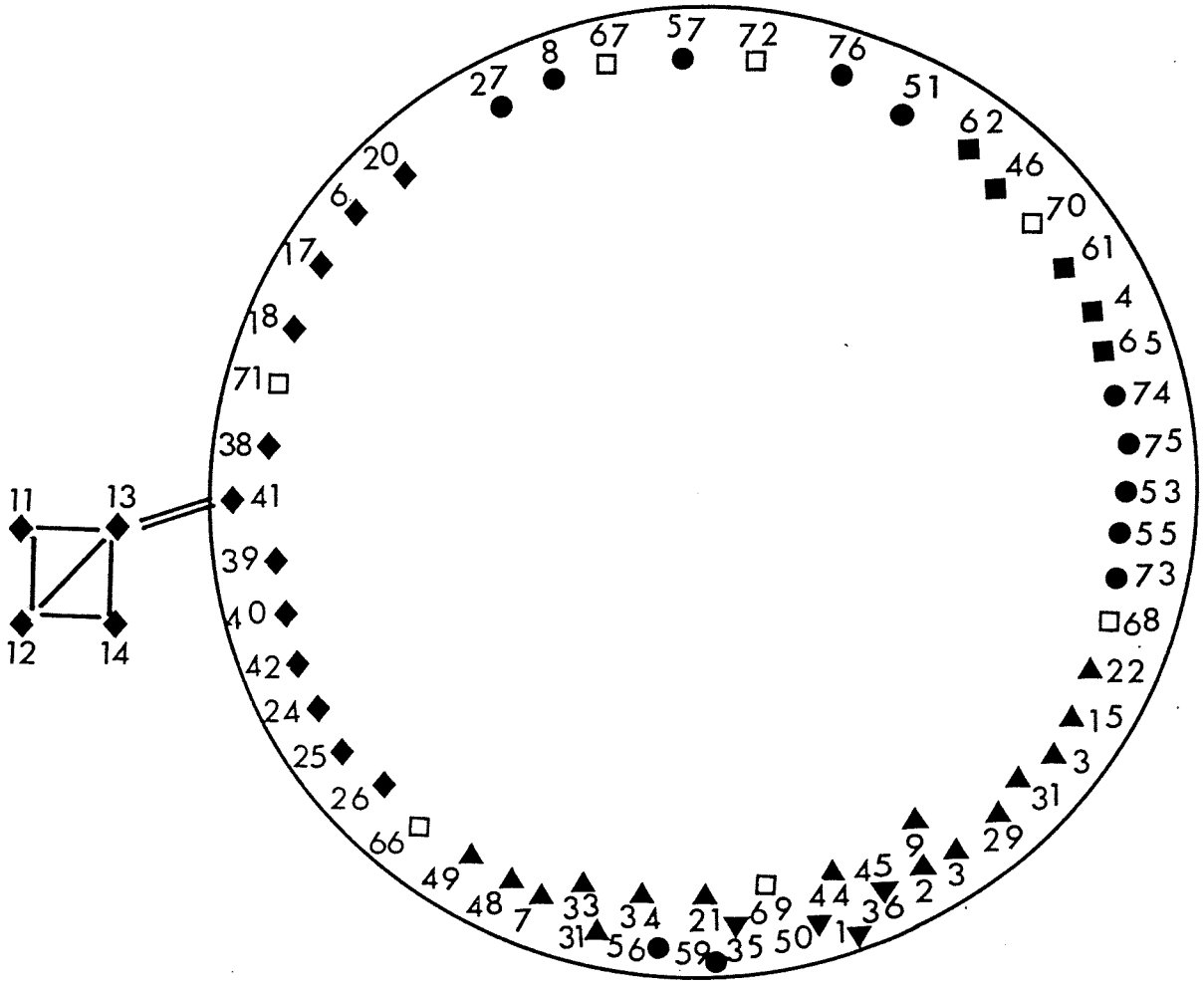
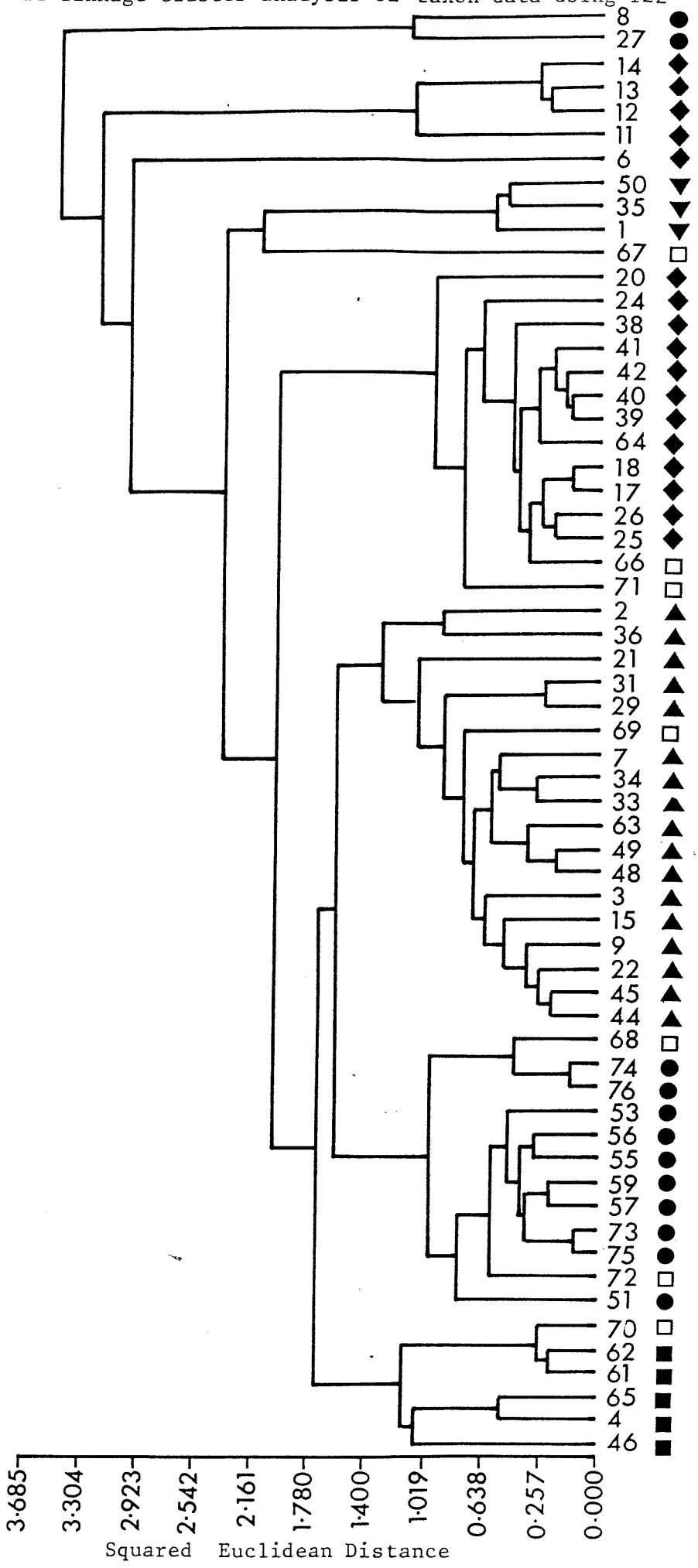


Figure 7.34. Centroid linkage cluster analysis of taxon data using 122 characters



The result of the median cluster analysis using the same data set is shown in Figure 7.35. The dendrogram indicates similar clusters to those seen in the previous analysis. There are, however, variations. The Peregrinae cluster is included within the Hypechusa cluster and V. barbazitae is included with V. grandiflora and V. qatmensis, rather than with V. pyrenaica and V. sativa as in the previous analyses. Having emphasised these differences, the analysis retains the sect. Atossa taxa in a distinct cluster, separates V. lathyroides and V. cuspidata from the other members of sect. Vicia and splits the sect. Faba taxa into three subclusters.

To undertake Ward's Method of cluster analysis and the principal components analysis, an older version of CLUSTAN (2.1) was used. These sub-routines had not successfully been implemented on the new version of CLUSTAN at Southampton University. Version 2.1 of CLUSTAN allows a maximum character number of 80 for principal components analysis. So the character set was reduced to 80, those that had a relatively high F value (obtained from the SPSS^{x} procedure DISCRIMINANT) or that had proved useful in the previous analyses were selected (character set 80 in Appendix 3).

The results of the Ward's Method of cluster analysis are shown in Figure 7.36. The groupings correspond to those formed by the methods of analysis used above, even though the data set had been reduced by one third from 122 to 80 characters. Ward's method tends to find minimum variance spherical clusters (Wishart, 1975) and this is shown by the dendrogram. Eight clusters can be seen: sect. Atossa; sect. Vicia; V. lathyroides and V. cuspidata; sect. Peregrinae; sect. Hypechusa; V. bithynica; the V. narbonensis complex; and V. faba. The one distinct group identified in the previous analyses, but not distinguished as a separate entity by Ward's method, is V. dionysiensis. In Figure 7.36, V. dionysiensis is linked with V. mollis and V. assyriaca. Both of these

Figure 7.35. Median linkage cluster analysis of taxon data using 122 characters.

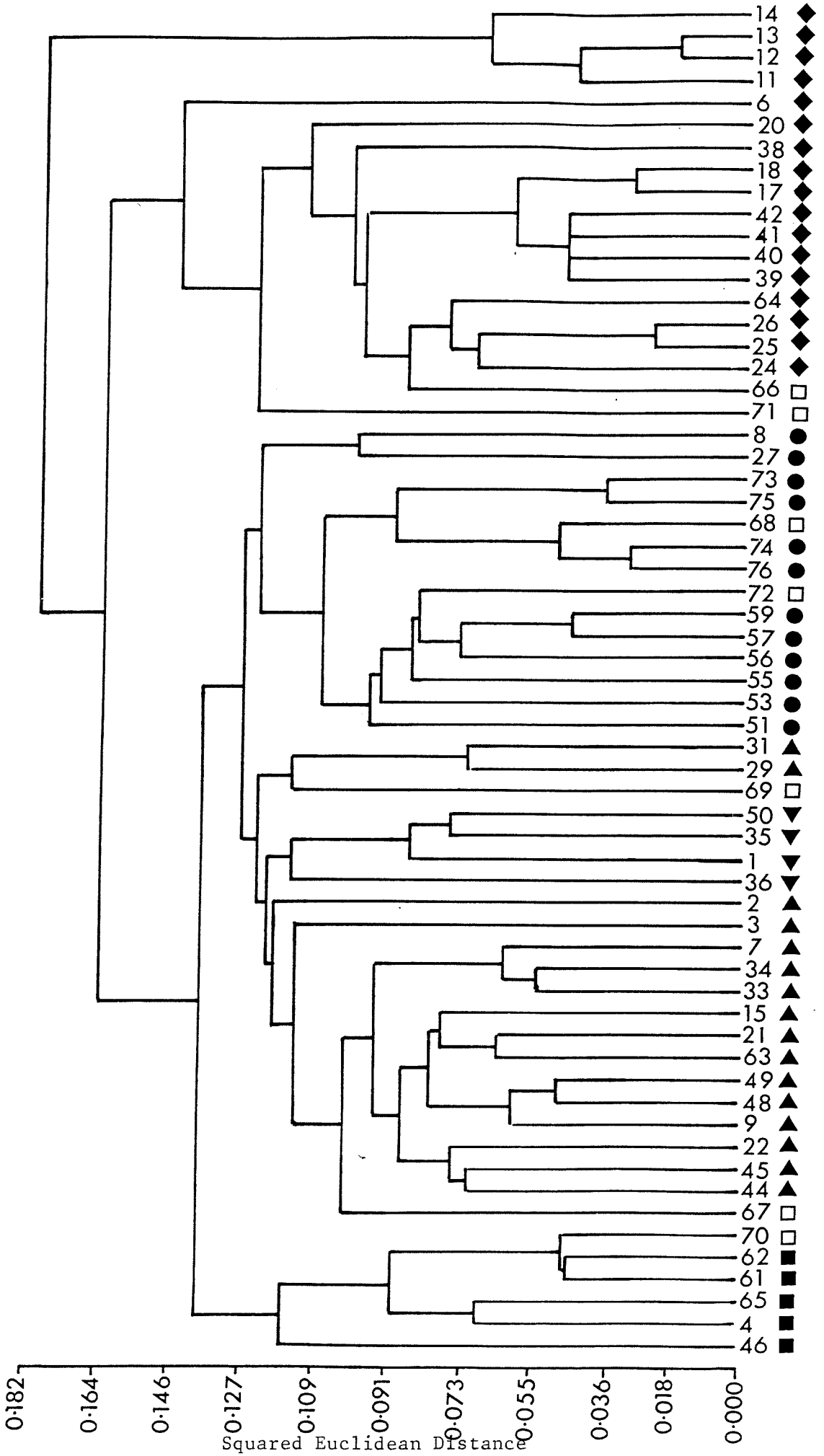
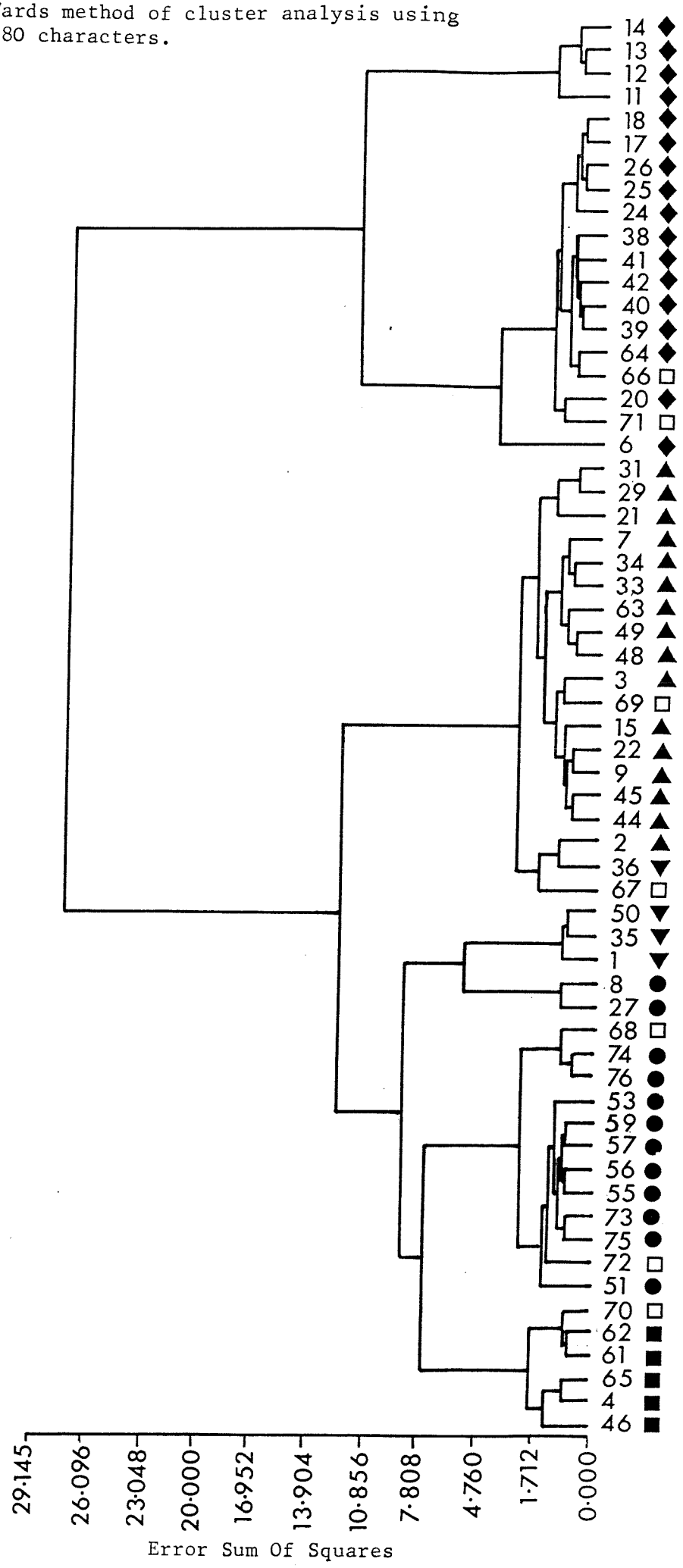


Figure 7.36. Wards method of cluster analysis using taxon data and 80 characters.

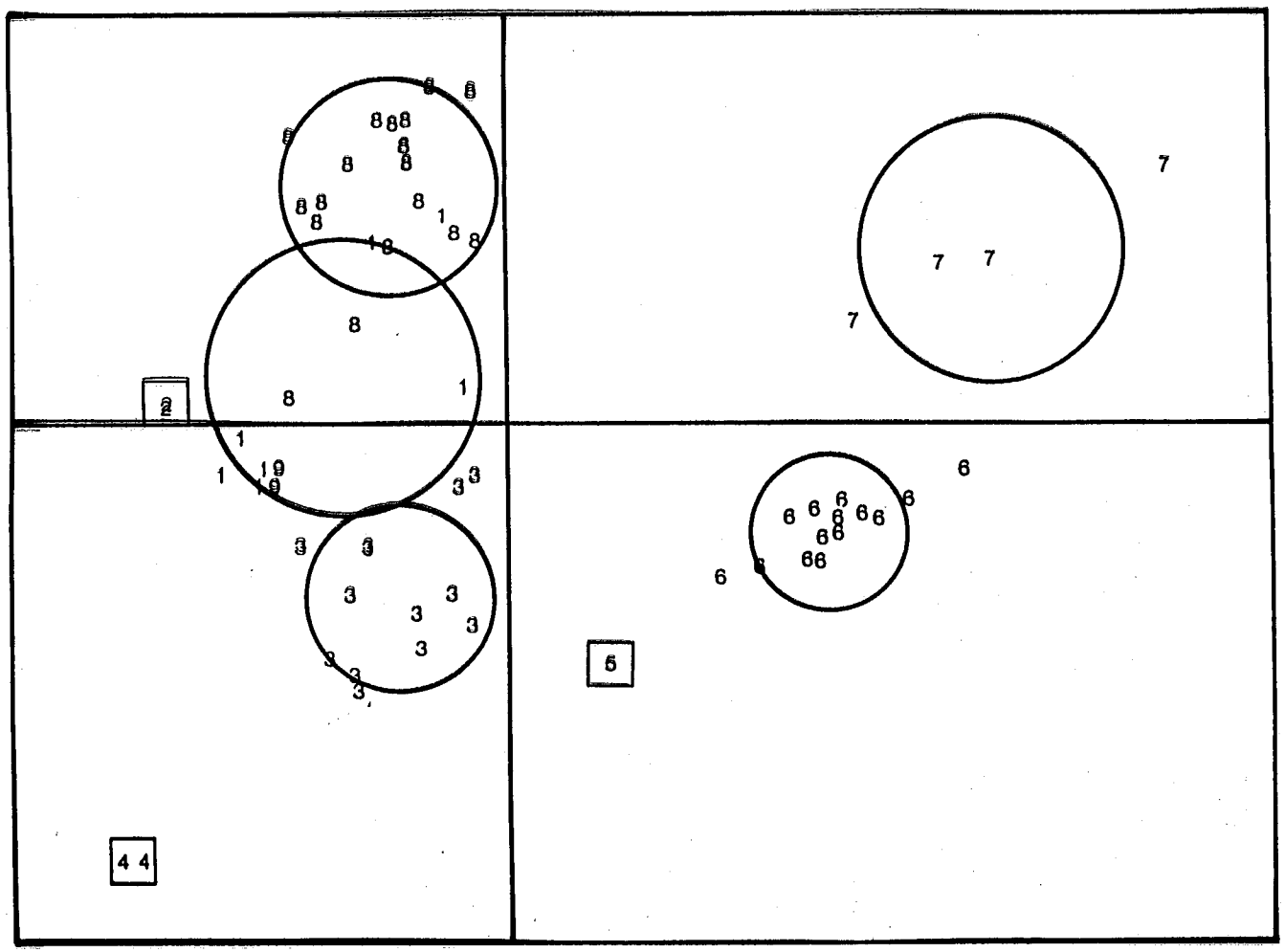


species were shown to be distant from V. dionysiensis by the other methods of analysis.

The results of the principal components analysis are shown in Figure 7.37, where the first two principal components are plotted. The first two principal components contain 26.03 and 14.97 percent of the variance respectively, which gives a cumulative variance of 41.00 percent. The clusters with minor alterations are similar to those identified above, although here the sect. Atossa taxa do not form a distinct cluster, but spread between sect. Hypechusa and sect. Peregrinae. If the minimum spanning tree is superimposed over Figure 7.37, then it can be seen that the sect. Atossa taxa do in fact link with one another, rather than linking with their closest spatial neighbours from other sections. The minimum spanning tree tends to link up taxa of clusters, established in the above methods of analysis, whose relationship has been distorted by the plotting of only the first two principal components.

The same grouping of taxa is also indicated by the second overlay of Figure 7.37, which shows the OTU cluster numbers and circles. If the nine cluster groupings obtained from Ward's method are used, then nine cluster numbers and circles can be superimposed on the PCA scatter diagram. The taxa are partitioned in accordance with the results of the above analyses. The largest amount of intra-cluster variance, indicated by the size of the cluster circle, is shown by sect. Atossa and V. faba, with the monospecific clusters and the cluster composed of V. lathyroides and V. cuspidata showing least intra-cluster variance.

The major groupings (sections) identified by the various methods of analysis are summarised in Table 7.5.


2
5
4 4

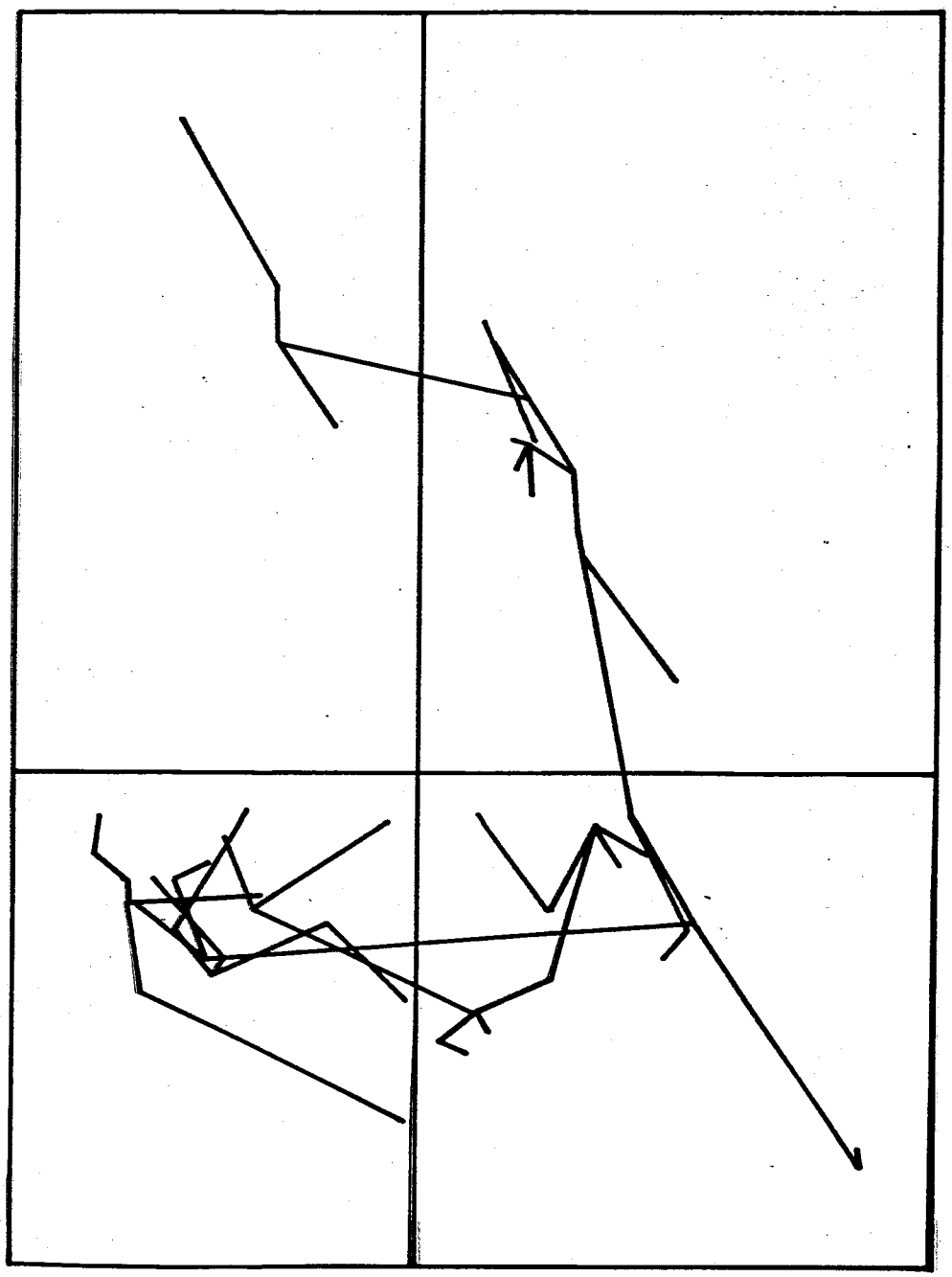


Figure 7.37. Principal components analysis of taxon data using 80 characters.

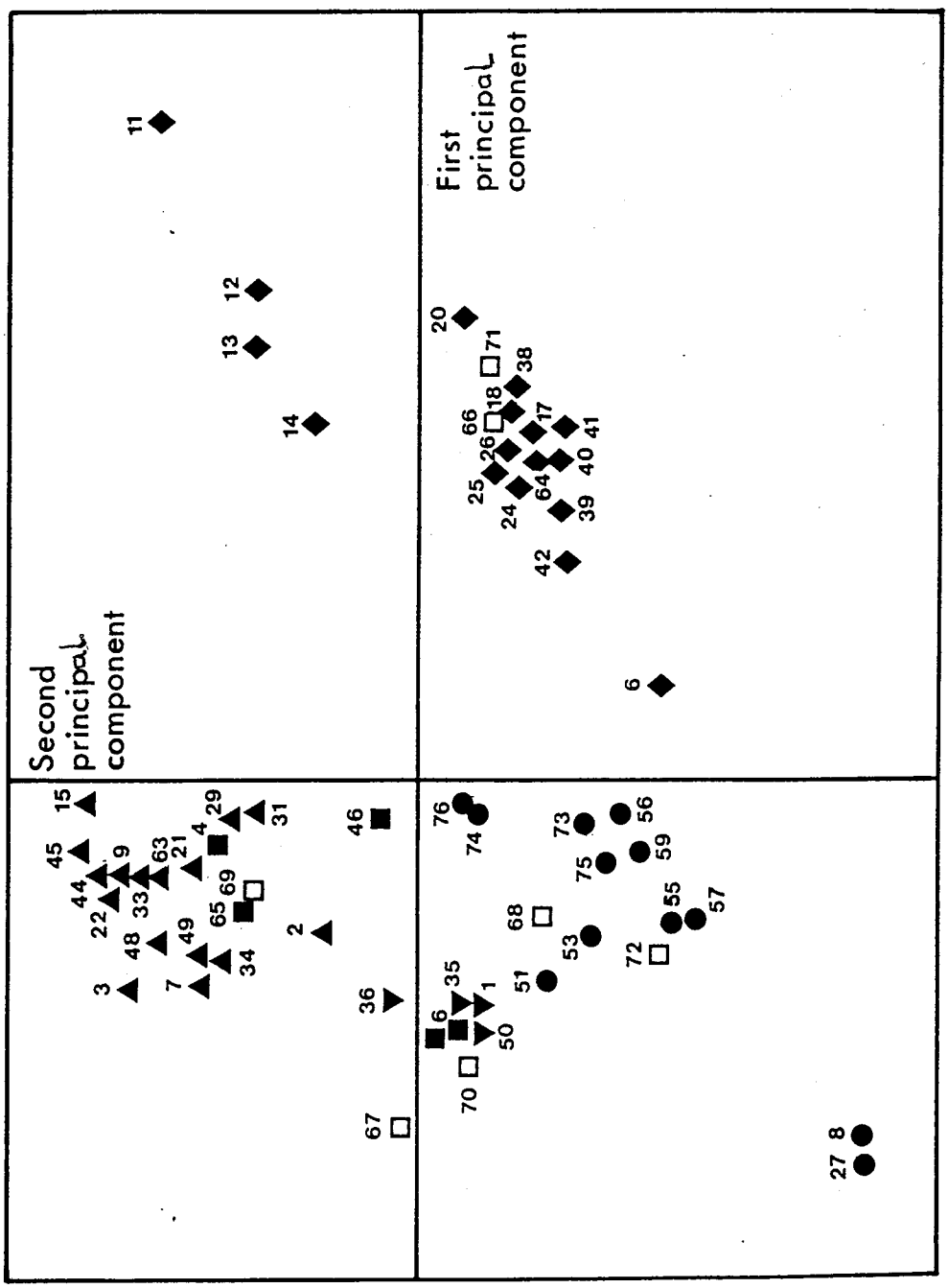


Figure 7.37. Principal components analysis of taxon data using 80 characters.

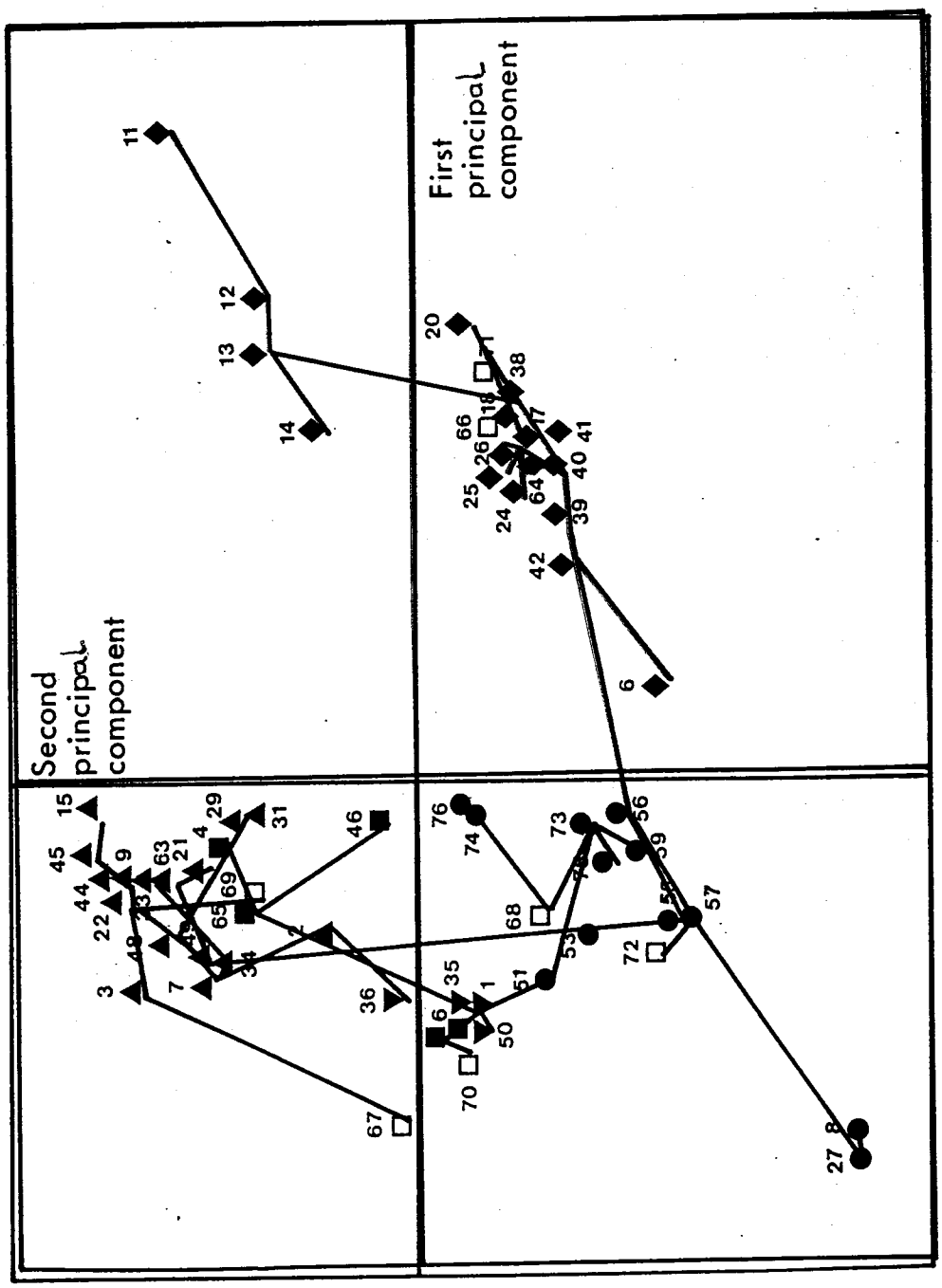


Figure 7.37. Principal components analysis of taxon data using 80 characters.

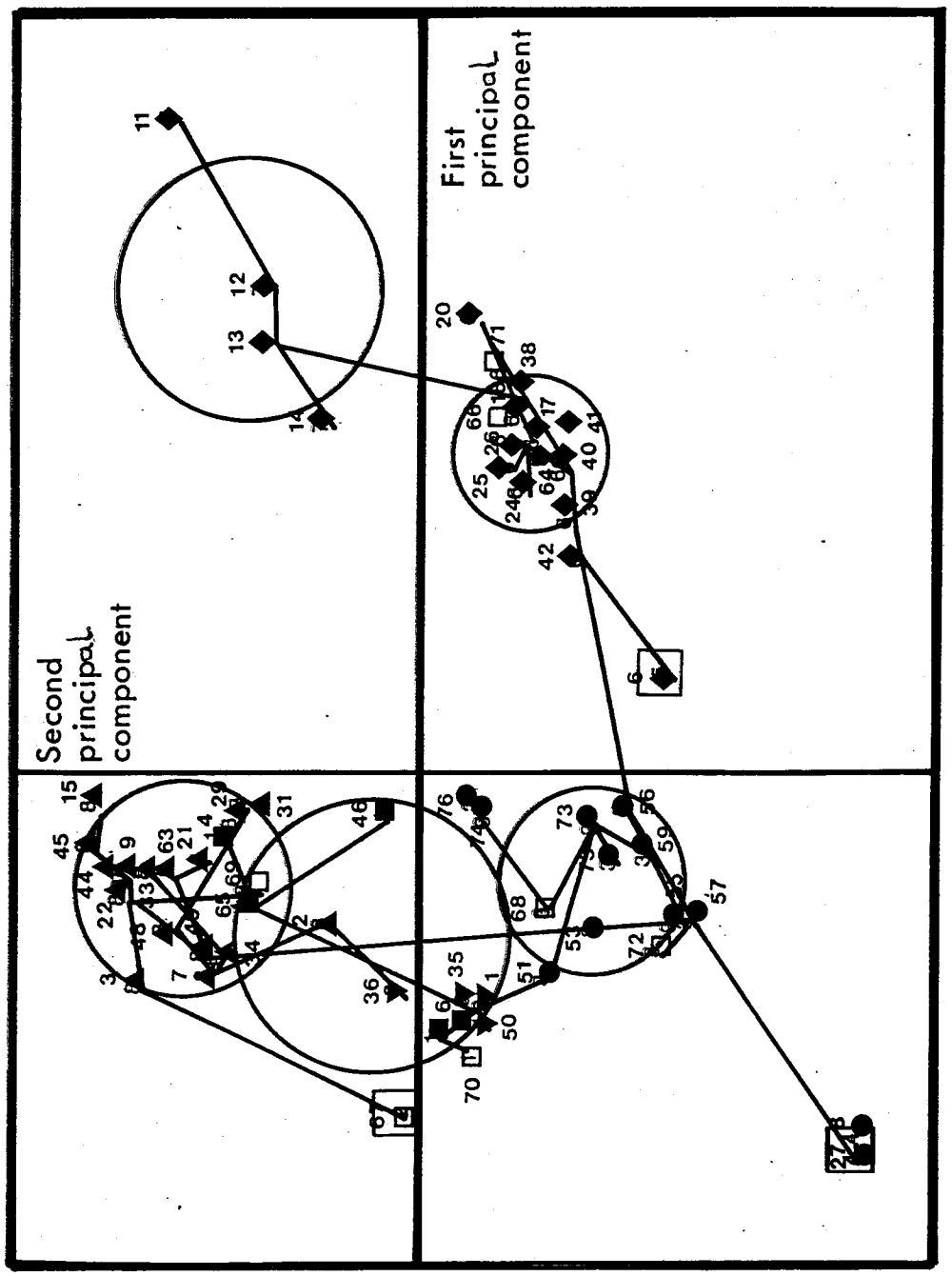


Table 7.5 Taxon membership of nine sectional clusters

Section Cluster	Taxa	Sensu Kupicha Section
1	V. oroboides, V. balansae, V. truncatula, V. sepium (var. sepium, var. montana & V. eriocalyx)	Atossa
2	V. dionysiensis	-
3	V. assyriaca, V. esdraelonensis, V. tigridis, V. galeata, V. hyrcanica, V. noeana (subsp. megalodonta & subsp. noeana), V. melanops (var. melanops & var. loiseaui), V. ciliatula, V. anatolica, V. mollis, V. pannonica (subsp. striata & subsp. pannonica), V. hybrida, V. sericocarpa, V. lutea (subsp. lutea & subsp. vestita)	Hypechusa
4	V. michauxii, V. aintabensis, V. peregrina	Peregrinae
5	V. cuspidata, V. lathyroides	Vicia
6	V. pyrenaica, V. sativa (subsp. nigra, subsp. amphicarpa, subsp. incisa, subsp. devia, subsp. sativa & subsp. macrocarpa), V. barbazitae (var. barbazitae & var. incisa), V. qatmensis, V. grandiflora (var. grandiflora & var. incisa)	Vicia
7	V. bithynica	Faba
8	V. eristalioides, V. kalakhensis, V. johannis (var. ecirrhosa, var. procumbens & var. johannis), V. galilaea (var. galilaea & var. faboidea), V. serratifolia V. narbonensis (var. salmonea, var. jordanica, var. affinis, var. aegyptiaca & var. narbonensis) V. hyaeniscyamus	Faba
9	V. faba (subsp. paucijuga & subsp. faba, var. minor, var. equina & var. faba)	Faba

7.4 Detailed Study of Closely Related Taxa

The subgeneric analysis indicates nine groupings or sections. Among these, two of the sections used by Kupicha (1976), Faba and Vicia, are split and a third, Hypechusa, is shown to contain a relatively large, complex of related taxa. In her discussion of the circumscription of these three sections, she comments that it is likely that these groupings could be subdivided further. This is supported by both the specimen and taxon analyses. The modification of her classification, suggested by the results of the analyses, requires the splitting of her sections. To confirm the analysis results for these three groups it was decided to reanalyse each section using character sets specifically selected for the purpose. Characters were selected on the basis of their F values and whether they were considered "good" discriminators for the taxa being analysed. Each of the data sets was analysed using single and centroid linkage cluster analysis, and principal components analysis.

7.4.1 Analysis of Section Faba Sensu Kupicha (1976)

Section Faba sensu Kupicha (1976) contains six species (V. faba, V. narbonensis, V. galilaea, V. hyaeniscyamus, V. johannis and V. bithynica). Subsequent to the publication of her classification two further species, V. kalakhensis and V. eristalioides were added to this section using her criteria. Khattab (1987) argues strongly for the inclusion of a ninth species, V. serratifolia, to be split away from V. narbonensis. The results of the above analysis support this and so V. serratifolia is included in this analysis. These nine species with their subordinate subspecies and varieties were included in the analyses, making a total data set of 19 taxa by 61 characters (character set F61 in Appendix 3).

The results of the single linkage cluster analysis are summarised in 18 linkage diagrams arranged in decreasing similarity from a similarity level of 0.9429 to 0.4715, when all 19 taxa are included in one cluster. Those diagrams which

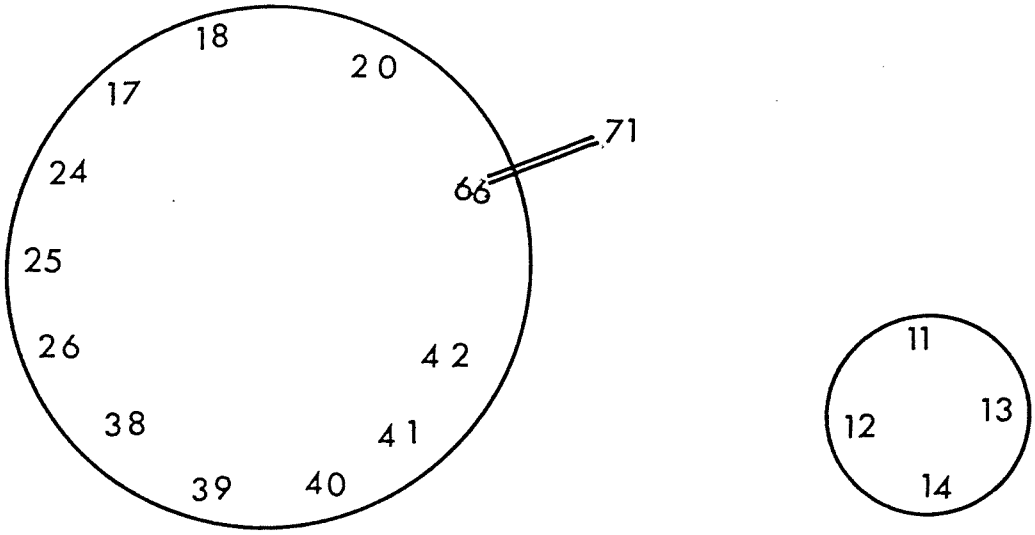
are most pertinent to elucidating the intra-sectional relations are shown in Figures 7.38-7.40. Figure 7.38 includes two clusters of taxa. The larger contains V. narbonensis, V. serratifolia, V. galilaea, V. hyaeniscyamus, V. johannis taxa encircled (the V. narbonensis complex taxa) and V. eristalioides, which forms a link with the encircled cluster via V. kalakhensis. The smaller second cluster contains the four V. faba taxa. V. bithynica has yet to form links with other taxa.

At the similarity level of 0.5325, two distinct clusters are still seen, see Figure 7.39. The larger cluster containing the V. narbonensis taxa with V. eristalioides as an outlier, forms a new link with V. bithynica. This species joins the large cluster via three V. narbonensis taxa and V. galilaea var. galilaea. At this similarity level V. eristalioides also forms multiple links with taxa included in the encircled cluster. There is a significant decrease in similarity to 0.4715 before the next inter-cluster link is made, drawn in Figure 7.40, this joins the V. narbonensis complex and V. faba clusters. The actual link is made between V. narbonensis var. jordanica and V.faba subsp. faba var. equina.

Figure 7.41 shows the results of the average linkage cluster analysis. The results are similar to those produced by the single linkage. The most closely related taxa are the two varieties of V. galilaea and two of the V. johannis varieties, var. johannis and var. procumbens. The third V. johannis taxon, var. ecirrhosa, is more remote and joins the cluster after the two sets of closely related varieties of the two species have linked. The data for V. johannis var. ecirrhosa is based on one collection and there are no legume or seed data available, which may explain its apparent misplacement. All five of the V. narbonensis varieties cluster together, with V. serratifolia forming an outlier of this group. V. hyaeniscyamus and V. kalakhensis are slightly

Figure 7.38. Single linkage cluster analysis of taxon data for sect. Faba using 61 characters.

Level 16. Similarity - 0.5578.



OTU 6 Not yet linked

Figure 7.39. Single linkage cluster analysis of taxon data for sect. Faba using 61 characters.

Level 17. Similarity - 0.5325.

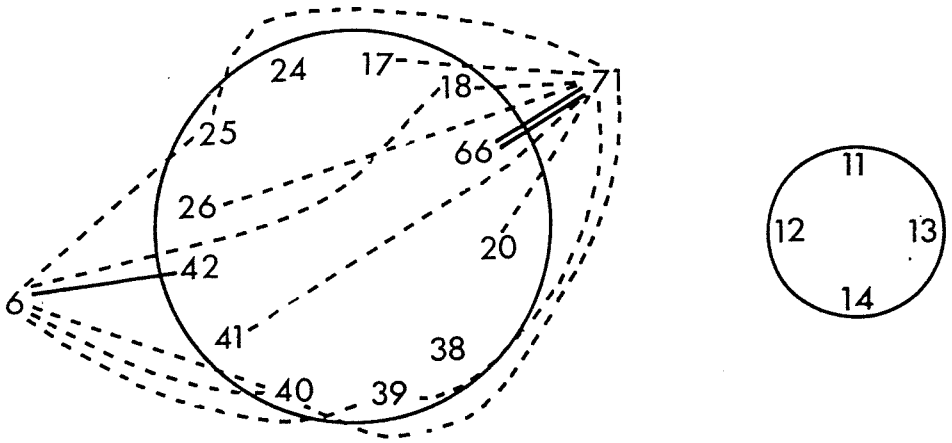


Figure 7.40. Single linkage cluster analysis of taxon data for sect. Faba using 61 characters.

Level 18. Similarity - 0.4715.

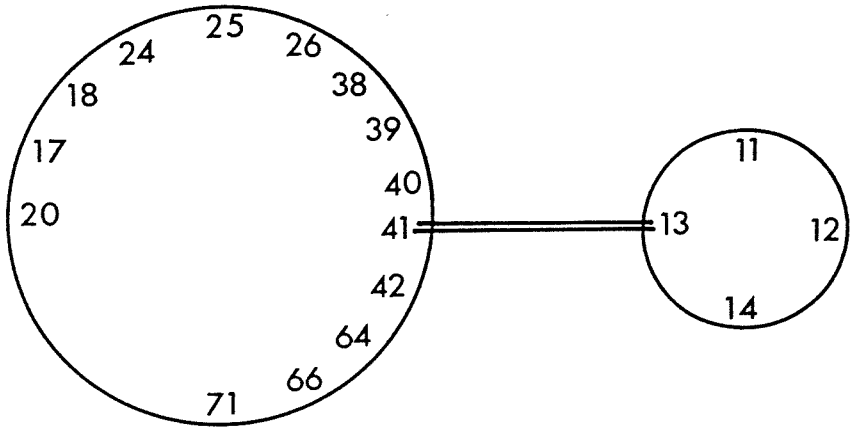
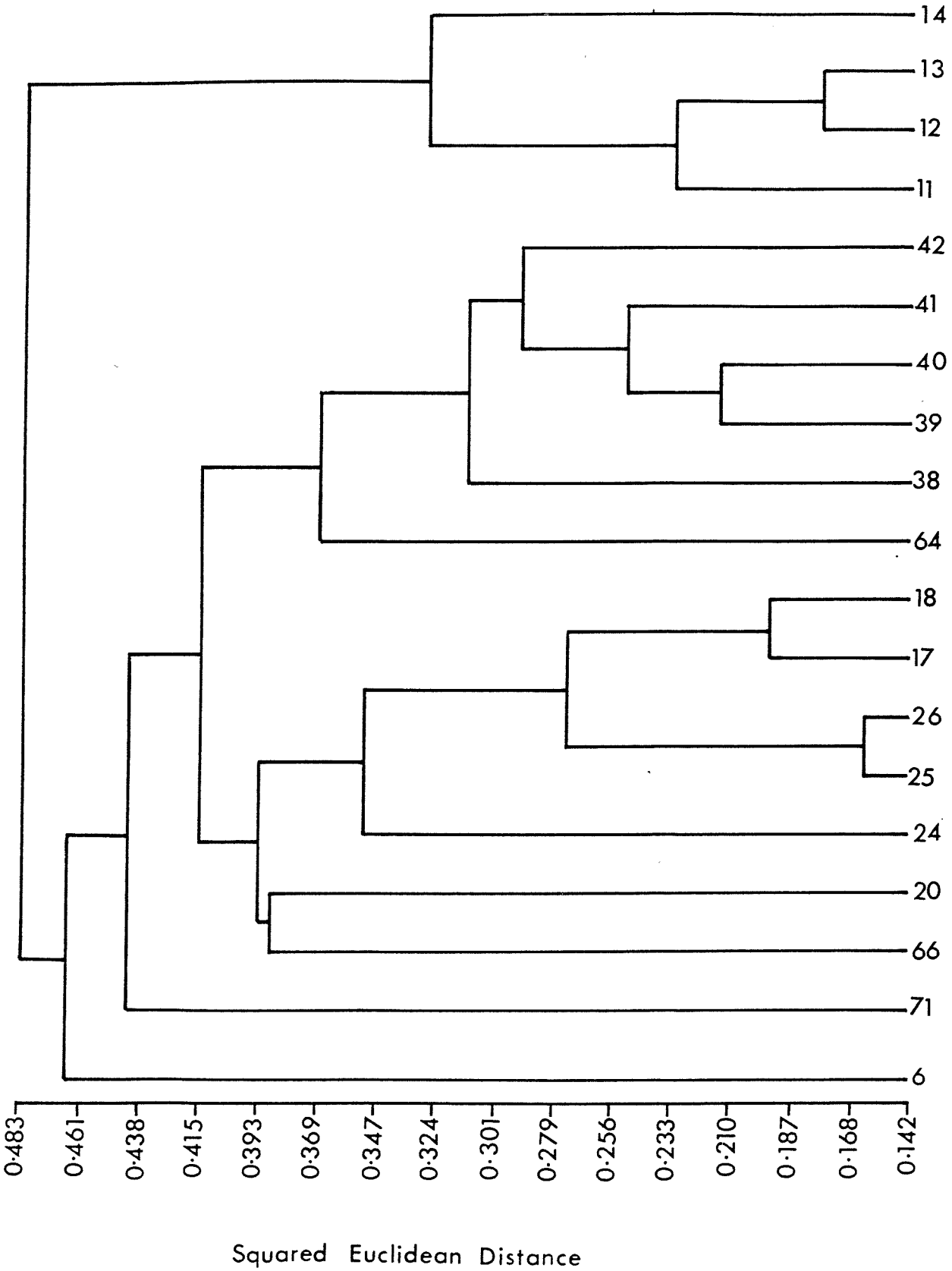


Figure 7.41. Average linkage cluster analysis of taxon data for sect. Faba using 61 characters.



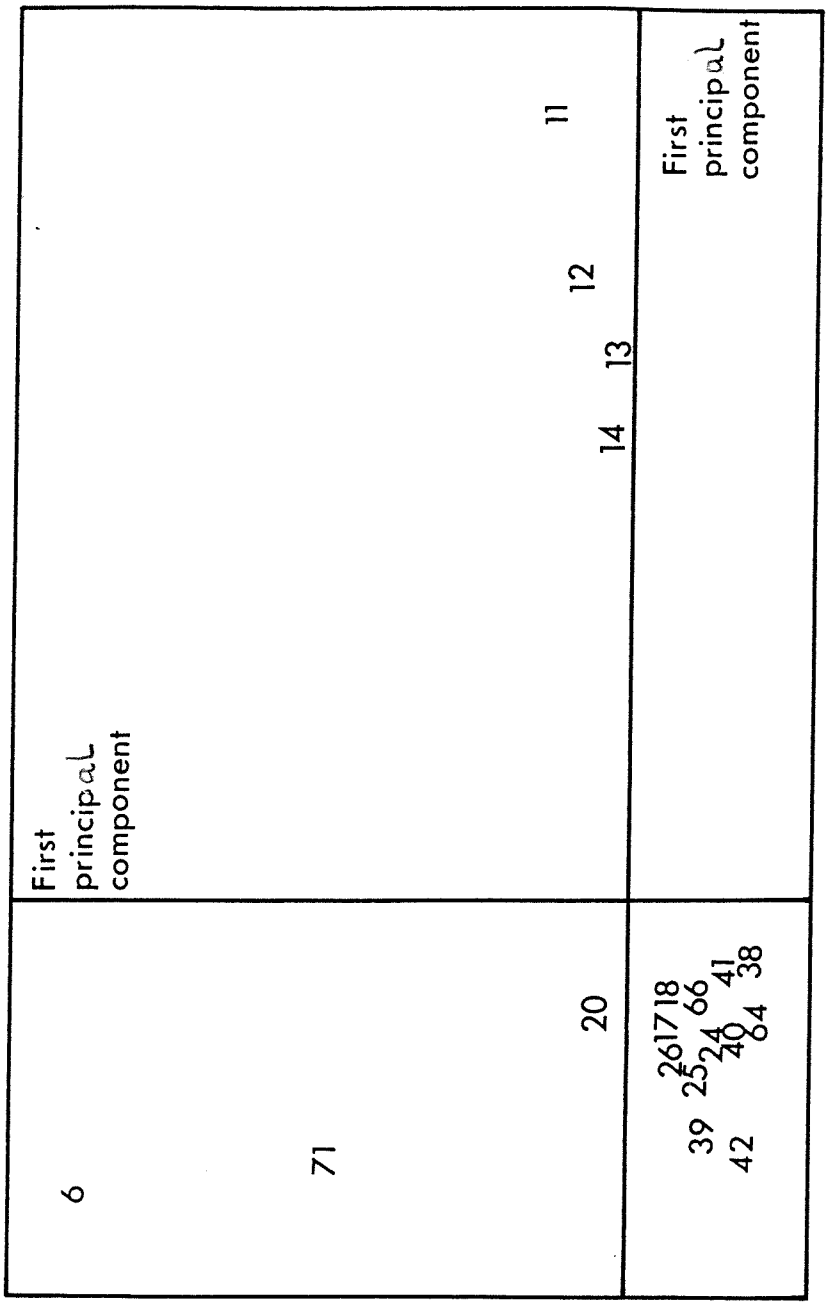
more closely related to each other than they are to V. galilaea and V. johannis. This inner grouping of six species forms the V. narbonensis complex of closely related taxa and within this group specific identification is often problematic, unless the complete set of key characters is available. The newly discovered species from Southern Turkey, V. eristalioides, is not a member of the complex, though it is closer to the complex than to the other species being analysed. V. bithynica is remote, both to the complex and V. eristalioides, and the use of distinct sectional rank may be appropriate. Finally, the most isolated cluster within sect. Faba sensu Kupicha is V. faba itself.

