UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

School of Biological Sciences

Linking transcript, QTL and association mapping to understand the genetic control of leaf size and shape in *Populus*

In one volume

By

Harriet Trewin

Oct 2008

Addendum

Whilst working on a paper concerning this work an error was brought to my attention. The diameter measurements taken in 2006 should be in the unit millimetres (mm) and not centimetres (cm).

The following are affected by this information.

Page	Affected
21	Equation 2-7
26	Figure 2-3
28	Figure 2-4 D
156	Figure 4-6

This will not have any effect to my conclusions.

Many thanks

Harriet Trewin

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE, HEALTH AND LIFE SCIENCES

Doctor of Philosophy

Linking transcript, QTL and association mapping to understand the genetic control of leaf size and shape in *Populus*

By Harriet Trewin

Leaf size in *Populus* is an adaptive trait and early indicator of biomass yield. In order to investigate the genetic variance contributing to this variation in leaf size a collection of *Populus nigra* L. were made from across Europe and were planted at a single site in Belgium, in a fully randomized and replicated trial, with leaf traits measured in three consecutive years and biomass estimated at one point in time. Results indicate that leaf traits vary with latitude of sample origin, with significant differences observed in leaf area, epidermal cell number and biomass, but not in leaf shape (leaf ratio), epidermal cell area, stomatal density and stomatal index. Overall a significant positive relationship between latitude of origin and leaf traits was observed with small-leaved genotypes containing fewer epidermal cells observed in the south west (Spain), and large-leaved genotypes occurring in the north and east (the Netherlands, Germany and Italy).

A sequence-based genetic study was conducted to identify Single Nucleotide Polymorphisms (SNPs) associated with leaf phenotype. Given that linkage disequilibrium (LD), decays rapidly ($r^2 = 0.09$) in *P.nigra*, a candidate gene approach for association is valid. Candidate genes were selected from Quantitative Trait Loci (QTL), from a microarray experiment and from bioinformatics and literature searches, identifying sixty robust genes. From this list eight candidate genes were selected for further analysis; *ASYMMETRIC LEAVES 1 (AS1), ASYMMETRIC LEAVES 2 (AS2), ACC OXIDASE (ACO), ERECTA (ER), PHABULOSA (PHAB), ANGUSTOFOLIA (AN), E2Fc* and *LEAFY*. Genetic association was conducted using a General Linear Model (GLM) both with and without population structure. The strongest genetic association was found in *AS1*, a gene involved in leaf initiation that acts by repressing *KNOX* genes to increase cell differentiation. Gene expression of the eight candidate genes were examined across extreme leaf genotypes using real time qPCR (RT-qPCR), at three growth stages. Extreme leaf genotypes consisted of five ‘small’ and five ‘big’ leaf genotypes selected from the association population. Significant differences in gene expression was seen between ‘small’ and ‘big’ genotypes in *AN, AS2* and *AS1*. These results suggest that *AS1* is a strong candidate gene for leaf size.

Table of Contents

List of Figures

List of Tables

Author's Declaration

The contents of this thesis are the result of original work by the author. I hereby declare that this thesis has not been submitted, either in the same or different form, to this or any other university for a degree.

Acknowledgement

I would like to thank my supervisors Professor Gail Taylor (University of Southampton) for providing and supervising this PhD funded partially by the Department for Environment, Food and Rural Affairs (DEFRA) and Popyomics. I would also like to thank Dr Michele Morgante (University of Udine) for his supervision and use of equipment such as the ABI sequencer during my time at Udine University while working on chapter 4, the association study.

I would like to acknowledge Caroline Dixon, Dr Penny Tricker, Dr Anne Rae, Dr Nathanial Street, Dr Laura Graham, Dr Matt Tallis, Maud Viger and Bradly Street (University of Southampton) for all their hard work between 2004 and 2006 joining me on those rainy days in Belgium to collect phenotype data for chapter 2.

For the work in chapter 3, I would first like to thank Dr Anne Rae (University of Southampton) for all her work in constructing the QTL maps and running analysis. I would also like to thank all members of the Taylor lab past and present for data collection of phenotypic traits to enable this study. Also I acknowledge input from Dr Nathanial Street and Dr Laura Graham (University of Southampton) for designing and carrying out microarrays and analysis, enabling me to combine these current genomic techniques to understand the underlying genetics of quantitative traits.

For the work in chapter 4, I would also like to thank Dr Guisi Zaina (University of Udine) and the other members within Michele's lab for their patience when I took my first steps into the world of genetics. For her lessons in DNA extraction, PCR, sequencing, sequence analysis and SNP detection. I would also like to acknowledge Jennifer De Woody (University of Southampton) for be able to pick her brain on the use of STRUCTURE and how to interpret results.

A special thanks to Dave Baddhams and Mick Cotton (University of Southampton) for their help and advice when growing poplar cuttings for the first time in a glasshouse environment. I acknowledge Maud Viger (University of Southampton) for also helping me collect all the phenotype data and supervising an early leaf collection with Dr Laura Graham, Dr Nathanial Street and Suzie Milner for RNA extraction.

Finally I would like to thank Dr Anne Rae, Dr Nathaniel Street, Dr Matt Tallis, Dr Carol Wagstaff, Dr Penny Tricker and Jennifer De Woody (University of Southampton) for reading this thesis, providing advice on editing and for keeping me motivated.

I would like to dedicate this thesis to my mother who has been a mountain of support even in her time of need whilst recovering from breast cancer last year. Words can't express my thanks and gratitude to my family, partner and friends for their support and love. I couldn't have done it without them.

List of abbreviations and symbols

Abbreviation	Definition
α	alpha
A	Adenine
A_0	First leaf area measurement
α_i	The population effect
A_i	Last leaf area measurement
ABA	Abscisic Acid
acl	acaulis
a[CO_2]	Ambient CO_2
ACO	ACC OXIDASE
AFLP	Amplified Fragment Length Polymorphism
AGR	Absolute Growth Rate (mm^2d^{-1})
AN	ANGUSTOFOLIA
ANOVA	Analysis of Variance
ANT	AINTEGUMENTA
AS1	ASYMMETRIC LEAVES 1
AS2	ASYMMETRIC LEAVES 2
β	Beta
BA	Basal Area
bp	base pair
$\beta_{j(l)}$	The clone effect
BOP1	BLADE-ON-PETIOLE
c	Cycle threshold
C	Cytosine
^oC	Degrees Celsius
CA	State of California
CBD	Cellulose Binding Domain
CC1	First Coppice Cycle

CC2	Second Coppice Cycle
CDK	Cyclin Dependent Kinase
CER1	ECERIFERUM 1
C_G	Genotype of candidate gene
CI	Confidence Intervals
CIM	Composite Interval Mapping
CLV1, CLV2 & CLV3	CLAVATA1, 2 &3
cm	Centimetre
cM	CentiMorgan
cm^2	Centimetre squared
CNPL	Cell Number Per Leaf
CO_2	Carbon Dioxide
cpDNA	Chloroplast DNA
Ct	Comparative Ct method is used to determine gene expression in qRT-PCR
Cy3	Cyanine 3-dNTP
Cy5	Cyanine 5-dNTP
CycA	A-type cyclins
CycD	D-type cyclins
d	day
D	Diameter (cm)
D’	LD measurement including recombinational history
DE	State of Delaware
DEFRA	Department for Environment, Food and Rural Affairs
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide - triphosphate
DPI	Dots per inch
Dro	Drought
ε	Error

e	The PCR efficiency
"E	Longitudinal coordinate east of Greenwich meridian
e[CO_2]	Elevated CO_2
ε_{ijkl}	The residual error
ECA	Epidermal Cell Area (μm^2)
ER	ERECTA
ESTABLISH	EU project to establish a network of excellence
ESTs	Expressed Sequence Tags
EU15	European Union (15 members)
EUFOGEN	European Forest Genetic Resource Programme
EXT	endoxyloglucan transferase
F_1	First Filial generation
F_2	Second Filial generation
FIL	FILAMENTOUS FLOWER
FP	Fluorescence Polarization
FRET	Fluorescence Resonance Energy Transfer
F_{ST}	A statistical term used to measure population differentiation
FSTAT	Computer package which estimates and tests gene diversity and differentiation statistics
F value	Fisher-Snedecor Distribution
g	Grams
G	Guanine
G1	Pre-synthetic interphase of the cell cycle
G2	Post-synthetic interphase of the cell cycle
GLM	General Linear Model
GMC	Guard Mother Cell
GMT	Greenwich Mean Time
GRF	Growth Regulating Factor
h	hour (s)
h^2	Broad Sense Heritability

H	Height (cm)
ha	hectare
H_2O	Water
HSP	High Scoring Segment Pairs
i	ith population
IBD	Identity By Descent
IEA	International Energy Agency
INDELS	Insertions or deletions of DNA sequences
INRA	French National Institute for Agricultural Research
j	jth clone
JGI	Joint Genome Institute
k	kth block
KAN1 & KAN2	KANADI1 &2
KD	Kilodaltons
KRP	Kip-related proteins
l	lth individual
LA	Leaf Area (mm^2)
LD	Linkage Disequilibrium
LD -r^2	LD measurement, involves recombinational and mutational history
LG	Linkage Group
LIGNOME	The French Genomics Initiative for Long-Lived Species
LL_0	First leaf length measurement
LLE	Leaf Length Extension
LL_i	Last leaf length measurement
Ln-1	First Fully Unfurled Leaf
LOD	Likelihood Of Odds
Log	Logarithim
LRR-kinase	Leu-rich repeat receptor kinase
LS	Least squares

LW_0	First leaf Width measurement
LWE	Leaf Width Extension
LW_i	Last leaf Width measurement
M	Mitosis
Ma	Millions of years ago
MAF	Minor Allele Fequency
MAPKK	Mitogen-Activated Protein Kinase
Mbp	Mega base pairs. A unit size of DNA, equal to a million base pairs in a double stranded nucleic acid
mg	Micro grams
Mg ha^{-1} yr^{-1}	Metric tons per hectare per year
MIM	Multiple Interval Mapping
mm	Millimeter
mm^2	Millimeter squared
MMC	Meristemoid Cell
Mol (M)	Mole
mP	Florescence Polarization value
mRNA	Messenger RNA
^oN/”N	Degrees north of the equator
N	Nitrogen
N_0	Initial number of DNA molecules
N_C	Number of DNA molecules at the threshold florescence
NCBI	National Center for Biotechnology Information
ns	Not significant
ODT	Oven Dry Ton
ORF	Open Reading Frames
OTC	Open Top Chamber
P	Predicted
pH	Potential of Hydrogen, a measure of the acidity or alkalinity of a solution.

PHAB	PHABULOSA
PHAN	PHANTASTICA
PICME	Platform for Integrated Clone Management
PNH	PINHEAD
PHV	PHAVOLUTA
Q	Population structure
Q_{ST}	A statistical measurement of genetic differentiation at quantitative traits
QTL	Quantitative Trait Loci
r^2	Radius
RAPD	Randomly Amplified Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RGR	Relative Growth Rate (mm^2d^{-1})
RIL	Recombinant Inbred Line
RNA	Ribonucleic Acid
ROT3	ROTUNDIFOLIA
rpm	Rotations per minute
RT	Room Temperature
RT-qPCR	Real Time Quantitative Polymerase Chain Reaction
S	DNA synthesis phase of the cell cycle
SAM	Shoot Apical Meristem
SBE	Single Base Extension
SD	Stomatal Density
SDD1	Stomatal density and distribution
SCMV	Sugarcane Mosaic Virus
SI	Stomatal Index
SLA	Specific Leaf Area ($mm^2\ g^{-1}$)
SNP	Single Nucleotide Polymorphisms. A variation in the genetic code whereby a single base in the DNA differs from the norm for that position
SRC	Short Rotation Coppice

SRF	Short Rotation Forest
SSR	Simple Sequence Repeat
STS	Sequence Tagged Site
T	Thymine
TAIR	The *Arabidopsis* Information Resources
Tr	Trait
TASSEL	Software package to evaluate traits associations, evolutionary patterns and LD
TDI-FP	Temple directed Dye terminator Incorporation with Florescence Polarization Detection
µ	The grand mean
UK	United Kingdom
μm^2	Micrometer squared
µmol	Micromole
µl	Microlitre
USA	United States of America
V	Stem Volume Index
VNTR	Variable Number of Tandem Repeats
WUS	WUSCHEL
XET	Endoxyloglucan transferase
XTH	Xylogulcan endotransglucosylase
γ_k	The block effect
YAB2 & YAB3	YABBY2 & 3
yr	Year
Z	Phenotype value
%	Percentage
Δ	Delta
Π	Pi
*	Statistically significant– $P<0.05$
**	Statistically highly significant – $P<0.01$

***	Statistically very highly significant – P<0.001
σ	Variance
σ_w^2	Genetic Variance
σ_E	Environmental Variance
>	Less than
<	More than
04	Year 2004
05	Year 2005
06	Year 2006

List of Roman Numerals

I	One
II	Two
III	Three
IV	Four
V	Five
VI	Six
VII	Seven
VIII	Eight
IX	Nine
X	Ten
XI	Eleven
XII	Twelve
XIII	Thirteen
XIV	Fourteen
XV	Fifteen
XVI	Sixteen
XVII	Seventeen
XVIII	Eighteen
XIX	Nineteen

1 . Literature Review

1.1 Introduction

Interest in the value of wood products has triggered studies in tree morphology and physiology down to the molecular level. Trees benefit our everyday lives in many ways, from their aesthetic beauty to their economic value, producing most of the world's terrestrial biomass and dominating most terrestrial ecosystems (Brunner *et al*., 2004, 2007). Their success is due to the formation of secondary xylem, having age-related phase changes in many aspects of morphology and physiology, coping with varying biotic and abiotic factors, transporting water, nutrients and macromolecules over long distances and possessing systems for coordinating development and environmental responses at the whole plant level *(Brunner et al*., 2004). Some tree species have survived for over 5000 years, showing remarkable developmental traits (Groover *et al*., 2004) and adaptations to become the most successful perennial plants. At the molecular level understanding differences between species can help us understand how trees have survived and adapted. This can also lead to breeding more efficient trees, in extreme environments for economic goals.

1.2 Poplar as a model tree

There are 40 members within the genus *Populus*, are commonly called poplars and include aspens and cottonwood (Sterky *et al*., 2004). Poplars are distributed in a wide environmental range across the Northern hemisphere from the tropics to above the Arctic Circle. Commercially-grown poplars can produce 10-25 Mg ha^{-1} yr^{-1} of dry woody biomass (Pellis *et al*., 2004) and these trees have a genuine wood value for timber, plywood, pulp, excelsior (packaging material) paper, pallets, soft board and hard board (Cervera *et al*., 2001a). Commitments to the Kyoto protocol have also resulted in an agreement to reduce UK emissions by 60 % by 2050. *Populus* hybrids used as bioenergy have the potential to provide global energy needs (Bunn *et al*., 2004). *Populus* species are dioecious and wind pollinated with dispersal in the early summer, are able to colonize disturbed sites and are found frequently in riverine floodplains (Bradshaw *et al*., 2000). Poplars are highly polymorphic for photoperiodic responses (Howe *et al*., 1995;Howe *et al*., 1998), crown architecture (Dunlap *et al*., 1995), cold hardiness (McCamant and Black, 2000) and wood structure. Poplar trees perform a range of ecological services including carbon sequestration, bioremediation, nutrient cycling, biofiltration and the creation of diverse habitats (Brunner *et al*., 2004).

Populus species is the model angiosperm for tree molecular biology and biotechnology (Bradshaw *et al*., 2000;Taylor, 2002;Brunner *et al*., 2004); they are readily transformable,

propagate vegetatively and display rapid growth and flowering. Poplar has a modest estimated genome size of ~550 Mega base pairs (Mbps) (Wullschleger *et al*., 2002) which is similar to rice and ~ 40 times smaller than pines (Brunner *et al*., 2004). Many members of the genus display high levels of ecological and intraspecific diversity. *Populus* trees were the first to be genetically transformed and regenerated and *Populus* is the tree genus with the most published studies (Taylor, 2002).

In 2006 the *Populus trichocarpa* genome sequence was released, a single female genotype 'Nisqually 1' using a whole shotgun sequence and assembly approach (Tuskan *et al*., 2006b). The *Populus* genome size was then estimated at 485 Mbp, divided into 19 chromosomes, which is four times larger than the genome of *Arabidopsis thaliana L.* (referred to as *Arabidopsis*) (Tuskan *et al*., 2006b). As *Populus* is predominantly outcrossing ,this species shows high levels of gene flow and heterozygosity (within individual genetic polymorphisms) (Tuskan *et al*., 2006b). The sequencing of 'Nisqually 1' identified 1,241,251 single nucleotide polymorphisms (SNPs) or insertion/deletion polymorphisms (Indels) at a rate of 2.6 polymorphisms per kilobase (Tuskan *et al*., 2006b). A present 45,555 protein coding gene loci have been identified in the *Populus* genome (Tuskan *et al*., 2006b), 89% of the gene models had homology to nonredundant set of proteins from the National Center for Biotechnology Information (NCBI) and 91% had homology to *Arabidopsis* genes (Tuskan *et al*., 2006b). *Populus* and *Arabidopsis* lineages diverged 100 to 120 million years ago (Ma), with a genome wide duplication event occurring in the salicoid-specific lineage 65 Ma (Tuskan *et al*., 2006b). As a result poplars are palepolypoids (polypoids that have undergone diploidization). *P.nigra, P.deltoides* and *P.trichocarpa* are the major species of poplar that are used for poplar breeding in Europe (Cervera *et al*., 2001a). The commercial clones that exist in Europe are derived from interspecific crosses between *P.deltoides* and *P.trichocarpa* and between *P.deltoides* and *P.nigra* and their backcrosses (Cervera *et al*., 2001a).

P.nigra, the European black poplar from the family *Salicaceae,* is a key species in floodplain forests around Europe (Storme *et al*., 2004b). The natural distribution of *P.nigra* ranges from North Africa to Ireland in the west and to Russia and China in the east (Cottrell *et al*., 2005). Cotrell *et al*. (2005) study on *P.nigra* chloroplast DNA (cpDNA) concluded that black poplar had an ice age refugium in Spain and a refugium in the Balkans (Smulders *et al*., 2008). *P.nigra* has an economic interest in soil protection, afforestation and domestic uses. Other uses, as its wood is lightweight, include as a raw material in clogs, fruit baskets, furniture and flooring (Cottrell *et al*., 2005). Finally *P.nigra* has a huge ecological importance as it is an

indicator species of riparian woodland and early successional stages of floodplain woodlands. It is a host to wildlife and important in flood control. *P.nigra* is similar to some species of balsam poplars in the section *Tacamahaca* and, oddly, its chloroplast genome is from the sympatric *P.alba*. This suggests that *P.nigra's* genome is a combination of a least two different species, giving rise to two scenarios for its evolution (Smith and Sytsma, 1990). Firstly, *P.nigra* may be misplaced taxonomically and actually derived from *Tomentosae* or secondly *P.nigra* may be derived from an ancient hybridization event in Eurasia involving an ancestor or relative of *P.alba* (Smith and Sytsma, 1990).

1.3 The Leaf

Interest has arisen in using biomass crops as an alternative to depleting oil reserves and to slow the increase of atmospheric carbon dioxide (CO_2). Commitments to the Kyoto protocol have also resulted in an agreement to reduce UK emissions by 60 % by 2050 (European commission, 2003). *Populus* hybrids used as bioenergy have the potential to provide global energy needs (Bunn *et al*., 2004). Therefore productivity of these crops is vital.

In the Devonian period 360 million years ago (Ma) two types of leaves evolved from the branch system: microphylls and megaphylls (Tsukaya, 2006). The leaf is defined as a lateral organ that develops on a shoot and has dorsoventrality (Tsukaya, 2006). Leaves play a major role in photosynthesis; they are the light capturing organ and the site of photochemical reactions. Leaves are responsible for most carbon fixation in a plant and therefore essential to plant productivity and survival (Sinha, 1999). The diversity of leaf form and function is substantial in nature, demonstrated by diverse examples such as leaves that form insect traps in carnivorous plants, leaf structure modified into thorns and spines, the hollow trunks of banana trees and non-photosynthetic storage organs in bulbs (Kerstetter and Poethig, 1998). For optimum absorption of light leaves have to be as wide as possible, whereas for optimum gas exchange leaves should be as flat and thin as possible (Tsukaya, 2006).

Traits such as leaf area development and leaf size are strongly linked to productivity in *Populus* (Bunn *et al*., 2004;Rae *et al*., 2004). An experiment investigating the productivity of poplar genotypes found that productivity was tightly correlated with stem and leaf traits such as length, diameter, total leaf area and individual leaf area (Monclus *et al*., 2005). Rae *et al.* (2004) also found traits that increased yield were number of leaves on the leading stem, plastochron index (an indicator of leaf production) and leaf area. At the cellular level, epidermal imprints show a strong correlation between large leaves and a high number of

epidermal cells (Bunn *et al*., 2004). Theory suggests that individual leaf size can be determined by cell division or cell size. Strong correlations between epidermal cell number per leaf with stemwood yield (Rae *et al*., 2004) suggests that determination of cell traits will be valuable knowledge for biomass crop research. Classical cell theory suggests that leaves are made by the sum and behavior of each cell, therefore cell division solely controls leaf size (Cookson *et al*., 2005). However organismal theory suggests that when cell division is impeded growth can be compensated by an increase in cell size (Cookson *et al*., 2005). Recently a Neo cell theory has been proposed whereby co-operative compensation between cell division and cell size determines leaf size (Cookson *et al*., 2005).

1.4 Leaf development

Variation in leaf shape and size is due to species–specific patterns in leaf development (Kerstetter and Poethig, 1998). In dicot species, leaf development initiates when leaf primordium first emerges as a small ridge or leaf buttress on the shoot apical meristem (SAM). The SAM contains the proliferative cells that give rise to the primary shoot and most importantly, gives rise to lateral organs such as leaves by consistently managing the balance between cell division and differentiation. The SAM is organized into zones; the central zone, peripheral zone and the morphogenic zone. The central zone is composed of "stem cells" which are undifferentiated, self-renewing cells that give rise to many cell populations. The central zone is surrounded by the peripheral zone composed of apical initial cells. Initial cells are undifferentiated, cytoplasmically-rich cells that divide frequently to provide cells for the morphogenic zone where organ primordial are formed on the flanks of SAM (Howell, 1998). Recent studies have shown that alterations in cell balance between zones alter the size of the SAM affecting lateral organ number and size (Nelissen *et al*., 2003). For example *CLAVATA1 (CLV1), CLV2* and *CLV3* genes promote differentiation of stem cells; *clv* mutations lead to accumulation of undifferentiated cells at shoot and flower meristems (Song and Clark, 2005b). A key component of the regulation of the size of SAM in *Arabidopsis* has been the identification of the *WUSCHEL* (*WUS*) gene, which enhances the expression of *CLV3* , which in turn causes *CLV3* to down-regulate the expression of *WUS*, resulting in a feedback loop to control SAM size (Song and Clark, 2005b). Growth on the flanks of the SAM by cell division and differentiation is coordinated along three principal axes. (i) proximal-distal axis, where the proximal petiole and the distal blade are positioned, (ii) the adaxial-abaxial axis, which defines the position of tissue and (iii) the medial-lateral axis, which defines the position of the central midrib relative to the lateral blade and leaf margins (Ohno *et al*., 2004).

Change from indeterminate meristem cells to determinate leaf primordium is characterized by the down regulation of class *1 KNOTTED1-LIKE HOMEOBOX (KNOX*) genes by *ASYMMETRIC LEAVES 1 (AS1), ASYMMETRIC LEAVES 2* (AS2) (Chalfun *et al*., 2005;Zgurski *et al*., 2005). *AS2* belongs to the *LATERAL ORGAN BOUNDARIES* (*LOB*) gene family and is involved in the development of a symmetrical expanded lamina (Iwakawa *et al*., 2007). Recent studies have identified a number of genes involved in adaxial–abaxial patterning where mutations cause radially symmetrical leaves. These genes include *PHANTASTICA* (*PHAN*) (Waites *et al*., 1998), *PHABULOSA* and *PHAVOLUTA* (*PHB* and *PHV*) and *REVOLUTA* (*REV*) (McConnell *et al*., 2001). Members of the *YABBY* gene family (Siegfried *et al*., 1999b) such as *KANADI1* (*KAN1*) and *KANADI2* (*KAN2*) (Eshed *et al*., 2001) are involved in the specification of abaxial cell fate in the leaf lamina (Iwakawa *et al*., 2007). Growing leaves often show a promimal- distal gradient of cell division (Donnelly *et al*., 1999) where the greatest amount of cell division occurs in the proximal region of the leaf. Genes identified as controlling cell division and differentiation in the leaf include: *BLADE- ON-PETIOLE 1* (*BOP1*) (Ohno *et al*., 2004), , *AINTEGUMENTA* (*ANT*) (Mizukami and Fischer, 2000), *JAG* (Ohno *et al*., 2004), the *GROWTH-REGULATING FACTOR* family (Horiguchi *et al*., 2005a), *ANGUSTIFOLIA* (*AN3*) , *ROTUNDIFOLIA* (*ROT3*), ROTUNDIFOLIA 4 (ROT4) (Horiguchi *et al*., 2005a) and Figure 1-1.

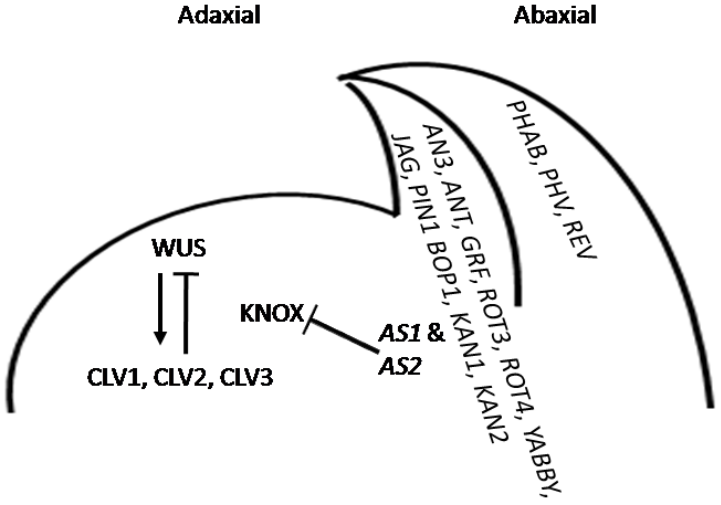


Figure 1-1: A schematic representation of the genetic control of leaf development. Class I KNOX transcription factors (STM, BP, KNAT2 and KNAT6) are expressed throughout the SAM, with CLAVATA1 (CLV1), CLV2 and CLV3 working as a negative feedback loop with WUSCHEL (WUS) producing a pool of founder cells for leaf initiation. ASYMMETRIC LEAVES 1 (AS1) and ASYMMETRIC LEAVES 2(AS2) repress KNOX genes to initiate leaf primordia growth. ANGUSTOFOLIA (AN3), AINTEGUMENTA (ANT), GROWTH-REGULATING FACTOR (GRF) family, ROTUNDIFOLIA3 (ROT3), ROTUNDIFOLIA (ROT4), YABBY, JAG, PINFORMED (PIN1) and BLADE-ON-PETIOLE (BOP1) control lamina growth. Whereas PHABULOSA (PHAB), PHAVOLUTA (PHV), REVOLUTA (REV) control adaxial –abaxial patterning.

1.5 Cell cycle

The ability to control when cells divide would have profound impacts on the development of an organism. Mersistems and meristematic regions are the primary locations within a plant where cell cycling occurs and this plays an important role in plant development including organ morphogenesis, cell proliferation within tissues and cell differentiation (Donnelly *et al.*, 1999).

The cell cycle is a process by which cells reproduce themselves and their genetic material, consisting of 4 phases: G1 (pre-synthetic interphase), S (DNA synthesis phase), G2 (post–synthetic interphase) and M (Mitosis). In G1 the nuclear DNA prepares for replication by assembling a pre-replication complex at the origins of replication along the chromatin. This is followed by the S phase where DNA is replicated, then in G2 cells prepare for mitosis in the final M phase.

Phosphoregulation of proteins is a major biochemical feature of the G1/S and G2/M transitions in the cell cycle and cyclin dependent kinases (CDKs) are the key players (Francis, 2007). Cyclin-dependent protein kinases are catalytic enzymes that, along with their activating subunits cyclins, regulate the cell cycle by controlling transition into each phase. Transition from G1 to S involves G1 cyclins, whereas transition from G2 to M involves mitotic cyclins. These protein kinases have been highly conserved throughout plant evolution, indicating differences must be found between species to give variation in leaf size and shape. Studies investigating the role of these genes involved in the cell cycle have shown differences in growth rates and small differences in leaf morphology (Wyrzykowska *et al*., 2002;Scarpella *et al*., 2004). In plants, *CDK*s have been identified and classified into five subtypes (A-E) (Rossi and Varotto, 2002). In the G1 phase D-type cyclins (CycD), some A-cyclins (CycA) and *CDK*s (CDKA, D and E) are expressed. At the G1/S transition *CDKA;1*, CDKC and *CDKE* are expressed, whereas at the S/G2 transition all the way to M phase *CDKA;1* is active.

In *Arabidopsis*, the *CDKA;1* gene is expressed mainly in dividing cells, however it is also present in non-dividing tissue. Its ‘partner- in- crime’ D-type cyclins seem to have an active role in development, evidence can be seen in snapdragon where *CycD* genes are localized in vegetative and floral meristems (Rossi and Varotto, 2002). More specifically, cyclins *D1* and *D3b* are expressed throughout the meristem. *D3a* is limited to the peripheral meristematic region and organ primordial and *CycD3* was found in proliferating tissue of the shoot meristem, young leaves and axillary buds (Rossi and Varotto, 2002). This suggests a

role of *CycDs* in particular in leaf development and growth. Transgenic tobacco lacking the *CDKA;1* and mutant plants of *Arabidopsis* with the *CycD2* gene show larger growth rates in cells, due to a reduction in the length of the G1 phase. Other mutants in *CycD3* caused morphological alterations in the SAM increasing leaf number and delaying senescence (Rossi and Varotto, 2002). Seven genes from *Arabidopsis* have been isolated called the Kip-related proteins (*KRPs)*, which show a cyclin-dependent kinase binding specificity (De Veylder *et al*., 2001) reducing cell cycling and serving as a checkpoint. Over-expression of *KPR2* genes inhibits cell cycle in leaf primordial, giving plants narrower, serrated leaves. The most interesting morphological feature in over-expressing *KRP* is the reduction in leaf area. In De Veylder *et al*.(2001) experiment leaf area was reduced by ~75 % with an increase in leaf thickness. This suggests a role of *KRP* genes in control of the cell cycle and again emphasizes the importance of cyclin-dependent kinases in leaf development.

One purposed theory in plant development is that genes influencing mitotic cell size therefore control plant development (Francis, 2007). A large proportion of evidence has arisen from studies in budding yeast, whereby coordination of division with growth occurs at START (a budding yeast term meaning that the mother cell has started to bud), where cells must reach their optimum size to then enter the cell cycle (Francis, 2007). Two genes have been identified within this mechanism, the first is *WEE1 kinase* and the second is *CDC25*. Studies in *Schizosaccharomyces pombe* have highlighted the importance of *WEE1* and *CDC25* in size regulation, by the mutants *spcdec25*^{oe} and s*pwee1*^{oe} resulting in short and long mitotic cells (Francis, 2007). In *Arabidopsis Arath:WEE1* is expressed in proliferative tissues and demonstrated a phenotype with long epidermal cells, slow root growth and reduced frequency of lateral roots (Francis, 2007). Alternatively Mizukami and Fischer. (2000) study shows that organ size is determined by internal developmental factors, cell number and not cell size. A study of the *Arabidopsis* transcription factor *AINTEGUMENTA* (*AINT*) showed that during organogenesis *ant-1* organs were smaller with fewer cells and that ectopic *AINT* expression allowed petals to proliferate for longer (Mizukami and Fischer, 2000). This leads to a discussion of the ongoing debate in plant development, “cell proliferation simply increases at the presumptive site of leaf formation and this leads to the formation of a new organ” (Fleming, 2006a), which has been discussed above, however things are rarely that simple. Although cell proliferation is certainly associated with leaf formation plant growth and morphogenesis is dependent on the physical characteristics of the cell wall.

1.6 Cell Wall

The cell wall is one of the main differences between plant and animal cells. It serves many purposes including structural support, defining cell shape, protection, storage of carbohydrates and metal ions and the control of signalling molecules. However its role in the size and shape of a cell is the most interesting in leaf development as this is what leads growth. In plant cells, extension of the cell wall can result in larger cells with modified cell shapes (Cosgrove, 1999).

The primary cell wall (the 'growing wall') is composed of cellulose microfibrils embedded in a hydrated matrix composed mostly of neutral and acidic polysaccharides and a small amount of proteins (Cosgrove, 1999). Wall enlargement requires the controlled spreading of the cellulose/matrix network, as a result of rearrangement of matrix polymers (Cosgrove, 1999). Cell wall extension has recently been thought to involve a 'loosening' and there are several mechanisms proposed involving viscoelastic properties and polymer rearrangements. Polymer rearrangements would lead to turgor-driven wall expansion by weakening non- covalent bonds between polysaccharides, cleavage of the backbone of the major matrix polymers (by endoglucanases, pectinases, transglucosylases and hydroxyl radicals) and breakage of crosslinks between matrix polymers (e.g. by esterases) (Cosgrove, 1999). However the protein expansin is the only agent so far shown to have catalyzed wall extension *in vitro* and can start and stop extension quickly without much change in structure, making it a strong candidate for rapid changes in cell wall growth such as hypocotyl elongation.

Expansins were first isolated in 1992 as the mediators of 'acid growth', this refers to the growth rate of cells after being placed in acid when the cell wall becomes more extensible at acidic pH. It has been purposed that expansins disrupt hydrogen bonds between cellulose microfibrils and cross-linking glycans in the wall to directly induce wall extension (Li *et al*., 2003) The model suggests that expansin in a primary wall-loosening factor inducing turgor driven wall extension, whereas xylogucan endotransglucosylases (XTHs) are secondary wall loosening factors rendering primary wall loosening to occur (Li *et al*., 2003). The *Arabidopsis* genome contains 38 open reading frames (ORFs) that encode expansin-like proteins (Li *et al*., 2003). There are three families of expansins in *Arabidopsis* including α-, β- and γ-expansin, where α-expansin is the largest sub-group and is the first to be cloned from cucumber that can induce wall extension in several different plant tissues (Li *et al*., 2003). Expansins' role in leaf development was highlighted by Fleming *et al*. (1997) in an experiment where ectopic expansin was applied to the flanks of tomato vegetative meristems leading to leaf initiation.

However these primordial did not grow on to form leaves, suggesting expansins may not be the only player in controlling cell wall extensibility.

Growth studies in plants have also focused on enzymes capable of breaking down the xyloglucan- cellulose network, such as xyloglucan a metabolizing enzyme (Rose *et al*., 2002). Two independent research groups originally described these enzymes; one group named them xyloglucan endotransglyase (XET), which cut and rejoin xyloglucans that tether adjacent cellulose microfibrils (Fry *et al*., 1992) and the other named them endoxyloglucan transferase (EXT) (Nishitani and Tominaga, 1992), adding an amount of confusion to the scientific community, at present a member of these genes is referred to as a xylogucan endotransglucosylase (*XTH*) (Rose *et al*., 2002). In vivo studies have observed a role of *XTH*s in wall restructuring and wall assembly, plus positive correlations have been shown between *XTH* protein activity and elongation growth in *Arabidopsis, tobacco* and *tomato* (Rose *et al*., 2002), however there is still an ongoing debate. In *Arabidopsis* there are 33 different *XTH* genes (Yokoyama and Nishitani, 2001), therefore it can be expected that different gene products are active in different aspects of cell wall metabolism. Studies have shown that *XTH* gene expression coincides with growth (Vissenberg *et al*., 2005) and that low gene expression of *AtXTH18* and *AtXTH27* results in phenotypic changes (Van Sandt *et al*., 2007), suggesting a role in leaf development.

1.7 Stomata

Stomata are special pores found on the epidermis of leaves, through which the diffusion of CO_2 takes place in all higher plants and many lower plants (exceptions are aquatic plants). After entry into the plant CO_2 is reduced by photosynthesis and stored as sugars or starch. This is then used by the plant to satisfy its energy needs, drive nitrogen assimilation and sulphate reduction and other aspects of the plants intermediary metabolism.Gas exchange between the leaf and air is dependent on diffusion controlled by the opening and closing of the stomatal complex, which is bordered by a pair of unique cells called guard cells, which in turn are surrounded by subsidiary cells. Guard cells act as hydraulically operated valves, by taking in water which causes swelling and therefore opening of the pore when CO_2 is required for photosynthesis. Stomata close (become flaccid) in response to water stress (Ellsworth, 1999). Therefore, in general, stomata open in low CO_2 concentrations and close in high CO_2 concentrations. Stomata are usually found on the abaxial surface of the leaf, however they are also present on the adaxial surface of many species. Variations are mainly due to species adaption to the environment, for example water lilies that have stomata present only on the

adaxial leaf surface. Typical stomata of dicots have two guard cells, which are kidney shaped, whereas monocots usually show an elongated dumb-bell shape (Hetherington and Woodward, 2003). The guard cells that make up the stomata contain very few chloroplasts and are not actively involved in photosynthesis.

Stomata are produced by a specialized cell lineage, found in developing shoot epidermis and after epidermal maturation (Bergmann and Sack, 2007). Therefore mutations in stomatal development genes can affect the physiology of the entire plant. For example the gene *STOMATAL DENSITY AND DISTRIBUTION 1* (*SDD1*) is expressed in meristemoids (a small stomatal precursor cell) and is involved in stomatal development, as overexpression of *SDD1* represses stomatal divisions. At the whole plant level *sdd1* plants have a higher stomatal density and can assimilate 30% more carbon than wild type plants (Schluter *et al*., 2003), this being a very appealing trait for biomass crops, therefore stomatal number is also under investigation in recent years.

1.8 Aims and Objectives

A recent boom in high through-put technologies has resulted in an increasing number of studies mapping quantitative trait loci (QTL) for biomass and leaf traits (Wullschleger *et al*., 2005;Rae *et al*., 2007), senescence (Rae *et al*., 2006), rust resistance (Yin *et al*., 2004b), osmotic potential (Tschaplinski *et al*., 2006) and bud set and bud flush (Frewen *et al*., 2000) in poplar. Alternatively to QTL analysis, high through-put microarrays have enabled scientists to analyse the expression of hundreds and thousands of genes in a single experiment quickly and efficiently. Most studies in poplar have been used to identify genes differentially expressed in contrasting environments (Moreau *et al*., 2005;Taylor *et al*., 2005;Street *et al*., 2006). Few studies have combined QTL and microarray techniques to identify candidate genes controlling leaf development. QTL and microarray studies theoretically could identify thousands of genes involved in a trait of interest, therefore combining techniques should be more powerful than using either method alone.

QTL studies are not possible in studies of natural populations; however Association approaches have been developed as an alternative (Table 1-1). First used effectively in human genetics it provides new benefits such as; the ability to examine unrelated individuals (i.e. a natural population), higher resolutions of genetic differences depending on Linkage Disequilibrium (LD), a larger number of alleles per locus can be tested and relatively rapid results since no mapping populations are required (Buckler and Thornsberry, 2002). There are

two applications of association analysis: genome scans and the candidate gene approaches. In genome scans Single Nucleotide Polymorphisms (SNPs) markers are placed across the genome at an appropriate density, whereas candidate-gene approaches involves sequencing candidate genes (Flint-Garcia *et al*., 2005).Success of either of these methods depends on the degree of LD and population size. Success in genome scans occurs with a species with a moderate to extensive LD (Flint-Garcia *et al*., 2005) because species with low LD need many markers to cover the genome therefore a candidate gene approach is used in this case.

Table 1-1 A comparison of Quantitative Trait Loci (QTL) and Association mapping.

Mapping technique	Advantages	Disadvantages
Quantitative trait Loci (QTL) mapping	- Requires inbred lines - Small population sizes possible, 300 individuals or more - Relatively inexpensive depending on markers used - Both co-dominant and dominant markers can be used	- Time consuming - Low resolution - Many markers needed
Association mapping based on Linkage Disequilibrium (LD)	- High resolution depending on LD - Requires natural populations - Candidate genes can be analysed without evidence of linkage - Rapid as no mapping population needed	- Large population sizes needed, 500 individuals or more - Affected by population structure - Expensive

The main aims of this study are:

- To quantify genetic variability in leaf size and shape within a natural population of *Populus nigra*
- To select candidate genes for leaf development by combining data from microarray analysis, QTL maps, literature searches and bioinformatics
- Use well characterized candidate genes to carry out association analysis combining phenotype, genotype and population structure data.
- To select extreme leaf size genotypes from the whole population study and carry out a detailed phenotype, growth and expression study in a controlled environment to verify findings in the association study.

2 . Phenotypic characteristics – leaf, cell and biomass of a natural collection of *Populus nigra* within Europe

2.1 Overview

In 2000, 6% of Europe's (EU15) primary energy came from renewable sources, and 3.7% of this came from biomass (Tuck *et al*., 2006). The amount obtained from biomass is likely to increase. Biomass is obtained from agriculture and forestry as a residual product from harvesting or from purpose-grown tree crops or other plants. At present short rotation forestry crops use fast growing tree species such as *Populus* (poplar) and *Salix* (willow), which are grown in rotations of 1-15 years to produce 10-15 dry tones $ha^{-1}yr^{-1}$(International Energy Agency ((IEA), 2005). For successful biomass crops future yields need to be as high as possible, therefore selecting naturally high yielding varieties is essential.

The leaf plays a major role in photosynthesis, is responsible for carbon fixation and therefore is essential for plant productivity and survival (Sinha, 1999). Leaves have adapted to give rise to a diversity of forms and functions in nature, for example, within poplar, variation in leaf size and area has been found in clones of differing parentage and hybrid groups (Al Afas *et al*., 2005). Studies have shown that leaves are an early indicator of yield in poplar (Pellis *et al*., 2004;Rae *et al*., 2004;Robinson *et al*., 2004). Therefore, characterization of both leaf and biomass traits is important to identify genotypes providing high yields.

In this study a short-rotation forest (SRF) of *Populus nigra* clones originating from a latitudinal gradient from Spain to the Netherlands, of differing leaf morphology and biomass, were studied over a three year growing season with the aim of identifying traits for high yield for breeding purposes.

2.2 Introduction

Biomass crops are one of the many suggested alternatives for energy production. Under the Kyoto Protocol, the European Union is committed to an 8% reduction in annual greenhouse gas emissions by the first commitment period (2008-2012) and the 'White Paper of the European Commission on renewable sources of energy' set the target at increasing renewable, including biomass, to 12% of the European gross energy consumption by 2010 (Walle *et al*., 2007). This commitment plus the release of agricultural land from set aside suggests that land dedicated to bioenergy will increase in the future. The major objective for biomass crops is to insure that maximum output (i.e. woody biomass) is achieved with minimum input (i.e. fertilization, site preparation) (Pellis *et al*., 2004).

Populus species (poplar and aspens) are one of the leading biomass crops due to their ease of propagation, vigorous sprouting after cutting and suitability for a variety of wood fiber products (Pellis *et al*., 2004). Hybrid poplar can produce 20-25 Mg $ha^{-1}yr^{-1}$ of dry woody mass in optimum conditions and in normal conditions yields are typically 10-15 Mg ha^{-1} yr^{-1} (Laureysens *et al*., 2005). Poplar can also achieve the most efficient use of land by combining close plant spacing, coppicing and short rotation cycles (Pellis *et al*., 2004). Pellis *et al*. (2004) when comparing clonal differences in biomass production between seventeen different clones belonging to six different parentages, discovered a *P.nigra L*. clone *Wolterson* to be the best performing short rotation coppice, with a mean biomass production of 8 Mg ha^{-1} yr^{-1} suggesting *P.nigra* to be a strong candidate for biomass crops, although very limited genetic improvement has yet been achieved in this species.

Leaf shape is a characteristic that enables us to distinguish between related species; differences are a result of adaptation to a particular environment for optimal light capture for photosynthesis. Photosynthetic surface area, duration and efficiency of photosynthetic activity throughout the growing season all affect the amount of carbon fixed and this in turn will affect the size of the tree (biomass). Strong positive correlations have been found between mature individual leaf area and biomass productivity (Rae *et al*., 2004;Robinson *et al*., 2004;Laureysens *et al*., 2005;Marron and Ceulemans, 2006). Theory suggests that individual leaf size is determined by cell division and cell size. Strong correlations between epidermal cell number per leaf with stemwood yield (Rae *et al*., 2004;Robinson *et al*., 2004) suggests that determination of cell traits will be valuable knowledge for biomass crop research enabling breeding programmes to identify early diagnostic traits indicative of yield. Classical cell theory suggests that leaves are made by the sum and behavior of each cell, therefore cell

division solely controls leaf size (Cookson *et al*., 2005). However organismal theory suggests blocking cell division is compensated by an increase in cell size (Fleming, 2006b). Recently however a Neo cell theory has been proposed whereby co-operative compensation between cell division and cell size determines leaf size (Cookson *et al*., 2005). Studies have also found that biomass production is significantly correlated to stem traits such as basal diameter, height (Verwijst and Telenius, 1999) and sylleptic branches (Scarascia-Mugnozza *et al*., 1999).

Recently breeding programmes have incorporated genetic studies to understand how high yielding clones vary, and whether traits are linked to a molecular marker. This information can be used in marker assisted selection practices, therefore traits with a high degree of heritability are favorable. Phenology can be defined as the study of annually recurring biological phenomena, such traits include bud burst and leaf fall, flowering and fruiting in the life cycle of plants (Pellis *et al*., 2004). These traits are adaptive as they determine the duration and timing of the growing season and period of reproduction (Pellis *et al*., 2004). Local adaptation results from a balance between natural selection and gene flow, and will occur if selection is stronger than gene flow (Chuine *et al*., 2000). Latitude offers a complex environmental gradient along which temperature, solar radiation and soil conditions vary, therefore it is not surprising during evolutionary history, species with wide distribution areas have adapted to the local growing conditions. Studies, for example by DeBussche *et al*. (2004) investigating variation in phenology and morphology in Mediterranean *Cyclamen* (common name Sowbread) found that peak leafing showed a bimodal phenology due to the Mediterranean climate, which is characterized by dry summers and cold winters. Therefore *Cyclamen* has adapted to these constraints. Bud set and bud flush are also adaptive traits. In northern and high elevation areas of Europe, trees tend to stop growing earlier in autumn (Skroppa *et al*., 1999), therefore genotypes that set bud early tend to perform less well and be at a competitive disadvantage (Riemenschneider and McMahon, 1993). In general harsh conditions produce inherently small ecotypes stature, which correlate to short development cycles, growing seasons and life spans (Li *et al*., 1998). In relation to biomass, plant size is important. A study by Li *et al*. (1998) investigated variation in *Arabidopsis* from a range of latitudes from 16^{o}N to 63^{o}N. Plant size traits included cotyledon width, rosette diameter, number of rosette leaves, size of the largest leaves, total leaf area and total dry weight per plant, a clinal pattern of latitudinal variation in plant size was seen whereby ecotypes decreased in size with increasing latitude of origin (Li *et al*., 1998). This pattern of variation in size is thought to be common in

plants as many species have followed similar patterns including *Carex aquatilis* (Water Sedge) and *Verbascum Thapsus* (Mullein) (Li *et al*., 1998).