The result of the principal components analysis for the 19 taxa is shown in Figure 7.42, where the first two principal components are plotted. They account for 33.26 and 12.54 percent of the variance respectively, which gives a cumulative variance of 45.80 percent. The four groupings identified in the above analyses are again apparent. The six species of the V. narbonensis complex form a fairly tight cluster (with V. hyaeniscyamus slightly peripheral), then V. eristalioides, V. bithynica and V. faba each increasingly isolated. It is interesting to note the intermediate position taken by V. eristalioides between V. bithynica and the V. narbonensis complex taxa. Prior to the discovery of V. eristalioides, V. bithynica was considered by Maxted and Khattab (unpublished) to have only a superficial relationship to section Faba. The fact that V. eristalioides acts as an intermediary in this manner indicates that, although V. bithynica is relatively remote from the V. narbonensis complex, it is more closely related to the complex than to any of the other subgeneric groupings.

7.4.2 Analysis of Section Vicia Sensu Kupicha (1976)

Kupicha (1976) includes six species (V. pyrenaica, V. sativa, V. grandiflora, V. barbazitae, V. lathyroides and V. cuspidata) in her sect. Vicia. She was unaware of V.

Figure 7.42. Principal components analysis of taxon data for sect. Faba using 61 characters



qatmensis, an endemic from Syria, which following her circumscription should also be included in section Vicia. These seven species with their subordinate subspecies and varieties were included in the analyses, making a total data set of 14 taxa by 72 characters (character set V72 in Appendix 3).

The results of the single linkage cluster analysis are summarised in 13 linkage diagrams arranged in decreasing similarity from a similarity level of 0.8611 to 0.4999, when all 14 taxa are included in one cluster. Those diagrams which are most useful in elucidating the intra-sectional relations are shown in Figures 7.43-7.44. Figure 7.43 shows two clusters, the larger containing V. pyrenaica, V. barbazitae, V. sativa, V. grandiflora and V. qatmensis and the smaller contains V. lathyroides and V. cuspidata. The larger cluster is formed by the union at this level of two subclusters made up of V. pyrenaica, V. barbazitae and V. sativa, and the yellow flowered and flat seeded V. grandiflora and V. qatmensis. The union of the two subclusters is made via the V. qatmensis and V. barbazitae var. incisa OTUs. When the similarity level is reduced to 0.4999, as is shown in Figure 7.44, both these subclusters are united in one encircled cluster and the second cluster containing V. lathyroides and V. cuspidata links to the main cluster. The link between the two main clusters is made between V. sativa subsp. nigra and V. lathyroides, two taxa which appear most similar apart from the difference in gross dimensions.

Figure 7.45 shows the results of the centroid linkage cluster analysis. As with the single linkage analysis, three main clusters are seen; one containing V. pyrenaica, V. barbazitae and V. sativa, the second containing V. grandiflora and V. qatmensis and the third more remote cluster of V. lathyroides and V. cuspidata. However, within the first cluster, this method of analysis does emphasise the distance between V. pyrenaica and V. sativa subsp. amphicarpa, and V.

Figure 7.43. Single linkage cluster analysis of taxon data for sect. Vicia using 72 characters.

Level 11. Similarity - 0.5948.

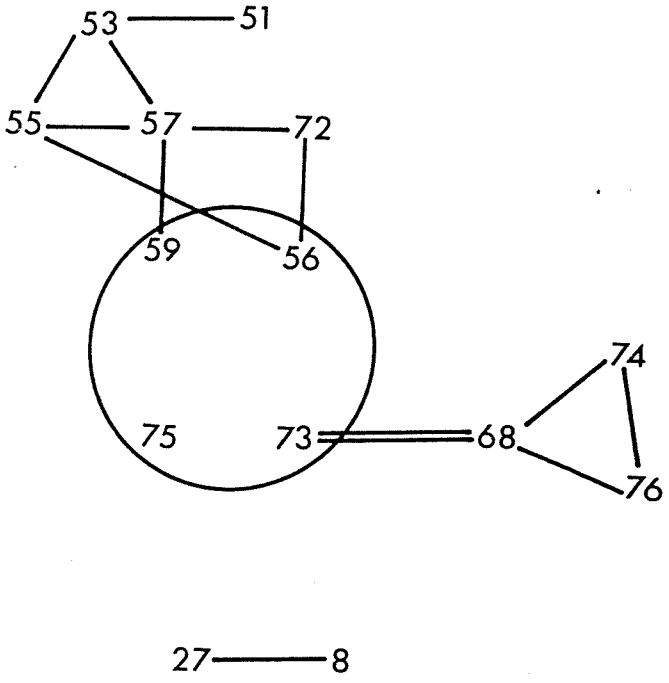


Figure 7.44. Single linkage cluster analysis of taxon data for sect. Vicia using 72 characters.

Level 12. Similarity - 0.4999.

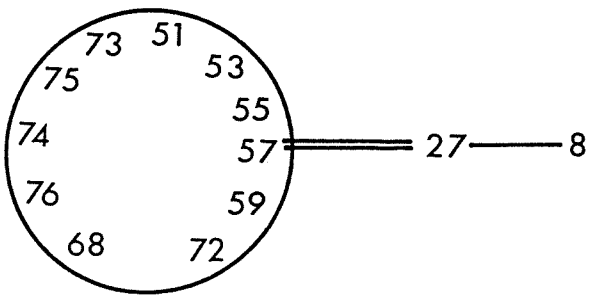
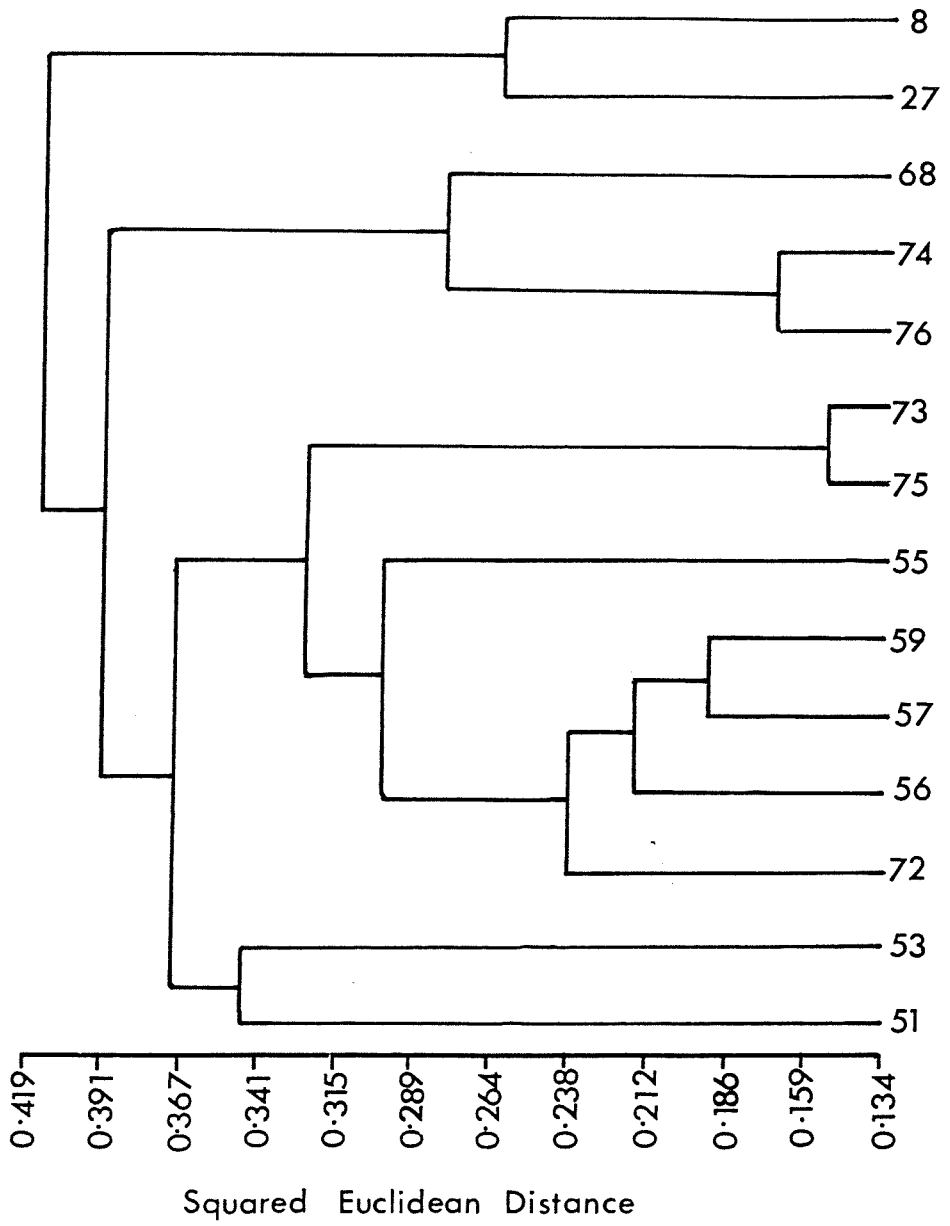


Figure 7.45. Centroid linkage cluster analysis of taxon data for sect. Vicia using 72 characters.



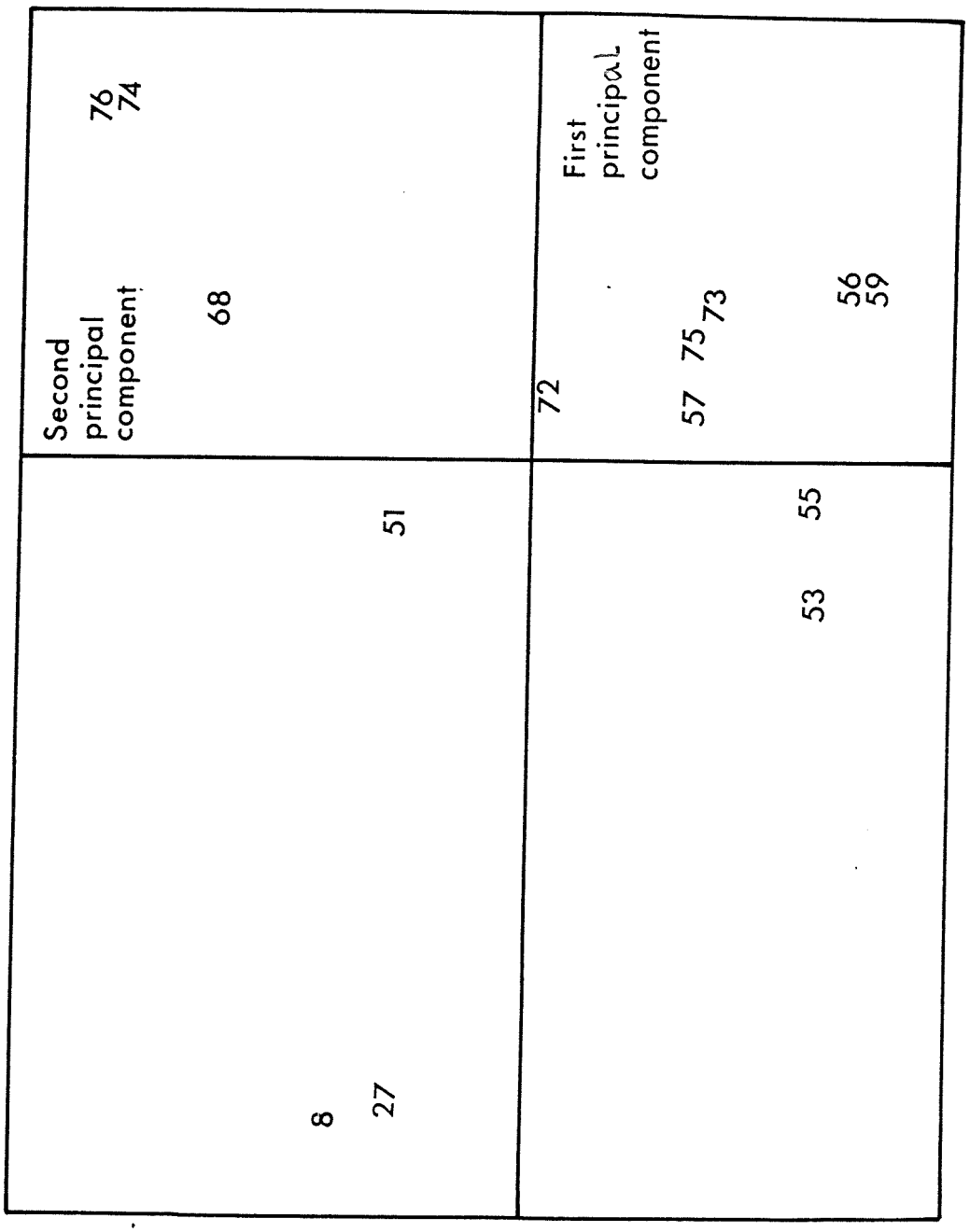
barbazitae and the other subspecies of V. sativa. V. pyrenaica, with its perennial life form, is likely to be isolated within this grouping, but the reason for the linking of subsp. amphicarpa to V. pyrenaica is unclear.

Interestingly in Figure 7.46, where the results of the principal components analysis shown, V. pyrenaica and V. sativa subsp. amphicarpa are not shown to be closely allied. The first two principal components, plotted in Figure 4.46, account for 29.62 and 17.38 percent of the variance respectively, which gives a cumulative variance of 47.00 percent. The three basic clusters seen with the other methods of analysis are seen with V. lathyroides and V. cuspidata being particularly distinct. However, the separation of the two lower level clusters is partially blurred by the intermediate placement of V. pyrenaica. If the minimum spanning tree is overlaid on Figure 7.46 this relationship is resolved. V. pyrenaica is shown to be more closely related to V. sativa subsp. nigra than V. qatmensis, its closest spatial ally in the scatter plot.

7.4.3 Analysis of Section Hypechusa Sensu Kupicha (1976)

Kupicha's circumscription of sect. Hypechusa includes 12 species: V. anatolica, V. assyriaca, V. ciliatula, V. esdraelonensis, V. galeata, V. hybrida, V. hyrcanica, V. lutea, V. melanops, V. noeana, V. pannonica and V. sericocarpa. V. tigridis, an endemic from Syria, should be added to this list, as it is covered by Kupicha's diagnosis of section Hypechusa. A fourteenth species, V. mollis should also be added to this list. Though placed by Kupicha in sect. Peregrinae, V. mollis has been consistently shown throughout the above analysis to form closer links with sect. Hypechusa than to sect. Peregrinae. Therefore it is included here. The data set for this analysis consists of 18 taxa and 27 characters (character set H27 in Appendix 3).

Figure 4.46. Principal components analysis of taxon data for sect. Vicia using 72 characters.



The results of the single linkage cluster analysis are summarised in 17 linkage diagrams arranged in decreasing similarity from a similarity level of 0.8301 to 0.5402. The diagram most useful in explaining the relationships within the section is shown in Figure 7.47. The linkage diagram shows two clusters, one cluster contains the species with a relatively broad standard: V. assyriaca, V. esdraelonensis, V. galeata, V. hyrcanica, V. noeana and V. tigridis, and the other contains species with a narrower standard: V. anatolica, V. ciliatula, V. hybrida, V. lutea, V. melanops, V. mollis, V. pannonica and V. sericocarpa. Having established that the taxa of this section can be split into two clusters, it is apparent that both clusters in the final linkage diagram still have outlying taxa not included in the encircled clusters. This indicates that the two clusters are composed of relatively heterogenous elements and within the two groups further subgroups may occur.

This pattern is also reflected in the results of the centroid linkage cluster analysis shown in Figure 7.48. As with the single linkage analysis, the same two broad clusters of taxa are seen. However, within the group with narrower standards, the cluster subdivides into two or three subgroups. Notably one of these distinct subclusters contains all the species which have a wing apex spot, V. anatolica, V. ciliatula and V. melanops. This analysis indicates a misleading separation of the two subspecies of V. pannonica, subsp. striata (49) and subsp. pannonica (48). These two subspecies are distinguished on the basis of corolla colour and size. Both these characters are included in the character set and it is likely that by using a relatively small character set of 27 characters, these characters are over weighted in comparison to the characters that would have united the two subspecies into one taxon.

The results of the principal components analysis are shown in Figure 7.49. The first two principal components are

Figure 7.47. Single linkage cluster analysis of taxon data for sect. Hypechusa using 27 characters.

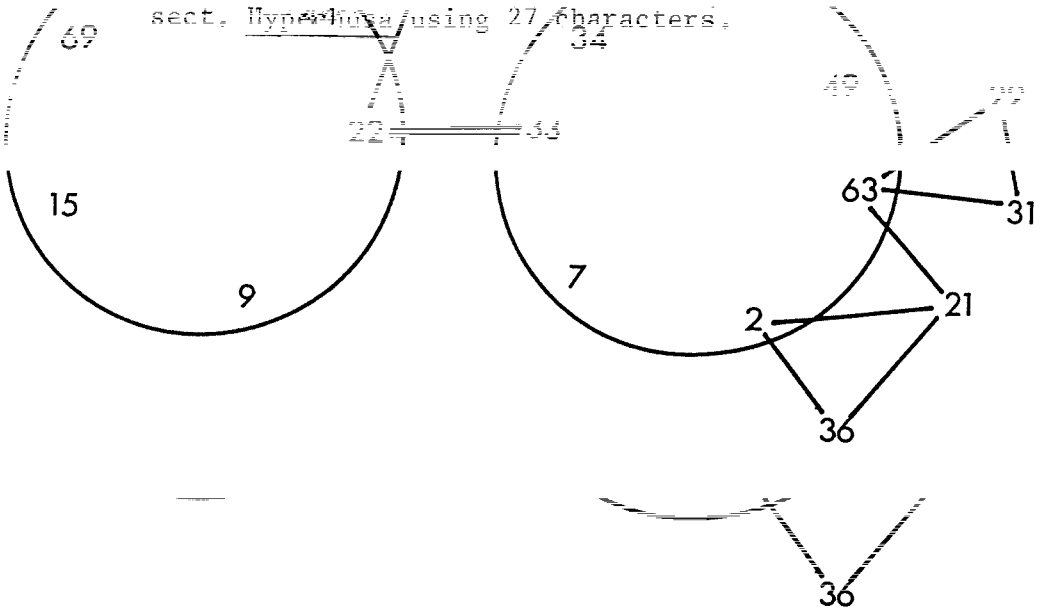
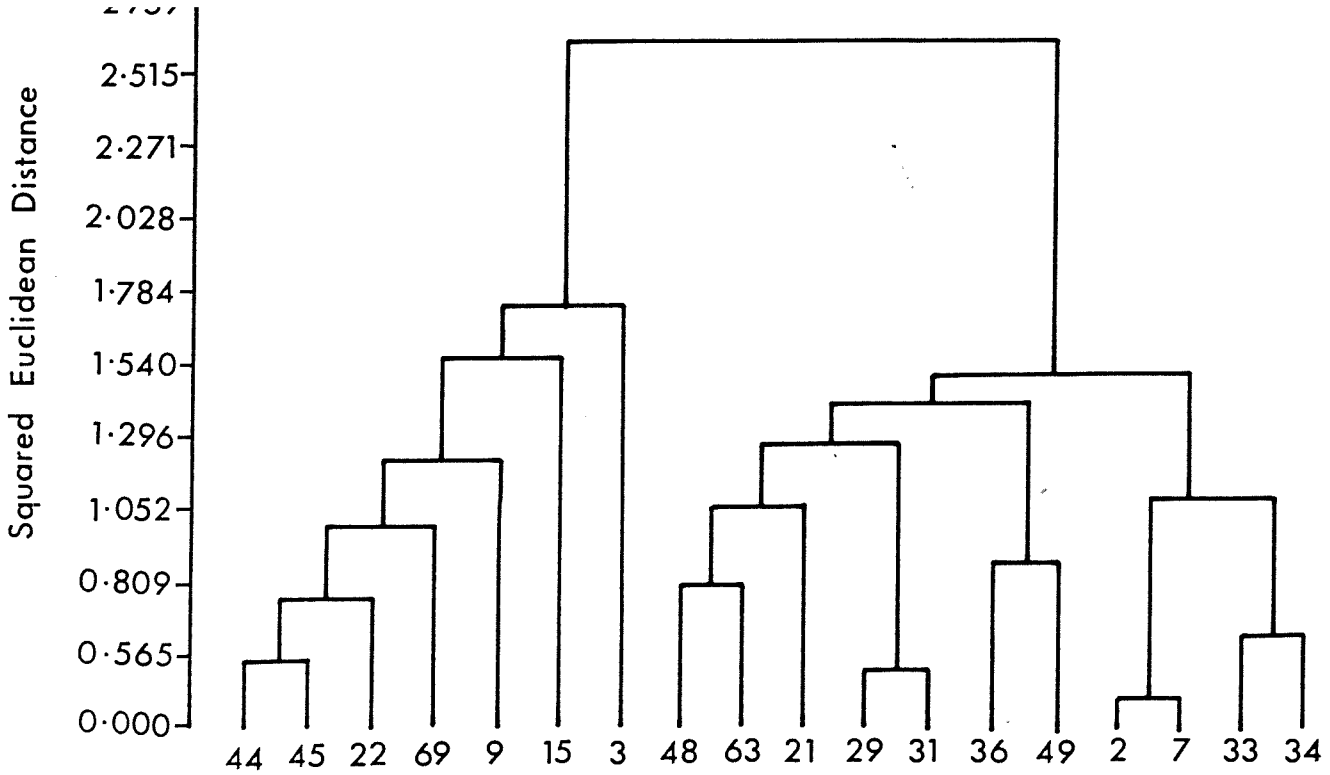
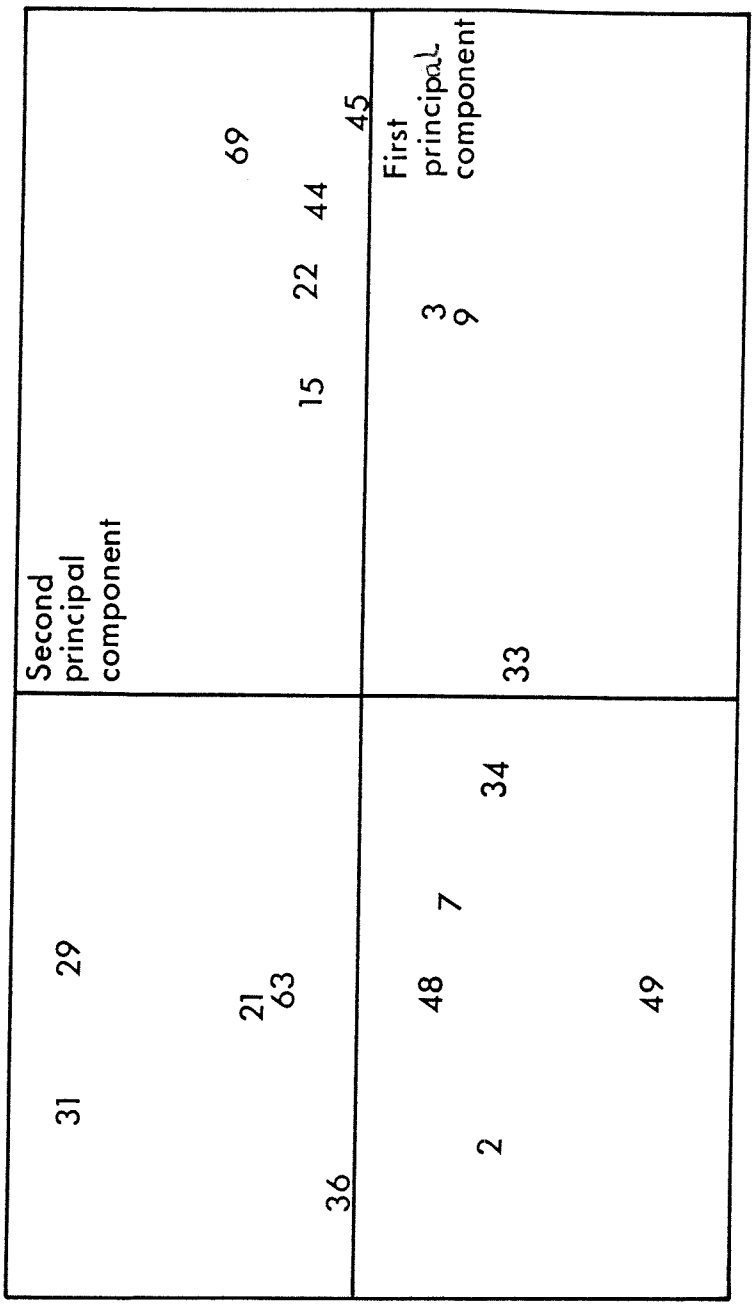


Figure 7.48. Centroid linkage cluster analysis of taxon data for sect. Hypechusa using 72 characters.



220

Figure 7.49. Principal components analysis of taxon data for sect. Hypechusa using 27 characters.



plotted and account for 40.84 and 12.36 percent of the variance respectively, which gives a cumulative variance of 53.21 percent. The two basic clusters can be identified and contain the same taxa as seen in the other methods of analysis. The taxa that have a broader standard can be seen to form a tighter cluster than the cluster containing the taxa with narrower standards.

CHAPTER EIGHT
LITERATURE-BASED TAXONOMIC EVIDENCE

8.1 Introduction

In the preceding chapter the evidence used to postulate relationships between taxa resulted from a novel morphological study. Time was not available to undertake complementary non-morphological studies. The taxonomic literature emphasises the importance of collecting information from a broad range of data types if the classifications produced are to have any practical stability (Davis & Heywood, 1973). Accordingly this chapter is an attempt to summarise the most pertinent research available from the literature.

Kupicha (1974) points out that the literature relating to non-morphological investigation of the Vicieae is copious. This is true only in a limited sense. The cytological, biosystematic and phytochemical investigations have been extensive, but other sources of evidence such as palynological and anatomical information have not been fully exploited. Within the restricted grouping of subg. Vicia, the information is much more limited, with the majority of authors only including V. faba and V. sativa as token Vicieae in family wide surveys. So the evidence available from the literature is not comprehensive for the 72 taxa included in the investigation.

This chapter is divided into nine sections: the introduction, seven sections each concerning a particular type of information, followed by a summary of the data and its relationship to existing patterns in subgenus Vicia taxonomy. Wherever lists of taxa are provided the specific arrangement and sectional names used will follow Kupicha (1976). Taxa not known to Kupicha will be listed in alphabetical order following the Kupicha taxa.

8.2 Phytogeographical Evidence

The genus Vicia is primarily Euro-Asiatic, with other distributional centres in North America, South America, East Africa and Hawaii (Kupicha, 1976; Hanelt & Mettin, 1989). The worldwide distribution of 42 subg. Vicia taxa is detailed by Allkin et al. (1983c) and is summarised in Table 8.1. The distributional information published by Allkin et al. is taken from the Vicieae Project Database (held at The Department of Biology, The University, Southampton, U.K.). The database contains only native records, excluding distributional information for wholly cultivated species and alien records.

A detailed study of the phytogeography of Vicia was undertaken by Kupicha (1974). She discusses Vicia phytogeography using distributional and isoflor mapping. In both cases she does not discuss the details of individual species, but rather sections as defined in her classification. Her sectional distribution map for the five sections of subg. Vicia is shown in Figure 8.1. For her treatment she includes taxa only if they are wild or fully naturalised in part of their range, i.e. she does not map V. faba, which is not known in the wild or map V. sativa in the Americas or Africa. She states that over half the North American species are introductions from the Old World (V. oroboides, V. sepium, V. lathyroides, V. sativa, V. grandiflora, V. lutea, V. pannonica, V. narbonensis and V. faba) and only includes these for their native areas.

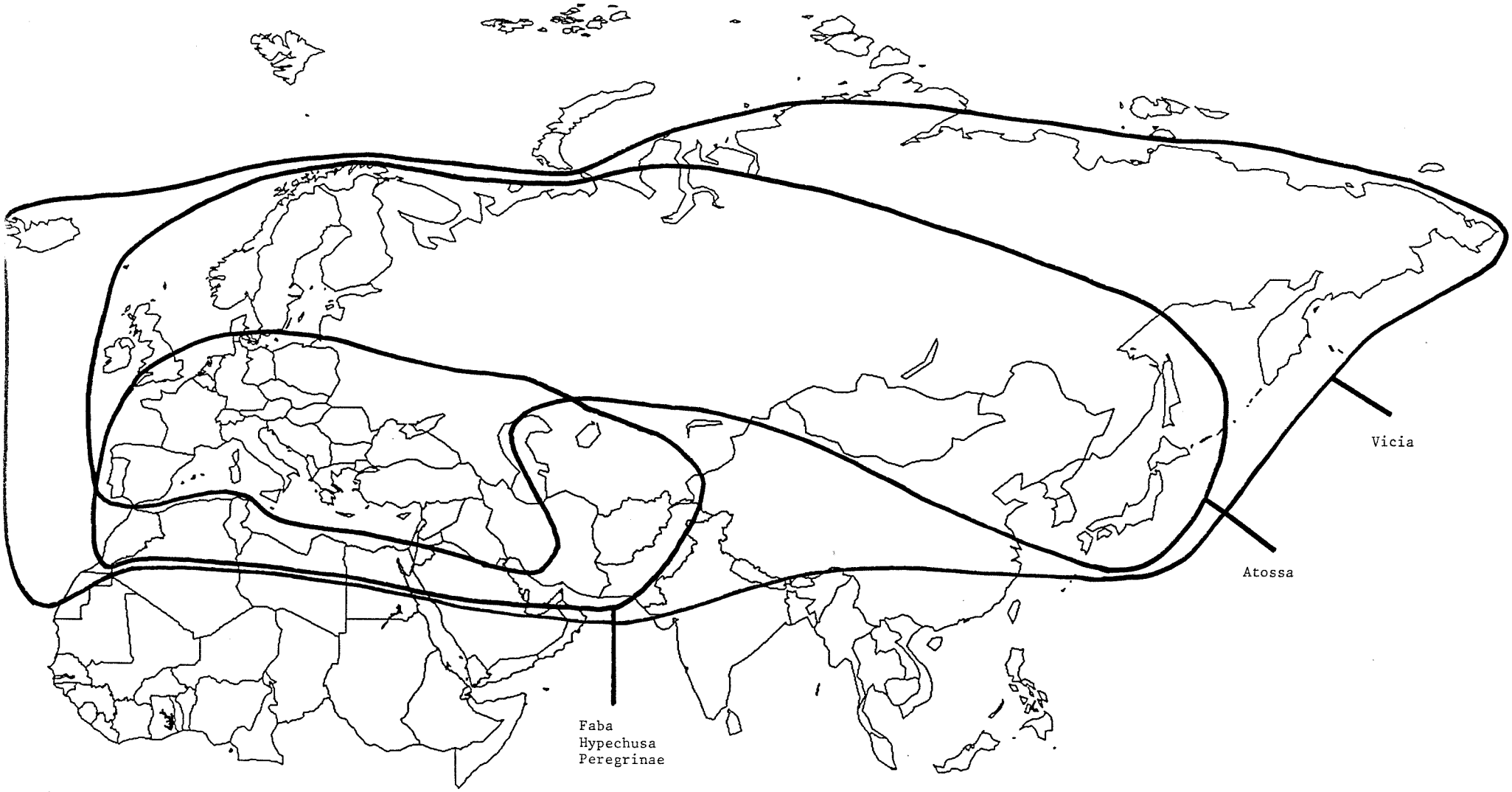
The distribution of the five sections of subg. Vicia sensu Kupicha (1976) (Figure 8.1) centre on the North East Mediterranean, with the highest concentration of species found in the fertile crescent countries of Turkey, Lebanon, Syria, Soviet Asia, Iran and Iraq. The distribution of sect. Peregrinae is restricted to this area. The other four sections, Hypechusa, Faba, Vicia and Atossa are more geographically dispersed, covering Europe, North Africa and spreading across to the extreme North-east of Asia. It should

Table 8.1. Geographical distribution of Vicia subgenus Vicia species.

Adapted from Kupicha (1974) and Allkin et al. (1983c).

Sect.	Species	Distribution
Atossa		
	V. oroboides	Central South East Europe
	V. sepium	European & Northern Asia
	V. balansae	Russia & Turkey in Asia
	V. truncatula	West Asia
Vicia		
	V. pyrenaica	France & Spain
	V. sativa	Europe, Asia & North Africa
	V. grandiflora	East Europe & West Asia
	V. barbazitae	South Europe & West Asia
	V. lathyroides	Europe, West Asia & Algeria
	V. cuspidata	South east Europe & West Asia
Faba		
	V. faba	Only cultivation
	V. narbonensis	Europe, West Asia & North Africa
	V. galilaea	Israel & Turkey
	V. hyaeniscyamus	Syria & Lebanon
	V. johannis	South Europe, West Asia & Libya
	V. bithynica	Europe, West Asia & Algeria
Hypechusa		
	V. anatolica	Iran, Russia & Turkey in Asia
	V. assyriaca	West Asia
	V. ciliatula	Russia & Turkey in Asia
	V. esdraelonensis	West Asia
	V. galeata	West Asia
	V. hybrida	Europe, West Asia & North Africa
	V. hyrcanica	South West Asia
	V. lutea	South Europe, West Asia & N. Africa
	V. melanops	South Europe & Turkey
	V. noeana	West Asia
	V. pannonica	South Europe, West Asia & N. Africa
	V. sericocarpa	West Asia
Peregrinae		
	V. aintabensis	West Asia
	V. michauxii	West Asia
	V. mollis	West Asia
	V. peregrina	South Europe, West Asia & N. Africa

Figure 8.1. Distributional map for the sections recognised by Kupicha, 1976.



be noted that in the case of Vicia and Atossa, there are two species, V. sativa and V. sepium, which greatly extend the distributional area covered by these sections. The sectional distribution and distributional concentration patterns provide little help in allying the subg. Vicia taxa, as all five sections have overlapping distribution patterns with the same centre of diversity.

Two authors, Schäfer (1973) and Khattab (1987) specifically studied the distributional pattern of sect. Faba species. Schäfer (1973) concluded that V. faba and the V. narbonensis complex are distributed throughout the Mediterranean region, but with a concentration of species in the Near East. Khattab (1987) pointed out that V. bithynica, V. serratifolia and V. narbonensis have the widest distributions from Britain through Europe and the Mediterranean Basin to Iran and Afghanistan, although Plitmann (1967) suggested V. bithynica had been introduced to Afghanistan. The distribution of V. johannis is more restricted and is focused on the Northern Mediterranean countries. Khattab records a few specimens from Poland and Afghanistan, both of which he regards as "introductions or casuals". However, I have personally collected from large populations of V. johannis on the Soviet - Afghanistan border, which appeared to be native. The remaining species are much more limited endemics of the Near East, V. galilaea to Turkey and Israel, V. hyaeniscyamus to Syria and Lebanon and V. kalakhensis to Syria. The existence of V. eristalioides was not known to Schäfer or Khattab, but this sect. Faba species is a narrow endemic of South-west Antalya province, Turkey.

Khattab (1987) concluded that the distribution of section Faba taxa is unhelpful in allying taxa. The species which are closely related and whose specific separation he investigated for his study, were sympatric (e.g. V. serratifolia and V. narbonensis are both distributed through Europe, Asia and North Africa). He also questioned whether V. narbonensis and

V. johannis are distinct entities, but these two species are also sympatric. The other three species he investigated (V. galilaea, V. hyaeniscyamus and V. kalakhensis) are much rarer and have a much more restricted distribution. However, until the areas of Turkey, Syria, Lebanon and Palestine where they grow have been collected more systematically, it is difficult to say whether V. galilaea is, in fact, allopatric with V. kalakhensis and V. hyaeniscyamus.

A detailed discussion of the distributional pattern for those sections identified during this revision is provided in Chapter Nine.

8.3 Cytotaxonomic Evidence

This is the source of taxonomically important data based on chromosome studies. The information may vary from simple chromosome numbers through to highly sophisticated techniques which elucidate chromosome banding patterns (see Fukuda 1984). Varying levels of cytological detail are available for different Vicia species, from detailed DNA banding patterns to species where the chromosome number has yet to be counted. As Kupicha (1974) points out, the Vicieae are a popular group for cytological study, because their haploid number is relatively low, individual chromosomes are large and the group contains temperate crop species, so the literature is relatively copious.

The first extensive review of Vicia cytotaxonomy was undertaken by Sveshnikova in 1927, who includes 28 species of Vicia in a key using solely chromosome morphological characters. She distinguishes two groups of species, which largely correspond to the classification of Ascherson and Graebner (1909). The first grouping contains the Ervum and Cracca species, while the second contains Vicia and Faba. She identifies the latter as possessing mainly asymmetrical, acrocentric chromosomes.

Within subg. Vicia Sveshnikova splits the species into V. faba and Euvicia, the latter is further subdivided into two subgroups, the first containing V. hybrida, V. pannonica, V. lutea, V. sepium, V. grandiflora, V. peregrina, V. angustifolia, V. sativa, V. macrocarpa, and V. bithynica and the second containing V. narbonensis and V. serratifolia. This distinction of V. narbonensis and V. serratifolia from Euvicia is a reflection of Boissier's (1872) classification based on morphology, but is not a scheme used by recent authors. In her view, the difference in relative chromosome arm lengths of V. sativa and V. angustifolia, and V. narbonensis and V. serratifolia is indicative of their independent specific status. However she adds that in such polymorphic species a more systematic coverage of geographic races should be investigated prior to drawing any final conclusions. She discusses in detail the problem of locating the specific allies of V. bithynica, whether it should be allied to V. narbonensis (Ascherson and Graebner, 1909; Koch, 1836) or allied to V. grandiflora and V. sativa (Schmalhausen, 1895) and on the basis of chromosome shape she supports the latter alliance.

In a general review of Legume cytology Senn (1938) established that the overall base count for Vicia was x=7. Senn proposes that the Vicieae arose by hypoploidy from primitive ancestral stock of x=8, a proposal which is still current (Goldblatt, 1981). Since Senn, numerous species chromosome counts have accrued, as summarised in Table 8.2. The full details of individual counts are held in the database files CHROMNOS and CHROMREF, discussed in Chapter Ten. Chromosome counts for V. hyaeniscyamus and V. kalakhensis are included from an unpublished study undertaken by Maxted & Khattab, reported in Khattab (1987). These latter counts were corroborated by Maxted, Callimassia and Bennett (In Prep.) and will be discussed in detail elsewhere.

Table 8.2. Chromosome numbers of subgenus Vicia species.

For details of references see the discussion of file CHROMO in Chapter ten. DNA values are taken from Chooi, (1971a). They are calculated as a proportion of V. faba which is given a standard value of 100. DNA quantities are taken from Raina & Rees (1983), measured in picograms of DNA. Counts, in brackets, are single reports for a species where several other reports indicate a different diploid number. * = novel count.

Sect. Species	2n	DNA Value	DNA Amount
Atossa			
V. oroboides	14	-	-
V. sepium	(12),14,(16),18	35	9.34
V. balansae	12,14	-	-
V. truncatula	-	-	-
Vicia			
V. pyrenaica	14	-	8.29
V. sativa	10,12,14	20	4.50
V. grandiflora	(12),14	24.9	9.05
V. barbazitae	14*	-	-
V. lathyroides	(10),12	19.7	6.47
V. cuspidata	12*	-	-
Faba			
V. faba	12,(14)	100	27.07
V. narbonensis	14	55.5	16.11
V. galilaea	14	-	-
V. hyaeniscyamus	14*	-	-
V. johannis	14	-	14.14
V. bithynica	14	34.3	9.98
Hypechusa			
V. anatolica	10	56.2	15.66
V. assyriaca	12,14	-	-
V. ciliatula	10	-	-
V. esdraelonensis	12	-	-
V. galeata	12,14	32.2	-
V. hybrida	12	51.1	16.46
V. hyrcanica	12,(14)	50.5	15.22
V. lutea	14	55.6	18.03
V. melanops	10	86.1	20.02
V. noeana	12	-	-
V. pannonica	12	50.9	13.52
V. sericocarpa	12	-	-
Peregrinae			
V. aintabensis	14	-	-
V. michauxii	(12),14	62.3	20.68
V. mollis	10*,(14)	-	-
V. peregrina	(12),14	71.1	19.15
Not circumscribed by Kupicha			
V. dionysiensis	12*	-	-
V. eristalioides	14*	-	-
V. kalakhensis	14*	-	-
V. qatmensis	14*	-	-
V. serratifolia	14	-	15.63
V. tigridis	12*	-	-

The most common diploid number is 2n=14, but 2n=10 and 12 are also common, as well as a few records of 2n=16 and 18. Polyploidy is found in a few Vicia species, but not in subg. Vicia. The records of 2n=10 are restricted to two sections, Vicia and Hypechusa; in the latter it is interesting to note that this count is recorded for three species that are obviously morphologically closely related. The new count for V. mollis of 2n = 10, supports the morphological evidence in suggesting the exclusion of V. mollis from sect. Peregrinae (2n = 14) and its transfer to sect. Hypechusa (2n = 10 or 12).

Sect. Hypechusa is distinct. It is the only section with a base number other than 7. A reduction in chromosome number Hanelt and Mettin (1966) believe indicates that the taxa are the most cytologically advanced. On this basis sect. Hypechusa, some sect. Vicia and V. faba are considered advanced. This corroborates Kupicha's (1976) view of their relative advancement based on morphological characters. Within section Hypechusa, V. lutea is an oddity, with consistent counts of 2n=14. If the reduction is equated with evolutionary advancement, it follows that V. lutea should be regarded as the most primitive sect. Hypechusa species.

The two novel counts for V. cuspidata (2n = 12) and V. barbazitae (2n = 14) tie these species to their morphologically based allies in sect. Vicia, V. lathyroides and V. sativa subsp. sativa respectively. Within sect. Faba, the base number is 7, V. faba stands out as being radically different in chromosome number and morphology. The three novel counts for this group, V. hyaeniscyamus, V. kalakhensis and V. eristalioides are all 2n = 14. Which links these species to the V. narbonensis complex and V. bithynica. Sections Atossa and Peregrinae have diploid counts of 2n = 14, but in both cases there are a significant number of 2n = 12 counts.

One species that is particularly difficult to discuss cytologically is V. sepium, which has recorded diploid counts of 12, 14, 16 and 18. There are two records of 2n=18 (Baksay, 1954; Mattick in Tischler, 1950). This range of counts seems unlikely to be explained by errors in counting. The species has a relatively wide distribution (throughout the Temperate region of Eurasia) and is often found in isolated habitats remote from other Vicia species. The different numbers may possibly be explained by long periods of spatial isolation resulting in the evolution of local chromosome races suited to particular local conditions.

Various authors have produced detailed karyotype data and interpretations for subg. Vicia taxa (Coutinho, 1940; Cincura, 1962, 1970; Mettin, 1962; Stankevich, 1970; Hanelt & Mettin, 1966, 1970a; Mettin & Hanelt, 1964, 1973; Schäfer 1973; Pignone & Attolico, 1980; Kuta, 1980; Yamamoto et al., 1982; Roti-Michelozzi & Caffaro, 1984). The results of Hollings and Stace (1974) should be considered when interpreting the karyotypes of the V. sativa aggregate. It is generally assumed that intra-taxon cytological variation will be minimal, but within sect. Vicia they found a relatively large amount of intra-taxon variation. They comment that this is likely for taxa with a broad geographic range. This type of chromosomal morphological variation within the V. sativa aggregate has also been recorded by several other workers, such as Sveshnikova (1927), Coutinho (1940), Watanabe & Yamada (1958), Mettin & Hanelt (1964, 1973), Hanelt & Mettin (1966), Yamamoto (1966), Ladizinsky (1978), Zohary & Plitmann (1979) and Raina & Rees (1983). In fact Hollings & Stace conclude that only one V. sativa chromosome could be consistently identified in 1974.

Schäfer (1973) reported a similar pattern of intra-specific variation in V. narbonensis. Yamamoto et al. (1982), studying the same species, distinguished only one chromosome type, although Yamamoto (1984) on reexamining the V.

narbonensis karyotype distinguishes three V. narbonensis karyotypes, corroborating Schäfer's original observations. Various studies indicate that extensive karyotype variation is limited to the V. sativa aggregate and the Faba group is in general more uniform. However, karyotype morphology will remain at least a crude source of taxonomic evidence until the position is clarified (Zohary & Plitmann, 1979). Perhaps it will prove more useful for comparison at and above the species level, than below it. Ladizinsky (1978) concludes that the variability in karyotype of V. sativa is not matched by a differentiation into corresponding distinguishable taxonomic units, except in V. sativa subsp. amphicarpa. In contrast, Zohary & Plitmann (1979) conclude, from cytological studies of the V. sativa aggregate, that although different members of the aggregate differ in chromosome number and karyotype, the chromosomes of the various subspecies retain considerable homology, pair in meiosis, form chiasmata and exchange chromosome segments.

Chooi (1971a) has, however, shown that the amount of DNA present does not vary significantly between accessions of the same taxon. Although basic chromosome architecture may vary, DNA content remains fairly consistent for particular taxa. In a comparative investigation of the quantity of nuclear DNA in 45 species of Vicia, Chooi found that, although the diploid chromosome number varies from 10 - 14 between species, the relative DNA content per cell varies approximately six-fold between the two extreme species, V. faba (100 - standard) to V. villosa subsp. eriocarpa (15.7). The relative DNA content per cell for the subg. Vicia species Chooi investigated is shown in Table 8.2. He found that the amount of DNA present was generally proportional to chromosome size and that the DNA content is particularly disjoint between subg. Vicia species. He concludes that increase in DNA is associated with evolutionary advancement (Chooi, 1971b). Chooi (1971a) also points out the striking difference in total DNA content, chromosome number and chromosome morphology, between the

morphologically allied V. faba and V. narbonensis. V. faba has the standard DNA quantity of 100, with 2n = 12 and five acrocentric and one very long symmetric chromosome. In contrast V. narbonensis has the relative DNA quantity of 54.5 units, 2n = 14 and the chromosomes are relatively symmetrical. Youssef & Hesemann (1985), in a similar study of the DNA content per cell of nine Egyptian species of Vicia, found similar relative DNA contents to those recorded by Chooi.

Raina & Rees (1983), in a broad survey of the absolute quantities of DNA per cell in 56 Vicia species, found the DNA content varied between four and twenty seven picograms. They found that increase in DNA content was achieved by equal increments to each chromosome and that increase in DNA content did not correlate with the accepted taxonomic patterns (they used Ball's 1968 classification). They concluded that speciation in Vicia has been accompanied by massive changes in chromosome size and DNA content, brought about by Robertsonian fusions and amplification of base sequences within each chromosome. As might be expected, patterns of DNA quantity varied between species in a similar manner to that found by Chooi (1971a), see Table 8.2.

One of the consistent problems throughout the taxonomic history of subg. Vicia has been the natural status and rank of the V. sativa complex taxa. Hanelt and Mettin (1964, 1966) suggest that this problem is insoluble using morphological data alone, but the distinct karyotypes indicate that V. angustifolia (2n=12), V. amphicarpa (2n=14), V. cordata (2n=10), V. sativa (2n=12), V. incisa (2n=12) and V. macrocarpa (2n=12) should each be regarded as independent species. Interestingly these are the subspecific taxa recognised by Davis & Plitmann (1970) on the basis of largely morphological studies of the species. Hanelt and Mettin (1964) go further and suggest on morphological, karyological and ecological evidence that V. angustifolia should be further split into subsp. angustifolia and subsp. segetalis. They

cite the variation in diploid number within the V. sativa complex as an example of a descending dysploid series, 2n=14 to 2n=12 to 2n=10 (Mettin and Hanelt, 1972; Hollings and Stace, 1974).

In a detailed, but limited investigation of three sect. Hypechusa species, Cincura (1970) argues that the reverse occurred evolving from a diploid number of 2n=10 to 2n=12. Ladizinsky & Temkin (1978) support this view following their examination of the V. sativa aggregate, where 2n = 10, 12, and 14 are known. V. sativa subsp. amphicarpa (2n = 14) would then be the most evolved subspecies. They argue that this species has amphicarpic pods, which are highly advantageous and an advanced characteristic of dry habitat legumes, a point also made by Plitmann (1973). So their argument would be that subsp. amphicarpa has a diploid number of 14, so 14 must be the most advanced number and the direction of the series would be 2n = 10 to 2n = 12 to 2n = 14.

Srivatava (1963), in a cytogenetic study of the 8 Indian species of Vicia, was the first author to note the alliance of V. narbonensis with V. faba on the shared secondary chromosomal constrictions. Plitmann (1967), in a review of the cytotaxonomy of annual species of Vicia of the Middle East, reports numerous novel chromosome counts. He describes section Vicia chromosomes as being typically sub-telocentric and section Faba as being mainly sub-metacentric (excluding V. faba which is telocentric). He compares the karyograms of the species he investigated and notes resemblances between V. galilaea and V. narbonensis var. narbonensis, but these karyotypes differ markedly from V. bithynica and V. narbonensis var. serratifolia. The karyograms suggest further close relationships between V. sericocarpa and V. mollis, which does not support Kupicha's placement of these species in two separate sections. The close relationship between V. sericocarpa and V. mollis is supported by the results presented in the previous chapter. The karyograms presented

by Plitmann also suggest a close relationship between V. bithynica and V. grandiflora, which supports the findings of Sveshnikova (1927), but contradicts the current accepted placement of V. bithynica. Plitmann shows that V. peregrina, V. aintabensis, and V. michauxii are cytologically allied, which is also suggested by morphological data. Plitmann concludes that his series Hyrcanicae is a cytologically heterogeneous group. Tantalisingly though he does not expand on this, but it would be interesting to know if his evidence supports a division of this group into two subgroups, as is suggested by the results of the phenetic analysis.

Hanelt et al. (1972) undertook a largely literature-based study of V. faba and the V. narbonensis complex in an attempt to elucidate the origins of the faba bean. A more detailed experimental investigation of this group was undertaken by Schäfer (1973), who provided morphological, cytological and systematic evidence. She concludes, from her cytological studies, that V. narbonensis (three types), V. galilaea, V. johannis, V. serratifolia and V. faba each have clearly definable karyotypes. She observes that V. faba is remote from the V. narbonensis complex species, although the latter are similar in respect to other Vicia species. Ladizinsky (1975a), investigating the origin of the broad bean, finds very little cytological difference between V. narbonensis (finding only one type), V. galilaea and the material he refers to as V. hyaeniscyamus, but these all differ markedly from the V. faba material examined. Ladizinsky refers to V. hyaeniscyamus but in fact his material is likely to be V. galilaea, see Khattab et al. (1988) for clarification and discussion.

Working on the subspecific taxa of V. faba, Pignone & Cottolico (1980) found that karyotypic variation between the four sub-specific taxa was slight. They considered that the differentiation among the four taxa is insufficient to support the sub-specific classification of Muratova (1931). However,

they note that V. faba subsp. paucijuga has the most distinct chromosome morphology. Perrino & Pignone (1981) undertook a broader survey of sect. Faba, using banding patterns induced by quinacrine and the benzimide derivative H33258. They conclude that the species investigated can be divided into two groups, V. faba and V. bithynica, and V. narbonensis, V. serratifolia, V. galilaea and V. johannis. The latter four species have karyotypes with similar chromosome shape, dimension and quantities of DNA.

Schäfer (1973) showed that these karyotypes can be derived from one another by single translocation. V. faba and V. bithynica share relatively similar chromosome shape and banding patterns, but their DNA quantity is quite different. Perrino & Pignone (1981) tentatively conclude that V. bithynica is the closest ally of V. faba in section Faba, a view unsupported by other authors. Specifically this conclusion has been contradicted by Yamamoto et al. (1982). The latter group suggest, on the basis of karyotype data, three groupings of species: V. faba; V. narbonensis, V. serratifolia, V. galilaea, V. johannis and V. hyaeniscyamus; and V. bithynica. They found that the chromosomes of the Narbonensis complex taxa have only minor variations of the short arm lengths of sub-median and subterminal chromosomes and in the relative size of satellite chromosomes. Following Stebbins's (1971) hypothesis, they suggest that the relative short to long arm length ratios indicate that V. galilaea is the most evolved and V. narbonensis is the most primitive Narbonensis complex species. It should be noted that like Ladizinsky, the V. hyaeniscyamus material used by Yamamoto et al. (1982) is likely to have been misidentified.

Singh & Lelley (1983), investigating V. faba and V. narbonensis using Giemsa C-banding, found very different banding patterns for the two species. These findings were corroborated by Ramsay (1984), who reports the use of Leishman's C-banding techniques in an attempt to clarify the

relationships between four sect. Faba and two sect. Hypechusa species. His results support the division of sect. Faba sensu Kupicha (1976) into three subgroups as suggested by Yamamoto et al. (1982). The two sect. Hypechusa species included, V. lutea and V. melanops, have similar banding patterns, which are quite different to the sect. Faba species. He concludes that the relationship between V. faba and other Vicia species remains "obscure", as there is no C-banding homology between V. faba and the other species examined.

Most of the above research concentrates on two groups of plants, sect. Vicia and Faba. This is probably due to the agronomic use of species in both sections and the presence of the cosmopolitan weed, V. sativa, in sect. Vicia. Other work in the subgenus is limited. Hanelt and Mettin (1970b) attempt to resolve the natural placement of V. oroboides. This species is enigmatic: it is the one oroboid species that, since Koch (1836), has been included as a peripheral member of the subg. Vicia. Hanelt and Mettin's karyotypic investigations show the presence of only acrocentric chromosomes, which links the species to sect. Vicia and makes it more remote from Lathyrus. They suggest the superficial similarity to Lathyrus is due to parallel evolution, caused by similar environmental factors operating in related evolutionary lines.

Roti-Michelozzi & Caffaro (1984) examined morphologically and cytologically two sect. Hypechusa species, V. hybrida and V. lutea. Their cytological evidence shows the species to be very closely allied. Both species have similar sized and shaped chromosomes, and quantities of DNA. The main difference between the species is the chromosome number, 2n = 14 in V. lutea and 2n = 12 in V. hybrida. Although V. hybrida has fewer chromosomes, those present are longer than in V. lutea. As the general trend in Vicia karyology is from 2n = 14 to 2n = 12, they suggest V. hybrida is probably derived from V. lutea. Further cytological examination of these two

species with V. faba was undertaken by Lavania & Sharma (1984), using Giemsa C banding analysis. They found distinctive specific banding patterns. However, as would be expected, V. hybrida and V. lutea produced relatively similar patterns when compared with V. faba.

Several authors (Plitmann, 1967; Mettin & Hanelt, 1973; Schäfer, 1973) note that the speciation process in Vicia is accompanied by karyological differentiation of chromosome number and morphology. Isolation between putative species is initiated by chromosomal rearrangements. The autogamy of most annual species favours this kind of quantum speciation, so the putative species may even be sympatric. The implication of this is that detailed cytological studies could possibly provide further information about Vicia evolution.

8.4 Biosystematic evidence

Biosystematic or hybridisation evidence provides very useful information for the plant breeder and taxonomist. For the former, inter-specific or inter-generic crosses may clarify the potential of gene pools that may be tapped for adaptive characters. For the latter, hybridisation experiments provide information on the degree of isolation of taxa and help establish a particular taxon's placement in the taxonomic hierarchy.

Hegi (1923-24) records three methods of pollination in Vicia, cross-, self-pollination and cleistogamy. The latter is restricted to the appropriately named V. sativa subsp. amphicarpa in subg. Vicia. Although the flowers of subg. Vicia are well adapted to insect pollination, there is a predominance of autogamy.

Plitmann (1967), using bagging experiments, found the majority of annual subg. Vicia were self-pollinated. The two exceptions were V. faba and V. bithynica. In the autogamous species, the anthers reach maturity when the flower is still

relatively small (2-4 mm. long). Hanelt & Mettin (1970a) recorded a cross-pollination level of 30-50% for V. faba. McVetty & Nugent-Rigby (1984), carrying out similar field experiments, found that cross-pollination levels varied between 8.5 and 58.8%, and were affected by local environmental conditions. Levels of cross-pollination were higher in drier conditions.

Proctor & Yeo (1973) and Kambal et al. (1976) have shown that insect visitors, breaking the stigmatic papillae, stimulate pollination even if the plant is subsequently self-pollinated. Autogamy is advantageous for annual weedy herbs, where plants may live in isolated environments and where obligate cross-pollination would limit distribution. Zohary & Plitmann (1979) comment that Vicia retains an elaborate floral mechanism to ensure the occasional cross-pollination and subsequent genetic mixing. Ladizinsky (1975) found that V. narbonensis, V. galilaea and V. hyaeniscyamus, grown in insect-free cages, showed no lack of seed production compared to a control. The pollination method favoured by the five perennial species of subg. Vicia is unclear, but these species occupy different niches to the annual Vicia and so the advantage of autogamy would appear to be less obvious.

Numerous hybridization experiments have been undertaken using subg. Vicia taxa. However, the majority of attempts have replicated earlier crosses and so will not be discussed in detail. The first extensive attempt of hybridization experiments in Vicia are those of Ascherson & Graebner (1909), who report attempted crosses between: V. sativa x V. lutea and the successful production of a hybrid plant from V. sativa subsp. nigra x V. lutea. Schoth (1916-1919), in attempting crosses within the V. sativa aggregate, reports limited success with the production of sterile hybrids. Tupikova's attempts at crosses are reported by Sveshnikova (1927). She crossed: 1) V. sativa subsp. sativa x V. sativa subsp. macrocarpa; 2) V. faba x V. narbonensis; 3) V. serratifolia x

V. narbonensis. She obtained no hybrid material and concluded that the difference in karyotype of these six taxa gives little hope of success in future hybridizations. This has proved prophetic. The limited success that has occurred has been within groups that are taxonomically difficult to distinguish.

Sveshnikova (1940) attempted further crosses in the V. sativa aggregate, producing F1 which showed hybrid vigour. The F2 plants returned to the parental type or were etiolated, lethal or sterile. Several other authors working within the same complex have achieved similar success (Coutinho, 1940; Yamamoto, 1950, 1955, 1966, 1968a, 1968b, 1969, 1974, 1977, 1980; Watanabe & Yamada, 1956, 1958; Sekizuka et al., 1960; Moriya, 1961; Mettin, 1962; Mettin & Hanelt, 1964, 1973; Hanelt & Mettin, 1966; Plitmann, 1967; Yamamoto, 1971, 1974a, 1974b, 1980; Ladizinsky & Temkin, 1978, Zohary & Plitmann, 1979; Ladizinsky, 1981). All of these authors conclude that, the taxa of the V. sativa aggregate are very closely related and, to varying degrees, are interfertile.

Thompson & Barton (1924) report the hybridisation of V. sativa subsp. nigra x V. lathyroides on sand dune areas in Britain. The natural progeny resemble one or other parent (usually small V. sativa subsp. nigra), while artificially produced hybrids all died prior to flowering.

Donnelly & Clark (1962) attempted a broader hybridization programme, incorporating six subg. Vicia species, see Table 8.3. As shown the success was limited to the V. sativa subsp. sativa x V. sativa subsp. nigra cross. They obtained seventy-four F1 plants which showed hybrid vigour. The F2 plants showed varying degrees of fertility and vigour, although soft seed coats leading to seed rot and lethality were found. They conclude, though, that via selection of the F2 material, vigorous, fertile lines with a high percentage of hard seed could be obtained.

Table 8.3. Summary of the hybridization experiments of Donnelly & Clark (1962).

	Female parent						
SPECIES Pollen parent	1	2	3	4	5	6	7
1 *V. sativa* subsp. *sativa*	/	+	-	/	/	-	-
2 *V. sativa* subsp. *nigra*	/	/	/	/	/	/	/
3 *V. grandiflora*	/	-	-	-	-	-	-
4 *V. serratifolia*	/	-	-	/	/	/	/
5 *V. galeata*	/	/	-	/	/	/	/
6 *V. lutea*	/	-	-	-	/	/	-
7 *V. pannonica*	/	-	-	/	-	/	/

"+" indicates a successful cross, "-" indicates an unsuccessful cross and "/" cross not attempted.

The historical success in hybridising *V. sativa* subspecies points to possible success in the current attempts being made to introduce amphicarpic genes from subsp. *amphicarpa* to subsp. *sativa* (ICARDA, 1988). Mettin (1962) also attempts a broader coverage of *Vicia*, using *V. sativa*, *V. lutea*, *V. pannonica*, *V. faba* and *V. melanops*, but the only successful cross was obtained between *V. sativa* subsp. *amphicarpa* x subsp. *obovata* (syn. subsp. *sativa*). Outside the *V. sativa* complex, Sekizuka *et al*. (1960) record the production of a hybrid from a *V. sativa* x *V. grandiflora* cross (a closely allied species on morphological evidence). The hybrid plant showed intermediacy between the parents with some hybrid vigour, but pollen fertility and seed set was poor. Most of the earlier hybridization experiments focused on the *V. sativa* complex, probably because spontaneous natural hybrids are found, but in recent years attempts have been made between *V. faba* and its allies.

Schäfer (1973), in an elaboration of her previous collaborative research (Hanelt *et al*., 1972), undertook

extensive morphological, karyotypical and hybridization studies of V. faba and the Narbonensis complex. She made crosses between and within V. narbonensis (2n=14), V. serratifolia (2n=14), V. johannis (2n=14), V. galilaea (2n=14) and V. faba (2n=12). She found relative ease at crosses within a taxon, some success at crosses between subspecific taxa, little success with crosses between Narbonensis complex species and no success in crossing V. faba with Narbonensis complex taxa. Her varying levels of success in crossing between Narbonensis complex taxa came from crosses of V. narbonensis x V. serratifolia, V. serratifolia x V. johannis and V. johannis x V. narbonensis. None of these crosses resulted in viable hybrid plants.

Schäfer did not include V. hyaeniscyamus in her cytological or hybridization experiments, but states that it, "is presumably conspecific with V. galilaea", which has recently been refuted by Khattab et al. (1988). She concludes, on the basis of morphological, cytological, biochemical and hybridization experiments, that V. faba is clearly distinct from the Narbonensis complex species. Though both are likely to have evolved from the same Vicia progenitor, gene transfer between the two is now impossible.

Ladizinsky (1975a), when undertaking similar hybridization experiments, had slightly more success in crossing the Narbonensis complex species than Schäfer. Though he still found a complete barrier between crossing V. faba with Narbonensis complex taxa. He suggests that the differences in morphology, cytology and crossability all indicate that V. faba is very distinct from the other species in sect. Faba sensu Ball (1968). Ladizinsky suggests that V. faba is so distinct that it may warrant generic rank as Faba bona Medik., putting it on a similar taxonomic level with all the rest of Vicia. He points out that the obvious barrier to hybridisation between V. faba and the Narbonensis complex species is the difference in chromosome number and chromosome

morphology (V. faba has 2n=12, much larger, extremely asymmetric chromosomes containing twice as much DNA as the wild species).

The barrier between V. faba and V. narbonensis seems to be very basic. There is difficulty in getting the pollen to germinate and grow down the other species' style (Van Cruchten, 1974). Yamamoto is quoted in Cubero (1982) saying that in his experiments a few crosses have resulted in embryo formation, but none have survived. Cubero reports the use of tetraploid V. faba and chemical treatments to try to induce crosses, but with no success. He also suggests using V. melanops crosses with V. faba, because Chooi (1971a) showed that the former species had high levels of DNA, closer to that found in V. faba, than in other Vicia species. This appears to be a misconception. Chooi's results do not imply because both species have relatively high levels of DNA, that both species have high levels of the same component DNA. The two species are quite distinct morphologically and this is likely to be reflected in their DNA compositions even if both species have it in relatively large quantities.

A comprehensive hybridization project was initiated in 1982 at Reading University. The goal of this study was to synthesise hybrids between V. faba and other Vicia species (Pickersgill et al., 1983). It was felt that the more primitive subspecific taxa of the faba bean would be the most amenable to crossing. So subsp. paucijuga and subsp. faba vars. minor and equina were used in crosses with V. narbonensis, V. galilaea, V. johannis and V. bithynica. As reported by previous authors, the pollen tubes grew slowly and a few embryos were subsequently fertilised, especially if V. faba was used as the seed parent. Ramsay et al. (1985) discuss the progress in their hybridization project. V. faba pollinated by V. galilaea or V. johannis, or V. faba used as a pollen parent with V. bithynica, resulted in many ovaries with one or more fertilised ovules. All of these embryos abort in

their early stages. There is a correlation between decreasing cross viability and larger seed size in V. faba. This correlation indicates that the large seeded forms have diverged further from the common ancestral progenitor. Interestingly, Ramsay & Pickersgill (1986) found that hybrid tissue, produced from V. faba x V. melanops or V. lutea, developed further than crosses with sect. Faba species. This suggests that the DNA composition of the two different groups of species is relatively similar, so that in this case quantity of DNA is a better guide to post-fertilisation interspecific compatibility, than established morphological relationships. These results tend to refute the point made above; that just because both groups of species have relatively high levels of DNA, it is not necessarily the same component DNA.

Yamamoto (1984) reported very little success in his attempts to cross between the species of sect. Faba. Although he did manage to obtain F1 plants from a V. johannis x V. hyaeniscyamus cross. As discussed above, the V. hyaeniscyamus material used was possibly confused with V. johannis or V. galilaea. Yamamoto comments that this cross produced normal growth and fertility, which would appear to indicate that the material had indeed been mistakenly identified.

Roupakias & Tai (1986), attempting crosses between V. faba and V. narbonensis, found temperature control was an important factor in controlling embryo viability. They managed to retain pods of 20 to 50mm. on plants for twenty days post fertilisation. Roupakias (1986) located the fertility barrier in the embryo sac, as did Ramsay & Pickersgill (1986). He suggests that embryo culture techniques will have to be developed, if the desired hybrid plants are to be obtained.

None of these hybridization projects, between V. faba and its supposed allies, has succeeded in one of their primary

aims, the identification of the wild progenitor of the faba bean. In fact, the opposite has occurred. These studies have shown that the accepted wild allies of V. faba do not include the elusive wild 12-chromosome progenitor (Zohary, 1977). Zohary also comments that the floristic detail of the Faba group is still sketchy in certain areas of the Mediterranean Basin and cites as an example the recent discovery of two new species, V. hyaeniscyamus (Mouterde, 1962) and V. galilaea (Plitmann & Zohary, 1965). This point is underscored by the discovery of two further members of this complex by members of the Vicieae Project group at Southampton, V. kalakhensis (Khattab et al., 1988) and V. eristalioides (Maxted, 1989).

8.5 Anatomical Evidence

There have been no anatomical studies that specifically or comprehensively cover Vicia. The following account, abstracted from the literature, will concentrate on relating the anatomical studies that have been undertaken to subg. Vicia taxonomy. This greatly limits this account, as there has been very little systematic exploitation of anatomy in the studies of Vicia.

The first comprehensive review of Vicieae anatomy was undertaken by Streicher (1902). He provided an overview of the tribe and then described generic and representative specific patterns in more detail. He described 13 subg. Vicia species, but commented that he found few characters which distinguished taxonomic subgroups. The presence of extra-floral nectaries were characteristic of all subg. Vicia species. Usually these extra-floral nectaries are hair covered, but hairs were absent in V. oroboides, while in V. sepium the hairs are longer and thinner than normal.

There are very references to Vicia in the Metcalfe & Chalk (1950) review of dicotyledonous anatomy. They do record the presence of extra-floral nectaries on the stipules of some Vicia species (the key character for distinguishing subg. Vicia

species). They comment that these nectaries are composed of stalked glands mixed with clothing hairs. Shah & Kothari (1973), in a study of stomata and hair characters in relation to Vicieae taxonomy, include two Vicia species, V. faba and V. sativa. These species are shown to be distinct for both stomata and hair characters, but this does not provide help in clarifying subg. Vicia taxonomy. As they point out there are many different hair types in the Vicieae. This could be just as true in subg. Vicia. The research for this revision has shown that these hair types will provide important characters for the future, especially in sections Faba and Hypechusa.

Toma et al. (1973) undertook a morphological and anatomical investigation of Vicia stipules. They found that size and shape of stipules varied greatly between species. Of the six subg. Vicia species studied, two have comparatively large stipules, V. narbonensis and V. serratifolia. Within species, variation was limited, but stipule lateral extension was more varied, especially in V. sativa and V. sepium, and to a lesser extent in V. lathyroides. They note the presence of hairs on the margin of the stipule in some species, e.g. V. narbonensis and V. serratifolia. Their observations of stipule epidermal cell shape and stomatal index divided the species into three groups; 1. V. sepium, 2. V. narbonensis and V. serratifolia and 3. V. sativa subsp. sativa, subsp. nigra, V. lathyroides, and V. grandiflora. In transverse section they found very little difference between the species, except that both V. narbonensis and V. serratifolia had a distinct palisade layer.

The most comprehensive systematic coverage of Vicia anatomy, was undertaken by Kupicha (1975), who studied the vascular anatomy. She found that the Vicieae was characterised by an unusual type of stele in which the lateral leaf-traces are present as cortical bundles in the internode below the insertion of their leaf. In most Vicia a fresh pair of cortical bundles is substituted at each node, but in ten

species (indicated in Table 8.4) there is only partial replacement. This has obvious taxonomic implications, clearly separating out the section Hypechusa species, from the species of the other four sections of subg. Vicia investigated.

Table 8.4 Nodal anatomy of subg. Vicia species, taken from Kupicha (1975).

Complete nodal cortical replacement	Partial nodal cortical replacement
Atossa	Hypechusa
V. sepium	V. anatolica
V. truncatula	V. assyriaca
	V. galeata
Vicia	V. hybrida
V. sativa	V. hyrcanica
V. grandiflora	V. lutea
V. cuspidata	V. melanops
	V. noeana
Faba	V. pannonica
V. faba	V. sericocarpa
V. bithynica	
Peregrinae	
V. aintabensis	
V. michauxii	
V. peregrina	

8.6 Chemotaxonomic Evidence

Vicia is a popular genus for chemotaxonomic study. Chemotaxonomic evidence is plentiful and the group's chemistry is better known than that of many taxa (Kupicha, 1974). However, as with other sources of data, the coverage of individual researchers is not comprehensive. Chemotaxonomists often include one or two Vicia species in a broad survey of the legumes. Such studies yield little information about the internal structure of subg. Vicia. The following section is an attempt to abstract data for subg. Vicia.

Initially, chemotaxonomic studies in Vicia seem to have focused on canavanine distribution. Birdsong et al. (1960) investigated its presence throughout the Leguminosae. They included five species of subg. Vicia. Their results were as follows:

Section	Species	presence / absence of canavanine
Vicia		
	V. sativa	-
	V. grandiflora	-
	V. lathyroides	-
Faba		
	V. faba	+
Hypechusa		
	V. pannonica	-

With such a small sample from Vicia species, it would be wrong to draw any fundamental conclusions, other than that it is interesting that V. faba has a different response to the other species investigated. However, this difference was not substantiated by a more detailed study of amino acid distribution in Vicia by Bell and Tirimanna (1965). They record canavanine as being undetectable from all subg. Vicia taxa except V. sepium.

The results of Bell and Tirimanna's (1965) survey, which show variation within subg. Vicia, are detailed in Table 8.5. There is no obvious link between the pattern of their results and established classifications of the subgenus, although species from the same sections do have similar amino acid distribution patterns. Sections Atossa, Vicia and Peregrinae are linked by the presence of the majority of the compounds investigated. A loose linking is suggested between Faba and Hypechusa by the presence of arginine, asparagine and the unidentified compound V.A3 and the absence of the other compounds investigated.

Bell et al. (1978) review the literature for canavanine distribution in the Papilionoideae, as well as undertaking some novel tests. Their overall conclusion is that canavanine is largely absent from subg. Vicia, the exceptions being V. faba, V. sativa and V. sepium. The records for V. faba and V. sativa may be incorrect as they come from single reports and there are multiple reports where no canavanine was found. Similar conclusions about the biochemically distinct nature of V. sepium were made by Tschiersch & Hanelt (1967). They extended the investigation of canavanine and amino acid distribution in Vicia and found that canavanine was absent from sect. Vicia and sect. Faba. The other sections that contained canavanine could be split in two. V. sativa, V. grandiflora and V. sepium possesses ß-cyanoalanin and its y-glutamyl derivative. Other subg. Vicia taxa lack non-protein amino acids, but have higher concentrations of arginine, which is present in smaller concentrations in both groups.

Table 8.5 Ninhydrin-positive compounds present in seed extracts of subg. Vicia taxa, taken from Bell & Tirimanna (1965).

Arg = arginine; y-OH Arg = y-hydroxyarginine; Asp(NH2) = asparagine; ß-CN-Ala = ß-cyanoalanine; Glu-ß-CN-Ala = y-glutamyl-ß-cyanoalanine; PIP = pipecolic acid; V.A1 - V.B2 represent unidentified compounds; T = trace; + = conc. approx. 1% of dry weight; ++ & +++ are proportionally greater concentrations; - = presence not recorded.

Taxa	Arg	yOH Arg	Asp (NH2)	ßCN Ala	Glu-ß-CN-Ala	PIP	V.A1	V.A3	V.A4	V.B2
Atossa										
V. sepium	+	+	T	+	+	-	-	-	+	+
Vicia										
V. sativa										
amphicarpa	+	-	+	+	++	-	-	-	+	+
cordata	+	+	-	+	+	-	-	-	+	+
nigra	+	-	+	+	+	+	-	-	+	+
sativa	T	+	T	+	++	+	-	-	+	+
V. grandiflora	+	+	+	+	++	-	-	-	+	+
V. lathyroides	+	T	+	+	++	-	-	-	+	+
Faba										
V. faba	++	-	T	-	-	-	-	-	-	-
V. narbonensis	++	+	T	-	-	-	+	++	-	-
V. bithynica	++	-	T	-	-	-	+	++	-	-
Hypechusa										
V. hybrida	+	-	+	+	++	-	-	-	+	-
V. hyrcanica	+++	-	T	-	-	-	-	++	-	-
V. lutea	+	-	-	-	-	-	-	++	-	-
V. pannonica	++	-	-	-	-	-	+	++	-	-
Peregrinae										
V. michauxii	+	+	+	+	++	+	-	-	+	-
V. peregrina	+	+	+	+	++	-	-	-	+	+

Boulter et al. (1967) found that various Vicia species have very similar globulin alkaline gel electrophoresis patterns. Boulter and Derbyshire (1971) examined the isoenzyme patterns of several Vicia species, and also found very little variation in pattern. They found that the gel patterns varied as much within taxa as between them. However, Derbyshire & Boulter (referred to in Boulter, 1981) note that V. narbonensis is the exception and has a unique distinctive pattern.

To complement Ladizinsky's (1975a) study of morphology, cytology and hybridization in Vicia sect. Faba, Ladizinsky (1975b) undertook an electrophoretic study of seed proteins in the same group. It is likely that he misidentified the material he referred to as V. hyaeniscyamus (see Khattab et al. 1988), which should be remembered when interpreting the results. The results show that each of the four species included (V. faba, V. narbonensis, V. hyaeniscyamus and V. galilaea) produces a distinctive albumin profile. The intra-specific variation was negligible. V. faba profiles have 14 distinctive bands, five of which could be matched to bands in the wild species. Among the three wild species, V. hyaeniscyamus and V. galilaea are shown to be relatively closely related, with V. narbonensis more remote. Abdalla & Gunzel (1979), reporting a similar study of seed protein electrophoresis, found little intra-specific variation within V. faba accessions, but large differences between the banding pattern of V. faba and V. narbonensis.

In a disc electrophoretic study of the salt-soluble seed proteins, five subg. Vicia taxa were examined (V. sepium, V. sativa subsp. sativa, subsp. nigra, V. lathyroides and V. faba), Tali (1975). These results are summarised in the following table of relatedness:

1	V. sepium	100				
2	V. lathyroides	86	100			
3	V. sativa subsp. sativa	78	81	100		
4	V. sativa subsp. nigra	87	100	81	100	
5	V. faba	87	87	69	87	100
		1	2	3	4	5

This data reflects the accepted taxonomy. V. faba is the most distinct species, followed by V. sepium and then the Sativa aggregate taxa are closely related. The results fail to distinguish V. sativa subsp. nigra and V. lathyroides. In a similar study of nine subg. Vicia taxa, Perrino et al. (1977), also present results which reflect the established taxonomy. V. narbonensis, the only sect. Faba species included, is isolated. So are the four V. sativa subsp. sativa and one subsp. cordata. One Algerian V. sativa accession, referred to as V. sativa tipo algerina (MG 652) has a very weak banding pattern in comparison to the other V. sativa accessions. The other six species share a few bands, but are each distinct. V. peregrina is separable as it only shows one clear band . They suggest a broader survey of species is required, to add validity to their study and help further clarify Vicia relatedness.

Using leaf or cotyledon tissue, Robeson & Harborne (1980) tested the phytoalexin response of Vicieae species. They included thirteen species from subg. Vicia. The results are shown in Table 8.6. The results proved more useful at clarifying generic level taxonomy than within subg. Vicia. It would be incautious to draw taxonomic conclusions from the six "not detectable" results, especially as the not detectable results are spread through the three sections where more species were tested. They, however, do record trace quantities of medicarpin presence only in V. faba, which distinguished this species from the other subg. Vicia species.

Table 8.6. Phytoalexin response in subg. Vicia species, taken from Robeson & Harborne (1980).

Sect.	Species	Wyerone	Wyerone epoxide
Atossa			
	V. sepium	+	+
Vicia			
	V. sativa	+	+
	V. sativa	+	-
	V. grandiflora	+	+
	V. lathyroides	+	-
Faba			
	V. faba	+	+
	V. narbonensis	+	+
	V. bithynica	+	-
Hypechusa			
	V. hybrida	+	+
	V. hyrcanica	+	-
	V. lutea	+	-
	V. pannonica	+	-
Peregrinae			
	V. michauxii	+	+
	V. peregrina	+	+

+ = present, - = not detectable

Yamamoto & Plitmann (1980), using isozyme polymorphism techniques, undertook a broad survey of Vicia accessions. Their results for subg. Vicia taxa are summarised in Table 8.7. Isozyme band patterns were recorded for amylase, esterase, glutamate oxaloacetic transaminase (GOT) and indophenol oxidase (IPO). Fifteen taxa are included from four subg. Vicia sections sensu Kupicha (1976). The results of the banding are complex, but show no distinct sectional taxonomic patterns. The replication indicated that there is as much variation within species as there is between species and there are few bands which can be used to define taxa. The exceptions are the esterase band 3, which is only found in the three V. hybrida replicates and esterase band 7 which is only found in some V. sativa accessions. The GOT 2 band is found in nearly all the taxa except V. bithynica and V. sativa subsp. incisa, which have no obvious close taxonomic relationship. GOT 3 is only found in the two sect. Peregrinae species (V. michauxii and V. peregrina) and so may possibly be used as a marker for the section. GOT 4 is only present for the V. melanops accessions, which suggests it may be a marker for this species. In a more sectional application of these techniques to Vicia sect. Faba, Yamamoto et al. (1982), found a clear separation of the species into three groups, V. faba, the Narbonensis complex species and V. bithynica. The results of the banding patterns are shown in Table 8.8 and are summarised in the similarity index shown below. The first three letters of each specific name are used as identification codes:

Species	Nar	Ser	Gal	Joh	Hya	Fab
Ser	62.5					
Gal	51.5	40.6				
Joh	37.0	55.0	43.3			
Hya	28.6	34.8	27.3	61.1		
Fab	27.5	28.6	39.0	25.7	26.5	
Bit	8.7	17.6	12.0	12.5	21.4	22.7

Table 8.7. Isozyme polymorphism results for Subg. <u>Vicia</u>. Taken from Yamamoto & Plitmann, 1980.

Only those bands which show variation in subg. <u>Vicia</u> are included. + = presence.

Sect. Species	Amylase Alpha						Amylase Beta				Esterase														GOT						IPO			
	1	2	3	4	5	6	1	2	3	4	1	2	3	4	6	7	a	b	c	d	e	f	g	h	1	2	3	4	5	6	4	6	8	9
Vicia																																		
V. sativa				+					+					+	+				+				+			+			+			+		
V. sativa				+					+			+			+					+			+			+			+			+		
V. sativa				+					+					+	+				+	+			+		+				+			+		
V. sativa				+					+					+		+				+			+			+			+			+		
V. sativa				+					+					+		+			+				+			+			+			+		
V. sativa ssp. amphicarpa				+		+				+				+		+				+			+			+			+			+		
V. sativa ssp. cordata				+		+			+					+	+					+			+			+			+			+		
V. sativa ssp. incisa				+						+				+	+					+			+		+				+			+		
V. lathyroides				+						+				+	+					+			+			+			+					
Faba																																		
V. narbonensis				+					+					+		+					+		+			+			+				+	
V. narbonensis				+					+					+		+					+		+			+			+				+	
V. bithynica																																	+	
Hypechusa																																		
V. esdraelonensis			+						+				+	+			+			+	+		+			+				+				+
V. hybrida				+					+				+			+		+		+		+				+			+					+
V. hybrida				+					+			+	+			+		+		+			+			+			+					+
V. hybrida				+					+			+	+			+		+		+			+			+			+					+
V. hyrcanica		+										+		+		+	+	+		+						+				+	+			+
V. lutea				+						+				+		+			+				+			+			+				+	
V. melanops			+					+				+		+		+	+				+		+			+		+			+		+	
V. melanops		+					+					+		+		+	+				+		+			+		+			+		+	
V. melanops			+						+		+	+		+		+	+				+		+			+		+			+		+	
V. pannonica			+						+					+		+		+		+		+				+			+					+
Peregrinae																																		
V. michauxii				+					+			+				+		+		+			+			+	+				+		+	
V. michauxii				+					+					+		+	+		+		+		+			+	+				+			+
V. peregrina	+								+					+		+		+		+		+				+	+				+		+	
V. peregrina			+						+			+				+		+		+				+		+	+				+		+	
V. peregrina					+				+			+				+		+		+		+				+	+				+		+	

Table 8.8 Isozyme polymorphism for sect. Faba, taken from Yamamoto et al. 1982.

Three letter codes designate species, + = presence.

Isozyme system	Band No.	Nar	Ser	Gal	Joh	Hya	Fab	Bit
GOT	1			+				
	2			+				
	3	+	+		+	+		
	4	+	+		+	+		
	5						+	
	6						+	
	7			+	+	+		
	8						+	
	9	+	+	+	+	+	+	
Esterase	1	+	+					
	2						+	
	3	+	+	+	+		+	
	4		+	+	+	+	+	
	5	+				+	+	
	6			+			+	
	7			+			+	
	8			+			+	
IPO	1					+		
	2	+	+	+	+		+	
	3	+		+				
	4							+
	5	+		+				
	6	+	+	+	+	+	+	
	7	+	+	+	+	+	+	
	8							+
	9	+	+				+	
	10							+
	11						+	
	12						+	
	13			+	+	+	+	
	14			+			+	
Peroxidase	+10	+		+				
	+9	+	+	+				+
	+8		+	+	+	+	+	+
	+7	+		+				
	+6	+	+	+				
	+5						+	+
	+4	+	+	+	+			
	+3						+	+
	+2					+	+	+
	+1						+	
	-1	+		+			+	
	-2	+	+	+			+	
	-3	+		+	+	+		
	-4						+	
	-6	+	+	+			+	
	-7			+	+			

V. bithynica is the most distinct species, with only five peroxidase bands in common with the other species investigated. The next most peripheral species is V. faba, which is slightly more distinct, than the Narbonensis complex species are to each other. Within the complex, the species form two pairs of species, V. narbonensis and V. serratifolia, and V. johannis and V. hyaeniscyamus. V. galilaea remains relatively remote from all the other species examined. The close alliance of V. johannis and V. hyaeniscyamus may be misleading, as the material he used as V. hyaeniscyamus was possibly V. johannis. He concluded that the furthest genetic distance is between V. faba and V. bithynica, with the Narbonensis species lying between these two extremes. This conclusion was reiterated later by Yamamoto (1984) and Khattab (1987).

Wolff (1980) investigated seed albumins and esterases distribution using gel electrophoresis and found marked pattern divergence among the accessions studied. Even accessions of the same species (e.g. V. johannis) produce different patterns, which showed little correlation with established subspecific taxonomy. The exception to this pattern was V. faba, which showed a consistent pattern for all varieties tested. He concluded that seed albumin and esterase gel electrophoresis is of limited use in clarifying Vicia taxonomy. Ladizinsky & Waines (1982), when undertaking a similar survey of the V. sativa aggregate, drew similar conclusions from their ambiguous results. They suggest this variation in seed proteins within taxa of the aggregate may be an important agent for acquiring ecological flexibility and so enhances this pernicious weed's colonisation abilities. Further confirmation of this seed protein variation was provided by Maplestone et al. (1985), who found that qualitative variation in the legumin subunit distribution pattern was less within the species V. faba, than the variation between all the wild Vicia species investigated.

Babac (1981) built an experimental chemosystematic database for the tribe Vicieae. This involved collating data from the literature on Vicieae chemotaxonomy, which he supplemented with novel investigations. His results are summarised for subg. Vicia species in Tables 8.9 to 8.11, which include the results of amino acids, leaf and flower phenolic aglycones and leaf phenolic glycosides investigations respectively. The results, however, do not show any clear systematic pattern. There is little correlation with the sectional groupings suggested by Kupicha (1976). Caffeic acid is present in all three sect. Faba taxa investigated, as well as V. grandiflora, so this could possibly be a helpful marker of sect. Faba taxa. Many of the chemical substances identified are present for only one species and it is tempting to infer that these are species specific. However, this could be erroneous. Further experimentation is required to replication existing findings and fill in gaps in the existing coverage before firm conclusions can be drawn.

Babac's results were incorporated, with many of the chemotaxonomic data discussed above, into the Vicieae database and are published in Vicieae Database Project Publications numbers 3 (Adey et al., 1983a) and 4 (Adey et al., 1983b).

Table 8.9. Amino acids of subg. Vicia, taken from Babac (1981).

Sect.	Species	a	c	d	e	f	g	h	j	k	o	q	r
Atossa													
	V. oroboides	+	-	-	+	+	-	+	-	-	-	-	+
	V. sepium	+	+	+	-	-	+	-	+	+	-	+	+
Vicia													
	V. pyrenaica	+	+	+	-	-	+	+	-	-	+	-	+
	V. sativa	+	+	+	-	-	+	-	+	-	+	+	+
Faba													
	V. narbonensis	+	-	-	+	+	-	-	-	-	-	+	+
	V. bithynica	-	-	-	+	+	-	+	-	-	-	+	+

+ = present, - = not detected
a = 4-hydroxyarginine, c = 3-cyanoalanine, d = 4-glutamyl-3-cyanoalanine, e - g = unidentified acidic amino acids, h = unidentified neutral amino acid, j - k = unidentified basic amino acid, o = pipecolic acid, q = asparagine, r = arginine.

Table 8.10. Leaf and flower phenolic aglycones of subg. Vicia, taken from Babac (1981).

Sect.	Species	Qr	Km	U1	U2	U5	U6	U7	U8	Fr	Cf	Le
Atossa												
	V. oroboides	-	+	+	-	-	-	-	+	-	-	-
Vicia												
	V. sativa	-	+	-	-	-	-	-	-	+	-	-
	V. grandiflora	+	+	+	-	-	-	-	-	+	+	-
	V. lathyroides	+	+	-	-	-	-	-	-	+	-	-
Faba												
	V. galilaea	+	-	+	-	-	-	-	+	+	+	-
	V. johannis	+	+	+	-	-	-	-	-	+	+	-
	V. bithynica	-	-	+	-	-	-	-	-	+	+	-
Hypechusa												
	V. hybrida	+	+	+	+	-	-	-	-	+	-	-
	V. lutea	-	+	-	-	-	-	-	-	+	-	-
	V. noeana	-	+	+	-	-	+	-	-	+	-	+
Peregrinae												
	V. peregrina	+	+	+	-	+	-	-	-	+	-	-

+ = present, - = not detected
Qr = quercetin, Km = kaempferol, U1 - U8 = unidentified leaf phenolic aglycones, Fr = ferulic acid, Cf = caffeic acid, Le = leuco-anthrocyanidins.

Table 8.11. Leaf phenolic glycosides of subg. Vicia, taken from Babac (1981).

Sect. Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Atossa																
V. oroboides	-	-	-	-	-	-	-	-	+	-	-	-	-	-	-	+
V. sepium	+	+	-	-	-	-	+	+	+	+	+	+	-	-	-	-
Vicia																
V. sativa	-	-	-	-	-	-	-	+	-	-	-	+	+	-	-	-
V. grandiflora	-	+	+	+	+	-	-	-	+	-	+	+	-	+	-	-
V. lathyroides	-	-	-	-	-	-	-	+	+	-	-	-	-	-	-	-
Faba																
V. galilaea	+	-	+	-	-	+	+	+	+	+	-	+	-	-	-	-
V. johannis	+	-	-	-	-	-	-	+	-	+	-	+	+	-	-	-
V. bithynica	+	-	+	-	-	-	-	+	-	+	+	+	-	-	-	-
Hypechusa																
V. hybrida	-	+	-	-	+	-	+	-	+	-	+	+	+	+	+	-
V. noeana	-	-	-	-	-	-	-	-	-	-	-	+	+	-	-	-
Peregrinae																
V. peregrina	+	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-

Sect. Species	17	18	19	20	21	22	23	24	25	26	27	28
Atossa												
V. oroboides	-	-	-	-	-	-	-	-	-	-	-	-
V. sepium	+	-	-	-	-	+	-	-	-	-	-	-
Vicia												
V. sativa	-	-	-	-	-	+	-	-	-	-	-	-
V. grandiflora	-	-	-	-	-	-	-	+	-	-	-	-
V. lathyroides	+	-	-	-	-	-	+	-	+	-	+	+
Faba												
V. galilaea	-	-	-	-	-	-	-	-	-	-	-	-
V. johannis	-	+	-	-	+	-	-	-	-	-	-	-
V. bithynica	-	-	+	+	+	-	-	-	-	+	-	-
Hypechusa												
V. hybrida	-	+	-	-	-	-	-	-	-	-	-	-
V. noeana	+	-	-	-	-	-	-	+	-	-	-	-
Peregrinae												
V. peregrina	-	-	-	-	-	-	-	-	-	-	-	-

+ = present, - = not detected
1 - 28 = unidentified leaf phenolic aglycones.

Khattab (1987), in a chemotaxonomic survey of sect. Faba, reported the presence of non-protein amino acids and phenolic compounds. The results of the investigation of non-protein amino acids showed that homoserine was always present in Narbonensis complex species and only occasionally present in accessions of both V. faba and V. bithynica. The occurrence of free aspartic acid showed no overall pattern within the section, but could be used to distinguish the two subspecies of V. faba, and V. hyaeniscyamus and V. kalakhensis from the other Narbonensis complex species. The results of his phenolic analysis indicate presence of quercetin 3,7-0 diglycosides linking V. faba and V. narbonensis. The latter species is considered by many authors (Linnaeus, 1753; Visiani, 1842; Taubert, 1894; Ball, 1968, and Tzvelev, 1980) the closest ally of V. faba. Quercetin was found to be present in V. faba subsp. faba and absent from subsp. paucijuga. Kaempferol 3,7-0 diglycoside was present in V. faba, distinguishing it from the V. narbonensis accessions that were investigated. He found that the V. narbonensis var. salmonea accessions examined appeared distinct from other V. narbonensis varieties. Only this variety showed the presence of an unidentified amino acid and the phenolic compound, myricetin 3-0 monoglycoside, which may indicate it is more distinct than was evident from morphological studies. He concluded that V. serratifolia is distinct from V. narbonensis, the former lacking many substances (protein and non-protein amino acids, and all flavonol and flavone compounds) found in the latter.

In a review of ultraviolet patterning and ultraviolet-absorbing pigments in legume flowers, Kay (1987) points out that there is an extensive range of visible colours and petal patterning, but UV patterns can either be absent or tend to reflect visible patterns. He found the most striking UV pattern in V. grandiflora, the visible colour of which is yellow with no patterning. Both V. hybrida and V. lutea, of sect. Hypechusa, have yellow flowers with moderate UV

reflection from the standard and deep UV reflection from the wings. Kugler (1963) briefly refers to Vicia, saying that V. sepium has a UV pattern that does not match its visible colour pattern. Like much of the phytochemical work described above, more detailed research is required before taxonomic conclusions can be drawn from flower colour patterns.

8.7 Other Non-morphological Evidence

This chapter section contains discussion of the research results that would not fit under the previous headings. All the types of taxonomic data discussed in this section are under-exploited and little taxonomic advantage has been taken of the possibility of obtaining data from these sources.

The pollen morphology of the Papilionoideae has not received as much attention as that of the Caesalpinioideae and Mimosoideae (Ferguson and Skvarla, 1981). Very little attention has been paid to examining Vicieae pollen and there has been no systematic coverage of any of the five genera included in the Vicieae. Gapotchka (1974), in attempting to find phylogenetic patterns within the temperate Fabaceae examined 26 Vicieae and 5 Vicia species (including V. sepium, V. grandiflora and V. faba). She found that each of the Vicia species examined was identical for the characters she recorded. Clark & Kupicha (1976) undertook a more detailed systematic survey of Vicieae pollen in an attempt to establish if the evidence from Cicer pollen supported its retention in or separation from the Vicieae group of genera. Vicia pollen was found to be very homogeneous. They provide a detailed description of Vicia pollen, but make no observations on intra-generic variation. Terziiski (1977) examining the pollen of two Vicia species, V. sativa and V. dumetorum found very similar sporodermal ornamentation and wall stratification for both these two taxonomically distant species. As all three studies detailed above indicate Vicia pollen to be uniform, the use of palynological evidence to infer relationships appears limited. However, until the genus is

investigated in a detailed systematic manner, no firm conclusions can be drawn.

The first seedling studies were carried out by De Candolle (1825b). These were expanded upon and added to by Compton (1912), but both these workers covered all the Leguminosae and so their coverage of Vicia is very limited. Nozzolillo (1977) made the only attempt to collect comprehensive seedling evidence for Vicia. She examined 58 species and found differences between species in leaflet shape, number of leaflets per eophyll (seedling leaf) and nodal position of the first eophyll. She recorded a large amount of intra-specific variation for seedling characters. She provides a key to the species examined, though it is possible due to this variation, that one species may key-out to several of her species groups. She comments that species that are related on adult morphological characters will also be related using seedling characters. This is, however, not borne out by her results. At least, as regards subg. Vicia, groupings are produced which are unlike any produced using other forms of evidence, e.g. V. bithynica is allied to V. biennis and V. sativa, and V. galeata is allied to V. neglecta and V. nigricans.

8.8 Literature Based Morphological Evidence

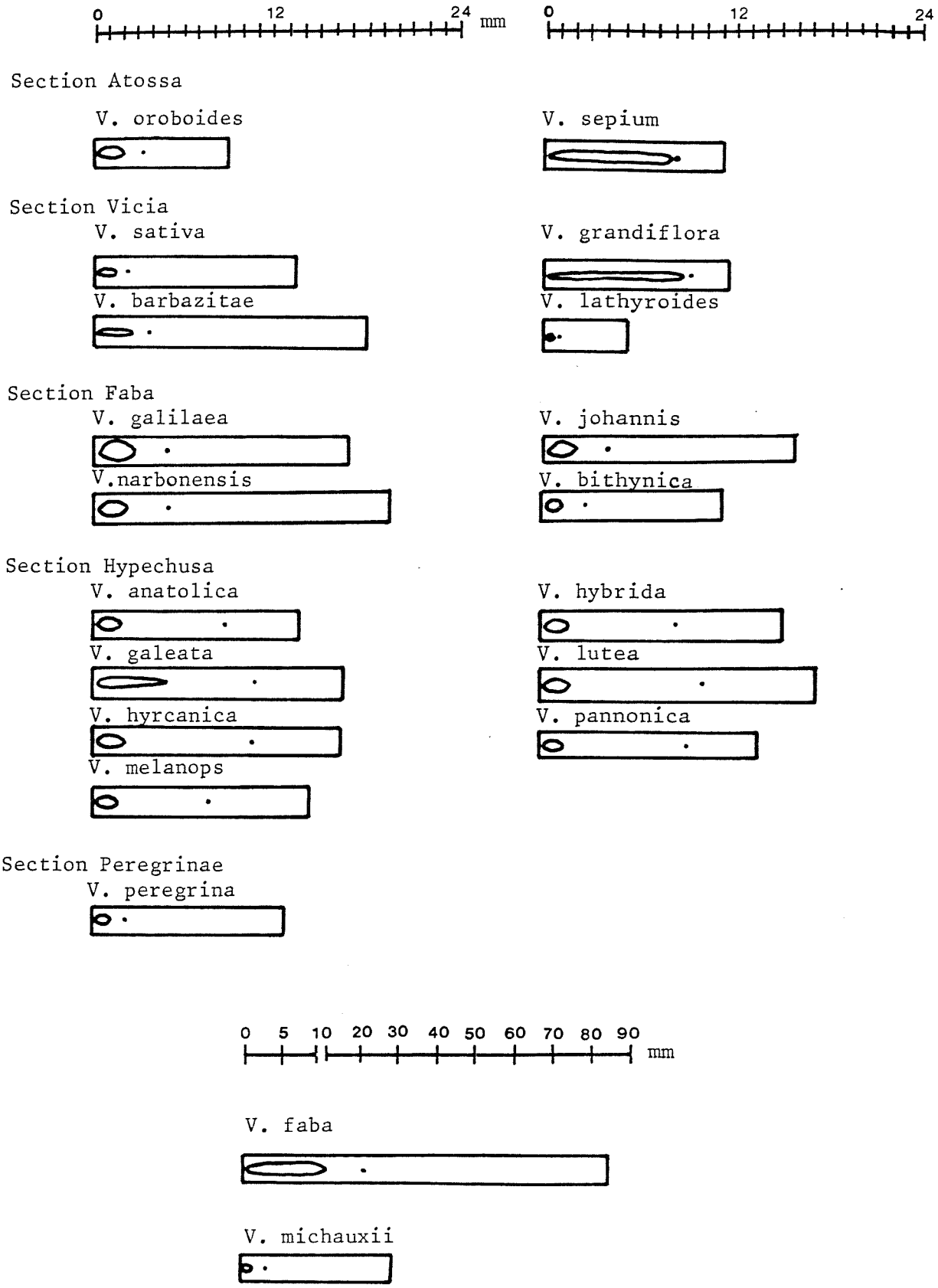
It is rarely possible for vegetative, inflorescence, floral, legume and seed characters to be scored from one herbarium specimen. The majority of herbarium specimens are made from flowering material, with few specimens having legumes and even fewer specimens with seed samples. So the results of the phenetic analysis reported in Chapter Seven, although including legume and seed characters, under-represent legume and seed data. To compensate for this general problem in phenetic studies of Vicia, several authors have specifically investigated Vicia legume and seed characters and then produced special purpose classifications and identification aids based on the results of their studies.

The first systematic attempt to identify Vicia species using seed characters alone was made by Swederski (1924). He examined 52 species of Vicia and found, by recording several seed measurements, that he could distinguish clusters of species and individual species. The most consistent character for identifying Vicia species seed was the relative length of hilum to circumference lengths. This character has been used extensively to identify species, but appears to have no general grouping significance, as groups of species with relatively long hila are distributed throughout Vicia. Relatively long hila are found in V. sepium, V. grandiflora and V. qatmensis. This early work was supplemented by Kostrakievicz (1951) and later by Zertova (1962). Both authors concurred with Swederski that the relative lengths of the seed hilum and circumference proved the most useful character for distinguishing individual species. Zertova provides detailed drawings and a key to the seeds of 18 Vicia (7 subg. Vicia) species.

In a study of the seeds of weedy vetches of European USSR, Leokene (1966) provides a key and description to 24 Vicia species. He expands the seed character set to include: seed shape; relative proportions; hilum colour; hilum trace colour; lens and micropyle position; hundred grain weight; diameter; testa colour and testa surface. However, like Gunn (1970 and 1971), the central objective was to provide a seed identification tool, via artificial keys, and not to attempt to relate seed characteristics to intra-generic groupings. In a similar study of the vetches of the Soviet Union, Voronchikhin (1981) provides a detailed key with diagrams, which uses legume and seed characters to distinguish taxa.

Gunn (1970) presents schematic diagrams for Vicia species. The subg. Vicia diagrams are presented in Figure 8.2. It is interesting to note that the characters used by Gunn are unable to distinguish individual species. Further more, the clusters of species produced do not show a clear

Figure 8.2. Schematic seed diagrams. Taken from Gunn 1970.



congruence with established taxonomic groups produced using broader character sets (an example of character set incongruence, see Chapter Six).

Lersten & Gunn (1982), in a survey of the Vicieae and their tribal relatives, attempt a more comprehensive, systematic approach in their study of external and internal seed topography. Unfortunately, for the purpose of this chapter, they concentrate their systematic observations above the subgeneric level. However, they comment that within Vicia there is little deviation from the standard type they describe. This is reflected in their data for subg. Vicia, which are summarised in Table 8.12. It is difficult to pick out groupings from these data, but distinctive species can be identified, such as V. grandiflora and V. faba.

Perrino et al. (1984) undertook a phenetic study of Vicia species using seed characters. They scored 52 species for 19 seed characters and found that seed circumference, relative hilum length, hilum shape and lens position were the most useful characters for distinguishing species. Cluster analysis of their seed data set grouped the Vicia species into eleven groups. Subg. Vicia species were dispersed throughout the groups. The only established taxonomic group retained in the results was sect. Hypechusa. The latter section is distinguished by Kupicha (1976) on the basis of relative lens position, so it is not surprising this group remains distinct. Perrino et al. summarise their results in the schematic form used by Gunn (1970) and their data are also incorporated into Table 8.12.

Table 8.12. Testa topography characteristics. Data taken from Lersten & Gunn (1982) and Perrino et al. (1984).

Taxa	Nr.Hilum	Midseed	Tracheid	Comment
Atossa				
V. sepium	Papillose	Papillose	-	
Vicia				
V. grandiflora	Papillose	Papillose	-	Rim aril associated with hilum
V. lathyroides	Papillose	Papillose	-	Surface rarely pusticular
V. sativa ssp. amphicarpa	Papillose	Papillose	Nonvestured to slightly vestured	
V. sativa ssp. sativa	Papillose	Papillose	Nonvestured to slightly warty	Surface rarely velvety
Faba				
V. bithynica	Papillose	Papillose	Nonvestured to slightly warty	
V. faba	Faintly papillose	Non papillose	Nonvestured to slightly vestured	
Hypechusa				
V. galeata	Papillose	Papillose	Nonvestured	
V. lutea	Papillose	Papillose	Nonvestured	Surface rarely velvety
V. melanops	Papillose	Papillose	Nonvestured to slightly warty	
V. pannonica	Papillose	Papillose	-	
Peregrinae				
V. michauxii	Papillose	Papillose	Nonvestured	

8.9 Summary of Literature-Based Taxonomic Evidence

Despite so much research being focused on Vicia, there is relatively little data that either correspond to any of the established patterns of taxon relationships or consistently indicate new taxon arrangements. What follows is a brief summary of the data that indicate a consistent set of taxon relationships.

The distribution of subg. Vicia sensu Kupicha (1976) centres on the Middle East. The taxa are especially plentiful in Turkey, Lebanon, Syria, Soviet Asia, Iran and Iraq. Two sections are largely restricted to this area, Hypechusa and Peregrinae. The other three sections, Faba, Vicia and Atossa are more geographically dispersed, covering Europe, North Africa and the latter section spreads across to the extreme North-east of Asia. It is interesting to note that the section with the greatest degree of internal variation (sect. Atossa) is the one that has disperse furthest from the centre of diversity of the subgenus.

The cytological evidence provides some clear patterns. The diploid number is 2n=14, but the aneuploids 2n=10 and 12 are common and are easily linked to established taxonomic groupings in sections Vicia and Hypechusa. Within the latter, three species are found with 2n = 10, V. anatolica, V. ciliatula and V. melanops, these species are also close morphological allies. As are the three species with counts of either 2n = 12 or 14, V.assyriaca, V. galeata and V. hyrcanica. The other close morphological ally of this group, V. noeana, has a count of 2n = 12 alone. V. lutea stands out in section Hypechusa, both morphologically, as it is the species with the shortest peduncle and cytologically, as it is the only species to have a consistent diploid count of 2n = 14. These points suggest this species may be more distinct than previously appreciated. V. faba is unique within section Faba with a count of 2n = 12, the other species are each 2n =

14. This supports the other sources of evidence, in placing V. faba further away from the rest of section Faba.

Chromosomal morphological studies have been restricted to sections Vicia and Faba. They have proved inconclusive as there was such a large degree of intra-taxon variability, especially within the V. sativa complex of species. There is some interesting evidence from chromosomal idiograms that suggests that V. bithynica is more closely related to section Vicia species than to section Faba, the latter alliance having been favoured on morphological grounds. These findings are further supported by chromosome banding studies. Chromosome banding patterns suggests alliances between V. faba and V. narbonensis, and V. sericocarpa and V. mollis. The latter two species are considered members of different sections by Kupicha (1976), but are shown in the previous chapter to be closely allied. Banding patterns indicate a close relationship between three section Peregrinae species, V. peregrina, V. aintabensis and V. michauxii. Several cytological studies of section Faba, reiterate the suggestion that the section can be subdivide into three subsections: V. faba; V. narbonensis, V. serratifolia, V. galilaea, V. johannis and V. hyaeniscyamus; and V. bithynica.

Several authors have investigated absolute DNA quantity per cell, but the results are difficult to interpret, as two species with similar quantities of DNA do not necessarily have similar quantities of the same DNA. The results do not reflect any established morphological grouping or point to distinct novel groupings. However, V. faba and V. sativa are shown to be distinctive in having much more and much less DNA than other subg. Vicia species respectively.

Numerous hybridization experiments have been undertaken between subg. Vicia taxa, but they have focused on sections Faba and Vicia. The experimental results may be summarised as, only producing viable hybrids between complexes of species

or subspecies which are difficult to distinguish clearly on morphological characters, e.g. the V. sativa and to a lesser extent the V. narbonensis complex taxa.

Relevant anatomical studies are of limited help within subg. Vicia, as so few studies have included enough subg. Vicia species to allow taxonomic conclusions to be drawn. However, at a broader level the presence of stipular nectaries and the associated anatomical features clearly separate the subg. Vicia taxa from other Vicia species. Stipule epidermal cell shape and stomatal index divide the species investigated into three groups; 1. V. sepium, 2. V. narbonensis and V. serratifolia and 3. V. sativa subsp. sativa, subsp. nigra, V. lathyroides, and V. grandiflora. These groupings correlate with sectional grouping (sections Atossa, Faba and Vicia respectively), but more detailed study of greater numbers of species is required. In most Vicia species a fresh pair of cortical bundles is substituted at each node, but section Hypechusa species are different and there is only partial replacement. This clearly separates the latter section from other subgenus species.

There has been extensive phytochemical investigation of Vicia. Within subg. Vicia, the results of amino acid investigations tend to link species of the same sections, although in sections Faba and Hypechusa the links are less strong and there is more internal variation. The results of canavanine distribution suggest it is generally absent from subg. Vicia, although it is found in V. sepium and occasionally in V. sativa. Section Atossa characteristically possesses ß-cyanoalanin and it s y-glutamyl derivative. Other subgenus taxa lack uncommon free amino acids, but have higher concentrations of arginine.

The results of protein gel electrophoresis are confusing. Some authors have found clear taxon-based patterns, while others have found greater variation within taxa than between

them. The profile of V. narbonensis, however, appears to be particularly distinctive. Disc electrophoretic studies of proteins have produced results which support the existing taxonomy, but a greater range of taxa and a larger replication of accessions is required, if the technique is to produce more conclusive results. The results do, however, support the division of section Faba into three subgroups, V. faba, the Narbonensis complex and V. bithynica. Both protein and non-protein amino acid studies suggest that V. narbonensis is distinct from its close ally V. serratifolia.

Palynological investigations suggest all Vicia pollen is uniform. Studies of seedling material showed that there is a large amount of intra-taxon variation and it is difficult to use the recorded data to identify taxa accurately. Relative proportions of the subg. Vicia seeds, and seed hilum to circumference lengths, have proved useful in identifying species, e.g. the relatively long hilum length compared to seed circumference is characteristic of V. grandiflora. The only seed character that can be consistently used to distinguish a group of Vicia species is position of the lens relative to the position of the hilum. The lens is always opposite the hilum in section Hypechusa species alone.

CHAPTER NINE
TAXONOMIC DISCUSSION

9.1 Introduction

The word 'conclusion' has been deliberately avoided, as this final taxonomic chapter is concluding, but not conclusive. This is the closing chapter of the taxonomic element of this thesis, but it will not be the closing chapter of Vicia subgenus Vicia taxonomic research. In the future, as more monographic databases become established, it is likely that a taxonomist will start and finish a revision with a monographic database. The revision process will involve the use and subsequent editing of this monographic database, in the light of the taxonomist's research. So the dynamic nature of the revision process will be evident, which makes the word conclusion inappropriate.