The objectives of this chapter were to; (i) examine genotype differences in leaf characteristics and leaf morphology for different poplar species, (ii) relate leaf characteristics and leaf morphology to biomass and (iii) estimate leaf traits and biomass traits for heritability.

2.3 Material & Method

2.3.1 Plant material and plantation layout

The association population consisted of one species of poplar, *P.nigra*, planted in a common garden experiment in Belgium near the Institute of Forestry and Game Management, Geraardsbergen (50^{O} 46'51.23"N). Unrooted hardwood cuttings were planted in the spring of 2004 derived from collections from the European Forest Genetic Resource Programme (EUFORGEN: (du Cros *et al*., 2001)) and the French National Institute for Agricultural Research (INRA (: (Villar *et al*., 1995)). The plantation consisted of 500 genotypes selected along river systems distributed along a latitudinal gradient. Genotypes originated from Spain, Germany, the Netherlands, France and Italy. The site was set out in six randomized blocks, each consisting of 479 cuttings at 0.75 x 2.0 metre spacing. A double row of the cultivar 'Muur' was planted around the entire trial at the same spacing to serve as a buffer. Cuttings were cut back in the spring of 2005 and side shoots cut back in June 2005 to leave one leading stem. No fertilization or irrigation was applied for the duration of the trial. Mechanical weed control was performed three times and trees treated against rust with fungicides every three weeks between March and September, throughout the experiment.

2.3.2 Leaf characteristics

Leaf traits were measured between 17^{th} and 26^{th}August of 2004, 2005 and 2006, from all genotypes and control trees. Traits included leaf area (mm^2), leaf length (mm) and leaf width (mm) of mature leaves. However, in the first growing season in 2004, semi-mature leaves were identified by counting down five leaves from the leaf just fully emerged; this was termed leaf age Ln-5.

Mature leaves were scanned using an Umax Astra 6700 scanner, at 200DPI, in black and white, and saved. The scanned images were then processed in Southampton using Image J (Image J.1.32j, Wayne Rasband, USA). Leaf outlines were selected by finding thresholds and then measured to obtain leaf area (mm^2). Leaf length and width were also measured and results used to calculate leaf ratio as shown in equation 2-1.

$$Leaf\ ratio = \frac{Leaf\ length\ mm}{Leaf\ width\ mm} \tag{2-1}$$

Leaves were collected into brown paper bags, were dried for 48 hr in an oven at 80^{o}C, and then weighed (mg) to obtain the dry mass. Specific leaf area (SLA) was then calculated using equation 2-2.

$$SLA = \frac{Leaf\ area\ mm^2}{Leaf\ dry\ weight\ mg} \quad (2\text{-}2)$$

2.3.3 Epidermal Cell Imprints

Cell imprints were taken in 2004 and 2006 on the tree from mature leaves; each individual tree was sampled. Imprints were taken from the abaxial (bottom) surface of the leaf on the basal section. An area approximately 1cm^2 was painted with clear nail varnish and left to dry for five minutes. Sellotape was then placed on the nail varnish with a little pressure from the thumb, and then peeled off gently. This left a cell imprint on the sellotape (Gardner *et al*., 1995) that was then placed on a glass microscope slide and labeled with the correct line, row and genotype name. The slides were placed in a dark container for later analysis in the laboratory.

The slides were viewed on a Zeiss microscope and images captured with a digital camera attached at x 400 magnification at a 100% zoom. Images were then imported for image processing and analysis using ImageJ for windows (Image J.1.32j, Wayne Rasband, USA). Number of epidermal cells and stomata were counted within image, then from counts stomatal density (SD) and stomatal index (SI) were calculated (equation 2-3 & 2-4). Average epidermal cell area (CA) was calculated by drawing around, 10 epidermal cells to get areas (mm^2) and the mean taken. Average cell area was then used to calculate cell number per leaf (CNPL) as shown in equation 2-5.

$$SD = \frac{\sum Stomata}{Field\ of\ view} \quad (2\text{-}3)$$

$$SI = \frac{\sum Stomata\ in\ field\ of\ view}{\sum CN + \sum Stomata}\ x\ 100 \quad (2\text{-}4)$$

$$CNPL = \frac{Leaf\ area\ mm^2}{ECA} \quad (2\text{-}5)$$

2.3.4 Biomass Traits

Tree height was measured in 2004 and in 2005 during (August) and end of the growing season (December), with a metre rule, as height is a strong indicator of biomass (Scarascia-Mugnozza *et al*., 1999;Rae *et al*., 2004). In 2006 stem diameter was measured one metre above ground level using manual callipers. Stem diameter was then converted to total basal stem area (equation 2-6), π is 3.142 and r^2 radius. Stem volume index (V) was then estimated using equation 2-7, whereby l is height (cm) from 2005 (Oct) and BA is basal area (cm^2) from 2006.

$$Basal\ area\ (BA) = \pi\ x\ r^2 \tag{2-6}$$

$$V = l\ x\ BA\ cm^2 \tag{2-7}$$

2.3.5 Phenotypic statistical analysis

Quantitative traits were analysed using the statistical package Minitab 15 for windows (Minitab Inc, Philadelphia USA). The relationship between phenotype and latitude of sample origin was analysed using a linear regression analysis, significant correlations were shown as *, P<0.05:*, P<0.01;**, P<0.001;***. A general linear ANOVA model was used to find the significance of population, genotype within population and block. Populations included genotypes within latitudes shown in Table 2-1.

Table 2-1*: P.nigra* populations. Populations were based on river system over a population gradient from southern Spain to the Netherlands. *P.nigra* genotypes were derived from collections from EUFROGEN and INRA and planted in 2004 in a common garden experiment in Belgium.

River system	Population Name	Location [a]	Number of genotypes
Loire Est Orl, France	Loire Est	47°09'15.54"N, 2°36'00.00"E	26
Loire W Orl, France	Loire WO	46°43'05.83"N, 0°09'22.74"E	21
Drome, France	Drome 1	44°25'01.34"N, 5°16'20.80"E	63
Drome, France	Drome 6	44°27'14.57"N, 4°33'11.46"E	63
Ain, Arc & Cher, France	Individual Clone F	45°12'58.90"N, 2°52'46.39"E	6
Durance, France	Durance	43°28'14.41"N, 5°18'01.70"E	12
Ebro, Spain	Ebro1	41°33'19.90"N, 1°12'40.27"E	2
Ebro, Spain	Ebro2	41°21'09.13"N, 0°26'28.85"E	26
Rhine, Germany	Rhine	49°29'34.56"N, 8°17'57.13"E	54
Ticino, Italy	Ticino (N)	45°10'18.54"N, 8°34'51.63"E	63
Ticino, Italy	Ticino (SN)	45°07'38.05"N, 9°02'08.05"E	44
Rhine, Netherlands	Netherlands	51°48'01.07"N, 5°23'09.29"E	50

a Median value are given if the original data give only the range

Variance of each phenotypic trait was estimated among genotypes within a population and among populations using the linear model taken from Hall *et al*. (2007) shown in equation (2-8).

$$Z_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \varepsilon_{ijkl} \quad (2\text{-}8)$$

Where Z_{ijkl} is the phenotype of the lth individual in the kth block from the jth clone from the ith population. In equation (2-8), μ is the grand mean, α_i is the population effect, $\beta_{j(i)}$ is the clone effect, γ_k is the block effect and ε_{ijkl} is the residual error. Significant effects are shown as *, P<0.05:*, P<0.01;**, P<0.001;***. Within population broad sense heritability was then calculated by dividing the genetic variance (σ_w^2) estimated from ($\beta_{j(i)}$) by the total variance as shown in equation 2-9 (Hall *et al*., 2007).

$$h^2 = \frac{(\sigma_w^2)}{(\sigma_w^2 + \sigma_E)} \quad (2\text{-}9)$$

σ_E is the environmental variance calculated from ε_{ijkl} in equation 2-8.

2.4 Results

Phenotypic variation observed in *P.nigra* indicates that leaf area and leaf width to length ratio are correlated with latitudes of population origin, but the significance of this relationship

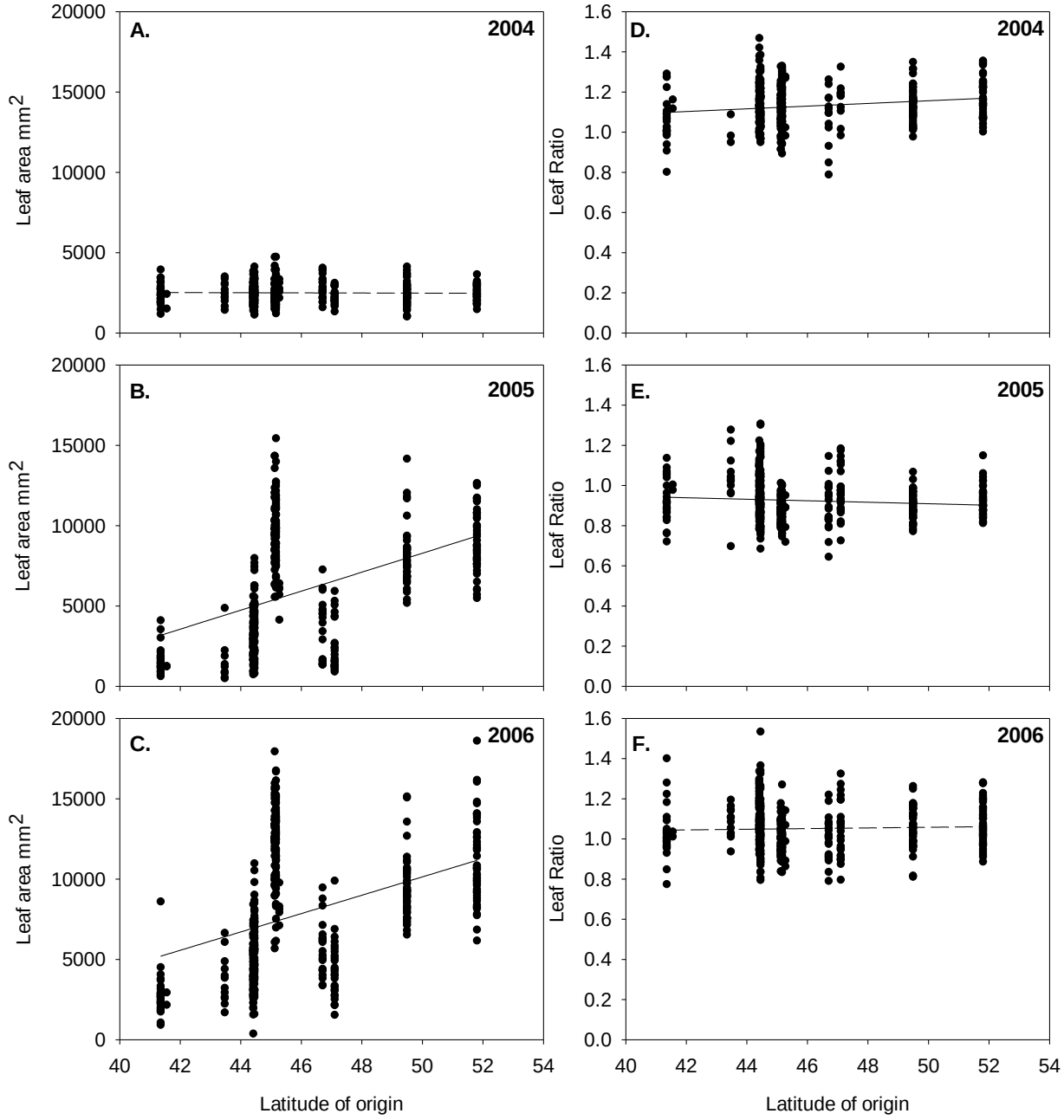


Figure 2-1: The relationship between leaf characteristics and latitude of origin in *P.nigra*.The relationship between leaf area and latitude of origin in the *P.nigra* association population in 2004 (A), 2005 (B) and 2006 (C) and the relationship between leaf ratio and latitude of origin in 2004 (D), 2005 (E) and 2006 (F). Solid lines represent significant regressions and dash lines represent non-significant regressions. Correlation coefficients together with their significance for the regression are A. r=0.02ns, B. r=0.46***, C. r=0.41***, D. r=0.19**, E. r=0.10*, F.r=0.04ns. Significance level is given by; ns non-significant, *P<0.05, **<0.01 and ***P<0.001.

varies with collection year (Figure 2-1).

These results indicate that leaf area in the second year (2005, r=0.46***) and third year of growth (2006, r=0.41***) increased significantly with latitude of origin, whereas no relationship was observed the first year of growth (2004, $r=0.02^{ns}$). Interestingly, inconsistent observations were observed in leaf ratio, an indicator of leaf shape, with significant correlations found in the first and second year of growth (2004: r=0.19** and 2005: r=0.10*) and not in the third (2006:$r=0.04^{ns}$, Figure 2-1D, E & F), suggesting that environment may have an overriding effect on leaf shape. Leaf ratio increases with latitude in the first year of growth indicating that leaf shape in higher latitudes are longer in length than in width (Figure 2-1D). However in the second and third year of growth, leaf shape decreases with latitude, indicating that leaves have greater width than length at higher latitudes. Together the differences between first year and subsequent years of growth demonstrate that phenotype can vary as trees become established.

A significant relationship was observed between cell number per leaf in the first and third year of growth, with cell number per leaf increasing with latitude of origin (Figure 2-2G & H). Average cell area showed a significant relationship to latitude in the first year of growth (r=0.27***) (Figure 2-2A), demonstrating a decrease with latitude, however in the third year of growth the relationship is not significant ($r=0.08^{ns}$) (Figure 2-2B). Together, these data suggest that, in lower latitudes, leaves were smaller with larger cell sizes, whereas in higher latitudes leaves were bigger with smaller cells determining size.

Stomatal traits such as stomatal density and stomatal index illustrate a similar pattern; however this is not consistent over growth years, suggesting again that environment had a large effect. The pattern shows that stomatal density and stomatal index decrease with latitude of origin in the first year of growth (Figure 2-2C & E) and increase with latitude of origin in the third year of growth (Figure 2-2D & F). However, some variance maybe explained by sampling procedures, as leaf age five was measured in 2004 but mature leaves were measured in 2006. Interestingly, in both the first and second year of growth, stomatal density and stomatal index showed significant correlation with latitude of origin.

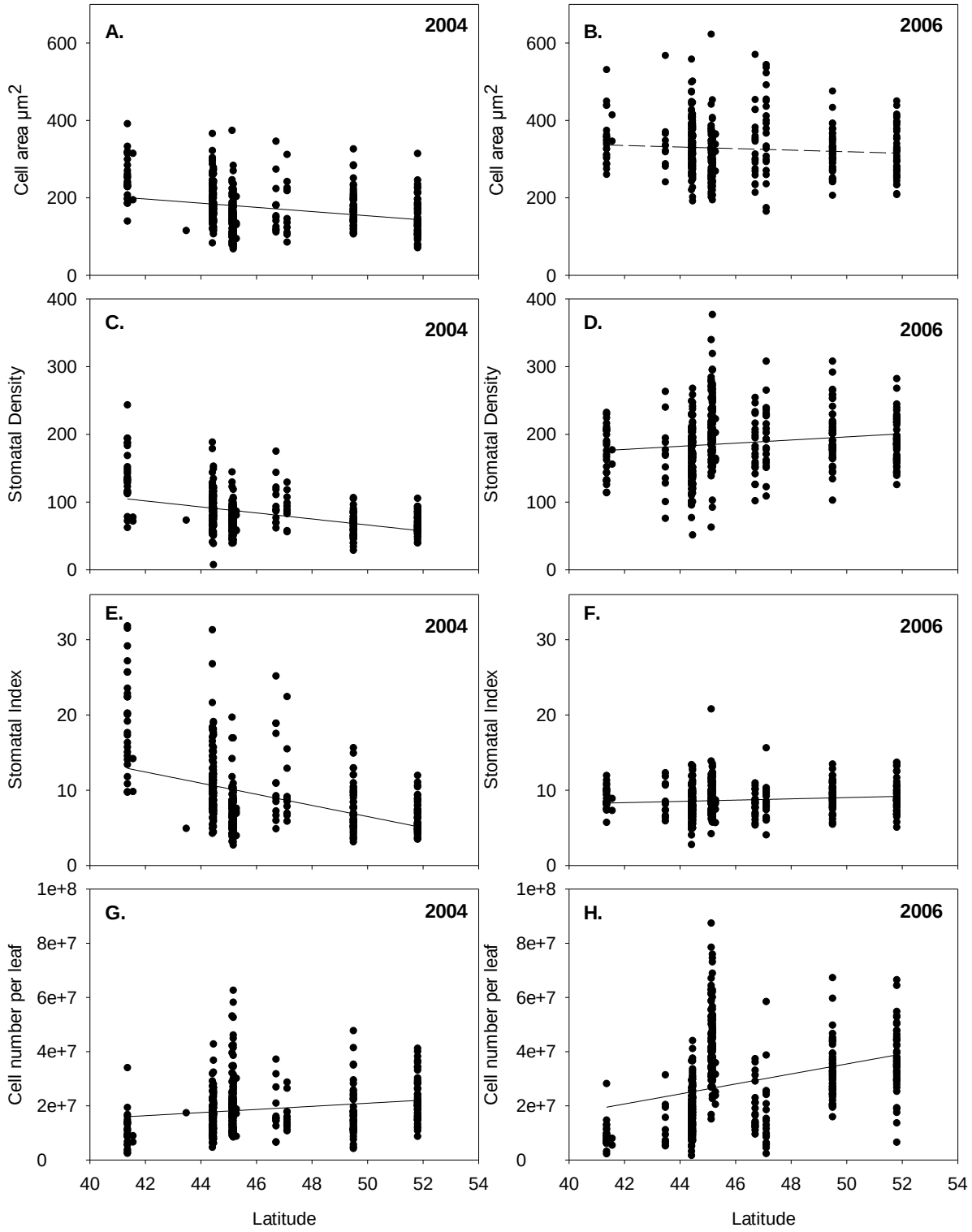


Figure 2-2: The relationship between cell traits and latitude of origin in *P.nigra*. The relationship between cell traits; average cell area (A &B), stomatal density (C & D), stomatal index (E & F) and cell number per leaf (G& H) and latitude of origin in the *P.nigra* association population after one year (A, C, E & G)and three (B, D, F & H) years of growth. Solid lines represent significant regressions and dash lines represent non-significant regressions. Correlation coefficients together with their significance for the regression are A. r=0.27***, B. r=0.08^{ns}, C. r=0.43***, D. r=0.14*, E. r=0.43***, F.r=0.11*, G. r=0.18** and H. r=0.31***. Significance level is given by; ns non-significant, *P<0.05, **<0.01 and ***P<0.001.

As with leaf and cell phenotypes, biomass traits were correlated with latitude of origin, though significance once again varied with year of growth. Height measurements showed a significant correlation in the second year of growth, whereby height increased with increasing latitude of

origin (Figure 2-3B & C). No significant correlation was seen in the first year of growth ($r=0.05^{ns}$) (Figure 2-3A).

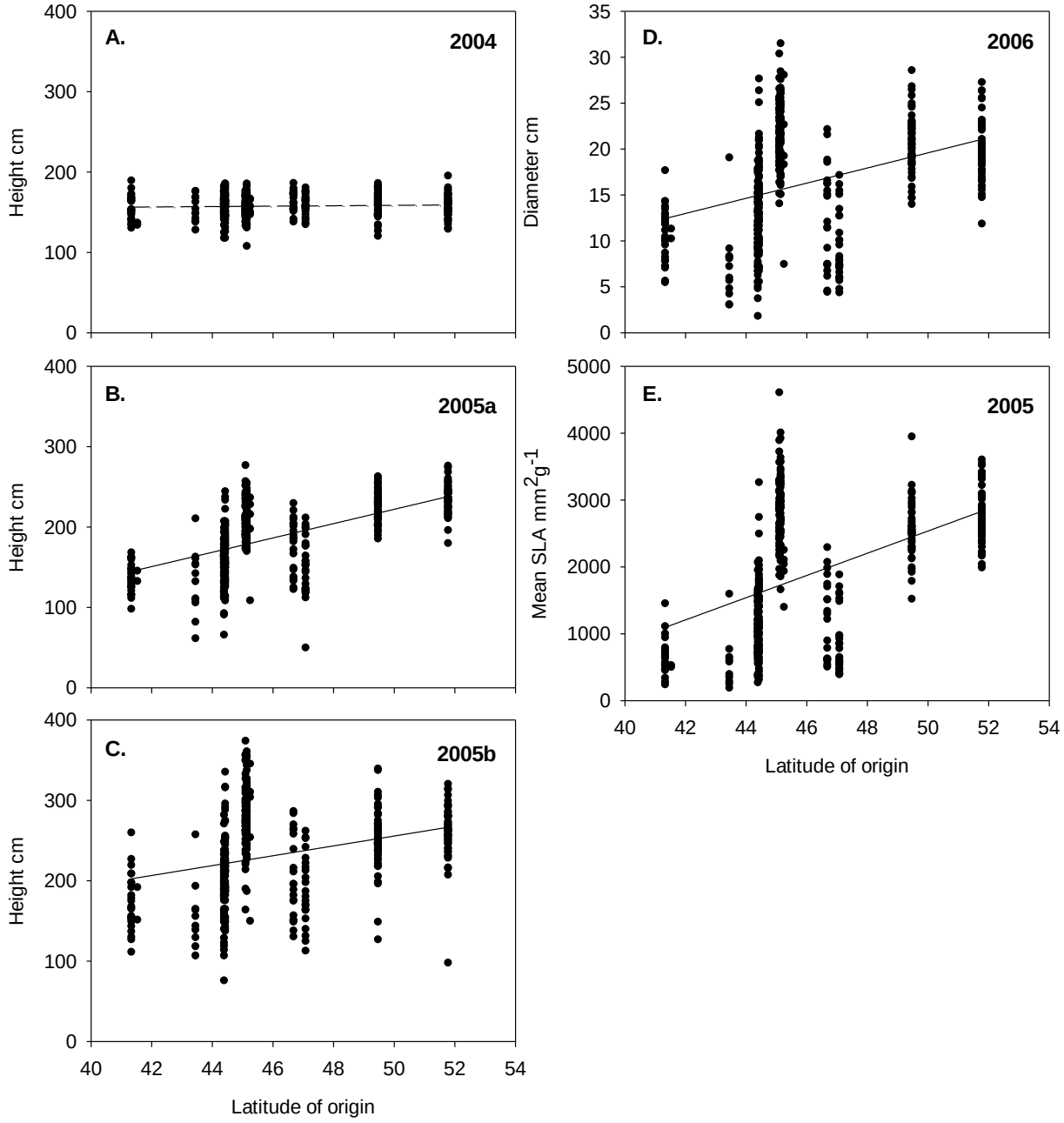


Figure 2-3: The relationship between biomass traits and latitude of origin in *P.nigra*.The relationship between biomass traits; height (A, B & C), diameter (D) and specific leaf area (SLA) (E) and latitude of origin in the *P.nigra* association population in 2004(A), 2005a (August), 2005b (December) (B, C & E) and 2006 (D). Solid lines represent significant regressions and dash lines represent non-significant regressions. Correlation coefficients together with their significance for the regression are A. $r=0.05^{ns}$, B. r=0.59***, C. r=0.31***, D. r=0.38*** and E. r=0.47***. Significance level is given by; ns non-significant, *P<0.05, **<0.01 and ***P<0.001.

Diameter showed a significant correlation with latitude of origin (r=0.38***) (Figure 2-3D) such that trees in higher latitudes of origin are larger, as diameter and height increase. SLA showed a significant relationship to latitude, where SLA increased with latitude of origin (r=0.47***) (Figure 2-3E).

Diameter measurements were taken solely in the final year of growth, therefore stem volume index was estimated only in the final growth year using height measured in December 2005 and diameter measured in 2006. Height was used as a indicator for biomass for 2004 and 2005 as height strongly correlates to biomass (Scarascia-Mugnozza *et al*., 1999;Rae *et al*., 2004). Leaf area has been suggested as a strong indicator of biomass, therefore it is interesting to see that leaf area strongly correlates to height in 2005 (Figure 2-4B & C) and stem volume index in 2006 (Figure 2-4D). The first year of growth resulted in no strong correlation to height for leaf area (Figure 2-4A), but a significant block effect and population difference (Table 2-2), indicating spatial differences in this year of growth.

Statistical analysis using general linear ANOVA showed similar results when population is considered in place of latitude, with significant differences between population detected for; leaf area, leaf ratio, cell area, stomatal density, stomatal index, cell number per leaf, height, diameter, specific leaf area and stem volume (Table 2-2). However when block and genotype within population is added to the model several traits are no longer significant such as leaf area 04, leaf ratio 04 & 06, height 04, cell area 04, stomatal density 04 & 06, stomatal index 04 & 06 and cell number per leaf 04 (Table 2-2), this could indicate that other influences are controlling these traits such as environment.

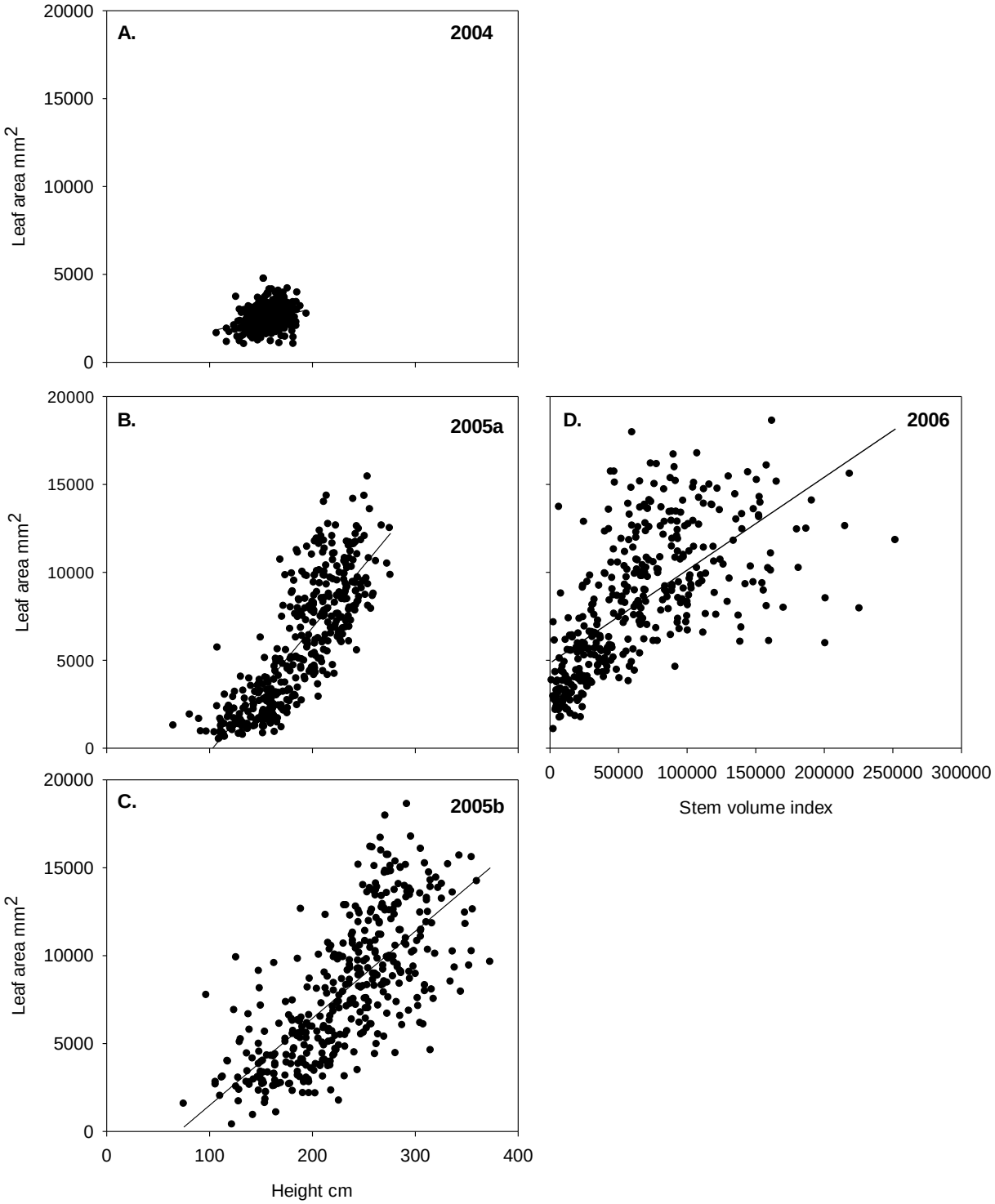


Figure 2-4: The relationship between biomass traits and leaf characteristics in *P.nigra*.The relationship between leaf area and height in the *P.nigra* association population in 2004 (A), 2005 (August -B & December -C) and the relationship between leaf area and stem volume in the *P.nigra* association population in 2006 (D). Solid lines represent significant regressions and dash lines represent non-significant regressions. Correlation coefficients together with their significance for the regression are A. r=0.28***, B. r=0.82***, C. r=0.71*** and D. r=0.64***. Significance level is given by; ns non-significant, *P<0.05, **<0.01 and ***P<0.001.

Table 2-2: Summary of phenotypic statistical analysis. Summary of general linear model ANOVA comparing variation between means of population, genotype and block for each quantitative trait. Populations were identified based on the median value of the original latitudinal range given by EUFROGEN and INRA. Values of heritability for all traits scored with replication.

Quantitative Trait	**Population**	**Clone (Population)**	**Block**	**Heritability**
Leaf traits				
Leaf area mm^2 04	ns	ns	<0.001***	0.5
leaf area mm^2 05	<0.001***	<0.001***	<0.05*	0.8
leaf area mm^2 06	<0.001***	<0.001***	<0.001***	0.7
leaf ratio 04	<0.001***	<0.001***	<0.001***	0.8
leaf ratio 05	<0.001***	<0.001***	ns	0.8
leaf ratio 06	<0.001***	<0.001***	ns	0.9
Cell traits				
abaxial cell area 04 (µm)	<0.001***	<0.001***	<0.05*	0.6
abaxial cell area 06 (µm)	<0.01**	<0.001***	<0.001***	0.7
abaxial stomatal density 04	<0.001***	<0.001***	ns	0.6
abaxial stomatal density 06	<0.001***	<0.001***	ns	0.6
abaxial stomatal index 04	<0.001***	<0.001***	ns	0.6
abaxial stomatal index 06	<0.001***	ns	ns	0.5
cell number per leaf 04	<0.001***	ns	<0.001***	0.5
cell number per leaf 06	<0.001***	<0.001***	<0.05*	0.7
Biomass traits				
height 04	<0.01**	ns	<0.001***	0.5
height 05 (February)	<0.001***	<0.001***	<0.001***	0.8
height 05 (October)	<0.001***	<0.001***	<0.001***	0.8
diameter 06	<0.001***	<0.001***	<0.001***	0.8
specific leaf area 05	<0.001***	<0.001***	<0.01**	0.8
stem volume index 06	<0.001***	<0.001***	<0.001***	0.7

Significance level is given by; ns non-significant, *P<0.05, **<0.01 and ***P<0.001.

2.5 Discussion

This study has examined genotypic variation in leaf characteristics with respect to latitude of origin. It is clear to see that there is a significant cline pattern of latitudinal variation in leaf traits; leaf area (Figure 2-1) and SLA (Figure 2-3), cell traits; cell number per leaf (Figure 2-2) and biomass traits; height and diameter (Figure 2-3). Height and diameter increased with increasing latitude, it is not surprising these traits showed similar patterns to cline variation as they are likely indices of similar biological processes. Results indicate that trees are larger in higher (northern) latitudes compared to lower latitudes (Southern). Strong correlations and moderate to high heritability estimates indicate that these traits have adapted on a complex environmental gradient, along which temperature, solar radiation, precipitation and soil conditions vary. At the climatic level investigations have concluded that low latitudes tend to have higher temperature, and longer frost free periods which speed up metabolic activity, cell growth, photosynthesis and hence the growth of the whole plant (Li *et al*., 1998). Other studies have looked at effects of drought and high temperature which resulted in a decrease in plant growth, leaf green area and leaf water potential (Xu and Zhou, 2006). This was found within lower latitudinal genotypes with small leaf areas, SLA, height and diameter, indicating less growth, although growth rates were not calculated and no conclusive comment can be made on growth. Variation in leaf area and SLA across latitude indicates that differences in photosynthetic capacity are present, which are strongly associated with nitrogen (N) (Xu and Zhou, 2006). It is interesting to note that, in general, leaf N and Phosphorous decline towards the equator as average temperature and growing season length increase (Reich and Oleksyn, 2004). This suggests that photosynthetic capacity may increase with latitude of origin due to an increase in N content (Xu and Zhou, 2006). N stress experiments have shown that cell number was reduced by 30% in expanding leaves of sugarbeet, and that mesophyll cell size is reduced (Trapani *et al*., 1999). Latitudinal differences in N, temperature and precipitation are likely to influence leaf area and SLA which are interesting to breeding programmes. A point must also be made on soil temperature as this also plays a major role in determining the growth rate of plants. Low soil temperatures can decrease root growth thereby affecting water and nutrient uptake. Decreases in water uptake can cause reductions in photosynthesis by inducing partial stomatal closure (Farquhar, 1989) and this is also seen in drought conditions. An experiment on silver birch showed that low soil temperatures reduced photosynthesis due to N starvation and reductions in leaf area (Aphalo *et al*., 2006) therefore lower latitudinal

genotypes may have been affected by the soil conditions in Belgium which are cooler, affecting root growth and nutrient uptake which leads to smaller leaves.

Further investigation into the morphology of the cells revealed that smaller leaves contained fewer cells per leaf that are larger, whereas larger leaves have more cells per leaf that are smaller. This could be due to differences in mechanisms controlling the cell production and cell expansion. For plant cells to expand the cell wall has to be capable of such action and the cellulose-hemicellulose network plays a leading role in determining this. Enzymes act on the network to control the process of cell growth (Li *et al*., 2003). In the 1990s two groups of proteins were identified acting as wall loosening enzymes:- endoxyloglucan transferase (XETs) and expansins (Li *et al*., 2003). Previous studies have shown that expansins play a role in leaf development. For example, a study by (Fleming *et al*., 1997) involving topical application of expansin to the flanks of a tomato's vegetative meristem, led to the initiation of leaf primordia. However the primordia did not grow to form a normal leaf, indicating that other factors control final leaf size. This developmental complexity is also indicated by mutational studies and knockout studies of expansins and XETs where a pronounced phenotype cannot be seen (Li *et al*., 2003). Expressional studies of XETs paint a different picture, where in *Arabidopsis* mutants with reduced internodal cell length in young leaves named *acaulis (acl*), expression of certain XETs such as *EXGT-A1* is reduced (Akamatsu *et al*., 1999) and XET can control cell size and therefore determine leaf size.

Cell production on the other hand is driven by cell division within the mitotic cell cycle. Previous studies have identified a number of cell cycle genes. Over-expression of a dominant –negative mutant from *Arabidopsis cdka* protein, which does not have kinase activity, caused a reduced cell number in tobacco plants (Hemerly *et al*., 1995). However cell cycle mutational studies have not been very successful, showing very few changes in phenotype. This is due to the large amount of compensatory mechanism which occurs within the cell cycle, making it very difficult to pin down the process to an individual gene determining leaf size.

Stomatal index and stomatal density results were inconsistent over the two years of study of cell traits. In the first year of growth stomatal index and density decreased with latitude of origin while in the second year of growth stomatal density and index was consistent across latitude of origin (Figure 2-2C, D, E & F). This could indicate a local environmental effect on stomata; however heritability estimates indicate that stomatal traits are genetically controlled. Stomatal differentiation occurs firstly by the asymmetric division of a protodermal cell, called

a meristemoid mother cell (MMC). MMC then produce a small triangular meristemoid, which differentiates into guard mother cells (GMC) or another meristemoid. The majority of stomata on mature leaves in *Arabidopsis* are formed from satellite meristemoids (Geisler and Sack, 2002). If clinal variation affects any of these processes, stomatal development would be affected. The GMC is also important in stomatal spacing, as guard cells are never found next to each other. Stomata have also been found to have higher densities at the margins and tips of leaves (Smith and McClean, 1989). There are three hypotheses for these differences; differentiation hypothesis, expansion hypothesis and mixed differentiation and expansion hypothesis (Poole *et al*., 1996); all hypotheses suggest that epidermal cell size dictates stomatal density.

Recently genes involved in stomatal spacing have been identified. One such is *MAPKKK (YODA*), which has been identified as a key intermediary in stomata formation, with repression of *YODA* leading to ectopic stomata formation and over expression of *YODA* leading to leaves lacking stomata (Larkin *et al*., 2003). *TOO MANY MOUTHS* (TMM) leads to the formation of groups of stomata, encodes an LRR-kinase, suggesting that the protein acts as a receptor for some signal which has not yet been identified (Larkin *et al*., 2003). *STOMATAL DENSITY AND DISTRIBUTION 1* (*SDD1*) controls stomatal density and distribution (Larkin *et al*., 2003).

Previous studies by Woodward *et al*. (2002) showed that stomatal density affected gas exchange, stomatal conductance and therefore water use efficiency. These studies have shown that stomatal density decreases when plants are exposed to elevated CO_2 concentrations and historical data from *Metasequoia glytostroboides* and *Ginkgo biloba* have shown that stomatal index reduces by 50% and 30% as concentrations of CO_2 in the atmosphere rise (Beerling and Royer, 2002), indicating a strong link to environment. Further evidence of stomatal regulation by local climatic adaptation was seen by Yin *et al*. (2004a): they found that two *Populus* species, *P. kangdingenesis* and *P. cathayana*, found at different altitudes, had interspecific differences in ABA-induced growth, with drought stress in higher altitudes causing reductions in leaf growth. This is interesting as ABA plays an important role in stomatal regulation. Other experiments have shown that stomatal development is influenced by mature leaves (Lake *et al*., 2002), where mature leaves detect environmental changes and relay the information to new developing leaves, causing an increase or decrease in number of stomata (Lake *et al*., 2002). Therefore changes in stomatal density and index due to environmental causes will be seen most in younger leaves.

Fate of stomata is dependent on signals received from the environment and from neighboring cells relatively late in development (Holroyd *et al*., 2002). Other theories suggest stomatal development is controlled largely by direct signals to pavement cells through the cell wall, by proteins such as wall associated kinases (Anderson *et al*., 2001) and in wax biosynthesis pathways (Gray *et al*., 2000). Therefore differences in genes controlling wax biosynthesis such as *CER1* may play a part in stomatal density and index. This evidence suggests that both stomatal density and stomatal conductance would be good parameters to determine genetic adaptation in these *P.nigra* genotypes.

Strong positive correlations between leaf area and stem volume index (Figure 2-4) indicate that leaf area is a robust indictor of biomass in *P.nigra*. This has been suggested in previous studies in poplar (Pellis *et al*., 2004;Rae *et al*., 2004;Marron and Ceulemans, 2006), however some studies suggest leaf area index to be a better indicator as a high number of small leaves have been shown in *P.nigra* to increase productivity (Laureysens *et al*., 2005). Mature leaves produce photosynthates for the production of woody biomass, therefore the larger their leaf size the greater the photosynthetic capacity and potential for larger growth. If trees are to be selected for biomass crops from the association population, the higher latitudinal trees should be chosen due to large leaf area and greater number of cells per leaf, indicating more light interception leading to faster growth.

2.5.1 Summary

Considerable variation in *P.nigra* architecture has been shown in leaf traits; (leaf size and SLA), cell traits; (cell number per leaf) and biomass traits; (height and diameter). This variation correlates with latitude of origin, indicating potential adaptation by *P.nigra* to an environmental gradient in solar radiation, precipitation, temperature and day lengths. Leaf area shows a strong correlation to height and stem volume. Stem volume is an estimate of yield therefore leaf area is a robust indictor of yield in natural populations. Moderate to high heritability scores of all quantitative traits suggest a strong genetic control which is essential for downstream analysis using association study approaches.

3 . Utilising parallel genetic approaches; QTL, microarray and bioinformatics to identify candidate genes involved in leaf size.

3.1 Overview

At the beginning of the 19th century discussions began between Mendelian geneticists and biometricians over the mechanisms involved in evolutionary processes. On one side the Mendelian view was that evolution was driven by variation in discrete characters through the appearance of mutations with large effect, whereas biometricians believed evolution was a result of natural selection acting upon continuously distributed characters. With this knowledge quantitative genetics progressed with Ronald Fisher (1918) and Sewall Wright (1921) at the forefront, stating that phenotypic expression is determined by shared genes and their environment.

Previous studies to identify genes involved in leaf morphology have used microarray analysis (Taylor *et al*., 2005) and QTL (Quantitative Trait Loci) analysis separately (Rae *et al*., 2006). However with the release for the *Populus trichocarpa* genome in 2005 combining multiple disciplines such as microarray and QTL is possible to detect candidate genes for leaf development. Therefore in this chapter I attempt to combine previous QTL analysis carried out on the pedigree family 331and microarray results obtained from *P.nigra* (a natural population), to co-locate genes to the family 331 genetic map and to identify hotspot areas for leaf development and morphology.

3.2 Quantitative Genetics

3.2.1 Introduction

Previous studies to understand genes that determine a phenotypic trait consisted of discovering a 'major gene' that had a large enough effect to be recognised; these genes arose from either spontaneous or induced mutation (Falconer & Mackay, 1996). However most of the natural variation observed in biology is due to minor genetic changes in many genes. QTL is an abbreviation for Quantitative Trait Loci, which are genes that underlie quantitative traits, whereas QTL analysis is the phrase used to study genetic variation and to locate genes responsible and explore their effects and interactions (Kearsey, 1998). A QTL can be defined as showing continuous variation in a population which is more or less normally distributed, has a large effect on phenotype compared to the environment, genotypes have recognizably different phenotypes and often one allele is non-functional or is very dysfunctional which results in a clear phenotype (Kearsey, 1998). In most cases underlying quantitative variation is due to allelic differences that occur in structural or regulatory genes producing small phenotypic effects. For QTL discovery it is essential to have both genotypic and phenotypic information, genetic information includes a genetic map consisting of genetic markers dispersed over the organisms genome. These molecular markers are typed within a mapping population of individuals, from which the phenotype data has also been collected. QTL analysis works on the principle that if a marker is in close physical linkage with a QTL the two will be in linkage disequilibrium within the mapping population, which generates a statistically significant association between the marker genotype and the trait variation (Slate, 2005). Therefore to carry out QTL analysis it is essential to have a genetic map of variable markers, a pedigree to follow the segregation of markers and phenotypic data.

3.2.2 Molecular Markers

The discovery of molecular markers has enabled molecular variation to be scored and this has been the major breakthrough in QTL analysis. Molecular markers give unambiguous single site genetic differences that can be scored and mapped (Kearsey, 1998). DNA-based markers satisfy all criteria for QTL analysis: they have to be highly polymorphic and abundant (Falconer & Mackay, 1996). Suitable genetic markers include variable number of tandem repeat (VNTR) or minisatellite loci, microsatellite (or simple sequence repeat, SSR) loci, SNPs (Single Nucleotide Polymorphisms) and AFLPs (Amplified fragment length polymorphisms).

The suitability of a marker does depend on the type of mapping population. In inbred line crosses an ideal marker should be biallelic and show fixed differences between the parental lines (Slate, 2005). Therefore AFLPs and SNPs are most frequently used in inbred line crosses. However both markers come with advantages and disadvantages; for example AFLPs can be generated in any organism producing many genotypes rapidly and cheaply, but they are usually unique to a mapping population and therefore comparative studies are difficult. SNPs on the other hand are abundant, found in both coding and non-coding regions, co-dominant, can be targeted to particular genes or regions in the genome and are conserved between mapping populations. Their disadvantages however are more financial; they are expensive compared to AFLPs and they need genome sequence data which is a costly process if the information is not already available. In outbred mapping populations microsatallites are usually used. This is due to parents in out breeding populations are not always homozygous at a marker. Therefore multiallelic markers are more suitable such as microsatallites. Once enough markers are collected across the whole genome, a linkage map can be created with the requirement of a pedigree population. The optimal number of markers depends on the organisms genome size, degree of linkage disequilibrium and recombination rate. There are several software packages available for this such as; MAPMAKER (Lander *et al*., 1987) and CRIMAP (Barker *et al*., 1987). Genetic maps show the order of loci along a chromosome and the relative distance between them created with recombination frequencies. Map distances are reported in centiMorgans (cM), a cM can correspond to a span between ten thousand to a million nucleotide base pairs (Lynch & Walsh, 1998). Several genetic maps have been already created in poplar; such as the family 13 consisting of 92 microsatellites and 24 AFLPs (Wullschleger *et al*., 2002), family 331 consisting of 343 RFLPs, STS and RAPD markers (Bradshaw *et al*., 1994;Frewen *et al*., 2000),a genetic map consisting of a back-cross between *P.deltoides* clone I-69 and *P.euramericana* clone I-45 (Yin *et al*., 2002) and a *P.nigra* map derived from an intraspecific cross between two *P.nigra* selected from natural Italian populations (Gaudet *et al*., 2007).

3.2.3 Mapping population

To identify and map QTL, a cross between lines that differ for a trait of interest is required (Falconer & Mackay, 1996). Inbred line crosses works on the assumption that all of the F_1 generation our genetically identical and in complete linkage disequilibrium for genes differing between lines. This is because the F_1 generation is created by crossing two inbred parental lines, therefore creating linkage disequilibrium between loci that differ between the lines, and

this in turn creates associations between maker loci and linked segregating QTLs. The F_1 is then used to construct the mapping population by two different designs; (i) F_2 design were F_1 individuals are inbred or (ii) a backcross design were F_1 individuals are mated to one of the parental populations. The F_2 design is prefered as it generates three genotypes at each marker locus, therefore enabling an estimation of the degree of dominance associated with the detected QTLs. Outbred populations can be mapped either using sibships (half-sibs or full-sibs) or using general pedigrees spanning several generations.

3.2.4 Phenotypic Data

All around us we see examples of traits that are quantitative. In local woodlands trees are not uniform in size and shape, at school or in the office people have varying heights, skin colour and weights. This is due to these traits being controlled by two or more genes and their environment. If you were to collect data on the heights of each individual in your department at University and plot as a histogram, the data would be normally distributed creating a bell shape curve. Therefore the next step in QTL analysis is to measure the trait of interest from each of the genotypes and test for normality. This is usually achieved using the Anderson-Darling test for normality, indicating departures from a normal distribution (bell-shaped curve) of data sets. However not all QTL mapping packages require normality.

3.2.5 QTL analysis

Until now we have discussed the raw material we need for QTL analysis. Here I would like to discuss the processes which links together these data sources to identify region(s) of the genome involved in the genetic control of the trait. The success of QTL analysis depends on; marker density and the size of the mapping population. Types of QTL analysis include; regression analysis, interval mapping, composite interval mapping (CIM) and multiple interval mapping (MIM).

3.2.6 Regression analysis

This is a basic single gene QTL model, which can be run by most statistical software packages. It uses a General Linear Model (GLM) to regress the phenotypic value onto the marker genotype. A strong regression is present when the marker is linked to a locus controlling a phenotypic trait. The strength of this regression is weakened with genetic distance and when markers lie on another chromosome. This technique is preferred in studies where the goal is to simply detect QTL linked to a marker.