A taxonomic study, of this kind, does produce answers to particular taxonomic problems and these are, hopefully, closer to the intrinsic natural truth of the taxon. These answers are summarised in the form of primary and secondary revision products, as discussed in Chapter Three. The primary product is the proposed classification and the secondary products are the taxon descriptions, keys, etc. The aim of this chapter is to propose and discuss the revision products for subgenus Vicia.

9.2 Proposed Classifications of Vicia Subgenus Vicia

The object of presenting a novel classification should be simply to suggest a better replacement, that is a better approximation of the abstract 'natural' classification. The classification proposed is an attempt to communicate the relationship between the taxa of subg. Vicia as accurately as possible given the currently available evidence.

The proposed classification of Vicia subgenus Vicia is presented in Table 9.1.

Table 9.1. Proposed classification of Vicia subgenus Vicia

Subgenus Vicia L.

Section	Series	Species	Infra-specific taxon
I Atossa (Alef.) Asch. & Graebner			
	A Pseudovicilla Maxted ser. nov.		
		1 V. oroboides Wulfen in Jacq.	
	B Truncatulae (B. Fedtsch. ex Radzhi) Maxted stat. nov.		
		2 V. balansae Boiss.	
		3 V. abbreviata Fischer ex Sprengel	
	C Atossa		
		4 V. sepium L.	
			i var. sepium
			ii var. eriocalyx Celak.
			iii var. montana Koch
II Microcarinae Maxted Sect. Nov.			
		5 V. dionysiensis Mout.	
III Hypechusa (Alef.) Asch. & Graebner			
	A Hyrcanicae B. Fedtsch. ex Radzhi		
		6 V. assyriaca Boiss.	
		7 V. esdraelonensis Warb. & Eig	
		8 V. tigridis Mout.	
		9 V. galeata Boiss.	
		10 V. hyrcanica Fischer & C. Meyer	
		11 V. noeana (Reuter in Boiss.) Boiss.	
			i subsp. megalodonta Rech. fil.
			ii subsp. noeana
	B Hypechusa		
		12 V. melanops Sibth. & Smith	
			i var. melanops
			ii var. loiseaui Alleiz.
		13 V. ciliatula Lipsky	
		14 V. anatolica Turrill	
		15 V. mollis Boiss. & Hausskn. ex Boiss.	
		16 V. pannonica Crantz	
			i subsp. striata (M. Bieb.) Nyman
			ii subsp. pannonica
		17 V. hybrida L.	
		18 V. sericocarpa Fenzl	
		19 V. lutea L.	
			i subsp. lutea
			ii subsp. vestita (Boiss.) Rouy
IV Peregrinae (B. Fedtsch. ex Radzhi) Maxted stat. nov.			
		20 V. michauxii Sprengel	
		21 V. aintabensis Boiss. & Hausskn. ex Boiss.	
		22 V. peregrina L.	
V Wiggersia (Alef.) Maxted stat. nov.			
		23 V. cuspidata Boiss.	
		24 V. lathyroides L.	

Section	Series	Species	Infra-specific taxon
VI Vicia			
	A Vicia		
		25 V. pyrenaica Pourret	
		26 V. sativa L.	
			i subsp. nigra (L.) Ehrh.
			ii subsp. amphicarpa (L.) Batt.
			iii subsp. incisa (M. Bieb.) Arcang.
			iv subsp. devia J.G. Costa
			v subsp. sativa
			vi subsp. macrocarpa (Moris) Arcang.
		27 V. barbazitae Ten. & Guss.	
			i var. barbazitae
			ii var. incisa (Orph.) Boiss.
	B Grandiflorae B. Fedtsch. ex Radzhi		
		28 V. qatmensis Gomb.	
		29 V. grandiflora Scop.	
			i var. grandiflora
			ii var. incisa Braun & Bouche
VII Bithynicae (B. Fedtsch. ex Radzhi) Maxted stat. nov.			
		30 V. bithynica (L.) L.	
VIII Narbonensis (B. Fedtsch. ex Radzhi) Maxted stat. nov.			
	A Rhombocarpae Maxted ser. nov.		
		31 V. eristalioides Maxted	
	B Narbonensis (B. Fedtsch. ex Radzhi) Maxted Stat. Nov.		
		32 V. kalakhensis Khattab, Maxted & Bisby	
		33 V. johannis Tamamschjan in Karyagin	
			i var. ecirrhosa (Popov) H. Schäfer
			ii var. procumbens H. Schäfer
			iii var. johannis
		34 V. galilaea Plitm. & Zoh. in Plitm.	
			i var. galilaea
			ii var. faboidea (Plitm. & Zoh. in Plitm.) H. Schäfer
		35 V. serratifolia Jacq.	
		36 V. narbonensis L.	
			i var. salmonea (Mout.) H. Schäfer
			ii var. jordanica H. Schäfer
			iii var. affinis Kornhuber ex Asch. & Schweinf.
			iv var. aegyptiaca Kornhuber ex Asch. & Schweinf.
			v var. narbonensis
		37 V. hyaeniscyamus Mout.	
IX Faba (Miller) Ledeb.			
		38 V. faba L.	
			i subsp. paucijuga Murat.
			ii subsp. faba
			ii/a var. minor Beck
			ii/b var. equina Pers.
			ii/c var. faba

9.3 Discussion of Proposed Classification

The proposed classification of Vicia subgenus Vicia will be compared to the classification proposed by Kupicha (1976), which was falsified during this revision. Other recent classifications of subgenus Vicia, most notably those undertaken by Russian botanists (Stankevich, 1970, 1982; Radzhi, 1971; Nikiforova, 1985), have attempted more detailed classifications of Vicia than Kupicha. None of these, however, encompassed Vicia worldwide and so are not as comprehensive as Kupicha's classification. The proposed classification and Kupicha's are summarised in Table 9.2. To avoid duplication, the broad outlines of the classification will be discussed here, but the more detailed discussion of the infra-specific taxa will be restricted to the conspectus (Appendix 5).

It is difficult to compare directly Kupicha's and the proposed classification as the latter includes several species unknown to Kupicha in 1974. Four of these species easily fit into Kupicha's conception, V. kalakhensis and V. eristalioides in her sect. Faba, V. tigridis in sect. Hypechusa and V. qatmensis in sect. Vicia. However, the fifth species V. dionysiensis, an endemic of the Jebel Druse, South West Syria, has the diagnostic stipular nectary of subg. Vicia, but does not fit within the circumscription of any of her five sections. So the inclusion of this species in subg. Vicia necessitates the erection of a monospecific section, Microcarinae Maxted Sect. Nov. This species is rare and is distinguished from other subg. Vicia species by having a keel markedly shorter than the wings and standard. It is stout and erect like V. abbreviata of sect. Atossa, but it has the general facies of a sect. Hypechusa species, like V. assyriaca. The latter species is its closest ally in the phenetic analysis. Thus this new section is placed in the classification between Atossa and Hypechusa.

Table 9.2. Classifications of Vicia subgenus Vicia

Proposed	Kupicha (1976)
Atossa	Atossa
V. oroboides	V. oroboides
V. balansae	V. sepium
V. abbreviata (syn. V. truncatula)	V. balansae
V. sepium	V. truncatula
Microcarinae	Vicia
V. dionysiensis	V. pyrenaica
Hypechusa	V. sativa
V. assyriaca	V. grandiflora
V. esdraelonensis	V. barbazitae
V. tigridis	V. lathyroides
V. galeata	V. cuspidata
V. hyrcanica	Faba
V. noeana	V. faba
V. melanops	V. narbonensis
V. ciliatula	V. galilaea
V. anatolica	V. hyaeniscyamus
V. mollis	V. johannis
V. pannonica	V. bithynica
V. hybrida	Hypechusa
V. sericocarpa	V. anatolica
V. lutea	V. assyriaca
Peregrinae	V. ciliatula
V. michauxii	V. esdraelonensis
V. aintabensis	V. galeata
V. peregrina	V. hybrida
Wiggersia	V. hyrcanica
V. cuspidata	V. lutea
V. lathyroides	V. melanops
Vicia	V. noeana
V. pyrenaica	V. pannonica
V. sativa	V. sericocarpa
V. barbazitae	Peregrinae
V. qatmensis	V. aintabensis
V. grandiflora	V. michauxii
Bithynicae	V. mollis
V. bithynica	V. peregrina
Narbonensis	Additional species unknown to Kupicha
V. eristalioides	V. dionysiensis
V. kalakhensis	V. eristalioides
V. johannis	V. kalakhensis
V. galilaea	V. qatmensis
V. serratifolia	V. tigridis
V. narbonensis	
V. hyaeniscyamus	
Faba	
V. faba	

Subgenus Vicia in the proposed classification contains nine sections, nine series, 38 species, 14 subspecies and 22 varieties, whereas Kupicha's classification divides the subgenus into five sections and 32 species. Many of the groupings identified by the phenetic analysis discussed in Chapter Seven have been established by earlier authors. It is interesting to note that similar detailed groups (if not given the same taxonomic status) have been identified by several recent Russian studies building on the initial work of Fedtschenko (1948).

Kupicha warns against the production of too fine a classification:

> "If, however, only convincingly natural assemblages of species (i.e. groups which share several traits) are accepted as sections, a too finely divided system is produced."

She does not use any subsectional taxa, although she does add that there are possible groupings within her sections. She comments:

> "No subsections are recognised, but sects. Vicilla, Cracca, Atossa, Vicia, Faba and Hypechusa could probably all be further subdivided."

The proposed classification acts on these comments and while attempting to reflect the taxa's natural relationships, it is not so fine that it loses its predictive value. Interestingly, the sections sensu Kupicha that are split into further sections and series are those she suggests: Atossa, Vicia, Faba and Hypechusa. Interestingly, the latter three sections each contain groups of closely related forms (e.g. the V. sativa agg., the V. narbonensis complex and the V. noeana complex), which suggests that they are currently undergoing a process of rapid speciation.

There is a problem in trying to represent the 'natural' relationship of taxa in the form of a list of taxa. This

problem of the 'squeezing and bending' of taxa to fit a list is faced by all taxonomists and here the job is no less brutal and has inevitably led to a loss of information. The order of the sections in Kupicha's classification goes from the primitive, erect, many flowered, shaded habitat species of sect. Atossa to the single flowered, adventitious weeds of sect. Peregrinae. In the proposed classification, the order of sections is markedly different to that proposed by Kupicha (1976). The starting point is sect. Atossa, but the few flowered, adventitious weeds of sect. Hypechusa, Peregrinae and Vicia are passed through to culminate in the more robust, crop-like forms of the V. faba and the V. narbonensis complex.

This proposed sectional order is directly derived from the results of the analysis discussed in thesis section 7.3, where the V. narbonensis complex and V. faba were consistently shown to be the most distinct groups within the subgenus. If a listing of taxa is required and sect. Atossa, the closest section to subg. Vicilla, is the starting point, then V. faba must logically come at the end of the list. The intervening sections between these two points are ordered in the manner suggested by the phenetic analysis results. However, this order of taxa cannot, for instance, reflect the similarity between V. sepium and V. qatmensis, both of which have laterally flattened legumes and seeds with hila more than half the seed circumference length.

The proposed classification retains sect. Atossa as delimited by Kupicha, although the arrangement of the species is altered. The section was originally derived from the Alefeld (1861a) genus Atossa. In the classification the correct name V. abbreviata is used for what Kupicha (1976) knew as V. truncatula. The priority of V. abbreviata over V. truncatula was established by Stankevich (1988) and is discussed in detail in the conspectus.

Section Atossa is internally heterogenous and the four species naturally fall into three series, with the closely allied V. balansae and V. abbreviata forming one series. Stankevich (1988) suggests that the latter two species are insufficiently distinct to warrant separate specific status and refers to hybrid forms. She has shown me these specimens (Busch & Busch 8.7.1936 and Busch & Busch 19.7.1929, both from VIR) and these two specimens do show some intermediate characteristics. However, the results of the analysis (see Figure 7.15) and field observations indicate that the majority of specimens can be easily attributed to one or other species and so the existence of two species is retained.

V. oroboides with its erect, oroboid form with broad leaflets is distinct within subg. Vicia. This species is the subg. Vicia species most closely allied to subg. Vicilla, although the phenetic resemblance may be the result of convergence (see discussion in the Conspectus). There is a remote alliance between the two robust, many flowered species V. oroboides and V. balansae, as well as between the two less robust species, V. abbreviata and V. sepium and so the latter two species are placed closer to the other subg. Vicia species.

The conception of sect. Hypechusa is basically the same as Kupicha's, and is derived from the genus Hypechusa erected by Alefeld (1860a). V. mollis was considered by Kupicha to belong to sect. Peregrinae, the grouping of V. mollis with V. peregrina and its allies being originally suggested by Boissier (1872). However, the analysis results consistently indicate a more natural affinity for sect. Hypechusa, to which it is transferred. This transfer of V. mollis to sect. Hypechusa sustains the view taken by Townsend (1967).

The sect. Hypechusa taxa are split into two series, Hyrcanicae and Hypechusa on the basis of peduncle length, corolla shape and size and standard pubescence. This division

of the species into two major subgroups is clearly indicated by the detailed analysis of sect. Hypechusa sensu Kupicha (1976), discussed in thesis section 7.4.3. However, the two series formed by this split are heterogenous. V. assyriaca is, for instance, peripheral to ser. Hyrcanicae and the series is more typically represented by species of the V. noeana complex.

Within ser. Hypechusa the species could be subdivided into several groupings: the species which have a spot on the apex of the wing, V. anatolica, V. ciliatula, V. melanops and V. mollis; the species with pubescent standards, V. anatolica, V. hybrida and V. pannonica; with V. lutea and V. sericocarpa remaining isolated. Fedtschenko (1948), for instance, splits the sect. Hypechusa sensu Kupicha species into three series and Plitmann (1967) splits the section into four series. This degree of subdivision of this group would lead to an excessively subdivided section. As stated above, it has been decided to adopt a central stance, not using such a fine classification that it loses its predictive value. Hence only the two major subsectional groupings are recognised.

As with sect. Hypechusa the conception of sect. Peregrinae is essentially the one followed by Kupicha, with the exclusion of V. mollis. Kupicha (1976, p. 323) includes four species in her sect. Peregrinae and she comments that the species have no peduncle. This is true for V. aintabensis, V. michauxii and V. peregrina. However, V. mollis does, in fact, possess a much reduced peduncle, so cannot be considered a natural member of this group. The remaining three species are closely allied and Ponert (1973) regards the three forms as subspecies of V. peregrina. The analysis results do suggest the forms are closely allied, but the existence of sets of correlated characters used elsewhere to distinguish species and used in sect. Peregrinae to distinguish the three forms indicates that they do warrant specific status.

With all the closely related groups studied there were problems of missing data. If a group of taxa are closely related, by definition, they are distinguished by a small number of correlated characters. If it is not possible to score these diagnostic characters then they are likely to be confused. The distinction between V. aintabensis and V. michauxii, in the above study, was a good illustration of this problem.

The results of the different methods of analysis each concur that Kupicha's sect. Vicia should be split in three. It is proposed that V. lathyroides and V. cuspidata are separated into a distinct section, Wiggersia, from the Vicia sensu stricto. These two species were recognised as a distinct grouping by Alefeld (1860b), who included them in his conception of the genus Wiggersia. This grouping was also separated from Vicia sensu stricto by Fedtschenko (1948) and Plitmann (1967). The two species are distinguished from sect. Vicia sensu stricto by the gross size of the plants, the subsessile legume and the presence of a sculptured seed testa. Stankevich (1982, 1983) also noted that V. lathyroides and V. cuspidata were distinct in comparison to other sect. Vicia sensu Kupicha species. She considered the two species sufficiently distinct from subg. Vicia taxa to include them in the reinstated genus, Ervum, as sect. Subsessiles Stankev. The results of the phenetic analysis do not suggest such a clear separation from the other sect. Vicia sensu Kupicha, but the separation is sufficient to warrant distinct sectional status for the two species.

The results of the analysis of sect. Vicia sensu stricto (discussed in thesis section 7.4.2) indicates that the species fall into two distinct groups. However, the division between these two groups is not as distinct as between sect. Wiggersia and sect. Vicia sensu stricto. So the two groups within Vicia sensu stricto are given series status, ser. Vicia, which contains V. pyrenaica, V. barbazitae and V.

sativa and ser. Grandiflorae, which contains V. grandiflora and V. qatmensis. V. grandiflora was first distinguished from the V. sativa agg. by Alefeld (1861a), who created the monospecific genus Cujunia. The need for a discrete taxon to include V. grandiflora was also seen by Fedtschenko (1948). He referred to the unit as ser. Grandiflorae, but did not validly publish the name. It was Radzhi (1971) who finally published ser. Grandiflorae. The two series are distinguished from each other by corolla colour, corolla size, legume shape, legume size and the relative lengths of the seed hilum to its circumference.

Kupicha's sect. Faba is split into three major subunits by the analysis. V. faba and V. bithynica are split into monospecific sections, while the remaining species are included in sect. Narbonensis. This subdivision of V. faba and its relatives into three sub-groups has been previously suggested by Fedtschenko (1948) and Plitmann (1967). The V. narbonensis complex species being placed between the two monospecific sections in a manner similar to Kupicha's listing of sect. Faba taxa. Many researchers, undertaking a taxonomic study of sect. Faba, have tended to approach the problem by investigating the relationship between the V. faba and its allies. By approaching the problem in this way they have found the fababeans closest ally is the V. narbonensis complex of species and so have placed V. faba together with them in a single section. This approach makes the results biased in favour of including V. faba in a single taxon with its closest allies. For a clear view of the relationship between V. faba and other Vicia species, a broader approach must be taken. The broader approach gives scale to the allocation of series, section and generic rank and so allows the position of V. faba to be more 'naturally' defined.

The three proposed sections are differentiated by plant habit, gross morphology, leaflet size and shape, whether their hairs are tubercular, flower colour and size, and legume and

seed shape and size. The less robust, narrow leaved V. bithynica has several characters which suggest a distant link to sect. Vicia and so this monospecific section is placed on the Wiggersia side of Sect. Narbonensis. The distinct nature of V. bithynica within subg. Vicia was noted as early as 1836 by Koch, who placed the species in a monospecific sub-sectional grouping. Subsequently it has been included with the V. narbonensis complex species, with sect. Vicia species or placed with the species related to V. cracca. The radical difference of opinion between Vicia taxonomists on the placement of this species within Vicia is largely due to the relative importance of peduncle length in the various classifications. V. bithynica has the overall features of Vicia subgenus Vicia sensu Kupicha (1976), but the peduncle is often longer than the flower, which makes it the exception in a group where the flowers are sessile to subsessile.

V. bithynica has, however, most commonly been linked with the V. narbonensis complex, with which it shares the large serrate edged stipule. Kupicha (1984) comments that she links V. bithynica with her sect. Faba species, because the species fits even less comfortably elsewhere. The results of the phenetic analysis indicate that V. bithynica is quite distinct from other subg. Vicia taxa and, as such, warrants the sectional rank it is given here.

Sect. Narbonensis comprises the species of the V. narbonensis complex and these seven species can be further split into two groups. The level of differentiation between these two groups is insufficient to warrant sectional rank and so is given series status. One species, V. eristalioides, is included in ser. Rhombocarpae and the other six more closely related species in ser. Narbonensis. These two series can be distinguished primarily by a clear difference in legume shape, V. eristalioides like V. bithynica has a rhomboid legume, as opposed to the linear-rectangular legume of the V. narbonensis complex species.

Sect. Narbonensis and sect. Faba, as defined here, were considered sufficiently distinct from the other Vicia species for Stankevich (1978, 1982), on the basis of morphological, anatomical, cytological and biochemical evidence, to split them into the respective genera Bona and Faba. The results of the phenetic analysis, considered with evidence from the literature do not suggest such a distinct separation, although a detailed study of the generic limits of Vicia, including Vicia and related genera would be required to confirm this view. However, broad studies of this kind undertaken by Hanelt et al. (1972) and Kupicha (1976) do not support Stankevich's thesis. Linnaeus (1753) comments that he examined the flowers of hundreds of plants of V. faba and other Vicia species and could not find any characters that differed significantly enough to warrant generic separation. This position seems as justified today as it did two and a half centuries ago.

9.4 The Revision Conspectus

The primary product of the revision is the classification of taxa, discussed above. The secondary products of a revision are many (see Chapter 3) and can most conveniently be summarised in a conspectus. A revision conspectus is broadly defined as a summation of the available taxonomic knowledge for the taxon being revised. In the terms of the revision paradigm, discussed in Chapter Three, the conspectus is a summary of the revision products. The conspectus for Vicia subg. Vicia is provided in Appendix 5.

The major innovation in this thesis is the use of database techniques to facilitate the revision process. The database technology was used to gather, order and analyse the masses of data associated with the taxonomic revision. The building of the Vicia subg. Vicia revision database enhanced the production of the conspectus. The raw data held in the database were ordered and then exported and the text was then edited, where necessary, to produce the conspectus. The

taxonomic notes were the only material to be included in the conspectus that was not extracted from the database. These notes were free text critical remarks, which reviewed and interpreted the data extracted from the revision database.

The following fields are included in the conspectus for each taxon: accepted taxon name; author(s); date of publication; where published; type location and provenance; iconography (reference to published descriptions and illustrations); common synonyms; taxon description; phenology; chromosome number; geographical distribution (two letter country ISO codes); distribution map; ecological notes, including altitude and habitat notes; taxonomic notes and specimen citations, for those specimens used during the revision. The taxon descriptions and keys to taxa were produced using the DELTA system (see discussion in Chapter Ten). The DELTA items files were produced by program from the raw specimen character scores held in the database.

The conspectus includes dot distribution maps for each species (infra-specific taxon distributions are included on a single species map), prepared using the program ATLAS GRAPHICS (Strategic Locations Planning, 1987). The input file, containing the latitude and longitude coordinates, used in the preparation of these maps was taken from the database. The database file ECOGEOG contains the latitude and longitude for each specimen borrowed. Once each specimen was attributed to a taxon, the file was indexed on taxon identity and the distributional data for the specimens representing that taxon were extracted from the database and imported into ATLAS GRAPHICS for the plotting of their distribution.

The conspectus has a double value: it should be read as a summary of the available knowledge on subg. Vicia, and it can be seen as one of the first examples of a taxonomic conspectus created almost directly from a revision database.

9.5 Conclusion of Phytogeographical study

Chapter eight provides a detailed survey of the literature and a discussion of the distribution of the subgenus Vicia groupings, as defined by Kupicha (1976). Now that a novel classification of this subgenus has been proposed, these conclusions will be reviewed. One of the programs, discussed in Chapter Ten, is PHYTOGEOG, which summarises taxon distributions based on representative specimen distributions. The program produces a dBASE file, GEODIST, that contains the geographic unit distribution for each taxon. These data are included in Appendix 6 and the file is provided on disc 1 of Appendix 7.

The distribution patterns of the 18 supra-specific taxa of subg. Vicia are summarised in Table 9.3, the sectional distribution map drawn in Figure 9.1 and the isoflor maps for the six multi-taxon sections are drawn in Figures 9.2 to 9.7.

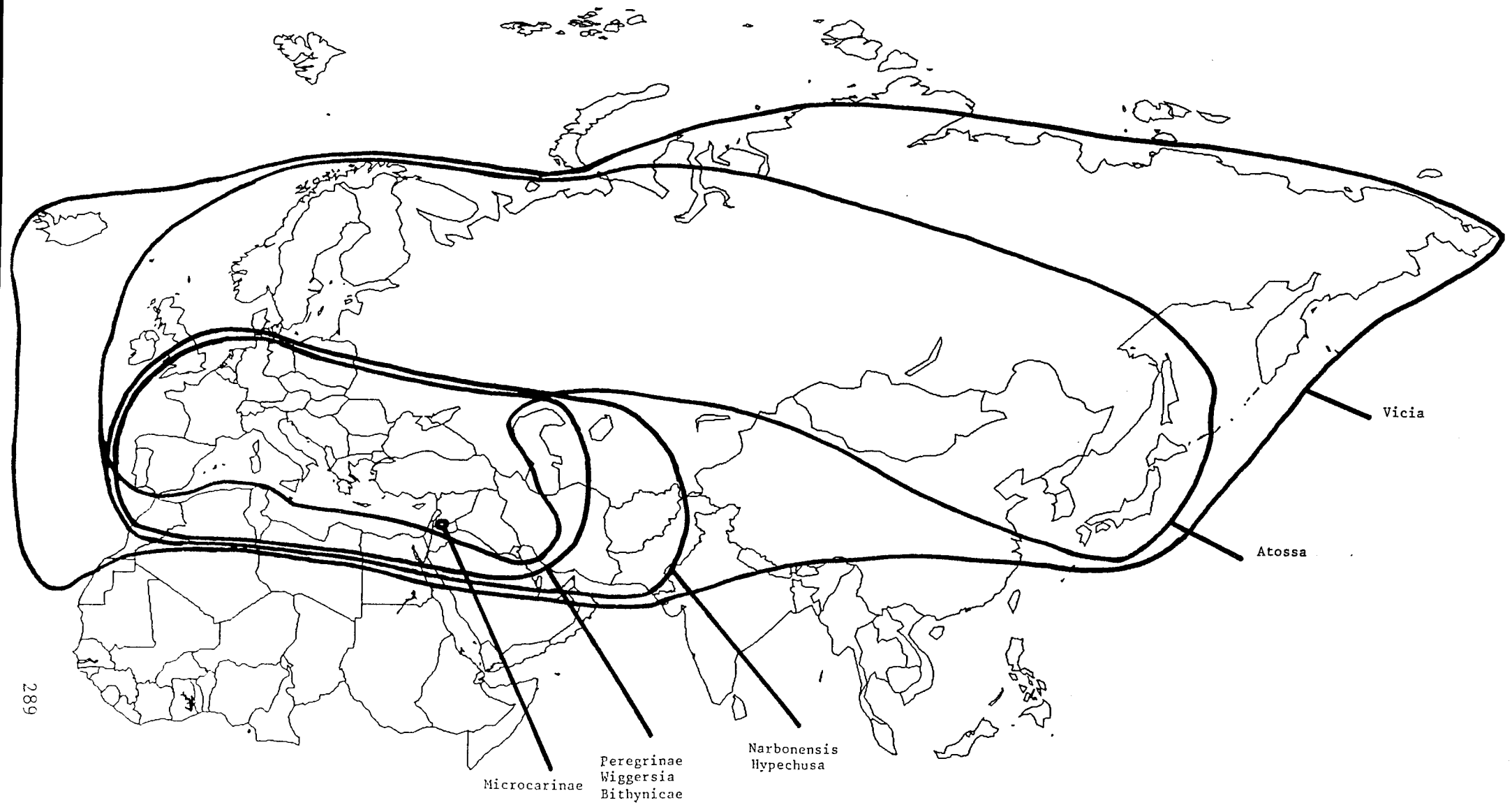
Isoflor maps do not show actual species distributions, but each line is a contour delimiting a greater or lesser concentration of species. The species distributions for the species included are plotted onto one sectional map. Contours are then drawn around areas of the map to indicate the concentration of species at any one point. The isoflor maps are intended to indicate patterns of distribution concentration, rather than actual distributional patterns. The three monospecific sections, sect. Microcarinae (V. dionysiensis), sect. Bithynicae (V. bithynica) and sect. Faba (V. faba), have not had isoflor maps drawn as they cannot be drawn for a single taxon. V. faba is not known in the wild and so the distribution cannot be plotted. However, the area of cultivation for V. faba is discussed in detail by Maxted et al. (1989).

The novel classification does not alter the patterns of distribution for the sections of subg. Vicia as all sections of the subgenus are predominantly Southern European and West

Table 9.3. Distribution of supra-specific taxa of *Vicia* subgenus *Vicia*

Taxa		Distribution
Section	Series	
Sect. Atossa		Europe, North & West Asia
	Ser. Pseudovicilla	Central, South & East Europe
	Ser. Truncatulae	South East Europe, North east Mediterranean, South USSR & North Iran
	Ser. Atossa	Europe and North Asia
Sect. Microcarinae		Syria (endemic)
Sect. Hypechusa		Europe, West Asia & North Africa
	Ser. Hyrcanicae	West Asia
	Ser. Hypechusa	Europe, West Asia & North Africa
Sect. Peregrinae		South Europe, West Asia & North Africa
Sect. Wiggersia		Europe, West Asia & Algeria
Sect. Vicia		Europe, Asia & North Africa
	Ser. Vicia	Europe, Asia & North Africa
	Ser. Grandiflorae	East Europe & West Asia
Sect. Bithynicae		Europe, West Asia & Algeria
Sect. Narbonensis		Europe, West Asia & North Africa
	Ser. Rhombocarpae	Turkey (endemic)
	Ser. Narbonensis	Europe, West Asia & North Africa
Sect. Faba		Only known in cultivation

Figure 9.1. Distribution of subgenus Vicia sections.



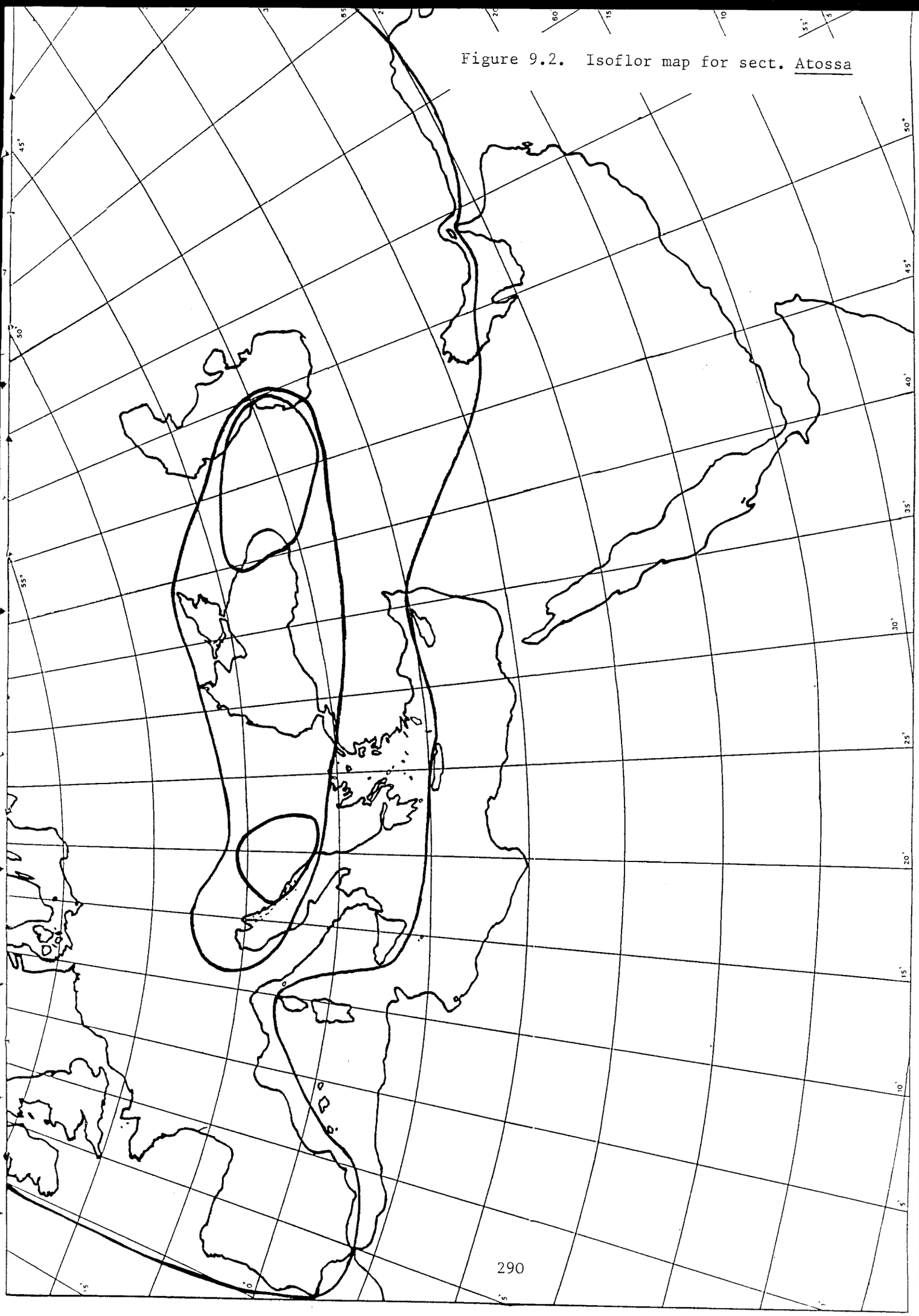


Figure 9.2. Isoflor map for sect. Atossa

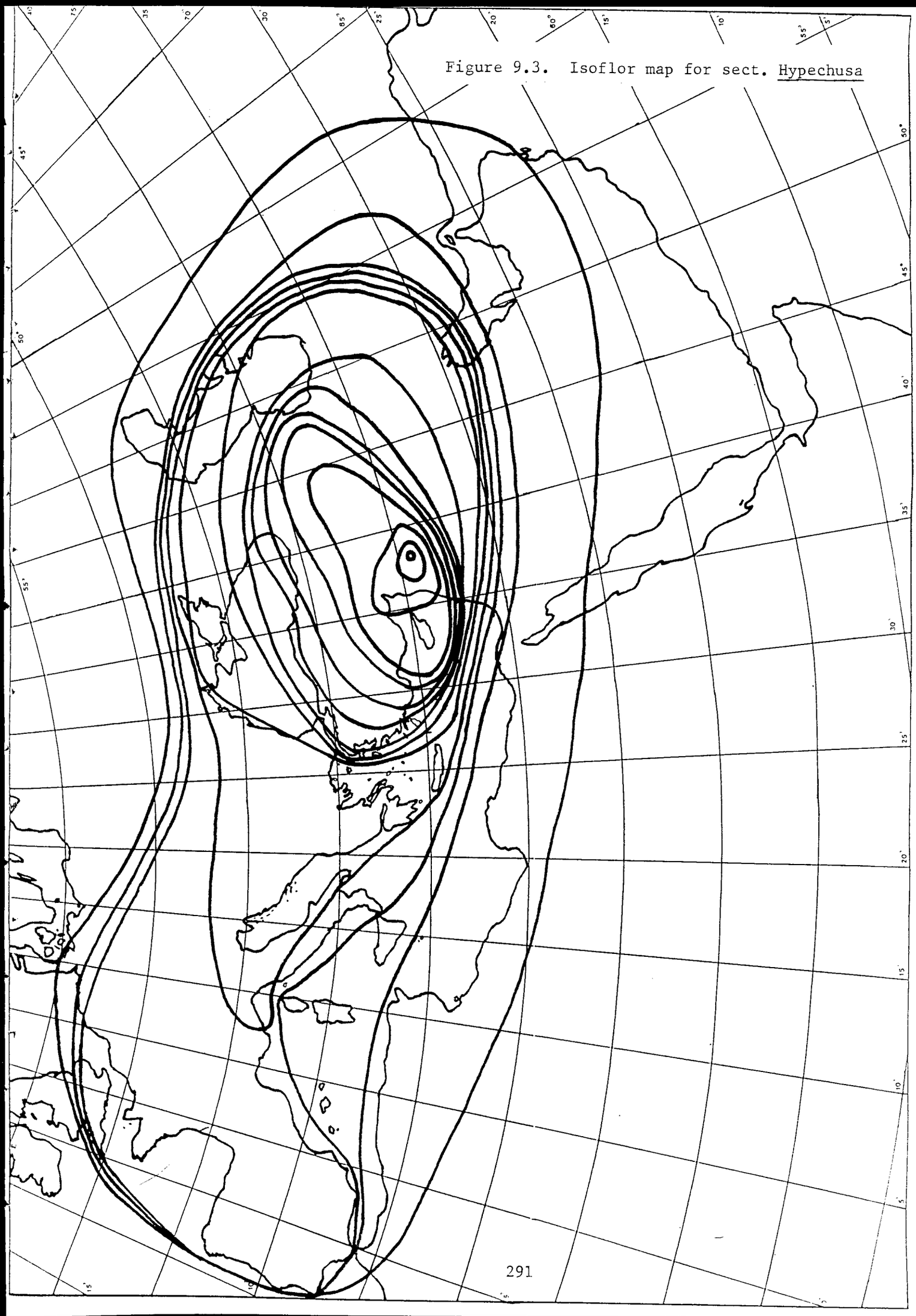


Figure 9.3. Isoflor map for sect. Hypechusa

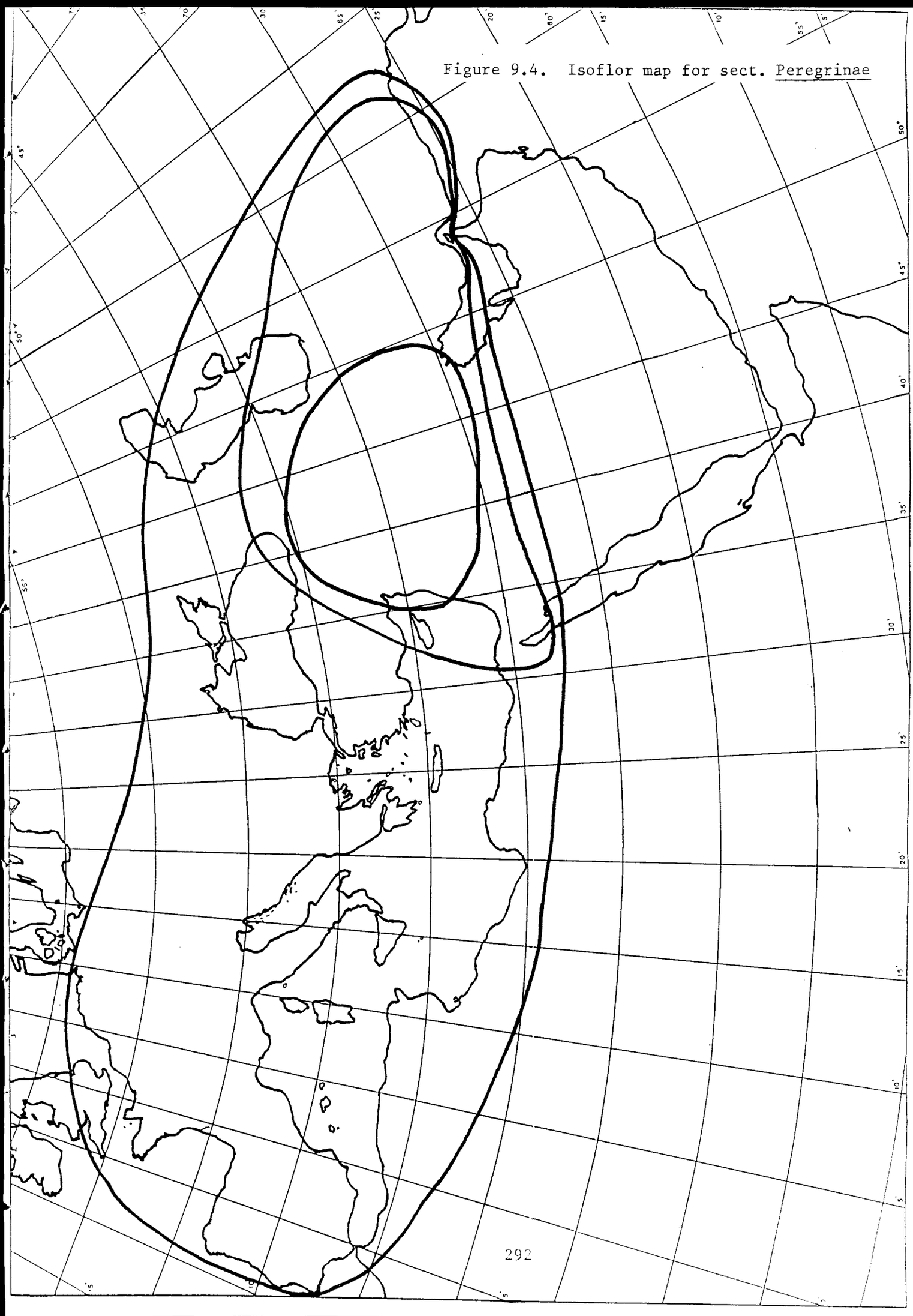


Figure 9.4. Isoflor map for sect. Peregrinae

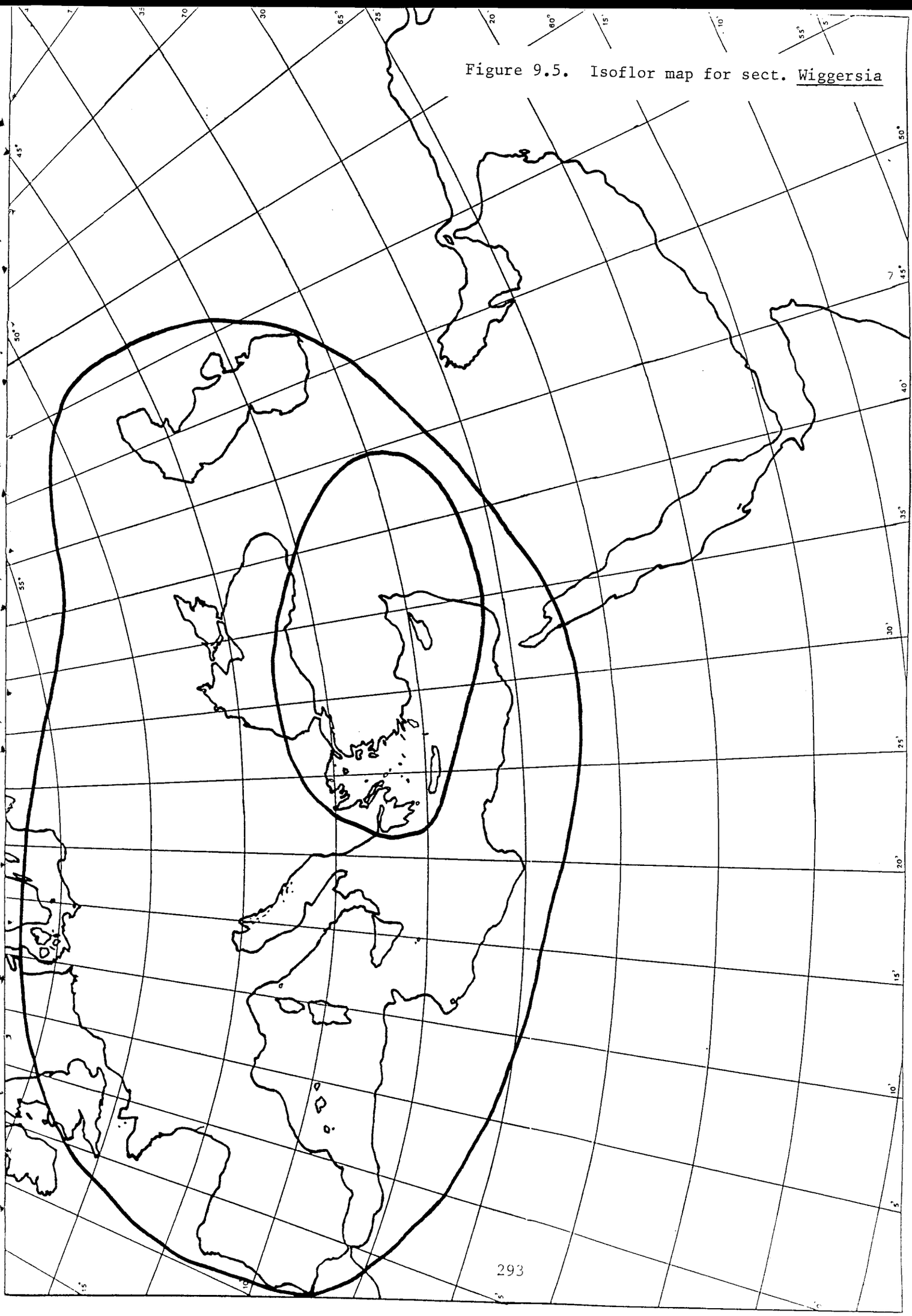


Figure 9.5. Isoflor map for sect. Wiggersia

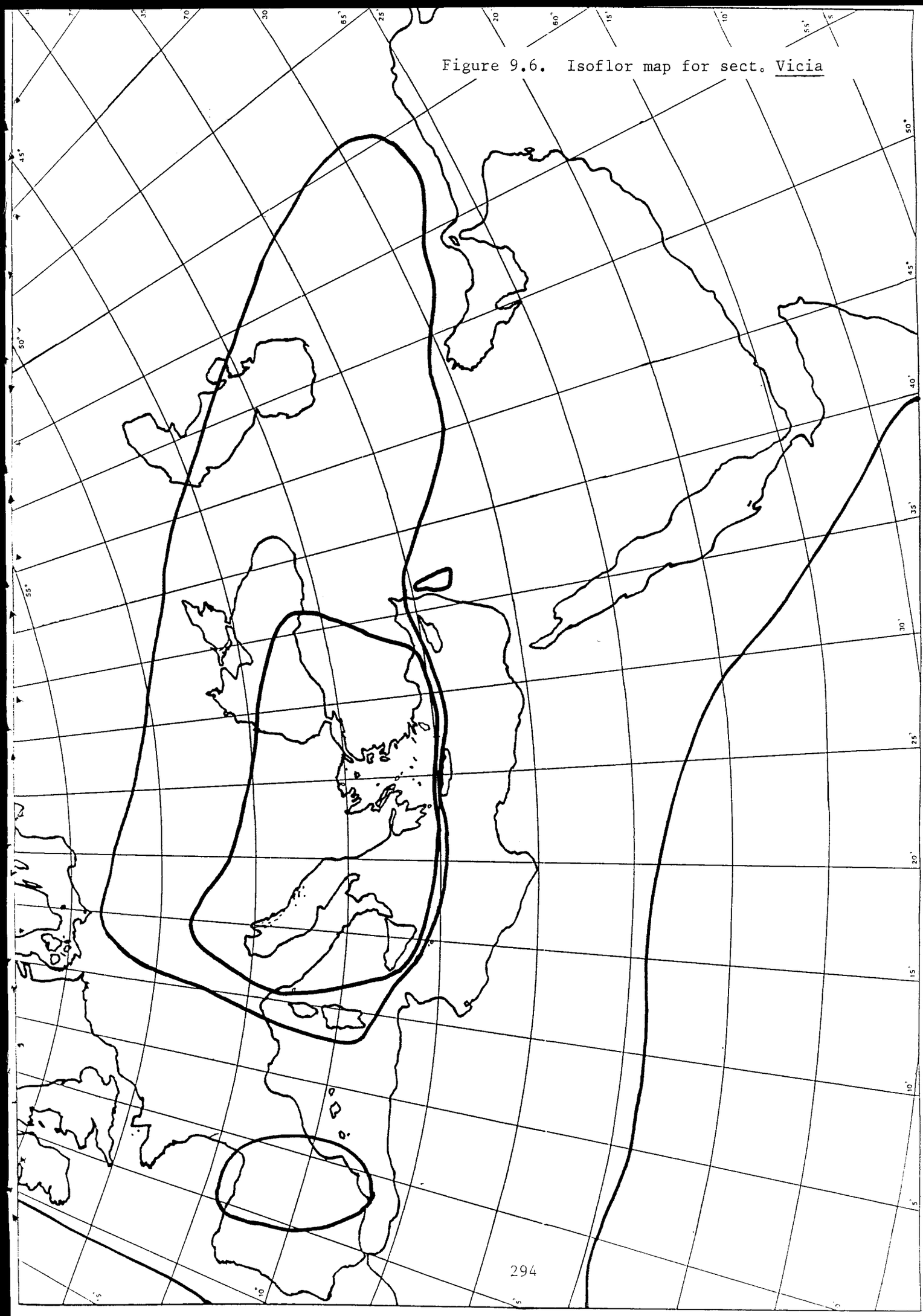


Figure 9.6. Isoflor map for sect. Vicia

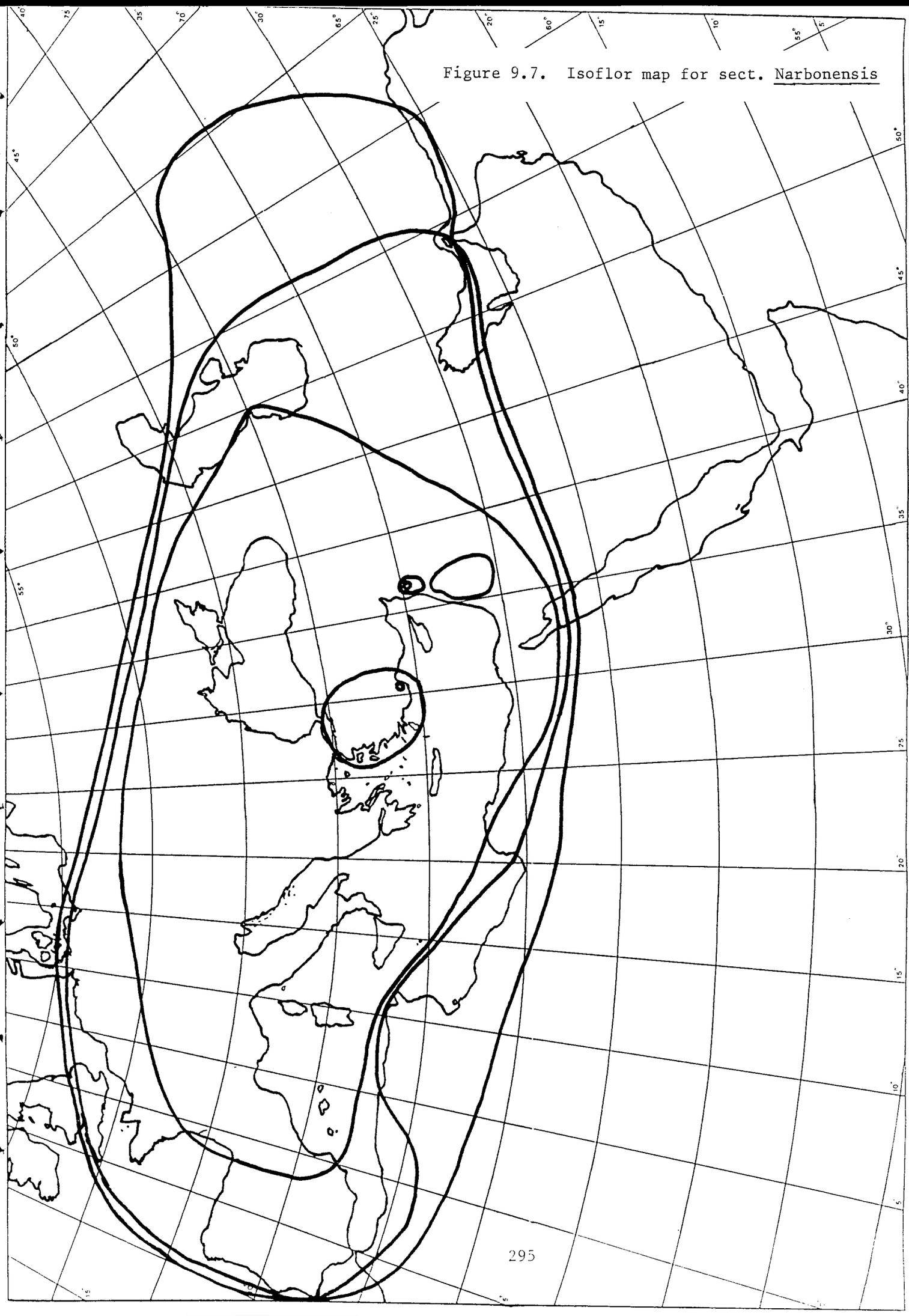


Figure 9.7. Isoflor map for sect. Narbonensis

Asiatic, with extensions into Northern Europe, Northern Asia and North Africa. The highest concentrations of taxa, and thus diversity, are found in the fertile crescent countries of Turkey, Lebanon, Syria, the Soviet Caucasus and Central Asian regions and Iran and Iraq. The majority of the sections and series can be seen to be widely distributed throughout Europe, West Asia and North Africa. The two exceptions are the taxa that each contain one restricted endemic species, sections Microcarinae (V. dionysiensis from Suweida, Syria) and ser. Rhombocarpae (V. eristalioides from Antalya, Turkey).

Isoflor maps indicate the centre of diversity of a group of taxa. The isoflor map for sect. Atossa, drawn in Figure 9.2, indicates dual centres of diversity concentrated in Northern Yugoslavia and the Soviet Caucasus. Within the section, the three series have quite distinct distributions. Ser. Pseudovicilla is centred on Middle Europe, ser. Truncatulae on the Caucasus and ser. Atossa is much more widely distributed in Europe and North Asia. The differences in distributional centres underline the taxonomic distance (based on morphological characteristics) between these three series. The level of intra-sectional variation suggests that the three series are relatively remote. It is argued above, however, that the section should be retained as a unit because of the phylogenetically important correlated characters (perennial life form, relatively long hilum length and the lack of legume pubescence) and also to avoid creating too fine a taxonomic hierarchy.

The isoflor map for sect. Hypechusa, drawn in Figure 9.3, shows that the centre of diversity is focused on the fertile crescent countries of South West Asia. The distributional pattern of two series within sect. Hypechusa is also centred on South West Asia, although ser. Hypechusa stretches more westerly to encompass Southern Europe than ser. Hyrcanicae. The section seems, both morphologically and geographically, to form a cohesive taxonomic unit. A similar position is

indicated by the isoflor map, drawn in Figure 9.4, for sect. Peregrinae. The distribution of the three included species is focused in South West Asia, with V. aintabensis and V. michauxii restricted to this area, but V. peregrina broadens the distribution of the section to Europe and North Africa. The two species of sect. Wiggersia, whose isoflor map is shown in Figure 9.5, shows a similar distribution pattern and centre of diversity to the two previously discussed sections. V. cuspidata is restricted to South East Europe and South West Asia and V. lathyroides broadens the distribution of the section to include North Africa and Western Europe.

Sect. Vicia is composed of five species, one of which is the pan-temperate and semi-tropical weed, V. sativa. The extent of the distribution of V. sativa causes the discussion of the distribution of sect. Vicia to be dominated by the one species. As V. sativa is so widespread, the precise native distribution of the species is unknown. Based on the distributional pattern of the subgenus and, particularly, sect. Vicia, V. sativa seems most likely to have originated in Southern Europe or South West Asia. Support for this hypothesis is provided by the distribution of three of the subspecies of V. sativa (subsp. amphicarpa, subsp. incisa and subsp. macrocarpa), which are endemic to this region.