3.2.7 Interval Mapping

Similar to regression analysis except that the regression is carried out across a marker interval, QTL are localized between two genetic markers (flanking markers). This model calculates the likelihood of odds (LOD) score, which is $\log_{10}$ of the likelihood of the odds ratio. LOD is the probability of a QTL being located within a marker interval measured against the probability of the result occurring by chance. An alternative has been suggested by (Haley and Knott, 1992) which is based on least-squares (LS) multiple regression which has widely available user friendly and software is found on the web.

3.2.8 Composite Interval Mapping (CIM)

CIM is an extension of interval mapping; however non-linked markers are included as cofactors in a multiple regression analysis. This reduces the influence of multiple or linked QTL, therefore reducing bias estimates. Also available in this model is the incorporation of multiple traits or QTL by trait interactions. Further consideration should be taken when selected markers as cofactors as there is potential for selection bias.

3.2.9 Multiple Interval Mapping (MIM)

MIM considers all the linked markers on a chromosome simultaneously therefore performing a single analysis for each chromosome. Similar to CIM, includes additional markers as cofactors in a multiple regression analysis. Many QTLs have been mapped to improve agriculturally important crop and animal species, such as yield traits in crosses between elite inbred maize strains, growth and fatness in pigs and milk production in cattle (Mackay, 2001). In poplar many QTL have been mapped for biomass traits (Wullschleger *et al*., 2005;Rae *et al*., 2007), senescence (Rae *et al*., 2006), rust resistance (Yin *et al*., 2004b), osmotic potential; (Tschaplinski *et al*., 2006) and bud set and bud flush (Frewen *et al*., 2000). Therefore the next step is to move from QTL analysis and look within QTL regions to find genes controlling these important traits and within the poplar community this is possible due to the whole genome sequence of *P.trichocarpa*.

3.3 Microarray Analysis

3.3.1 Introduction

Jansen and Nap, (2001) proposed genetical genomics, an approach that blends QTL mapping and microarray analysis to identify associations between the allelic state of a genomic region

and a gene's transcript abundance (Wang and Nettleton, 2006). Success has already been achieved by using this approach to reduce numbers of candidate genes. An example of this is an experiment by Yagil and Yagil, (2006) who combine QTL analysis resulting in 1102 genes and microarray analysis resulting in 2470 transcripts to find 7 novel candidate genes for hypertension (Yagil and Yagil, 2006). Microarrays have enabled scientists to analyse the expression of hundred and thousands of genes in a single experiment quickly and efficiently. Microarrays work by exploiting the ability of a given mRNA molecule to bind specifically to, or hybridize to, the DNA template from which it originated. In a single experiment the expression levels of many genes can be measured with the aid of computers measuring the amount of mRNA bound to each spot on the array, generating a profile of gene expression. Microarrays were first described by Schena *et al*., (1995). Using microarrays it is possible to develop a complete overview of all the genes in a genome that are up-regulated or down-regulated in response to some factor of interest (Roelofs *et al*., 2008). The use of microarrays includes diagnostics where transcription profiling can provide insights into the functional performance of an already well-known genome and explorative uses where the array is used to discover transcripts that respond to some factor, usually in organisms where the whole genome has not been sequenced. An explorative approach was undertaken in this study to identify transcripts up-regulated or down-regulated in large or small leaves within the association population (2.3.1).

3.3.2 Microarray technology

In a cDNA microarray many probes are used at once, as they are produced robotically by spotting multiple probes representing specific genes or ESTs (Expressed Sequence Tags). Within the *Populus* community several microarrays have been developed; 25,000 element cDNA array developed in Sweden (Sterky *et al*., 2004), a 27,000 element PICME cDNA array developed in France (Dejardin *et al*., 2004) and a 15,400 element Treenomix cDNA array developed in Canada (Jansson and Douglas, 2007). Sequencing of the whole poplar genome has seen the development of the Affymetrix Gene Chip ® Polar Genome array which interrogates 56, 000 transcripts. To carry out microarray analysis, firstly mRNA is isolated from two samples; one RNA sample is labeled with a green fluorescent dye (Cyanine 3-dNTP, Cy3) and the other is labeled with a red fluorescent dye (Cyanine 5-dNTP, Cy5), this is then used to generate cDNA with a fluorescent tag attached. The tags are used to differentiate between the samples in subsequent steps. The two samples are then mixed and incubated with a microarray containing all genes spotted onto a solid support such as a glass microscope slide.

Labeled molecules will then bind to the sites on the array corresponding to the genes expressed in each sample, this process is called hybridization. The microarray is then placed into a scanner, which with the aid of a specific laser, camera and microscope produces a digital image of the microarray slide. Computer programs are then used to calculate the different concentrations between the two samples on a scale from green (Cy3 >Cy5) via yellow (Cy3 = Cy5) and red (Cy 5 > Cy 3). Colour differences indicate up or down regulation of the particular gene compared to another (Ouborg and Vriezen, 2007) (Figure 3-1) Microarrays have been successfully used within the *Populus* community to identify genes differentially expressed in contrasting environments (Moreau *et al*., 2005;Taylor *et al*., 2005;Street *et al*., 2006). A study by Andersson *et al*. (2004) has enabled insight into autumn senescence showing shifts in gene expression coinciding with chlorophyll degradation.

Within this study we hoped to identify gene expression differences between large and small leaf area extremes (Figure 3-1) to select out, with the aid of QTL analysis, candidate genes for the association study.

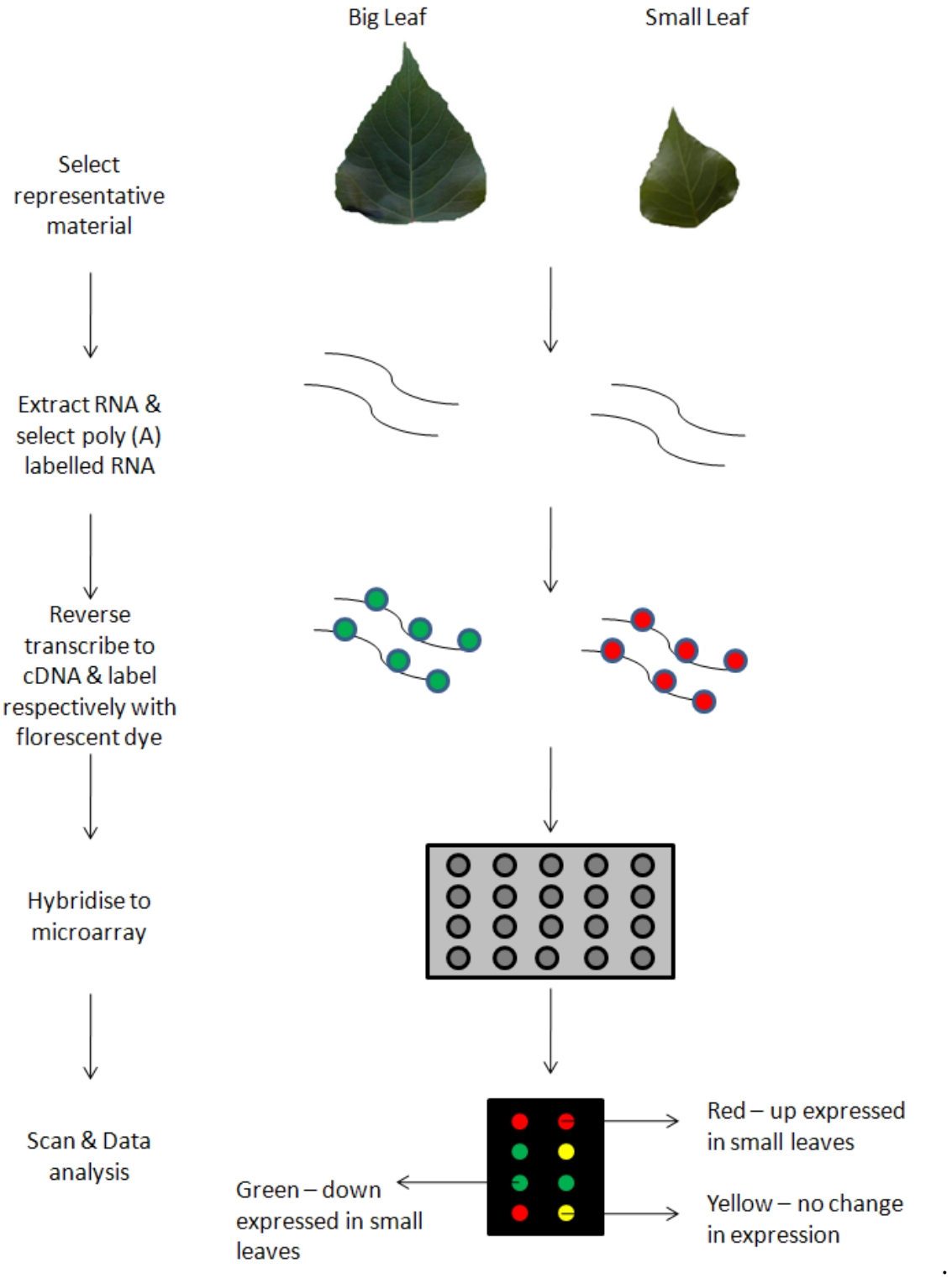


Figure 3-1.Schematic representation of transcript analysis using cDNA microarray technology. In this case plant material is selected from 'big' and 'small genotypes, RNA is then extracted and samples are labeled; one with a green fluorescent dye and the other with a red fluorescent dye. Samples are then hybridized to a cDNA microarray and then scanned and anlaysed. Colour differences indicate up (red), down (green) or no charge (yellow) in expression.

3.3.3 Combining QTL & Microarray techniques

Few studies have combined QTL and Microarray techniques to identify candidate genes controlling leaf development. QTL and microarray studies can pull out thousands of genes involved in your trait of interest, therefore combining techniques should enable clustering of genes into QTL regions implying they are strong candidates. Wang *et al*. (2007) combined the bioinformatics and QTL analysis to identify strong candidate genes for defence response homologs, which included a variety of signal transduction and biochemical reactions within higher plants for defence against pathogens. In this study they combined genes found within maize databases and EST databases and then mapped them to a linkage map created by a cross between two elite inbred maize (Zea mays L) lines "Zang" and "87-1" creating 294 Recombiant Inbred Lines (RILs). They found defence response genes not to be equally spaced but found in clusters; the highest number on chromosome 4 and the lowest found on chromsome 3. Also they found half of their defence response genes to locate within chromosomal regions whereby major genes or QTL had already been mapped. Shi *et al*., (2005) combined macroarray and QTL approaches to locate candidate genes involved in sugarcrane mosaic virus (SCMV) resistance in maize and found on chromosome arms 6S and 3L.

In this chapter I will utilize the wealth of QTL regions already identified in the Taylor lab in family 331, the whole genome sequence of *P.trichocarpa* and a previous microarray experiment comparing expression levels of different *P.nigra* to pin point strong candidate genes for leaf development and morphology.

3.4 Material & Methods

3.4.1 QTL analysis

3.4.2 Plant material

In this study a three-generation *Populus* mapping pedigree was generated by first crossing a female *P.trichocarpa* (Clone 93-968 from western Washington) and a male *P.deltoides* (Clone ILL-129 from central Illinois) in 1981. The resulting F_1 family (family 53) produced two siblings 53-246 (male) and 53-242 (female) which were crossed in 1988 to produce 90 genotypes and again in 1990 to produce 320 genotypes , creating the F_2 family named family 331 (Bradshaw and Stettler, 1993;Bradshaw *et al*., 1994).

3.4.3 Data collection

Family 331 has been extensively phenotyped by research workers within Professor Gail Taylors' lab, therefore a wealth of information was available for this comparative study. Here I will outline three previous studies within the Taylor lab used in this study; a drought, a short rotation coppice (SRC) and a CO_2 experiment

3.4.4 Drought Experiment

Two hundred and ten genotypes of family 331 were included in this study. Six replicate cuttings of each genotype were planted in April 2003 at two separate sites in a randomized six block design spacing 75 cm by two metres. One site was located in the UK at the Forestry Commission site, Headley, UK (51^o07'N, 0^o50'W) and the other in Cavallermaggiore, Italy (44^o21'N, 8^o17'E). Traits measured included; leaf growth, leaf characteristics and cell counts such as stomatal and trichome density. Full details of the experiment and traits measured can be found in Rodríguez-Acosta *et al*., (2008) in preparation (Table 3-1).

3.4.5 Short Rotation Coppice (SRC) Experiment

Three hundred genotypes of family 331 were used in this study. Three replicate cuttings of each genotype were planted in spring 2000 at the Forestry Commission site, Headley UK (51^o07'N, 0^o50'W), in a randomized block design spacing one by one metres (Rae *et al*., 2004). To initiate the first coppice cycle (CC1) single stem plants were cut back January 2001. CC1 was harvested in the winter of 2002, which initiated the second coppice cycle (CC2) which was harvested in winter 2005, after four years of growth. Traits were measured in 2001 and 2006 including; biomass traits and leaf traits. Full details of the experiment and traits

measured in 2006 and 2004 can be viewed in (Rae *et al*., 2004) and Rae *et al*., (2008) in preparation (Table 3-1)

3.4.6 CO_2 experiment

Two hundred and eighty-nine members of family 331 were used in this study. The experiment was conducted in 16 open top field chambers (OTC) at the forestry commission field site, Headley, UK (51°07'N, 0°50'W) (Rae *et al*., 2006). Eight of the chambers received ambient CO_2 (a[CO_2]) while the other eight received elevated levels of CO_2 (e[CO_2]) at a concentration of 600μmol mol^{-1} CO_2. One genotype was randomly placed in each of the eight chambers consisting of e[CO_2] and a[CO_2] and therefore each chamber had 36 genotypes. Traits measured included; leaf growth, leaf plasticity & elasticity, leaf senescence and petiole length. Measurements were conducted throughout the growing season. Full details of this study can be view in (Rae *et al*., 2006) (Table 3-1).

3.4.7 QTL analysis

All three experiments used the same genetic linkage map, produced at Qak Ridge National Laboratory consisting of 91 simple sequence repeats (SSRs) genotyped on 350 individuals and 92 fully informative amplified fragment length polymorphisms (AFLPs) genotyped on 165 individuals (Rae *et al*., 2006). Linkage groups were orientated by blasting SSR primers against the poplar genome sequence (Tuskan *et al*., 2006b) (Rae *et al*., 2006).Normal distribution of trait data was tested using the Anderson-Darling test, which tests departures from normality in data sets. In cases were data was not normally distributed the Box-Cox transformation was carried out (Rae *et al*., 2006), Rodríguez-Acosta *et al*., 2008 in preparation and Rae *et al*., 2008 in preparation). Data analysis for QTL used the linear regression approach put forward by (Haley *et al*., 1994). The freely available web-based software QTLExpress (Seaton *et al*., 2002) was used using the out-breeding module. This method determines the identity –by-descent (IBD) probabilities from multiple marker data, then fits statistical models to the observations and IBD coefficients (Rae *et al*., 2006). Chromosome-wide permutation tests with 1000 iterations determined p-values, a significance threshold of 0.05 was taken as evidence of a QTL (Churchill and Doerge, 1994). Confidence intervals (CIs) showing the position of a QTL were calculated from the critical F value using an F-two drop off (the cM distance taken from the peak F value to drop by two either side) (Rae *et al*., 2008 in preparation).

Table 3-1: Phenotypic traits measured in the Taylor lab between 2004 & 2008.Details of biomass, leaf growth , leaf morphology and cell morphological traits measured from family 331 (*Populus trichocarpa* x *Populus deltoides* pedigree), within three separate experiments within the Taylor lab ; Drought (Dro), Short Rotation Coppice (SRC) and CO_2 experiments indicated by "x". At the right of the table references are displayed with "[P]" indicating the journal is in preparation.

		Experiment			
	Trait	**Dro**	**SRC**	**CO_2**	**Reference**
Biomass	No. of sylleptic branches		x		Rae *et al.*, 2004,
	Maximum Stem height (m)		x		Rae *et al.*, 2004, Rae *et al.*, 2008[P]
	Stem extension increment (mm)		x		Rae *et al.*, 2004,
	Basal stem diameter (mm)		x		Rae *et al.*, 2004, Rae *et al.*, 2008[P]
	Stem diameter 1 metre above ground		x		Rae *et al.*, 2008 [P]
	No. of stems on stool		x		Rae *et al.*, 2004, Rae *et al.*, 2008[P]
	Whole-tree dry mass (ODT $ha^{-1}y^{-1}$)		x		Rae *et al.*, 2004, Rae *et al.*, 2008[P]
Leaf growth	Leaf production rate ($no.day^{-1}$)		x		Rae *et al.*, 2004
	No. of leaves on leading stem		x	x	Rae *et al.*, 2004, Rae *et al.*, 2006
	Leaf extension rate ($mm\ day^{-1}$)	x	x	x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2004, Rae *et al.*, 2006
	Leaf expansion ($mm\ day^{-1}$)			x	Rae *et al.*, 2006
Leaf morphology	Leaf elasticity (% reversible extension per 10g load)			x	Rae *et al.*, 2006
	Individual leaf area (mm^2)	x	x	x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2004, Rae *et al.*, 2006
	Specific Leaf area (mm^2g^{-1})	x	x	x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2004, Rae *et al.*, 2006
	Leaf senescence index (%)			x	Rae *et al.*, 2006
	Petiole Length (mm)	x	x	x	Rodríguez-Acosta *et al.*, [P] ,Rae *et al.*, 2004, Rae *et al.*, 2006
	Petiole Width (mm)	x			Rodríguez-Acosta *et al.*, 2008 [P]
	Leaf length (mm)	x		x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2006

Table 3-1: Continued...

	Leaf Width (mm)	x		x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2006
	Leaf width to length ratio	x		x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2006
Cell morphology	Abaxial stomatal density	x			Rodríguez-Acosta *et al.*, 2008 [P]
	Adaxial stomatal index	x		x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2006
	Abaxial stomatal index	x			Rodríguez-Acosta *et al.*, 2008 [P]
	Adaxial epidermal cell area (μm^2)	x	x		Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2004
	Abaxial epidermal cell area (μm^2)	x			Rodríguez-Acosta *et al.*, 2008 [P]
	Adaxial cell number	x		x	Rodríguez-Acosta *et al.*, 2008 [P], Rae *et al.*, 2006
	Abaxial cell number	x			Rodríguez-Acosta *et al.*, 2008 [P]
	Number of adaxial epidermal cells per leaf ($x10^7$)		x	x	Rae *et al.*, 2004
	Number of adaxial epidermal cells per leaf ($x10^7$)		x	x	Rae *et al.*, 2004

3.5 Microarray Analysis

3.5.1 Plant material

A natural population of *P.nigra* was used as plant material for microarray analysis. The trees were grown in an experiment conducted at the Popyomics facility in Belgium as described in 2.3.1.

Leaf area extremes were identified from leaf area measurements collected in August 2004 from the LD population (2.3.2). Average leaf area was calculated for each genotype, sorted in ascending order and the five lowest leaf area genotypes and the five highest leaf area genotypes were selected for microarray analysis (Figure 3-2) and named 'extremes'. Analysis of 2005 and 2006 mature leaf area showed that the selected extremes were not absolute, however a great deal of separation can still be seen between years (Figure 3-2).

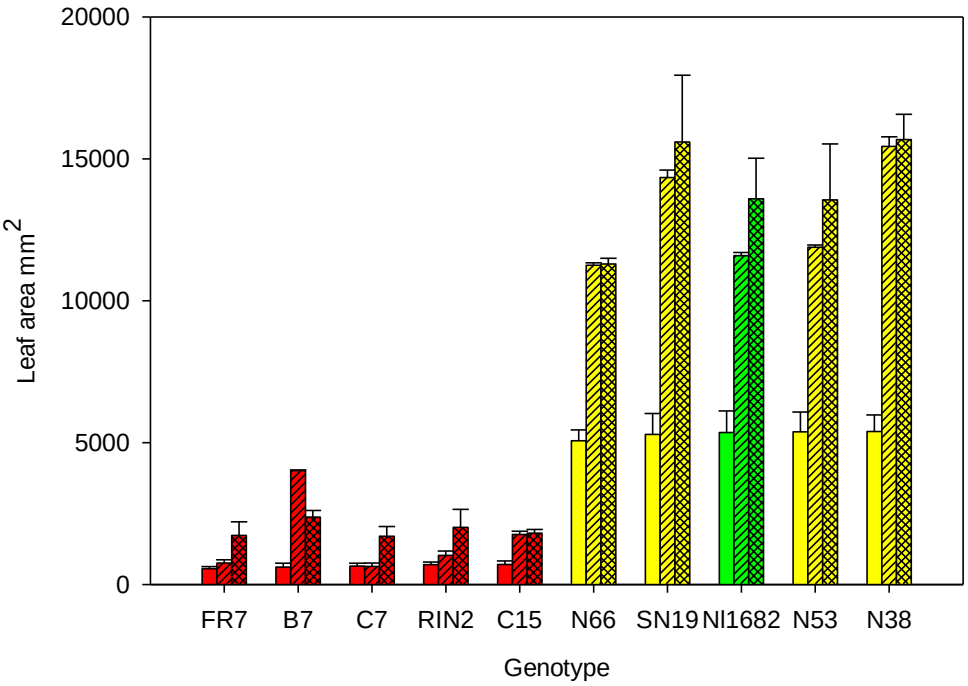


Figure 3-2: *P.nigra* leaf area extremes. Five "Big" and five "Small" leaf area extreme genotypes selected from 2004 phenotype data (chapter 2: Results) (solid colour). Average leaf area of these genotypes shown from data collected in 2005 (slash symbols) and 2006 (criss crossed symbols). Colours represent country of origin; red is Spain, yellow is Italy and green is the Netherlands.Data points are the mean leaf area for each genotype.

3.5.2 Leaf collection

Extremes were sampled from the LD population as described in 2.3.1. Leaf age was defined by counting down from the first fully unfurled leaf termed Ln-1. Ln-5 was sampled on the 22nd

August 2005 between 15:00 and 18:00 GMT. Leaves were immediately flash-frozen into liquid N and stored at -80^{o}C until RNA extraction.

3.5.3 Microarray design, preparation, hybridization and analysis

Laura Graham carried out the design (Figure 3-3), preparation and hybridization of the samples as described in Taylor *et al*. (2005). The only exception to the protocol is that a PICME *Populus* microarray was used in this study, composed of 28,000 elements, including 23,500 cDNAs (Rinaldi, 2007). This set of cDNAs corresponds to 10,000 different gene models in the *P.trichocarpa* genome sequence (Tuskan *et al*., 2006a). The ESTs printed on the PICME poplar arrays were produced by INRA-Nancy (Rinaldi, 2007), INRA-Orleans (Dejardin *et al*., 2004), and University of Helsinki (Brosche *et al*., 2005) within the framework of the LIGNOME and ESTABLISH programme respectively. Dr Nathaniel.Street executed the image and data analysis as described by Taylor *et al*. (2005). ESTs (Expressed Sequence Tags) found to be differentially expressed were located within JGI (Joint Genome Institute) (Tuskan *et al*., 2006a) to obtain Linkage group (LG) position base pairs (bp) and annotation (Table 3-2), using a custom-written R package developed by Dr Nathaniel Street. Scripts are available from Dr Nathanial Street on request and run on R version 2.5. 1 (The R foundation for statistical computing, 2007).

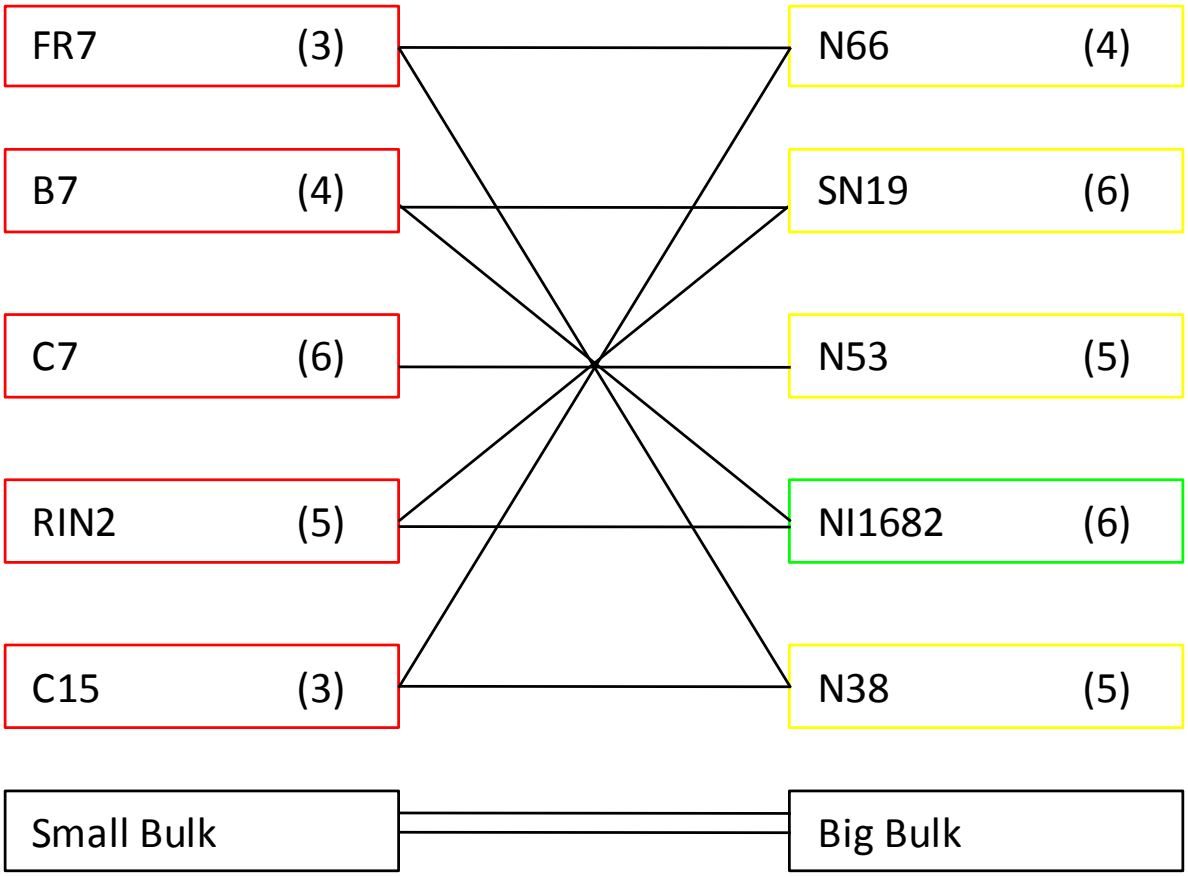


Figure 3-3.A schematic representation of the design of cDNA microarray. Ten genotypes were selected for microarrays, five "big" (right hand side) and five "small" (Left hand side) leaf area genotypes from 2004 growing season. One box represents one genotype and outline represents country (red = Spain, yellow = Italy and green = Netherlands). Numbers in brackets indicate the number of biological replicates for each genotype (Max = 6). Black lines connecting boxes are each hybridization. Bulk analysis is represented by the last two white boxes connected by two black lines, this analysis consists of all biological replicates for the 'small' leaf area genotypes and all of those for the 'big' leaf area genotypes.

Table 3-2: Summary of resulting ESTs from microarray analysis. ESTs classed as significantly differential expressed in 'small' relative to 'big' leaf genotypes. Closest *Arabidopsis* homologs to the PICME EST are shown by TAIR (Swarbreck *et al*., 2008) accession numbers (AGI) and Genbank number. Closest *Arabidipsis* homologs to *Populus tricocharpa* are described with linkage group and gene model in JGI (GM poplar).

Picme EST	AGI no.	Genbank no.	LG	GM poplar	Short description
CA821760	AT1G01100	839410	II	eugene3.00021666	60S acidic ribosomal protein P1 (RPP1A)
(AJ778399, CA821837, CF228665, CF229407, CF232761, CF234679)	AT1G01620	839235	III	estExt_Genewise1_v1.C_LG_III0271	PIP1C (PLASMA MEMBRANE INTRINSIC PROTEIN 1;3)
CF235061	AT1G02180	839543	VI	eugene3.00061544	Ferredoxin-related
(CF230100, CF231249, CF233454)	AT1G04040	839325	II	estExt_Genewise1_v1.C_LG_II4012	Acid phosphatase class B family protein
CA821768	AT1G04270	839564	II	estExt_fgenesh4_pg.C_LG_II0398	RPS15 (RIBOSOMAL PROTEIN S15)
(CA826138, CA826280, AJ770898, AJ774960)	AT1G05010	839345	II	eugene3.00002047	EFE (ethylene forming enzyme)
			XIV	eugene3.00141061	EFE (ethylene forming enzyme)
CF232861	AT1G05570	837059	I	estExt_fgenesh4_pg.C_LG_I0109	CALS1 (CALLOSE SYNTHASE 1)
(AJ769443, CA821838)	AT1G07440	837256	I	estExt_Genewise1_v1.C_LG_I1127	Tropinone reductase, putative

Table 3-2: continued…

(AJ778111, CA823505)			VII	grail3.0019030801	
(CA822865, CA823858, CA823872, CA824281, CF227467)			VIII	estExt_fgenesh4_kg.C_LG_VIII0019	
(CA823650, CA823658, CA824081, CF236272)			X	grail3.0022033301	
AJ769085	AT1G07660	837279	XVIII	eugene3.00180824	Histone H4
(CA826311, CA821551, CA823946)			XVIII	grail3.0020016401	
(CA821310, CA825393, CA825457, CA826260)			XVIII	estExt_fgenesh4_pm.C_LG_XVIII0287	
CA821767	AT1G07920	837307	scaf_28	grail3.0028013201	Elongation factor 1-alpha / EF-1-alpha
CF231397	AT1G08880	837409	XIII	grail3.0016024101	Histone H2A, putative
AJ779580			I	fgenesh4_kg.C_LG_I000011	
CA824004			III	eugene3.00031570	
(CA821628, CA821670)	AT1G09200	837440	scaf_14074	grail3.14074000101	Histone H3
CA822298			scaf_663	estExt_Genewise1_v1.C_6630004	

Table 3-2: Continued…

CF231147	AT1G10155	837553	XV	eugene3.00150996	Similar to ATPP2-A9 (Phloem protein 2-A9)
CF229744	AT1G14720	838037	VIII	estExt_Genewise1_v1.C_LG_VIII2102	XTR2 (XYLOGLUCAN ENDOTRANSGLYCOSYLASE RELATED 2)
CF229935	AT1G14890	838054	VIII	gw1.VIII.1035.1	Invertase/pectin methylesterase inhibitor family protein
CA825479	AT1G19020	838483	XV	grail3.0005005601	Similar to unknown protein
(CF233068, CF234448, CF234800, CF235265, CF236628, CF236831, CF237115)	AT1G22480	838854	II	estExt_fgenesh4_pg.C_LG_II0927	Plastocyanin-like domain-containing protein
CA822641	AT1G24020	839014	scaf_77	estExt_fgenesh4_pg.C_770059	Bet v I allergen family protein
CA823615	AT1G24575	839072	X	eugene3.00100566	Unknown protein
AJ772923	AT1G24620	839076	II	eugene3.00020820	Polcalcin, putative / calcium-binding pollen allergen
(AJ777130, AJ778072, CF231166, CF231436)	AT1G28290	839723	IV	grail3.0046017801	Pollen Ole e 1 allergen and extensin family protein
(CF232020, CF232260, CF232373, CF233223, CF235364, CF235838)	AT1G29380	839813	I	estExt_fgenesh4_pg.C_LG_I2498	Similar to glucan endo-1,3-beta-glucosidase-related
AJ769925	AT1G30900	839974	III	estExt_fgenesh4_pm.C_LG_III0520	Vacuolar sorting receptor, putative
CF233613	AT1G36240	840530	scaf_66	estExt_fgenesh4_pg.C_660113	60S ribosomal protein L30 (RPL30A)

Table 3-2: Continued…

AJ776572	AT1G44191	5007774	I	grail3.0032009701	DNA binding / ligand-dependent nuclear receptor
CF234754	AT1G52720	841705	I	estExt_Genewise1_v1.C_LG_I9453	Similar to unknown protein
(CA821360, CF230463, CF230781, CF231538)	AT1G54575	2745830	XIII	grail3.0016027101	Unknown protein
(CF231041, CF234722)	AT1G54690	841910	scaf_70	estExt_fgenesh4_pg.C_700170	Histone H2A, putative
(AJ769227, CA820871)	AT1G60470	842342	VIII	grail3.0009037801	ATGOLS4 (*ARABIDOPSIS THALIANA* GALACTINOL SYNTHASE 4)
AJ767459			X	estExt_fgenesh4_pg.C_LG_X0618	
(AJ770874, CF228570, CF229106, CF231874, CF232306)	AT1G62480	842545	VI	estExt_fgenesh4_pg.C_LG_VI0110	Vacuolar calcium-binding protein-related
CF231242			XVI	estExt_fgenesh4_pg.C_LG_XVI0075	
CA821764	AT1G67920	843120	XV	grail3.0043013701	Similar to unknown protein
CA823789	AT1G68710	843201	VIII	estExt_Genewise1_v1.C_LG_VIII2173	Haloacid dehalogenase-like hydrolase family protein
(CB239763, CF234601, CF235391)	AT1G69230	843254	scaf_82	grail3.0082000101	SP1L2 – SPIRAL1-LIKE 2

Table 3-2: Continued…

CF233285	AT1G72750	843607	III	fgenesh4_pg.C_LG_III000193	*Arabidopsis thaliana* translocase inner membrane subunit 23-2
CA823771	AT1G73500	843685	scaf_122	gw1.122.164.1	*Arabidopsis thaliana* MAP kinase kinase
CF230930	AT1G73620	843696	XV	gw1.XV.1016.1	Thaumatin-like protein, putative
CA823702	AT1G75380	843874	II	grail3.0003027102	Wound-responsive protein-related
AJ778221	AT1G75500	843886	V	estExt_fgenesh4_pg.C_LG_V1470	Nodulin MtN21 family protein
AJ776096	AT1G77120	844047	II	estExt_fgenesh4_pg.C_LG_II0662	ADH1 (ALCOHOL DEHYDROGENASE 1)
CA826266	AT1G78520	844188	I	eugene3.00012781	Glycosyl hydrolase family protein 17
(CA820794, CA821564, CA824325, CA825553, CF227321, CF227379, CF230335, CF230793, CF230825, CF231020, CF231029, CF231048, CF231099, CF231243, CF231272, CF231432, CF234939)	AT2G02130	814744	XIX	estExt_Genewise1_v1.C_LG_XIX0269	LCR68/PDF2.3 (Low-molecular-weight cysteine-rich 68)
CF236926	AT2G03510	814880	X	gw1.X.834.1	Band 7 family protein
CA821761	AT2G18020	816314	VII	estExt_Genewise1_v1.C_LG_VII3915	EMB2296 (EMBRYO DEFECTIVE 2296)
CA822759	AT2G18660	816381	XVIII	gw1.XVIII.1554.1	Expansin family protein (EXPR3)

Table 3-2: Continued…

CF229317	AT2G18910	816407	scaf_121	estExt_fgenesh4_pg.C_1210049	Hydroxyproline-rich glycoprotein family protein
CA824409	AT2G19800	816499	scaf_145	estExt_Genewise1_v1.C_1450155	MIOX2 (MYO-INOSITOL OXYGENASE 2)
(CA823628, CF228804)	AT2G23810	816913	XVIII	gw1.XVIII.3026.1	TET8 (TETRASPANIN8)
CF234786	AT2G25490	817087	VI	estExt_fgenesh4_pg.C_LG_VI0499	EBF1 (EIN3-BINDING F BOX PROTEIN 1); ubiquitin-protein ligase
AJ767665	AT2G26080	817149	VI	estExt_fgenesh4_pm.C_LG_VI0678	Glycine dehydrogenase (decarboxylating), putative / glycine decarboxylase, putative / glycine cleavage system P-protein, putative
CA825043	AT2G27580	817304	IV	grail3.0045025401	Zinc finger (AN1-like) family protein
CF228786	AT2G27980	817342	II	eugene3.00002366	Protein binding / zinc ion binding
(AJ777094, CF230867, CF232223, CF235463)	AT2G36830	818255	XVI	grail3.0025002201	GAMMA-TIP (Tonoplast intrinsic protein (TIP) gamma)
CF228792	AT2G37170	818293	scaf_28	estExt_fgenesh4_pg.C_280159	PIP2B (plasma membrane intrinsic protein 2;2)
CF230523	AT2G38430	818424	scaf_453	estExt_fgenesh4_pg.C_4530001	Similar to unknown protein
CA821401	AT2G38470	818429	XIII	grail3.0092008201	WRKY33 (WRKY DNA-binding protein 33)
CF235201	AT2G38800	818462	III	grail3.0074005201	Calmodulin-binding protein-related
AJ780726	AT2G40840	818682	XVI	gw1.XVI.1067.1	DPE2 (DISPROPORTIONATING ENZYME 2)

Table 3-2: Continued…

AJ768646	AT2G41415	5007952	scaf_123	grail3.0123004301	Encodes a Maternally expressed gene family protein
(CA821523, CF229634, CF232570, CF236103)	AT2G46170	819224	XIV	eugene3.00140364	Reticulon family protein (RTNLB5)
CF229224	AT2G46370	819244	II	grail3.0039014201	JAR1 (JASMONATE RESISTANT 1)
AJ767433	AT3G01500	821134	I	estExt_Genewise1_v1.C_LG_I2426	CA1 (CARBONIC ANHYDRASE 1); carbonate dehydratase
(AJ773859, CF229882, CF229903, CF230093, CF230218, CF230399, CF230811, CF230914, CF230915, CF231183, CF231325, CF231919, CF231927)	AT3G03430	821255	XII	grail3.0015017401	Polcalcin / calcium-binding pollen allergen, putative
AJ769752	AT3G07310	819919	II	eugene3.00002530	Similar to unknown protein
AJ775600	AT3G08030	819994	I	eugene3.00011733	Similar to unknown protein
CF234314	AT3G13130	820501	I	eugene3.00012637	Similar to unknown protein
CA824310	AT3G13720	820581	scaf_64	eugene3.00640124	Prenylated rab acceptor (PRA1) family protein

Table 3-2: Continued…

CA825088	AT3G15210	820752	VII	estExt_Genewise1_v1.C_LG_VII0556	ATERF-4 (ETHYLENE RESPONSIVE ELEMENT BINDING FACTOR 4)
(CA823499, CA824905, CA826110, CF229085, CF229921, CF230524, CF230751, CF230752, CF230762, CF231328)	AT3G15353	820771	XI	eugene3.00110909	MT3 (METALLOTHIONEIN 3)
CF231342	AT3G15480	820787	scaf_107	estExt_fgenesh4_pg.C_1070075	Similar to unknown protein
CB239377	AT3G16300	820877	I	fgenesh4_pg.C_LG_I001440	Integral membrane family protein
AJ767666	AT3G22120	821775	VI	gw1.VI.1805.1	CWLP (CELL WALL-PLASMA MEMBRANE LINKER PROTEIN); lipid binding
CA822510	AT3G25170	822109	scaf_5222	eugene3.52220001	RALFL26 (RALF-LIKE 26)
CF231331	AT3G26510	822258	VIII	estExt_fgenesh4_pg.C_LG_VIII1649	Octicosapeptide/Phox/Bem1p (PB1) domain-containing protein
(AJ777450, CA825182)	AT3G27060	822324	I	fgenesh4_pg.C_LG_I002334	TSO2 (TSO2); ribonucleoside-diphosphate reductase
CF235704	AT3G27110	822330	I	gw1.I.3059.1	Peptidase M48 family protein
CF233718	AT3G43740	823485	I	estExt_Genewise1_v1.C_LG_I4040	Leucine-rich repeat family protein
CF230544	AT3G44140	823535	XII	fgenesh4_pg.C_LG_XII000630	Unknown protein
(AJ780294, CA820926)	AT3G45310	823669	scaf_28	grail3.0028002001	Cysteine proteinase, putative
CA822494	AT3G45640	823706	IX	estExt_fgenesh4_pm.C_LG_IX0462	MITOGEN-ACTIVATED PROTEIN KINASE 3

Table 3-2: Continued…

(CA822846, CA824181, CF228594)	AT3G46030	823746	II	eugene3.00002335	Histone H2B, putative
CB239375	AT3G46430	823793	X	grail3.0022027101	Similar to unknown protein
(CA821474, CF230163)	AT3G49780	824140	XIV	eugene3.00140037	ATPSK4 (PHYTOSULFOKINE 4 PRECURSOR)
(CA824034, CA825259)	AT3G51030	824267	VII	gw1.VII.3777.1	ATTRX1 (*Arabidopsis thaliana* thioredoxin H-type 1)
(CF227550, CF236126)	AT3G53650	824533	X	estExt_Genewise1_v1.C_LG_X2891	Histone H2B, putative
CF230962	AT3G54260	824593	VIII	eugene3.00080178	Similar to unknown protein
CF235403	AT3G54770	824642	VIII	grail3.0049010601	RNA recognition motif (RRM)-containing protein
AJ773529	AT3G60830	825254	IX	estExt_fgenesh4_pm.C_LG_IX0484	ACTIN-RELATED PROTEIN 7
(CA822742, CF227933, CF230358)	AT3G61640	825337	I	eugene3.00010810	AGP20 (ARABINOGALACTAN PROTEIN 20)
CA825898		825337	III	grail3.0018039001	
AJ768479	AT3G62410	825414	XIV	estExt_Genewise1_v1.C_LG_XIV2304	CP12-2
(CA823148, CA824766, CA824859, CF229072, CF232082)	AT4G02380	828053	II	estExt_Genewise1_v1.C_LG_II1841	SAG21 (SENESCENCE-ASSOCIATED GENE 21)

Table 3-2: Continued…

AJ771208	AT4G12570	826870	XVI	grail3.0101001001	UPL5 (UBIQUITIN PROTEIN LIGASE 5)
(AJ776342, CA824361, CF227616, CF228057, CF231366, CF234647)	AT4G19200	827660	scaf_86	estExt_fgenesh4_pg.C_860158	Proline-rich family protein
(CA823722, CA823747)	AT4G19950	827739	V	estExt_Genewise1_v1.C_LG_V0127	Similar to unknown protein
CA821762	AT4G21960	828285	IV	grail3.0111002302	PRXR1 (peroxidase 42); peroxidase
CF230846	AT4G22010	828290	scaf_252	eugene3.02520012	SKS4 (SKU5 Similar 4) oxidoreductase
CA822314	AT4G25130	828616	scaf_232	estExt_Genewise1_v1.C_2320013	Peptide methionine sulfoxide reductase, putative
AJ767770	AT4G30190	829142	XVIII	estExt_Genewise1_v1.C_LG_XVIII2227	AHA2 (*Arabidopsis* H(+)-ATPase 2); ATPase
CF234682	AT4G31890	829319	VI	estExt_fgenesh4_pm.C_LG_VI0817	Armadillo/beta-catenin repeat family protein
CF235128	AT4G31985	829329	XVIII	eugene3.00181016	60S ribosomal protein L39 (RPL39C)
AJ768434	AT4G35090	829661	V	estExt_fgenesh4_pm.C_LG_V0171	CAT2 (CATALASE 2)
CF233031	AT4G36220	829779	scaf_57	grail3.0057011701	FAH1 (FERULATE-5-HYDROXYLASE 1)
AJ773411	AT4G38460	830002	IX	grail3.0001025501	GERANYLGERANYL REDUCTASE
CF229635	AT4G38660	830022	IX	estExt_Genewise1_v1.C_LG_IX1261	Thaumatin, putative
(AJ770832, AJ774029)	AT4G38970	830052	IV	estExt_Genewise1_v1.C_LG_IV0774	Fructose-bisphosphate aldolase, putative
			IX	estExt_fgenesh4_pm.C_LG_IX0211	Fructose-bisphosphate aldolase, putative

Table 3-2: Continued…

CF235633	AT4G39730	830128	V	estExt_Genewise1_v1.C_LG_V2845	Lipid-associated family protein
(CA822190, CA823238, CA823286, CA823451, CA823828, CA824106, CA824445, CA825598, CB239335, CF229897, CF231444, CF232357, CF235693)	AT5G02560	831891	VI	estExt_Genewise1_v1.C_LG_VI0741	Histone H2A, putative
CF227503	AT5G03650	831769	scaf_28	eugene3.00280310	SBE2.2 (STARCH BRANCHING ENZYME 2.2)
CF230091	AT5G05110	830393	XVI	grail3.0101010901	Cysteine protease inhibitor, putative / cystatin, putative
CF233053		830594	I	estExt_Genewise1_v1.C_LG_I7028	Pepsin A
CA823963	AT5G07730	830666	X	eugene3.00100098	Similar to unknown protein
CB239378	AT5G08530	830752	X	estExt_fgenesh4_pg.C_LG_X0463	NADH-ubiquinone oxidoreductase 51 kDa subunit
(AJ768493, CF229785)	AT5G09660	830825	IX	grail3.0001073802	PEROXISOMAL NAD-MALATE DEHYDROGENASE 2
CF236941	AT5G10160	830880	III	eugene3.00030074	Beta-hydroxyacyl-ACP dehydratase, putative
AJ768910	AT5G10770	830944	XVIII	gw1.XVIII.2006.1	Chloroplast nucleoid DNA-binding protein, putative
CA821443	AT5G12020	831075	VI	estExt_fgenesh4_pm.C_LG_VI0650	17.6 kDa class II heat shock protein (HSP17.6-CII)

Table 3-2: Continued…

(AJ777667, CA821515, CA825362, CA826097)	AT5G13930	831241	XIV	eugene3.00140920	CHS (CHALCONE SYNTHASE)
CA820816	AT5G14700	831322	scaf_123	estExt_Genewise1_v1.C_1230130	Cinnamoyl-CoA reductase-related
CF230114	AT5G14740	831326	X	estExt_fgenesh4_kg.C_LG_X0015	CA2 (BETA CARBONIC ANHYDRASE 2)
CF228450				grail3.0059004501	GASA4 (GAST1 PROTEIN HOMOLOG 4)
(CF231000, CF231220, CF232375)	AT5G15230	831375	XVII	estExt_fgenesh4_pg.C_LG_XVII0378	GASA4 (GAST1 PROTEIN HOMOLOG 4)
CB239376	AT5G15600	831412	XIX	grail3.0094003501	SP1L4 (SPIRAL1-LIKE4)
CF232678	AT5G16250	831485	X	estExt_Genewise1_v1.C_LG_X1460	Similar to unknown protein
CA821766	AT5G20165	832139	VIII	eugene3.00081645	Similar to Os02g0299600
(CF230519, CF231969)	AT5G22580	832321	IX	grail3.0001018301	Identical to the *Arabidopsis* stable protein 1-related
CF231416			scaf_66	estExt_Genewise1_v1.C_660086	Identical to Protein At5g22580
(AJ768966, CA821005)	AT5G37600	833738	IV	estExt_fgenesh4_pm.C_LG_IV0266	*Arabidopsis thaliana* glutamine synthase clone R1
CA824239	AT5G40780	834078	I	estExt_fgenesh4_pg.C_LG_I2255	LHT1 (LYSINE HISTIDINE TRANSPORTER 1)
(CA822738, CF231815)	AT5G42380	834244	scaf_40	grail3.0040020901	Calmodulin-related protein, putative
CB239369	AT5G42650	834273	IV	eugene3.00041130	AOS (ALLENE OXIDE SYNTHASE)
CF231085	AT5G48490	834905	XIV	estExt_Genewise1_v1.C_LG_XIV3201	Lipid transfer protein (LTP) family protein

Table 3-2: Continued…

CF230068	AT5G53300	835411	IV	eugene3.00041353	UBC10 (ubiquitin-conjugating enzyme 10)
CF229210	AT5G54940	835585	X	estExt_Genewise1_v1.C_LG_X0725	Eukaryotic translation initiation factor SUI1, putative
(CF230983, CF231244, CF231908)	AT5G57020	835805	XII	fgenesh4_pg.C_LG_XII000841	NMT1 (N-MYRISTOYLTRANSFERASE 1)
CA821534	AT5G59720	836093	VIII	eugene3.00080557	HSP18.2 (HEAT SHOCK PROTEIN 18.2)
CB239372	AT5G59820	835805	IX	eugene3.00091367	RHL41 (RESPONSIVE TO HIGH LIGHT 41)
CA823712	AT5G60030	836125	X	estExt_fgenesh4_pg.C_LG_X1782	Similar to unknown protein
(AJ767532, CF230476, CF230498)	AT5G60530	836174	I	estExt_fgenesh4_pg.C_LG_I0819	Late embryogenesis abundant / LEA protein-related
CA821759	AT5G61170	836238	IV	eugene3.00041113	40S ribosomal protein S19
CF234346	AT5G62550	836375	scaf_129	estExt_fgenesh4_pg.C_1290017	Similar to unknown protein
AJ777792	AT5G62740	836395	XVII	estExt_fgenesh4_pg.C_LG_XVII0326	Band 7 family protein
(CF233734, CF236061, CF237249)	AT5G64310	836552	XVII	eugene3.00170110	AGP1 (ARABINOGALACTAN-PROTEIN 1)

ATG is the *Arabidopsis* gene product Identification. CA, CB, AJ & CF is the PICME accession codes. Scaf is the portion of the genome sequence composed of contigs & gaps.