The isoflor map for sect. Vicia, drawn in Figure 9.6, excludes V. sativa as it is found throughout the region and so provides little help in interpretation of the distributional pattern. The centre of distribution, indicated by the map, is South East Europe and South West Asia, with two restricted species (V. pyrenaica and V. qatmensis) forming isolated pockets of distribution in the Iberian Peninsula and Syria. The discontinuous distribution of these four species could suggest that they have evolved separately from a V. sativa form. The most likely example is the evolution of V. pyrenaica. The morphological similarity between this species and V. sativa suggests the former is simply a form of the

latter, that has evolved, due to geographic isolation, in a high alpine environment.

Sect. Narbonensis is composed of four species which are restricted to South West Asia and three species which are more widely distributed throughout Europe, West Asia & North Africa. So not surprisingly, the isoflor map for the section, drawn in Figure 9.7, indicates that the section's centre of diversity and possible origin is South West Asia. This hypothesis is supported, at least for V. narbonensis, by archaeological evidence (Schultze-Motel, 1972; Renfrew, 1973) from neolithic levels at Beidha in Israel.

The sectional distribution and distributional concentrations, described above, provide little help in clarifying subg. Vicia taxa relationships, as all nine sections have overlapping distribution patterns with the same centre of diversity. The sectional distributional pattern does, however, indicate that the subg. Vicia sections are more closely related to each other, than they are to sections whose distribution is endemic to the Americas, e.g. Americanae, Leucophaeae and Australes.

The natural distribution of subg. Vicia species is restricted to Europe, North and West Asia and North Africa. In interpreting this distribution, it is worth noting the remarks of Brenan (1965):

> "The genus is a particularly useful unit for studies of this sort (Phytogeographical), since the number of species of a given genus occurring in different parts of the world often indicate its centre of origin and directions of spread."

The derivation of centres of origin are just as likely to be reflected in the distribution of subgeneric taxa. Thus it can be argued on the basis of relative species concentrations (as indicated by the isoflor maps) that the centre of origin for subg. Vicia is South Eastern Europe and South Western Asia.

Over half of the included taxa are endemic to this area . Taxa of the subgenus have subsequently spread West and North, and, to a lesser extent, East and South from this area.

This region is very important floristically as it is at the cusp of three major temperate phyto-geographical regions, Euro-siberian, Mediterranean and Irano-turanian (Takhtadjan, 1969). Stebbins (1967) points out that if areas with different environmental and floristic characters are juxtaposed, it may act as a catalyst to evolution. Subgenus Vicia would appear to be a good example of a taxon to have evolved in such an evolutionary "melting pot".

There is evidence, however, that taxa with centres of distribution in the Balkans, South-west and South-east Asia are generally centred in these areas due to climatic deterioration and floristic movement from more northerly distributed Floras (Takhtadjan, 1969). The Boreal-tertiary Flora was driven South by the Quaternary Ice Age towards these three floristic refugia in the South. Kupicha (1974) concludes that Vicia (and by implication subg. Vicia) evolved in the early tertiary and the centre of origin is likely to have been much further North than is indicated by contemporary concentrations of taxa.

Stebbins (1967) suggests that the most genotypically and phenotypically plastic species (e.g. V. sativa), will evolve rapidly away from the centre of diversity, while the more primitive species will remain at the centre. Contrary to this hypothesis, however, Willis (1921) argues that evolution and plant dispersal occur at the same rate. Therefore the oldest and most primitive species are most widespread and are currently found on the distributional periphery, while the most recently evolved species occur at the centre of origin and are localised. It would explain the widespread distribution of V. sepium, which is regarded as a primitive member of subg. Vicia, because of its perennial habit,

relatively long peduncle and numerous flowers per peduncle. This hypothesis, though, clearly does not explain the distribution of V. sativa, which is regarded as an advanced species. The fact that three sections (Atossa, Vicia and Narbonensis) have a significantly wider distribution indicates these are likely to be the more advanced (or more primitive if Willis's thesis is followed) than the sections which are endemic to the centre of diversity.

9.6 Suggested Future Taxonomic Research

The suggestions for further research on Vicia subgenus Vicia, related both to those taxonomic problems that remain unresolved and to the numerous gaps that exist in the non-morphological evidence, they are:

1. V. sativa is undoubtably the most ubiquitous and diverse Vicia species. As such this species warrants a specific revision (as is currently being undertaken by Y. Potokina at the N.I. Vavilov Institute, Leningrad, U.S.S.R.). V. sativa is in chaos, both taxonomically and nomenclaturally. There are several times more names attached to this one species, than to any other Vicia species.

   I recognise six subspecies of V. sativa and comment that it is likely that a seventh 'lentil seeded' form exists. However, I have been unable to distinguish two other widely used groupings: subsp. cordata (Wulfen ex Hoppe) Arc. and subsp. nigra var. segetalis (Thuill.) Ser. ex DC. Many authors have attempted a specific clarification of V. sativa, but none, including myself, has undertaken a detailed enough study to provide a stable intra-specific classification. A combination of cytological, biosystematic, and phytochemical investigations are required if the infra-specific taxonomy is to be clarified.

2. Vicia Sect. Hypechusa contains 14 species divided into two series. However, this group is heterogeneous and contains more distinct clusters of taxa than the two major groupings used in the proposed classification. The two other sections containing a complex of closely related species (Vicia and Faba sensu Kupicha, 1976) have been extensively studied, but there has been no detailed study of this section. For this reason a study focused on this section is required.

3. As with the lack of a detailed study of sect. Hypechusa, the three sect. Peregrinae species have never been studied together. Though this section is thought to contain three species here, Ponert (1973) uses five subspecies to describe the variation. Uniform garden experiments are required to resolve this problem.

4. Recent work by Stankevich (1982) has suggested that the V. narbonensis complex species and V. faba should be split into two distinct genera, Bona and Faba respectively. The results discussed here do not support this view, but as this study focused on one subgenus of Vicia, a more detailed study of the generic limits of Vicia as a whole is required to clarify the position of these two groups.

5. The revision discussed here concentrated on one of the subgenera of Vicia identified by Kupicha (1976). Subg. Vicilla is equally deserving of this level of study. Field experience during the course of this project has identified problems in the taxonomy of subg. Vicilla; the complex of species around V. cracca certainly requires clarification and numerous sect. Cracca endemics from the Soviet Union require examination to see if they do warrant specific rank.

6. This study has concentrated on a phenetic study of subg. Vicia. No attempt has been made to undertake a phylogenetic study or to draw phylogenetic conclusions for the subgenus. It is suggested that such a phylogenetic study of subg. Vicia be undertaken.

7. Having undertaken extensive field work in the centre of diversity of subg. Vicia, vouchered seed of all but a few taxa is available from the Vicieae Genebank (Southampton, U.K.). Such a comprehensive seed stock is rare for any wild plant group, so the suggestion is that this material should be used in conjunction with the proposed classification and conspectus to undertake detailed phytogeographic, cytological, biosystematic, palynological, anatomical and phytochemical investigations of subg. Vicia. A relatively well known group that has a readily available seed source may also prove useful as experimental material in the development of new taxonomic techniques.

SECTION

THREE

DATABASE RESULTS

AND DISCUSSION

CHAPTER TEN
APPLICATION OF DATABASE TECHNOLOGY TO THE EXEMPLAR REVISION

10.1 Introduction

The central element to the "Dream" (Bisby, 1984a) was the production of an automatic taxonomic information system, in which a central computerised taxonomic database can be used to automatically produce a series of taxonomic products. This chapter is concerned with a particular practical implementation of this "Dream"; how database techniques could be used with other computer techniques to enhance a revision.

The discussion divides into three sections: the first concerns the choice of the proprietary software used in the database element of the project; the second concerns the design and implementation of a computer-aided system which facilitates the exemplar revision; and the final section discusses how database and other computer-based techniques were used to facilitate the exemplar revision.

10.2 Choice of Proprietary Software

It is an over-simplification to imply that only database software was used to facilitate the revision. The software used can be separated into three components: the database management system (DBMS), the computer-aided product generation programs and the phenetic analysis packages. Use of all three components facilitated the revision. Though each of these components has been previously used in revisions, it is the linked usage of these components that provides the major innovation of this project. The choice of the two former components is discussed below. The third component, phenetic analysis packages, is now an established feature of effective contemporary revisions and the choice of the packages was discussed in detail in Chapter Six.

10.2.1 Database Management Systems

The implementation of computer aided techniques to the revision process required experimentation with the use of a proprietary DBMS. This necessitated the choice of a DBMS package. Freeston (1984) provides a review of the features that should be considered when choosing a DBMS, listing such considerations as: cost, portability, ease of use, ease of use of programming language, available applications, multi-user access, ease of data protection and security. Having listed these considerations, the choice in practice was essentially between two systems, dBASE ll and KNOWLEDGEMAN and, in January 1985 at the start of the project, the decision was taken to use dBASE II (Ashton-Tate, 1984).

Of the two systems considered, the former was selected for the following reasons:

a) it is a relational database management system of proven reliability,

b) it is widely used throughout the industrialised world. Over 2000 copies of dBASE ll were being sold monthly in 1984,

c) it was considered a relatively versatile system, with its own programming language, allowing the user to write specific functions and programs,

d) dBASE ll was considered more user friendly than KNOWLEDGEMAN although the latter allowed a larger number of fields per record and more than 2 database files open at any one time,

e) dBASE ll was already available in the Dept. of Biology, The University, Southampton and so could be used at the commencement of the project,

f) and because dBASE 11 was available in the Department of Biology, there was some expertise available in using this DBMS.

Details of the operating instructions for constructing and manipulating files are provided in the dBASE 11 manual (Ratliff 1982). This was used in conjunction with a simple guide (Townsend, 1984) to obtain sufficient competency in using dBASE II.

Lang (1983) discusses the best use of dBASE II, while highlighting the limitations of this DBMS. Two of the restrictions limited the design freedom in development of the revision database. Firstly, dBASE II only allows 2 database files to be open at any one time and, secondly, the number of fields per record is limited to 32. Both these problems will be discussed in detail in relation to the database implementation in section 10.3 of this chapter. Although using dBASE II did restrict the precise way in which database files were used within the revision, it did not restrict the overall design and implementation of the database assisted revision paradigm.

10.2.2 Computer Aided Taxonomic Product Generation

As was discussed in Chapter Two the object of a taxonomic revision is to synthesise various primary and secondary products based on a detailed taxonomic study of a particular taxon. The core products of a revision are fivefold: a circumscription of accepted taxa in the revision taxon, a classification of taxa, descriptions of the taxa, a summary of nomenclatural information associated with the taxa and various identification aids. Computers were first used to aid classifications by undertaking phenetic or numerical analysis of character data sets representing the circumscribed taxa. This was discussed in Chapter Six. Since the early 1960's computers have been used increasingly to help produce

taxonomic descriptions and identification aids (Pankhurst, 1975a).

Computers have been used to assist in the production of full descriptions, diagnostic descriptions, online identification, identification via matching, batch key construction, online key construction, typesetting of descriptions and conversion to other data formats for further analyses e.g. CLUSTAN or PAUP. The following brief review will concentrate on the use of computers in the production of these products from taxonomic data sets.

Morse (1968, 1971) was one of the first workers to produce computer programs that assisted in taxonomic product generation. He described a suite of FORTRAN programs (Morse, 1974), using a time-sharing main-frame computer which could be used in specimen identification. The system uses a particular taxonomic data matrix in conjunction with six sub-programs that utilise the data to produce specimen identification, key construction, descriptive printing, taxon-taxon comparison, preparation of inverted descriptions and production of punch-card field keys. The package of programs also contains a simple data editing and manipulation module. The basic limitations of this early system were its inflexibility and its strong ties to mainframe computers (Keller & Crovello, 1975).

An important element in Morse's work was the production of taxonomic products straight from the coded taxonomic data matrix. Subsequent authors have expanded the programs available for botanical product generation. For example: Pankhurst (1971), Watson & Milne (1972), Hall (1970, 1975), Dallwitz (1974), and Payne (1975a, 1975b) have produced diagnostic key production programs; Pankhurst & Aitchison (1975a), and Johnston (1980) have produced programs which aid the construction of polyclaves (diagnostic keys using punched cards); Pankhurst (1975b) has produced an identification aid

which matches the specimen with the taxa included in the taxonomic dataset; Pankhurst & Aitchison (1975b), and Forget et al. (1986) have produced online identification programs that allows specimen identification via a question and answer procedure at the computer terminal. These programs are reviewed in more detail by Pankhurst (1978) and Abbott et al. (Chapter 9, 1985).

Two suites of product generation programs dominate contemporary taxonomic product generation, PANKEY written by Pankhurst (1986) and the DELTA based system written by Dallwitz (1979). As the software associated with ALICE (Allkin, 1988) is expanded it will also play an increasingly important role in taxonomic product generation, but currently the Pankhurst and Dallwitz suites of programs are most universal. Both the Pankhurst and Dallwitz systems can be run on IBM compatible micro-computers and use as a basis for their product generation the data format DELTA (Dallwitz, 1980; Dallwitz & Paine, 1986).

DELTA (DEscriptive Language for TAxonomy) is a generalized system for the concise coding of all kinds of characters developed by Dallwitz from one used by Watson & Milne (1972). Data written in this format can be utilised by a series of application programs. The system was written to be user friendly and, as such, it allows character dependencies and comments to be incorporated within the data. DELTA is capable of encoding the many kinds of characters (unordered multistate, ordered multistate, integer numeric and real number) commonly used in identification and classification. Dallwitz (1980a) comments that the system is versatile enough to replace natural-language descriptions and lists the following general advantages of using the system:

"1. The coded descriptions occupy less space than natural-language descriptions.

2. It is easier to ensure that descriptions are consistent and complete.

3. Translation into different natural languages is easier.

4. The data are easily accessible to computer programs (for information retrieval, classification, key construction, etc.)."

Dallwitz (1984) adds a further thirteen more specific advantages of using DELTA:

"1. It is a "free" format i.e. the data do not have to be placed in particular positions in the lines;

2. there is no limit on the lengths of the character descriptions (in most formats, each feature and state is limited to one line of 70 or 80 symbols);

3. the feature and state descriptions of each character are adjacent, and the states are numbered in a straightforward way;

4. all the main types of character are catered for (unordered and ordered multi-state, integer and real numeric);

5. the characters may be placed in any order (most formats require characters of the same type to be placed together);

6. comments may be included in the character descriptions;

7. there is no limit on the length of each taxon name;

8. comments may be included with the taxon name;

9. the name and the coded description of each taxon are adjacent;

10. distinction may be made between "unknown", "variable" and "inapplicable";

11. partial variability, i.e. the possession of more than one but not all of the states of a character, is coded in a straightforward way (some formats represent the presence of a state by the presence of the corresponding bit (binary digit) in a number, e.g. "state 1 or state 3" would be represented as "5");

12. distinction may be made between "or", "and", and "to";

13. comments may be included in attributes."

A more general advantage of using DELTA, is that the data is recorded in a highly structured manner which forces the taxonomist to think logically and consistently about data manipulation and taxonomic product synthesis. For all these reasons it is not surprising that DELTA has recently (Bisby, 1989b) been adopted as the International Working Group on Taxonomic Database for Plant Sciences (TDWG) format standard for coded taxonomic descriptions.

Precise details of the data format are provided in Dallwitz & Paine (1986). Initially two programs were available for use with the DELTA format, CONFOR and KEY. CONFOR (Dallwitz, 1979) converts the data in DELTA format to other formats, performs renumbering of characters, produces natural-language descriptions and translates the DELTA data into KEY format. KEY is a key generation program (Dallwitz, 1978).

Subsequently, these programs have been altered and further programs have been added to the Dallwitz suite of programs. Dallwitz (1988) lists the currently available programs as follows:

"1. CONFOR - Converts DELTA into natural-language descriptions, or into other formats. The other formats currently available are:

   a. KEY, DIST and INTKEY formats for other programs in the package.
   b. PAUP format for cladistic analysis.
   c. DELTA format, to tidy the data files, to re-order characters or character states, or to produce subsets of the data.

2. KEY. Constructs diagnostic keys.

3. DIST. Constructs a distance matrix for phenetic analysis.

4. INTKEY. Interactive identification and information retrieval."

To these basic programs Dallwitz has added TYPSET (Dallwitz, 1980b, 1984) a computer typesetting program, which enables the conversion of DELTA data into natural-language descriptions and keys that can be directly printed in book format. Watson & Dallwitz (1980) have used this facility to produce a monograph on Australian grass genera.

Examples of DELTA format and the use of Dallwitz's programs in the exemplar revision are provided in section 10.4 of this chapter.

Pankhurst in a series of publications (Pankhurst, 1971; 1975b; 1978a; 1983c; 1988c; Pankhurst & Aitchison, 1975a) has developed a suite of programs, PANKEY, that produce taxonomic products. The use of PANKEY is reviewed by Pankhurst (1986). Initially Pankhurst used his own data format for the basis of PANKEY, but this format was superseded by the conversion of his programs to use a version of the DELTA data format. The PANKEY programs currently available are listed by Pankhurst (1988a):

"1. KEY3M2. Construction of diagnostic keys. Reads DELTA data and constructs keys without user guidance.

2. KCON1. Interactive key construction. Writes keys with the user making his own choice at every stage to produce a very polished key.

3. ONLIN6. Interactive expert identification program. Question and answer at the keyboard. Now includes colour graphics character images.

4. DESCRIP3. Converts DELTA to written descriptions in natural language.

5. SPD1. Finds diagnostic descriptions, the smallest character set which will distinguish a taxon from all others.

6. MATCH. Uses similarity coefficients to make an accurate comparison, suitable for numerous similar taxa.

7. Conversion programs SC3 to give a matrix of similarity coefficients for cluster analysis, and DELPAUP1 to convert to PAUP format for cladistics.

8. DEDIT, DELTA editor. Current version is simply a utility for re-ordering and deleting characters and tidying up data."

Pankhurst (1988b) adds that he has two further programs, PREP and COND4 which are DELTA data capture programs.

During the exemplar revision the Dallwitz suite of taxonomic product generation programs were used. The reason for this choice is twofold: firstly Dallwitz's programs are accompanied by a comprehensive user's guide and secondly this system was being extensively used by a colleague in the Department of Biology, Southampton (Mr. R. Crust), so permitting access to advice if unforseen problems arose.

10.3 Exemplar Revision Database Management System

The following section details how contemporary computer aided technology was used during the revision process, i.e. the design and implementations of the database management system (DBMS), the use of the computer aided product generation programs and the use of phenetic analysis packages. This use of computer technology was exploratory and so, by definition, the way in which the technology was applied is not necessarily the most economic or utilitarian. A posteriori the computer aided techniques experimented with during the revision could have been more elegantly implemented. So what follows is essentially an experiment with computer aided techniques and not a "computer aided revision paradigm". The latter will be discussed in the Chapter Eleven.

Discussion of the revision database system can essentially be split into five components: the database files,

the structure and links between the database files, the database programs, the computer aided product generation and finally the overall database system implementation. The latter component will attempt to explain the interaction between the four preceding components within the process of the exemplar revision. Thus this section of the thesis is subdivided into five sections in which specific parts of the database system are discussed. It should be noted, however, that the components are essentially as inter-linked as pieces in a puzzle, as will be shown in the final section.

10.3.1 Database File Structure

As discussed in detail in Chapters Two and Three, the revision process involves dealing with four types of data: curatorial, descriptive, bibliographic and nomenclatural. So this section will be subdivided into four, each sub-section will correspond to the discussion of the database file structures for one of the particular kinds of data.

It is stressed in Chapter 3 that no two revisions use exactly the same methodology and no two revisions will base their classification on the same mixture of sources of taxonomic evidence. So the files detailed in this section of the thesis are those used in this case, not necessarily those that a theoretically ideal revision would use.

Following initial familiarization with the DBMS, dBASE ll was used to set up the database files that were used for data storage and retrieval during the course of the revision. Files were created and managed as described in Ratliff (1982). The database files detailed below are essentially revision products, which summarise the results of the research. This emphasises the dynamic nature of the revision process. Not all the data components could be identified at the beginning of the project and so while experimenting with the database techniques during the revision the dBASE file structures evolved. dBASE files are relatively malleable and so

structures were altered as requirements and conceptions changed. The database file structures that follow are those structures in use at the completion of the project. It is important to appreciate that the files detailed below, though discussed individually, overlap and interlink and so should not be considered as isolated data files.

Before detailing individual dBASE files, the file format should be introduced. A dBASE II file is basically a rectangular table of fields by records. Each record in a given file has identical characteristics. It contains a number of fields of data, each field is named and has a fixed length and type (character or numeric). When discussing the files used during the revision a standard format will be used, the overall contents of the file will be outlined, the contents of individual fields will be explained, followed by a summary of the file structure and sample of the first five file records taken direct from dBASE. In the latter in order to aid recognition of individual fields and records each field will be separated by a " / " and records will be separated by a blank line.

A. Curatorial Structure

The dBASE file **SPECIMEN** contains curatorial information on the herbarium specimens used during the exemplar revision. SPECIMEN contains 2008 records for the specimens borrowed from BM, CAI, CAIM, E, ERE, G, HUJ, K, LE, MADM, MEXU, MO, MPU, RNG, W and WIR, as well as those held at SPN. Each specimen was given a unique accession code, included in the field ACCNOS, e.g. S01001, S34007, S64099: S = specimen; 01, 34, 64 = taxon identification code at the beginning of the project; 001, 007, 099 = number of that specimen of that taxon.

BEGNOS and ENDNOS relate to taxon acceptance and specimen identification, i.e. the taxon that each specimen belongs to at the beginning and at the end of the study. The specimen may be identified as belonging to taxon A at the beginning of

the revision. Following the revision this taxon may be split into taxon A and taxon B or taxon A may be regarded as synonymous with taxon X, so that taxon A no longer exists. The difference between beginning and end taxa is indicated in the code, so that all BEGNOS have the prefix B and end taxa codes the prefix T, e.g. B010, B230 and B650 are beginning taxa codes, and T010, T230 and T650 are end taxa codes. This method of having a beginning and ending taxon code enables the taxonomist to keep track of changes in the taxon circumscription and its relationship to specimen identification. The contents of the other fields is explained below.

The structure and a sample print-out of SPECIMEN is provided below. The file SPECIMEN.DBF is included in Appendix 7 on Disc 1:

SPECIMEN file structure -

Structure for file: C:SPECIMEN.DBF
Number of records: 02008
Date of last update: 01/21/89
Primary use database

Fld	Name	Type	Width	Dec	Field Name Explanation
001	ACCNOS	C	006		- specimen accession code
002	BEGNOS	C	004		- beginning taxon code
003	ENDNOS	C	004		- end taxon code
004	HERBCODE	C	004		- loan herbarium code
005	COLLNAME	C	030		- collector's name
006	COLLCODE	C	003		- collector's name code
007	COLLNOS	C	015		- collection number
008	VEGCHAR	C	001		- vegetative char. presence
009	INFCHAR	C	001		- inflor. char. presence
010	LEGCHAR	C	001		- legume char. presence
011	SEDCHAR	C	001		- seed char. presence
** Total **			00070		

Sample of first five records from SPECIMEN -

S01001 / B010 / T010 / E / Davis / 122 / 42294 / + / + / + / -

S01002 / B010 / T010 / K / Haussknecht / 233 / 24-4-1865 / + / + / + / +

S01003 / B010 / T010 / K / Samuelsson / 446 / 3671 / + / - / + / +

S01004 / B010 / T010 / BM / Davis et Hedge / 127 / 28087 / + / - / + / -

S01005 / B010 / T010 / BM / Davis et Hedge / 127 / 27919 / + / + / + / -

File **HERBARIA** contains information on the addresses of the herbaria from which specimens were borrowed. HERBARIA is related to SPECIMEN, via the field HERBCODE which denotes the Index Herbarium (Holmgren & Schofield, 1979) code for that herbarium. The field HERBARIA matches these codes to the address of the herbarium.

The structure and a sample print-out of HERBARIA is provided below. The file HERBARIA.DBF is included in Appendix 7 on Disc 1:

HERBARIA file structure -

Structure for file: C:HERBARIA.DBF
Number of records: 00018
Date of last update: 03/12/88
Primary use database

Fld	Name	Type	Width	Dec	
001	HERBCODE	C	004		- Index Herbarium code
002	HERBADDS	C	150		- Address of herbarium
** Total **			00155		

Sample of first five records from HERBARIA -

BM / British Museum (Natural History), Cromwell Rd, London, SW7 5BD, England, U.K.

CAI / Herbarium, Department of Botany, Faculty of Science, A'in Shams University, Abbassia, Cairo, Egypt.

CAIM / Herbarium, Flora and Phytotaxonomy Researches, Ministry of Agriculture, Dokki, Cairo, Egypt.

E / Herbarium, Royal Botanic Garden, Inverleith Row, Edinburgh, EH3 5LR, Scotland, U.K.

ERE / Dept. Of Plant Taxonomy & Geography, Botanical Institute of the Academy of Science of the Armenian SSR., Yerevan 63, 375063, U.S.S.R.

File **COLECTOR** links the name(s) of the collector(s) of each specimen with a code which represents each combination of collectors. COLECTOR relates to SPECIMEN via the field COLLCODE.

The structure and a sample print-out of COLECTOR is provided below. The file COLECTOR.DBF is included in Appendix 7 on Disc 5:

COLECTOR file structure -

```
Structure for file: C:COLECTOR.DBF
Number of records: 00711
Date of last update: 03/12/88
Primary use database
Fld   Name    Type   Width   Dec
001     COLLCODE     N     003       - collector(s) name code
002     COLLNAME     C     030       - collector(s) name
** Total **              00034
```

Sample of first five records from COLECTOR -

1 / A.,C. et W.

2 / Adamovic

3 / Adamson

4 / Adey

5 / Ahles et Jackson

B. Descriptive Structure

MORPDATA is a composite of seven files. These files contain 1539 records of specimen score data produced from the morphological scoring of herbarium specimens. The Subgeneric character set contained 171 characters but as dBASE II has a limit of 32 fields per file and as each character was given a unique field, the scores for one specimen could not be placed in a single file. Each of the seven files has a similar structure, the specimen accession code followed by sequential character scores. MORPDATA relates to SPECIMEN via the specimen code, though to be succinct in MORPDATA the prefix "S" is missing from the specimen code.

The structure and a sample print-out of MORPDATA is provided below. The four files DATAVEGA.DBF, DATAVEGB.DBF, DATAINFA.DBF and DATAINFB are included in Appendix 7 on Disc 2, the other three MORPDATA files DATAINFC.DBF, DATALEGA.DBF and DATASEDA.DBF are included in the Appendix 7 on Disc 3:

DATAVEGA file structure -

Structure for file: C:DATAVEGA.DBF
Number of records: 01539
Date of last update: 09/07/88
Primary use database

Fld	Name	Type	Width	Dec	
001	CS	N	005	-	specimen accession code
002	C1	N	001	-	
003	C2	N	001	-	
004	C3	N	002	-	
005	C4	N	002	-	
006	C5	N	003	-	
007	C6	N	001	-	
008	C7	N	001	-	
009	C8	N	001	-	
010	C9	N	001	-	characters scores
011	C10	N	001	-	
012	C11	N	001	-	
013	C12	N	001	-	for characters 1-19
014	C13	N	003	-	
015	C14	N	002	-	
016	C15	N	002	-	
017	C16	N	002	-	
018	C17	N	002	-	
019	C18	N	003	-	
020	C19	N	003	-	
** Total **			00039		

Sample of first five records from DATAVEGA -

1001 / 2 / 2 / 3 / 4 / 75 / 1 / 2 / 1 / 2 / 2 / 1 / 4 / 38 / 3 / 7 / 6 / 2 / 20 / 24

1002 / 2 / 2 / 0 / 0 / 0 / 1 / 2 / 1 / 1 / 0 / 1 / 0 / 55 / 2 / 6 / 7 / 2 / 35 / 46

1003 / 2 / 2 / 5 / 6 / 83 / 1 / 2 / 1 / 2 / 2 / 1 / 4 / 34 / 5 / 10 / 26 / 2 / 27 / 20

1004 / 2 / 2 / 3 / 2 / 150 / 1 / 2 / 1 / 1 / 2 / 1 / 4 / 60 / 4 / 11 / 12 / 2 / 39 / 37

1005 / 2 / 2 / 4 / 4 / 100 / 1 / 2 / 1 / 1 / 1 / 1 / 4 / 44 / 5 / 7 / 10 / 1 / 29 / 35

File **MORPCHAR** contains the character list for the 171 characters of the subgeneric character set. The field TYPE refers to whether a character is M - multistate or C - continuous, NOSTATE for a continuous character is set at a maximum value of 999 (discussed further under the program LICEA), and DB refers to which of the seven MORPDATA files contains the character scores for that particular character. MORPCHAR relates to MORPDATA in that MORPCHAR contains the character definitions for the character scores included in MORPDATA.

The structure and a sample print-out of MORPCHAR is provided below. The file MORPCHAR.DBF is included on Disc 3:

MORPCHAR file structure -

```
Structure for file: C:MORPCHAR.DBF
Number of records: 00171
Date of last update: 03/12/88
Primary use database
Fld   Name      Type    Width    Dec
001     CHARNOS       C     004        - character number
002     NAME          C     045        - character name
003     ONE           C     020        - character state one
004     TWO           C     020        - character state two
005     THREE         C     020        - character state three
006     FOUR          C     020        - character state four
007     FIVE          C     020        - character state five
008     SIX           C     020        - character state six
009     TYPE          C     001        - character type
010     NOSTATE       N     003        - nos. of character states
011     DB            C     010        - database with character
** Total **                 00184
```

Sample of first five records from MORPCHAR -

C001 / GROWTH HABIT / ERECT / ASCENDING / PROCUMBENT / M / 3 / C:DATAVEGA

C002 / STIPULE SHAPE / ENTIRE / SEMIHASTATE / SEMISAGGITATE / LACINIATE / M / 4 / C:DATAVEGA

C003 / STIPULE LENGTH / 0.5 MM / C / 999 / C:DATAVEGA

C004 / STIPULE WIDTH / 0.5 MM / C / 999 / C:DATAVEGA

C005 / STIPULE LENGTH TO WIDTH RATIO / C / 999 / C:DATAVEGA

File **GEODIST** contains information on the geographical unit (as defined in Allkin et al., 1983c) distribution of each accepted taxon and is produced by the program PHYTOGEOG (discussed below). The file is a table of countries (or subunits of countries as defined by Allkin et al.) by taxa. Within the file "-" indicates absence and a number indicates presence from that country. The latter is a count of the number of specimens for that taxon included in the study from that geographical unit. As with MORPDATA, the restriction of 32 fields per file meant that GEODIST was split into three files, GEODISTA (taxa 01 - 30), GEODISTB (taxa 31 - 60) and GEODISTC (taxa 61 - 72).

The structure and a sample print-out of GEODISTA is provided below, the three files GEODISTA.DBF, GEODISTB.DBF and GEODISTC.DBF are included in Appendix 7 and on Disc 1:

GEODISTA file structure -

Structure for file: C:GEODISTA.DBF
Number of records: 00113
Date of last update: 01/19/89
Primary use database

Fld	Name	Type	Width	Dec	
001	GEOGCODE	C	004	-	geographical unit code
002	T010	C	001	-	
003	T020	C	001	-	
004	T030	C	001	-	
005	T040	C	001	-	
006	T050	C	001	-	
007	T060	C	001	-	
008	T070	C	001	-	
009	T080	C	001	-	
010	T090	C	001	-	
011	T100	C	001	-	
012	T110	C	001	-	
013	T120	C	001	-	
014	T130	C	001	-	
015	T140	C	001	-	end taxa
016	T150	C	001	-	
017	T160	C	001	-	
018	T170	C	001	-	
019	T180	C	001	-	
020	T190	C	001	-	
021	T200	C	001	-	
022	T210	C	001	-	
023	T220	C	001	-	
024	T230	C	001	-	

025	T240	C	001	-
026	T250	C	001	-
027	T260	C	001	-
028	T270	C	001	-
029	T280	C	001	-
030	T290	C	001	-
031	T300	C	001	-
** Total **			00035	

Sample of first five records from GEODISTA -

Ab / - / - / - / - / - / - / - / - / - / - / - / - / - / - / -
Af / - / - / - / - / - / - / - / - / - / - / 2 / 1 / - / - / -
Ag / - / - / - / - / - / - / - / - / - / - / - / - / - / - / -
Ak / - / - / - / - / - / - / - / - / - / - / - / - / - / - / -
Al / - / - / - / - / - / 2 / - / - / - / - / - / - / - / - / -

File **COUNTRYS** contains the explanation of the coding for the 113 geographical units used in GEODIST, the codes follow Allkin <u>et al</u>. (1983c). GEODIST and COUNTRYS relate via the field GEOGCODE, which contains the country codes. The two letter ISO code is included in the field ISOCODE.

The structure and a sample print-out of COUNTRYS is provided below. The file COUNTRYS.DBF is included in Appendix 7 on Disc 1:

Structure for file: C:COUNTRYS.DBF
Number of records: 00113
Date of last update: 01/19/89
Primary use database

Fld	Name	Type	Width	Dec	
001	GEOGCODE	C	004		- geographical unit code
002	ISOCODE	C	002		- ISO alpha-2 country code
002	GEOGUNIT	C	035		- geographical unit
** Total **			00042		

Sample of first five records from COUNTRYS -

Ab / SA / Saudi Arabia, Kuwait & Gulf States

Af / AF / Afghanistan

Ag / DZ / Algeria

Ak / US / Alaska

Al / AL / Albania

File **ECOGEOG** contains the ecogeographic information for the herbarium specimens listed in SPECIMEN. Here ecogeographic information is defined as being the specimen passport data, details about the collection site location, ecology and geography. The files SPECIMEN and ECOGEOG have two linking fields ACCNOS and ENDNOS. There is no theoretical need for ECOGEOG to contain ENDNOS, but the dBASE II limitation, of only allowing two files open at any one time, required that ENDNOS was present in this file if the program PHYTOGEOG (discussed below) was to be used. ECOGEOG also relates to COUNTRYS as it holds the explanation to the geographical unit codes used in ECOGEOG.

The structure and a sample print-out of ECOGEOG is provided below. The file ECOGEOG.DBF is included in Appendix 7 on Discs 4 and 5:

```
Structure for file: C:ECOGEOG.DBF
Number of records: 02008
Date of last update: 01/19/89
Primary use database
Fld   Name      Type    Width    Dec
001     ACCNOS        C     006        - specimen accession code
002     ENDNOS        C     004        - end taxon code
003     COLLDATE      C     010        - collection date
004     GEOGCODE      C     004        - geog. unit of provenance
005     ISOCODE       C     002        - ISO 2-alpha country code
006     PROVINCE      C     025        - major area within country
007     LOCALITY      C     035        - locality within region
008     LATITUDE      C     008        - latitude
009     LONGTUDE      C     008        - longitude
010     ALTITUDE      C     006        - altitude
011     SOILTYPE      C     026        - soil type of coll. site
012     HABITAT       C     045        - habitat of coll. site
013     COMMENTS      C     045        - any other comments
** Total **                 00225
```

Sample of first five records from ECOGEOG -

S01001 / T010 / 01-05-1966 / Tuas / TR / Urfa / Urfa to Viransehir. / 37 10N / 39 27E / 500m / Fallow fields

S01002 / T010 / 24-04-1865 / Tuas / TR / Gaziantep / Gaziantep / 27 05N / 27 10E / 650m / Agricultural weed / Type at G

S01003 / T010 / 20-04-1933 / Sy / SY / Hama / Khan Sheikhun, Hama. / 35 28N / 36 50E / 400m / In cultivated fields

S01004 / T010 / 13-05-1957 / Tuas / TR / Urfa / Akcakale, Urfa. / 36 58N / 38 54E / 500m / Fallow fields

S01005 / T010 / 14-05-1957 / Tuas / TR / Gaziantep / Birecik, Nisib. / 37 02N / 37 58E / 450m / Vineyard weed

CHROMNOS contains information on the diploid chromosome counts for subgenus Vicia taxa. This file relates to TAXNOM2 for the end taxon identification and CHROMREF for the reference containing the chromosome count. The reference is coded as the first four letters of the author's name (or first author's name) plus the last two digits of the year of publication. All publication dates were later than 1900.

The structure and a sample print-out of CHROMNOS is provided below. The file CHROMNOS.DBF is included in Appendix 7 on Disc 6:

```
Structure for file: C:CHROMNOS  .DBF
Number of records: 00311
Date of last update: 01/20/89
Primary use database
Fld   Name    Type    Width    Dec
001     ENDNOS        C     004       - end taxon code
002     DIPNOS        N     002       - diploid chromosome count
003     REFCODE       C     006       - reference code
** Total **                 00013
```

Sample of first five records from CHROMNOS -

T010 / 14 / Plit67

T020 / 10 / Fedo69

T020 / 10 / Cinc70

T020 / 10 / Kesa82

T020 / 10 / Stan70

C. Bibliographic Structure

File **CHROMREF** is the only dBASE file to contain bibliographic data, the bulk of the references used in the course of the revision were held in a word processing (Word Perfect 5.0) file. CHROMREF contains the bibliographic detail which relates to CHROMNOS via the field REFCODE.

The structure and a sample print-out of CHROMREF is provided below. The file CHROMREF.DBF is included in Appendix 7 on Disc 6:

```
Structure for file: C:CHROMREF  .DBF
Number of records: 00097
Date of last update: 01/20/89
Primary use database
Fld   Name      Type    Width    Dec
001     REFCODE         C      006        - reference code
002     AUTHNAME        C      046        - author(s)
003     YEARPUB         N      004        - year of publication
004     TITLE           C      125        - title & journal name
** Total **                    00182
```

Sample of first five records from CHROMREF -

AlMa77 / Al-Mayah, A.A. & Al-Shehbaz / 1977 / Chromosome numbers for some Leguminosae from Iraq. Bot. Notiser, 130: 437-440.

Baks54 / Baksay, L. / 1954 / Chromosomenstudien an den ungarischen Vicia-arten. Ann. Hist.-Nat. Mus. Natl. Hungarici, Ser. Nova, 5: 139-148.

Baks56 / Baksay, L. / 1956 / Cytotaxonomic studies on the flora of Hungary. Ann. Hist.-Nat. Mus. Natl. Hungarici, Ser. Nova, 7: 321-334.

Baks61 / Baksay, L. / 1961 / A kerteszeti margitvirag polyploid sorozata. Kulonenyomat a kerteszeti kutatointezet IV. Evkonyvebol (Budapest): 517-526.

Berg58 / Berger, C.A., Witkus, E.R. & McMahon, R.M. / 1958 / Cytotaxonomic studies in the Leguminosae. Bull. Torrey Club, 85(6): 405-414.

D. Nomenclatural Structure

The nomenclatural information relating to the taxa included in the revision is held in four files TAXNOM1,

TAXNOM2, SECTNOME and GENNOME. These relate to taxon, sectional and generic names as is suggested by their titles. There are two files containing data for taxon names **TAXNOM1** and **TAXNOM2**. There is necessarily some duplication of data between these two files as TAXNOM1 relates names to accepted taxa at the commencement of the revision and TAXNOM2 relates accepted names and synonyms to taxa on completion of the revision. The two files have similar file structures, the difference being TAXNOM1 has the field BEGNOS = beginning taxon codes, while TAXNOM2 has the field ENDNOS = revision conclusion taxon codes. TAXNOM1 contains the names of the 65 taxa recognised at the commencement of the revision and TAXNOM2 contains the 72 accepted taxon names and 451 associated synonyms recognised on completion of the revision. Comparison of these two files facilitates understanding of pre and post-revision conceptions of subg. Vicia.

The taxa included in TAXNOM1 were taken from the revision of Vicia by Kupicha (1976) and from Vicieae Project pub. No.7 (Allkin et al., 1986), which together formed the starting point for the revision. Following the revision TAXNOM2 was produced, this was based on the TAXNOM1 data set which was adapted to incorporate the novel classification and expanded to include data on synonyms. The extra data were obtained from Allkin et al., (1986), from the Vicia Nomenclator (Gunn, Hollis & Bisby, in preparation), Dyomina (1973), Leokene (1973) Greuter et al. (1989) and from personal observations. The bulk of the synonyms included in the list are taken from these publications and many require valid typification.

Gunn (1981) produced the "Vicia Nomenclator", which required the collection and synthesis of a large amount of nomenclatural details. This work has not yet been completed, but many of the nomenclatural details were collected into the Gunn nomenclatural collection which provided a source for much of the data included. Bondarenko (1973) edited a

nomenclatural list for the economic groups of the Leguminosae, within which Dyomina prepared the nomenclatural data for Faba and Leokene prepared the data for Vicia. It is well established that Soviet botanists have a different conception of taxonomic rank than that current in the West. However Adamova lists proved useful in clarifying some nomenclatural problems.

TAXNOM2 evolved throughout the revision process, but was finalised following the conclusion of the taxonomic study. The nomenclatural details included relate to the accepted taxa of the proposed classification and their synonyms. The data held in the file is divided into two levels of detail. All the data (where known) are included for the seventy-two accepted taxa, the synonyms have less detailed data recorded, only the taxon name, author, publication details and basionym.

The status of an ENDNOS is coded into the ENDNOS itself, e.g. T080, where T = end taxon, 08 = V. cuspidata Boiss., 0 = this name is accepted - whereas for T654, T = end taxon, and 654 = V. chlorantha Heuffel ex Nyman, the last digit of the code is higher than 0, which indicates this name is not accepted and the basionym it relates to is T650 (V. abbreviata Fischer ex Bieberstein). So any ENDNOS code ending in a 0 is an accepted name, any other ending in a digit other than 0 is a synonym.

The field SECTCODE contains the section code which links each species, subspecies or variety with the section it belongs to. The codes for SECTCODE link TAXNOM2 to SECTNOME, the sectional nomenclatural file. Taxa are only given SECTCODE codes if they are accepted at the conclusion of the revision. GENCODE is recorded as the abbreviated generic code used in the file GENNOME. SPECNAME is the specific epithet for that species. AUTHOR refers to the author(s) of a name and the abbreviations used follow Meikle (1984) the TDWG's standard. STATUS was recorded as A = accepted or S = synonym.

Originally it was envisaged that a third category of provisional would be used for names whose position remained unclear. However, as TAXNOM2 was assembled as a revision product, no taxa were left in the provisional state. TYPETYPE was recorded as H = holotype, L = lectotype I = isotype or S = syntype and HERBCODE was the Index Herbarium (Holmgren & Schofield, 1979) code for the herbarium where the type is deposited.

TAXNOM2, via the field ENDNOS, is related to SPECIMEN, GEODIST, ECOGEOG and CHROMNOS, as ENDNOS holds the key to the identification of the accepted taxon codes. TAXNOM2 also links via the field SECTCODE to SECTNOME and via GENCODE to GENNOME.

The structure and a sample print-out of TAXNOM2 is provided below. TAXNOM1 has a similar structure to TAXNOM2 and is therefore not listed. Files TAXNOM1.DBF and TAXNOM2.DBF are included in Appendix 7 on Disc 6:

```
Structure for file: C:TAXNOM2.DBF
Number of records: 00523
Date of last update: 03/04/89
Primary use database
Fld   Name      Type    Width    Dec
001     ENDNOS        C     005       - end taxon code
002     SECTCODE      C     005       - section code
003     GENCODE       C     002       - genus code
004     SPECNAME      C     018       - species name
004     SUBSP         C     020       - subspecies name
005     VAR           C     020       - variety name
006     AUTHOR        C     050       - author(s) of name
007     DATEPUB       N     004       - year of publication
008     SOURCEPUB     C     050       - place of publication
009     STATUS        C     001       - tax.stat.accepted/synonym
010     TYPETYPE      C     001       - type of type
011     TYPECOLL      C     030       - collectors name & number
012     HERBCODE      C     005       - code for herbarium
** Total **            00216
```

Sample of first five records from TAXNOM2 -

```
T0100 / SECTP / V. aintabensis / Boiss. et Hausskn. ex Boiss.
/1872  / F. Orient. 2: 577. / A / H  / Anon. / W
```

T0101 / V. peregrina / aintabensis / (Boiss. et Hausskn. ex Boiss.) Ponert / 1973 / Feddes Repert., 83(9-10): 634. / S

T0200 / SECTH / V. anatolica / Turrill, W.B. / 1927 / Kew Bull., 1: 8. / A / H / Lindsay-15 / K

T0201 / V. hajastana / Grossheim, A.A. / 1930 / Beih. Bot. Centralbl., 44(2): 358. / S

T0300 / SECTH / V. assyriaca / Boissier, E. / 1849 / Diagn. Pl. Or. Nov. Ser. 1(9): 123. / A / H / Kotschy-213 / W

The file **SECTNOME** contains the sectional nomenclatural details for sections of subgenus Vicia.

The structure and a sample print-out of SECTNOME is provided below. The file SECTNOME.DBF is included in Appendix 7 on Disc 6:

```
Structure for file: C:SECTNOME.DBF
Number of records: 00018
Date of last update: 08/26/90
Primary use database
Fld   Name       Type    Width    Dec
001     SECTCODE      C     005        - section code
002     TAXONORDER    C     006        - taxonomic order code
003     SECORSER      C     005        - whether section or series
004     SECSERNAME    C     015        - section or series name
005     AUTHNAME      C     050        - author of name(s)
006     DATEPUB       N     004        - year of publication
007     SOURCEPUB     C     050        - place of publication
008     TYPESP        C     035        - type species
** Total **              00171
```

Sample of first five records from SECTNOME -

SECTA / I / Sect. / Atossa / (Alef.) Aschers & Graebner / 1909 / Syn. Mitteleur. Fl., 6(2): 949. / V. sepium L.

SECTA / I-A / Ser. / Pseudovicilla / Maxted / 1990 / ser. nov. / V. oroboides Wulfen in Jacq.

SECTA / I-B / Ser. / Truncatulae / (B. Fedtsch. ex Radzhi) Maxted / 1990 / stat. nov. / V. abbreviata Fischer ex Sprengel

SECTA / I-C / Ser. / Atossa / (Alef.) Aschers & Graebner / 1909 / Syn. Mitteleur. Fl., 6(2): 949. / V. sepium L.

SECTM / II / Sect. / Microcarinae / Maxted / 1990 / sect. nov. / V. dionysiensis Mout.

The file **GENNOME** contains the generic nomenclatural details for 16 genera. Vicia is the only accepted genus and the other 15 generic names used have species which are synonymous with Vicia subg. Vicia taxa. The code used for the genus is usually the first letter of the genus name, except where there is more than one genus name starting with the same letter e.g. Ervum and Ervilia.

The structure and a sample print-out of GENNOME is provided below. The file GENNOME.DBF is included in Appendix 7 on Disc 6:

```
Structure for file: C:GENNOME .DBF
Number of records: 00016
Date of last update: 02/15/89
Primary use database
Fld   Name      Type    Width    Dec
001     GENUS          C     011       - genus name
002     GENCODE        C     002       - genus code
003     AUTHNAME       C     050       - author of name(s)
004     DATEPUB        N     004       - year of publication
005     SOURCEPUB      C     050       - place of publication
006     STATUS         C     001       - tax. status
007     TYPESP         C     035       - type species
008     TYPSPTYPE      C     001       - type of type
009     TYPSPINF       C     200       - type species data
** Total **                 00355
```

Sample of first five records from GENNOME -

Vicia / V / Linn. / 1753 / Sp. Pl., 734. / A / V.sativa L. / L / vide N.L. Britton et A. Brown III Fl. N.U.S. Ed. 2(2): 408. 7 June 1913.

Arachus / Ar / Medikus, F.C. / 1787 / Vorles. Churpf. Phys.-Okon. Ges., 2: 360 / S / Ar.vicioides / Medikus / H / Arachus vicioides Medikus, nom. illeg. (Lathyrus bithynicus L.) Syn = V.bithynica L.

Atossa / A / Alefeld, F. / 1861 / Bonplandia, 9: 100. / S / Type non designatus

Bona / B / Medikus, F.C. / 1787 / Vorles. Churpf. Phys.-Okon. Ges., 2: 360. / S / B.narbonensis (L.) Medikus / H / B.narbonensis (L.) Medikus. Syn = V.narbonensis L.

Cujunie / C / Alefeld, F. / 1861 / Bonplandia, 9: 101. / S / C.grandiflora / H / C.grandiflora (Scop.) Alefeld. Syn = V.grandiflora Scop.

10.3.2 Exemplar Revision Database File Relations

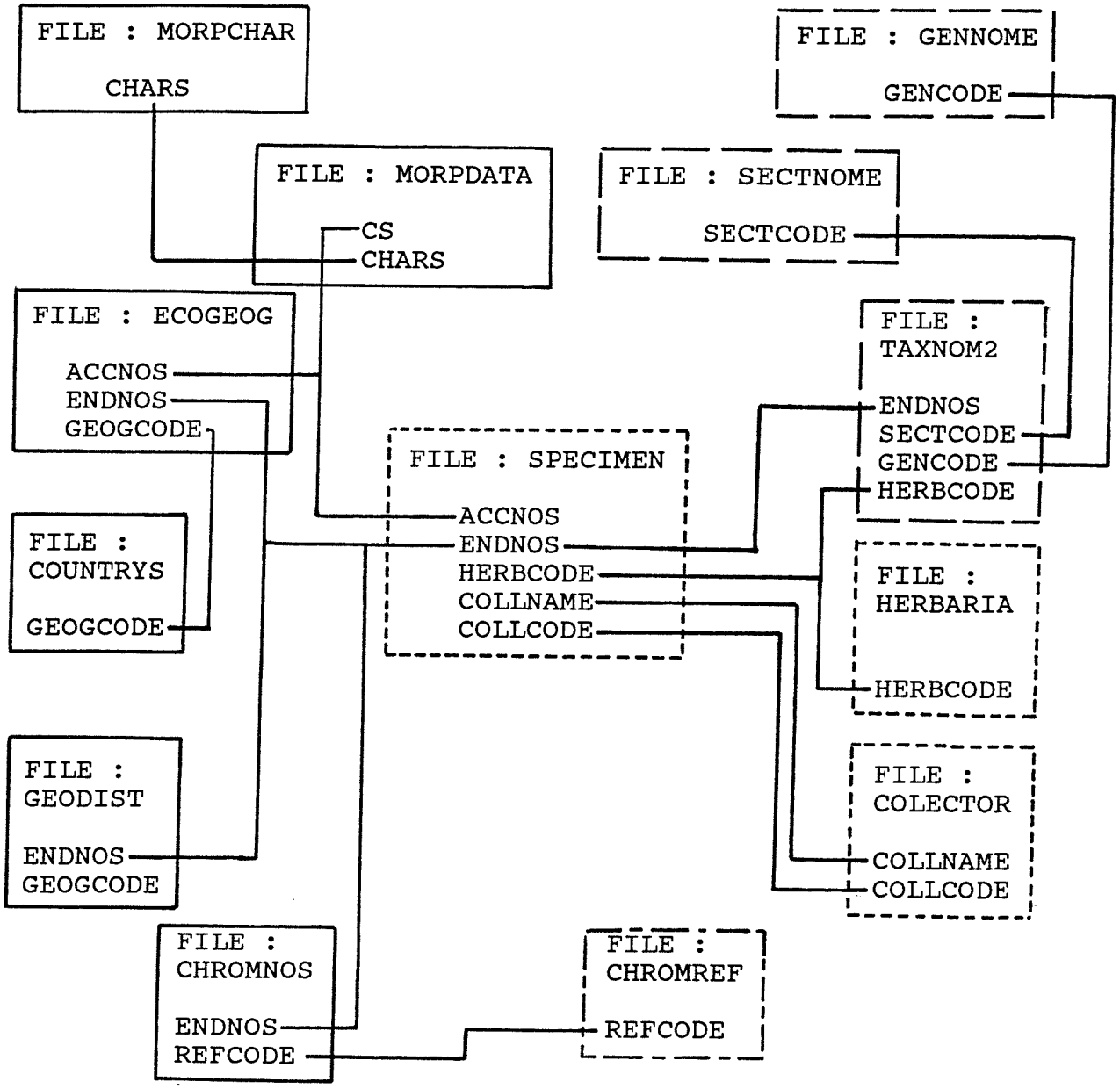
The database files should not be considered as isolated entities, they each form related units within the overall system. The revision database was designed using the relational approach discussed in Chapter Two. The twelve database files described above are therefore linked by item codes (fields). The specific relations between the twelve database files and their linking codes is shown in Figure 10.1. The figure shows fields which have relations linked by lines. Where the linked items have identical field structures, the fields in both files have identical names, e.g. the field ENDNOS occurs in five files, but in each case refers to the end taxon identification. Where a line links two fields with different names, the fields are related but not identical, e.g. field ACCNOS (S01001 or S64036) in file SPECIMEN and field CS (01001 or 64036) in MORPDATA.

One file, MORPCHAR, does not have direct links with the other files in the structure. This file does relate to MORPDATA. As mentioned in thesis section 10.3.1, MORPDATA is split into seven subfiles due to the dBASE II limitation of the number of fields per file. Thus the MORPCHAR records of which there are one for each character, acts as a morphological character list and a key to MORPDATA explaining which character scores are held in which MORPDATA subfile.

Figure 10.1. Exemplar revision file relations.
(Only fields within files that are linked to fields within separate files are shown. Lines connect linked fields).

Key to box coding

Descriptive file _____
Curatorial file - - - - - -
Nomenclatural file — — —
Bibliographic file —.—.—



10.3.3 Database Software

The use of database techniques in a taxonomic revision required the design and writing of several dBASE II programs. The three largest of these programs were CELIA, LICEA and CALIE. Each of these form one step in a chain of data synthesis stretching from the scoring of individual specimens to automatically produced descriptions for accepted taxa. All programs discussed below are listed in Appendix 8. Letters in brackets following capitalised program names indicate the program listing sequence in Appendix 8.

A. Data Capture Program - CELIA

CELIA, the data capture program (written by Maxted & Williams, 1985), was used to place the specimen score data directly into the MORPDATA file (discussed above). The program CELIA can be split into three component sub-programs MENU (a), CELIPROG (b) and SCREEN (c).

MENU is the menu program which presents the basic choices initially presented to the taxonomist.

```
********** CELIA **********
SPECIMEN CHARACTER SCORE INPUT PROGRAM

     0     EXIT TO OPERATING SYSTEM
     1     EXIT TO dBASE II
     2     COMMENCE SCORING SPECIMENS
```

If option "2" is selected then the program "CELIPROG" is called. This program initially asks the taxonomist to input a specimen code number. The specimen code number entered must contain only five digits or the program will refuse to accept the number.

```
Please input specimen number:S
```

The specimen code number is then checked against the specimens previously scored and held in the MORPDATA file. If the

specimen has been previously scored, the program will inform the taxonomist,

```
Specimen previously scored
```

then ask if you wish to score further specimens.

```
DO YOU WISH TO SCORE ANOTHER SPECIMEN y/n
```

If the specimen has not been previously scored then a blank record is added to the DATA file and the specimen number is inserted. CELIPROG then calls the SCREEN program, which produces a sequential character list composed of seven vegetative (VEG1 - 7), eleven inflorescence (INF1 - 11), five legume (LEG1 - 5) and five seed (SED1 - 5) character screens. Programming these screens is a relatively complicated procedure in dBASE II and so the dBASE II utility program ZIP was used to 'draw screens'. These screens are automatically coded, so reducing the complex programming required. An example of the first screen character score sheet is provided in Figure 10.2.

As each screen is seen, the program initially asks whether the taxonomist wishes to enter data for that group of characters. If no data are to be entered then the program will skip to the next screen.

```
Enter any key to score data or 0 to skip a screen
```

Whether the screen is skipped or data are entered, "0" is stored to each data variable, "0" being the missing data signifier. Each character is listed with its character number, character states and a place to enter data for the specimen being scored. If the character can be scored from the herbarium specimen, the "0" is replaced by the score for that character. Once the individual screen has been completed

Figure 10.2. First SCREEN data capture screen.

Enter any key to score data or 0 to skip screen

SUBGENERIC CHARACTER SET

No.	Characters	State	Description
1	Growth habit	1 2 3	Erect Ascending Procumbent 0
2	Stipule shape	1 2 3	Entire Semi-hastate Semi-sagittate 0
3	Stipule length (A)		0.5mm 00
4	Stipule width (B)		0.5mm 00
5	Stipule length-width ratio (A divided by B)		000
6	Stipule edge form	1 2	Entire Uneven with swollen hairs 0

and the scored data items entered in MORPDATA, the SCREEN program calls up the next SCREEN sub-program. This process continues until the final SCREEN sub-program SED5 is completed and then CELIPROG prompts the taxonomist by asking:

```
********** CELIA **********
SPECIMEN CHARACTER SCORE INPUT PROGRAM
DO YOU WISH TO SCORE ANOTHER SPECIMEN y/n
```

If the answer is "yes" the process of specimen scoring via SCREEN is repeated or if "no" MORPDATA is reindexed and CELIA is completed.

B. Specimen to Taxon Data Synthesis Programs - LICEA / CALIE

A central innovation of the project was the attempt to produce generalised information about taxa direct from details of representative specimens. The two programs involved in this data synthesis process are LICEA (d) and CALIE (e). LICEA uses the specimen character scores held in MORPDATA for a taxon and synthesises general information for that taxon. It uses the morphological specimen character score data to produce taxon descriptions. CALIE then takes the LICEA output files for a taxon and converts them to DELTA format which can subsequently be used in taxonomic product generation.

LICEA was written by Maxted & Williams in 1986. It requires two basic input files: SPECIMEN contains the information concerning specimen identification and MORPDATA contains the specimen score data that are used in the synthesis of taxon based information. After calling up LICEA the program prompts the taxonomist and asks whether:

```
********** LICEA **********
Do you wish to produce descriptions for a
Series of taxa, or One particular taxon ?
```

The program then prompts one to enter the taxa codes for the taxa which are to have synthesised descriptions produced. So if, for example, one wished to synthesise descriptive data for the first sixteen accepted taxa from V. aintabensis (taxon code T010) to V. galilaea (taxon code T160), one would be presented with this screen:

```
********** LICEA **********
Please input starting taxon code T00
Please input finishing taxon code T00
```

The cursor is under the starting taxon code waiting for one to enter the code. When entered the cursor would move to the finishing taxon code position and wait for entry of the finishing taxon code, as shown below:

```
********** LICEA **********
Please input starting taxon code T01
Please input finishing taxon code T16
```

It should be noted that the taxon codes, as discussed above, have three digits, but in this case the final digit of the taxon code is not used. The third digit is used to indicate whether a name is accepted or is regarded as a synonym. As only accepted taxa have representative specimens scored, so the final digit of the code is superfluous.

Once the scope of the data synthesis has been set, the program works iteratively to produce descriptive data for the included taxa via a series of program loops. These can best be understood by referring to the program summary provided in Figure 10.3. The outer loop is the taxon loop, within which the specimen loop and the character loop exist. Firstly, LICEA takes the data for character 1, specimen 1, taxon 01, then character 2, specimen 1, taxon 01 and so on. In the example cited above this would continue until character 171, specimen 52 of taxon 16.

Figure 10.3. LICEA program summary

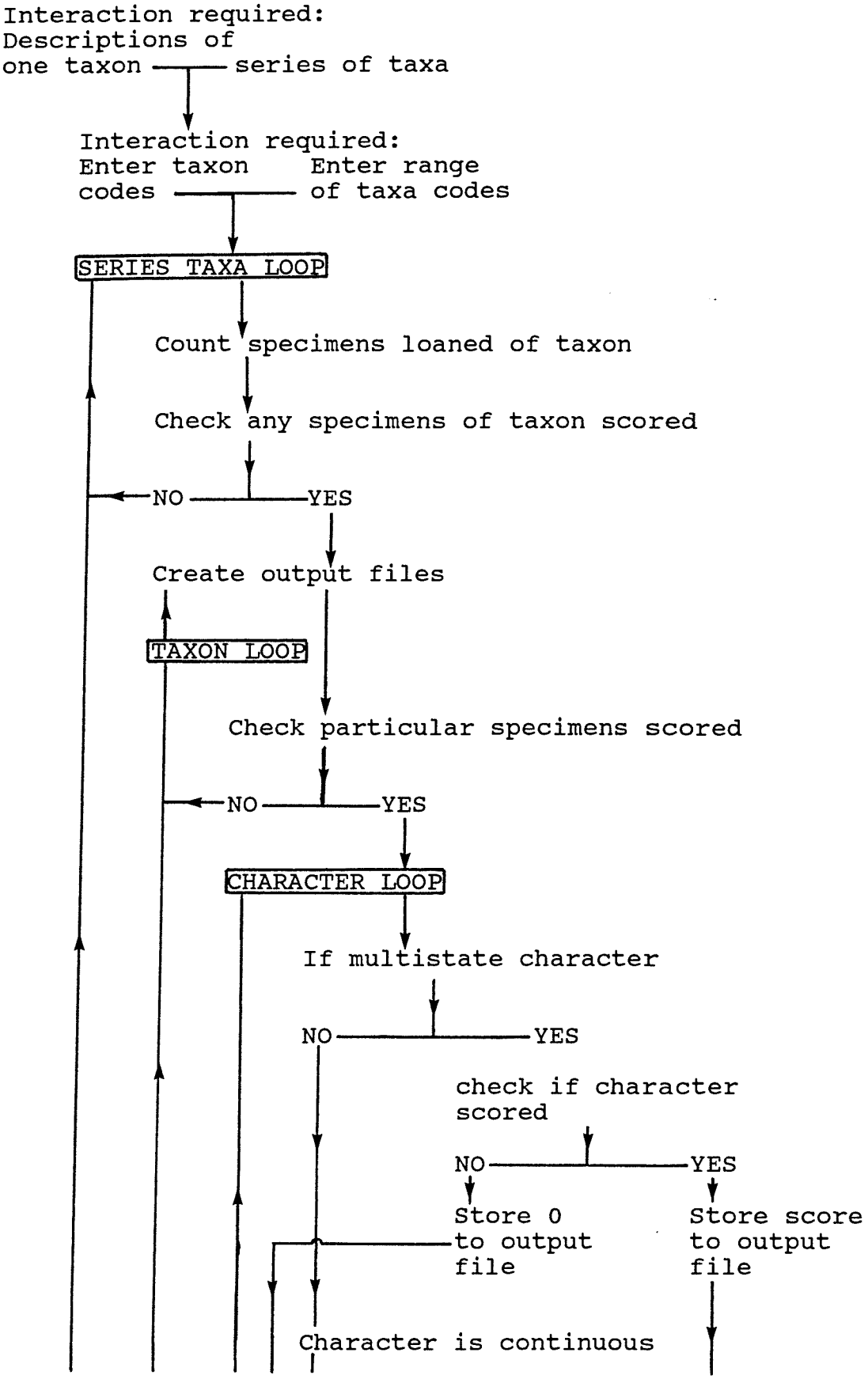
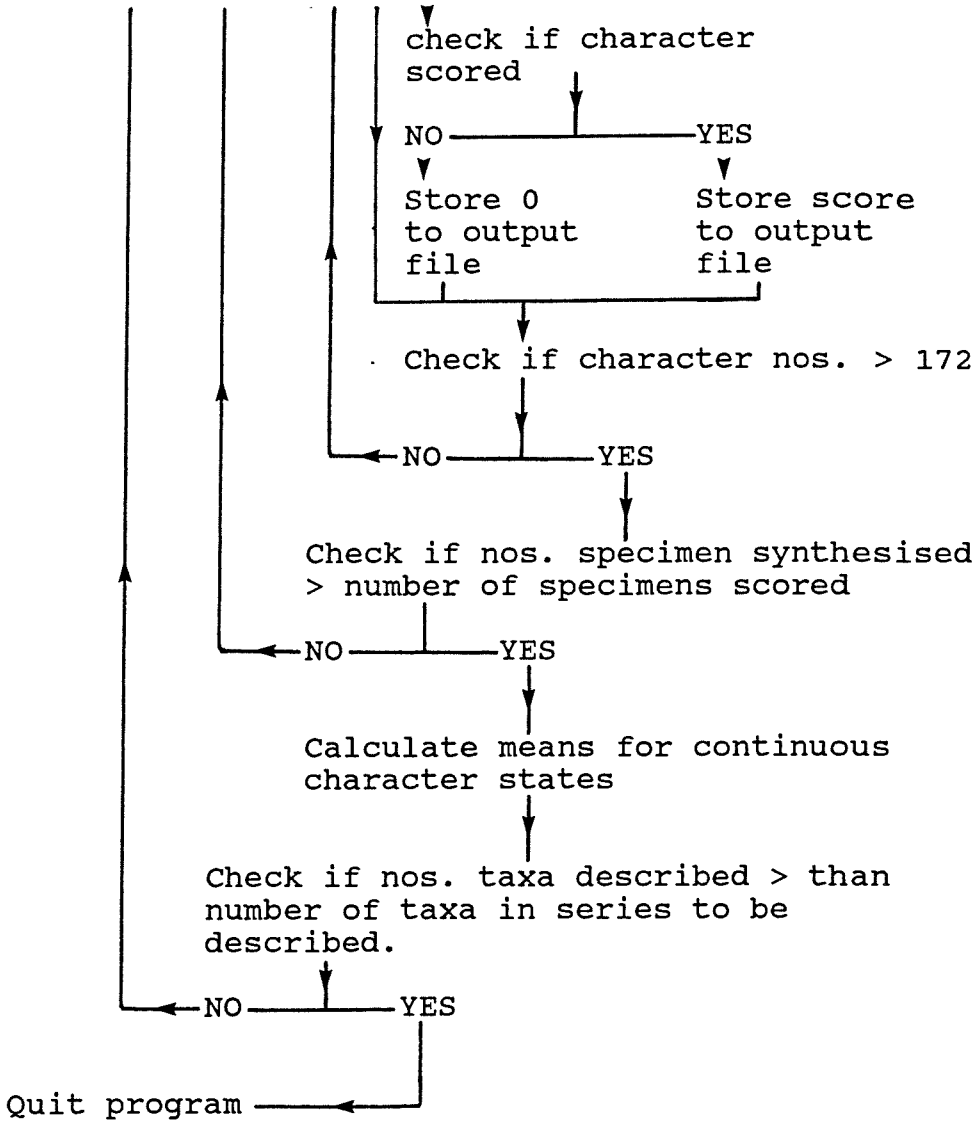


Figure 10.3. Continued.



The program stores the variation pattern for each taxon, representative specimen and character, thus using representative specimen data to synthesise information about the taxon. This process is complicated by the use of two basic character types, qualitative and quantitative, both of which require a different approach. As the manner of synthesis is different for the two character types, LICEA produces two output files for each taxon, one for each character type. To explain how the synthesis is actually undertaken examples for each will be provided and discussed.

For the qualitative character example, character 68 standard face colour is used:

68	Standard face colour	1 2 3 4 5 6 7	cream yellow yellow-pink yellow-green lilac violet purple

For Vicia sativa subsp. sativa the scores for this character are:

Specimen Code No.	Score	Specimen Code No.	Score	Specimen Code No.	Score
59001	0	59013	5	59025	5
59002	5	59014	0	59026	0
59003	0	59015	0	59027	5
59004	0	59016	5	59028	0
59005	5	59017	5	59029	5
59006	5	59018	5	59030	5
59007	5	59019	5	59031	5
59008	0	59020	5	59032	5
59009	0	59021	5	59033	5
59010	5	59022	5	59034	5
59011	5	59023	1	59035	5
59012	5	59024	5	59036	5
				59037	0

In total it can be seen that there are thirty seven specimens of this taxon. Ten have no scores recorded (missing data),

one has cream and the remaining twenty six have lilac standard face colour. These results are synthesised into the following example of the output file for multistate characters:

CHARNOS	MISS	ONE	TWO	THREE	FOUR	FIVE	SIX	SEVEN	EIGHT
C068	10	1	0	0	0	26	0	0	0

For the quantitative character example, character 124, legume length is used:

124	Legume length		mm.

For Vicia sativa subsp. sativa the scores for this character are:

Specimen Code No.	Score	Specimen Code No.	Score	Specimen Code No.	Score
59001	23	59013	47	59025	0
59002	0	59014	38	59026	41
59003	38	59015	35	59027	40
59004	42	59016	0	59028	43
59005	0	59017	39	59029	0
59006	0	59018	38	59030	0
59007	0	59019	35	59031	0
59008	48	59020	42	59032	0
59009	44	59021	25	59033	0
59010	0	59022	0	59034	0
59011	0	59023	37	59035	37
59012	0	59024	0	59036	42
				59037	27

As with the qualitative character there are thirty seven specimens of this taxon. There are seventeen which have no scores recorded and twenty for which scores were available. The minimum legume length recorded for Vicia sativa subsp. sativa was 23 mm., the maximum was 49 mm., the total length of all the legumes measured was 789 mm. and the mean length for Vicia sativa subsp. sativa recorded is 39 mm. These results are synthesised into the following example of the output file for continuous characters:

CHARNOS	MISS	MIN	MAX	MEAN	NUMBER	TOTAL
C124	17	23	49	39	20	789

The second program in the "Specimen To Taxon Description Synthesis" is CALIE, the data conversion program, which takes the two files produced for each taxon by LICEA and converts them into one file containing DELTA format data.

CALIE uses the same initial code as LICEA to question the scope of the data conversion; is one taxon being converted or many, and what are the taxon codes for the taxa being converted? Examples of the first ten records for the two Vicia sativa subsp. sativa LICEA output files are provided below:

M59 - qualitative LICEA output / CALIE input file

QUALITATIVE CHARACTERS

	MISS	1	2	3	4	5	6	7	8
C001	0	0	37	0	0	0	0	0	0
C002	0	2	17	18	0	0	0	0	0
C006	0	22	15	0	0	0	0	0	0
C007	0	2	5	11	19	0	0	0	0
C008	0	21	12	3	1	0	0	0	0
C009	0	36	1	0	0	0	0	0	0
C010	0	6	31	0	0	0	0	0	0
C011	0	37	0	0	0	0	0	0	0
C012	0	0	0	9	28	0	0	0	0
C020	0	0	1	4	16	16	0	0	0

C59 - quantitative LICEA output / CALIE input file

QUANTITATIVE CHARACTERS

	MISS	MIN	MAX	MEAN	N	TOTAL
C003	0	6	24	12	37	474
C004	0	3	20	10	37	399
C005	0	57	275	134	37	4990
C013	0	23	79	49	37	1838
C014	0	1	7	3	37	125
C015	0	6	17	9	37	367
C016	0	11	38	19	37	709
C017	0	2	13	5	37	186
C018	0	12	89	42	37	1580
C019	0	15	76	44	37	1652

CALIE uses the character state presence for the qualitative characters and the minimum and maximum records for the quantitative characters for inclusion in the DELTA format. Firstly, CALIE abstracts the required data from the two LICEA output files, links the abstracted data to their character numbers, then orders the characters on character number, copies the DELTA data for each taxon into a DELTA file and finally copies all the separate taxon DELTA files into one taxon series DELTA file. The DELTA file includes the DELTA format data for each taxon listed sequentially. An example of a sample of the CALIE output file for the first twenty characters of V. sativa subspecies sativa is shown below:

LINE NOS	CHARACTER NUMBER, CHARACTER STATES
59.01	1,2 2,1/2/3 3,6-24 4,3-20
59.02	5,57-275 6,1/2 7,1/2/3/4 8,1/2/3/4
59.03	9,1/2 10,1/2 11,1 12,3/4
59.04	13,23-79 14,1-7 15,6-17 16,11-38
59.05	17,2-13 18,12-89 19,15-76 20,2/3/4/5

The CALIE output files can be used directly as DELTA "items files".

C. Geographic Distribution Program - PHYTOGEOG

PHYTOGEOG (f) is a dBASE II program written by Maxted in 1986, which produces distribution information for taxa based directly on the distributional data for specimens of that taxon. PHYTOGEOG uses two input files: ECOGEOG and GEODIST. Due to the limit of 32 fields in dBASE II file, GEODIST is for practical reasons split into three subfiles (GEODISTA, GEODISTB and GEODISTC).

The file GEODIST is a table of geographical units by taxa, where each taxon is a field and each country a record. GEODIST is initially set so that each unit in the country matrix is set to minus ("-"). PHYTOGEOG uses the ECOGEOG field ENDNOS to distinguish which specimens belong to which taxon and then stores the number of specimens encountered in

that geographical unit for each taxon, so that a taxon not found in a geographical unit will register absent ("-"). A crude estimate of the concentration of taxa in that particular geographical unit can be obtained from the number of representative specimens from that geographical unit. A high concentration of representative specimens roughly reflects a high taxon concentration in that geographical unit. However, it must be emphasised that this estimate of concentration is obviously crude, because the specimens sampled during the revision may not be completely representative of the taxon's distribution, i.e. Britain is sampled very well in comparison to Albania.

File GEODIST prior to running PHYTOGEOG is shown below:

		TAXA		
COUNTRY	T010	T020	T030	T040
Cy	-	-	-	-
Ir	-	-	-	-
Le	-	-	-	-
Sy	-	-	-	-
Tuas	-	-	-	-
Rssa	-	-	-	-

File GEODIST after running PHYTOGEOG is shown below:

		TAXA		
COUNTRY	T010	T020	T030	T040
Cy	-	-	1	-
Ir	2	2	1	-
Le	1	-	-	-
Sy	2	-	2	-
Tuas	4	6	2	3
Rssa	-	2	-	6

From this example it can be seen that V. aintabensis is found in Iran, Lebanon, Syria and Turkey in Asia; V. anatolica is found in Iran, Turkey in Asia and Soviet Asia; V. assyriaca is found in Cyprus, Iran, Syria and Turkey in Asia and V.balansae is found in Turkey in Asia and Soviet Asia. Using the specimens sampled, there is a crude indication that the highest concentration of V. anatolica is found in Turkey in

Asia and the highest concentration of V. balansae is in Soviet Asia.

D. Label Production Programs - LABELS

LABELS is a herbarium utility program used to produce various types of labels for use in the taxonomic environment. During the course of the revision over 1000 herbarium and seed specimen collections of Vicia subgenus Vicia taxa were made from Britain, France, Spain, Syria, Turkey the Soviet Union and Yugoslavia. To avoid confusion of specimens careful labelling is required. LABELS was written by Maxted in 1986, it is menu driven (MENULAB, g) and produces herbarium (HERBLAB, h), determination (DETLAB, i) and seed container labels (SEEDLAB, j).

The menu options screen presents the herbarium curator with the following options:

```
            MENU OPTIONS

0     EXIT TO OPERATING SYSTEM
1     EXIT TO dBASE II
2     PRODUCE HERBARIUM SPECIMEN LABELS
3     PRODUCE DETERMINATION LABELS
4     PRODUCE SEED CONTAINER LABELS

PLEASE SELECT OPTION :
```

The separate label producing programs are called from MENULAB. HERBLAB is slightly more complicated than the other two label producing programs as it uses two input files, one containing specimen accession data and the other specimen site location data. However, HERBLAB, DETLAB and SEEDLAB are basically very similar programs. They pass through the accession file searching for tagged records. If a tagged record is located, the data are stored in memory variables and then printed sequentially. The program then skips through the file until another tagged record is encountered or the end of the file is reached. If HERBLAB encounters a tagged record, it stores the

required accession data in memory variables then switches to the site detail file, stores the required site details in memory variables and then prints out the herbarium specimen label. Examples of the three labels for the type collection of V. eristalioides are provided below:

Example of determination label produced by LABELS

```
Maxted, Kitiki & Allkin  4256

Vicia eristalioides Maxted

Det.  Nigel Maxted     05.viii.87
```

Example of seed container label produced by LABELS

```
VICIEAE  SEED  COLLECTION
    Turkish Germplasm

    Vicia eristalioides

Our Accession Number : 877531
Original Seed.     Year : 1987
```

Example of herbarium label produced by LABELS

```
                University of Southampton

                     Flora of Turkey

Vicia eristalioides Maxted

Province : Antalya   Nearest Settlement : Cavus
FT Grid Sq : C/3   Location : Olimpos Beydaglari
National Park, 14 Kms. North west of Kumluca.

Habitat : Enclosed plantation of young pine.
Altitude 600m   Latitude : 36 17N  Longitude : 30 25E

Collectors : Maxted, Kitiki and Allkin  4256
Date : 25/04/87
```

10.3.4 Exemplar Revision Product Generation Using DELTA

The product from the series of dBASE programs CELIA - LICEA - CALIE (described above) is a series of synthesised taxon descriptions in DELTA item file format. DELTA (DEscription Language for TAxonomy) is a data format system for the concise representation and manipulation of taxonomic data. This data format system and the accompanying programs for data manipulation are discussed in detail by Watson & Dallwitz (1981) and Dallwitz & Paine (1986, 1989).

The DELTA system requires three basic files: character file, containing the characters and character states; items file containing the individual taxon descriptive data and the specification file, containing information describing the nature of the data. These three files are required for most manipulations of the DELTA data. Two sets of DELTA files were used in the revision, SUBG containing the data for the 74 species and infra-specific taxa and SS containing data for the 18 supra-specific taxa. All six files are included on disc 7. Extracts from these files, showing the type of data they contain, are provided below. A full understanding of the coding used is given in Dallwitz & Paine (1986):

CHARACTER FILE - Characters and Character States

```
*COMMENT File SUBG.CHR
*SHOW : Vicia subgenus Vicia - SUBGENERIC CHARACTER SET
*Characters List

1.01      #1.  <Growth habit>/
1.02           1.   Erect/
1.03           2.   Ascending/
1.04           3.   Procumbent/

2.01      #2.  <Stipule shape>/
2.02           1.   Entire/
2.03           2.   Semi-hastate/
2.04           3.   Semi-sagittate/

3.01      #3.  <Stipule length (A)>/
3.02           mm. long/

4.01      #4.  <Stipule width (B)>/
4.02           mm. wide/
```

ITEMS FILE - Individual Taxon Descriptive Data

*COMMENT File SUBG.ITM
*SHOW: Subgenus Vicia (Taxa) 7, December 1989.
*ITEM DESCRIPTIONS

V. aintabensis <Boiss. & Hausskn. ex Boiss.>/
1,1 2,2 3,20-50(-80) 4,1-3 5,1-4 6,75-250 7,2 8,1 9,2 10,1
11,1 12,1/2 13,1/2 14,3/4 15,32-60 16,2-6 17,6-13 18,6-26
19,1-4 20,20-39 21,20-56 22,10-61 23,2 24,2/4 25,1 26,2
27,8-14 28,1 29,1 30,2 31,1/2/3/4/5 32,1 33,1 34,1/2/3 35,1
36,1/2/3 37,1/3/4 38,2/3 39,3/4 40,3 41,1/2<rarely> 42,1 43,U
44,U 45,2-6 46,10-24 47,U 48,U 49,U 50,U 51,1 52,2/3 53,2/3/4
54,2-4 55,1-4 56,1-3 57,3-6 58,0.5-0.88 59,2 60,2 61,1 62,2
63,1 64,3 65,2/3 66,2/3/4 67,2 68,1/2 69,9-18 70,4-11 71,3-8
72,5-10 73,3-7 74,172-333 75,1-2.33 76,81-175 77,1 78,2<or
pale yellow> 79,2 80,2 81,1/2 82,1 83,2 84,1 85,3 86,8-16
87,3.5-9 88,3.5-7 89,1.5-5 90,300-700 91,70-142 92,2 93,1 94,1
95,2/3 96,2/4 97,2 98,2 99,1 100,6-11 101,2-4.5 102,3.5-7
103,2.5-3.5 104,48-71 105,83-120 106,1 107,2 108,1
109,2<rarely>/4 110,1 111,4.5-9.5 112,1-2 113,325-700 114,1
115,1 116,4-6.5 117,2.5-4.5 118,1-2 119,3 120,2 121,1 122,1/2
123,3 124,2 125,4-8 126,18-26(-40) 127,7-10 128,3-5
129,2.3-4.12 130,160-300 131,1 132,1 133,1 134,4 135,2 136,1
137,1 138,2 139,1 140,1/2 141,2 142,1 143,2/3/4 144,2/3
145,2/3 146,4 147,2 148,1 149,2-6 150,4-5.5 151,4-5.5
152,3-4.5 153,12-17 154,1-2 155,1 156,0.9-1 157,111-133
158,0.07-0.1 159,2 160,1 161,3 162,2 163,3 164,1 165,2 166,2
167,1 168,2 169,1/3 170,1/2 171,1 172,2 173,1 174,1

SPECIFICATIONS FILE - Descriptive and Directives Referring to the Nature of the Data

*COMMENT File SUBG.SPC
*SHOW: Subgenus Vicia 7, December 1989.
*MAXIMUM NUMBER OF ITEMS 75
*NUMBER OF CHARACTERS 174
*MAXIMUM NUMBER OF STATES 8
*CHARACTER TYPES 2,OM 3-5,RN 6,IN 7,OM 9-10,OM 13-14,OM
15-22,RN 24,OM 27,IN 28,OM 31-34,OM 36-40,OM 42,OM 43-46,RN
47-48,IN 49,RN 50-53,OM 54-58,RN 59,OM 63-68,OM 69-73,RN 74,IN
75,RN 76,IN 78-79,OM 82,OM 85,OM 86-89,RN 90-91,IN 92-93,OM
95-97,OM 100-103,RN 104-105,IN 108-109,OM 111-112,RN 113,IN
116-118,RN 119,OM 123-124,OM 125,IN 126-129,RN 130,IN 132,OM
134-135,OM 138,OM 140,OM 143-148,OM 149,IN 150-158,RN 159,OM
161,OM 164-165,OM 167-168,OM 172,OM
*NUMBERS OF STATES 2,3 7,4 9-10,4 13,3 14,4 23,3 24,5 28,4
29,3 31,8 32,5 33-34,3 36-40,4 42,4 50-53,4 59-60,3 63-64,3
65,5 66,4 67,3 68,4 78,8 79,7 82,4 85,3 92,8 93-94,3 95-96,5
97,3 108,5 109,4 119,3 121,3 123-124,5 132,4 133,3 134,4 135,3
138,3 140,3 141,4 143,5 144-145,4 146,6 147-148,4 159,3 161,4
163,3 164-165,4 167,3 168-169,4 170,3 172,4
*CHARACTER RELIABILITIES 1,9 2,7 3,6 4,7 5,6 6,2 7,6 9,8 10,8
11,6 12,4 13,8 14,6 17,4 21,4 22,4 23,8 24,7 26,3 27,7 28,8
29,8 30,7 34,3 35,2 36,4 37,4 40,4 41,8 42,10 43,6 44,3 46,8
47,4 48,4 49,7 51,8 52,4 53,4 54,7 55,4 56,4 57,7 58,7 59,9

60,9 61,7 64,7 67,8 69,8 70-76,4 77,6 78,8 79,8 80,6 81,7 83,8 84,8 85,3 86,7 87,4 88,4 89,6 90,4 91,4 92,8 93,8 94,8 97,7 98,3 99,3 100,7 101-109,4 112-113,4 114,7 115,7 118,8 121,4 122,6 123,7 125,6 126,8 127,8 128,4 129,7 130,3 131-136,7 137,4 139-140,7 141,4 147,8 148,4 149,6 150-151,7 152,3 153-156,7 158,7 163,4 164,7 165,7 166-168,3 170,4 171,6 172,8 173,6 174,7

Once the taxonomic data are in DELTA format they can be converted into other formats and used to manufacture taxonomic products by the programs: CONFOR, used for data reorganisation and production of natural language descriptions; KEY, used to produce tabular and bracketed keys; INTKEY, used as an interactive polyclave for identification and DIST, used to produce a distance matrix. All of these programs were used during the course of the revision. Each of these programs is used via a directive file which contains the directives required for a particular operation. Examples of the directive files (e.g. TONAT.DIR, TOKEY.DIR, TOINT.DIR, TODIS.DIR) used during the revision are included in Appendix 9 and are included on disc 6.

The natural language descriptions of all the 92 taxa included in the conspectus were produced using the program CONFOR and the directive files TONAT.DIR and SSNAT.DIR. Initially both sets of descriptions were tautological and unnecessarily prolix. However, with some reconstruction of the character file and editing of the output the descriptions were improved. An example of an edited description is provided below for Vicia sect. Atossa:

> "Life form perennial; erect, or climbing; stem slender, or stout. Stipules entire, or semi-hastate; length less than 3.5mm, or 3.5 to 5.5mm; edge entire, or with 1-2 teeth, or with 3-5 teeth. Leaf apex tendrilous, or mucronate; leaflet less than 20mm, or 20-30mm, or longer than 30mm; with 1-4 pairs, or more than 4 pairs. Leaflets symmetric; margins entire. Number of flowers per inflorescence 1 to 2, or 3 to 4, or more than 4; peduncle 3-6mm, or peduncle longer than 6mm. Calyx mouth oblique; lower tooth longer than upper; base gibbous. Pedicel shorter or equal to 3mm. Flowers shorter than 15mm, or 15 to 20mm, or longer than 20mm; standard colour yellow, or blue or purple; shape platonychioid; claw bowing absent;

upper standard surface glabrous. All petals of approximately equal length. Wing marking absent; wing limb kinking absent, or with slight kinking. Legume length less than 30mm, or 30 to 50mm; width 5 to 10mm; oblong; laterally flattened; sutures parallel, or curved; valve hairs absent; septa absent; number of seeds per legume less than 7. Seed length less than 3.4mm, or 3.5 to 6.0mm; round, or oblong; not laterally flattened; hilum over half seed circumference; lens positioned near hilum; testa surface smooth."

The production of natural language descriptions using DELTA data and CONFOR was easily achieved. The one problem encountered was the linking of characters into organ based sentences, e.g. describing plant habit, stipule, wing, etc. characteristics. CONFOR includes a LINK CHARACTERS directive for this purpose. However, when used it excluded the character description from the text produced, including only the character state description, leading to grammar and sense problems within the text. This problem was resolved by avoiding the LINK CHARACTERS directive and instead using the search and replace option of the word-processing package.

Both bracketed and tabular keys were produced for the 18 sections and series, the 38 species and the 12 species with infra-specific taxa using KEY. CONFOR converts the three basic DELTA files into KEY format files using a directive file e.g. TOKEY.DIR and these are then accessed by the program KEY with a further directive file, e.g. SUBKEY.DIR. As with the natural language descriptions, the character file was edited iteratively to produce the least tautological or prolix natural language key.

However, it proved impossible to produce a key to the species of subg. Vicia which could be used reliably to identify the taxa included. The reason was possibly the relatively high level of intra-taxon variability within subg. Vicia. This variability may be a side effect of the method of production of the taxon descriptive data, which were produced direct from the specimen descriptive data by the dBASE programs CELIA - LICEA - CALIE. Synthesising taxon

descriptions direct from specimen descriptions means that every character state recorded for representative specimens is incorporated into the taxon description. This is thorough but may not duplicate the normal taxonomic process. It could be argued that the taxonomist is actively looking for specimen associations rather than simply recording the innate pattern of variation. KEY uses the logically correct choice of characters, but cannot take account of the judgement or expertise of the taxonomist and thus often produced key dichotomies that were unexpected or which used characters with low reliabilities.

KEY also requires the continuous characters to be converted to multi-state characters using the KEY STATES directive. This presents a further problem. The range of variation must be divided up into sub-ranges which are then used as character states. This division of the range of variation is taxonomically arbitrary, i.e. the segregation of a continuous to multi-states characters may be useful in the division of two taxa, but may unnaturally split a third taxon. For this reason the keys produced by the program KEY were not used intact but formed "skeletons", which were then edited to produce the final keys included in the conspectus.

INTKEY is an interactive program for the identification of specimens. It carries out identifications by comparing the attributes of individual specimens with stored descriptions of taxa (Dallwitz & Paine, 1989). The three basic DELTA files containing the descriptive information for the 74 taxa of subg. Vicia were converted to the two INTKEY format files SUBINT.ITM and SUBINT.CHR using the program CONFOR and the directive file TOINT.DIR. These two files could then be accessed by the second directive file INTKEY.INI and the program INTKEY. The program was used with the subg. Vicia species, subspecies and variety data sets. The files INTKEY.INI, SUBINT.ITM and SUBINT.CHR are available on disc 7. INTKEY can be used, via a series of key words, to produce full

descriptions of a taxon, diagnostic descriptions, lists of differences between taxa, lists of similarities between taxa and to suggest which characters are best for identification. It can also be used to allow the user to include or exclude subsets of taxa or characters for a particular identification session, as well as other functions (see Dallwitz & Paine, 1989).

The subg. Vicia item and specification files were used to prepare the distance matrix for subg. Vicia species and infra-specific taxa. SUBG.ITM and SUBG.SPC were converted to an intermediate format using CONFOR. This intermediate file SUBG.SIM was then converted to the distance matrix using the program DIST. This process of data transformation is described by Dallwitz & Paine (1989). The distance matrix was calculated using the Gower metric. The matrix produced for subg. Vicia was then analysed using CLUSTAN procedure CLUSTER, described in Chapters Six and Seven. The results of this cluster analysis indicated an anomalous pattern of relatedness between the taxa, i.e. one that did not reflect their natural relationships. The overall pattern reflects that discussed in Chapters Seven and Nine, but individual clusters contained aberrant taxa. The reason for these anomalies was thought to be the method of conversion of the multistate and continuous characters to an intermediate format that could be used with the Gower metric. The method of computing the intermediate format is described by Dallwitz & Paine (1989) as follows:

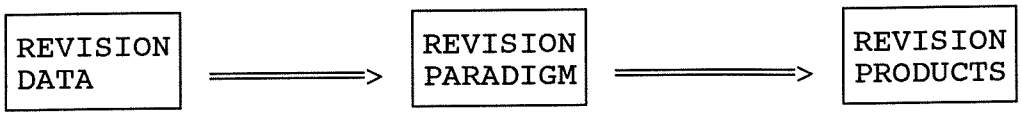
> "For ordered multistate and numeric characters, a 'central' value Xik is calculated for each masked-in item i and character k. For ordered multistate characters, Xik is the mean of all the coded values. For numeric characters, only 'normal' values are used: the 'extreme' values (those enclosed in parenthesis) are ignored. Xik is the middle normal value, if present, or otherwise the mean of the normal values."

This method of using the 'central' or mean data value may be valid with simple character sets with little internal

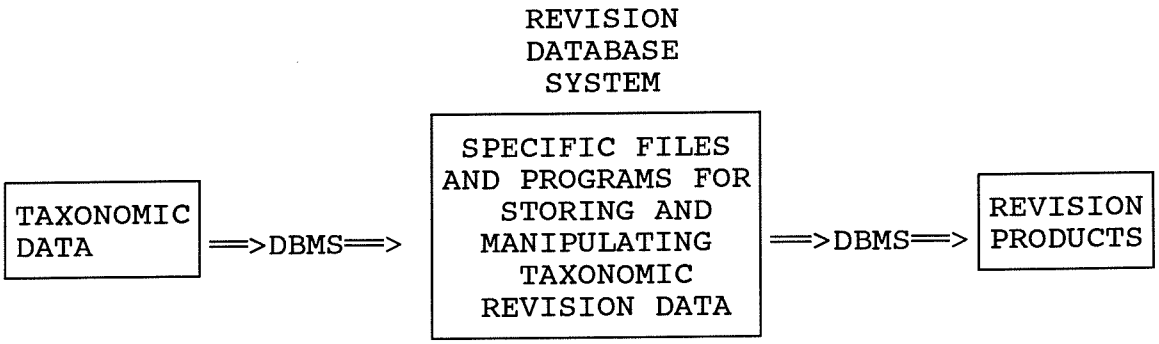
variation and normally distributed character states. As discussed above, the synthesised subg. Vicia taxon data were produced directly from the representative specimen data. The data were highly variable within taxa and the multistate or continuous characters were not normally distributed. For this reason it was decided that the assumptions made by DIST were unacceptable and this program was not used to generate the input for the cluster and ordination analysis.

10.3.5 Exemplar Revision Database System Implementation

This thesis section explains how the database files, database programs and product generation programs were used together to aid the revision process. This overall subject of system implementation will be introduced via a series of models, grading from the general to the specific. The basic processes of a taxonomic revision, as shown in Chapter Three, can be summarised in the following general revision flow model:



If database technology is applied to this generalised flow model, the following "database enhanced" revision flow model is produced.



In this model it can be seen that the taxonomist still starts with the same basic data and finishes with the same products, but the revision paradigm is mediated and enhanced by the application of database and other computer aided technology.

If the central element of the above model is expanded, the model can be used to further clarify the computer aided revision process as indicated in the model shown in Figure 10.4. This model contains five elements, the four component data type files included in a revision and the database programs used to manipulate these data. All five elements of the model sit upon the DBMS. Lines are deliberately not used

to connect files (as in the models proposed by Bisby, 1984a, Fig. 2 and 3). If connections are explicitly specified they, by implication, limit the connections to those files joined by lines. All the files included in the model are interdependent.

If the model shown in Figure 10.4 is applied to the exemplar revision, then the specific database system model used in the exemplar revision is produced, Figure 10.5. For the exemplar revision fourteen database files and five dBASE application programs were used in the DBMS environment. If this model is further developed to include the product generation programs and the products themselves, the exemplar computer assisted revision paradigm is produced, as shown in Figure 10.6. It can be seen that this model relates directly to the general revision paradigm (Figure 3.1). The specific model can be split into three components, Revision Data, Revision Database System and Revision Products. As the taxonomist requires the Revision Data and Revision Products remain unaltered by the application of database technology, the innovative element of the project is concentrated on the central Revision Database System, that is the process between the revision input and output.

The fourteen dBASE files and the five dBASE programs are shown to convert the revision data into the revision products. The phenetic analysis of MORPDATA is central to the system, as it is the results which tie the taxa, identified during the course of the analysis, to their representative specimens and subsequently enable generalisations to be made about the taxa based on the specimen data. Thus in the model MORPDATA is analysed using phenetic methods and the results are used the produce a classification of the specimens. This classification of specimens provides the basis of the conversion of the initial taxonomic hierarchy (TAXNOM1) into the novel taxonomic hierarchy (TAXNOM2). As a result of the analysis the list of and relationship between the accepted

Figure 10.4. General revision database system model.

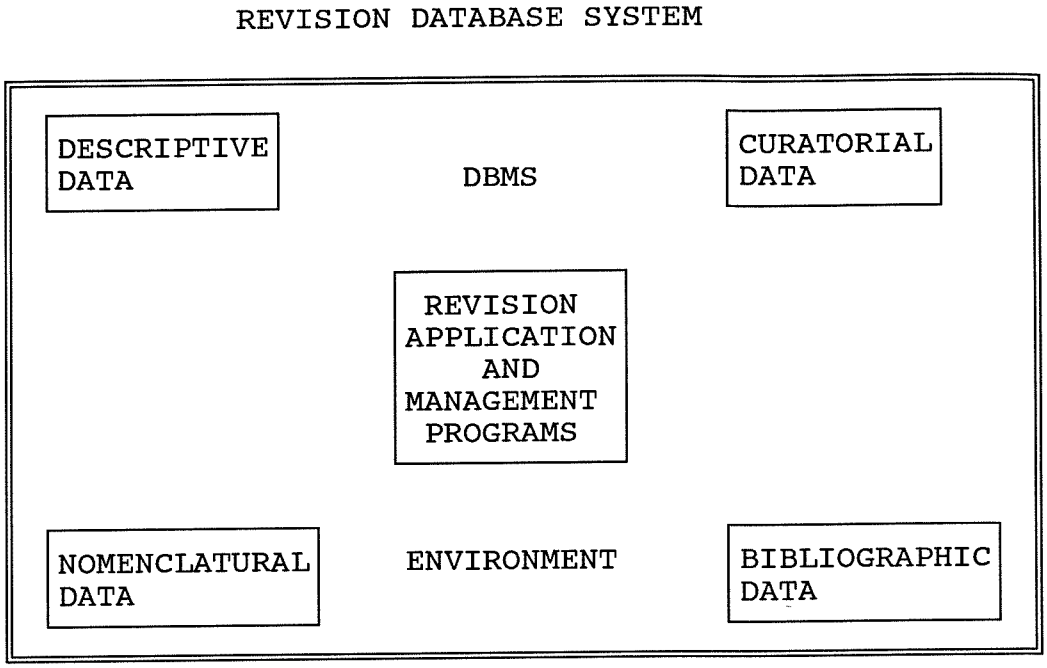


Figure 10.5. Specific exemplar revision database system model.

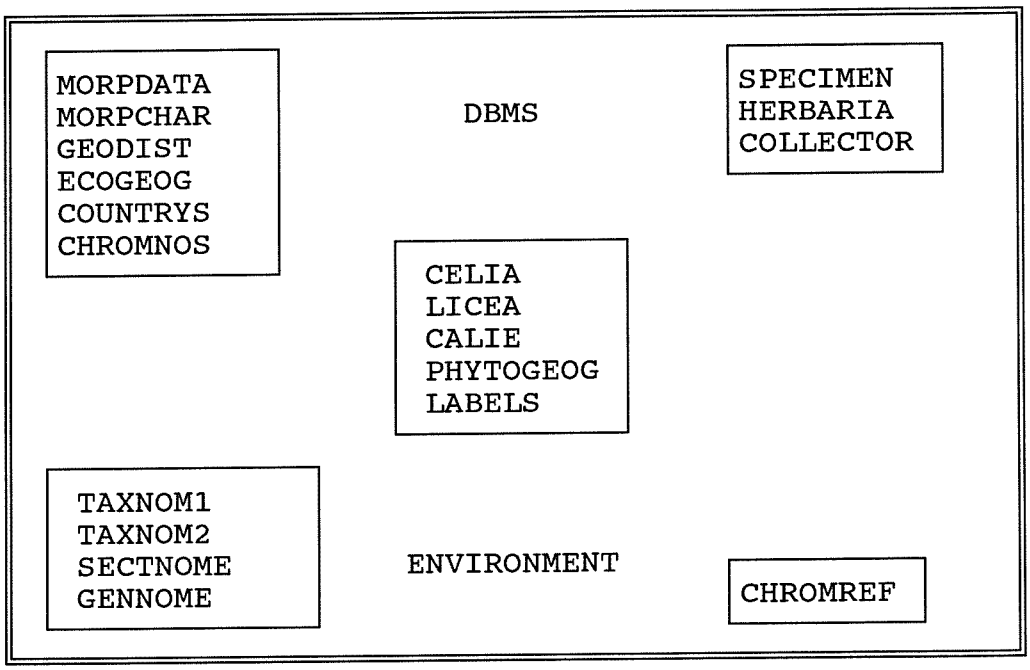
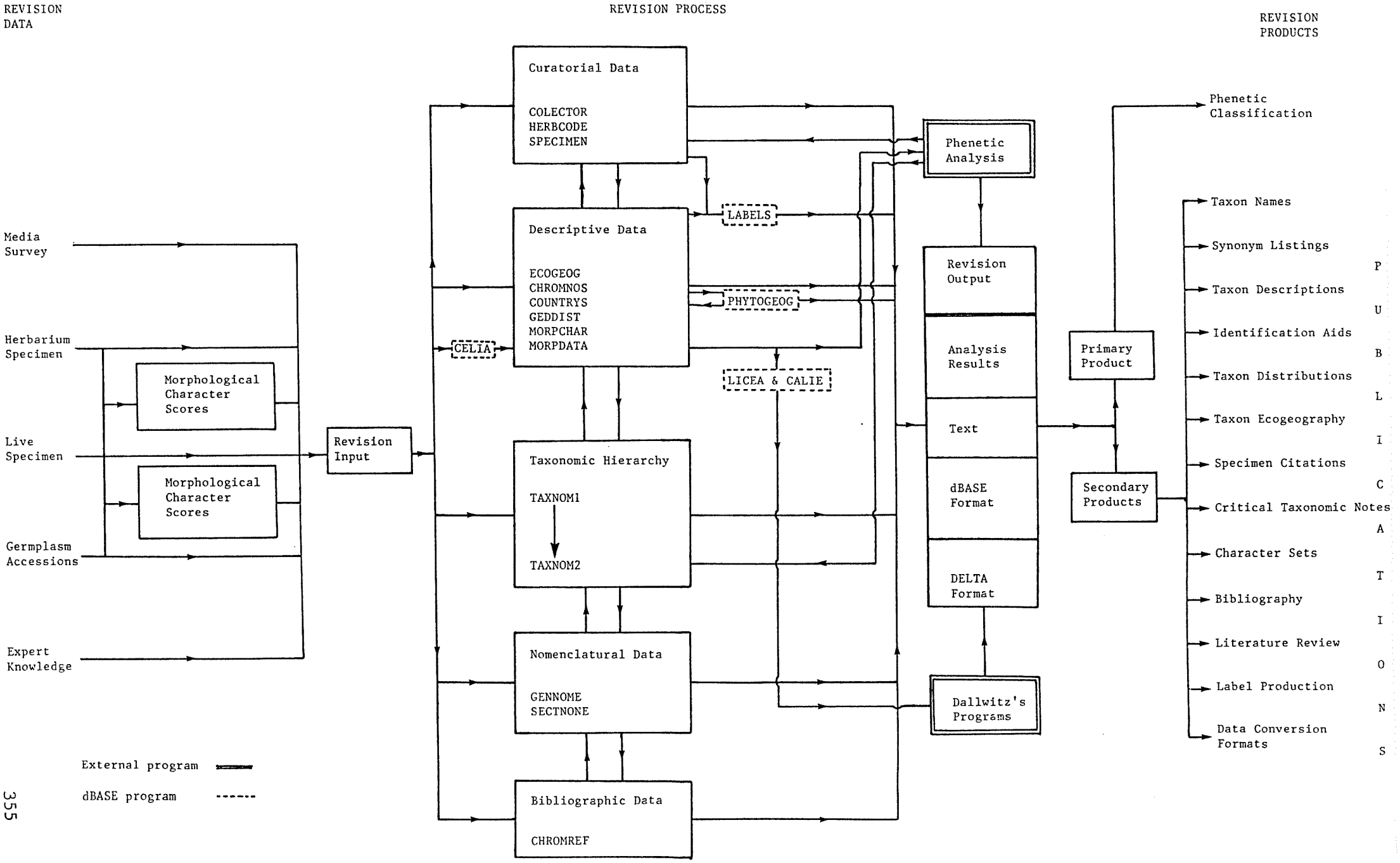


Figure 10.6 Exemplar Computer Assisted Revision Paradigm



taxa is likely to change and this is reflected in the differences between the content of TAXNOME1 and TAXNOME2. These changes will in turn affect the other database files via the field ENDNOS, so that the novel taxonomic hierarchy is inculcated into the results of the revision.

The model shows that the revision output is produced in three formats: text, dBASE format and DELTA format. The only direct output in the form of text is the herbarium, determination and seed container labels, although dBASE files can be queried or listed with text output (see Ratliff, 1982). Examples of the three label types produced by labels are provided in thesis section 10.3.3. Examples of the other revision products produced from DELTA and dBASE formats are provided in Chapter Nine.

The central revision database system and the structure depicted in the model (Figure 10.6) should be considered essentially as having a dynamic structure that is available to future taxonomists (see Chapter Eleven). For this reason there is not one database (or wealth of taxonomic knowledge) that exists at the commencement of a revision and another concluding database when the revision is completed. The database is not static, taxonomists are constantly accruing novel data, which are incorporated into the database and result in further taxonomic refinements. In this project the results of the revision of Vicia subg. Vicia are presented in this thesis, but they will, via the database, also be available to future workers to build upon, expand, correct and complement. A copy of the database will be held, for future use by Vicia taxonomists, by the Legume Section of the Royal Botanic Garden, Kew, U.K.

One of the primary innovations explicit in this application of computer technology to the revision paradigm is the use of automatic taxonomic hierarchy shifts (ATHS). The ATHS uses the fact that data referable to one level in the

taxonomic hierarchy have automatic implications for data at another level of the hierarchy. In the exemplar revision ATHS's were used to produce taxon descriptive data and taxon geographical distributions based on representative specimen data.

The first example of this ATHS is demonstrated in Figure 10.7. Individual specimen character scores are placed in the database file MORPDATA via the data capture program CELIA. These data are then analysed using a numerical analysis package and specimens are clustered into accepted taxa. The program LICEA uses the character score data from MORPDATA, for representative specimens of a particular taxon, and synthesises output files, containing quantitative or qualitative data, for that taxon. These two synthesised dBASE files are converted to a single DELTA format file via the program CALIE. Once a DELTA file is available, it can be used to produce a taxon description or keys to taxa, etc., as described above.

The second example of an ATHS is illustrated in Figure 10.8. Here the geographical distribution data for individual specimens is entered into the dBASE file, ECOGEOG. As with the above example, the MORPDATA file is analysed so that specimens can be accurately attributed to taxa. Once specimens are identified as belonging to certain taxa, the program PHYTOGEOG can be used to synthesise taxon distributions. This program uses the distributional pattern of representative specimens to produce generalised information about the distributional pattern of taxa.

Although these are the only two ATHS's that directly use dBASE programs and files to produce their output, the concept of taking specimen data and using them to generalise about particular taxa is central to taxonomy. In the conspectus (Appendix 5) information is summarised on habitat preferences from ECOGEOG and on diploid chromosome number from CHROMNOS.