3.6 Combining QTL and Microarray analysis

3.6.1 Candidate gene selection

3.6.2 Literature Searches

Genes involved in leaf morphology and physiology were first found from literature. Literature searches were conducted in the Web of Science databases (Thomson Reuters, USA) using keywords; leaf, poplar, leaf genes and biomass. Genes were then identified by name in *Arabidopsis* using The *Arabidopsis* Information Resource database (TAIR) (TAIR curators, USA). Data collected from TAIR included the accession number, sequence length, function and link to the National Center for Biotechnology Information (NCBI) (National Library of Medicine, USA) providing the full genomic nucleotide sequence for each gene. The full genomic sequence was then blasted in *P. trichocarpa* sequence database : Joint Genome Institute (JGI) (Tuskan *et al*., 2006a) using a heuristic database-searching method called tBLASTx to find high scoring segment pairs (HSP) from the original *Arabidopsis* nucleotide sequence query. The highest statistically significant sequence alignment relative to the query sequence was then saved and information collected including linkage group (LG), base pair position and annotation (Table 3-3).

Table 3-3: Leaf development candidate genes.Details of candidate genes involved in leaf morphology, found in The *Arabidopsis* Information Resource (TAIR curators, USA) name, AGI number and National Center for Biotechnology Information (NCBI, National Library of Medicine, USA) number shown. *Populus trichocrapa* homologs to *Arabidopsis* found by BLASTing using tBLASTx in Joint Genome Institute (JGI) (Tuskan *et al*., 2006a) labeled by linkage group, gene model (GM poplar) and description.

Gene & (NCBI no.)	**AGI no.**	**LG**	**GM poplar**	**Short Description**	**Reference**
PHABULOSA (818036)	AT2G34710	I	estExt_Genewise1_v1.C_LG_I4896	Dominant PHB mutations cause transformation of abaxial leaf fates into adaxial leaf fates.	(Ohno *et al*., 2004)
RBR1 (820408)	AT3G12280	I	grail3.0100006103	This protein is involved in G1/S cell cycle transition in plants. transcription for progression into S phase.	(Sabelli *et al*., 2005)
TORNADO1 (835648)	AT5G55540	I	fgenesh1_pg.C_LG_I002802	In trn mutants, leaf laminas were asymmetric and narrow because of a severely reduced cell number.	(Cnops *et al*., 2006)
YABBY (YAB3) (827914)	AT4G00180	II	grail3.0033028501	YABBY gene family member, likely has transcription factor activity, involved in specifying abaxial cell fate.	(Siegfried *et al*., 1999b;Ohno *et al*., 2004)
KANADI1 (831518)	AT5G16560	II	fgenesh4_pg.C_LG_II002170	Encodes a KANADI protein (KAN) that regulates organ polarity in *Arabidopsis*.	(Ohno *et al*., 2004)
CYCA3;2 (841124)	AT1G47210	II	estExt_Genewise1_v1.C_LG_II0565	Cyclin family protein, a core cell cycle gene	(Dewitte and Murray, 2003)

Table 3-3: Continued…

CYCD4;1 (836667)	AT5G65420	II	fgenesh1_pg .C_LG_II000 954	Encodes a D-type cyclin that physically interacts with CDC2A and is expressed during vascular tissue development, embryogenesis, and formation of lateral root primordia.	(Dewitte and Murray, 2003)
E2Fc (841202)	AT1G47870	II	eugene3.000 20210	Member of the E2F transcription factors, (cell cycle genes), key components of the cyclin D/retinoblastoma/E 2F pathway.	(Sozzani *et al*., 2006)
ATGRF7 (835447)	AT5G53660	III	grail3.0018 022201	Growth regulating factor encoding transcription activator. Involved in leaf development and expressed in shoot and flower.	(Kim *et al*., 2003)
YABBY (YAB2) (3766682)	AT1G08465	III	grail3.0018 017701	Member of the YABBY family of *Arabidopsis* proteins involved in the abaxial cell fate specification in lateral organs	(Siegfried *et al*., 1999a;Ohno *et al*., 2004)
CYCD5;1 (829917)	AT4G37630	III	eugene3.000 30604	Core cell cycle genes	(Dewitte and Murray, 2003)

Table 3-3: Continued…

TORNADO2 (834713)	AT5G46700	III	grail3.0018 000201	Encodes a transmembrane protein of the tetraspanin (TET) family. Required for the maintenance of both the radial pattern of tissue differentiation.	(Cnops *et al.*, 2006)
KANADI (KAN1) (831518)	AT5G16560	IV	fgenesh1_pm .C_LG_IV000 191	*Arabidopsis thaliana* myb family transcription factor (KAN1). Involved in leaf development.	(Emery *et al.*, 2003)
PLL5 (837276)	AT1G07630	IX	estExt_Gene wise1_v1.C_ LG_IX4231	Encodes a protein phosphatase 2C like gene, similar to POL. Involved in leaf development.	(Song *et al.*, 2006)
CKS1 (817340)	AT2G27960	IX	estExt_fgen esh1_pg_v1. C_LG_IX1496	Catalytic subunit of cyclin dependent kinase 1. Role in the regulation of the cell cycle	(Dewitte and Murray, 2003)
E2F1 (832283)	AT5G22220	IX	fgenesh1_pm .C_LG_IX000 037	Member of the E2F transcription factors, key components of the cyclin D/retinoblastoma/E 2F pathway. Promotes cell division and shortens cell doubling time, inhibits cell growth.	(Sozzani *et al.*, 2006)

Table 3-3: Continued…

KRP7 (841386)	AT1G49620	IX	eugene3.00090629	Kip-related protein (KRP) gene, encodes CDK (cyclin-dependent kinase) inhibitor (CKI), negative regulator of cell division.	(De Veylder *et al.*, 2001)
AINTEGUMENTA(829931)	AT4G37750	V	eugene3.00050574	Arabidopsis thaliana ovule development protein aintegumenta (ANT)	(Mizukami and Fischer, 2000)
CDKB2;1 (843987)	AT1G76540	V	estExt_fgenesh1_pg_v1.C_LG_V1708	Encodes a cyclin-dependent protein kinase involved in regulation of the G2/M transition of the mitotic cell cycle.	(Dewitte and Murray, 2003)
LEUING (829390)	AT4G32551	VI	estExt_fgenesh1_pg_v1.C_LG_VI0219	Arabidopsis thaliana WD-40 repeat family protein (LEUNIG)	(Sinha, 1999)
PLL4 (817438)	AT2G28890	VI	fgenesh1_pg.C_LG_VI000683	Encodes a protein phosphatase 2C like gene, similar to POL. Involved in leaf development.	(Song and Clark, 2005a)
AS1 (818340)	AT2G37630	VI	estExt_fgenesh4_pm.C_LG_VI0283	Encodes a MYB-domain protein involved in specification of the leaf proximodistal axis.	(Zgurski *et al.*, 2005)
PHAVOLUTA (839928)	AT1G30490	VI	estExt_fgenesh1_pm_v1.C_LG_VI0133	Dominant PHV mutations cause transformation of abaxial leaf fates into adaxial leaf fates.	(Ohno *et al.*, 2004)

Table 3-3: Continued…

BOP1 (824880)	AT3G57130	VI	fgenesh1_pg.C_LG_VI000366	Lines carrying recessive mutations exhibit a number of visible defects, most pronounced being ectopic outgrowths of in leaf petioles of rosette leaves.	(Ohno *et al.*, 2004)
XTH9 (828024)	AT4G03210	VI	fgenesh1_pg.C_LG_VI000550	Encodes a member of xyloglucan endotransglucosylase/hydrolases (XTHs) that catalyze the cleavage and molecular grafting of xyloglucan chains function in loosening and rearrangement of the cell wall.	(Akamatsu *et al.*, 1999)
GPA1 (817170)	AT2G26300	VI	fgenesh4_pg.C_LG_VI001473	Positive regulator in abscisic acid (ABA) inhibition of stomatal opening.	(Huang *et al.*, 2006)
WUSCHEL (816305)	AT2G17950	VII	grail3.0019031001	Homeobox gene controlling the stem cell pool.	(Kieffer *et al.*, 2006)
CURLY LEAF (816870)	AT2G23380	VII	estExt_Genewise1_v1.C_LG_VII2923	Polycomb group protein CURLY LEAF Protein INCURVATA 1)	(Goodrich *et al.*, 1997)
ATGRF2 (829930)	AT4G37740	VII	estExt_Genewise1_v1.C_LG_VII0082	Growth regulating factor encoding transcription activator. Mutants result in smaller leaves indicating the role of the gene in leaf development. Expressed in root, shoot and flower	(Kim *et al.*, 2003)

Table 3-3: Continued…

ASL1 (836821)	AT5G66870	VII	fgenesh1_pg.C_LG_VII000913	Encodes LOB domain protein whose overexpression results in KNOX gene repression.	(Chalfun *et al.*, 2005)
CYCD4;1 (836667)	AT5G65420	VII	fgenesh1_pg.C_LG_VII001256	Encodes a D-type cyclin that physically interacts with CDC2A and is expressed during vascular tissue development, embryogenesis, and formation of lateral root primordia.	(Dewitte and Murray, 2003)
DRL1 (837946)	AT1G13870	VIII	fgenesh1_pm.C_LG_VIII000356	Encodes a homolog of the yeast TOT4/KTI12 protein. Yeast TOT4/KTI12 associates with Elongator, a multisubunit complex that binds the RNA polymerase II transcription elongation complex.	(Nelissen *et al.*, 2003)
KORRIGAN (835035)	AT5G49720	VIII	estExt_fgenesh1_pg_v1.C_LG_VIII1292	Endo-1,4-beta-glucanase KORRIGAN (KOR) / cellulase)	(Dewitte and Murray, 2003)
EXGT-A3 (XTH27) (814716)	AT2G01850	VIII	estExt_Genewise1_v1.C_LG_VIII2102	EXGT-A3 has homology to xyloglucan endotransglucosylases/hydrolases (XTHs).	(Akamatsu *et al.*, 1999)
CYCB3;1 (838202)	AT1G16330	VIII	eugene3.00080764	Core cell cycle genes	(Dewitte and Murray, 2003)

Table 3-3: Continued…

CYCD3 (829564)	AT4G34160	VIII	eugene3.00081371	Encodes a cyclin D-type protein involved in the switch from cell proliferation to the final stages of differentiation.	(Dewitte and Murray, 2003)
PPD1 (827123)	AT4G14713	VIII	grail3.0090002601	*Populus tremula x Populus tremuloides* pumilio domain-containing protein PPD1 (PPD1)	(White, 2006)
AS2 (842873)	AT1G65620	VIII	grail3.0010011101	Required for formation of a symmetric flat leaf lamina.	(Zgurski *et al.*, 2005)
DIM (821519)	AT3G19820	X	fgenesh1_pg.C_LG_X001540	Arabidopsis thaliana cell elongation protein / DWARF1 / DIMINUTO (DIM)	(Takahashi *et al.*, 1995)
ZWILLE (834403)	AT5G43810	X	estExt_Genewise1_v1.C_LG_X5681	Along with WUS and CLV genes, controls the relative organization of central zone and peripheral zone cells in meristems.	(Sinha, 1999)
JAG (843177)	AT1G68480	X	fgenesh1_pg.C_LG_X001157	Encodes a putative zinc finger transcription factor that is necessary for proper lateral organ shape and is sufficient to induce the proliferation of lateral organ tissue.	(Ohno *et al.*, 2004)

Table 3-3: Continued…

H2B (824533)	At3G53650	X	estExt_fgenesh1_pg_v1.C_LG_X2080	Histone H2B, putative	(Okada *et al.*, 2005)
STM (842534)	AT1G62360	XI	estExt_fgenesh4_pm.C_LG_XI0028	Class I knotted-like homeodomain protein that is required for shoot apical meristem (SAM) formation during embryogenesis and for SAM function throughout the lifetime of the plant.	(Scofield and Murray, 2006)
H2A (841528)	AT1G51060	XI	grail3.0014021301	Histone H2A, putative	(Okada *et al.*, 2005)
EXGT-A1 (XTH4) (815247)	AT2G06850	XI	fgenesh1_pm.C_LG_XI000043	Xyloglucan:xyloglucosyl transferase / xyloglucan endotransglycosylase / endo-xyloglucan transferase (EXT) (EXGT-A1)	(Akamatsu *et al.*, 1999)
AGAMOUS (827631)	At4g18960	XI	eugene3.00110505	DNA binding, transcription factor activity, specification of carpel identity, carpel development, stamen development, maintenance of floral organ identity, ovule, stamen primordium, carpel primordium	(Gregis *et al.*, 2008)
ACC oxidase (839345)	AT1G05010	XI	eugene3.00110176	1-aminocyclopropane-1-carboxylate oxidase / ACC oxidase / ethylene-forming enzyme (ACO)	(Chen and McManus, 2006)

Table 3-3: Continued…

PFS2 (814678)	AT2G01500	XII	grail3.0042003201	PFS2 encodes a homeodomain gene that is a member of the WUS clade of transcription factors. It delays differentiation and maturation of primordia and regulates ovule patterning.	(Park *et al.*, 2005)
ATGRF8 (828515)	AT4G24150	XII	eugene3.00120277	Growth regulating factor encoding transcription activator. One of the nine members of a GRF gene family, containing nuclear targeting domain. Involved in leaf development and expressed in shoot and flower.	(Kim *et al.*, 2003)
SWELLMAP2 (SMP2) (829866)	AT4G37120	XII	fgenesh1_pm.C_LG_XII000222	Encodes a zinc finger containing protein, plants make smaller organs having reduced cell numbers but increased cell size.	(White, 2006)
ATGRF3 (818213)	AT2G36400	XIII	eugene3.00130769	Growth regulating factor encoding transcription activator. Mutants result in smaller leaves indicating the role of the gene in leaf development. Expressed in root, shoot and flower.	(Kim *et al.*, 2003)

Table 3-3: Continued…

ATGRF6 (815176)	AT2G06200	XIII	eugene3.00130769	Growth regulating factor encoding transcription activator. Mutants result in smaller leaves indicating the role of the gene in leaf development. Expressed in root, shoot and flower.	(Kim *et al.*, 2003)
AN3 (832968)	AT5G28640	XIII	eugene3.00130435	Encodes a protein with similarity to mammalian transcriptional coactivator that is involved in cell proliferation during leaf and flower development.	(Horiguchi *et al.*, 2005b)
TCH4, XTH22 (835860)	AT5G57560	XIII	eugene3.00130049	Encodes a cell wall-modifying enzyme, rapidly upregulated in response to environmental stimuli	(Akamatsu *et al.*, 1999)
STRUWWELPETER (819634)	AT3G04740	XIII	gw1.XIII.1514.1	Encodes a protein with similarities to subunits of the Mediator complex, required for RNA polymerase II recruitment at target promoters in response to specific activators. Lines carrying loss of function mutations in the gene have reduced cell numbers in aerial organs.	(White, 2006)
ANGUSTIFOLIA (839401)	AT1G01510	XIV	eugene3.00140349	Leaf development.	(Tsuge *et al.*, 1996)

Table 3-3: Continued…

PROLIFERA (828153)	AT4G02060	XIV	eugene3.00140663	Arabidopsis thaliana prolifera protein (PRL) / DNA replication licensing factor Mcm7 (MCM7)	(Springer *et al.*, 2000;Sozzani *et al.*, 2006)
ATGRF1 (816815)	AT2G22840	XIV	fgenesh1_pg.C_LG_XIV0000039	Growth regulating factor encoding transcription activator. Mutants result in smaller leaves indicating the role of the gene in leaf development. Expressed in root, shoot and flower.	(Kim *et al.*, 2003)
H3 (830903)	AT5G10390	XIV	fgenesh1_kg.C_LG_XIV	histone H3	(Okada *et al.*, 2005)
CYCD3;3 (824169)	AT3G50070	XIV	eugene3.00140027	core cell cycle genes	(Dewitte and Murray, 2003)
ATGRF9 (819156)	AT2G45480	XIX	eugene3.00190480	Growth regulating factor encoding transcription activator. One of the nine members of a GRF gene family, containing nuclear targeting domain. Involved in leaf development.	(Kim *et al.*, 2003)
KNAT6 (838946)	AT1G23380	XV	fgenesh1_pg.C_LG_XV000528	required for shoot apical meristem (SAM) formation during leaf development	(Belles-Boix *et al.*, 2006)
LEAFY (836307)	AT5G61850	XV	estExt_fgenesh4_pm.C_LG_XV0337	Involved in floral mersitem, over expression results in changes in leaf size	(Rottmann *et al.*, 2000)

Table 3-3: Continued…

LOB (836429)	AT5G63090	XV	grail3.0005001801	Arabidopsis thaliana LOB domain protein / lateral organ boundaries protein (LOB), involved in lamina formation	(Lin *et al.*, 2003)
ATGRF5 (820609)	AT3G13960	XV	eugene3.00150073	Growth regulating factor encoding transcription activator. One of the nine members of a GRF gene family, containing nuclear targeting domain. Involved in leaf development.	(Kim *et al.*, 2003)
ACC synthase (837082)	AT1G01480	XV	eugene3.00151103	a member of the 1-aminocyclopropane-1-carboxylate (ACC) synthase gene family, involved in leaf development	(Wang *et al.*, 2005)
DEL1 (823971)	AT3G48160	XV	estExt_fgenesh1_pg_v1.C_LG_XV0448	E2F-like protein, an inhibitor of the endocycle, preserves the mitotic state of proliferating cells by suppressing transcription of genes that are required for cells to enter the DNA endoreduplication cycle.	(Vlieghe *et al.*, 2005)

Table 3-3: Continued…

DEL3 (821221)	AT3G01330	XV	estExt_fgenesh1_pg_v1.C_LG_XV0448	Member of the E2F transcription factors, key components of the cyclin D/retinoblastoma/E2F pathway.	(Vlieghe *et al.*, 2005)
SWELLMAP1 (SMP1) (842877)	AT1G65660	XV	fgenesh1_pm.C_LG_XV000071	Encodes a CCHC zinc finger protein that may function as a step II splicing factor. In an epigenetic allele of SMP1 (in which SMP1 and SMP2 mRNA is reduced) organs are smaller and contain fewer cells.	(White, 2006)
UV14 (818827)	AT2G42260	XVI	eugene3.00160434	Encodes a novel plant-specific protein of unknown function. The UVI4 gene is expressed mainly in actively dividing cells.	(Hase *et al.*, 2006)
EXP6 (817444)	AT2G28950	XVI	eugene3.00160918	expansin, putative (EXP6), similar to expansin.	(Li *et al.*, 2003)
DPA (830987)	AT5G02470	XVI	estExt_Genewise1_v1.C_LG_XVI2943	core cell cycle genes	(Sozzani *et al.*, 2006)
DPB (831847)	AT5G03415	XVI	estExt_Genewise1_v1.C_LG_XVI2943	Encodes a homolog of the animal DP protein. DP, in animals, forms a heterodimer with E2F and plays a central role in G1/S transition in the cell division cycle.	(Sozzani *et al.*, 2006)

Table 3-3: Continued…

E2F3 (818174)	AT2G36010	XVI	eugene3.00160662	Member of the E2F transcription factors, (cell cycle genes), key components of the cyclin D/retinoblastoma/E2F pathway.	(Sozzani *et al.*, 2006)
KRP3 (834940)	AT5G48820	XVII	estExt_Genewise1_v1.C_LG_XVII0448	Kip-related protein (KRP) gene, encodes CDK (cyclin-dependent kinase) inhibitor (CKI), negative regulator of cell division.	(De Veylder *et al.*, 2001)
ERECTA (817173)	AT2G26330	XVIII	gwl.XVIII.1052.1	Homologous to receptor protein kinases. Involved in specification of organs originating from the shoot apical meristem.	(Shpak *et al.*, 2003)
KNAT3 (832593)	AT5G25220	XVIII	estExt_fgenesh1_pm_v1.C_LG_XVIII0152	Arabidopsis thaliana homeobox protein knotted-1 like 3 (KNAT3) (At5g25220) mRNA, complete cds	(Fleming, 2003)
CYCA2;1 (832610)	At5G25380	XVIII	fgenesh1_pm.C_LG_XVIII000101	core cell cycle genes	(Fleming, 2003)
H4 (836090)	At5G59690	XVIII	grail3.0020016401	histone H4	(Okada *et al.*, 2005)

3.6.3 Functional characterisation of literature and microarray candidate genes

Candidate genes identified from microarray (Table 3-2) and literature searches (Table 3-3) were collected and sorted according to their accession number into three groups; literature, microarray up-regulated and microarray down-regulated. These groups were then analysed separately using the web based tool: bulk data retrival and analysis using GO annotations, which is freely available from TAIR. This tool enabled *Arabidopsis* genes to be functionally classified into four groups; cellular, development, metabolic and biological for comparisons.

3.6.4 Mapping Genes & ESTs

Using linkage group and base pair position of each gene enabled the co-location to a pedigree genetic linakge map produced by Tuskan *et al* (personal communication, http://www.ornl.gov/sci/ipgc/ssr_resource.htm), consisting of 91 SSRs genotyped on 350 individuals and 92 fully informative AFLPs genotyped on 165 individuals. The programme called Map Chart (Voorrips, 2002) was used to display the genetic map, QTLs, ESTs and candidate genes. QTLs were separated into two separate groups for candidate gene selection; (i) leaf morphological traits and biomass traits, biomass was included due to biomass traits being highly correlated with leaf size (2.4), (ii) cell traits. Candidate genes were selected based on; (i) found within QTL region containing either three (major hotspot) or two experiments (minor hotspot) and (ii) know function in leaf development (i.e. transformations cause differences in phenotype).

3.7 Results

3.7.1 Comparative functional categorisation of literature and microarray candidate genes.

Comparison of the number of genes in specific physiological processes between literature and microarray searches, demonstrate the major differences. Literature searches consisted for a large majority of candidate genes involved in developmental and cellular process (Figure 3-4A). This is not unexpected due to the specific search engines and restraints put in place. Microarray analysis shows similar results in both up (Figure 3-4C) and down (Figure 3-4B) regulated genes, whereby the majority of genes function in metabolite processes followed by cellular processes. This indicates that the two search engines provide very different outlooks on genes involved in leaf size. However combining both techniques provides a powerful tool when co-location to QTL analysis.

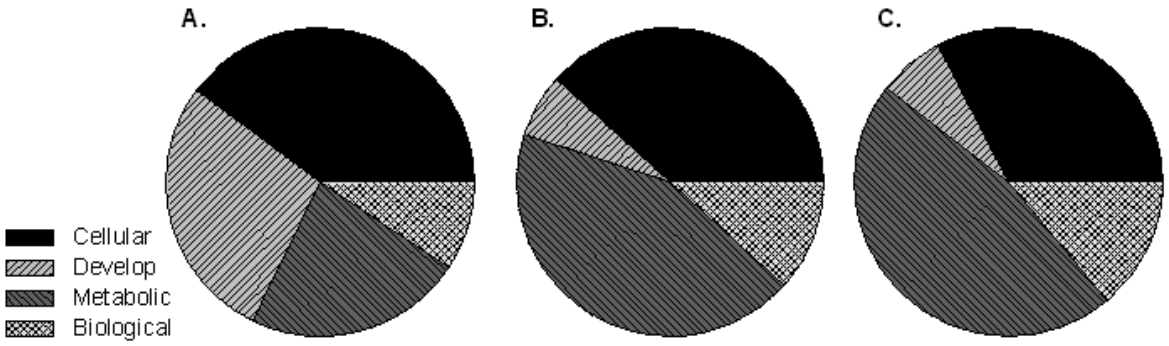


Figure 3-4: The functional diversity of candidate genes. A comparison of the functional diversity of candidate genes identified from; literature searches (A), microarray down-regulated (B), microarray up-regulated (C). Pie charts show proportional size of each process; cellular, developmental, metabolic and biological.

3.7.2 Co-location of candidate genes to QTL

16 leaf, 18 biomass and 12 cell traits were submitted to QTL collocation mapping from plants grown in three different experiments; drought, CO_2 and SRC. A total of 274 leaf and biomass QTL and 129 cell QTL were mapped in the pedigree, QTLs mapped to all linkage groups for both trait sets. The highest number of QTL (>20) for leaf and biomass traits was seen on linkage groups III, IV, VIII and XI (Figure 3-5C, D & H), whereas the highest number of QTL (>15) for cell traits was seen on linkage groups; III and XII (C & L). Major hotspot regions of the genetic map for leaf and biomass traits are shown on linkage groups; I, IV, V, VI, VIII, IX,

X, XIII, XIV, XV and XVII (Figure 3-5A, D, E, F, H, I, J, M, N, O & Q). Minor hotspot regions of the genetic map for leaf and biomass traits are shown on linkage groups; I, II, III, VI, VII, XI, XII, XV, XVI, XVIII and XIX (Figure 3-5A, B, C, F, G, K, N, P, R & S). On linkage groups IV, IX and XVII; QTL for leaf area is found in all three experiments and located in the same area of the genetic map (Figure 3-5D, I, Q), indicating good areas for finding candidate genes for leaf size. Linkage groups VIII and X have three experiments present at a QTL for Specific Leaf Area (SLA) located in the same area of the genetic map (Figure 3-5H & J). All linkage groups showed QTL collocating to the same area of the genetic map for at least two experiments (Figure 3-5) Many QTL co-located for correlated traits such as individual leaf area with biomass traits; height, biomass and stem volume index on linkage groups VI, VIII, IX, XIII, XVI, XVIII & XIX (Figure 3-5F, H, I, M, N, Q & S),indicating that candidate genes found within these regions could determine both traits. One major hotspot for cell traits is shown on linkage group V (Figure 3-6E), whereas minor hotspot regions are shown on linkage groups I, II, III, IV, V, VI, VIII, IX, XI, XII, XIII, XIV and XVII (Figure 3-6A, B, C, D, E, F, H, I, K, L, M, N & Q). No hotspots for cell traits are shown on linkage groups; X, XV, XVI, XVIII and XIX (Figure 3-6J, O, P, R & S). On linkage groups I, III, IV, VIII, IX, XI and XII; QTL for cell number is found in both the drought and CO_2 experiment (Figure 3-6A, C, D, H, I, K & L). Cell area QTL is found on linkage groups II, III, VIII, XII and XIV for two experiments (Figure 3-6B, C, H, L & N). Stomatal index and stomatal density QTL co-locate to the same regions on linkage groups XII and XIII (Figure 3-6L & M), whereas as they both co-locate to linkage group III but at different regions (Figure 3-6C).

3.7.3 Parallel approaches; literature, microarray and QTL, yield strong candidate genes for leaf development

Combining three QTL experiments, a microarray experiment resulting in 293 transcripts differently expressed and literature searches resulting in 79 candidate genes making a total of 372 possible candidate genes. Selecting hotspot regions has reduced this number to 132 leaf and biomass candidates and 68 cell candidates (Figure 3-5& Figure 3-6). Strong candidate genes were found in areas of the genetic map consisting of QTL from all three experiments known as major hotspots, 10 are seen for leaf and biomass traits (Figure 3-5), whereas one is shown for cell traits (Figure 3-6). Candidates genes ranged from 11 to 1 within these 'major hotspots', the largest number of candidate genes for leaf and biomass were found on linkage group XIV between markers CACTG-25 and p_571 (Figure 3-5N),consisting of the genes; *CHALCONE SYNTHASE* (eugene3.00140920), *lipid transfer protein (LTP) family protein*

(estExt_Genewise1_v1.C_LG_XIV3201), *ethylene forming enzyme* (eugene3.00141061), CYCD3;3 (eugene3.00140027), PHYTOSULFOKINE 4 PRECURSOR (eugene3.00140037), ATGRF1 (fgenesh1_pg.C_LG_XIV000039), ANGUSTIFOLIA (eugene3.00140349), reticulon family protein (eugene3.00140364), Histone 3 (fgenesh1_kg.C_LG_XIV), CP12-2 (estExt_Genewise1_v1.C_LG_XIV2304) and PROLIFERA (eugene3.00140663) (Figure 3-5N). Interestingly minor hotspots for cell traits are found in the same regions as major hotspots for leaf and cell traits in linkage groups I, IV, V, VI, VIII, IX, X, XIII, XIV, XV and XVII (Figure 3-5 & Figure 3-6 A, D, E, F, H, I, J, L, M, N & P). Therefore genes described above are strong candidates not only for leaf and biomass, but also cell traits. Table 3-6 shows the final selection of candidate genes found within minor hotspots for leaf,biomass and cell traits QTL, which resulted in 86 'minor' candidate genes for leaf development.(Table 3-6). Whereas linkage group V was the only major hotspot for cell traits which consisted of two candidate genes; nodulin MtN21 family protein (poplar GM: estExt_fgenesh4_pg.C_LG_V1470) and CDKB2;1 (poplar GM: estExt_fgenesh1_pg_v1.C_LG_V1708). Major hotspots for leaf and biomass are also seen in linkage group I consisting of three candidate genes; LEA protein-related (estExt_fgenesh4_pg.C_LG_I0819), ARABINOGALACTAN PROTEIN 20 (eugene3.00010810) and a unknown protein (estExt_Genewise1_v1.C_LG_I9453) (Figure 5A), linkage group IV consisting of five candidate genes; 40S ribosomal protein S19 (eugene3.00041113), ALLENE OXIDE SYNTHASE (eugene3.00041130), fructose-bisphosphate aldolase, putative (estExt_Genewise1_v1.C_LG_IV0774), ubiquitin-conjugating enzyme 10 (eugene3.00041353) and zinc finger (AN1-like) family protein (grail3.0045025401) (Figure 3-5D), linkage group V consisting of four candidate genes; AINTEGUMENTA (eugene3.00050574), a unknown protein (estExt_Genewise1_v1.C_LG_V0127), nodulin MtN21 family protein (estExt_fgenesh4_pg.C_LG_V1470) and CDKB2;1 (estExt_fgenesh1_pg_v1.C_LG_V1708) (Figure 3-5E), linkage group VI consisting of five candidate genes; CELL WALL-PLASMA MEMBRANE LINKER PROTEIN (gw1.VI.1805.1), vacuolar calcium-binding protein-related (estExt_fgenesh4_pg.C_LG_VI0110), BOP1 (fgenesh1_pg.C_LG_VI000366), EIN3-BINDING F BOX PROTEIN 1(estExt_fgenesh4_pg.C_LG_VI0499) and XTH9 (fgenesh1_pg.C_LG_VI000550) (Figure 5F), linkage group VIII consisting of five candidate genes; HEAT SHOCK PROTEIN 18.2 (eugene3.00080557), KORRIGAN (estExt_fgenesh1_pg_v1.C_LG_VIII1292), AS2 (fgenesh1_pg.C_LG_VIII000690), CYCB3;1 (eugene3.00080764) and DRL1 (fgenesh1_pm.C_LG_VIII000356) (Figure 3-5H),

linkage group IX consisting of three candidate genes; PLL5 (estExt_Genewise1_v1.C_LG_IX4231), *RESPONSIVE TO HIGH LIGHT 41* (eugene3.00091367) and *CKS1* (estExt_fgenesh1_pg_v1.C_LG_IX1496) (Figure 3-5I), linkage group X consisting of four candidate genes; three *unknown proteins* (estExt_Genewise1_v1.C_LG_X1460, estExt_fgenesh4_pg.C_LG_X1782 & grail3.0022027101) and *histone H4* (grail3.0022033301) (Figure 5J), linkage group XIII consisting of three candidate genes; *STRUWWELPETER* (gw1.XIII.1514.1), *AN3* (eugene3.00130435) and *ATGRF3* (eugene3.00130769) (Figure 5M), linkage group XV consisting of one candidate gene; *LEAFY* (estExt_fgenesh4_pm.C_LG_XV0337) (Figure 5O) and linkage group XVII consisting of five candidate genes*; ARABINOGALACTAN-PROTEIN 1* (eugene3.00170110*), KRP3* (estExt_Genewise1_v1.C_LG_XVII0448), *band 7 family protein* (estExt_fgenesh4_pg.C_LG_XVII0326), *GAST1 PROTEIN HOMOLOG 4* (estExt_fgenesh4_pg.C_LG_XVII0378 & grail3.0059004501) (Figure 3-5Q). Less favorable candidate genes were found in regions of the genetic map consisting of QTL from two experiments known as minor hotspots. Interestingly minor hotspots for cell traits are found in the same regions as major hotspots for leaf and cell traits in linkage groups I, IV, V, VI, VIII, IX, X, XIII, XIV, XV and XVII (Figure 3-5 & Figure 3-6 A, D, E, F, H, I, J, L, M, N & P). Therefore genes described above are strong candidates not only for leaf and biomass, but also cell traits. Table 3-6 shows the final selection of candidate genes found within minor hotspots for leaf,biomass and cell traits QTL, which resulted in 86 'minor' candidate genes for leaf development. Table 3-6).

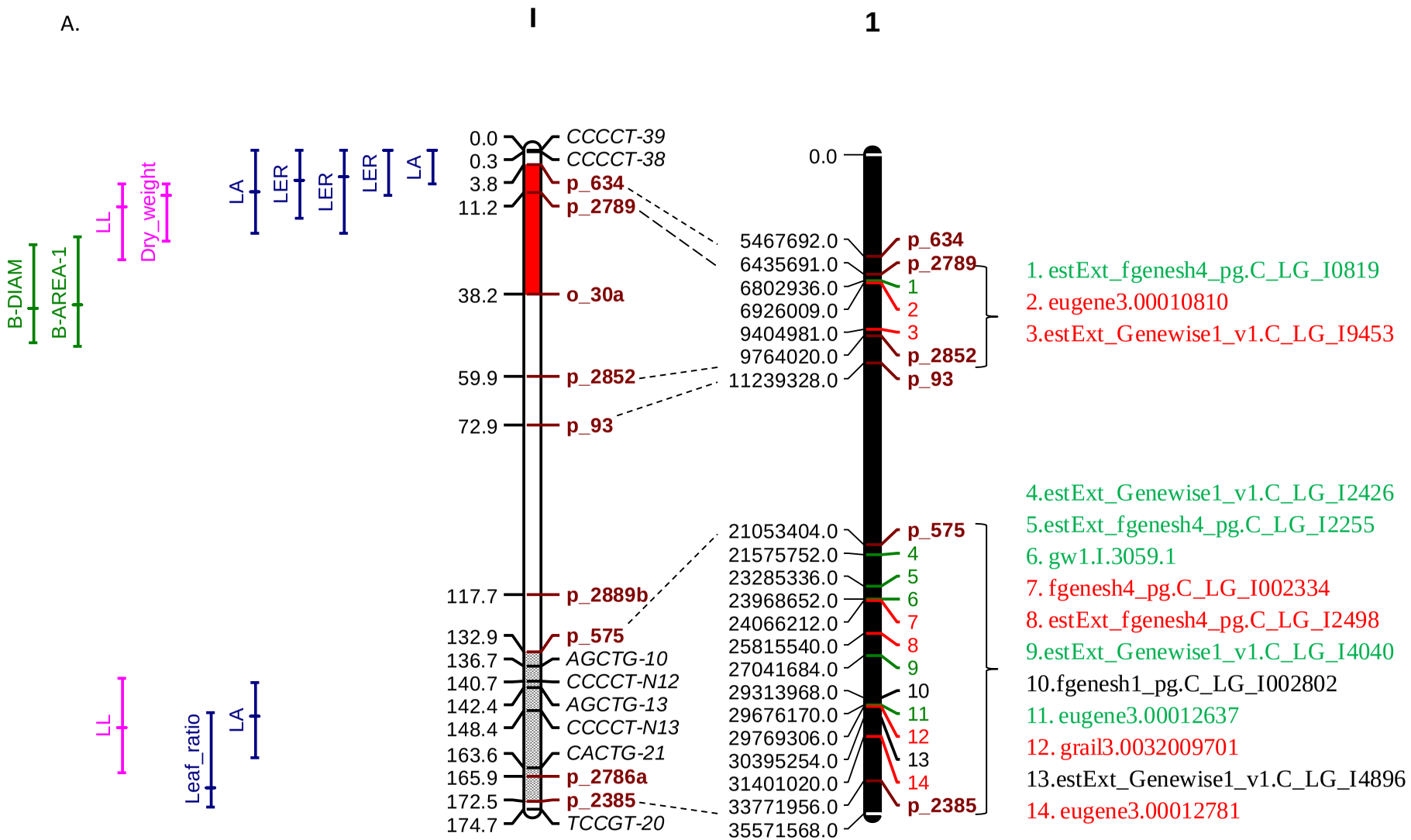
Figure 3-5:QTL map of leaf and biomass traits. Co-location of ESTs, literature candidate genes and QTL for leaf and biomass traits recorded in three experiments within the Taylor lab; drought, CO_2 and short rotation coppice .QTL positions are shown by confidence intervals as determined by F_2 drop off, colour represents experiment: drought (pink), CO_2 (blue) and short rotation coppice (green), abbreviations are described in Table 3-4.Marker names are shown to the right of the linkage group (SSR markers – in bold brown and AFLPs – in italics) and cM distances to the left of the clear bar representing the genetic map. Solid black bar represents physical sequence of *Populus trichocarpa*, with base pair positions of markers and candidate genes to the left. Linkage group number is shown in Roman numerals above bar, whereas chromosome is represented by numbers. Hotspots are shown on the genetic map; solid fill red is a major hotspot, whereas red criss-crossed is a minor hotspot. Candidate genes within hotspots are shown to the right of the genetic map linked by brackets using gene model names from JGI, colour of candidate gene represent origin: microarray genes up-regulated in big leaves (red), down regulated in big leaves (green) and, literature searches (black), candidate genes are also represented on the physical sequence using colour key already stated, number on chromosome corresponds to gene model names to the right that are in brackets.

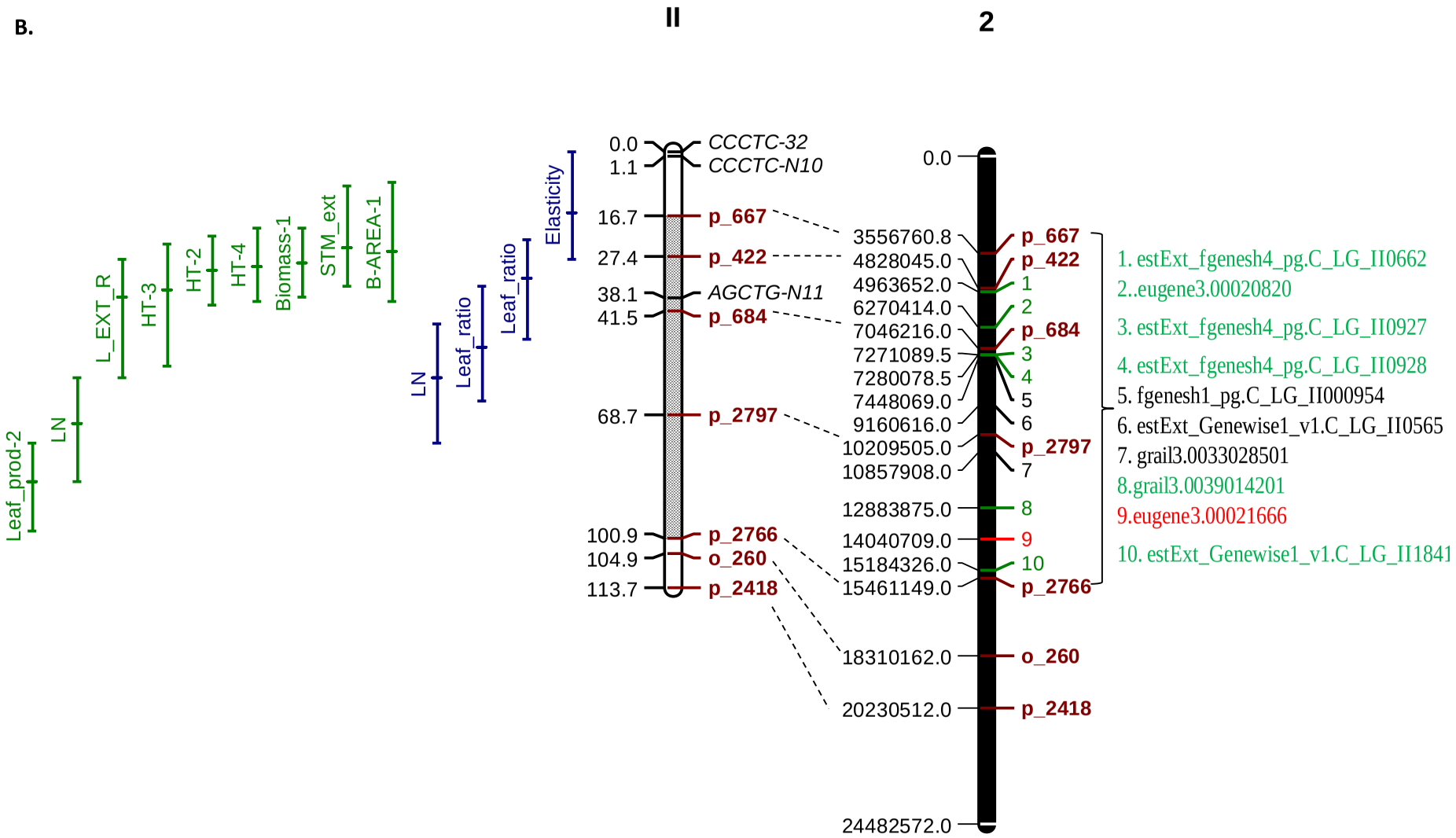
Table 3-4: List of QTL map trait abbreviations. Description of abbreviations used for leaf and biomass QTL maps.

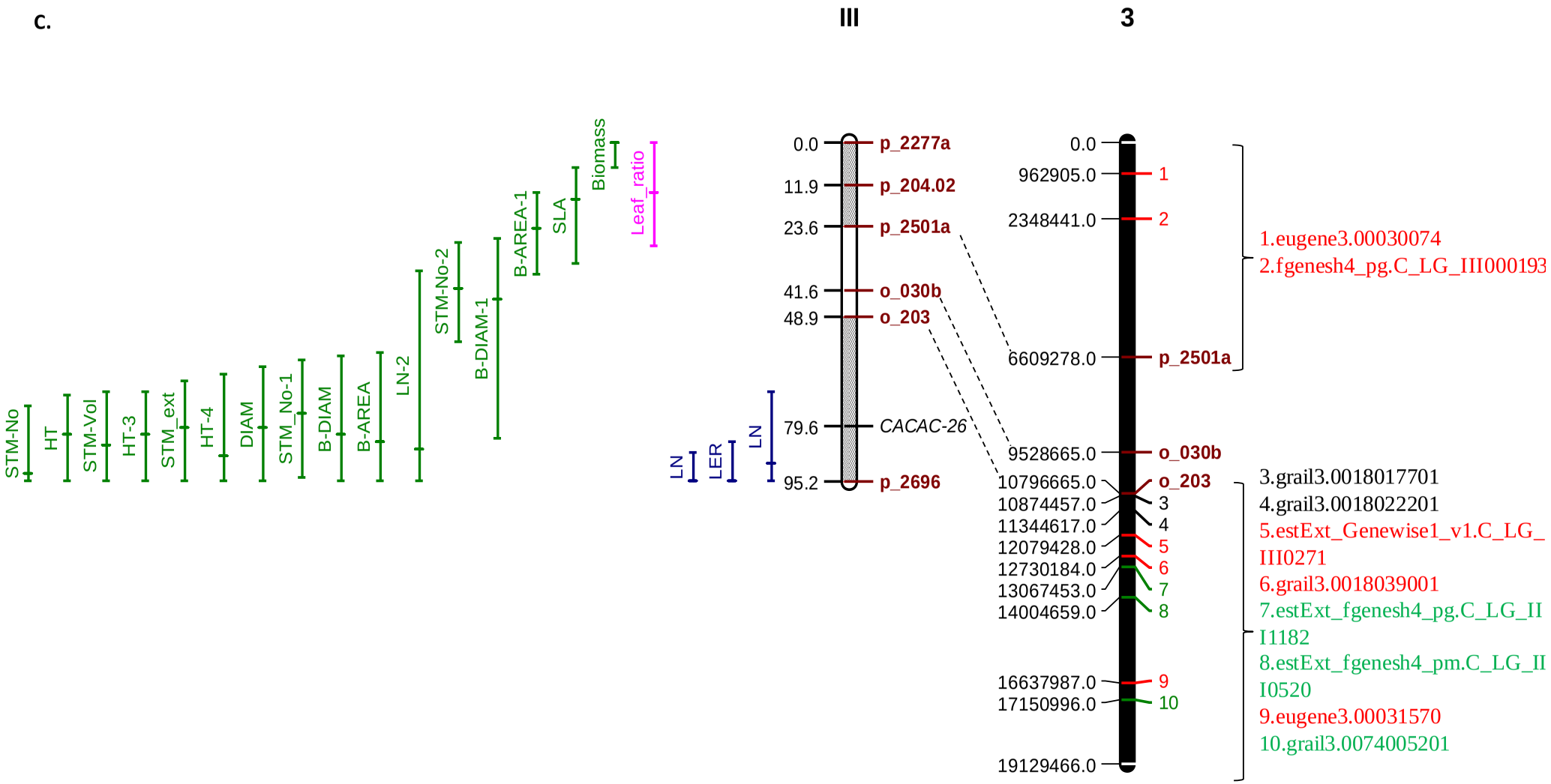
Abbreviation	**Description**
L_EXT_R	Leaf extension rate (mm day^{-1})
LA	Leaf area mm^2
LA-1	Leaf area 1st coppice (mm^2)
LA-2	Leaf area 2nd coppice (mm^2)
LDW	Leaf dry weight (mg)
LPR	Leaf production rate ($no.day^{-1}$)
Leaf_prod-1	Leaf production 1st coppice ($no.day^{-1}$)
Leaf_prod-2	Leaf production 2nd coppice ($no.day^{-1}$)
LER	Leaf expansion rate (mm day^{-1})
LL	Leaf length (mm)
LW	Leaf width (mm)
Leaf ratio	Leaf width to length ratio
LN	Number of leaves on leading stem
LN 2	Number of leaves on leading stem 2nd coppice
SLA	Specific Leaf area (mm^2g^{-1})
Elasticity	Leaf elasticity (% reversible extension per 10g load)
B-AREA	Basal stem area, 2nd coppice

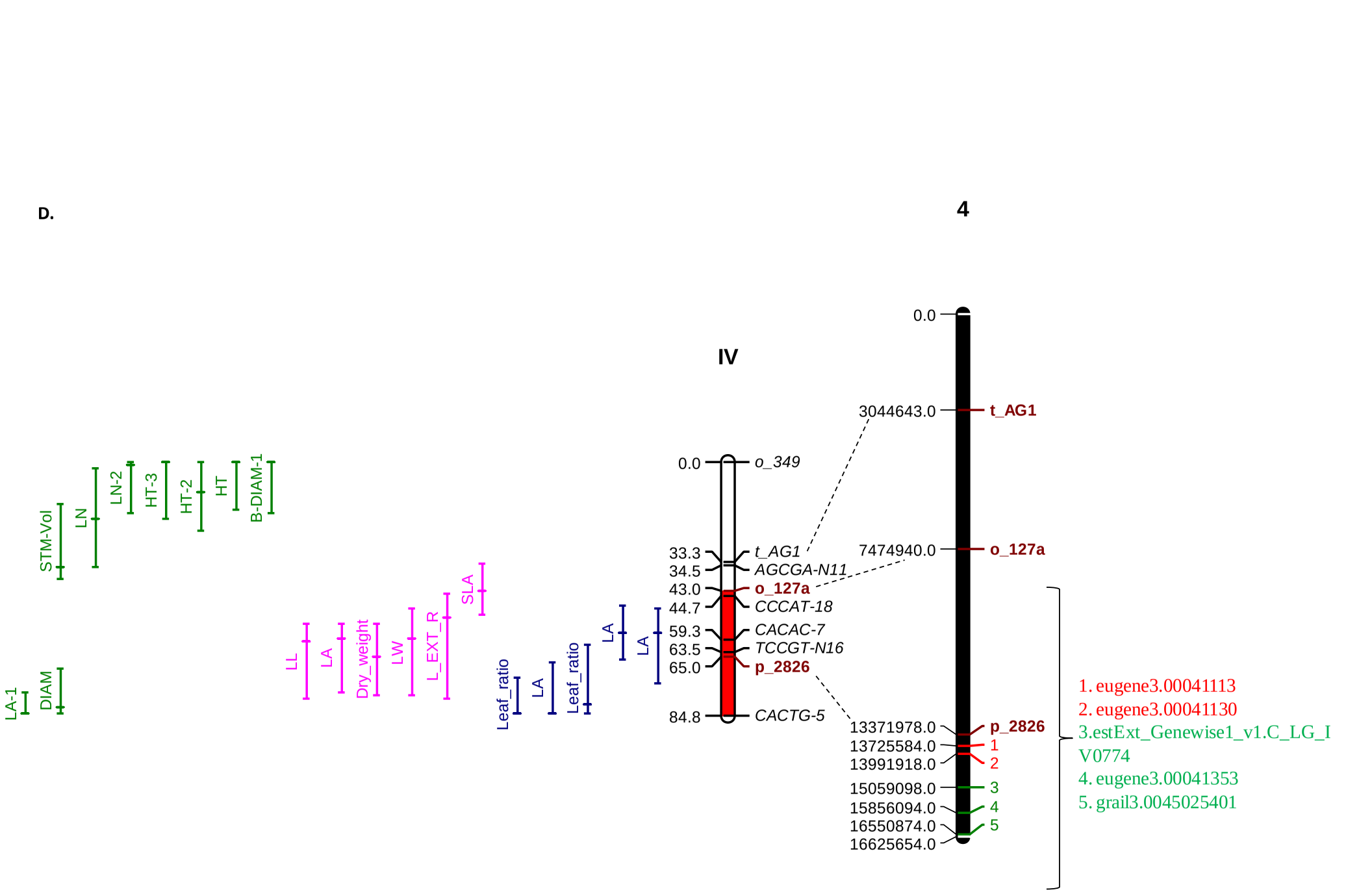
Table 3-4: Continued...

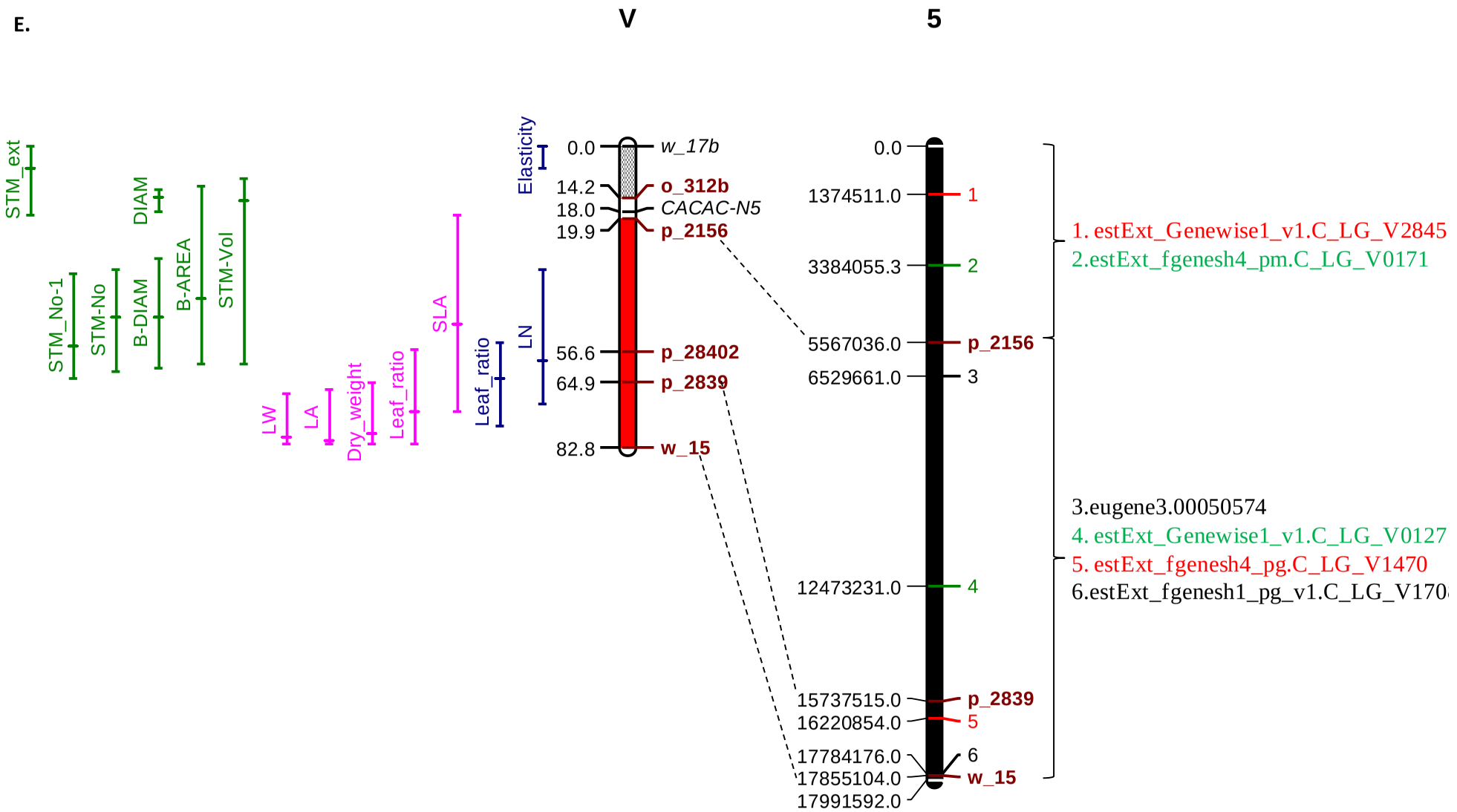
B-AREA-1	Basal stem area, 1st coppice
B-DIAM	Basal stem diameter, 2nd coppice (mm)
B-DIAM-1	Basal stem diameter, 1st coppice (mm)
DIAM	Stem diameter 1 metre above ground (mm)
Biomass	Biomass 2nd coppice
Biomass-1	Biomass 1 st coppice
Dry weight	Whole-tree dry mass (ODT $ha^{-1}y^{-1}$)
HT	Maximum Stem height (m)
HT-2	Maximum Stem height, 2nd coppice (m)
HT-3	Maximum Stem height, 3 years of growth (m)
HT-4	Maximum Stem height,4 years of growth (m)
STM_ext	Stem extension increment (mm)
STM_No-1	No. of stems on stool, 1st coppice
STM-No	No. of stems on stool
STM-No-2	No. of stems on stool, 2nd coppice
STM-Vol	Stem volume index
SYL	No. of sylleptic branches


A.
I
1
B-DIAM
B-AREA-1
LL
Dry_weight
LA
LER
LER
LER
LA
LL
Leaf_ratio
LA
0.0
0.3
3.8
11.2
38.2
59.9
72.9
117.7
132.9
136.7
140.7
142.4
148.4
163.6
165.9
172.5
174.7
CCCCT-39
CCCCT-38
p_634
p_2789
o_30a
p_2852
p_93
p_2889b
p_575
AGCTG-10
CCCCT-N12
AGCTG-13
CCCCT-N13
CACTG-21
p_2786a
p_2385
TCCGT-20
0.0
5467692.0
6435691.0
6802936.0
6926009.0
9404981.0
9764020.0
11239328.0
21053404.0
21575752.0
23285336.0
23968652.0
24066212.0
25815540.0
27041684.0
29313968.0
29676170.0
29769306.0
30395254.0
31401020.0
33771956.0
35571568.0
p_634
p_2789
1
2
3
p_2852
p_93
p_575
4
5
6
7
8
9
10
11
12
13
14
p_2385
1. estExt_fgenesh4_pg.C_LG_I0819
2. eugene3.00010810
3.estExt_Genewise1_v1.C_LG_I9453
4.estExt_Genewise1_v1.C_LG_I2426
5.estExt_fgenesh4_pg.C_LG_I2255
6. gw1.I.3059.1
7. fgenesh4_pg.C_LG_I002334
8. estExt_fgenesh4_pg.C_LG_I2498
9.estExt_Genewise1_v1.C_LG_I4040
10.fgenesh1_pg.C_LG_I002802
11. eugene3.00012637
12. grail3.0032009701
13.estExt_Genewise1_v1.C_LG_I4896
14. eugene3.00012781

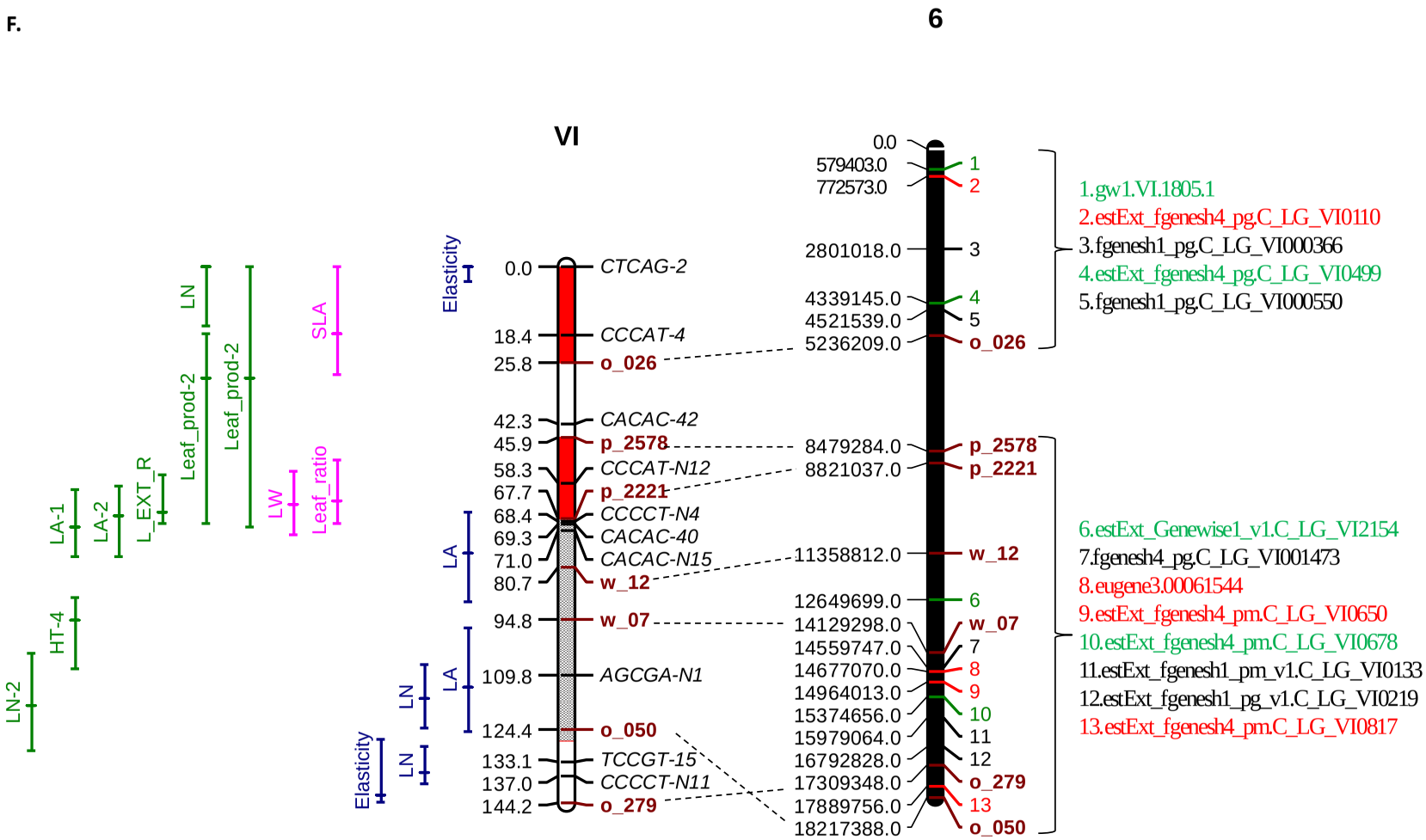

B.
II
2
Leaf_prod-2
LN
L_EXT_R
HT-3
HT-2
HT-4
Biomass-1
STM_ext
B-AREA-1
LN
Leaf_ratio
Leaf_ratio
Elasticity
0.0
1.1
16.7
27.4
38.1
41.5
68.7
100.9
104.9
113.7
CCCTC-32
CCCTC-N10
p_667
p_422
AGCTG-N11
p_684
p_2797
p_2766
o_260
p_2418
0.0
3556760.8
4828045.0
4963652.0
6270414.0
7046216.0
7271089.5
7280078.5
7448069.0
9160616.0
10209505.0
10857908.0
12883875.0
14040709.0
15184326.0
15461149.0
18310162.0
20230512.0
24482572.0
p_667
p_422
1
2
p_684
3
4
5
6
p_2797
7
8
9
10
p_2766
o_260
p_2418
1. estExt_fgenesh4_pg.C_LG_II0662
2..eugene3.00020820
3. estExt_fgenesh4_pg.C_LG_II0927
4. estExt_fgenesh4_pg.C_LG_II0928
5. fgenesh1_pg.C_LG_II000954
6. estExt_Genewise1_v1.C_LG_II0565
7. grail3.0033028501
8.grail3.0039014201
9.eugene3.00021666
10. estExt_Genewise1_v1.C_LG_II1841


c.
III
3
STM-No
HT
STM-Vol
HT-3
STM_ext
HT-4
DIAM
STM_No-1
B-DIAM
B-AREA
LN-2
STM-No-2
B-DIAM-1
B-AREA-1
SLA
Biomass
Leaf_ratio
LN
LER
LN
0.0
11.9
23.6
41.6
48.9
79.6
95.2
p_2277a
p_204.02
p_2501a
o_030b
o_203
CACAC-26
p_2696
0.0
962905.0
2348441.0
6609278.0
9528665.0
10796665.0
10874457.0
11344617.0
12079428.0
12730184.0
13067453.0
14004659.0
16637987.0
17150996.0
19129466.0
1
2
p_2501a
o_030b
o_203
3
4
5
6
7
8
9
10
1.eugene3.00030074
2.fgenesh4_pg.C_LG_III000193
3.grail3.0018017701
4.grail3.0018022201
5.estExt_Genewise1_v1.C_LG_III0271
6.grail3.0018039001
7.estExt_fgenesh4_pg.C_LG_III1182
8.estExt_fgenesh4_pm.C_LG_III0520
9.eugene3.00031570
10.grail3.0074005201

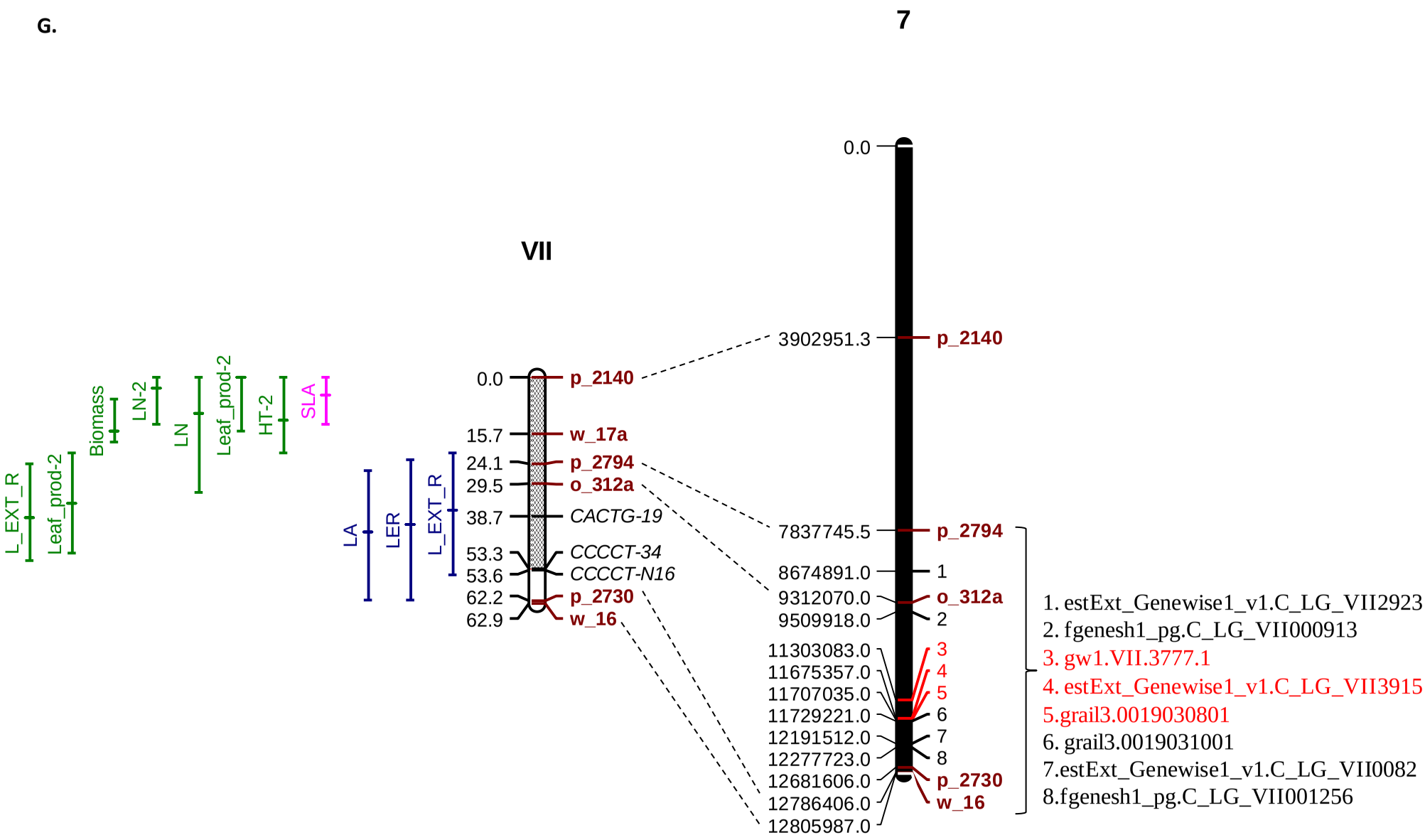

D.
4
IV
0.0
3044643.0
7474940.0
13371978.0
13725584.0
13991918.0
15059098.0
15856094.0
16550874.0
16625654.0
t_AG1
o_127a
p_2826
1
2
3
4
5
0.0
33.3
34.5
43.0
44.7
59.3
63.5
65.0
84.8
o_349
t_AG1
AGCGA-N11
o_127a
CCCAT-18
CACAC-7
TCCGT-N16
p_2826
CACTG-5
1. eugene3.00041113
2. eugene3.00041130
3.estExt_Genewise1_v1.C_LG_IV0774
4. eugene3.00041353
5. grail3.0045025401
LA-1
DIAM
STM-Vol
LN
LN-2
HT-3
HT-2
HT
B-DIAM-1
LL
LA
Dry_weight
LW
L_EXT_R
SLA
Leaf_ratio
LA
Leaf_ratio
LA
LA

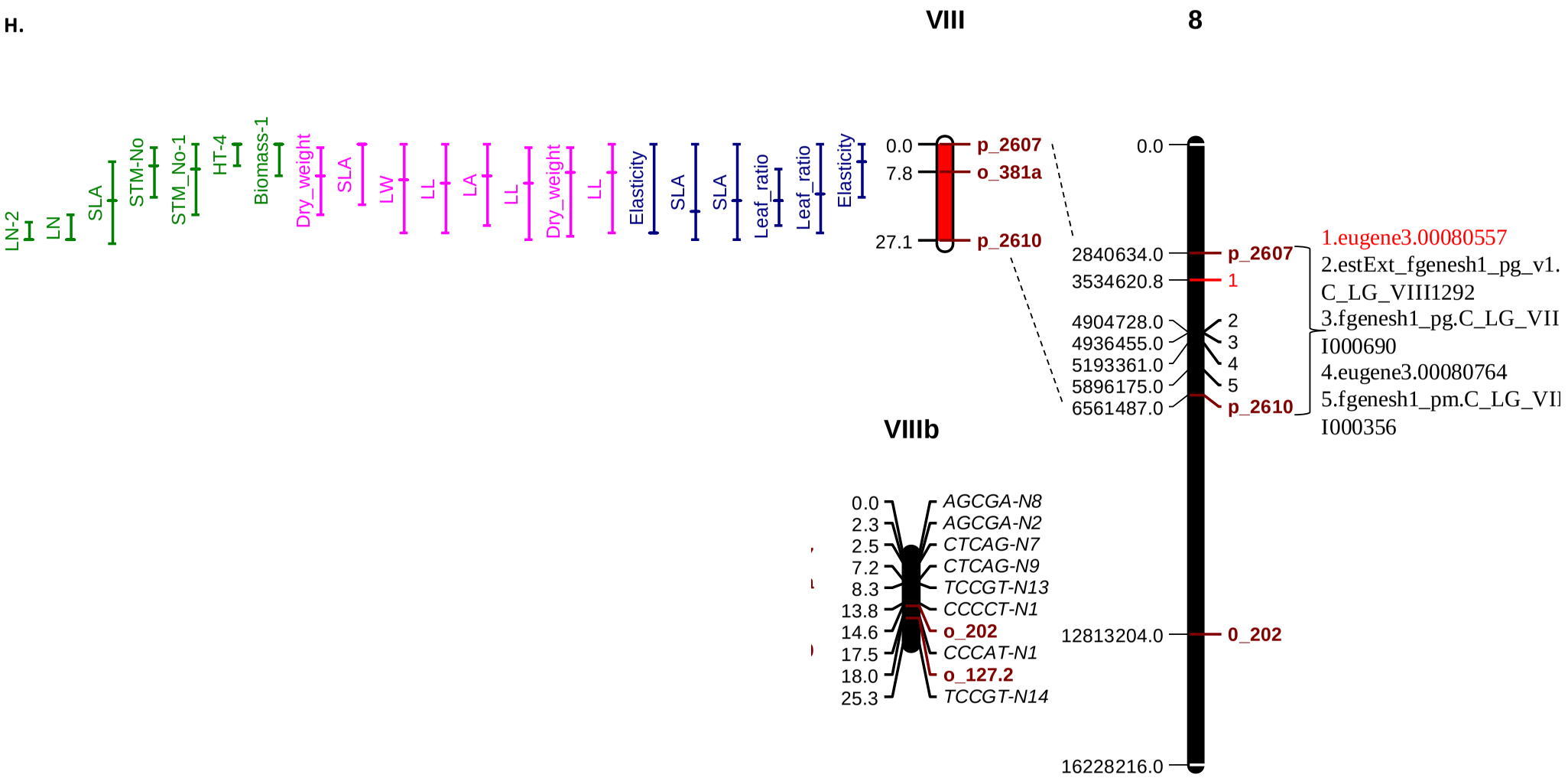

E.
V
5
STM_ext
STM_No-1
STM-No
B-DIAM
DIAM
B-AREA
STM-Vol
LW
LA
Dry_weight
Leaf_ratio
SLA
Leaf_ratio
Elasticity
LN
0.0
w_17b
14.2
o_312b
18.0
CACAC-N5
19.9
p_2156
56.6
p_28402
64.9
p_2839
82.8
w_15
0.0
1374511.0
1
3384055.3
2
5567036.0
p_2156
6529661.0
3
12473231.0
4
15737515.0
p_2839
16220854.0
5
17784176.0
6
17855104.0
w_15
17991592.0
1. estExt_Genewise1_v1.C_LG_V2845
2.estExt_fgenesh4_pm.C_LG_V0171
3.eugene3.00050574
4. estExt_Genewise1_v1.C_LG_V0127
5. estExt_fgenesh4_pg.C_LG_V1470
6.estExt_fgenesh1_pg_v1.C_LG_V170

F.


VI
6
LA-1
LA-2
L_EXT_R
LN
Leaf_prod-2
Leaf_prod-2
LW
Leaf_ratio
SLA
HT-4
LN-2
Elasticity
LA
LN
LA
LN
Elasticity
LN
0.0 CTCAG-2
18.4 CCCAT-4
25.8 o_026
42.3 CACAC-42
45.9 p_2578
58.3 CCCAT-N12
67.7 p_2221
68.4 CCCCT-N4
69.3 CACAC-40
71.0 CACAC-N15
80.7 w_12
94.8 w_07
109.8 AGCGA-N1
124.4 o_050
133.1 TCCGT-15
137.0 CCCCT-N11
144.2 o_279
0.0 1
579403.0 2
772573.0
2801018.0 3
4339145.0 4
4521539.0 5
5236209.0 o_026
8479284.0 p_2578
8821037.0 p_2221
11358812.0 w_12
12649699.0 6
14129298.0 w_07
14559747.0 7
14677070.0 8
14964013.0 9
15374656.0 10
15979064.0 11
16792828.0 12
17309348.0 o_279
17889756.0 13
18217388.0 o_050
1.gw1.VI.1805.1
2.estExt_fgenesh4_pg.C_LG_VI0110
3.fgenesh1_pg.C_LG_VI000366
4.estExt_fgenesh4_pg.C_LG_VI0499
5.fgenesh1_pg.C_LG_VI000550
6.estExt_Genewise1_v1.C_LG_VI2154
7.fgenesh4_pg.C_LG_VI001473
8.eugene3.00061544
9.estExt_fgenesh4_pm.C_LG_VI0650
10.estExt_fgenesh4_pm.C_LG_VI0678
11.estExt_fgenesh1_pm_v1.C_LG_VI0133
12.estExt_fgenesh1_pg_v1.C_LG_VI0219
13.estExt_fgenesh4_pm.C_LG_VI0817

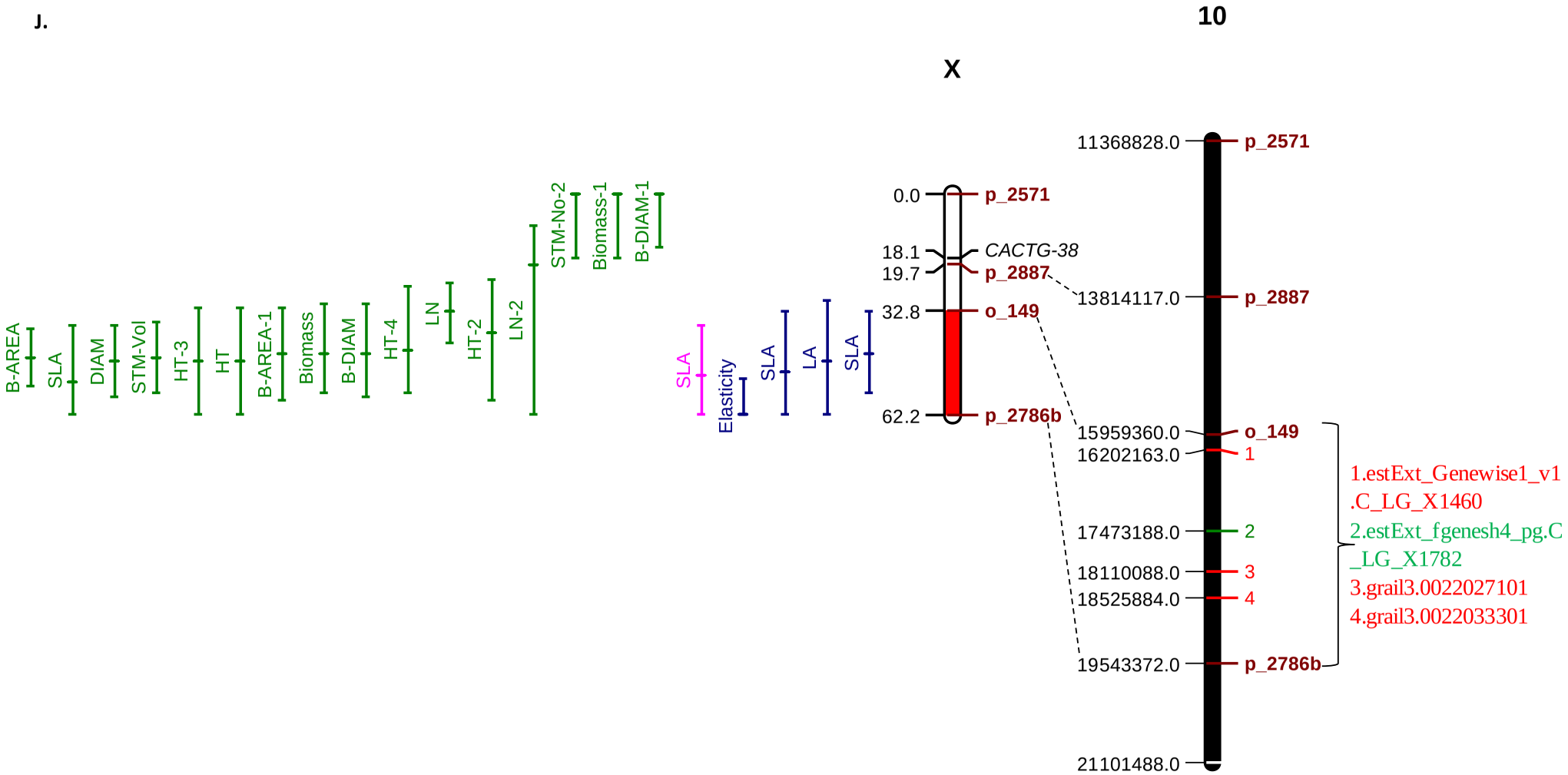
G.


7
VII
0.0
3902951.3
p_2140
7837745.5
p_2794
8674891.0
1
9312070.0
o_312a
9509918.0
2
11303083.0
3
11675357.0
4
11707035.0
5
11729221.0
6
12191512.0
7
12277723.0
8
12681606.0
p_2730
12786406.0
w_16
12805987.0
0.0
p_2140
15.7
w_17a
24.1
p_2794
29.5
o_312a
38.7
CACTG-19
53.3
CCCCT-34
53.6
CCCCT-N16
62.2
p_2730
62.9
w_16
L_EXT_R
Leaf_prod-2
Biomass
LN-2
LN
Leaf_prod-2
HT-2
SLA
LA
LER
L_EXT_R
1. estExt_Genewise1_v1.C_LG_VII2923
2. fgenesh1_pg.C_LG_VII000913
3. gw1.VII.3777.1
4. estExt_Genewise1_v1.C_LG_VII3915
5.grail3.0019030801
6. grail3.0019031001
7.estExt_Genewise1_v1.C_LG_VII0082
8.fgenesh1_pg.C_LG_VII001256

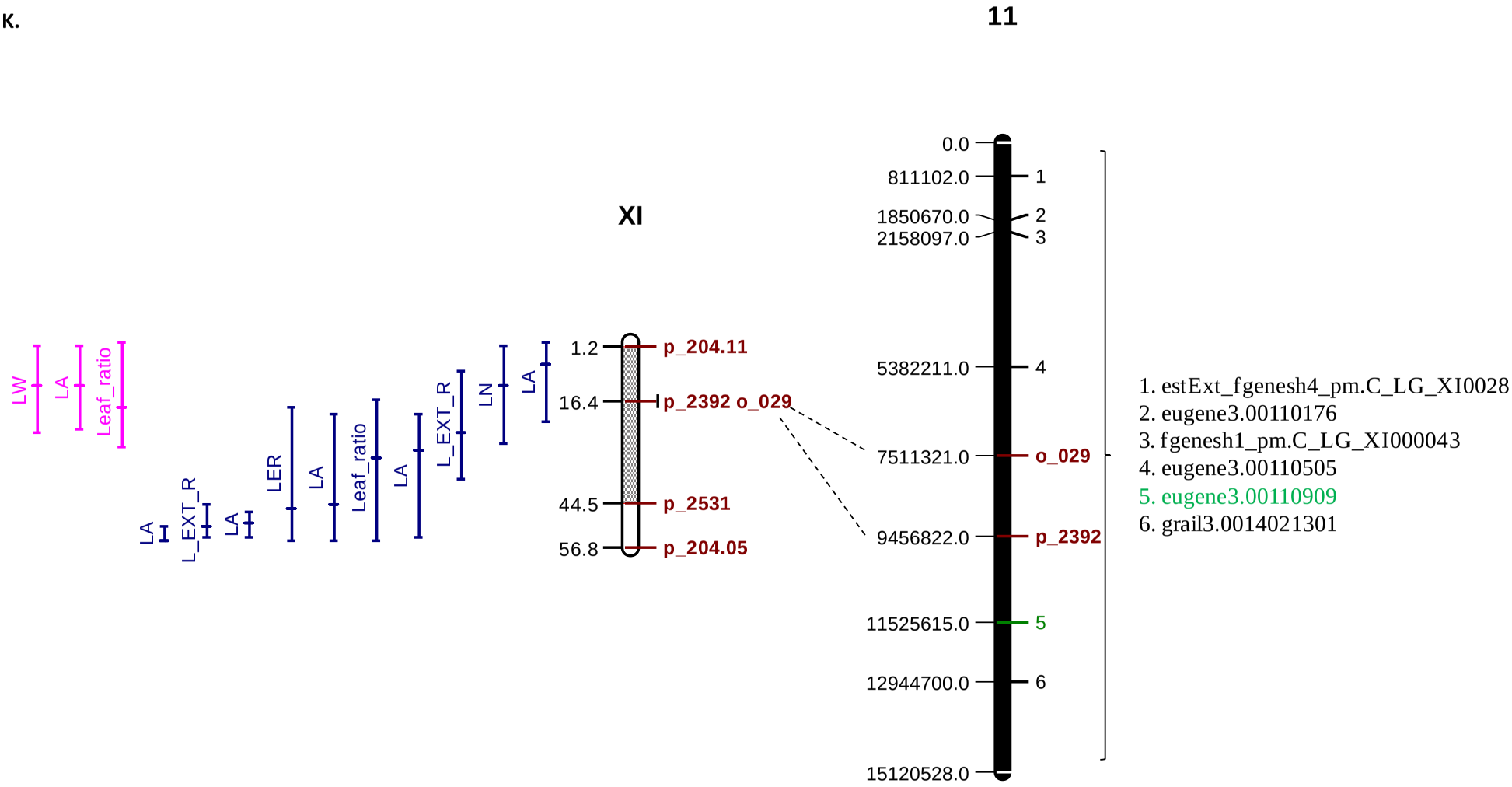

H.
VIII
8
LN-2
LN
SLA
STM-No
STM_No-1
HT-4
Biomass-1
Dry_weight
SLA
LW
LL
LA
LL
Dry_weight
LL
Elasticity
SLA
SLA
Leaf_ratio
Leaf_ratio
Elasticity
0.0
7.8
27.1
p_2607
o_381a
p_2610
0.0
2840634.0
3534620.8
4904728.0
4936455.0
5193361.0
5896175.0
6561487.0
12813204.0
16228216.0
p_2607
1
2
3
4
5
p_2610
0_202
1.eugene3.00080557
2.estExt_fgenesh1_pg_v1.
C_LG_VIII1292
3.fgenesh1_pg.C_LG_VII
I000690
4.eugene3.00080764
5.fgenesh1_pm.C_LG_VII
I000356
VIIIb
0.0
2.3
2.5
7.2
8.3
13.8
14.6
17.5
18.0
25.3
AGCGA-N8
AGCGA-N2
CTCAG-N7
CTCAG-N9
TCCGT-N13
CCCCT-N1
o_202
CCCAT-N1
o_127.2
TCCGT-N14

9
0.0
8388569.0
8743715.0
8798608.0
8993060.0
11883027.0
12525049.0
p_2522
1
o_023
2
3
1.estExt_Genewise1_v1.C_LG_IX4231
2. eugene3.00091367
3.estExt_fgenesh1_pg_v1.C_LG_IX14
96
IX
0.0
6.9
9.3
p_2522
o_023
AGCGA-22
L_EXT_R)
SLA
LA
LA
Leaf_ratio
LW
Dry_weight
LA
B-AREA
B-AREA
B-DIAM
Biomass
LA-2
Leaf_prod-2
LN-2
STM_No-1
STM-Vol

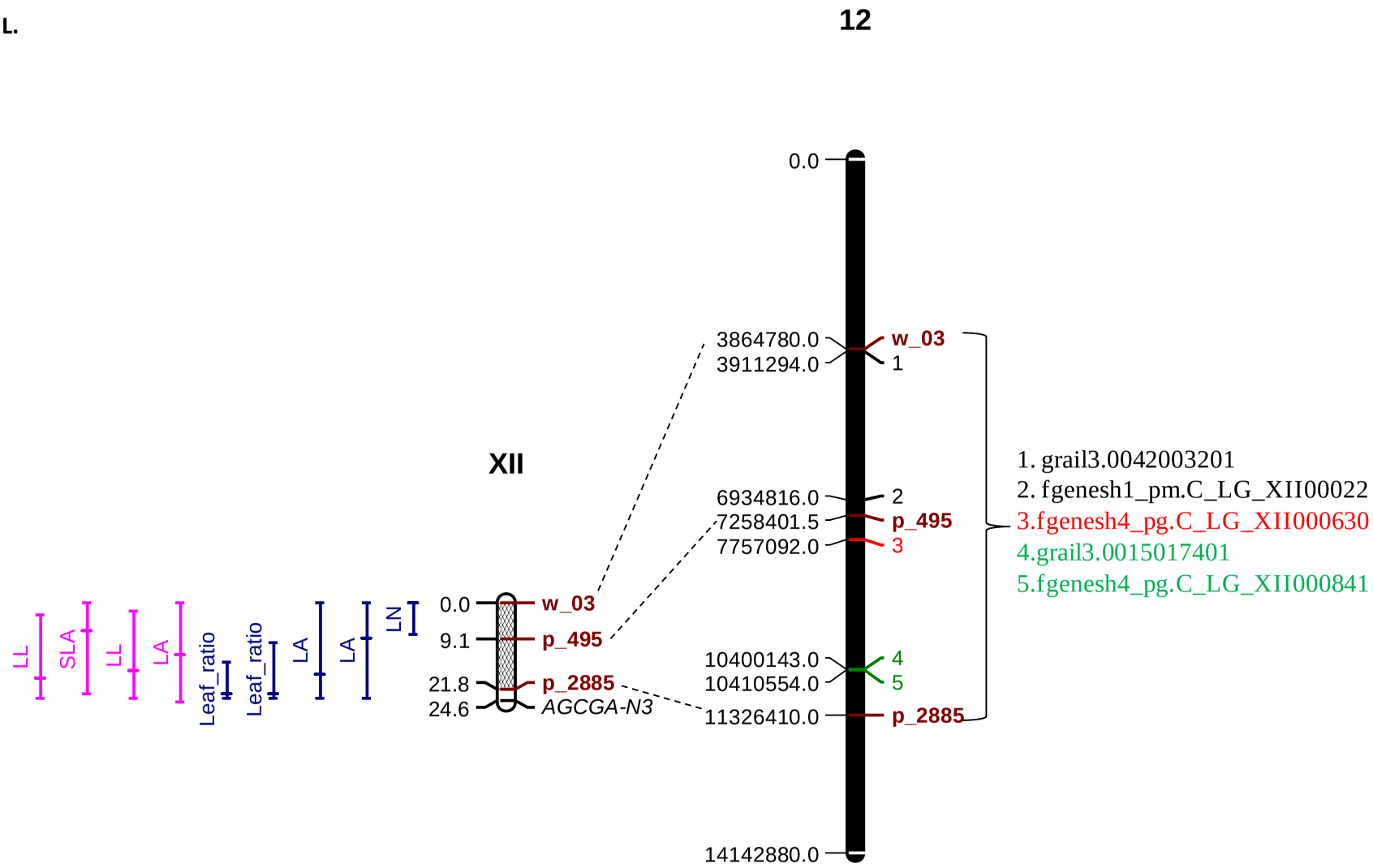
J.


B-AREA
SLA
DIAM
STM-Vol
HT-3
HT
B-AREA-1
Biomass
B-DIAM
HT-4
LN
HT-2
LN-2
STM-No-2
Biomass-1
B-DIAM-1
SLA
Elasticity
SLA
LA
SLA
X
0.0
18.1
19.7
32.8
62.2
p_2571
CACTG-38
p_2887
o_149
p_2786b
10
11368828.0
13814117.0
15959360.0
16202163.0
17473188.0
18110088.0
18525884.0
19543372.0
21101488.0
p_2571
p_2887
o_149
1
2
3
4
p_2786b
1.estExt_Genewise1_v1
.C_LG_X1460
2.estExt_fgenesh4_pg.C
_LG_X1782
3.grail3.0022027101
4.grail3.0022033301

K.


11
XI
LW
LA
Leaf_ratio
LA
L_EXT_R
LA
LER
LA
Leaf_ratio
LA
L_EXT_R
LN
LA
1.2
16.4
44.5
56.8
p_204.11
p_2392 o_029
p_2531
p_204.05
0.0
811102.0
1850670.0
2158097.0
5382211.0
7511321.0
9456822.0
11525615.0
12944700.0
15120528.0
1
2
3
4
o_029
p_2392
5
6
1. estExt_fgenesh4_pm.C_LG_XI0028
2. eugene3.00110176
3. fgenesh1_pm.C_LG_XI000043
4. eugene3.00110505
5. eugene3.00110909
6. grail3.0014021301

L.