Figure 10.7. Automatic taxonomic hierarchy shift 1 - Specimen to taxon descriptions

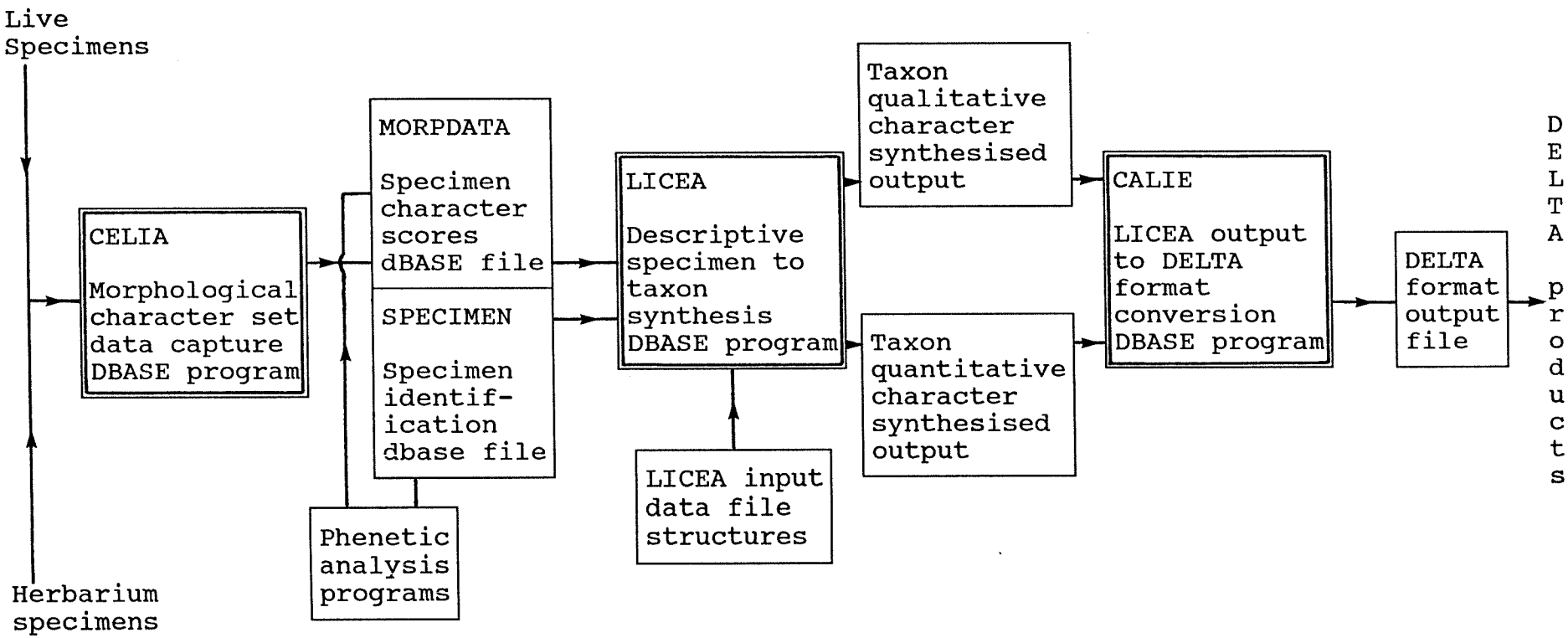
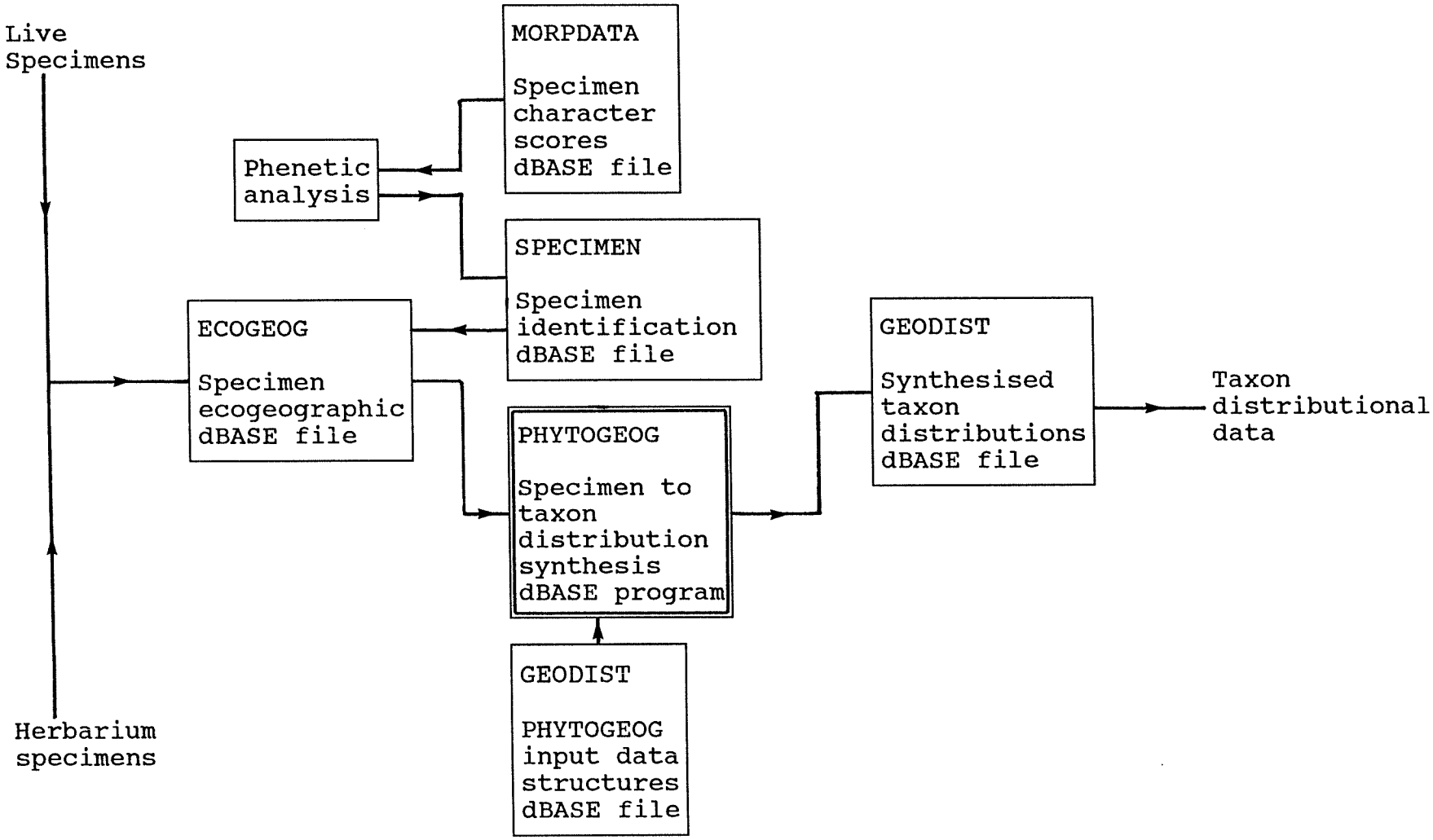


Figure 10.8. Automatic taxonomic hierarchy shift 2 - Specimen to taxon distributions



10.4 Discussion Of Exemplar Revision Database System Implementation

The initial decision to use dBASE II for the project was made in January 1985. This DBMS has now been superseded by better, more flexible DBMS. DBMS technology and associated software has advanced very swiftly. The decision was made not to upgrade from dBASE II during the course of the project to a more flexible DBMS such as dBASE III, dBASE III PLUS, dBASE IV, R:BASE, REVELATION, PICK, ORACLE, etc. The reason for not upgrading the system was partly because of hardware limitations. The project was restricted to a Research Machines 380Z micro-computer with dual 8" floppy disc drives, using the CP/M operating system. Much time would have been lost upgrading the entire database structures and programs through the series of dBASE upgrades produced since 1985. The major objective of the project was to test the application of database techniques. While dBASE II did cause specific limitations within the project, it did not crucially restrict the experimentation with a DBMS and so the entire project was undertaken using dBASE II.

The development of improved DBMS's is advancing very rapidly at this time. New DBMS's packages are appearing almost weekly and it is beyond the scope of this project to outline the advantages and disadvantages of each system. However, two recent reveiw articles in the computing press broadly cover this field, PC Week (1990) for full range of DBMS's and .EXE for dBASE like DBMS's (1990)

The project described here was experimental. The application of a DBMS within the revision paradigm was being tested and no generalised software was expected to result. Time and expertise were not available for the writing of generalised software for taxonomic usage. For this reason the system described remains relatively crude in terms of software development. However, if more time had been available, software should have been added to unify the programs and

databases described above, and to provide a query module for the revision database.

The system described above must remain outside Allkin's (1988) definition of a taxonomic database program. As he envisaged it a taxonomic database program would comprise three basic elements:

TAXONOMIC DATABASE	=	TAXONOMIC	+	TAXONOMIC	+	TAXONOMIC
PROGRAMS		DATA STRUCTURES		ALGORITHMS		INTERFACE

The taxonomic interface is absent from the system described. Although Allkin & White (1988) comment that a taxonomic database system will be a sophisticated taxonomic program, if it is to work both efficiently and effectively.

> "Since the data is complex and requires sophisticated management to support nomenclatural, classifactory and diagnostic activities, it is impractical to expect each taxonomist to design and develop his or her own data management environment."

In the system described above the taxonomist undertaking the revision was also the program designer and wrote much of the software. So it could be argued that a taxonomic interface was not essential in the experimental stage, although it would be an essential element of a generalised system.

Within the codes used for various items, taxonomic meaning was incorporated into the code. This was very helpful in identifying particular items, but presents problems if the meaning of the item codes changes during the course of the project. For example the BEGNOS code was included as an element of the ACCNOS code, so that specimen S37002 was initially identified as Vicia narbonensis (BEGNOS code = B370), but was later reindentified as V. johannis (ENDNOS code = B230). At the conclusion of the phenetic analysis the taxon code (ENDNOS) was different to the code included within the accession code (ACCNOS). This did not cause problems in the

simple dataset used in this revision, but it should be avoided in a generalised system.

It is generally accepted that the relational database model is preferable in designing and implementing database systems (see Chapter Two). Not adhering to the relational model can lead to replication of data items in separate files, leading to confusion and loss of internal consistency, when correcting errors in the database. However, the use of a fully relational model in practice, while saving computer space, requires a complete reliance on accessing programs. As shown above, the system developed was partially relational as some data items are replicated in more than one database file. There were three reasons for this use of a partially relational system. Firstly, inexperience in database design and programming at the commencement of the project, secondly, storage space was not a limiting factor and thirdly, the programming limitation of using dBASE II (see thesis section 10.2.1). As Pankhurst (1983) observes:

> "The use of such extra tables (dictionaries) does however complicate database management. The simplest flat file works and can be easily implemented."

and Allkin (1988) adds:

> "The complexity of much taxonomic information will mean that the number of files containing raw and meta data (data about data item relations) and the number of types of file in systematic databases is necessarily large. Look-up tables and dictionaries (e.g. gazetteers, authors' abbreviations) permit verification of data input and ensure more efficient use of computer storage, but further add to the number of files used, making file management increasingly cumbersome."

Thus a pragmatic approach to data structuring was adopted as opposed to following a more idealistic relational approach. However, in a generalised system a more purely relational approach should be adopted.

One problem associated with the programs used was "hard wiring", which is synonymous with a loss of program

flexibility. A hard wired program is one written to interact with a specific data structure and so its use is limited to interactions with that data structure. All five dBASE programs written were hard wired for the project. This was not a problem for LABELS, PHYTOGEOG, LICEA and CALIE, which were programs used once or used more than once with similar data structures. However, CELIA, the specimen morphological data capture program, was by definition used throughout the phenetic, data capture element of the project. CELIA was re-written and the MORPDATA files were re-structured several times during the project as the morphological character set changed. The hard wiring thus proved undesirable and should be avoided in a generalised system.

The most significant problem encountered during the experimentation with database technology was the problem associated with the analysis of MORPDATA (see Chapter Seven). Such a comparatively large data set (171 characters x 1539 specimens) presented major practical analysis problems. Pheneticists do not commonly deal with character sets of this size. So programs that are theoretically satisfactory for the analysis of these data sets are not generally available. This should not be seen as a fault of using database techniques, but must be taken as a pointer to taxonomic programmers. In the future larger data sets will commonly be used and suitable analysis programs must be made available.

The database system described was restricted by the use of DELTA format for the output of the synthesised specimen data. With minor adjustments LICEA could be enhanced to calculate more detailed data than the basic character details that could be coded using DELTA. The format allows the minimum and maximum dimensions for quantitative characters, but the mean, mode and median might also be included. While for qualitative characters, rather than simply recorded presence for individual character states, it would be advantageous to record the percentage of specimens with each

character state. The advent of taxonomic DBMS's and ATHS programs will necessitate the revision of DELTA format to enable a more flexible approach.

During this project over half of the research time was devoted to becoming familiar with dBASE II, then designing and implementing the database structures and writing the taxonomic programs. As Allkin (1988) points out this process is labour intensive and the software developed in this way is unlikely to be cost effective. To test the application of database software within the revision paradigm, extensive time was necessarily devoted to experimenting with the DBMS. It would be wrong, however, to suggest that future taxonomists undertaking a revision should partition such a large proportion of their research time to non-taxonomic activities. The amount of effort and time expended by an individual taxonomist in building a personal (as opposed to an institutional or general) RDMS would not be redressed by its ultimate usefulness. Having reached this conclusion, it follows that proprietary non-taxonomic DBMS will remain a superficial organisational tool for the average taxonomist until funds are made available for the development of a general purpose taxonomically intelligent DBMS.

CHAPTER ELEVEN

PROPOSED COMPUTER ASSISTED TAXONOMIC REVISION PARADIGM

11.1 Introduction

The previous chapter discussed a specific application of computer techniques to a revision. This chapter is concerned with the general application of computer techniques to the revision process.

It is a truism that taxonomists can no longer afford to squander data (Taylor, 1971; Shetler, 1974; Heywood, 1974, 1984; Raven, 1977). Taxonomic resources are scarce, which forced Allkin (1988) to conclude that the obvious approach to avoid data wastage and make better use of limited taxonomic resources was the "development of intelligent, easily used, and reliable programs" to facilitate the taxonomic process. Pankhurst (1988c) points out, however, "that even if databases are a convenient means to begin a Flora project, and the analysis of DELTA files an effective way to end it, there is still a gap between them which needs to bridged." Computers are already being used piecemeal to bridge this gap in various taxonomic processes. There is a need for these various applications of computers, enhanced with the latest database technology, to be developed into one unified system. This could then be used to manage and enhance the entire revision process (Bisby, 1984b). This general system is referred to as a revision database management system (RDMS).

The general availability of proprietary DBMS's will encourage each computer-literate taxonomist producing his or her own RDMS. As these taxonomists will rarely have a knowledge of systems programming, their programs are likely to be crude and inefficient. It will mean, as each taxonomist develops a personal system, that the developmental process will prove wasteful and repetitive. For this reason it is essential that a sophisticated RDMS system is developed by professionals. Taxonomists would thus avoid wasting time

undertaking developmental research outside their field of expertise. A professionally designed and programmed RDMS would be more like to run efficiently and effectively.

The following discussion of the principles of such a system involves listing its requirements, defining the proposed computer assisted taxonomic revision paradigm, discussing how such a system could be implemented, and discussing where future research should be directed. This chapter is the concluding one of this thesis, consequently, overall conclusions are drawn from the experimentation detailed above.

11.2 Requirements of the Revision Database Management System

One of the primary aims of this research project was to experiment with applying database technology to a taxonomic revision, in effect building a revision database management system (RDMS). Though the system, described in Chapter Ten, is simplistic, it has helped clarify what would be the requirements of a more sophisticated RDMS. These requirements are listed below. This list incorporates the specification for such a system listed by other authors (Allkin, 1988; Allkin & Bisby, 1984, 1988; Allkin & White, 1982, 1988; Crovello et al. (1984), Pankhurst 1983a, 1988c; Watson & Dallwitz, 1981), as well as listing the requirements evident from the above study.

Before detailing the specific requirements of a RDMS, it should be noted that the creators of ALICE are currently involved in the development of a more general, taxonomically intelligent DBMS (BAOBAB). The latter system would be a taxonomic information system rather than a RDMS. The major difference between the two is that BAOBAB will include a built-in capability to manipulate a taxonomic hierarchy or classification and the RDMS could be used specifically to falsify an existing classification and replace it with one closer to the taxon's natural classification. However, the

requirements of both systems are similar and Allkin (1988) provides a list of three general requirements for such a taxonomic information system:

> "1. they need a set of data structures that genuinely reflect the complexity of the data and describe taxonomic data relationships;
>
> 2. they must incorporate algorithms to undertake commonly performed taxonomic operations with all of the necessary checks and adjustments to the defined data structures;
>
> 3. they require a user interface to hide the underlying complexity of the database while allowing taxonomists to manipulate their data as flexibly as possible using terminology and concepts already familiar to them."

These three general requirements underlie the specific ones of the RDMS listed below. It should also be noted that such requirements listed are those of an ideal system. Building such a system would involve solving some of the most basic problems of taxonomy (e.g. those associated with character dependency and reportage of intra taxon variance). So it is not envisaged that the RDMS will meet all the requirements outlined. Taxonomy should, however, aim for the ideal and so the requirements of the RDMS are that the system should:

1. include codified data - to allow more efficient storage in the database, multiple indexing of the data, and thus, once stored a more flexible use of the data;

2. use the relational database model - so that data are only stored once, avoiding data inconsistencies or duplication, and making errors easier to correct.

3. be a workbench tool - the taxonomist must have easy access to the RDMS. It should able to run on the range

of commonly used desk-top micro-computers, using a range of operating systems;

4. provide a file manager - the system must provide an efficient data file manager. It should be as convenient and easy to operate, and incorporate the same facilities, as a proprietary DBMS;

5. provide a taxonomic and database interface - so that the taxonomist may use a taxonomically intelligent system, not requiring extensive computing experience to operate. The RDMS should shield the taxonomist from the inner workings of the system, while incorporating an element of taxonomic common sense, thus enabling the use of familiar terminology such as 'genus' or 'character';

6. be dynamic - the system, when implemented for one particular revision, must be available to subsequent users and taxonomists revising the same group. The process of a computer assisted revision should mirror a traditional revision. The revision is an evolving summation of taxonomic knowledge about a particular taxon and using the RDMS will make it possible to avoid the replication of data gathering and codification;

7. include a data capture module - which allows the taxonomist to sit at his work bench observing herbarium specimens, at the same time being prompted to enter coded taxonomic data directly into the system. The RDMS should facilitate data checking and validation, and be able to recognise and deal appropriately with character dependencies on entry. It is likely that the structure of primary and meta files would be complex. Manual data entry could lead to data inconsistencies and so should be avoided. There is a requirement for a ditto facility. This means that if the data being placed into the database are very similar, (e.g. curatorial data for

specimens collected during the same expedition) then, rather than type repetitive details, the same data could be presented to be edited for the second record;

8. include scrolling screen input - during the present project the data capture program CELIA presented the taxonomist with individual screens of characters. The switching between screens is time-consuming and so the system should offer a scrolling morphological data capture screen. This may be a side effect of the choice of proprietary DBMS, dBASE is conspicuously slow, but this should be considered during RDMS development;

9. produce format output for inter database transfer - the revision products must be easily transferable from individuals RDMS's (preferably using XDF format the International Working Group on Taxonomic Database for Plant Sciences, TDWG standard) to larger institutional database systems and vice versa. The individual databases could feed into large databases, thus avoiding unnecessary duplication. Regular exchange between databases will force the use of data standards, such as are being discussed by TDWG's, as a means to aid consistent data definitions;

10. produce output for product generation programs - it must be possible to take the various kinds of revision data and export data sets from the system for further use in, for example, phenetic and phylogenetic analysis, computer typesetting and word processing. Many of these requirements can be satisfied by the production of output in DELTA format followed by the use of the taxonomic analysis or product generation programs;

11. include usage of automatic taxonomic hierarchy shifts - this concept is introduced in Chapter Ten and means that generalisations about a taxon can be automatically drawn

from data about representative specimens of that taxon. For example, when traditional taxonomists publish habitat preferences for a species A, they would have recorded specimen label data for representative specimens of A, they would then state species A is found in X, Y and Z habitats. This type of generalisation for taxa via ATHS is applicable for all kind of descriptive taxonomic data;

12. allow flexible usage - as established above, no two revisions are identical. The type of revision undertaken is dictated by the purposes of that revision. However, there are certain types of data that are always used in a revision e.g. specimen curatorial data, specimen identifications, synonymy. Thus the RDMS must be flexible enough to allow various combinations of revision data to be gathered and synthesised, then allow the production of various combinations of taxonomic products. The taxonomist must be able to declare which type of data will be included in the revision and the program must be sufficiently flexible to allow changes to the data structures during the course of the revision;

13. accommodate multiple specimen identifications - herbarium specimens often carry with them an identification history. At different times various authors have identified the specimen and each has attached a determination slip. The RDMS must be able to mirror this and store multiple identifications for one specimen, each with the name of the person who identified the specimen and the date of determination;

14. not be limited by size - the Rubus database built by Pankhurst (1983a), discussed in Chapter Two, is relatively small with 100 specimens and 50 taxa. However, the revision of Vicia subgenus Vicia included 2000 specimens from 72 taxa, which can be assumed is closer to the usual dimensions of a revision. The system

must be able to run easily with large sized data sets, avoiding any problems in re-indexing large data files;

15. deal with the problems associated with morphological characters - such as different kinds of character (quantitative and qualitative) and variation within specific characters for a particular taxon. Problems of character dependencies should be resolved, to avoid illogicality, e.g. V. peregrina being scored as having no peduncle and yet having a peduncle length of 7 mm;

16. allow easy data editing and/or addition - if atypical specimens arise, it should be simple to amend the description of the various higher taxonomic levels to allow the inclusion of these fresh data. This could be done automatically by a ATHS;

17. produce comparable products - the system should use a basic template for revision products, thus forcing revision products to be comparable. This point tends to counter-balance the earlier requirement for system flexibility. However, at least within a revision, the products should be uniform, e.g. descriptions should be directly comparable, using the same basic character set. More generally applicable data types can be more directly comparable, e.g. by using one set of recognised geographical unit codes. This type of coordination of acceptable data standards will naturally arise out of the efforts of the International Working Group on Taxonomic Database for Plant Sciences, TDWG (Bisby, 1989b);

18. allow character subset abstraction - so that if a subset of the complete data set is required, it can be easily extracted for use by the taxonomist or for tailored customer products. For example, if a morphological character subset is required for analysis or to undertake a detailed analysis of a subgroup of taxa or to produce a

key based on a subgroup of characters or to produce ecogeographic data for a subset of taxa;

19. allow the user to extract data via synonymous names - so that the user, for example, trying to find the distribution of Faba speciosa, will be told the accepted name for this taxon is Vicia narbonensis and the taxon is found across Europe and the Near East;

20. facilitate customer directed production - revision products are produced for various levels of market, other taxonomists, other professionals and amateurs. It should be possible to direct production specifically for each one of these markets. For example, a botanist going plant collecting in Spain and a hobby gardener going on holiday to Malaga will both require plant identification products. These products can be produced from the same descriptive database, but the level of sophistication required of the two products will differ greatly. The botanist will require a more technical product, while the gardener may prefer a glossy guide with numerous pictures and short diagnoses. Both the botanist- and the gardener-orientated products would certainly be enhanced by the presence of drawings, photographs or paintings. Consumer demand is likely to place an increasing requirement for image storage and management within the database to aid non-specialists in identification;

21. be universally available - the use of one RDMS worldwide would provide or at least enable a unification and collaboration within the taxonomic community. This spirit of collaboration has already been shown in the acceptance of TDWG's standards internationally. It would be ideal if the funding of a RDMS were international, so that like the DELTA associated programs, it could be distributed free of charge;

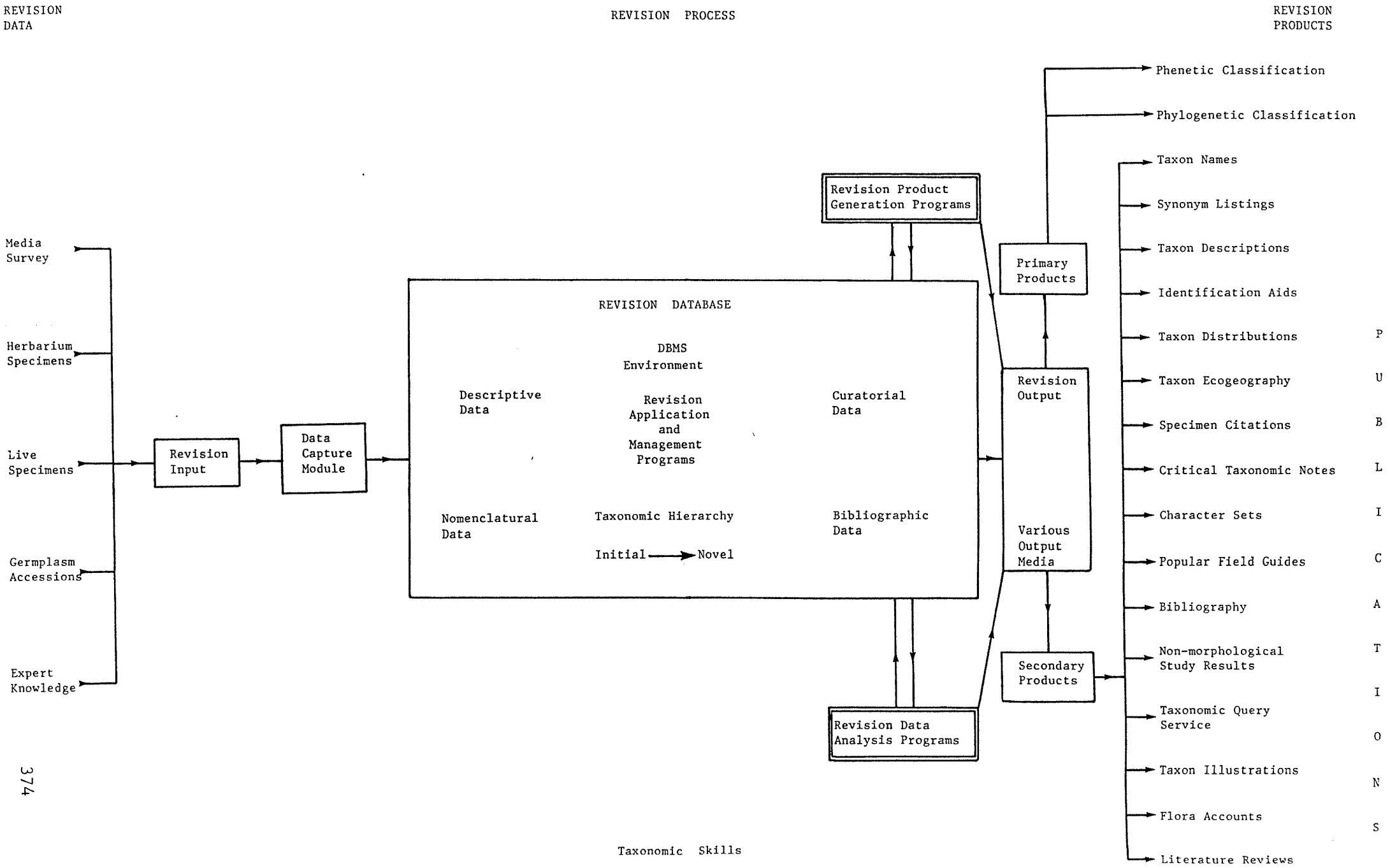
22. include a taxonomic expert system - the rules of nomenclature and taxonomy are complex and so the taxonomist should be prompted to use correct taxonomic procedures by the system. An expert system may be defined as a computer program which exhibits a high level of performance, comparable to that of a human expert. It should also be capable of explaining its reasoning to the user and it should be amenable to updating the database containing the knowledgebase as new knowledge accrues (Hayes-Roth, Waterman and Lenat, 1983). Aid of this kind supplied by the system would prove extremely useful to the inexperienced taxonomist.

11.3 Revision Database Management System Paradigm

Progress has already been made in outlining solutions to some of the above requirements. Allkin (1988) and Allkin and White (1988) specifically discuss the development of a general taxonomic database management system (BAOBAB). Many of the solutions presented by these authors are just as relevant to a less general RDMS. However, drawing concrete conclusions about the detailed structure of the RDMS from the experiment described above and from the work of Allkin, Bisby and White is premature. This work provides a basis from which the implementation of the RDMS will undoubtedly evolve via further experimentation. It is possible, however, to speculate about the content of such a system. The following discussion indicates general design features of the RDMS which are based on the results of the experiment discussed here and other published work.

The detailed models, discussed in Chapter Three for the revision paradigm (Figure 3.1) and in Chapter Ten for the specific computer assisted revision paradigm (Figure 10.6), are the basic units from which a general RDMS paradigm can be derived. This general RDMS paradigm is shown in Figure 11.1. It uses the same input data and produces the same revision products as those listed in the model for the revision

Figure 11.1 The General Revision Database Management System Paradigm



During digitisation it was noted that this numbered page is missing from the Library's copy.

The author has confirmed that the pagination is incorrect and there is no missing textual page.

paradigm (Figure 3.1). However, the central portion of the model, where the revision data are collated, stored and synthesised, is the component most enhanced by the application of database and other computer aided techniques. This central portion of the paradigm composes six basic elements: revision input; data capture program; revision database; revision product generation programs; revision descriptive data analysis programs and revision output.

Two of these six elements, revision input and output, are "funnels" through which data passes between the external taxonomic environment and the revision database. The data capture program is a program designed to place the raw data collected by the taxonomist in the appropriate database tables. Cutbill (1971) emphasises the importance of recording data in the coded form to avoid transcription costs and to reduce the chance of introducing errors during data format conversion. So the data capture program should present the taxonomist with menus that prompt him or her to enter the data in coded form. This method of data entry was used with CELIA, the morphological data capture program (discussed in Chapter Ten). In the simple example described in Chapter Ten, a data capture module was used for the morphological character scores alone, the other data were entered to the database using the DBMS. Entering data, via the DBMS, in a fully relational RDMS is impractical and error-prone due to the complexity of the file structures.

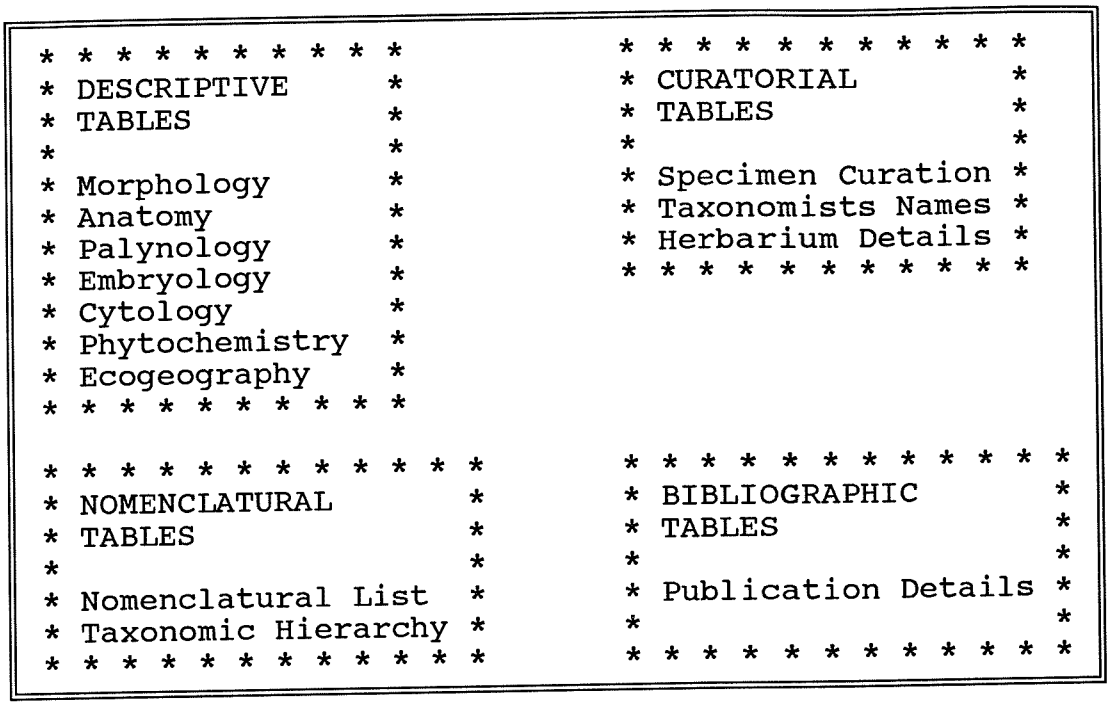
Two other elements in the central portion of the model are sets of proprietary programs external to the revision database, but which manipulate the data held. These programs are used to edit the data and to produce revision products. The revision product generation programs use the descriptive data from the database to generate identification aids and descriptions. The revision data analysis programs use the descriptive data to produce phenetic and phylogenetic classification of datasets held in the database. The results

of this analysis may affect the internal contents of the database, e.g. the results of phenetic analysis may alter the taxa accepted by the revision, forcing the taxonomist to alter the content of the nomenclatural files.

The core of the RDMS is the revision database and its associated programs. If their contents and workings are to be understood, these require further expansion from the model shown in Figure 11.1. The revision database contains four sets of primary data tables, one for each of the data types, as is shown in Figure 11.2. It must be stressed that the word table is used rather than file, as the word file implies a particular physical storage structure, whereas a table may contain several files. The actual files would be created using the relational approach, so one type of data need not correspond to one actual database file. The sorts of data included in the primary tables are outlined in the four sets of files listed below. These tables contain primary data, that is the raw data, as opposed to meta tables, which only contain data about data. Allkin (1988) makes the point that the meta files (tables) may well out-number the primary files (tables) in the database system.

Each of the descriptive tables is split into a pair, one pair for each source of taxonomic evidence. One table contains the accession codes, followed by coded descriptor states, and the second table contains the key to the descriptor state codes (like MORPDATA and MORPCHAR in Chapter Ten). The term accession code is deliberately used to describe the contents of the first table, as it implies non-specific taxonomic ranking. Therefore this code could represent any taxonomic rank from an individual specimen to a family. Both the descriptive tables for each kind of evidence would have author and publication codes, so any evidence could be tied to a reference.

Figure 11.2. General revision database data areas



The general use of DELTA format in the product generation programs suggests that, if possible, the morphological data held in the database should be also held in DELTA format. This would obviate the necessity for data conversion programs. Although, White (pers. comm.), has pointed out that data held in DELTA format would be very slow to search or index.

The general format of the basic descriptive files is shown below:

	Fields
Descriptive File One	Accession code Accession character states Author Code Publication Code

	Fields
Descriptive File Two	Character Character states codes Author Code Publication Code

The curatorial data can be split into three basic table types: specimen curation, taxonomists' names and herbarium detail. The first contains the basic curatorial identification information about specimens. The taxonomists' names table is simply taxonomists' names and their codes, similar to COLECTOR (described in Chapter Ten). Although in COLECTOR the names refer to the specimen collector name combinations, in this more general sense it could be used to include the names of taxonomists in whatever context they are used, e.g. specimen collectors, specific authorities, authors of bibliographic references or as a source of expert advice. Pankhurst (1988c) suggests all names should have individual codes and various code combinations should be used to form what he refers to as a "committee" to represent multiple author combinations or multiple collector combinations. It would also be necessary to store not only the collector name but the standard abbreviation of the name, taken from Meikle (1984) the TDWG standard.

The herbarium code table could simply be based on HERBARIA, discussed in Chapter Ten. The code used throughout the database to refer to herbaria would be the standard Index Herbariorum (Holmgren & Schofield, 1979) code and HERBARIA would match these codes to the Herbarium addresses.

The general format of the three basic curatorial files is shown below:

	Fields
Curatorial File One	Specimen code Specimen identification code Collector code Ecogeographic codes Herbarium code
Curatorial File Two	Fields Taxonomist's name code Taxonomist's name Taxonomist's name abbreviation
Curatorial File Three	Fields Herbarium name code Herbarium name Herbarium address

The nomenclatural data can be held in two basic tables, one with details about taxonomic names and the other containing coded details of the taxonomic hierarchy. The first table would be very similar to the nomenclatural files discussed in Chapter Ten (TAXANOM1, TAXANOM2, SECTNOME and GENNOME). In those examples the different files contain accepted and synonymous name data for the three levels of the hierarchy investigated, that is species, sections and genus. However the splitting of these data into three files is not strictly necessary or desirable. The data for all the taxonomic levels could be better held in the following file structure:

	Fields
Nomenclatural File One	Taxon name code Taxon name Taxonomic rank code Taxonomic status code Author Code Publication Code

The second suggested nomenclatural table provides the key to the taxonomic hierarchy. Within the context of the RDMS the treatment of the taxonomic hierarchy or classification is complex. The primary aim of undertaking the revision is to falsify the existing classification and replace it with a better approximation of the natural relationships of the taxon. Thus the revision starts with one classification and concludes with a novel classification. The two classifications exist outside the revision, as revision input and output. For this reason the incorporation of the taxonomic hierarchy within the RDMS is not explicitly required as a component of the RDMS. However, as each revision progresses from an existing classification, it would be useful at the commencement of the revision for the RDMS to contain a summary of the starting taxon classification.

This hierarchy could be provided via a matrix of taxonomic rank against coding. This may be illustrated with reference to V. faba and its four varieties:

V. faba subsp. faba var. faba

V. faba subsp. faba var. equina

V. faba subsp. faba var. minor

V. faba subsp. paucijuga

this could be coded as:

	Taxonomic Rank		
C	C1	B2	A2
o			
d	C1	B2	A4
i			
n	C1	B2	A5
g			
	C1	B3	

In the coded example, the letter prefix indicates the taxonomic rank , A being the lowest rank and working alphabetically through the alphabet. The number following the

letter indicates the taxon code and thus identification, so that C1 is C = species and 1 = V. faba or A5 is A = variety and 5 = minor. It can be concluded from the hierarchy that A5 is a constituent of B2 and C1. The key to the taxon code number would then need to be identified in a separate file thus:

Code	Taxon
1	V. faba
2	faba
3	paucijuga
4	equina
5	minor

Following the revision, the products could be assessed by appropriate experts. If the results of the revision are accepted by the taxonomic community, then the taxonomic hierarchy could be amended to encode the new classification.

The bibliographic data can be contained in one table, similar in structure to the file CHROMREF (discussed in Chapter Ten). The general structure of which could be:

	Fields
Bibliographic File One	Publication code
	Author code
	Year of publication
	Title
	Journal code
	Journal issue & pages

The programs which construct the database files and manage the database could be developed using four different approaches (Allkin, 1988):

1. program the RDMS using conventional programming languages, such as Fortran or C to write the entire program,

2. program the RDMS so that it sits upon a proprietary DBMS. The user would see a taxonomically intelligent database program, written for example in dBASE, Oracle or C,

3. program the RDMS using the FORTH language, as discussed by White (1984),

4. or program the RDMS using an artificial intelligence language, such as Lisp.

Allkin & White (1988) favour adopting of a combination of 1 and 2 for BAOBAB development, they suggest using "a high level language, C, with calls to a general database management system". They believe this option is preferable as it allows flexibility in the interface format and lower memory requirement. Lang (1983) highlights the problems of fixed field length, which are not restricted to dBASE. The use of fixed field length wastes computer storage space as it forces both fixed and variable length data into a rigid format. If the information is of variable length then the field length must be set a priori to the maximum data item length. This is difficult to estimate a priori and is wasteful of space. For these reasons, if a general database management system is to be used, one that allows variable field lengths is desirable.

Allkin and White (1988) stress that the software for their taxonomic information system contains two elements, the algorithms and the interface. These two elements are also central to the RDMS, which must have the algorithms to do the work of the revision, whilst using the interface to shield the taxonomist from the algorithms. The algorithms need to undertake four basic tasks: data management, data conversion, data synthesis and product generation. Simple data management will encompass data addition, deletion, querying, editing, etc. The basic structure of the data held in the database requires careful consideration. Should the morphological data capture program store the morphological data in DELTA format or in an intermediate format, like that used in MORPDATA (see Chapter Ten)? Whichever one was considered preferable, it is likely that, to undertake certain functions, the RDMS needs to

convert the data from one format to another, either for use within the revision database or for use in external analysis or product generation programs. This process of data conversion is illustrated in Chapter Ten by the program CALIE.

The RDMS program must be capable of data synthesis or generalisation, taking information about specimens representing particular taxa and synthesising from these generalised information about particular taxa. This type of process referred to in Chapter Ten as an Automatic Taxonomic Hierarchy Shift is applicable to all descriptive data and at all levels of the taxonomic hierarchy. Descriptive specimen data can be synthesised to produce descriptive data for higher level taxa. This is illustrated in Chapter ten for two types of descriptive data, morphological and distributional, via the programs LICEA / CALIE and PHYTOGEOG respectively.

Finally, the RDMS program would need to be able to undertake simple product generation tasks, such as file listings and various types of labels production, not covered by the external product generation programs. There are numerous existing programs that will undertake these tasks (Pankhurst, 1983b; Regalado et al., 1986), but each requires specific data structures. Revision products of this kind can be assembled using relatively simple programs and so rather than adapt the data structure to suit specific programs, it would be easier if simple product generation programs were included in the RDMS.

One of the major problems of constructing a RDMS is associated with intra-taxon morphological character variation and dependency (see Chapters Two and Ten). One approach that would possibly avoid this problem would be to include in the database only the lowest taxonomic denominator. This would mean including only data for the lowest taxonomic level, usually individual specimens. Generalised taxon data, such as descriptions or distributions could be produced by automatic

taxonomic hierarchy shifts. This approach would avoid the problem of intra-taxon variability as any variability would be automatically incorporated in the products.

Such an approach avoids the problem discussed in Allkin and Bisby (1988):

> "the observation that "genus X has white flowers" describes, by implication, the flower colour of all species belonging to that genus. Should a species belonging to genus X subsequently be described as having blue flowers, the taxonomist would need to be aware that the generic description was no longer accurate and required amendment."

If the description for the genus X is synthesised from the specimen data of the species of genus X, then the addition of a blue flowered species via its specimens would automatically record the flower colour as white or blue. This solution is implied by Allkin and White (1988), who comment that "variable taxa are simply incorporated, therefore, by entering multiple records in the database". Using the specimen data to produce taxon data is a natural progression of this approach.

A possible argument against adopting this approach is that routine use of ATHS would become increasingly cumbersome and time-consuming the larger the number of specimens included in the database became. This should not be used to justify negating the use of ATHS's. Computer hardware is developing very rapidly, computers are obtaining larger RAM and are working much faster and so in the longer term this may not present a problem. Adoption of this approach would make changing taxon circumscriptions extremely easy, e.g. merging two existing taxa would simply involve changing the specimen identification codes for the new taxon's representative specimens.

A solution to the problem of character dependencies is discussed by Allkin and White (1988). They suggest linking of dependent characters to the data values whose observation

permits the expression of that character state. Allkin and White provide the example of the master character, leaf presence and the dependent character, leaf length. In this case the dependent character can only be used if the master character is scored as present. The use of master and one or multiple dependent characters in this way avoids recording a leaf length of 5 cm. on a plant that has no leaves.

Hawksworth and Bisby (1988) list as one of the products of taxonomy, botanical illustration. The rapid development of computer storage of images will result in the routine incorporation of images to complement botanical descriptions and other identification aids. Hawthorne (1985) and Crust (Pers. Comm.) point out that schematic or outline drawings provide an effective means of communicating descriptive detail of form, shape and colour. So the development of the RDMS should permit the incorporation of various forms of images as well as codified and textual data.

There is currently a movement away from the use of standard identification aids, requiring a degree of expertise to operate, to simpler identification aids. This new generation of aids utilises more computer and image based identification systems. The International Board for Plant Genetic Resources (Reid, Pers. Comm.) is currently developing a range of field guides for its field collectors. These will use computer multi-access polyclaves in conjunction with heavily illustrated printed guides, rather than the more traditional botanical keys. The point that should be taken from this movement, by the RDMS developers, is that the revision products must be tailored to the evolving customer requirements. The RDMS must be capable of producing the products that are required by the market. Bisby (1984b, 1988) and Hawksworth and Bisby (1988) stress the need for taxonomists to supply external users (the non-taxonomists) with easily usable revision products. Some recent applications of this kind are the development of an expert

identification system called EXPERT KEY by Atkinson and Gammerman (1987) or the Orchid identification system developed by Pankhurst (1988a). Both of these were specifically designed with the non-specialist in mind.

11.4 Suggested Future Research Priorities

The previous thesis sections include discussion of the various problems associated with the production of an efficient RDMS. The suggested areas of future research focus on finding a solution to these problems. Thus the priorities are as follows:

1. The most immediate priority is to obtain sufficient funds so that research can be focused on developing the algorithms, programs and data structures that will form the basis of a flexible RDMS. Shetler (1974) warns against the perils of underestimating the cost of developing database systems. Allkin (Pers. Comm.) estimates that the development cost of ALICE (up till 1989) was £200,000 sterling and, with the added complexities inherent in the development of a RDMS, the cost would be likely to be several times more expensive.

   This will be an expensive burden for the taxonomic community, but as Bisby (1984b) points out when discussing the more specific desirability of using a codified database:

   "The very extensive labour of designing and constructing the codified database would then be amortized over the large number of uses to which it would be put."

   The same applies to the development of the RDMS. The cost of design and implementation would be amortized by the resultant enhancement of the revision process and the large numbers of system users.

2. Allkin and White (1982), Allkin (1984, 1988) and Pankhurst (1988c) all point out that there are numerous

programs available for revision product generation and multivariate or cladistic analysis, but that no program yet exists to manage and retrieve taxonomic descriptive data. Within the RDMS the research priority must be the development of software for the revision database, the central element of the paradigm described above. This element of the paradigm is so central to efficient taxonomy that once completed funds to complete the other modules of the RDMS should be easier to obtain.

3. The development of a RDMS will result in the routine use of much larger data sets than taxonomists currently use. This project has shown a need for the development of analysis and product generation software that is not restricted in size.

4. The results of this experiment with the concept of a generalised RDMS have outlined certain requirements for, and the problems associated with designing, such a system. Further research should be focused on clarifying both these requirements and the associated design problems. The clearer the scientific question, the easier it is to provide an answer.

11.5 Project Conclusions

The principal findings of the project described above are:

1. Taxonomists use large complex data sets during a taxonomic revision, which they synthesise to produce various revision products.

2. Database technology can be used to facilitate and enhance revision data management and synthesis.

3. The adoption of a revision database management system (RDMS) will encourage the taxonomist to adopt a logical methodology resulting in revision products being more complete and comparable.

4. There is a requirement by taxonomists for a generalised RDMS, as it will provide an invaluable curation and research tool for individual taxonomists.

5. The amount of effort and time expended by individual taxonomists in building a personal (as opposed to an institutional or general) RDMS is not compensated by a corresponding usefulness.

6. A clarified revision paradigm is presented, which enables a more explicit definition of the required algorithm, a prerequisite for designing the RDMS.

7. Until the development of a sophisticated RDMS, the full potential of database technology will not be available to the individual taxonomist.

REFERENCES

Abbott, L.A., (1976). EXIR User's manual. University of Colorado, Boulder.

Abbott, L.A., Bisby, F.A. & Rogers, D.J., (1985). Taxonomic analysis: the use of computers, models and data bases in biological taxonomy. Columbia University Press, New York.

Abdalla, M.M.F. & Gunzel, G., (1979). Protein content and electrophoresis of seed proteins of certain Vicia faba L. stocks and their assumed ancestors. Z. Pflanzenzücht., 83: 148-154.

Adanson. M., (1763). Familles Des Plantes. Vincent, Paris.

Adey, M.E., Allkin, R., Bisby F.A., & White R.J., (1984). The Vicieae Database: An experimental taxonomic monograph. In "Databases In Systematics". Eds. R. Allkin and F.A. Bisby. pp. 175-188. Academic Press, London and Orlando.

Adey, M.E., Babac, M.T., Bisby F.A., White R.J., & Allkin, R., (1983a). Distribution of chemicals among species in the Vicieae, Part 1. The whole plant. Vicieae Database Project, Publication No. 3, ISSN 0263-8517.

Adey, M.E., Babac, M.T., Bisby F.A., White R.J., & Allkin, R., (1983b). Distribution of chemicals among species in the Vicieae, Part 2. Seeds and leaves. Vicieae Database Project, Publication No. 4, ISSN 0263-8517.

Alefeld, F., (1859). Ueber Die Vicieen. Öst. Bot. Zeitschr., 9: 352-366.

Alefeld, F., (1860a). Hypechusa, nov. gen. Viciearum. Öst. Bot. Zeitschr., 19: 165-166.

Alefeld, F., (1860b). Ueber Vicieen. Bonplandia, 9(5-6): 66-72.

Alefeld, F., (1861a). Ueber Vicieen. Bonplandia, 9(7): 99-105.

Alefeld, F., (1861b). Ueber Vicieen. Bonplandia, 9(8-9): 116-131.

Alefeld, F., (1861c). Ueber Vicieen. Bonplandia, 9(8-9): 138-153.

Alefeld, F., (1866). Vicieen in "Landwirth Schaftliche Flora". Berlin.

Ali, S.I., (1967). Revision of the genus Vicia Linn. from West Pakistan. Bot. Notiser, 120: 46-56.

Alleizette D', C., (1958). Une varieté du Vicia melanops S. et Sm. en Auverge. Bull. Soc. Bot. France, 105: 360-362.

Allkin, R., (1980). Some difficulties of computer-stored taxonomic descriptions. (INIREB document no. 80 30036.) INIREB, Xalapa, Mexico.

Allkin, R., (1984). Handling taxonomic descriptions by computer.In "Databases In Systematics". Eds. R. Allkin and F.A. Bisby. pp. 263-278. Academic Press, London and Orlando.

Allkin, R., (1988). Taxonomically intelligent database programs. In "Prospects in Systematics". Ed. D.L. Hawksworth. pp. 315-331. Clarendon Press, Oxford.

Allkin, R. & Bisby, F.A., (1984). Databases in systematics. The Systematics Association special volume no. 26. Academic Press, London and Orlando.

Allkin, R. & Bisby, F.A., (1988). The structure of monographic databases. Taxon, 37(3): 756-76.

Allkin, R., Goyder, D.J., Bisby, F.A. & White, R.J., (1986). Names and synonyms of species and subspecies in the Vicieae: Issue 3. Vicieae Database Project, Publication No. 7, ISSN 0263-8517.

Allkin, R., Macfarlane, T.D., White, R.J., Bisby, F.A. & Adey, M.E., (1983a). List of species and subspecies in the Vicieae. Vicieae Database Project, Publication No. 1, ISSN 0263-8517.

Allkin, R., Macfarlane, T.D., White, R.J., Bisby, F.A. & Adey, M.E., (1983b). Names and synonyms of species and subspecies in the Vicieae. Vicieae Database Project, Publication No. 2, ISSN 0263-8517.

Allkin, R., Macfarlane, T.D., White, R.J., Bisby, F.A. & Adey, M.E., (1983c). The geographical distribution of Vicia. Vicieae Database Project, Publication No. 5, ISSN 0263-8517.

Allkin, R., Macfarlane, T.D., White, R.J., Bisby, F.A. & Adey, M.E., (1985). The geographical distribution of Lathyrus Vicieae Database Project, Publication No. 6, ISSN 0263-8517.

Allkin, R. & White, R.J., (1982). Design criteria for a computer program to facilitate the acquisition, storage, retrieval and reformatting of biological descriptions. Unpublished Southampton University Research Fund Papers, Southampton, U.K.

Allkin, R. & White, R.J., (1988). Data management for biological classification. In "Classification and related methods of data analysis". Ed. H.H. Bock. pp. 653-660. Elsevier Science Publishers, Amsterdam, Netherlands.

Ascherson, P., (1884). Amphikarpie bei der ein heimischen Vicia angustifolia. Ber. Deutsch. Bot. Ges., 2: 235-245.

Ascherson, P. & Graebner, P., (1909). Synopsis der Mitteleuropäischen Flora. Vol. 6, Part 2. Wilhelm Engelmann, Leipzig.

Ashton-Tate., (1984). dBASE II for your 16-bit PC. Culver City, California.

Atkinson, W.D. & Gammerman, A., (1987). An application of expert systems technology to biological identification. Taxon, 36(4): 705-714.

Avazneli, A., (1981). Vicia. In "Flora of Georgia", pp 422-465. Botanical Institute of the Academy of Sciences, Tblisi, Georgia, U.S.S.R.

Babac, M.T., (1981). An experimental chemotaxonomic data-base for the tribe Vicieae. Ph.D. thesis, University of Southampton.

Babac, M.T. & Bisby, F.A., (1984). A chemotaxonomic database. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 210-218. Academic Press, London and Orlando.

Baksay, L., (1954). Chromosomenstudien an den ungarischen Vicia Arten. Ann. Hist.-Nat. Mus. Natl. Hungarici, Ser. Nova, 6: 167-176.

Ball, P.W., (1968). Vicia L., in Flora Europaea, Vol. 2. pp. 129-136 Cambridge University Press, Cambridge and London.

Barron, D.W., (1984). Current database design - the user's view. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 35-42. Academic Press, London and Orlando.

Barulina, H., (1930). Lentils of the USSR. and of other countries. Bull. Appl. Bot. Genet. & Pl. Breed. Leningrad, Suppl. 40.

Bassler, M., (1973). Revision der eurasiatischen Arten von Lathyrus L. sect. Orobus (L.) Gren. & Godr. Feddes Repert., 84: 329-447.

Bell, E.A., (1960). Canavanine in the Leguminosae. Biochem. J., 75: 618-620.

Bell, E.A., Lackey, J.A.and Polhill, R.M., (1978). Systematic significance of canavanine in the Papilionoideae (Faboideae). Biochem. Syst. Ecol., 6: 201-212.

Bell, E.A. & Tirimanna, S.L., (1965). Associations of amino acids and related compounds in the seeds of forty-seven species of Vicia: their taxonomic and nutritional significance. Biochem. J., 97: 104-111.

Benson, L.D., (1961). Plant taxonomy: methods and principles. Ronald Press, New York.

Bentham, G., (1865). Leguminosae, in Bentham and J.D. Hooker, "Genera Plantarum", Vol 1: 434-600. Lovell Reeve and Co., London.

Birch, A.N.E, Tithecott, M.T. and Bisby, F.A., (1985). Vicia johannis and wild relatives of the Faba Bean : A Taxonomic Study. Econ. Bot., 39(2): 177-190.

Birdsong, B.A., Alston, R. & Turner, B.L., (1960). Distribution of canavanine in the family Leguminosae as related to phyletic groupings. Canad. J. Bot., 38: 499-505.

Bisby, F.A., (1970). The evaluation and selection of characters in Angiosperm taxonomy: An example from Crotalaria. New Phytol., 69: 1149-1160.

Bisby, F.A., (1973). The role of taximetrics in Angiosperm taxonomy. 1. Empirical comparisons of methods using Crotalaria L. New Phytol., 72: 699-726.

Bisby, F.A., (1981). Vicieae Database Project. Taxon, 30: 732.

Bisby, F.A., (1984a). Automated taxonomic information systems. In "Current concepts in plant taxonomy". Eds. V.H. Heywood and D.M. Moore. pp. 301-322. Academic Press, London and Orlando.

Bisby, F.A., (1984b). Information services in taxonomy. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 17-33. Academic Press, London and Orlando.

Bisby, F.A., (1984c). The Vicieae Database Project: products and services. Webbia, 38: 639-644.

Bisby, F.A., (1987). ILDIS (International Legume Database and Information Service). Reunion de coordination et d' information concernant les Legumineuses alimentaires dans le programme STD 1. Centre Tecnique de Cooperation Agricole et Rural (C.T.A.), Wagningen, Netherlands.

Bisby, F.A., (1988). Communications in taxonomy. In "Prospects in Systematics". Ed. D.L. Hawksworth. pp. 277-291. Clarendon Press, Oxford.

Bisby, F.A., (1989a). Databases, information systems, and legume research. In "Advances in Legume Biology", Eds. C. H. Stirton and J.L. Zarucchi. Monogr. Syst. Bot. Missouri Bot. Gard., 29: 811-825.

Bisby, F.A., (1989b). Highlights of TDWG 4 at St. Louis. TDWG Newsletter 2: 3.

Bisby, F.A., Goyder, D.J. & Khattab, A., (1986). Close relatives of Vicia faba. Vicieae Database Project, unpublished document prepared for the International Food Legume Research Conference.

Bisby, F.A. & Nicholls, K.W., (1977). Effects of varying character definitions on classifications of Genisteae (Leguminosae). Bot. J. Linn. Soc., 74: 97-121.

Bisby, F.A., Sutton, D.A. & Russell, G.F., (in press). Taxonomic Databases Working Group for Plant Sciences - Report of Third Meeting, Royal Botanic Gardens, Edinburgh, 19-21 October 1987. Huntia.

Bisby, F.A., White, R.J., Macfarlane, T.D. & Babac, M.T., (1983). The Vicieae Database Project: experimental uses of a monographic taxonomic database for species of vetch and pea. In "Numerical Taxonomy". Ed. J. Felsenstein. pp. 625-629. Springer-verlag, Berlin & Heidelberg.

Boissier, E., (1872). Flora Orientalis, Vol. 2. H. Georg, Geneva & Basel.

Bond, D.A., (1976). Field beans. In "Evolution of crop plants", pp. 179-182. Ed. N.W. Simmonds. Longman, London.

Bondarenko, O.N., (1973). Taxa of grain legumes and their wild relatives. Supp. 122, Bull. Genet. Pl. Breed. (Leningrad).

Bornmuller, J., (1940). Symbolae ad Floram Anatolicam. Fedde Repert., 89(1): 208.

Boswell Syme, J.T., (1873). English botany or coloured figures of British plants, Vol. 3. George Bell & Sons, London.

Boulomov, L., (1930). Flore du Liban et de la Syrie. Vigot Freres, Paris.

Boulter, D., (1981). Proteins of Legumes. In "Advances in Legume Systematics". Eds. R.M. Polhill and P.M. Raven. pp. 501-512. Royal Botanic Gardens, Kew.

Boulter, D. & Derbyshire, E., (1971). Taxonomic aspects of the structure of legume proteins. In "Chemotaxonomy of the Leguminosae". Eds. J.B. Harborne, D. Boulter & B.L. Turner. pp. 285-308. Academic Press, London and New York.

Boulter, D., Thurman, D.A. & Derbyshire, E., (1967). A disc electrophoretic study of globulin proteins of legume seeds with reference to their systematics. New Phytol., 66: 27-36.

Brenan, J.P.M., (1965). The geographical relationships of the genera of Leguminosae in Tropical Africa. Webbia, 19(2): 545- 578.

Brenan, J.P.M., Franks, J.W., Raynal, J. & Cullen, J., (1975). Report of working party on electronic data processing in major European plant taxonomic collections. Adansonia, ser. 2, 15(1): 7-24.

Brill, R.C., (1971). "The TAXIR primer". Institute of Arctic and Alpine Research (Occasional Paper No. 1), University of Colorado, Boulder, U.S.A.

Brill, R.C., (1983). "The TAXIR primer, MTS version". 4th. Edition. University of Michigan Computer Center, Ann Arbor, U.S.A.

Brill, R.C. & Estabrook, G.F., (1984). Management of almost flat files in systematic biology using TAXIR. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 53-68. Academic Press, London and Orlando.

Bronn, H.R., (1822). De formis plantarum leguminosarum primitivis et derivatis. Groos, Heidelberg.

Brough, D.R. & Alexander, I.F., (1986). The fossil expert system. Expert Systems, 3(2): 76-83.

Bueno Perez, A., (1979). Taxonometria y cariologia en el genero Vicia. Ph.D. Thesis, Instituto Nacional De Investigaciones Agrarias, Madrid.

Caesalpinus, A., (1583). De Plantis Libri. Florentiae.

Cain, A.J., (1959). Deductive and inductive methods in post-Linnaean taxonomy. Proc. Linn. Soc. London., 170: 185-217.

Candolle, A.P. de, (1821). Théorie élémentaire de la botanique. W. Blackwood, Edinburgh.

Candolle, A.P. de, (1824-1873). Prodromus systematis naturalis regni vegetabilis. 17 vols and 4 index vols. Treuttel et Wurtz, Paris.

Candolle, A.P. de, (1825a). Prodromus systematis naturalis regni vegetabilis. Vol. 2. Treuttel et Wurtz, Paris.

Candolle, A.P. de, (1825b). Mémoirs sur la famille des légumineuses. pp.1-525. A. Belin, Paris.

Cardenas, A.F., (1985). Data base management systems. Allyn and Bacon Inc, New York.

Chooi, W.Y., (1971a). Variation in nuclear DNA content in the genus Vicia. Genetics, 68: 195-211.

Chooi, W.Y., (1971b). Comparison of the DNA of six Vicia species by the method of DNA-DNA hybridization. Genetics, 68: 212-230.

Chrtkova-Zertova, A., (1979). Flora Iranica, Vol. 140 - Papilionaceae 1 - Vicia and Faba. pp. 16-57. Akademische Druck- u. Verlagsanstalt, Graz, Austria.

Chrtkova-Zertova, A., (1969). Bemerkungen zur Taxonomie von Vicia sepium L. Folia Geobot. Pytotax. Bohem., 4: 303-311.

Cincura, F., (1962). Poznamky k cytologii druhov rodu Vicia L. zo slovenskych nalezisk - II. Acta Fac. Rerum. Nat. Univ. Comenianae, Bot., 7: 349-388.

Cincura, F., (1970). Bemerkungen zur zytotaxonomie einiger fremdlandischer Arten der Gattung Vicia L. (Wicke). Acta Facult. Rer. Nat. Univ. Comenianae, Bot., 18: 17-35.

Clark, G.C.S., & Kupicha, F.K., (1976). The relationship of Cicer L. - the evidence from pollen morphology. Bot. J. Linn. Soc., 72: 35-44.

Compton, R.H., (1912). An investigation of the seedling structure in the Leguminosae. Bot. J. Linn. Soc. 41: 1-122.

Corner, E.J.H., (1951). The leguminous seed. Phytomorphology 1: 117-150.

Coutinho, L.A., (1940). Racas cariologicas na V. sativa. L. Agron. Lusit., 2: 379-403.

Cronquist, A., (1975). Some thoughts on Angiosperm phylogeny and taxonomy. Ann. Mo. Bot. Gard., 62: 517-520.

Crosby M.R. & Magill, R.E., (1988). TROPICOS, a botanical database system at the Missouri Botanical Garden - "The Booklet". Missouri Botanical Garden, St. Louis, Missouri, U.S.A.

Crovello, T.J., (1967). Problems in the use of electronic data processing in biological collections. Taxon, 16: 481-494.

Crovello, T.J., Hauser, L.A. & Keller, C.A., (1984). Brassband (The Brassicaceae Database at Notre Dame): an example of database concepts in systematics. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 219-233. Academic Press, London and Orlando.

Crovello, T.J. & MacDonald, R.D., (1970). Index of EDP-IR projects in systematics. Taxon, 19: 63-76.

Cubero, J.I., (1973). Evolutionary trends in Vicia faba L. Theor. App. Genet., 43: 59-65.

Cubero, J.I., (1982). Interspecific hybridization in Vicia. In "Faba bean improvement". Proceedings of the Faba Bean conference held in Cairo, Egypt, March 7-11, 1981. Eds. G. Hawtin & C. Webb. Martinus Nijhoff, The Hague, Netherlands.

Cubero, J.I., (1984). Taxonomy, distribution and evolution of the faba bean and its wild relatives. In "Genetic resources and their exploitation - chickpeas, faba beans and lentils". Eds. J.R. Witcombe & W. Erskine, pp 137-145. Martinus Nijhoff, The Hague, Netherlands.

Cutbill, J.L., (1971). New methods for handling biological information. Bot. J. Linn. Soc., 3(3): 253-260.

Dallwitz, M.J., (1974). A flexible computer program for generating diagnostic keys. Syst. Zool., 23(1): 50-57.

Dallwitz, M.J., (1978). User's guide to KEY-a computer program for generating identification keys. CSIRO Aust. Div. Entomol. Rep. No. 4, Canberra, Australia.

Dallwitz, M.J., (1979). User's guide to the DELTA system-a general system for coding taxonomic descriptions. CSIRO Aust. Div. Entomol. Rep. No. 13, Canberra, Australia.

Dallwitz, M.J., (1980a). A general system for coding taxonomic descriptions. Taxon, 29: 41-46.

Dallwitz, M.J., (1980b). User's guide to TYPSET - a computer typesetting program. CSIRO Aust. Div. Entomol. Rep. No. 18, Canberra, Australia.

Dallwitz, M.J., (1984). Automatic typesetting of computer-generated keys and descriptions. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 279-290. Academic Press, London and Orlando.

Dallwitz, M.J., (1988). Dallwitz's Package. DELTA Newsletter, 1: 1.

Dallwitz, M.J. & Paine, T.A., (1986) User's guide to the DELTA system - a general system for coding taxonomic descriptions. Third edition, CSIRO Aust. Div. Entomol. Rep. No. 13, Canberra, Australia.

Dallwitz, M.J. & Paine, T.A., (1989) INTKEY version: 25-JUL-89, CSIRO Aust. Div. Entomol. Unpublished Report, Canberra, Australia.

Darwin, C.R., (1859). The origin of species by means of natural selection. J.M. Dent & Sons, London.

Date, C.J., (1986). An introduction to Database Systems, Vol. 1. Addison Wesley, Reading, Massachusetts.

Davis, P.H. & Heywood, V.H., (1963). Principles of angiosperm taxonomy. Oliver & Boyd. Edinburgh and London.

Davis, P.H. & Plitmann, U., (1970). Vicia L., in Flora of Turkey, Vol. 3: 274-325. Edinburgh University Press, Edinburgh.

Davies, R.G., (1985). Comparative methods of numerical analysis. Unpublished Systematics Association 1985 Annual Lecture.

Deen, S.M., (1977). Fundamentals of data base systems. Macmillan Educational Ltd, London.

Dietrich, D.N.E., (1847). Synopsis Plantarum Seu Enumeratio Systematica, Sectio Quarta. B.F. Voigtii, Vimariae.

Donnelly, E.D. & Clark, E.M., (1962). Hybridization in the genus Vicia. Crop Sci., 2: 141-145.

Duncan, T. & Baum, B.R., (1981). Numerical phenetics: its uses in botanical systematics. Ann. Rev. Ecol. Syst., 12: 387-404.

Dunn, G. & Everitt, B.S., (1982). An introduction to mathematical taxonomy. Cambridge University Press, Cambridge.

Dyomina, R.B., (1973). Faba Mill. nomenclature. In "Taxa of grain legumes and their wild relative". Ed. Bondarenko, O.N. pp. 21. Supp. 122, Bull. Genet. Pl. Breed. (Leningrad).

Edwards, M., Morse, D.R. & Fielding, A.H., (1987). Expert systems: frames, rules or logic for species identification?. Comp. App. Biosci. 3(1): 1-7.

Engels, R.W., (1974). A tutorial on database organisation. In, "Annual Review of Automatic Programming". Eds. A. Halpern & N. Mcgee. Pergamon Press, Elmsford, New York.

Estabrook, G.F., (1967). An information theory model for character analysis. Taxon, 16: 86-97.

Estabrook, G.F., (1979). A TAXIR data bank of seed plant types at the university of Michigan herbarium. Taxon, 28(1,2/3): 197-203.

Estabrook, G.F. & Brill, R.C., (1969). The theory of the TAXIR accessioner. Math. Biosci., 5: 327-340.

Estabrook, G.F. & Rogers, D.J., (1966). A general method of taxonomic description for a computed similarity measure. Bio. Science, 16: 789-793.

Everitt, B.S., (1974). Cluster analysis. SSRC and Heinemann, London.

.EXE, (1990). dBASE World. .EXE, 5(3): 12-30.

Fabre, J.H., (1855). Observations sur les fleurs et les fruits hypogées du Vicia amphicarpa . Bull. Soc. Bot. France, 2: 503-509.

Farr, E.R., Leussink, J.A. & Stafleu, F.A., (1979). Index Nominum Genericorum (Plantarum), Vol. 1-3. Bohn, Scheltema & Holkema, Utrecht dr. W. Junk, The Hague, Netherlands.

Fedtschenko, B.A., (1948). Vicia L., in Flora of U.S.S.R., Vol. XIII: 406-475. Izdatel'stvo Akademii Nauk SSSR, Moscow, U.S.S.R.

Ferguson, I.K. & Skvarla, J.J., (1981). The pollen morphology of the subfamily, Papilionoideae (Leguminosae), In "Advances in Legume Systematics". Eds. R.M. Polhill and P.H. Raven, pp.859-896. Royal Botanic Gardens, Kew.

Fleming, H.S. & Appan, S.G., (1971). Program CHARANAL (character analysis) software documentation. Biology and Chemistry teams GURC/MTF environmental data management teams document no. 2. Boulder, Colorado, U.S.A.

Fleming, H.S. & Appan, S.G., (1972). Cluster analysis using graph theory. Gulf Universities Research Consortium (Appendix to task 4). Boulder, Colorado, U.S.A.

Forget, P.M., Lebbe, J., Puig, H., Vignes, R. & Hideux, M., (1986). Microcomputer-aided identification: an application to trees from French Guiana. Bot. J. Linn. Soc., 93: 205-223.

Freeston, M.W., (1984). The implementation of databases on small computers. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 43-52. Academic Press, London and Orlando.

Fukuda, I., (1984). Chromosome banding and biosystematics. In "Plant Biosystematics". Ed. W.F. Grant. pp. 97-116. Academic Press, London and Orlando.

Gams, H., (1924). Illustrierte Flora von Mittel-Europa, Vol. 4(3). Lehmann, Munich.

Gapotchka, G.P., (1974). On the palynomorphology of the species of the tribe Vicieae from the family Fabaceae. Vest. Mosk. Univ. Ser. 6 Biol., 3: 93-98.

Gibbs-Russell, G.E. & Gonsalves, P., (1984). PRECIS-a curatorial and biogeographic system. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 137-153. Academic Press, London and Orlando.

Gillett, J.B., (1971). Introduction to the Flora of Tropical East Africa. In J.B. Gillett, R.M. Polhill and B. Verdcourt, "Flora of Tropical East Africa, Leguminosae". Whitefriars Press, London and Tonbridge.

Godron, D.A., (1849). Vicieae. In "Flore de France, Vol.1". Eds. J.C.M. Grenier & D.A. Godron. J.B. Bailliere, Paris.

Goldblatt, P., (1981). Cytology and the phylogeny of Leguminosae. In "Advances in Legume Systematics". Eds. R.M. Polhill and P.M. Raven. pp.427-463. Royal Botanic Gardens, Kew.

Gómez-Pompa, A., Moreno, N.P., Gama, L. & Sosa, V., (1984). Flora of Veracruz: progress and prospects. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 165-174. Academic Press, London and Orlando.

Gower, J.C. and Ross, G.J.S., (1969). Minimum spanning trees and single linkage cluster analysis. App. Stat., 18: 54-64.

Greuter, W., Burdet, H.M. & Long, G., (1989). Med-Checklist, Vol. 4 - Dicotyledons (Lauraceae - Rhamnaceae). Med-Checklist Trust of OPTIMA, Geneva.

Gunn, C.R., (1970). A key and diagrams for the seeds of one hundred species of Vicia (Leguminosae). Proc. Int. Seed Test. Ass., 35(3): 773-790.

Gunn, C.R., (1971). Seeds of native and naturalised vetches of North America. USDA Agricultural handbook no. 392., Beltsville, Maryland, U.S.A.

Gunn, C.R., Hollis, S. & Bisby, F.A., (in preparation). Identification of Vicia nomenclator: Vicia nomenclator.

Gunn, C.R. & Kluve, J., (1976). Androecium and pistil characters for tribe Vicieae (Fabaceae). Taxon 25(5/6): 563-575.

Hall, A.V., (1970). A computer based system for forming identification keys. Taxon, 19: 12-18.

Hall, A.V., (1975). A system for automatic key-forming. In "Biological Identification with computers", Ed. Pankhurst, R.J. pp 55-72. Academic Press, London and New York.

Hämet-Ahti, L., (1970). Taxonomy of Vicia sepium L. (Leguminosae) in Finland. Ann. Bot. Fenn., 7: 170-176.

Hanelt, P. & Mettin, D., (1964). Cytosystematische Untersuchungen in der Artengruppe um Vicia sativa L. I. Kulturpflanze, 12: 163-171.

Hanelt, P. & Mettin, D., (1966). Cytosystematische Untersuchungen in der Artengruppe um Vicia sativa L. II. Kulturpflanze, 14: 137-161.

Hanelt, P. & Mettin, D., (1970a). Über die systematische Stellung temperater und meridionaler Sippen der Gattung Vicia L. Feddes Repert. 81: 147-161.

Hanelt, P. & Mettin, D., (1970b). Zur systematischen Stellung von Vicia oroboides Wulf. Kulturpflanze, 18: 179-188.

Hanelt, P. & Mettin, D., (1989). Biosystematics of the genus Vicia L. (Leguminosae) Annual Rev. Ecol. Syst., 20: 199-223.

Hanelt, P., Schäfer, H. & Schultze-Motel, J., (1972). Die Stellung von Vicia faba L. in der Gattung Vicia L. und Betrachtungen zur Entstehung dieser kulturart. Kulturpflanze, 20: 263-275.

Hawksworth, D.L. & Bisby, F.A., (1988). Systematics: the keystone of biology. In "Prospects in Systematics". Ed. D.L. Hawksworth. pp. 3-30. Clarendon Press, Oxford.

Hayes-Roth, F., Waterman, D.A. and Lenat, D.B., (1983). Building expert systems. Addison-Wesley, Reading, Massachusetts.

Hawthorne, W.D., (1985). Images in botanical databases. Unpublished University Research Fund Report, Biology Department, Southampton University, U.K.

Hegi, G., (1924). Illustrieste Flora von Mitteleuropa. Vol. 4(3): 1506-1562. Lehmann, Munchen.

Heywood, V.H., (1973). Ecological data in practical taxonomy. In, "Taxonomy and Ecology". Ed. V. H. Heywood. 329-347. Academic Press, London and Orlando.

Heywood, V.H., (1974). Systematics - the stone of Sisyphus. Biol. J. Linn. Soc., 6: 169-178.

Heywood, V.H., (1978). Flowering plants of the World. Oxford University Press, Oxford.

Heywood, V.H., (1984). Electronic data processing in taxonomy and systematics. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 1-15. Academic Press, London and Orlando.

Heywood, V.H. & Derrick, L.N., (1984). The European taxonomic, floristic and biosystematic documentation system. Norrlinia, 2: 33-54.

Hollings, E. & Stace, C.A., (1974). Karyotype variation and evolution in Vicia sativa aggregate. New Phytol., 73: 195-208.

Hollings, E. & Stace, C.A., (1978). Morphological variation in the Vicia sativa aggregate. Watsonia, 12: 1-14.

Hollis, S., (1988). ALICE tutorial. Unpublished ILDIS document, Southampton, U.K.

Holmgren, P.K. & Schofield, E.K., (1981). Index Herbarium I: The herbaria of the world (ed. 7) Bohn, Scheltema and Holkema, Utrecht and Antwerpen, Netherlands.

Howe, D.R., (1985). Data analysis for data base design. Edward Arnold, London.

Hutchinson, J., (1964). The genera of flowering plants, Vol 1. Oxford University Press, Oxford.

IBPGR, (1987). IBPGR Annual Report. International Board for Plant Genetic Resources, Rome, Italy.

ICARDA, (1987). ICARDA Annual Report. International Centre for Agricultural Research in Dry Areas, Aleppo, Syria.

ICARDA, (1988). ICARDA Annual Report. International Centre for Agricultural Research in Dry Areas, Aleppo, Syria.

Jacquin, N., (1778). Flore Austriaca, Vol. 5: 30. J.M. Gerold, Vienna.

Janchen, E., (1957). Vicia. In "Catalogus florae Austriae, Vol. 1(2)". Eds. Hofler, K. & Knoll, F. Vienna.

Jardine, N. & Sibson, R., (1971). Mathematical taxonomy. Wiley and Sons Ltd., London.

Johnston, B.C., (1980). Computer programs for constructing polyclave keys from data matrices. Taxon, 29(1): 47-51.

Jones, S.B., & Luchsinger, A.E., (1987). Plant Systematics. Second Edition. Mcgraw-Hill, New York.

Jussieu, A.L. de, (1789). Genera plantarum. Herissant et Barrois, Paris.

Just, T., (1954). Generic synopses and their role in modern botanical research. Taxon, 3: 201-202.

Kambal, A.E., Bond, D.A. & Toynbee-Clarke, G., (1976). A study on the pollination mechanisms in field beans (Vicia faba L.). J. Agric. Sci. (Cambridge), 87: 519-526.

Kay, Q.O.N., (1987). Ultraviolet patterning and ultraviolet absorbing pigments in flowers of Leguminosae. In "Advances in Legume Systematics, Part 3". Ed. C.H. Stirton. pp. 317-354. Royal Botanic Gardens, Kew.

Kedler, C.A. & Crovello, T.J., (1973). Problems and procedures in the use of floristic data in a computerized data bank. Proc. Indiana Acad. Sci., 82: 116-122.

Khattab, A.M.A., (1987). Taxonomic studies on the close wild relatives of the faba bean (Vicia faba L.) in section Faba. Ph.D. Thesis, University of Southampton.

Khattab, A., Maxted, N. & Bisby, F.A., (1988). Close relatives of the fababean from Syria: a new species of Vicia and notes on V. hyaeniscyamus (Leguminosae). Kew Bull., 43(3): 535-540.

Kiffman, R., (1952). Morphologie und Systematik der landwirtschaftlich bedeutsamen Wicken- (Vicia) und Lisen- (Lens) arten. Z. Acker Pflanz., 94: 422-427.

Koch, D.G.D.J., (1836). Florae Germanicae et Helveticae, Sectio Prior. Friederici Wilmans, Frankfurt.

Kostrakievicz, K., (1951). Studia systematiczne nad polskimi gatunkami rodzaju Vicia L. Polska Akademia umiejetnosci, Materialy do fizjografii kraju, 27: 1-71.

Kugler, H., (1963). UV-Musterungen auf bluten und ihr zustandekommen. Planta, 59: 296-329.

Kuta, E., (1980). Karyological studies on the genus Vicia L. I. Acta. Biol. Cracov. Sér. Bot., 22(1): 81-99.

Kupicha, F.K., (1974). Taxonomic studies in the tribe Vicieae (Leguminosae). Ph.D. thesis, University of Edinburgh.

Kupicha, F.K., (1975). Observations on the vascular anatomy of the tribe Vicieae (Leguminosae). Bot. J. Linn. Soc., 70: 231-242.

Kupicha, F.K., (1976). The infrageneric structure of Vicia. Notes Roy. Bot. Gard. Edinburgh., 34: 287-326.

Kupicha, F.K., (1981). Vicieae. In "Advances in Legume Systematics". Eds. R.M. Polhill and P.M. Raven. pp. 377-381. Royal Botanic Gardens, Kew.

Kupicha, F.K., (1983). The infrageneric structure of Lathyrus. Notes Roy. Bot. Gard. Edinburgh., 41(2): 209-244.

Lackey, J.A., (1977). A revised classification of the tribe Phaseoleae (Leguminosae, Papilionoideae), and its relation to canavanine distribution. Bot. J. Linn. Soc., 74: 163-178.

Ladizinsky, G., (1975a). On the origin of the broad bean, Vicia faba L. Isr. J. Bot., 24: 80-88.

Ladizinsky, G., (1975b). Seed protein electrophoresis of the wild and cultivated species of section Faba of Vicia. Euphytica, 24: 785-788.

Ladizinsky, G., (1978). Chromosome polymorphism in wild populations of Vicia sativa L. Caryologia, 31: 233-241.

Ladizinsky, G., (1981). Consequences of hybridization in Vicia sativa aggregate. Heredity, 47(3): 431-438.

Ladizinsky, G. & Temkin, R., (1978). The cytogenetic structure of Vicia sativa aggregate. Theor. Appl. Genet., 53: 33-42.

Ladizinsky, G. & Waines, G., (1982). Seed protein polymorphism in Vicia sativa agg. (Fabaceae). Pl. Syst. Evol., 141: 1-5.

Lang, K., (1983). Getting the best out of dBASE II. Soft, 1(1-3): 62-66.

Lavania, U.C. & Sharma, A.K., (1984). Giemsa C-banding in Vicia L. J. Indian Bot., 63: 404-407.

Lawrence, G.H.M., (1951). Taxonomy of vascular plants. Macmillan Publishing Co., Inc., New York.

Ledebour, C.F. von, (1842). Flora Rossica, Vol 1. E. Schweizerbart, Stuttgart.

Leenhouts, P.W., (1968). A guide to the practice of herbarium taxonomy. Regnum Veg., 58: 1-60.

Leokene, L.V., (1966). Morphological seed peculiarities of distributed vetch species. Bull. Appl. Bot. Pl. Breed. (Leningrad), 38(1): 32-49.

Leokene, L.V., (1973). Vicia L. nomenclature. In "Legume Nomenclature". Ed. Bondarenko, O.N. pp. 163-197. Supp. 122, Bull. Genet. Pl. Breed. (Leningrad).

Lersten, N.R., (1981). Testa topography in Leguminosae, subfamily Papilionoideae. Proc. Iowa Acad. Sci., 88(4): 180-191.

Lersten, N.R., (1982). Tracheid bar and vestured pits in Legume seeds (Leguminosae: Papilionoideae). Amer. J. Bot., 69: 98-107.

Lersten, N.R. & Gunn, C.R., (1982). Testa characters in Tribe Vicieae, with notes about tribes Abreae, Cicereae, and Trifolieae (Fabaceae). USDA Technical Bulletin Number 1667, Beltsville, Maryland, U.S.A.

Lightowlers, P., (1988). The Antarctic plant database and the BAS herbarium. Taxon, 37(2): 378-380.

Linnaeus, C., (1737a). Genera Plantarum, p. 221. Apud Conradum Wishoff, Leyden.

Linnaeus, C., (1737b). Hortus Cliffortianus, p. 369. Med. Doct & Ac. Imp. N. C. Soc., Amsterdam, Netherlands.

Linnaeus, C., (1753). Species Plantarum, Vol 2. Salvius, Stockholm.

Lock, M., (1989). Legumes of Africa, a check-list. Royal Botanic Gardens, Kew.

Lockerman, R.H., Bisby, F.A. & Van Der Maesen, L.J.G., (1988). Workshop: integration of information on plant diversity. In, "World crops: cool season food legumes". Ed. R.J. Summerfield. Kluwer, Dordrecht, Boston and London.

Lucas, G.Ll., (1984). Databases in systematics: a summing up. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 321-324. Academic Press, London and Orlando.

McVetty, P.B.E. & Nugent-Rigby, J., (1984). Natural cross pollination of faba beans (Vicia faba L.) grown in Manitoba. Canad. J. Plant Sci., 64: 43-46.

Magill, R., (1987). TROPICOS: a botanical database system at the Missouri Botanical Garden. Internal document. Missouri Botanical Garden, St. Louis, Missouri, U.S.A.

Maplestone, P., Allison, J., Hussein, E.H.A., Gamel El-Din, A., Gatehouse, J.A. & Boulter, D., (1985). Variation of the legumin seed storage protein amongst Vicia species. Phytochem., 24: 1717-1723.

Maxted, N., (1984). The inter and intra-generic relationships between Psophocarpus spp. and their allies. M.Phil. thesis, University of Southampton.

Maxted, N., (1989). A new Vicia from South-west Turkey. Notes Roy. Bot. Gard. Edinburgh, 45(3): 453-456.

Maxted, N., Khattab, A. and Bisby, F.A., (1989). Domesticated legumes and their wild relatives: newly discovered relatives of Vicia faba L. do little to resolve the enigma of its origin. OPTIMA conference (Delphi, Greece), paper to be included in the conference volume.

Meikle, R.D., (1984). Draft index of author abbreviations compiled at the Herbarium, Royal Botanic Gardens, Kew.

Meikle, R.D., (1977). Flora of Cyprus, Vol. 1. Bentham-Moxon Trust, Kew.

Metcalfe, C.R. & Chalk, L., (1950). Anatomy of the dicotyledons. Vol. 1. Oxford University Press, London.

Mettin, D., (1962). Bastardierungsversuche in der Gattung Vicia. I. Über das zytogenetische verhalten der bastarde von V. sativa subsp. amphicarpa x subsp. obovata. Züchter 32: 146-155.

Mettin, D. & Hanelt, P., (1964). Cytosystematische Untersuchungen in der Artengruppe um Vicia sativa L. I. Kulturpflanze 12: 163-225.

Mettin, D. & Hanelt, P., (1973). Über Speziationsvorgange in der Gattung Vicia L. Kulturpflanze 21: 25-54.

Michener, C.D. & Sokal, R.R., (1957). A quantitative approach to a problem in classification. Evolution 11: 130-162.

Moriya, A., (1961). Hybrid vetch obtained by species crossing between common and Japanese wild vetches. Crop Sci., 1: 229-230.

Morris, J.W., (1980). Encoding the national herbarium (PRE) for computerised information retrieval. Bothalia, 13(1&2): 149-160.

Morris, J.W. & Glen, H.F., (1978). PRECIS, the national herbarium of South Africa (PRE) computerised information system. Taxon, 27(5/6): 449-462.

Morse, L.E., (1968). Construction of identification keys by computer (Abstract). Amer. J. Bot., 55: 737.

Morse, L.E., (1971). Specimen identification and key construction with time sharing computers. Taxon, 20: 269-282.

Morse, L.E., (1974). Computer programs for specimen identification, key construction and using taxonomic data matrices. Michigan State University, Biological Series, 5(1): 1-128.

Morse, L.E., (1975). Recent advances in biological specimen identification. In "Biological identification with computers". Ed. Pankhurst, R.J. pp. 11-54. Academic Press, London and New York.

Mouterde, P., (1962). Contribution à l'étude de la flore syrienne et libanaise (suite). Bull. Soc. Bot. France, 108: 310-316.

Mouterde, P., (1966). Vicia in Nouvelle Flore du Liban et de la Syrie, Vol. 2: 396-414. Librairie Orientale, Beirut, Lebanon.

Muratova, V.S., (1931). Common beans (Vicia faba L.). Supp. 50, Bull. Genet. Pl. Breed. (Leningrad), pp. 248-298.

Nikiforova, O.D., (1985) .The system of the genus Vicia (Fabaceae) in Siberia. Bot. Zhurn. (Leningrad), 70(5): 604-611.

Norusis, M.J., (1988) SPSS^x, User's guide, 3rd. Edition. SPSS Inc. Chicago, USA.

Nozzolillo, C., (1976). Identification of Vicia seedlings. Canad. J. Bot., 55: 2439-2462.

Pankhurst, R.J., (1970). A computer program for the generation of diagnostic keys. Computer J., 13(2): 145-151.

Pankhurst, R.J., (1971). Botanical keys generated by computer. Watsonia, 8: 357-368.

Pankhurst, R.J., (1975a). Biological identification with computers. Academic Press, London and New York.

Pankhurst, R.J., (1975b). Identification by matching. In "Biological identification with computers". Ed. Pankhurst, R.J. Academic Press, London and New York. pp. 79-91.

Pankhurst, R.J., (1975c). Identification methods and the quality of taxonomic descriptions. In "Biological identification with computers". Ed. Pankhurst, R.J. Academic Press, London and New York. pp. 237-247.

Pankhurst, R.J., (1978a). The printing of taxonomic descriptions by computer. Taxon, 27(1): 35-38.

Pankhurst, R.J., (1978b). Biological Identification: The principles and practice of identification methods in biology. Edward Arnold, London.

Pankhurst, R.J., (1983a). The construction of a floristic database. Taxon, 32(2): 193-202.

Pankhurst, R.J., (1983b). Preparation of collection labels with a microcomputer. Curator, 26(4): 293-300.

Pankhurst, R.J., (1983c). An improved algorithm for finding diagnostic taxonomic descriptions. Math. Biosci., 65: 209-218.

Pankhurst, R.J., (1986). A package of computer programs for handling taxonomic databases. Comp. App. Biosci., 2(1): 33-39.

Pankhurst, R.J., (1988a). PANKEY programs. DELTA Newsletter, No. 1. pp. 3-4.

Pankhurst, R.J., (1988b). Notes on PANKEY. DELTA Newsletter, No. 2. pp. 5-6.

Pankhurst, R.J., (1988c). Database design for monographs and floras. Taxon, 37(3): 733-746.

Pankhurst, R.J., (1988d). An interactive program for the construction of identification keys. Taxon, 37(3): 747-755.

Pankhurst, R.J. & Aitchison, R.R., (1975a). A computer program to construct polyclaves. In "Biological identification with computers". Ed. Pankhurst, R.J. pp. 73-78. Academic Press, London and New York.

Pankhurst, R.J. & Aitchison, R.R., (1975b). An on-line identification program. In "Biological identification with computers". Ed. Pankhurst, R.J. pp. 181-194. Academic Press, London and New York.

Partridge, T.R., Dallwitz, M.J. & Watson, L., (1986). A primer for the DELTA system on VMS, MS-DOS, and PRIMOS. Division of entomology report no. 38, CSIRO, Canberra, Australia.

Payne, R.W., (1975a). GENKEY: a program for constructing diagnostic keys. In "Biological Identification With Computers". Ed. Pankhurst, R.J. pp. 65-72. Academic Press, London and New York.

Payne, R.W., (1975b). GENKEY. University of Edinburgh Program Library Unit, Inter-University / Research Councils Series Report No. 24. University of Edinburgh, Edinburgh.

PC Week, (1990). Databases. PC Week, 6(29): 8-12.

Perrino, P., Maellaro, I. & Blanco. A., (1977). Patterns di elettroforesi in specie del genere Vicia. Genet. Agrar., 31: 121-130.

Perrino, P., Maruca, G., Lester, R.N. & Hanelt, P., (1988). A chromatographic approach to the taxonomy of Vicia L. Kulturpflanze, 36: 391-404.

Perrino, P. & Pignone, D., (1981). Contribution to the taxonomy of Vicia species belonging to the section Faba. Kulturpflanze, 29: 311-319.

Perrino, P., Yarwood, M., Hanelt, P. & Battista Polignano, G., (1984). Variation of seed characters in selected Vicia species. Kulturpflanze, 32: 103-122.

Pickersgill, B., Jones, J.K., Ramsay, G. & Stewart, H., (1983). Problems and prospects of wide crossing in the genus Vicia for the improvement of faba bean. Proceedings of the International Workshop on Faba Beans, Kabuli Chickpeas and Lentils in the 1980's, 16-20 May 1983, ICARDA, Aleppo, Syria.

Pignone, P., & Attolica, M., (1980). Chromosome banding in four groups of Vicia faba L. Caryologia, 32: 283-288.

Plitmann, U., (1965). Three new legumes from Israel. Is. J. Bot., 14: 90-96.

Plitmann, U., (1967). Biosystematical study in the annual species of Vicia of the Middle East. The Hebrew University of Jerusalem.

Plitmann, U., (1973). Biological Flora of Israel, 4. Vicia sativa subsp. amphicarpa (Dorth.) Aschers. & Graebn. Is. J. Bot., 22: 178-194.

Polhill, R.M. & Raven, P.H., (1981). Advances in Legume Systematics. Royal Botanic Gardens, Kew.

Polhill, R.M., Raven, P.H., & Stirton, C.H., (1981). Evolution and systematics of the Leguminosae. In "Advances in Legume Systematics". Eds. R.M. Polhill and P.H. Raven. pp.1-26. Royal Botanic Gardens, Kew.

Ponert, J., (1973). Combinationes novae, stat novi et taxa nova non tantum specierum turcicarum. Feddes Repert., 83(9/10): 617-644.

Ponert, J., (1975). Phytogeographische und indirekte genetische Analyse innerhalb der Subsektion und die Artenbildung in der Turkei und Nachbarlandern. Feddes Repert., 86(1/2): 43-56.

Prance, G.T., Rogers, D.J., & White, F., (1969). A taximetric study of an angiosperm family: generic delimitation in the Chrysobalanaceae. New Phytol., 68: 1203-1234.

Proctor, M. & Yeo, P., (1973). The pollination of flowers. Collins, London.

Radford, A.E., (1986). Fundamentals of plant systematics. Harper & Row, New York.

Radzhi, A.D., (1971a). Evolution of the genera of the tribe Vicieae Adans. Bot. Zhurn. 56(7): 978-981.

Radzhi, A.D., (1971b). Conspectus systematis specierum Caucasicarum Generis Vicia L. Novit. Syst. Pl. Vasc. (Leningrad), 7: 228-240.

Raina, S.N. & Rees, H., (1983). DNA variation between and within chromosome complements of Vicia species. Heredity, 51: 335-346.

Ramsay, G., (1984). C-banding in Vicia species. In "Systems for cytogenetic analysis in V. faba". Eds., G.P. Chapman & S.A. Tarawah. pp. 28-39. Martinus Nijhoff, The Hague, Netherlands.

Ramsay, G. & Pickersgill, B., (1984). Interspecific hybridization in Vicia section Faba: comparison of selfed and hybrid embryo and endosperm. In "Systems for cytogenetic analysis in V. faba". Eds., G.P. Chapman & S.A. Tarawah. pp. 138-140. Martinus Nijhoff, The Hague, Netherlands.

Ramsay, G. & Pickersgill, B., (1986) Interspecific hybridization between Vicia faba and other species of Vicia: approaches to delaying embryo abortion. Biol. Centralbl., 105: 171-179.

Ramsay, G., Pickersgill, B., Jones, J.K., Hammond, L. & Stewart, M.H., (1984). Barriers to interspecific hybridization between Vicia faba and other species of section Faba. Unpublished paper presented at the EEC faba bean seminar, University of Nottingham, Sept. 1983.

Ratliff, W., (1982). dBASE II - User manual. Ashton- Tate, California.

Raven, P.H., (1977). The systematics and evolution of higher plants. In "Changing Scenes in Natural Sciences, 1776-1976", pp. 59-83. Academy of Natural Sciences, Philadelphia [Spec. Publ. 12], Lancaster, U.S.A.

Ray, J., (1682). Methodus Plantarum Nova. pp.166. Faitborne and Kersey, London.

Raynaud, C., (1976). Monographie et iconographie du genre Vicia L. au Maroc. Bull. Inst. Sci. No. 1., Algiers.

Regalado, J.C., Rabeler, R.K. and Beaman, J.H., (1986). LABELS3 user's manual: Guide to development of a collection database. Beal - Darlington Herbarium, Michigan State University, East Lansing, Michigan, U.S.A.

Renfrew, J., (1973). Palaeoethnobotany. Columbia University Press, New York.

Robeson, D.J. & Harborne, J.B., (1980). A chemical dichotomy in phytoalexin induction within the tribe Vicieae of the Leguminosae. Phytochem., 19: 2359-2365.

Rohlf, F.J., (1968). Stereograms in numerical taxonomy. Syst. Zool., 17: 246-255.

Ross-Craig, S., (1948). Vicia. In "Illustrated Drawings of British Plants", Vol. 3: 55-72.

Roti-Michelozzi, G. & Caffaro, L., (1984). Cytotaxonomic studies in wild populations of Vicia hybrida L. and Vicia lutea L. Folia Geobot. Phytotax. Bohem., 19: 169-176.

Roupakias, D.G., (1986). Interspecific hybridization between Vicia faba (L.) and Vicia narbonensis (L.): Early pod growth and embryo-sac development. Euphytica, 35: 175-183.

Roupakias, D.G. & Tai, W., (1986). Interspecific hybridization in the genus Vicia under controlled environment. Z. Pflanzenzücht., 96: 177-180.

Rouy, G.C.C., (1899). Vicieae. In "Flore de France, Vol. 5". Eds. G.C.C., Rouy, J. Foucaud & E.G. Camus. Deyrolle, Paris.

Rowlands, D.G., (1959). A case of mimicry in plants - Vicia sativa L. in lentil crops. Genetica, 30: 435-436.

Sarasan, L., (1981). Why museum computer projects fail. Museum News, 59: 40-49.

Sarasan, L. & Neuner, A.M., (1983). Museum collections and computers. Association of systematic collections, Lawrence, Kansas, U.S.A.

Schäfer, H.I., (1973). Zur Taxonomie der Vicia narbonensis - Gruppe. Kulturpflanze, 21: 211-273.

Schoth, H.A., (1916-1919). Third annual report on the cooperative vetch experiments at the Oregon Experimental Station. Corvallis, Oregon, U.S.A.

Schmalhausen, J.T., (1895-97). Flora ssrednej i jushnoj Rossii. 2 Vols. Kiev, U.S.S.R.

Schultze-Motel, J., (1972). Die archaologischen Reste der ackerbohne Vicia faba L., und die Genese der Art. Kulturpflanze, 19: 321-358.

Sekizuka, S., Yoshiyama, T., Moriya, A. & Akabane, M., (1960). New vetch strains obtained by species crossing between Japanese and foreign vetches. Kanto-Tosan Agr. Exp. Sta. J., 16: 291-308.

Senn, H.A., (1938). Chromosome number relationships in the Leguminosae. Bibliogr. Genet., 12: 175-336.

Seringe, N.C., (1825). Vicieae. In Prodromus systematis naturalis regni vegetabilis, Candolle, A.P. de, Treuttel et Wurtz, Paris.

Singh, V.P. & Lelley, T., (1983). Giemsa C-banding karyotype of V. narbonensis as compared to V. faba. FABIS, 4: 22-23.

Shah, G.L. & Kothari, M.J., (1973). On the structure of stomata and hairs and its bearing on the systematics in the Tribe Vicieae (Papilionaceae). Flora Bd., 162: 533-548.

Schur, P.J.F., (1866). Vicieae. Enumeratio plantarum transsilvaniae. Braumuller, Vienna.

Shetler, S.G., (1974). Demythologizing biological data banking. Taxon, 23(1): 71-100.

Shetler, S.G., (1975). A generalised descriptive data bank as a basis for computer-assisted identification. In "Biological Identification with computers". Ed. Pankhurst, R.J. pp. 197-235. Academic Press, London & New York.

Shultz, L., (1983). Computerised literature file. Medical electronics, 14: 72-78.

Smartt, J., (1984). Evolution of grain legumes. 1. Mediterranean pulses. Expl. Agric., 20: 275-296.

Smith, J.E., (1798). Vicia laevigata. In "English Botany". Ed. Sowerby, J. George Bell and Sons, London.

Sneath, P.H.A., (1961). The mathematical assessment of taxonomic similarity including the use of computers. Taxon, 10: 98.

Sneath, P.H.A. & Sokal, R.R., (1973). Numerical taxonomy. Freeman and Co., San Francisco.

Sokal, R.R. & Sneath, H.A., (1963). Principles of numerical taxonomy. Freeman and Co., San Francisco.

Srivatava, L.M., (1963). Cytogenetical studies in certain species of Vicia. Cytologia, 28: 154-169.

Stace, C.A., (1980). Plant taxonomy and biosystematics. Edward Arnold, London.

Stankevich, A.K., (1970). On clarification of the Vicia L. genus systematics. Tr. Prikl. Bot. Genet. Sel., 43: 110-125.

Stankevich, A.K., (1978a). On clarification of Vicia sativa L. taxonomy. Byull Vses. Ord. Lenina Inst. Rast. N.I. Vavilova, 81: 3-11.

Stankevich, A.K., (1978b). The systematic position of Vicia faba L. Nov. Syst. Vysh. Rast., 15: 181-183.

Stankevich, A.K., (1982). On taxonomic position of some sections of the genus Vicia L. Bull. App. Bot. Genet. Pl. Breed. (Leningrad), 72: 21-27.

Stankevich, A.K., (1983). The genus Vicia L. and its position in the tribe Fabeae (Fabaceae). Bull. App. Bot. Genet. Pl. Breed. (Leningrad), 79: 3-10.

Stankevich, A.K., (1988). On the priority of the name Vicia abbreviata Fisch. ex Spreng. to V. truncatula Fisch. ex Bieb. and on the volume of this species (Fabaceae). Bull. App. Bot. Genet. Pl. Breed. (Leningrad), 117: 85-90.

Stearn, W.T., (1971). A survey of the tropical genera Oplonia and Psilanthe (Acanthaceae). Bull. Brit. Mus. (Nat. Hist.) Bot., 4: 261.

Stebbins, G.L., (1967). Adaptive radiation and trends of evolution in higher plants. Evol. Biol., 1: 101-140.

Stebbins, G.L., (1971). Chromosome evolution in higher plants. Edward Arnold, London.

Stone, N.D., Coulson, R.N., Frisbie, R.E. & Loh, D.K., (1986). Expert systems in entomology: three approaches to problem solving. Bull. Entomol. Soc. Am., 32: 161-166.

Strategic Locations Planning, (1987). ATLAS GRAPHICS Manual. STSC. Inc., Rockville, Maryland, U.S.A.

Streicher, O., (1902). Beitrage zur vergleichenden Anatomie der Vicieen. Beih. Bot. Centralbl., 12: 483-538.

Stuessy, T.F., (1990). Plant taxonomy: the systematic evaluation of comparative data. Columbia University Press, New York and Oxford.

Sundgren, B., (1985). Data bases and data models. Chartwell-Bratt, New York.

Sveshnikova, I.N., (1927). Karyological studies on Vicia. Bull. Appl. Pl. Breed. (Leningrad), 17: 63-72.

Sveshnikova, I.N., (1940). Cytogenetical analysis of heterosis in hybrids of Vicia. J. Hered., 31: 349-360.

Swederski, W., (1924). O okreslaniu nasion gatunkow Vicia L. wedlug wielkosci znaczka. Roczniki Nauk Rolniczych, 11(1): 44-51.

Takhtadjan, A., (1969). Flowering plants: origin and dispersal. Oliver & Boyd, Edinburgh and London.

Tali, V., (1975). Disc electrophoretic study of salt-soluble seed proteins of various Vicia species. Eesti. NSV Tartu Riikl. Ülik. Toimet. Biol. Teadused, 24(1): 35-41.

Tamamschjan, S., (1954). Flora of Azerbaijan, Vol. 5: 495.

Taubert, P., (1894). Leguminosae. In Die Naturlichen Planzenfamilien III. Eds. A. Engler and H. Prantl, 10: 350-384, Leipzig.

Taylor, R.L., (1971). The Flora North America project. BioScience, 21: 521-523.

Terziiski, D., (1977). Scanning and transmission electron microscopic study on pollen morphology and sporodermal ultrastructure in Vicia sativa L. and Vicia dumentorum L. Fitol., 8: 20-30.

Tischler, G., (1950). Die chromosomenzahlen der gefasspflanzen Mittleeuropas. W. Junk, The Hague, Netherlands.

Toma, C., Rodica, R. & Rodica, D., (1973). Contributii la studiul morfologic si histo-anatomic al stipelelor de la specii de Vicia L. din Flora Romaniei. Ann. St. Univ. Lasi. Sect. IIA, 19: 113-121.

Thompson, H.S. & Barton, W.C., (1924). V. lathyroides L. Rep. Watson B.E.C., 3: 250.

Tournefort, J.P. de, (1694). Elemens de botanique, 3 Vols. L'Imprimerie Royale, Paris.

Tournefort, J.P. de, (1700). Instituitiones Rei Herbariae. Ed. 3., Paris.

Townsend, C., (1984). Using dBASE II. Osborne/McGraw-Hill, Berkeley, California.

Townsend, C.C., (1967). Contributions to the Flora of Iraq: V., Notes on the Leguminosales. Kew Bull., 21: 435-458.

Townsend, C.C., (1974). Vicia L. in Flora of Iraq, Vol. 3. pp. 512-544. Ministry of Agriculture and Agrarian Reform of Iraq, Baghdad, Iraq.

Trabut, L., (1910). L'indigénat de la fève en Algérie. Bull. Soc. Fr., 57(5): 424-429.

Trabut, L., (1911). L'indigénat de la fève en Algérie (Note 2). Bull. Soc. Fr., 58(1): 3-7.

Trautvetter, E.R., (1871). Catalogus Viciearum rossicarum. Acta Horti Petrppolitani, 3: 33-83.

Tschiersch, B., (1961). Über das Vorkommen von Canavanine. Flora Abt. A. Bd., 153: 73-86.

Tschiersch, B. Von & Hanelt, P., (1967). Die freien Aminosauren der Samen von Vicia L. und die systematische Gliederung der Gattung. Flora Abt. A. Bd., 157: 389-406.

Turner, B.L., (1977). Chemosystematics and its effect upon the traditionalists. Ann. Mo. Bot. Gard., 64: 235-242.

Tzvelev, N., (1980). Systema specierum generis Vicia L. in parte Europaea URSS. Nov. Sist. Vyssh. Rast., 17: 200-208.

Van Cruchten, C., (1974). Etude de la croissance des tubes polliniques dans les crisements intra et inter specifiques de Vicia faba et Vicia narbonensis. Internal Report, Station D'Amelioration des Plantes, Dijon, France

Visiani, R., (1842). Flora Dalmatica, Vol. 1. Freidricum Hofmeister, Leipzig.

Voronchikhin, V.V., (1981). Identification of certain species of the genus Vicia L. from their fruits and seeds. Vestnik Mosk. Univ. Biol., 36(2): 22-29.

Ward, J.H., (1963). Hierarchical grouping to optimize an objective function. J. Am. Statist. Ass., 58: 236-244.

Watanabe, K. & Yamada, T., (1956). Studies on the interspecific hybridization of common vetch, Vicia sativa L. and Yahazuendo, V. angustifolia L. var. segetalis. I. Morphological characterisation and fertility in F1 and F2 plants. Bull. Nat. Inst. Agr. Ser. G., 12: 1-18.

Watanabe, K. & Yamada, T., (1958). Studies on the interspecific hybridization of common vetch, Vicia sativa L. and Yahazuendo, V. angustifolia L. var. segetalis. II. On karyotype observed in hybrid progenies and their relation to phenotypes. Bull. Nat. Inst. Agr. Ser. G., 15: 109-147.

Watson, L., (1970). Representation of taxonomic data. In "Data Representation". Eds. R.S. Anderson & M.R. Osborne. pp. 108-112. University of Queensland Press, Brisbane, Australia.

Watson, L., (1971). Basic taxonomic data: the need for organisation over presentation and accumulation. Taxon, 20(1): 131-136.

Watson, L., (1981). An automatic system of generic descriptions for Caesalpinioideae, and its application to classification and key-making. In "Advances in Legume Systematics". Eds. R.M. Polhill and P.H. Raven. pp.65-80. Royal Botanic Gardens, Kew.

Watson, L. & Milne, P., (1981). A flexible system for automatic generation of special purpose dichotomous keys, and its application to Australian grass genera. Aust. J. Bot., 20: 331-352.

Watson, L. & Dallwitz, M.J., (1980). Australian grass genera - anatomy, morphology and keys. The Australian National University, Research School of Biological Sciences, Canberra, Australia.

Watson, L. & Dallwitz, M.J., (1981). An automatic data bank for grass genera. Taxon, 30: 424-429.

Watson, L., Dallwitz, M.J. & Johnston, C.R., (1986) Grass genera of the world: 757 descriptions from an automated database. Aust. J. Bot., 34: 223-230.

Williams, W.T. & Lance, G.N., (1965). Logic of computer based intrinsic classifications. Nature (London), 207: 159-161.

Williamson, M.H., (1978). The ordination of incidence data. J. Ecol., 66: 911-920.

Willis, J.C., (1921). Age and area; a study in geographic distribution and origin of species. Cambridge University Press, Cambridge, U.K.

Winfield, P.J., Allkin, R. & Bisby, F.A., (1987) ALICE biological checklist and species diversity database system. Agricultural scientific services, East Craigs, Edinburgh.

Winfield, P.J. & Green, F.N., (1984). The use of a descriptive database as an aid to assessing the distinctness of pea cultivars. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 189-200. Academic Press, London and Orlando.

Wirth, M., Estabrook, G.F. & Rogers, D.J., (1966). A graph theory model for systematic biology, with an example for the Oncidiinae (Orchidaceae). Syst. Zool., 15: 59-69.

Wishart, D., (1975). CLUSTAN 1C. User Manual. Computer Centre, University College, London.

Wishart, D., (1987). CLUSTAN 3, Users Manual. Computer centre, University of Edinburgh.

Wittgenstein, L. (1961). Tractatus Logico-Philosophicus. Routledge & Kegan Paul, London, U.K.

White, R.J., (1984). Implementing small database systems with specialized features. In "Databases in Systematics". Eds. R. Allkin and F.A. Bisby. pp. 291-308. Academic Press, London and Orlando.

Wolff, G., (1980). Investigations on the relations within the family Papilionaceae on the basis of electrophoretic banding patterns. Theor. Appl. Genet., 57: 225-232.

Woolley, J.B. & Stone, N.D., (1987). Application of artificial intelligence to systematics: SYSTEX-a prototype expert system for species identification. Syst. Zool., 36(3): 248-267.

Yamamoto, K., (1966). Studies on the hybrids between the Vicia sativa L. and its related species. Mem. Fac. Agr. Kagawa U., 21: 1-104.

Yamamoto, K., (1968a). On the interspecific hybrid between Vicia sativa and V. macrocarpa. Japan. J. Breed., 18: 156-166.

Yamamoto, K., (1968b). On the interspecific hybrid between Vicia amphicarpa and V. macrocarpa. Japan. J. Breed., 18: 283-290.

Yamamoto, K., (1969). On the interspecific hybrid between Vicia macrocarpa and V. angustifolia var. segetalis. Tech. Bull. Fac. Agr. Kagawa Univ., 20: 101-107.

Yamamoto, K., (1971). On the cytogenetical analysis of flower colour and pod colour in the interspecific hybrid between Vicia sativa and V. macrocarpa. Japan. J. Breed., 21: 204-210.

Yamamoto, K., (1974a). On the interspecific hybrid between Vicia pilosa and V. angustifolia, V. pilosa and V. macrocarpa. Tech. Bull. Fac. Agr. Kagawa Univ., 25: 179-190.

Yamamoto, K., (1974b). Hybrid plants between two races of Vicia amphicarpa having 2n = 10 and 2n = 12 chromosomes. Japan. J. Breed., 24(2): 73-80.

Yamamoto, K., (1977). On the interspecific hybrid between Vicia amphicarpa and V. sativa. Tech. Bull. Fac. Agr. Kagawa Univ., 28(1): 23-31.

Yamamoto, K., (1980). On the interspecific progenies between Vicia pilosa M.B. and V. amphicarpa Dorth. Japan. J. Breed., 30(4): 351-357.

Yamamoto, K., (1984). A note on interspecific hybridization between V. narbonensis and its related species. In "Systems for cytogenetic analysis in V. faba". Eds., G.P. Chapman & S.A. Tarawah. pp. 141-142. Martinus Nijhoff, The Hague, Netherlands.

Yamamoto, K., Moritani, O. & Ando, A., (1982). Karyotypic and isozymatic polymorphism in species of the section Faba (genus Vicia). Tech. Bull. Fac. Agr. Kagawa Univ., 34(1): 1-12.

Yamamoto, K. & Plitmann, U., (1980). Isozyme polymorphism in species of the genus Vicia (Leguminosae). Jap. J. Gene., 55(3): 151-164.

Youssef, S.S. & Hesemann, C.U., (1985). Nuclear DNA contents of some species of Vicia L. and some Egyptian varieties of Vicia faba L. Egypt J. Genet. Cytol., 14: 111-121.

Zertova, A., (1962). Ein Schlussel zur Bestimmung der tschechoslowakischen Arten der Gattung Vicia L. nach den morphologischen Merkmalen der Samen. Acta Horti Bot. Prag., 113-118.

Zohary, M., (1972). Vicia in Flora Palaestina, Vol. 2: text 194-209, plates 278-298. Israel Academy of Sciences and Humanities, Jerusalem, Israel.

Zohary, D., (1977). Comments on the origin of cultivated broad bean, Vicia faba L. Isr. J. Bot., 26: 39-40.

Zohary, D. & Hopf, M., (1973). Domestication of pulses in the old world. Science, 182: 887-894.

Zohary, D. & Plitmann, U., (1979). Chromosome polymorphism, hybridization and colonization in the Vicia sativa group (Fabaceae). Pl. Syst. Evol., 131: 143-156.