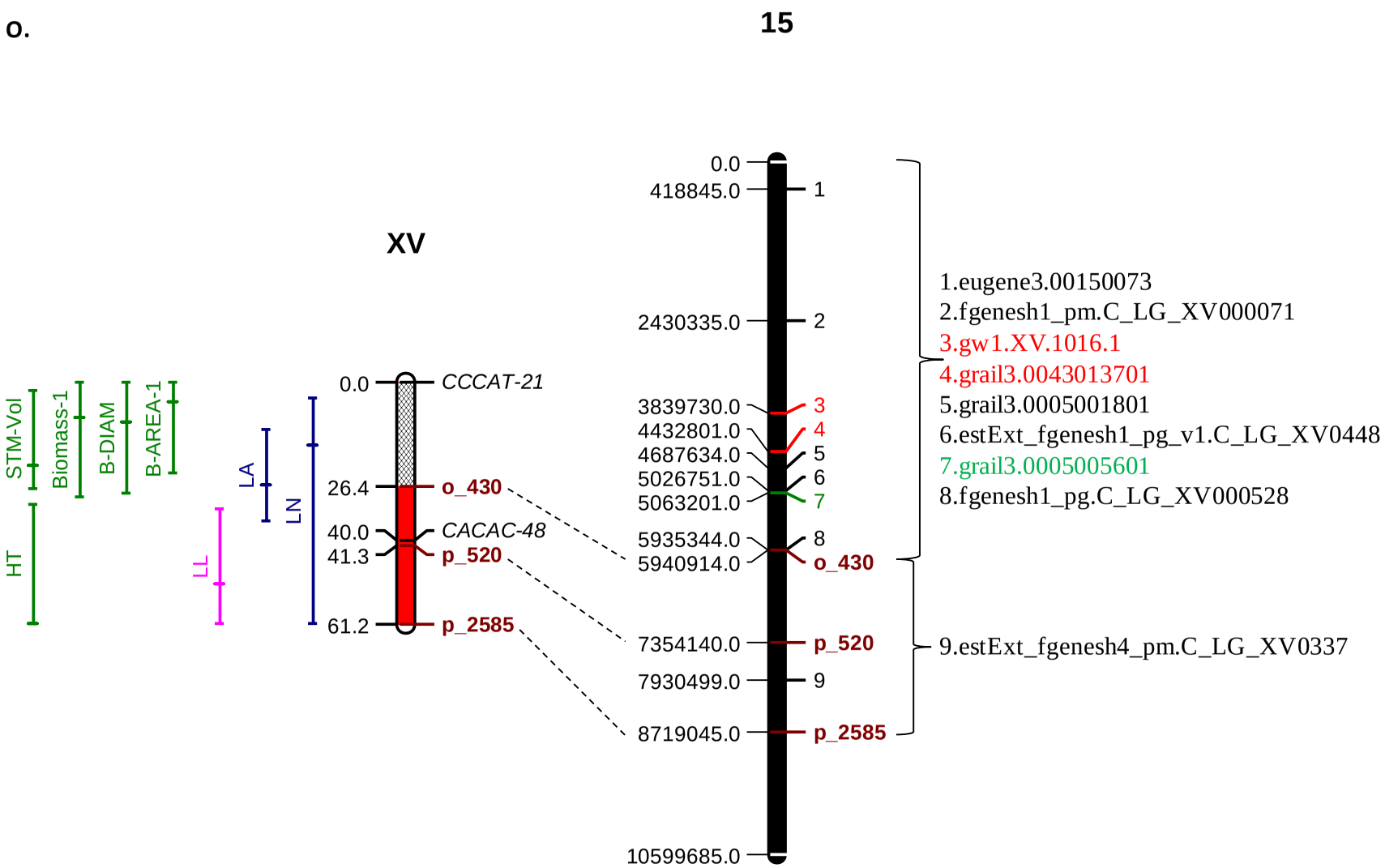

12
XII
0.0
3864780.0
3911294.0
6934816.0
7258401.5
7757092.0
10400143.0
10410554.0
11326410.0
14142880.0
w_03
1
2
p_495
3
4
5
p_2885
1. grail3.0042003201
2. fgenesh1_pm.C_LG_XII00022
3.fgenesh4_pg.C_LG_XII000630
4.grail3.0015017401
5.fgenesh4_pg.C_LG_XII000841
0.0
9.1
21.8
24.6
w_03
p_495
p_2885
AGCGA-N3
LL
SLA
LL
LA
Leaf_ratio
Leaf_ratio
LA
LA
LN

M.
LA-1
HT
STM_No
STM_No-1
HT-3
SLA
B-DIAM
B-AREA-1
B-DIAM-1
Biomass-1
HT-4
DIAM
Dry_weight
LW
LA
SLA
Leaf_ratio
Elasticity
0.0
16.3
28.6
30.4
35.0
38.2
40.3
41.5
42.6
42.9
51.8
54.8
57.5
58.0
62.6
62.7
64.4
69.6
71.7
71.9
81.1
85.3
89.0
XIII
AGCTG-N13
p_14
CACAC-N25
CACAC-49
CCCCT-N3
o_016
CACAC-51
CACAC-50
AGCGA-11
AGCGA-N5
p_2658
CCCTC-13
CACAC-33
CCCTC-12
AGCGA-31
CTCAG-31
p_649
p_2420
o_417
CCCTT-N2
p_2277b
CCCTT-23
CACAC-N1
13
0.0
669135.0
2501123.0
2837724.0
3051497.0
4443546.0
6732505.0
10038122.0
13101108.0
p_14
o_016
1
2
p_2658
3
p_649
1.gw1.XIII.1514.1
2. eugene3.00130435
3.eugene3.00130769

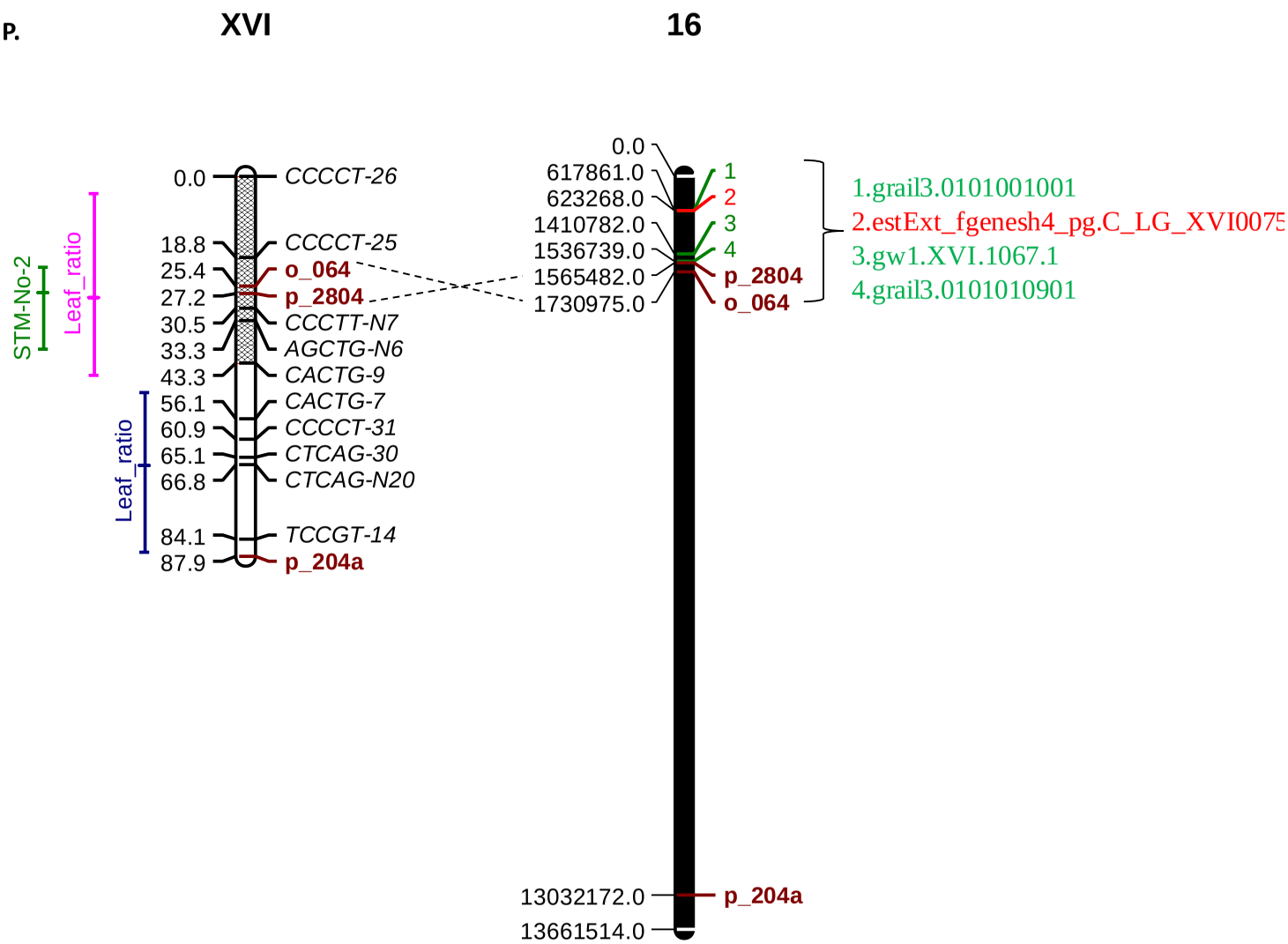
14
1.eugene3.00140027
2.eugene3.00140037
3.fgenesh1_pg.C_LG_XIV000039
4.eugene3.00140349
5.eugene3.00140364
6.fgenesh1_kg.C_LG_XIV
7.estExt_Genewise1_v1.C_LG_XI
V2304
8.eugene3.00140663
9.eugene3.00140920
10.estExt_Genewise1_v1.C_LG_
XIV3201
11.eugene3.00141061
p_571
p_2515
0.0
194589.0
254128.0
275371.0
2823895.0
2930186.0
3372932.0
5247859.0
5267911.0
6689090.0
7151030.5
7578689.0
8489708.0
12174321.0
14699530.0
XIV
CACTG-25
CCCTC-4
p_2515
CACAC-1
p_571
0.0
5.6
6.5
8.1
28.6
LA
LER
L_EXT_R)
L_EXT_R
Dry_weight
Leaf_ratio
LW
Leaf_ratio
B-AREA
Leaf_prod-2
SLA
STM-No-2
STM-Vol
B-DIAM
STM_No-1
B-AREA-1
B-DIAM-1

N.

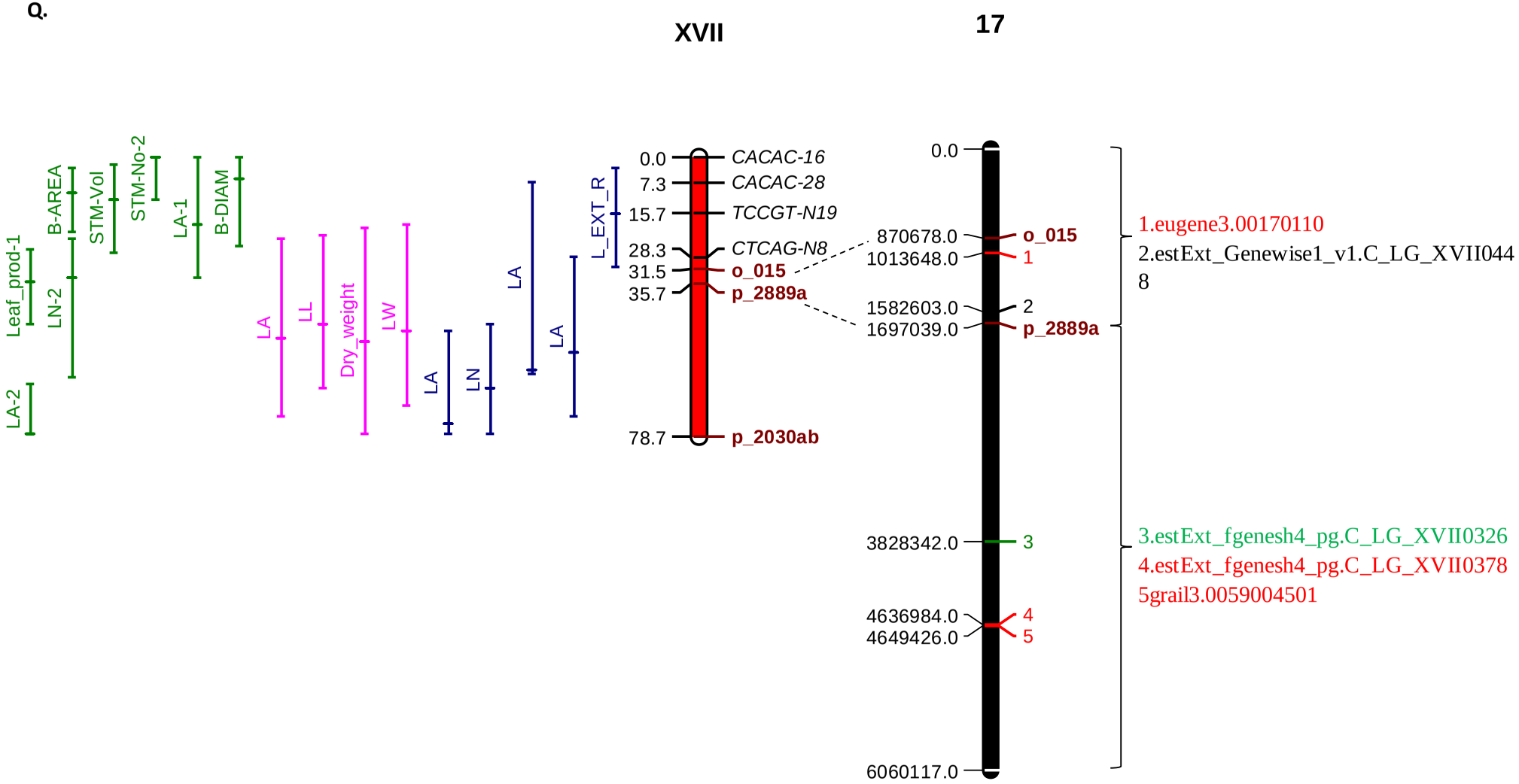
o.


15
XV
STM-Vol
Biomass-1
B-DIAM
B-AREA-1
HT
LL
LA
LN
0.0
26.4
40.0
41.3
61.2
CCCAT-21
o_430
CACAC-48
p_520
p_2585
0.0
418845.0
2430335.0
3839730.0
4432801.0
4687634.0
5026751.0
5063201.0
5935344.0
5940914.0
7354140.0
7930499.0
8719045.0
10599685.0
1
2
3
4
5
6
7
8
o_430
p_520
9
p_2585
1.eugene3.00150073
2.fgenesh1_pm.C_LG_XV000071
3.gw1.XV.1016.1
4.grail3.0043013701
5.grail3.0005001801
6.estExt_fgenesh1_pg_v1.C_LG_XV0448
7.grail3.0005005601
8.fgenesh1_pg.C_LG_XV000528
9.estExt_fgenesh4_pm.C_LG_XV0337

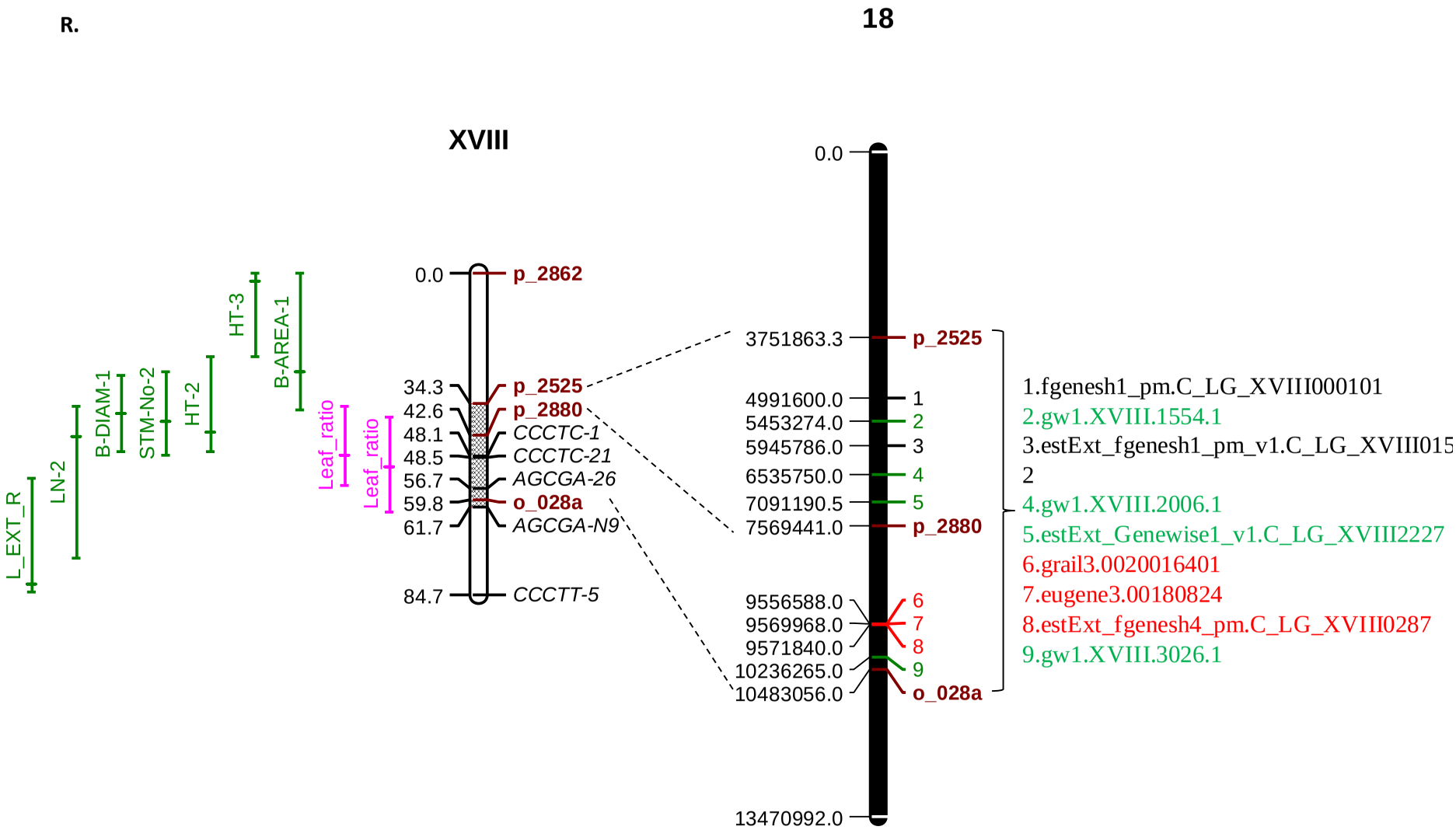
P.


XVI
16
STM-No-2
Leaf_ratio
Leaf_ratio
0.0 CCCCT-26
18.8 CCCCT-25
25.4 o_064
27.2 p_2804
30.5 CCCTT-N7
33.3 AGCTG-N6
43.3 CACTG-9
56.1 CACTG-7
60.9 CCCCT-31
65.1 CTCAG-30
66.8 CTCAG-N20
84.1 TCCGT-14
87.9 p_204a
0.0
617861.0 1
623268.0 2
1410782.0 3
1536739.0 4
1565482.0 p_2804
1730975.0 o_064
13032172.0 p_204a
13661514.0
1.grail3.0101001001
2.estExt_fgenesh4_pg.C_LG_XVI0075
3.gw1.XVI.1067.1
4.grail3.0101010901

Q.


XVII
17
Leaf_prod-1
LN-2
LA-2
B-AREA
STM-Vol
STM-No-2
LA-1
B-DIAM
LA
LL
Dry_weight
LW
LA
LN
LA
LA
L_EXT_R
0.0
7.3
15.7
28.3
31.5
35.7
78.7
CACAC-16
CACAC-28
TCCGT-N19
CTCAG-N8
o_015
p_2889a
p_2030ab
0.0
870678.0
1013648.0
1582603.0
1697039.0
3828342.0
4636984.0
4649426.0
6060117.0
o_015
1
2
p_2889a
3
4
5
1.eugene3.00170110
2.estExt_Genewise1_v1.C_LG_XVII0448
3.estExt_fgenesh4_pg.C_LG_XVII0326
4.estExt_fgenesh4_pg.C_LG_XVII0378
5grail3.0059004501

R.


18
XVIII
L_EXT_R
LN-2
B-DIAM-1
STM-No-2
HT-2
HT-3
B-AREA-1
Leaf_ratio
Leaf_ratio
0.0
34.3
42.6
48.1
48.5
56.7
59.8
61.7
84.7
p_2862
p_2525
p_2880
CCCTC-1
CCCTC-21
AGCGA-26
o_028a
AGCGA-N9
CCCTT-5
0.0
3751863.3
4991600.0
5453274.0
5945786.0
6535750.0
7091190.5
7569441.0
9556588.0
9569968.0
9571840.0
10236265.0
10483056.0
13470992.0
p_2525
1
2
3
4
5
p_2880
6
7
8
9
o_028a
1.fgenesh1_pm.C_LG_XVIII000101
2.gw1.XVIII.1554.1
3.estExt_fgenesh1_pm_v1.C_LG_XVIII015
2
4.gw1.XVIII.2006.1
5.estExt_Genewise1_v1.C_LG_XVIII2227
6.grail3.0020016401
7.eugene3.00180824
8.estExt_fgenesh4_pm.C_LG_XVIII0287
9.gw1.XVIII.3026.1

S.

XIX **19**

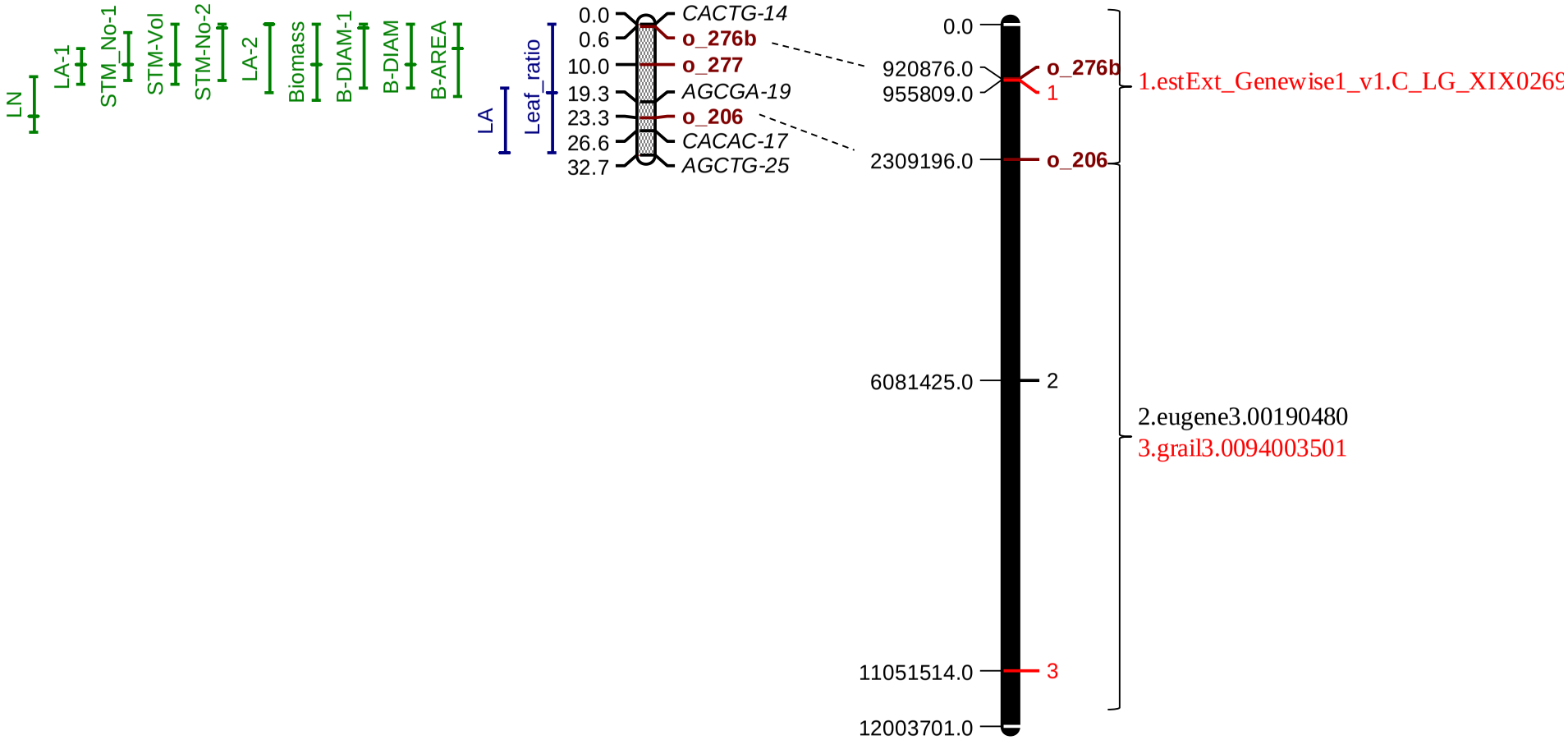
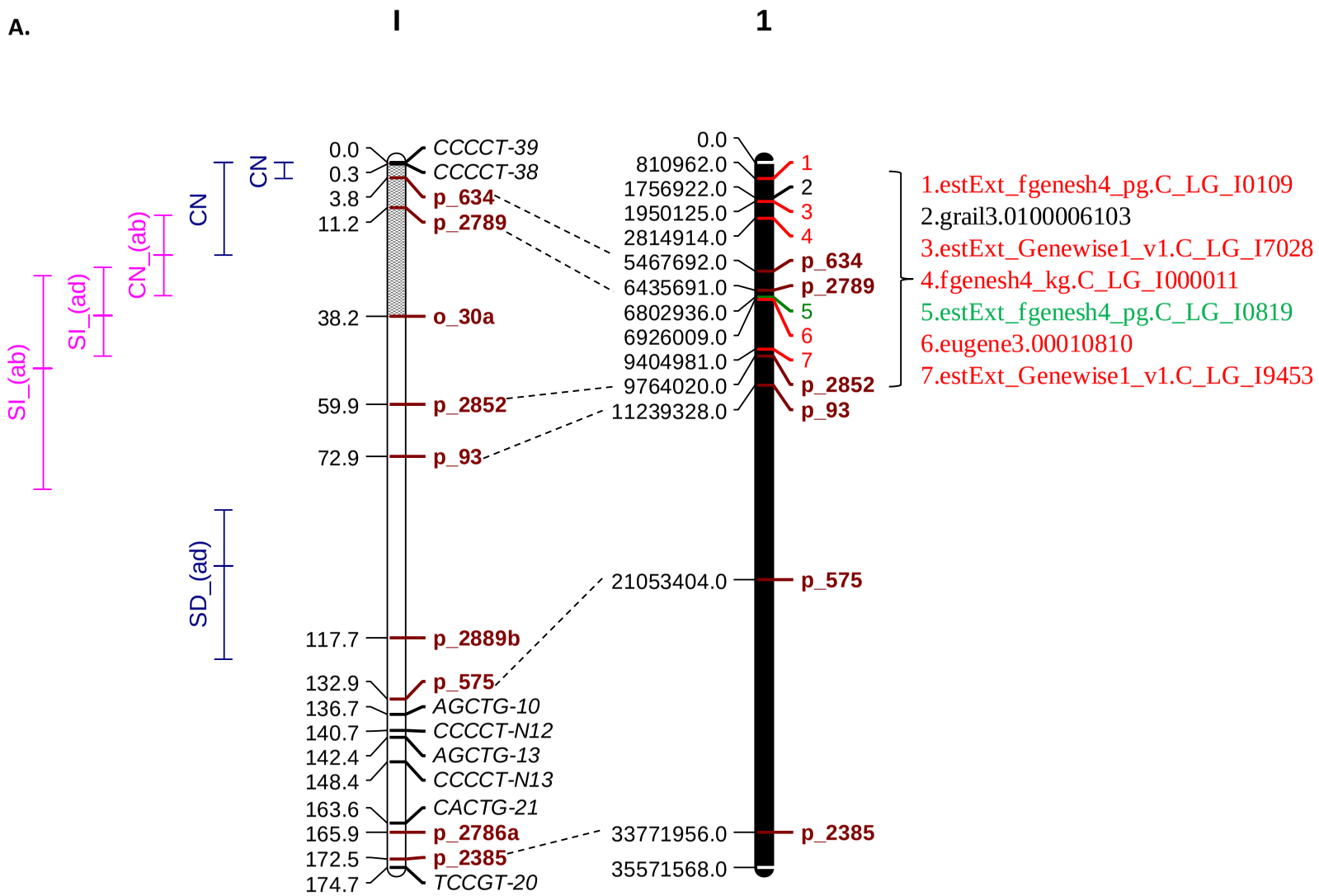

LN
LA-1
STM_No-1
STM-Vol
STM-No-2
LA-2
Biomass
B-DIAM-1
B-DIAM
B-AREA
LA
Leaf_ratio
0.0
0.6
10.0
19.3
23.3
26.6
32.7
CACTG-14
o_276b
o_277
AGCGA-19
o_206
CACAC-17
AGCTG-25
0.0
920876.0
955809.0
2309196.0
6081425.0
11051514.0
12003701.0
o_276b
1
o_206
2
3
1.estExt_Genewise1_v1.C_LG_XIX026
2.eugene3.00190480
3.grail3.0094003501

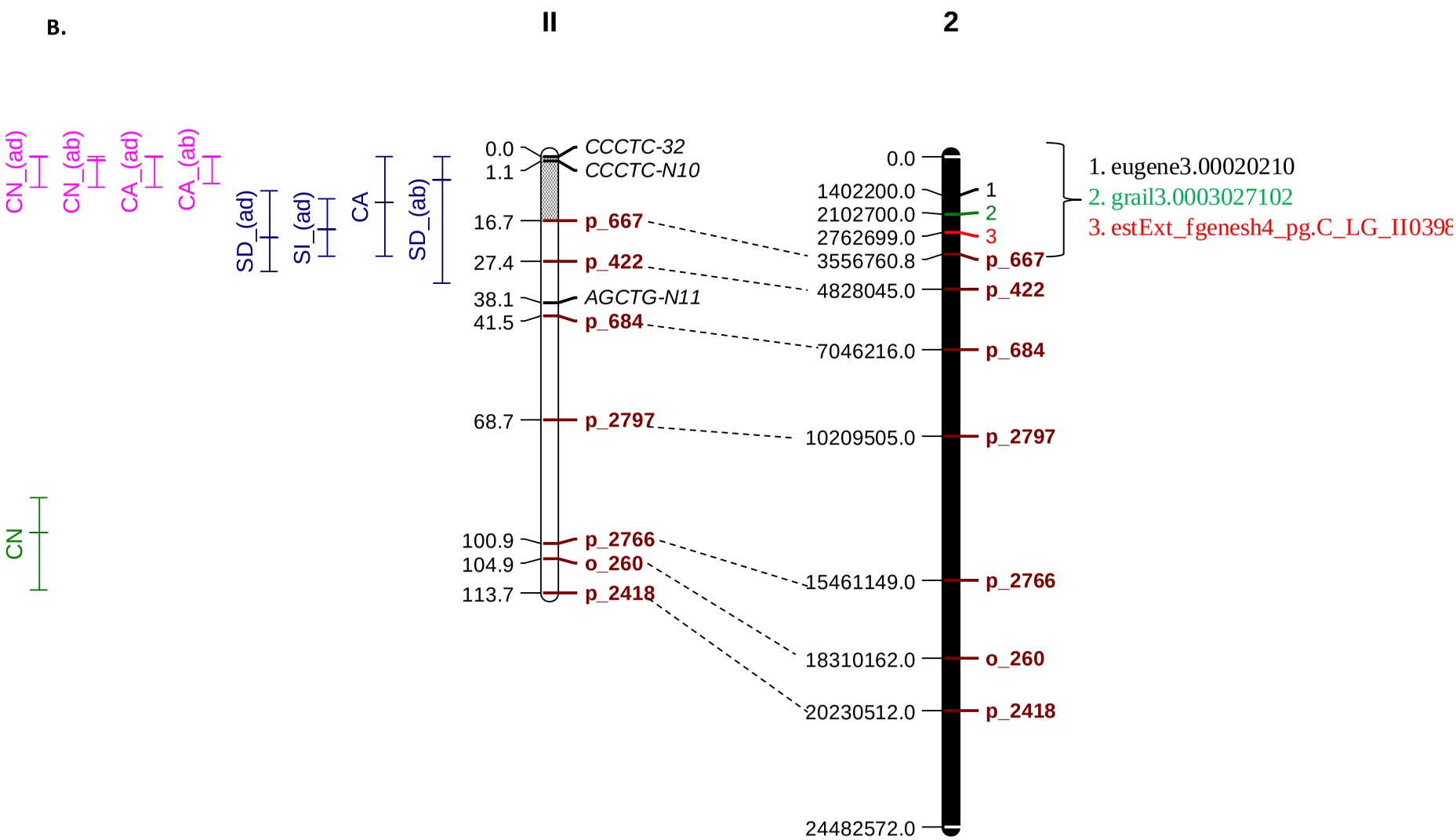
Figure 3-6: QTL map of cell traits.Co-location of ESTs, literature candidate genes and QTL for cell traits recorded in three experiments within the Taylor lab; drought, CO_2 and short rotation coppice .QTL positions are shown by confidence intervals as determined by F_2 drop off, colour represents experiment: drought (pink), CO_2 (blue) and short rotation coppice (green), abbreviations are described in Table 3-5. Marker names are shown to the right of the linkage group (SSR markers – in bold brown and AFLPs – in italics) and cM distances to the left of the clear bar representing the genetic map. Solid black bar represent physical sequence of *Populus trichocarpa*, base pairs are shown to the left Linkage group number is shown in Roman numeral above bar and chromosome number above physical map. Hotspots are shown on the genetic map; solid fill red is a major hotspot, whereas red criss-crossed is a minor hotspot.Candidate genes within hotspots are shown to the right of the genetic map linked by brackets using gene model names from JGI, colour of candidate gene represent origin: microarray genes up-regulated in big leaves (red), down regulated in big leaves (green) and, literature searches (black), candidate genes are also represented on the physical sequence using colour key already stated, in order show in brackets.

Table 3-5: List of QTL map trait abbreviations. Description of abbreviations used for cell traits QTL map.

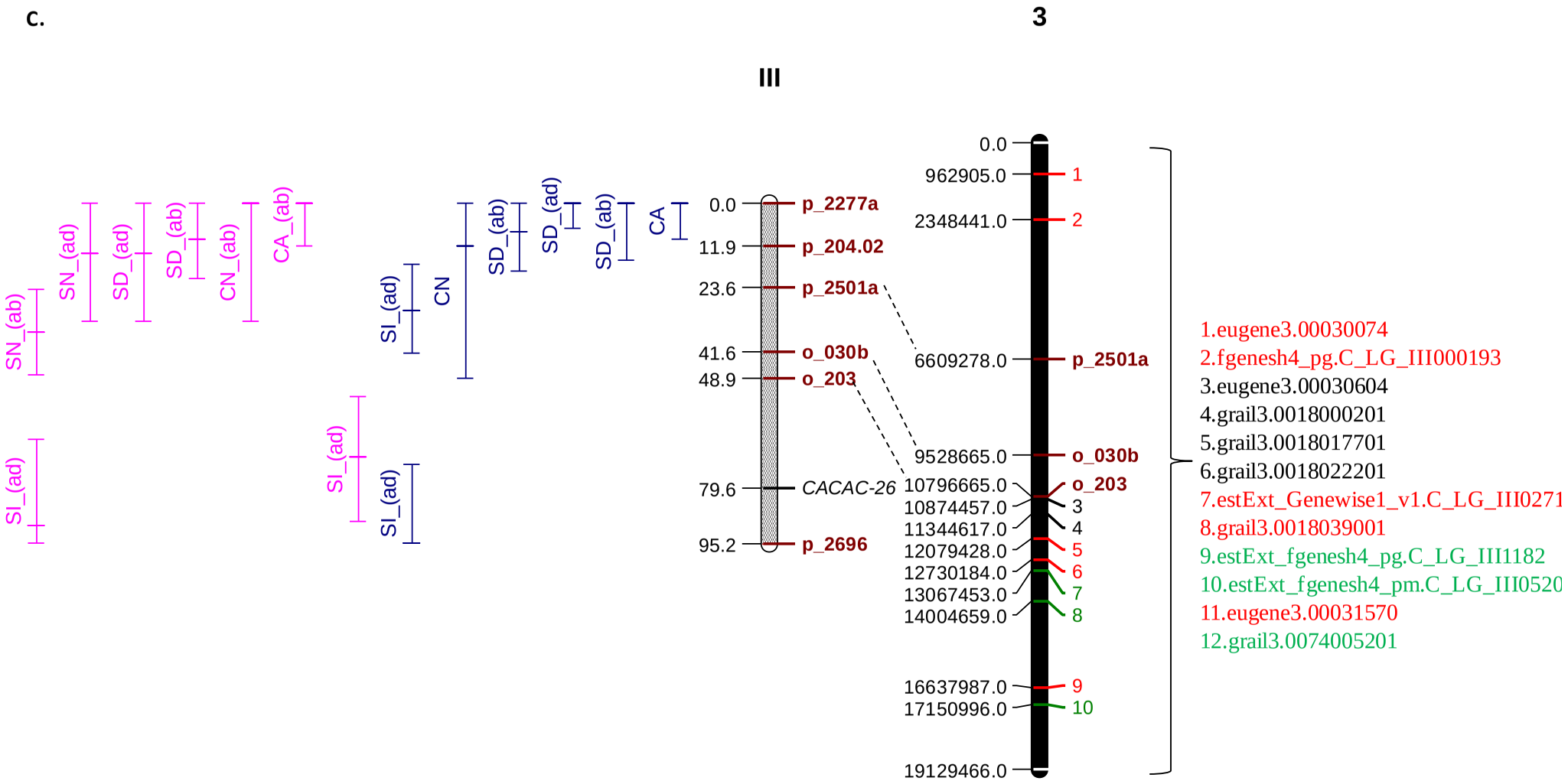
Abbreviation	**Description**
CA	Cell area (μm^2)
CA_(ab)	Abaxial epidermal cell area (μm^2)
CA_(ad)	Adaxial epidermal cell area (μm^2)
CN	Epidermal Cell Number
CN_(ab)	Abaxial Epidermal Cell Number
CN_(ad)	Adaxial Epidermal Cell Number
SD_(ab)	Abaxial Stomatal Density
SD_(ad)	Adaxial Stomatal Density
SI_(ab)	Abaxial Stomatal Index
SI_(ad)	Adaxial Stomatal Index
SN_(ab)	Abaxial stomatal Number
SN_(ad)	Adaxial Stomatal Number


A.
I
1
CN
CN
CN_(ab)
SI_(ad)
SI_(ab)
SD_(ad)
0.0 CCCCT-39
0.3 CCCCT-38
3.8 p_634
11.2 p_2789
38.2 o_30a
59.9 p_2852
72.9 p_93
117.7 p_2889b
132.9 p_575
136.7 AGCTG-10
140.7 CCCCT-N12
142.4 AGCTG-13
148.4 CCCCT-N13
163.6 CACTG-21
165.9 p_2786a
172.5 p_2385
174.7 TCCGT-20
0.0 1
810962.0 2
1756922.0 3
1950125.0 4
2814914.0 p_634
5467692.0 p_2789
6435691.0 5
6802936.0 6
6926009.0 7
9404981.0 p_2852
9764020.0 p_93
11239328.0
21053404.0 p_575
33771956.0 p_2385
35571568.0
1.estExt_fgenesh4_pg.C_LG_I0109
2.grail3.0100006103
3.estExt_Genewise1_v1.C_LG_I7028
4.fgenesh4_kg.C_LG_I000011
5.estExt_fgenesh4_pg.C_LG_I0819
6.eugene3.00010810
7.estExt_Genewise1_v1.C_LG_I9453

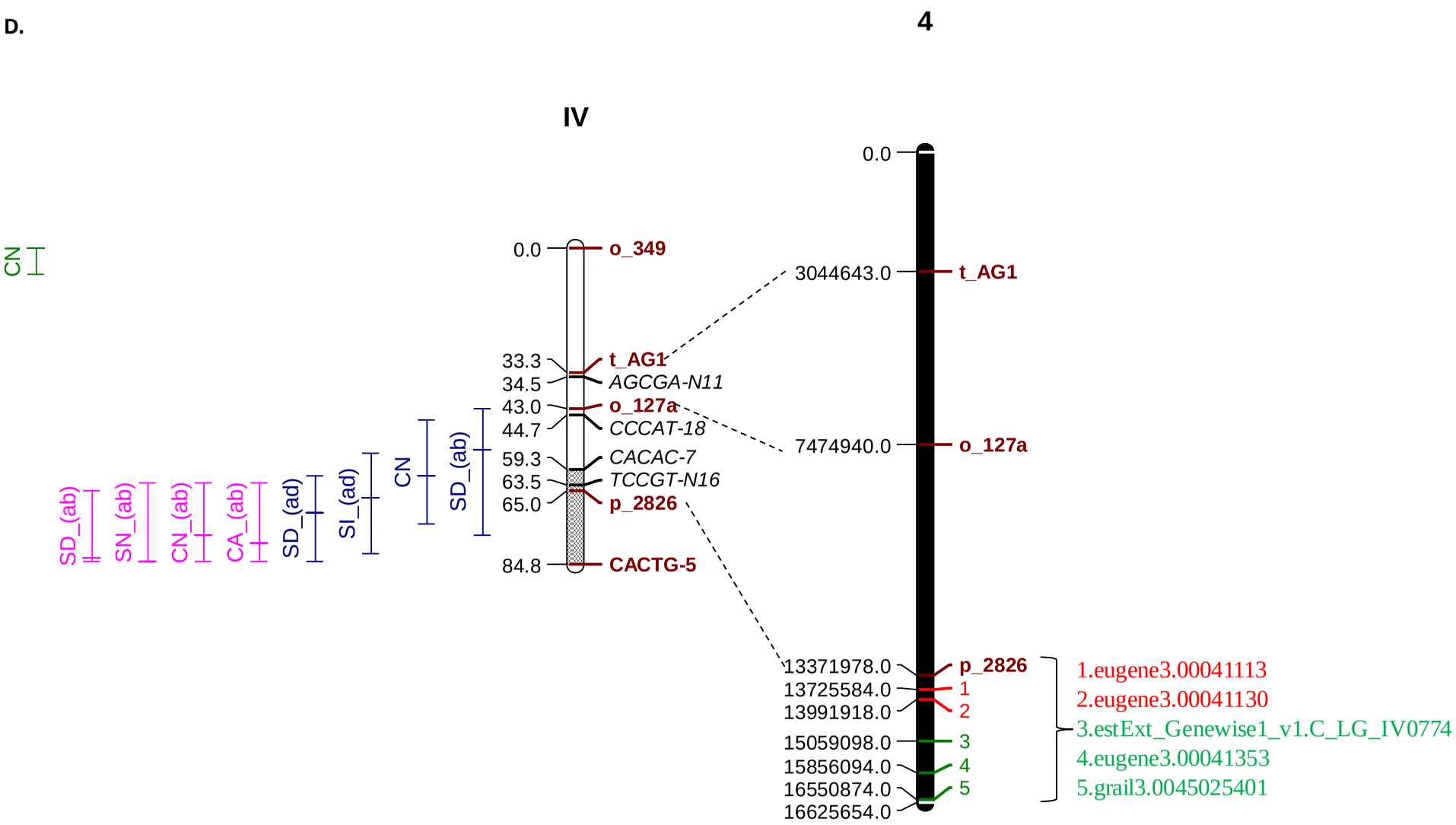
B.

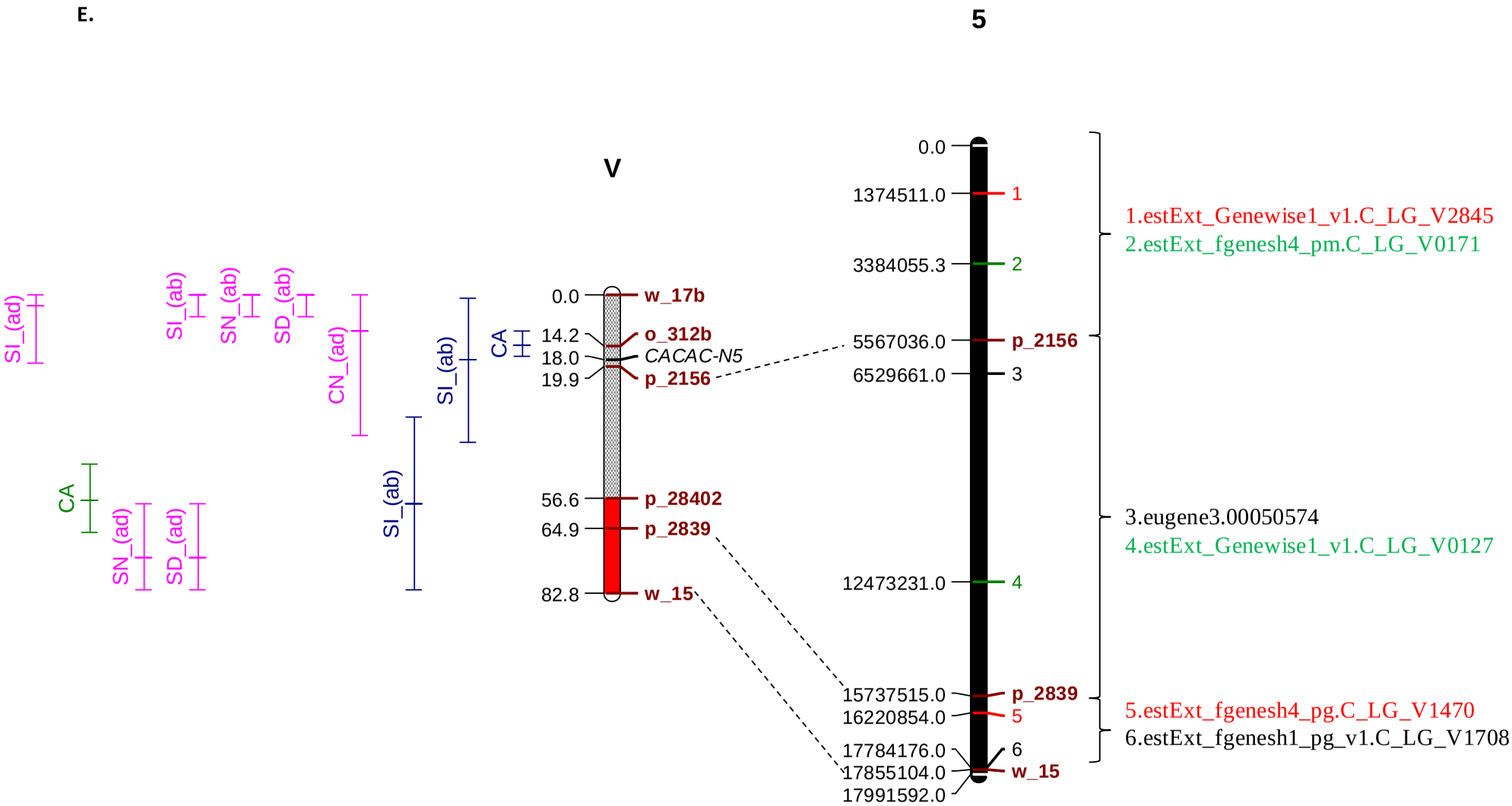

II
2
CN_(ad)
CN_(ab)
CA_(ad)
CA_(ab)
SD_(ad)
SI_(ad)
CA
SD_(ab)
CN
0.0
1.1
16.7
27.4
38.1
41.5
68.7
100.9
104.9
113.7
CCCTC-32
CCCTC-N10
p_667
p_422
AGCTG-N11
p_684
p_2797
p_2766
o_260
p_2418
0.0
1402200.0
2102700.0
2762699.0
3556760.8
4828045.0
7046216.0
10209505.0
15461149.0
18310162.0
20230512.0
24482572.0
1
2
3
p_667
p_422
p_684
p_2797
p_2766
o_260
p_2418
1. eugene3.00020210
2. grail3.0003027102
3. estExt_fgenesh4_pg.C_LG_II039

c.

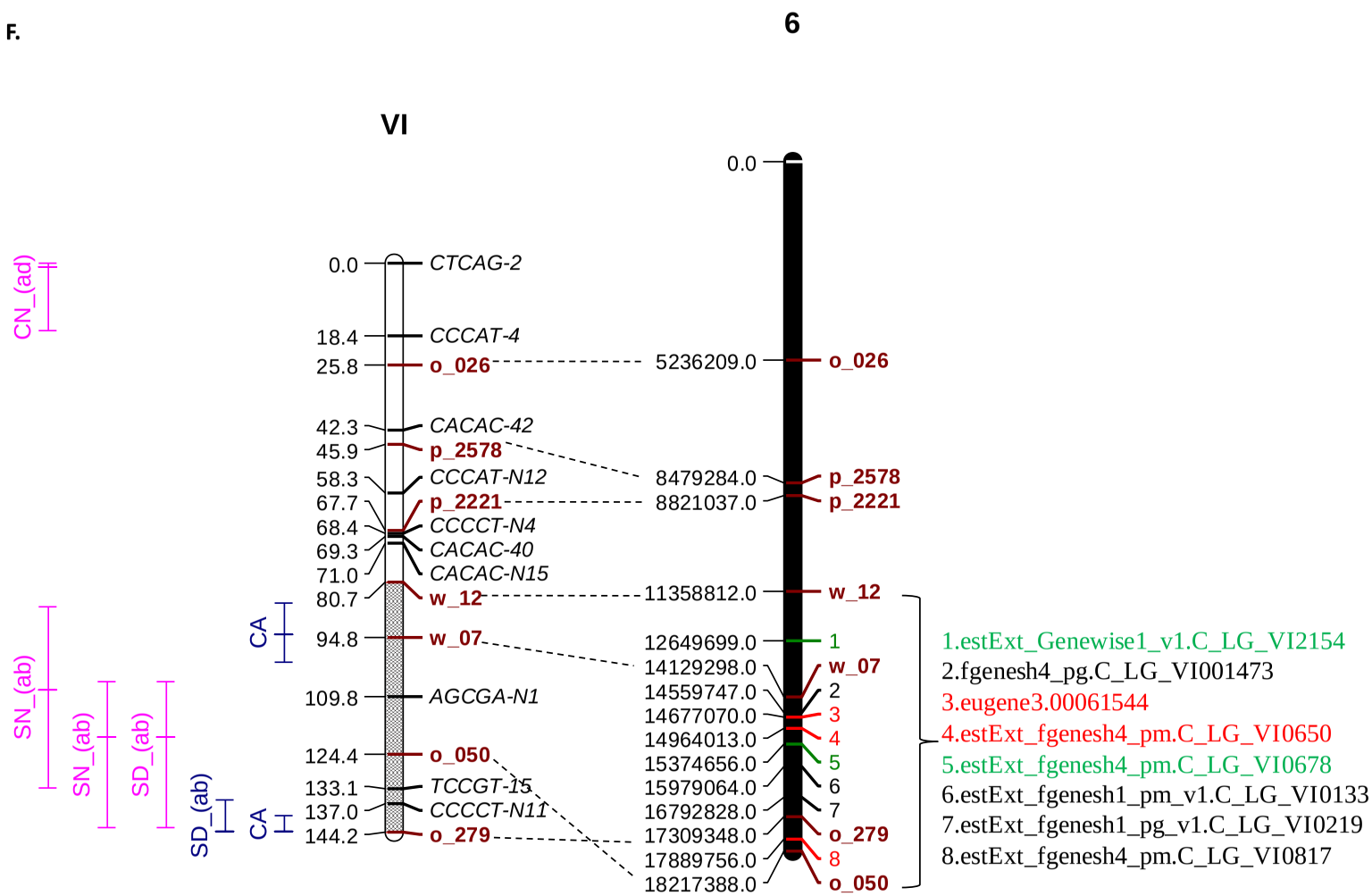

3
III
SN_(ab)
SI_(ad)
SN_(ad)
SD_(ad)
SD_(ab)
CN_(ab)
CA_(ab)
SI_(ad)
SI_(ad)
CN
SD_(ab)
SD_(ad)
SD_(ab)
CA
0.0
11.9
23.6
41.6
48.9
79.6
95.2
p_2277a
p_204.02
p_2501a
o_030b
o_203
CACAC-26
p_2696
0.0
962905.0
2348441.0
6609278.0
9528665.0
10796665.0
10874457.0
11344617.0
12079428.0
12730184.0
13067453.0
14004659.0
16637987.0
17150996.0
19129466.0
1
2
p_2501a
o_030b
o_203
3
4
5
6
7
8
9
10
1.eugene3.00030074
2.fgenesh4_pg.C_LG_III000193
3.eugene3.00030604
4.grail3.0018000201
5.grail3.0018017701
6.grail3.0018022201
7.estExt_Genewise1_v1.C_LG_III0271
8.grail3.0018039001
9.estExt_fgenesh4_pg.C_LG_III1182
10.estExt_fgenesh4_pm.C_LG_III0520
11.eugene3.00031570
12.grail3.0074005201

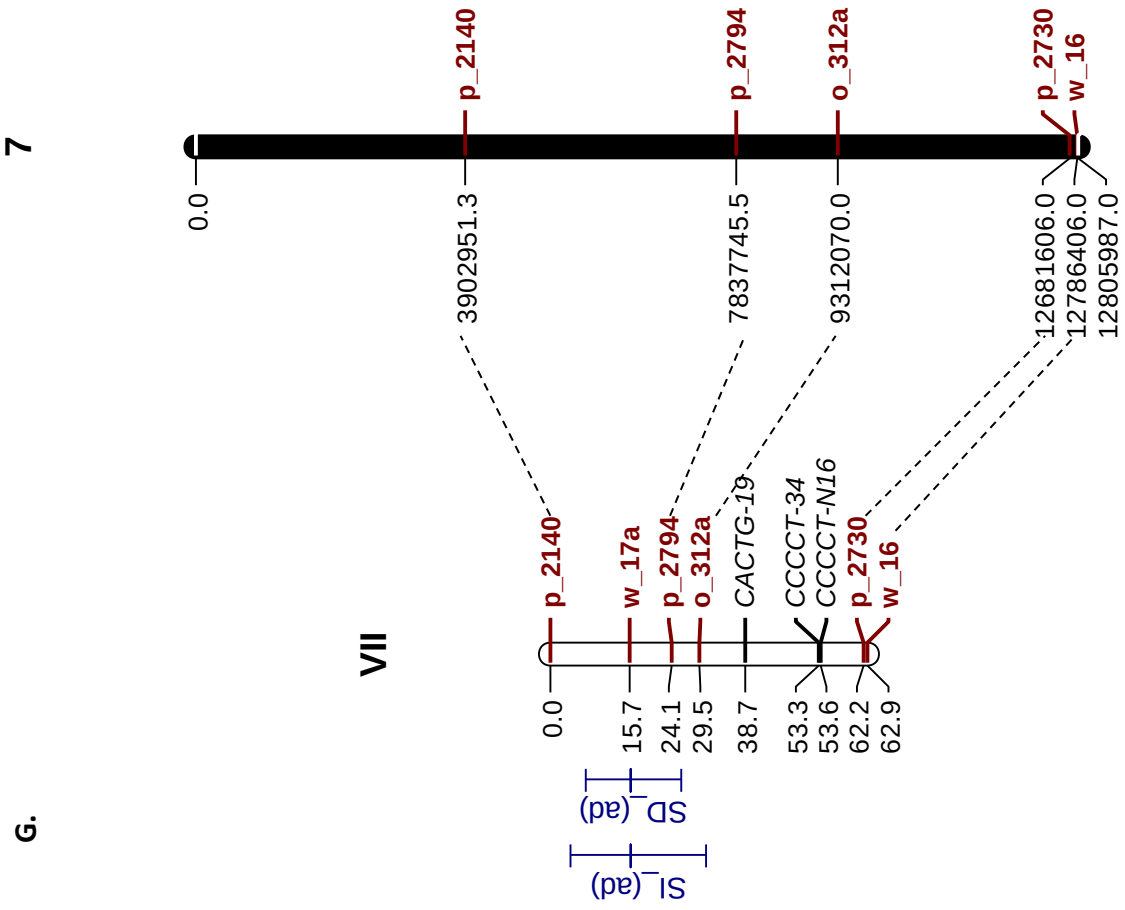
D.


4
IV
0.0
o_349
33.3
t_AG1
34.5
AGCGA-N11
43.0
o_127a
44.7
CCCAT-18
59.3
CACAC-7
63.5
TCCGT-N16
65.0
p_2826
84.8
CACTG-5
CN
SD_(ab)
SD_(ad)
SI_(ad)
CN
SD_(ab)
SN_(ab)
CN_(ab)
CA_(ab)
0.0
t_AG1
3044643.0
o_127a
7474940.0
p_2826
13371978.0
1
13725584.0
2
13991918.0
3
15059098.0
4
15856094.0
5
16550874.0
16625654.0
1.eugene3.00041113
2.eugene3.00041130
3.estExt_Genewise1_v1.C_LG_IV0774
4.eugene3.00041353
5.grail3.0045025401

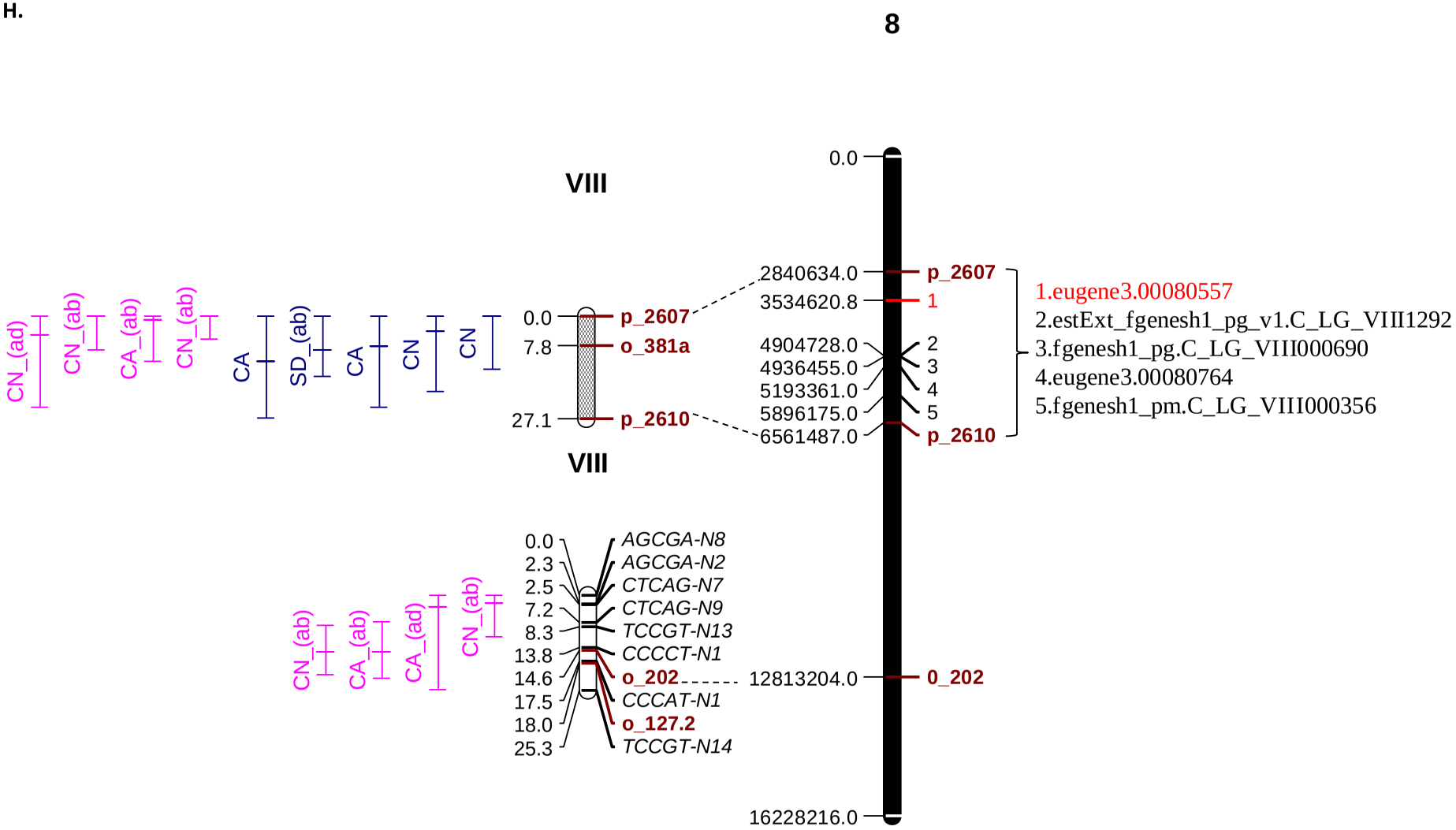

E.
V
5
SI_(ad)
SI_(ab)
SN_(ab)
SD_(ab)
CN_(ad)
CA
SN_(ad)
SD_(ad)
SI_(ab)
SI_(ab)
CA
0.0
14.2
18.0
19.9
56.6
64.9
82.8
w_17b
o_312b
CACAC-N5
p_2156
p_28402
p_2839
w_15
0.0
1374511.0
3384055.3
5567036.0
6529661.0
12473231.0
15737515.0
16220854.0
17784176.0
17855104.0
17991592.0
1
2
p_2156
3
4
p_2839
5
6
w_15
1.estExt_Genewise1_v1.C_LG_V2845
2.estExt_fgenesh4_pm.C_LG_V0171
3.eugene3.00050574
4.estExt_Genewise1_v1.C_LG_V0127
5.estExt_fgenesh4_pg.C_LG_V1470
6.estExt_fgenesh1_pg_v1.C_LG_V1708

F.

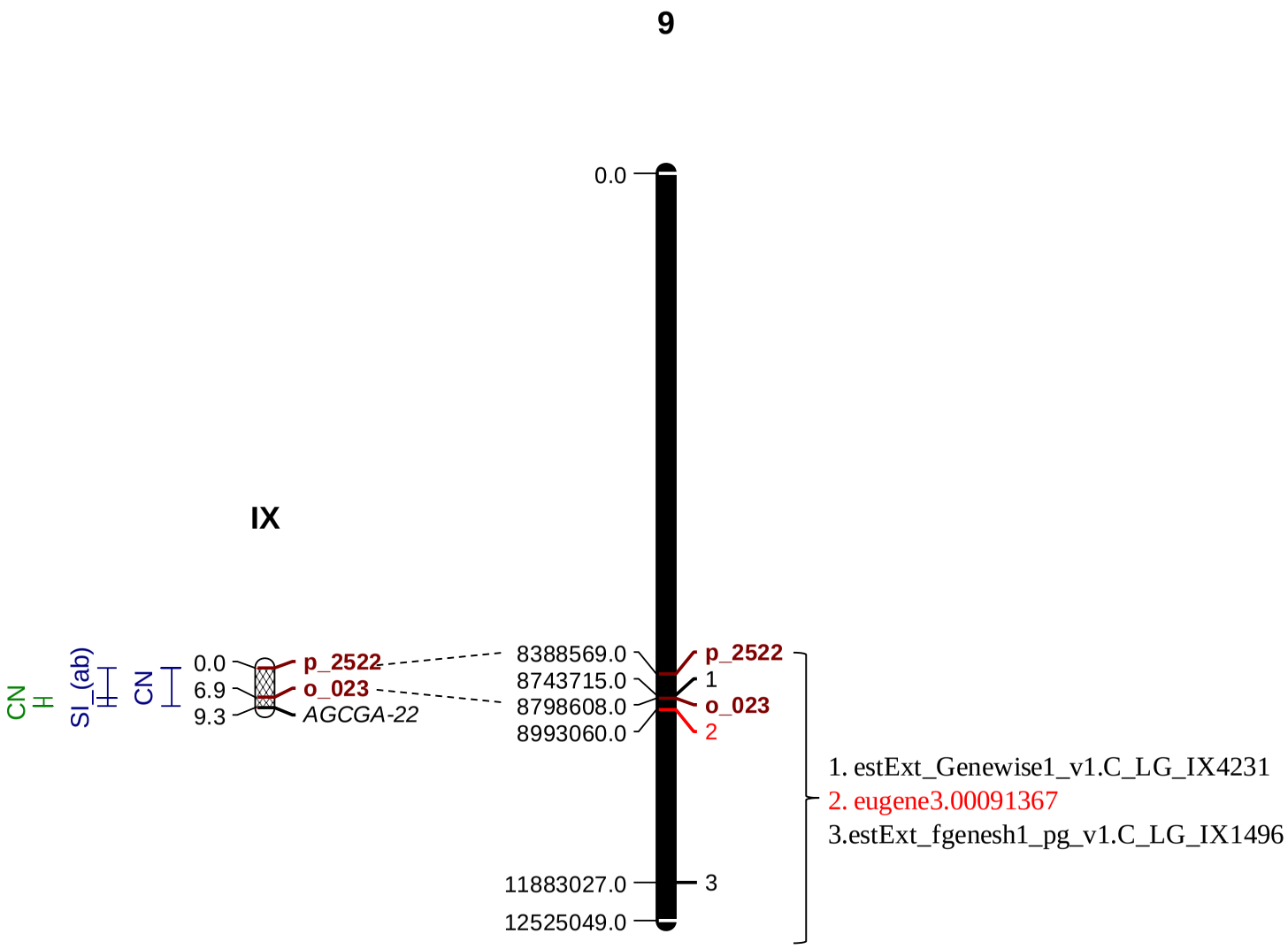

VI
6
0.0
CTCAG-2
18.4
CCCAT-4
25.8
o_026
42.3
CACAC-42
45.9
p_2578
58.3
CCCAT-N12
67.7
p_2221
68.4
CCCCT-N4
69.3
CACAC-40
71.0
CACAC-N15
80.7
w_12
94.8
w_07
109.8
AGCGA-N1
124.4
o_050
133.1
TCCGT-15
137.0
CCCCT-N11
144.2
o_279
CN_(ad)
SN_(ab)
SN_(ab)
SD_(ab)
SD_(ab)
CA
CA
5236209.0
8479284.0
8821037.0
11358812.0
12649699.0
14129298.0
14559747.0
14677070.0
14964013.0
15374656.0
15979064.0
16792828.0
17309348.0
17889756.0
18217388.0
1.estExt_Genewise1_v1.C_LG_VI2154
2.fgenesh4_pg.C_LG_VI001473
3.eugene3.00061544
4.estExt_fgenesh4_pm.C_LG_VI0650
5.estExt_fgenesh4_pm.C_LG_VI0678
6.estExt_fgenesh1_pm_v1.C_LG_VI0133
7.estExt_fgenesh1_pg_v1.C_LG_VI0219
8.estExt_fgenesh4_pm.C_LG_VI0817


G.
7
0.0
3902951.3
7837745.5
9312070.0
12681606.0
12786406.0
12805987.0
p_2140
p_2794
o_312a
p_2730
w_16
VII
0.0
15.7
24.1
29.5
38.7
53.3
53.6
62.2
62.9
p_2140
w_17a
p_2794
o_312a
CACTG-19
CCCCT-34
CCCCT-N16
p_2730
w_16
SD_(ad)
SI_(ad)

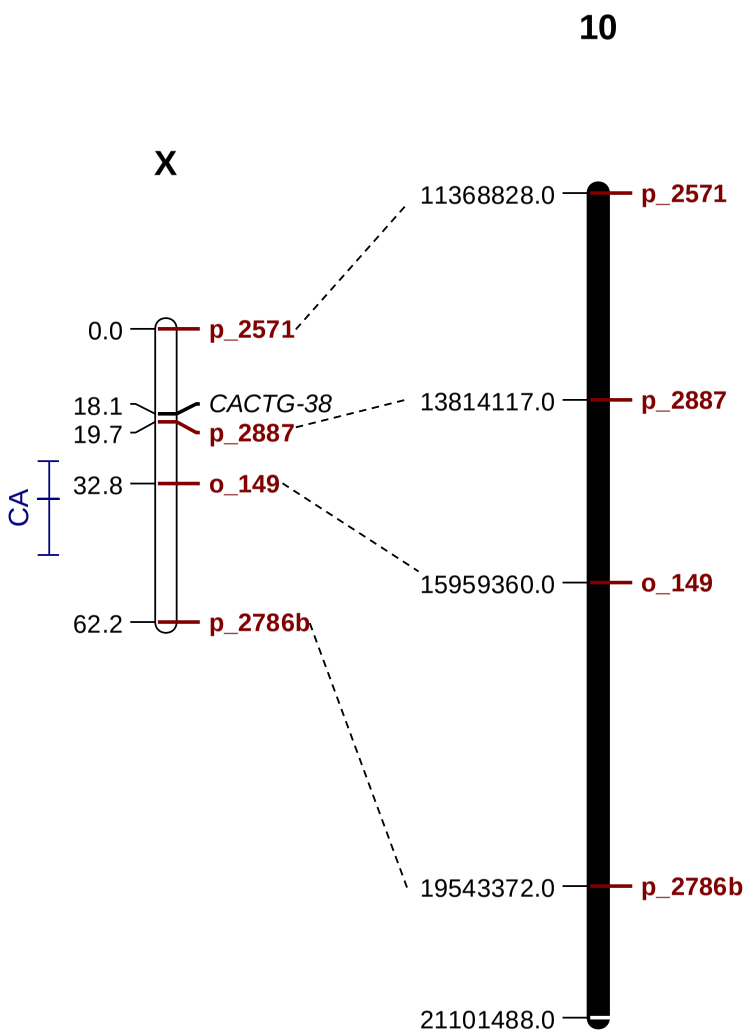
H.


8
VIII
0.0
p_2607
7.8
o_381a
27.1
p_2610
CN_(ad)
CN_(ab)
CA_(ab)
CN_(ab)
CA
SD_(ab)
CA
CN
CN
2840634.0
p_2607
3534620.8
1
4904728.0
2
4936455.0
3
5193361.0
4
5896175.0
5
6561487.0
p_2610
1.eugene3.00080557
2.estExt_fgenesh1_pg_v1.C_LG_VIII1292
3.fgenesh1_pg.C_LG_VIII000690
4.eugene3.00080764
5.fgenesh1_pm.C_LG_VIII000356
VIII
0.0
AGCGA-N8
2.3
AGCGA-N2
2.5
CTCAG-N7
7.2
CTCAG-N9
8.3
TCCGT-N13
13.8
CCCCT-N1
14.6
o_202
17.5
CCCAT-N1
18.0
o_127.2
25.3
TCCGT-N14
CN_(ab)
CA_(ab)
CA_(ad)
CN_(ab)
12813204.0
0_202
16228216.0

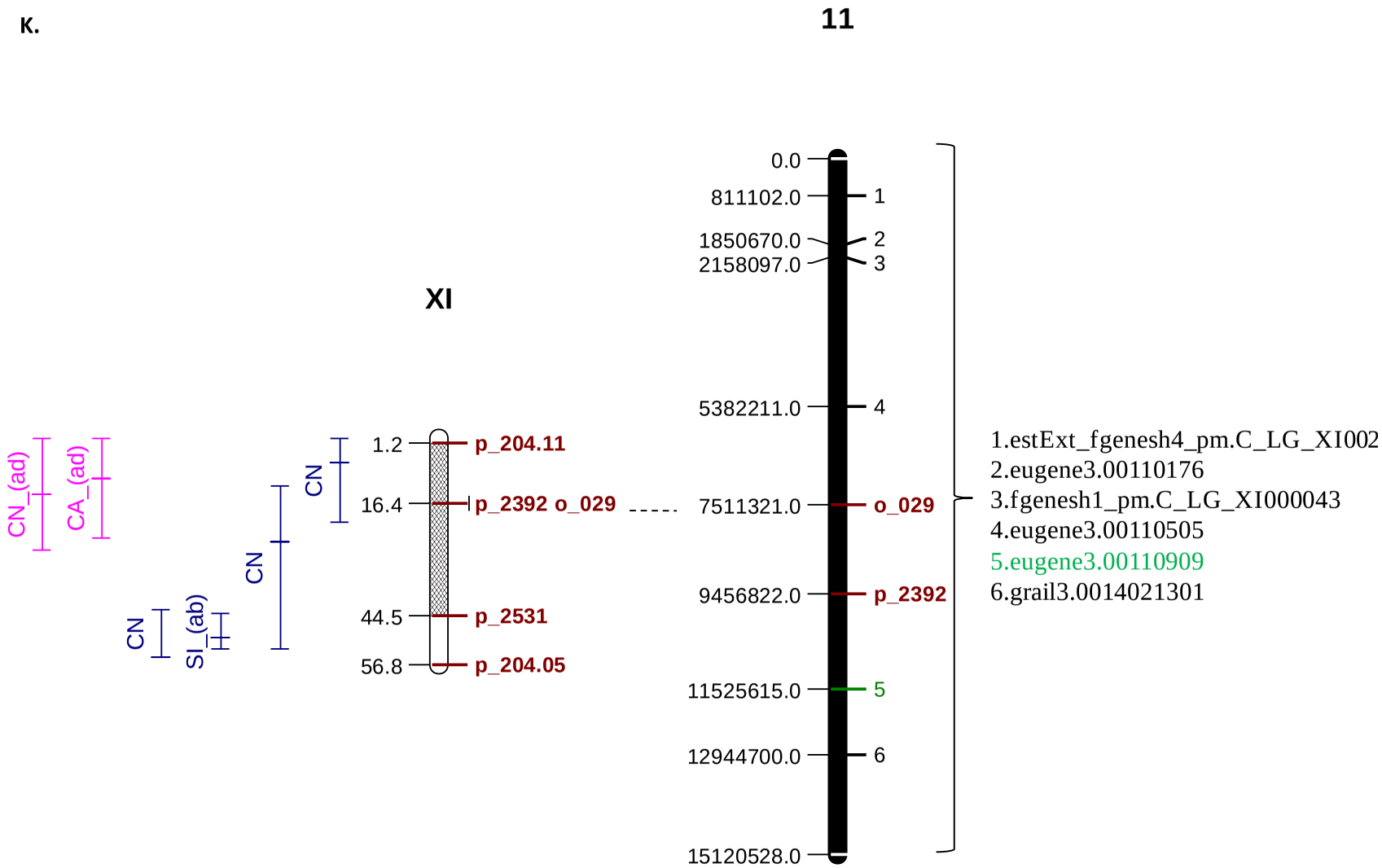
I.


9
0.0
8388569.0
8743715.0
8798608.0
8993060.0
11883027.0
12525049.0
p_2522
1
o_023
2
3
IX
CN
H
SI_(ab)
H
CN
0.0
6.9
9.3
p_2522
o_023
AGCGA-22
1. estExt_Genewise1_v1.C_LG_IX4231
2. eugene3.00091367
3.estExt_fgenesh1_pg_v1.C_LG_IX1496

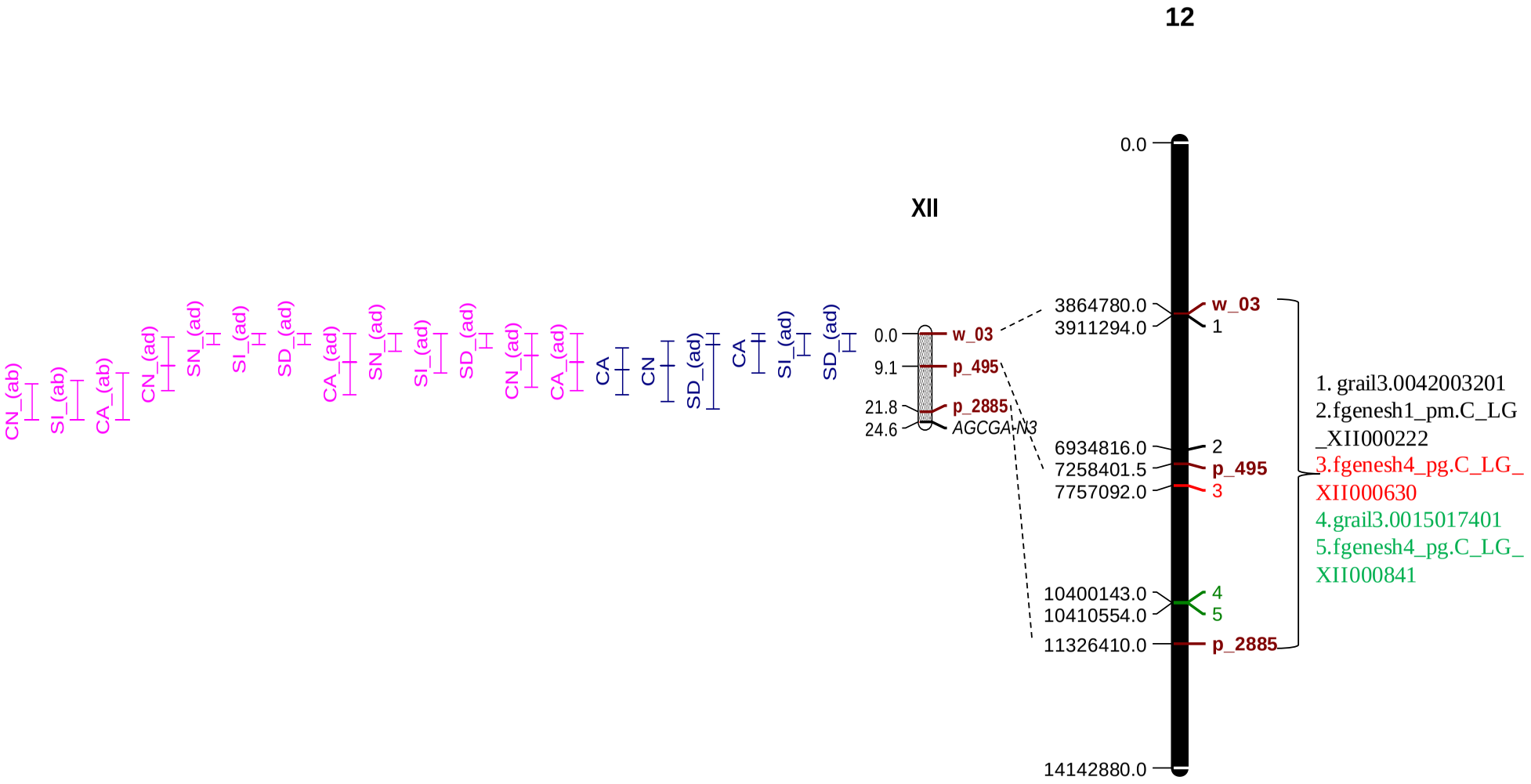
J.


10
X
11368828.0
p_2571
0.0
p_2571
18.1
CACTG-38
19.7
p_2887
13814117.0
p_2887
CA
32.8
o_149
15959360.0
o_149
62.2
p_2786b
19543372.0
p_2786b
21101488.0

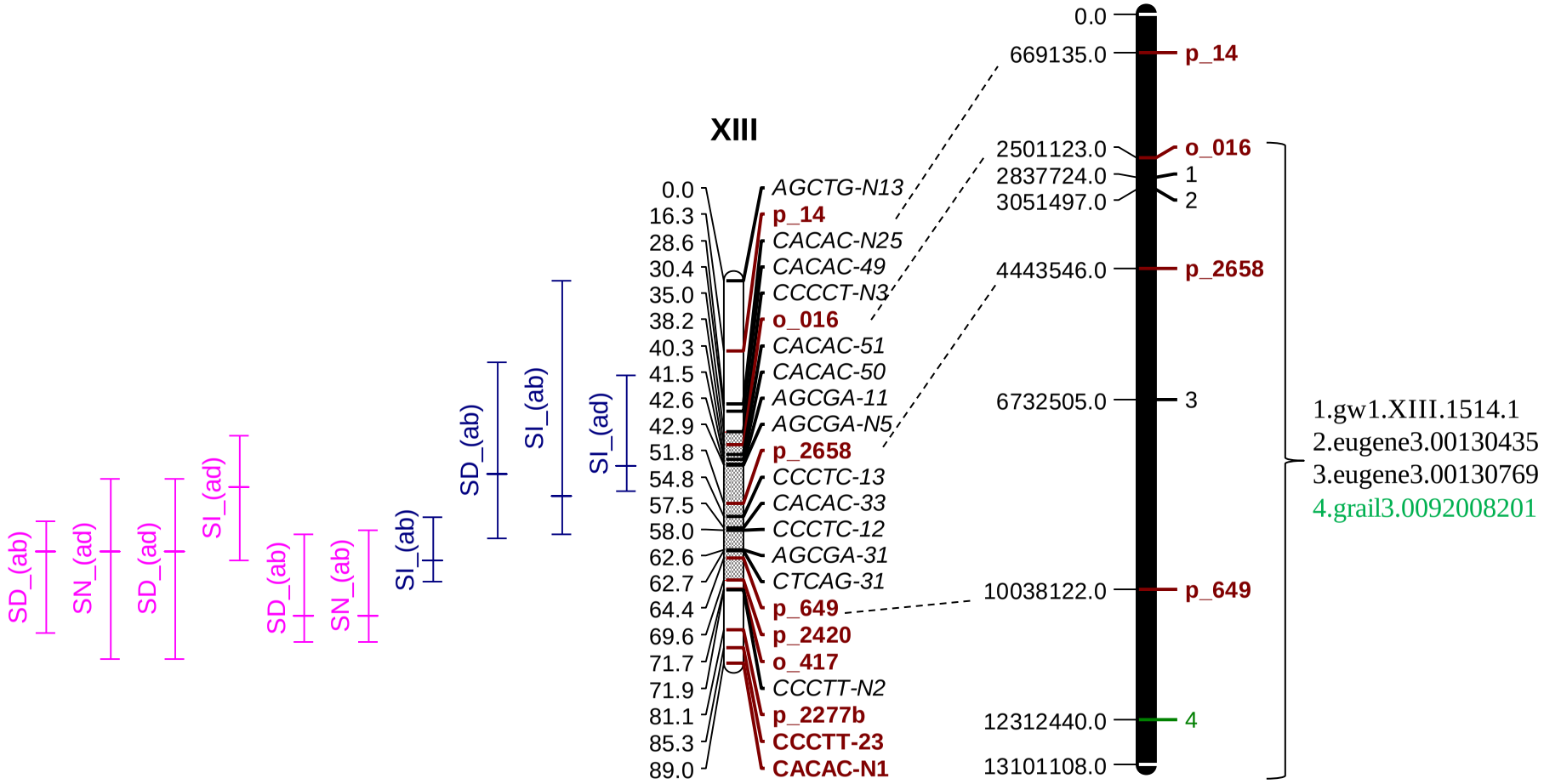
K.


11
XI
0.0
811102.0
1850670.0
2158097.0
5382211.0
7511321.0
9456822.0
11525615.0
12944700.0
15120528.0
1
2
3
4
o_029
p_2392
5
6
1.2
16.4
44.5
56.8
p_204.11
p_2392 o_029
p_2531
p_204.05
CN_(ad)
CA_(ad)
CN
SI_(ab)
CN
CN
1.estExt_fgenesh4_pm.C_LG_XI002
2.eugene3.00110176
3.fgenesh1_pm.C_LG_XI000043
4.eugene3.00110505
5.eugene3.00110909
6.grail3.0014021301

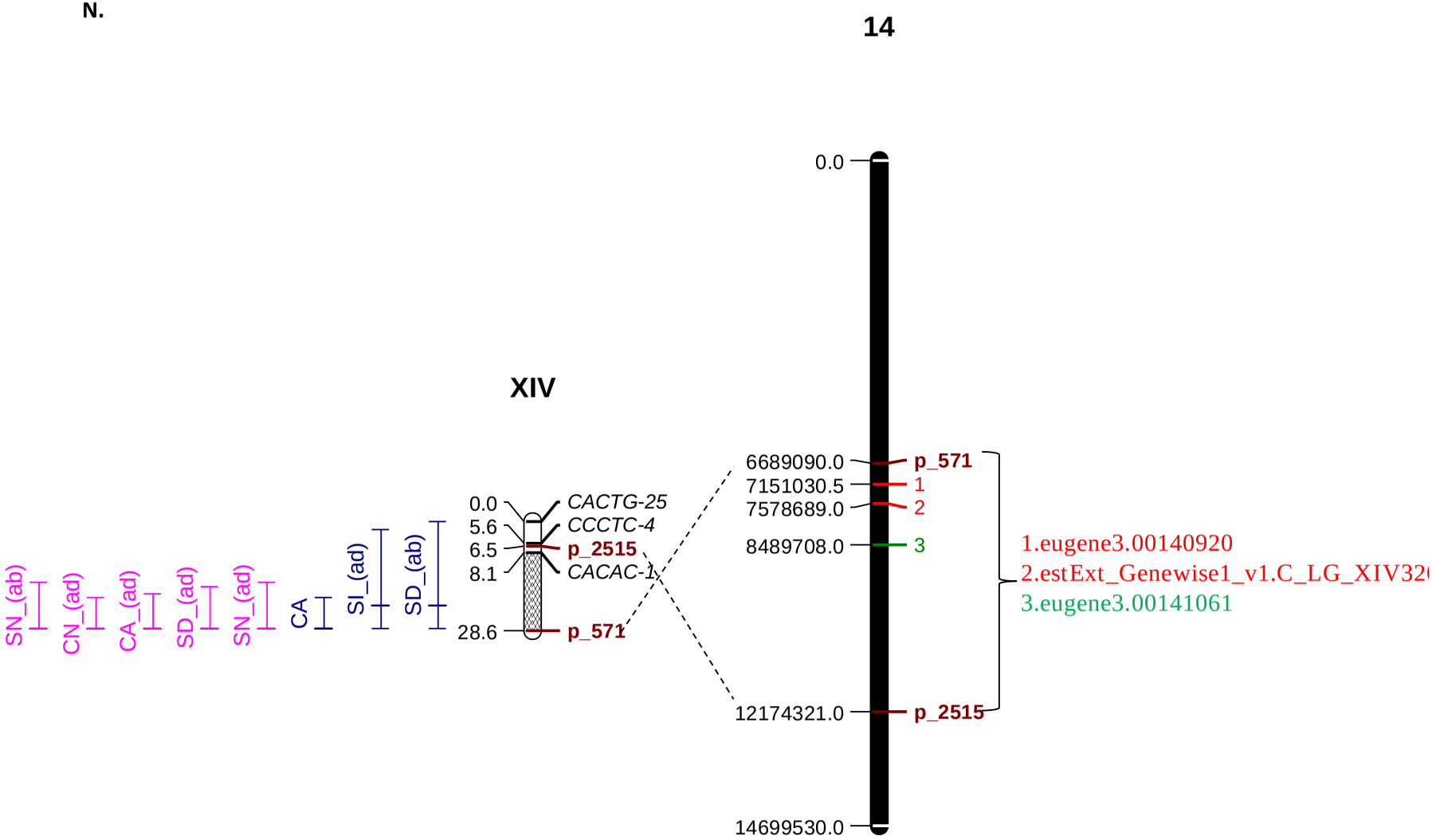
L.


12
XII
CN_(ab)
SI_(ab)
CA_(ab)
CN_(ad)
SN_(ad)
SI_(ad)
SD_(ad)
CA_(ad)
SN_(ad)
SI_(ad)
SD_(ad)
CN_(ad)
CA_(ad)
CA
CN
SD_(ad)
CA
SI_(ad)
SD_(ad)
0.0
9.1
21.8
24.6
w_03
p_495
p_2885
AGCGA-N3
0.0
3864780.0
3911294.0
6934816.0
7258401.5
7757092.0
10400143.0
10410554.0
11326410.0
14142880.0
w_03
1
2
p_495
3
4
5
p_2885
1. grail3.0042003201
2.fgenesh1_pm.C_LG
_XII000222
3.fgenesh4_pg.C_LG_
XII000630
4.grail3.0015017401
5.fgenesh4_pg.C_LG_
XII000841

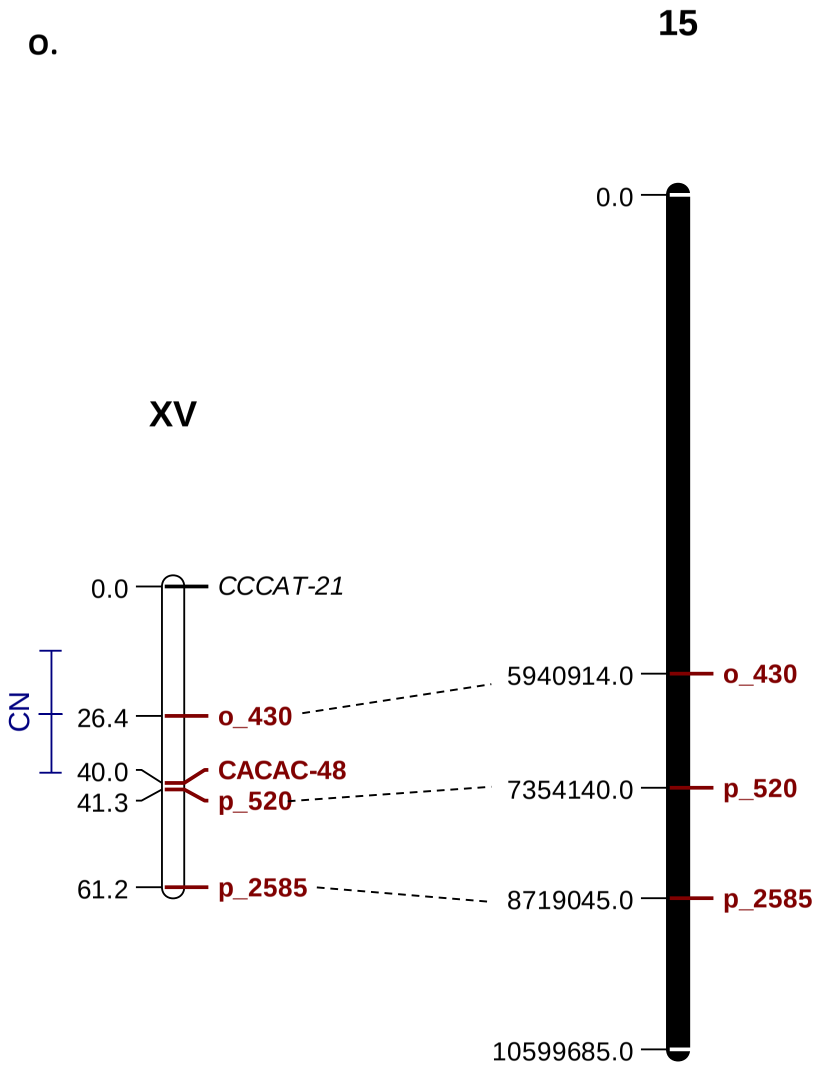
M.


13
XIII
SD_(ab)
SN_(ad)
SD_(ad)
SI_(ad)
SD_(ab)
SN_(ab)
SI_(ab)
SD_(ab)
SI_(ab)
SI_(ad)
0.0
16.3
28.6
30.4
35.0
38.2
40.3
41.5
42.6
42.9
51.8
54.8
57.5
58.0
62.6
62.7
64.4
69.6
71.7
71.9
81.1
85.3
89.0
AGCTG-N13
p_14
CACAC-N25
CACAC-49
CCCCT-N3
o_016
CACAC-51
CACAC-50
AGCGA-11
AGCGA-N5
p_2658
CCCTC-13
CACAC-33
CCCTC-12
AGCGA-31
CTCAG-31
p_649
p_2420
o_417
CCCTT-N2
p_2277b
CCCTT-23
CACAC-N1
0.0
669135.0
2501123.0
2837724.0
3051497.0
4443546.0
6732505.0
10038122.0
12312440.0
13101108.0
p_14
o_016
1
2
p_2658
3
p_649
4
1.gw1.XIII.1514.1
2.eugene3.00130435
3.eugene3.00130769
4.grail3.0092008201

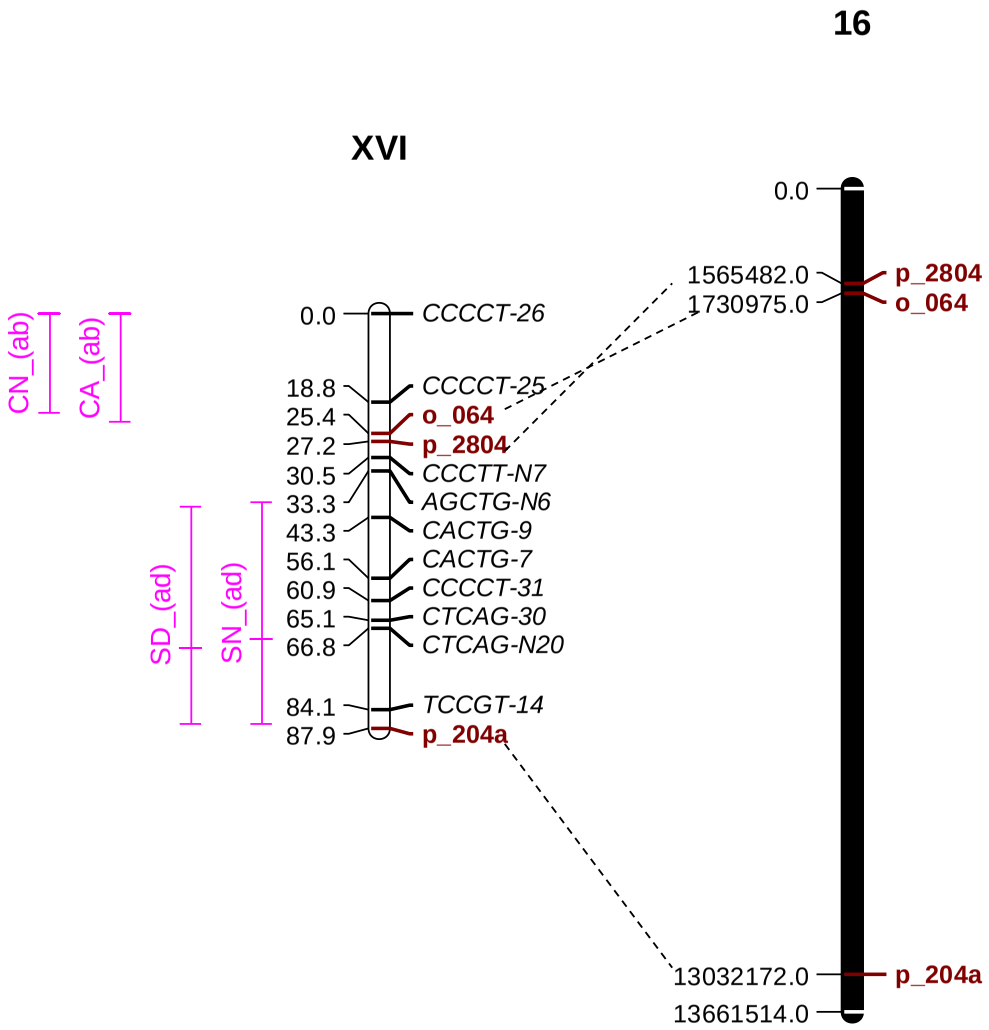
N.

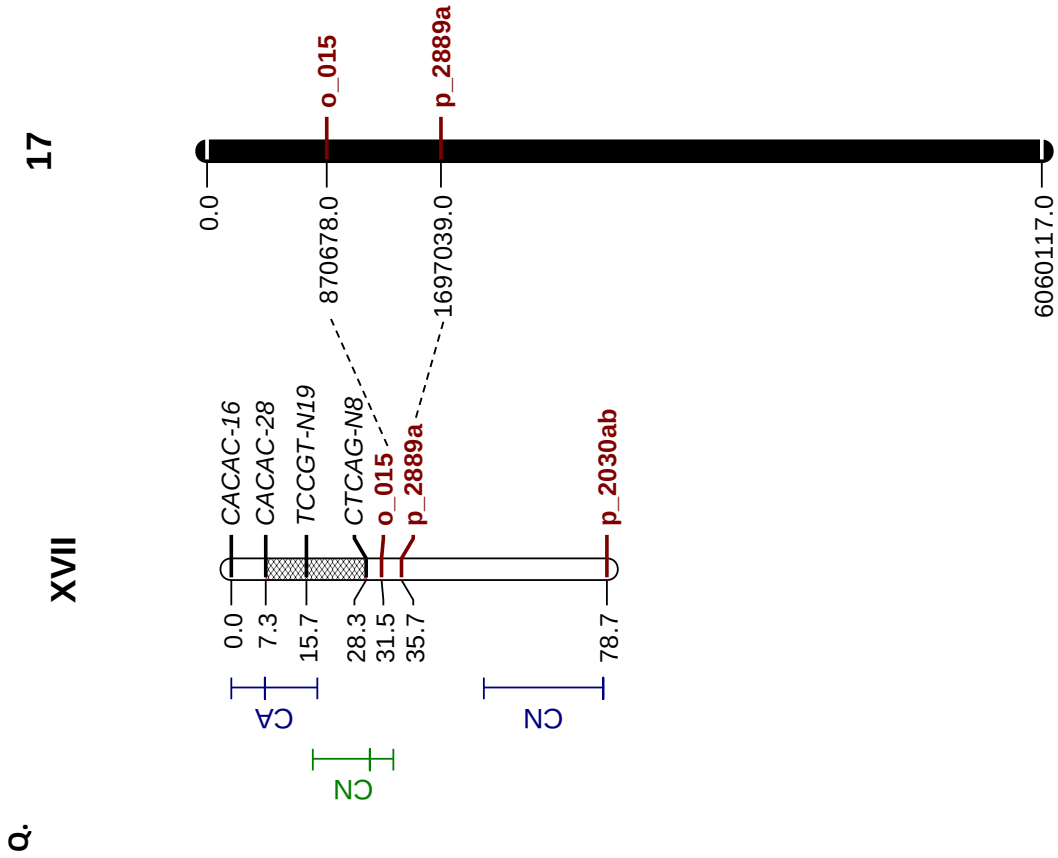

14
0.0
6689090.0
p_571
7151030.5
1
7578689.0
2
8489708.0
3
12174321.0
p_2515
14699530.0
1.eugene3.00140920
2.estExt_Genewise1_v1.C_LG_XIV32
3.eugene3.00141061
XIV
0.0
5.6
6.5
8.1
28.6
CACTG-25
CCCTC-4
p_2515
CACAC-1
p_571
SN_(ab)
CN_(ad)
CA_(ad)
SD_(ad)
SN_(ad)
CA
SI_(ad)
SD_(ab)

o.

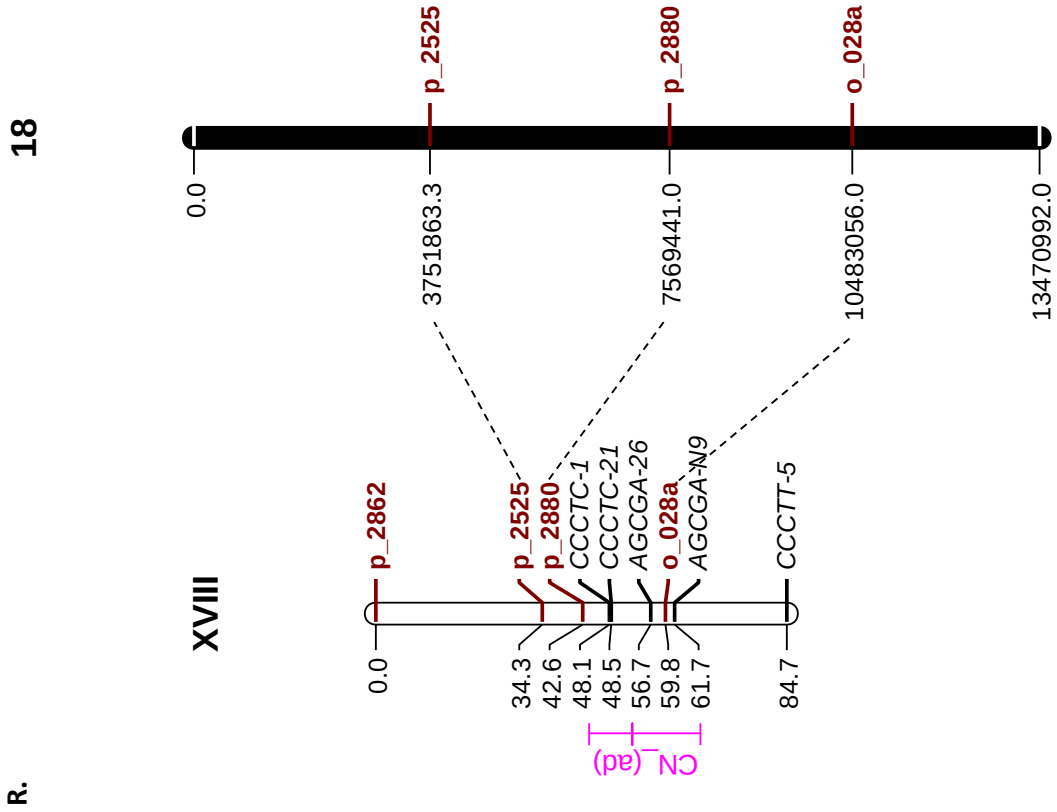

15
0.0
5940914.0
o_430
7354140.0
p_520
8719045.0
p_2585
10599685.0
XV
0.0
CCCAT-21
CN
26.4
o_430
40.0
CACAC-48
41.3
p_520
61.2
p_2585

P.


16
XVI
0.0
1565482.0
1730975.0
p_2804
o_064
13032172.0
13661514.0
p_204a
CN_(ab)
CA_(ab)
SD_(ad)
SN_(ad)
0.0
CCCCT-26
18.8
CCCCT-25
25.4
o_064
27.2
p_2804
30.5
CCCTT-N7
33.3
AGCTG-N6
43.3
CACTG-9
56.1
CACTG-7
60.9
CCCCT-31
65.1
CTCAG-30
66.8
CTCAG-N20
84.1
TCCGT-14
87.9
p_204a


17
o_015
p_2889a
0.0
870678.0
1697039.0
6060117.0
XVII
CACAC-16
CACAC-28
TCCGT-N19
CTCAG-N8
o_015
p_2889a
p_2030ab
0.0
7.3
15.7
28.3
31.5
35.7
78.7
CA
CN
CN

Q.


R.
18
0.0
p_2525
3751863.3
p_2880
7569441.0
o_028a
10483056.0
13470992.0
XVIII
0.0
p_2862
34.3
p_2525
42.6
p_2880
48.1
CCCTC-1
48.5
CCCTC-21
56.7
AGCGA-26
59.8
o_028a
61.7
AGCGA-N9
84.7
CCCTT-5
CN_(ad)

S.

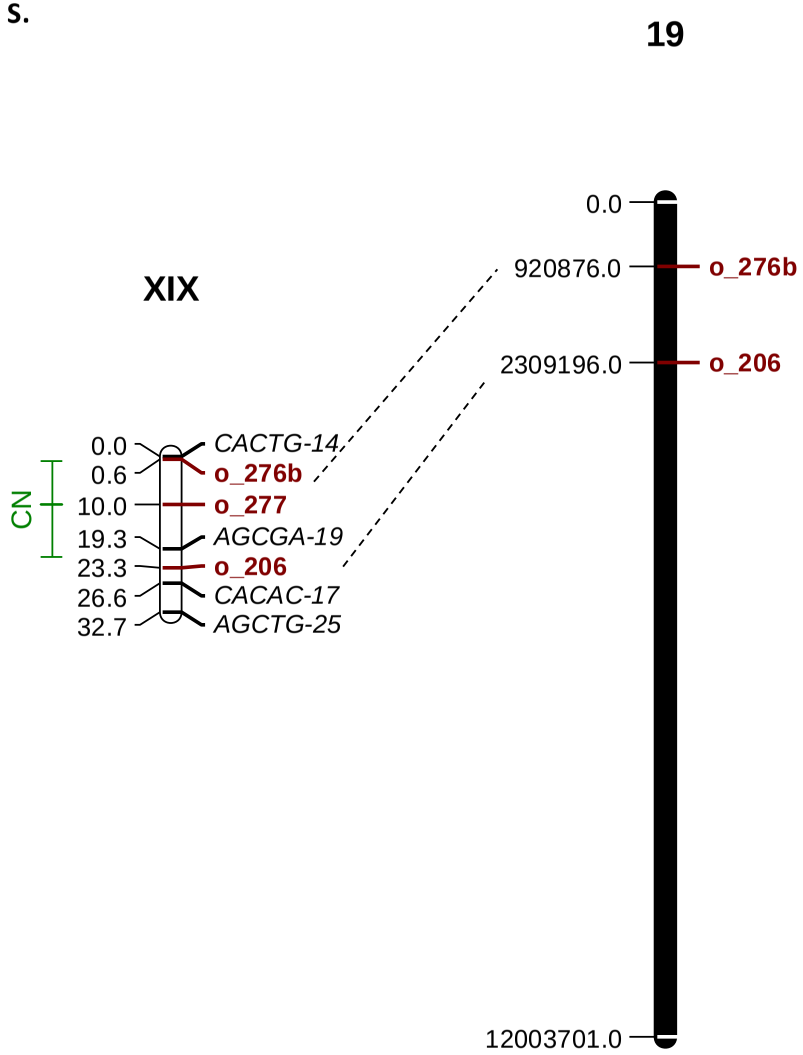

19
0.0
920876.0
o_276b
2309196.0
o_206
12003701.0
XIX
CN
0.0
0.6
10.0
19.3
23.3
26.6
32.7
CACTG-14
o_276b
o_277
AGCGA-19
o_206
CACAC-17
AGCTG-25

Table 3-6: Candidate gene selection.Summary ‘minor’ candidate gene selected for leaf, biomass and cell traits as found within regions of the genetic map consisting of two QTL experiments termed a ‘minor’ hotspot. ‘X’ represents whether candidate gene found within leaf & biomass QTL (XX), cell QTL (xx) or both (Xx). LG represents linkage group.

LG	Poplar gene model	Origin of QTL	Short Description
I	estExt_Genewise1_v1.C_LG_I2426	XX	CA1 (CARBONIC ANHYDRASE 1)
I	estExt_fgenesh4_pg.C_LG_I2255	XX	LHT1 (LYSINE HISTIDINE TRANSPORTER 1)
I	gw1.I.3059.1	XX	Peptidase M48 family protein
I	fgenesh4_pg.C_LG_I002334	XX	TSO2 (TSO2); ribonucleoside-diphosphate reductase
I	estExt_fgenesh4_pg.C_LG_I2498	XX	Similar to glucan endo-1,3-beta-glucosidase-related
I	estExt_Genewise1_v1.C_LG_I4040	XX	Leucine-rich repeat family protein
I	fgenesh1_pg.C_LG_I002802	XX	TORNADO1 -changes leaf shape by reducing cell number.
I	eugene3.00012637	XX	Similar to unknown protein
I	grail3.0032009701	XX	DNA binding / ligand-dependent nuclear receptor
I	estExt_Genewise1_v1.C_LG_I4896	XX	PHABULOSA - abaxial/ adaxial cell fate
I	eugene3.00012781	XX	Glycosyl hydrolase family protein 17
II	estExt_fgenesh4_pg.C_LG_II0662	XX	ADH1 (ALCOHOL DEHYDROGENASE 1)
II	eugene3.00020820	XX	Polcalcin, putative / calcium-binding pollen allergen
II	estExt_fgenesh4_pg.C_LG_II0927	XX	Plastocyanin-like domain-containing protein

Table 3-6 continued …

II	estExt_fgenesh4_pg.C_LG_II0928	XX	Plastocyanin-like domain-containing protein
II	fgenesh1_pg.C_LG_II000954	XX	CYCD4;1 - core cell cycle gene
II	estExt_Genewise1_v1.C_LG_II0565	XX	CYCA3;2 - a core cell cycle gene
II	grail3.0033028501	XX	YABBY (YAB3) involved in specifying abaxial cell fate.
II	grail3.0039014201	XX	JAR1 (JASMONATE RESISTANT 1)
II	eugene3.00021666	XX	60S acidic ribosomal protein P1 (RPP1A)
II	estExt_Genewise1_v1.C_LG_II1841	XX	SAG21 (SENESCENCE-ASSOCIATED GENE 21)
III	eugene3.00030074	Xx	Beta-hydroxyacyl-ACP dehydratase, putative
III	fgenesh4_pg.C_LG_III000193	Xx	*Arabidopsis thaliana* translocase inner membrane subunit 23-2
III	grail3.0018017701	Xx	YABBY (YAB2) - involved in the abaxial cell fate
III	grail3.0018022201	Xx	ATGRF7 - Growth regulating factor
III	estExt_Genewise1_v1.C_LG_III0271	Xx	PIP1C (PLASMA MEMBRANE INTRINSIC PROTEIN 1;3)
III	grail3.0018039001	Xx	AGP20 (ARABINOGALACTAN PROTEIN 20)
III	estExt_fgenesh4_pg.C_LG_III1182	Xx	Similar to unknown protein
III	estExt_fgenesh4_pm.C_LG_III0520	Xx	Vacuolar sorting receptor, putative
III	eugene3.00031570	Xx	Histone H3
III	grail3.0074005201	Xx	Calmodulin-binding protein-related
III	eugene3.00030604	xx	CYCD5;1 - core cell cycle gene
III	grail3.0018000201	xx	TORNADO2 - maintenance of radial tissue differentiation

Table 3-6: Continued…

V	estExt_Genewise1_v1.C_LG_V2845	Xx	Lipid-associated family protein
V	estExt_fgenesh4_pm.C_LG_V0171	Xx	CAT2 (CATALASE 2); catalase
VI	estExt_Genewise1_v1.C_LG_VI2154	Xx	TET8 (TETRASPANIN8)
VI	fgenesh4_pg.C_LG_VI001473	Xx	GPA1 -Positive regulator in abscisic acid (ABA) inhibition
VI	eugene3.00061544	Xx	Ferredoxin-related
VI	estExt_fgenesh4_pm.C_LG_VI0650	Xx	17.6 kDa class II heat shock protein (HSP17.6-CII)
VI	estExt_fgenesh4_pm.C_LG_VI0678	Xx	Glycine dehydrogenase (decarboxylating), putative
VI	estExt_fgenesh1_pm_v1.C_LG_VI0133	Xx	PHAVOLUTA - abaxial/ adaxial cell fate
VI	estExt_fgenesh1_pg_v1.C_LG_VI0219	Xx	LEUING - *Arabidopsis thaliana* WD-40 repeat family protein
VI	estExt_fgenesh4_pm.C_LG_VI0817	Xx	Armadillo/beta-catenin repeat family protein
VII	estExt_Genewise1_v1.C_LG_VII2923	XX	CURLY LEAF - Polycomb group protein
VII	fgenesh1_pg.C_LG_VII000913	XX	ASL1 - repression of KNOX genes in the SAM
VII	gw1.VII.3777.1	XX	ATTRX1 (*Arabidopsis thaliana* thioredoxin H-type 1)
VII	estExt_Genewise1_v1.C_LG_VII3915	XX	EMB2296 (EMBRYO DEFECTIVE 2296)
VII	grail3.0019030801	XX	Histone H4
VII	grail3.0019031001	XX	WUSCHEL - Homeobox gene controlling the stem cell pool
VII	estExt_Genewise1_v1.C_LG_VII0082	XX	ATGRF2 - Growth regulating factor

Table 3-6: Continued…

VII	fgenesh1_pg.C_LG_VII001256	XX	CYCD4;1 - core cell cycle gene
XI	estExt_fgenesh4_pm.C_LG_XI0028	Xx	STM - required for SAM formation
XI	eugene3.00110176	Xx	ACC oxidase - ethylene forming enzyme
XI	fgenesh1_pm.C_LG_XI000043	Xx	EXGT-A1 (XTH4) - xyloglucan
XI	eugene3.00110505	Xx	AGAMOUS - maintenance of floral organ identity
XI	eugene3.00110909	Xx	MT3 (METALLOTHIONEIN 3)
XI	grail3.0014021301	Xx	H2A - core histone
XII	grail3.0042003201	Xx	PFS2 - member of the WUS transcription factors
XII	fgenesh1_pm.C_LG_XII000222	Xx	SWELLMAP2 (SMP2) - zinc finger containing protein
XII	fgenesh4_pg.C_LG_XII000630	Xx	Unknown protein
XII	grail3.0015017401	Xx	Polcalcin / calcium-binding pollen allergen, putative
XII	fgenesh4_pg.C_LG_XII000841	Xx	NMT1 (N-MYRISTOYLTRANSFERASE 1)
XIX	estExt_Genewise1_v1.C_LG_XIX0269	XX	LCR68/PDF2.3 (Low-molecular-weight cysteine-rich 68)
XIX	eugene3.00190480	XX	ATGRF9 - growth regulating factor
XIX	grail3.0094003501	XX	SP1L4 (SPIRAL1-LIKE4)
XV	eugene3.00150073	XX	ATGRF5 - growth regulating factor
XV	fgenesh1_pm.C_LG_XV000071	XX	SWELLMAP1 (SMP1) - encodes a CCHC zinc finger protein
XV	gw1.XV.1016.1	XX	Thaumatin-like protein, putative
XV	grail3.0043013701	XX	Similar to unknown protein

Table 3-6: Continued…

XV	grail3.0005001801	XX	LOB - lateral organ domain protein
XV	estExt_fgenesh1_pg_v1.C_LG_XV0448	XX	DEL1 - E2F-like protein
XV	grail3.0005005601	XX	Similar to unknown protein
XV	fgenesh1_pg.C_LG_XV000528	XX	KNAT6 - required of SAM formation
XVI	grail3.0101001001	XX	UPL5 (UBIQUITIN PROTEIN LIGASE 5)
XVI	estExt_fgenesh4_pg.C_LG_XVI0075	XX	Vacuolar calcium-binding protein-related
XVI	gw1.XVI.1067.1	XX	DPE2 (DISPROPORTIONATING ENZYME 2)
XVI	grail3.0101010901	XX	Cysteine protease inhibitor, putative
XVIII	fgenesh1_pm.C_LG_XVIII000101	XX	CYCA2;1 - core cell cycle gene
XVIII	gw1.XVIII.1554.1	XX	Expansin family protein (EXPR3)
XVIII	estExt_fgenesh1_pm_v1.C_LG_XVIII0152	XX	KNAT3 - homeobox protein
XVIII	gw1.XVIII.2006.1	XX	Chloroplast nucleoid DNA-binding protein, putative
XVIII	estExt_Genewise1_v1.C_LG_XVIII2227	XX	AHA2 (Arabidopsis H(+)-ATPase 2); ATPase
XVIII	grail3.0020016401	XX	Histone H4
XVIII	eugene3.00180824	XX	Histone H4
XVIII	estExt_fgenesh4_pm.C_LG_XVIII0287	XX	Histone H4
XVIII	gw1.XVIII.3026.1	XX	TET8 (TETRASPANIN8)

3.8 Discussion

3.8.1 Comparative study of functional diversity between candidate gene selection approaches literature and microarray

Both literature and microarray approaches have shown huge potential towards the selection of candidate genes for leaf development. Microarray analysis resulted in a high through-put output of 293 potential candidate genes for leaf development from a possible twenty seven thousand spots on the array. On the other hand literature searches using a relatively standard key word search engines resulted in 79 potential candidate genes. There is a substantial difference between the two search engines, however if time was not a limitation and the parameters for literature searches were more loosely set, an even larger gene list could be possible. Literature searches resulted in a large number of candidate genes from *Arabidopsis* that have a known function in leaf development, therefore functional analysis resulted in a large proportion of genes involved in cellular and developmental processes (Figure 3-4A). On the other hand candidate genes selected from microarrays were obtained by a less biased approach enabling a wide variety of possible genes, resulting in a large amount of candidates involved in metabolite processes (Figure 3-4B & C). Firstly I would like to point out why *Arabidopsis* accession numbers are used and not popular gene models to investigate function. One reason is due to the limitation of similar bulk retrieval analysis in poplar databases, as the genome is relatively new and the annotations are not always correct. Secondly literature searches resulted in candidate genes found mostly in *Arabidopsis* and ESTs for the poplar microarray were designed with a homology to *Arabidopsis*, therefore to compare the two gene lists effectively a common denominator was needed. *Arabidopsis* and poplar also share a evolutionary path were they both started out with around 12,000 genes followed by two shared genome duplications, poplar however had a third genome duplication (Sterck *et al*., 2007). Comparative studies between poplar and *Arabidopsis* genes have showed 80% of genes to be identical (Douglas and Ehlting, 2005), therefore utilising homologies between species is effective.

3.8.2 Combining QTL experiments to indentify 'hotspots'

This study has resulted in several interesting regions worthy of further study in particular nine major hotspots for leaf and biomass traits on linkage groups I, IV, V, VI, VIII, IX, X, XIII, XIV, XV and XVIII (Figure 3-5). Other QTL studies in poplar such as a study by Wullschleger *et al*. (2005) showed QTL for leaf biomass to locate on linkage groups VI and IV and total biomass on linkage groups IV, XII and XIII. in a (*P.trichocarpa* x *P.delotides)* x

P.deltoides hybrid poplar backcross pedigree similar to the hotspots described in this study. QTL regions for leaf traits were also found in this study to be located on every linkage group, suggesting that loci involved in leaf development maybe involved in multiple developmental pathways (pleiotropy). This has been found in *Drosophilia* were genes affecting bristle development are also involved in the development of the central nervous system, sex determination, embryonic pattern formation and eye and wing development (Mackay, 2001). However this could also be a fault of the genetic map itself. In an ideal world a map should be constructed with as many molecular markers as possible equally spaced, but this cannot always be achieved due to the cost. Individual effects of QTL are small, the more QTL present in a population the smaller their individual contribution and the true position maybe in a range of 10-20cM and it is difficult to reduce confidence intervals. Confidence intervals can be misleading as they contain hundreds or thousands of genes, therefore to pinpoint one particular area is difficult. Ways to improve the QTL analysis includes increasing population size, increasing replicates, by reducing the environmental variation and combining analysis of several traits. Even with these pitfalls QTL analysis has been effective in identifying genes controlling phenotypic variation. For example in Rice, Massahiro Yano's group have isolated the genes for five heading date QTLs, genes Hd1 –Hd5 were found to be 0.5, 0.3, 0.0, 2.6 and 1.2cM from the original QTL identifying by single mapping (Yano *et al*., 2000). Wang *et al*.(2007) investigated defence response genes in maize by mapping to the maize genome using molecular markers and bioinformatics and found clustering of defence response genes in QTL regions.

3.8.3 Co-location of candidate genes to QTL

I have discussed the pitfalls of using just QTL regions to locate candidate genes and microarrays have been useful in obtaining a large amount of candidate genes. However they do have their downfalls, such as they are only as good as the number of genes spotted and in this case they are only as good as the number of leaf development genes spotted. Also they work on the assumption that genes that show genotype –species differences in their level of expression could be comparative agent for the variation in a trait (Morgante and Salamini, 2003). However this study combined several approaches making a powerful tool to understanding leaf development. Combining QTL, literature and microarray approaches has enabled the reduction of the candidate gene list to 86 possible candidates for leaf, biomass and cell phenotypic variation. Other studies have used a similar approach in poplar due to the genome sequence being widely available. For example (Frewen *et al*., 2000) combined QTL

analysis and candidate gene co-location to identify regions controlling bud set and bud flush. Rae *et al*. (2006) combined QTL analysis, microarray and candidate gene co-location to identify regions controlling leaf traits in elevated CO_2.

Co-location is only an estimate, there were many occasions when limitations were met due to the incompleteness of the biological databases and the software used. Annotations in biological databases are not always reliable, with multiple entries and unrelated genes bearing the same name; caution has to be exercised. It is also essential to use the appropriate BLASTing format; within this study we used tBLASTx, which is preferable when encoding DNA sequences as it uses nucleotide sequences as queries and translates them in all six reading frames to produce translated protein sequences which are used to query a protein sequence. Another cautionary note is within the QTL analysis and regions. QTL identified are based on a different species to the species under study; therefore it can only be used as a tool. Once QTL have been identified by linkage mapping it is difficult to determine a gene or genes that are responsible for the variation in phenotype. However it has become common in species with their genomes sequenced to place candidate genes onto QTL maps to look for areas of co-location. Co-location is a valued indicator of candidate genes involved in leaf development and finally combining genes indentified within this study and running association analysis will consummate our findings. The association study approach to identify genes involved in phenotypic differences is costly, so eight candidate genes were selected. The selection process consisted of genes found within major and minor QTL hotspots. Understanding how these genes directly affected our trait of interest was essential. Genes selected included:-

AS1 (ASYMMETRIC LEAVES 1) & AS2 (ASYMMETRIC LEAVES 2)

AS2 co-located to linkage group VIII between a major QTL hotspot for leaf and biomass traits (Figure 3-5H). *AS1* was found to co-locate to linkage group VI between markers O_026 and p_2578, a hotspot is not seen between these markers on either the leaf and biomass or cell trait co-location QTL analysis, (Figure 3-5F & F). Although *AS1* did not co-locate to hotspot QTL for either leaf, biomass or cell traits it was selected due to its relationship with *AS2* in the control of leaf initiation. *AS1* is a member of a small MYB-related gene family, which also contains *ROUGHT SHEATH2* (*RS2*) and *PHANTASTICA* (*PHAN*). Whereas *AS2* is a member of the *AS2* gene family also known as *LATERAL ORGAN BOUNDARIES* (*LOB*) domain gene family (Chalfun *et al*., 2005). Expressional studies have found these genes play a role in repressing *KNOX* genes in leaves to retain the differentiated state in the lateral organs

(Schneeberger *et al*., 1998). Other studies have also found *AS2* to be involved in lateral organ polarity and *AS1* and *AS2* to be positive regulators of the *LOB* gene involved in the establishment of boundaries between the meristem and the differentiated lateral organs (Naito *et al*., 2007).

ERECTA

ERECTA co-located to linkage group XVIII between markers p_2862 and p_2525, however this gene was not found within a major or minor hotspot QTL region for leaf, biomass or cell traits (Figure 3-5R & Figure 3-6 R), but was selected due to its function solely. *ERECTA* is the *Arabidopsis* receptor like kinase (RLK) which regulates inflorescence architecture (Shpak *et al*., 2004).Expressional studies have resulted in *ERECTA* being highly expressed in the SAM and developing leaves, making it an ideal candidate gene for leaf development. Loss of function mutations in *ERECTA* result in short lateral organs and internodes and reduced cell numbers in the cortex cell files (Shpak *et al*., 2004). It encodes a leucine rich receptor like serine/ threonine kinase (LRR-RLK) which is a subfamily of signalling receptors in plants, it is therefore suggested that *ERECTA* mediates cell-cell signals that sense and coordinate organ growth.

PHABULOSA

This co-location study placed *PHABULOSA* on linkage group I between markers p_575 and p_ 2385 within a minor QTL region (Figure 3-5A). *PHABULOSA* works in association with *PHAVOLUTA* to determine abaxial or adaxial polarity, both genes encode START domain HD-ZIP proteins (Fleming, 2003). Dominant mutations in *phabulosa* (*phb*) and *phavoluta* (*phv*) cause a dramatic transformation of abaxial leaf fates into adaxial leaf fates (McConnell *et al*., 2001). This is achieved by altering the sterol or lipid binding domains of *ATHB14* and *ATHB9,* which are members of a plant specific class homeodomain –leucine zipper (HD-ZIP) containing proteins.

ANGUSTIFOLIA (AN)

Co-location of *ANGUSTOFOLIA* placed it between markers CACTG 25 and p_571 on linkage group XIV (Figure 3-5), co-locating to a major hotspot QTL. *AN* is a unique gene as it is the first in the family of *Ct/BARS* like protein genes to be identified in a plant genome. It is involved in the regulation of polarized growth of leaf cells by controlling the arrangement of cortical microtubules (MTs) in epidermal cells (Kim *et al*., 2003). It is able to regulate polar

elongation in the leaf width direction making it an important gene in leaf size. Mutants result in altered leaf shape due to decreased lamina expansion (Folkers *et al*., 2002;Kim *et al*., 2003).

E2Fc

E2Fc collocated to the top of linkage group II between markers CCCTC 32 and p_ 667 (Figure 3-6B),a minor hotspot for cell traits. *E2Fc* was selected due to its known function in the cell cycle as this gave a border selection of candidate genes form all areas known to be involved in leaf size. The retinoblastoma (*RB*)-*E2F* pathway is one of the most important regulatory pathways that control and couple cell division and cell differentiation in both animals and plants (Stevaux and Dyson, 2002). *E2F* and *DP* proteins interact to form active transcription factors that bind to gene promoters to regulate the expression of genes required for cell cycle progression (del Pozo *et al*., 2006). The *E2F/DP* family in *Arabidopsis* is made up of six *E2F* and two *DP* proteins, which are divided into subfamilies due to their structure. Overexpressional studies of *E2Fc* proteins have shown delays in cell division and to represses the expression of S-phase genes (del Pozo *et al*., 2006).

ACC_oxidase (ACO)

ACO co-located to minor hotspots for both leaf and biomass and cell traits on linkage group XI between markers p_204 and p_2392 (Figure 3-5K & Figure 3-6K). *ACO* is involved biosynthetic pathway of the hormone ethylene in plants. The pathway begins with the conversion of S-adenosylmethionine to 1-aminocyclopropane-1-carboxylate (*ACC*); this reaction is catalyzed by *ACC* synthase. *ACC* is then converted to ethylene in a reaction catalyzed by *ACC* oxidase (*ACO*). Three *ACO* genes have been shown to be expressed differentially during leaf development in white clover (Trifolium repens L). This includes *TR-ACO1* expressed in the apical structure of the stolon, *TR-ACO2* in newly initiated leaves and *TR-ACO3* is expressed in senescent leaf tissue (Chen and McManus, 2006), making *ACO* a very interesting gene for leaf development.

LEAFY

LEAFY collocated on linkage group XV between markers o_430 and p_ 2585 to a major QTL for leaf and biomass traits (Figure 3-5O). *LEAFY* is a floral meristem identity gene in *Arabidopsis*, its role in the vegetative to reproductive phase transition has been increasing in interest (Rottmann *et al*., 2000). *LEAFY* has been isolated in *Populus trichocarpa* which affected the inflorescence to floral meristem transition (Coen *et al*., 1990). Studies have shown

over expression of *LEAFY* within Poplar have resulted in a hybrid cottonwood 184-402 line 31 to produce smaller and deformed leaves, whereas hybrid aspen developed bushier growth with significantly smaller leaves (Rottmann *et al*., 2000) , indicating *LEAFY* to be an important gene for leaf development in poplar.

3.8.4 Summary

Using parallel approaches; microarray analysis and literature searches resulted in two hundred and ninety six candidate genes for leaf development, with further accumulation of techniques such as collocation to QTL enabled the reduction for this list to eighty six. I have shown that this approach is therefore effective in increasing the capacity to select candidate genes. The candidate genes; *ASYMMETRIC LEAVES 1* (*AS1*), *ASYMMETRIC LEAVES 2* (*AS2*), *ERECTA*, *PHABULOSA*, *ANGUSTOFOLIA*, *E2Fc*, *ACC oxidase* (*ACO*) and *LEAFY* have been selected for the association study described in chapter 4.

4 . A comparison of neutral markers, candidate genes and phenotypic traits using association genetics in *Populus nigra.*

4.1 Overview

Recently association tests have become a powerful technique for dissecting complex adaptive traits in humans and plants. Association mapping based on Linkage Disequilibrium (LD) offers an alternative method to QTL mapping as it offers; fine scale mapping due to historical recombination, investigation of phenotypic and genetic variation in a single experiment and multiple markers associations to be discovered in a single experiment. This study identified candidate genes for leaf development through microarray analysis and literature searches, as described in 3.7.3. Also due to the publication of the *Populus trichocarpa* sequence and previous QTL analysis undertaken in Taylor lab, candidate genes can now be co-located to QTL regions, identifying 'hotspots' for leaf development and biomass. Within chapter 3.8.3 eight candidate genes for leaf development were selected for association genetics based on co-location and known function.

In this chapter I attempt to identify genes controlling leaf development using fluorescence polarization technologies and software packages TASSEL (Yu *et al.*, 2006). SNP genotyping using fluorescence polarization detection will be described and advantages and disadvantages discussed.

4.2 Introduction

A new revolution in genetical genomics is in full force due to the increased efficiency of technology involved in the direct sequencing of Polymerase Chain Reaction (PCR) products. High throughput sequencing has enabled rapid Single Nucleotide Polymorphisms (SNP) detection. SNPs are a single difference in base pair from the common form, which constitute most variation in a population (Roelofs *et al*., 2008) Limitations in QTL mapping at present, using segregating families, have been that population sizes are too small, with only two alleles at a locus sampled (Gupta *et al*., 2005) and complex ecosystems with long- lived species such as trees, especially hardwoods, are impossible to map as it would take too long to produce the population (Rafalski, 2002). Association mapping using Linkage Disequilibrium (LD) has been proposed to overcome these limitations. Association mapping refers to a significant association of molecular markers with a phenotypic trait, whereas LD refers to the non-random association between two alleles (Gupta *et al*., 2005). Association approaches were first used effectively in human genetics; it provides new benefits such as; the ability to examine unrelated individuals (i.e. a natural population), higher resolutions produced (depending on LD), larger number of alleles per locus can be tested and relatively rapid results due to use of natural populations (Buckler and Thornsberry, 2002).

There are two applications of association analysis: genome scans and the candidate gene approaches. In genome scans SNP markers are placed across the genome at an appropriate density, whereas candidate-gene approaches involves sequencing candidate genes (Flint-Garcia *et al*., 2005). Success of either of these methods depends on the degree of LD and population size. Genome scans are appropriate for species with a moderate to extensive LD (Flint-Garcia *et al*., 2005) because species with low LD need many markers to cover the genome, thus a candidate gene approach is more useful for species with low LD. Association analysis consists of five basic components: germplasm choice, estimation of population structure, trait evaluation, identification of candidate polymorphisms and statistical analysis (Flint-Garcia *et al*., 2005) .Figure 4-2 shows the steps taken within this study.

Germplasm

Populus was chosen as the germplasm of study because it has shown a great deal of phenotypic variation from previous studies (Tuskan *et al*., 2004;Marron and Ceulemans, 2006) and within this study (Chapter 2). The next step in an association study is to quantify LD within your species of interest, as this will determine the approach. LD is defined by Flint-

Garcia *et al*. (2003) as the "non-random association of alleles at different loci.". It is due to linkage and is the net result of all the recombination events that occurred since the origin of an allele by mutation, thus providing a lower opportunity for recombination to take place between any two closely linked loci (Gupta *et al*., 2005). There are several measures of LD, but D' and LD-r^2 are commonly used, these statistics simply indicate the difference between the observed and expected haplotype frequencies. Figure 4-1 shows the behaviour of D' and LD-r^2. D' takes into account only recombinational history whereas LD-r^2 involves recombinational and mutational histories. Figure 4-1 shows three different scenarios of how linked polymorphisms can show different LD results. Figure 4-1A represents absolute LD when two polymorphisms are completely correlated, due to linked mutations occurring at a similar point in time, with no recombination. Figure 4-1B shows polymorphisms that are not completely correlated, with no evidence of recombination, which could be due to mutations on different allelic lineages, therefore different mutation histories but the same recombination history, resulting in low values of LD-r^2. Finally, Figure 4-1C illustrates linkage equilibrium, where the two sites are unlinked because of recombinational events.

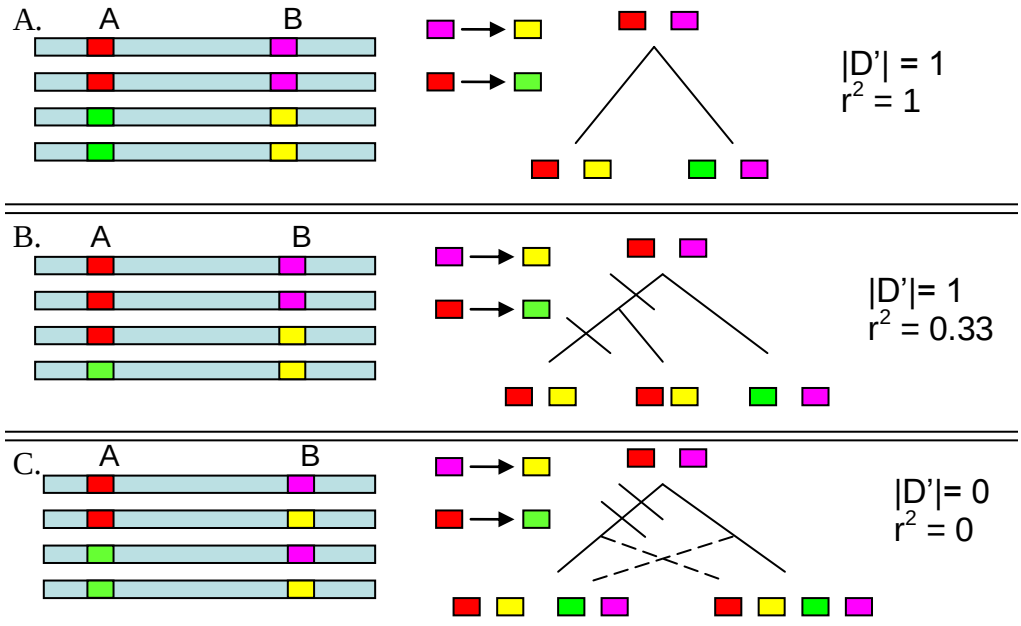


Figure 4-1: Linkage disequilibrium explained. Hypothetical scenarios of linkage disequilibrium (LD) between linked polymorphisms caused by different mutational and recombination histories illustrating the behaviour of D' and LD-r^2 statistics, adapted from Flint-Garcia *et al*., 2003. Blue bars to the left represent individuals and colours represent allelic state (red- A, green -T, pink - G and yellow - C), whereas the right represents possible evolutionary lineages causing differences in LD (solid lines represent mutation events whereas dash lines represent recombination events). (A) Absolute LD – occurs when two loci share similar mutational histories with no recombination. (B) Mutations occur on different lineages without recombination between loci and (C) Linkage equilibrium produced when there is recombination between loci, regardless of mutational history.

Most processes in population genetics affect LD, such as recombination rate, mating systems, genetic isolation, population subdivision, population admixture, natural and artificial selection, population size, mutation rate and stochastic effects (chance) (Rafalski and Morgante, 2004). Taking population subdivision as an example, LD increases because, in small populations, genetic drift causes a consistent loss of rare allelic combinations. In the context of association genetics, when LD is low, a whole genome scan is impractical as many markers are needed and therefore a candidate gene approach is suitable. Collaborations with Udine University, Italy, have enabled us to obtain LD measurements for *P.nigra* which have revealed the most suitable approach for the association study. Zaina & Morgante., (2008 unpublished data) assessed genetic diversity of 31 loci from 12 different genotypes of *P.nigra*. They found that SNPs occurred on average every 122 bp (1 SNP occurred on average every 102 bp in non-coding regions and every 195 bp in coding regions) and that LD decayed rapidly over large distances of 100 kb (LD-r^2 = 0.09), suggesting a candidate gene approach is most appropriate in this species. This consistent with by a recent study by Ingvarsson *et al*. (2006), in which 24 trees of *P.tremula* from four different sites in Europe were selected to study LD in five loci. LD rapidly decayed, LD-r^2 <0.05 in <500 bp (Ingvarsson *et al*., 2006) supporting a candidate gene approach in poplar (Cheng *et al*., 2006)

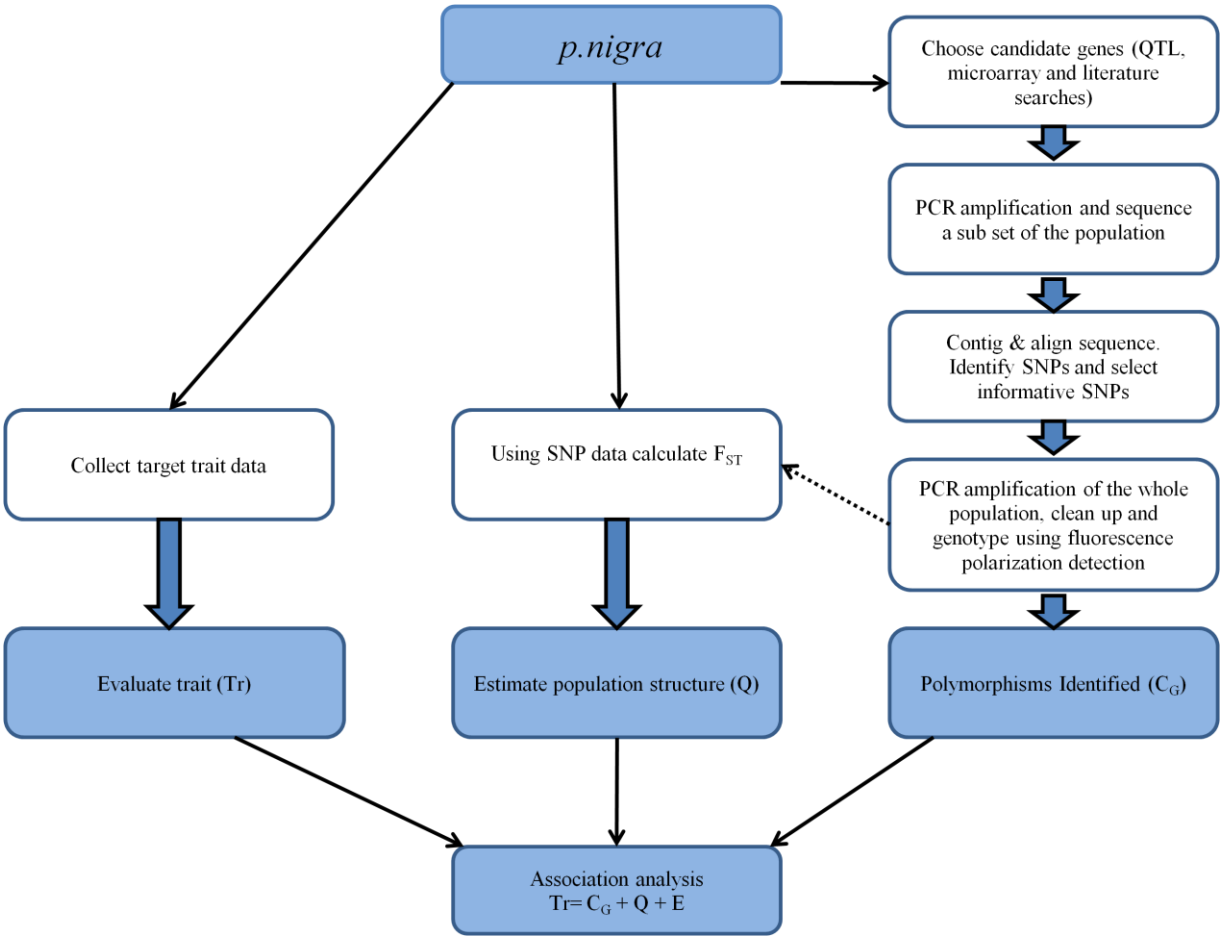


Figure 4-2 : How to carry out an association study. Flow chart adapted from Flint-Garcia *et al.* (2005), illustrating the steps involved in association mapping. Each box represents a step in a association study, key steps in all studies are filled by blue, such as selection of germplasm (*P.nigra*) based on linkage disequilibrium, evaluation of phenotypic traits (Tr), estimation of population structure (Q), identification of genetic variation (C_G) and finally combining Q, C_G and error (E) to find out the genetic bases controlling phenotypic variation (Tr).

Population structure & trait heritability

Population structure can cause artificial results in association studies therefore it is has to be taken into account. If a population is divided into subpopulations less heterozygosity (many genes which cause the same trait) occurs. Therefore Wrights (1951) developed a system describing the properties of hierarchically subdivided natural population termed F-statistics. F-statistic are also known as fixation indices which estimate the differences in allele frequencies due to subdivision within a population (Wright, 1949). F_{ST} measures the difference between the mean heterozygosity among the subdivisions in a population, and the potential frequency of heterozygotes if all members of the population mixed freely and non-assortatively (Weir and Hill, 2002). Wright's model for F_{ST} ranges from zero to a theoretical maximum of one, zero indicates no differentiation between the subpopulations in the overall populations (Weir and Cockerham, 1984).

Complementary to Wrights F-statistic (Spitze, 1993) derived a statistic for quantitative traits equivalent to F_{ST} termed Q_{ST}, which is the comparison between genetic differentiation at quantitative traits (Spitze, 1993). When Q_{ST} and F_{ST} differ significantly from each other, genetic differentiation among populations in quantitative traits cannot solely be explained by drift, and therefore it is hypothesised that natural selection is the driving force. When Q_{ST} is greater than F_{ST} directional selection is assumed and when Q_{ST} is less than F_{ST} a uniform selection process is taking place over the populations.

Another approach to understanding population structure was devised by Pritchard and Rosenberg. (1999),who considered how genetic information could be used to detect the presence of 'cryptic' population structure i.e. population structure that is difficult to detect using visible characters (Pritchard *et al*., 2000). The method identifies actual subpopulations and assigns individuals to these populations. A Bayesian clustering approach is taken in which there are K populations (where K maybe unknown), each of which is characterised by a set of allele frequencies at each locus (Pritchard *et al*., 2000). Individuals are assigned to a cluster based on their genotypes, while simultaneously estimating population allele frequencies (Pritchard *et al*., 2000). This approach makes two assumptions; one is that markers (microsatellites or SNPs) are neutral and the second is that there is Hardy-Weinberg equilibrium within populations. Two types of clustering methods are used; distance based methods and model-based methods. Distance-based methods calculates a pairwise distance matrix, whose entries give the distance between every pair of individuals and clusters are assigned by eye. A model-based method assumes observations from each cluster are randomly drawn from some parametric model. Inference for the parameters corresponding to each cluster is then done jointly with inference for cluster membership of each individual using either maximum-likelihood or Bayesian method (Pritchard *et al*., 2000).

Both F_{ST}, Q_{ST} and K were used to estimate population structure in this study. Traits chosen for association studies should have high heritability scores or else a large number of replicates are needed to get meaningful trait data (Flint-Garcia *et al*., 2005). Chapter two shows the broad sense heritability scores (Falconer and Mackay, 1996) for all traits, which are moderate to high. Therefore, all traits are suitable for association analysis.

SNP genotyping and association analysis

In this study I have chosen a multi-disciplinary approach to select candidate genes. This is described in chapter three and results are shown. The candidate gene sequences were obtained

from JGI and overlapping primers pairs designed to amplify both non-coding and coding regions of the gene. Candidate genes in a sub set of the population were amplified using PCR, purified and then sequenced using capillary electrophoresis. Sequences are then assessed using the software Poly Phred Phrap Consed (Gordon, 2003) and informative SNPs identified. Amplification and purification was then repeated using the whole population, SNPs are genotyped using fluorescence polarization detection assigned to each individual.

Finally the association analysis to identify trait-marker relationships was performed. The type III sums of squares were tested using the generalized linear model (GLM) (equation 4-1) (Flint-Garcia *et al*., 2005).

$$Tr = C_G + Q + \varepsilon \quad (4\text{-}1)$$

Where Tr is the trait, C_G is the genotype of the candidate gene, Q is population structure and ε is error. Permutation analysis is used to determine significance thresholds (Flint-Garcia *et al*., 2005).

Examples of successful association studies include an experiment by Palaisa *et al*. (2003) where 78 out of 81 informative SNPs and Indel polymorphisms in the Y1 gene were found associated with endosperm colour when genotyped over a set of 41 yellow or orange endosperm lines and 34 white endosperm lines. Also, the most recognised example is likely to be the association found with flowering time and the *dwarf8* (*d8*) gene in maize in which they used the test statistic Λ for quantitative traits, as modified by Pritchard *et al*. (2000), to find significant associations (Thornsberry *et al*., 2001).

4.2.1 SNP Genotyping with Fluorescence Polarization Detection

SNPs are potential markers in association studies and LD mapping and the usefulness of SNPs is determined by the extent of LD between them. Most studies have been done so far in humans, however, SNPs can be used to identify genes involved in complex traits in any species. SNPs are mostly bi-allelic, DNA sequence variations (polymorphisms) which occur when a single nucleotide base, Adenine (A), Thymine (T), Cytosine (C), or Guanine (G) in a genome is altered. SNPs can occur in coding and non-coding regions of a genome so that some affect function whereas others do not. SNPs are also evolutionarily stable so are a great tool in population studies; they can be used in identifying multiple genes associated with complex traits. Results from the Human Genome Project showed that SNPs account for 90 % of all human genetic variation occurring on average every 100-300 bp along the three billion base pair human genome (Lander *et al*., 2001). The hope is that SNPs can be used to identify

multiple genes associated with complex disease traits such as cancer, diabetes and mental illness. The most direct approach to the discovery of DNA polymorphisms is by direct sequencing of PCR products, which is normally achieved best when individuals are homozygous so that individual polymorphisms are easy to detect and haplotypes (two or more SNPs found together) can be determined (Rafalski, 2002). Rafalski, (2002), by direct sequencing of PCR products, has found the frequency of polymorphisms in the US maize elite germplasm to be very high, on average 1 SNP per 48 bp in non-coding regions and 1 SNP per 131 bp in the coding regions (Bhattramakki *et al*., 2002). SNP markers have huge application possibilities as they can be used in constructing high resolution genetic maps, mapping traits, genetics diagnostics, analysis of the genetic structure of populations, phylogentic analysis and association studies based on linkage disequilibrium.

Within this study SNPs were detected using Template directed Dye terminator Incorporation with Florescence Polarization detection (TDI-FP). A thermostable polymerase adds one or two fluorescent dye labelled terminators to an oligonucleotide primer that ends immediately upstream of the SNP position. Which dye labelled terminator (Acyclo Terminators ^{TM} and AcycloPol ^{TM}) is incorporated will determine the genotype for each individual. The identity of the base added is determined by the increased fluorescence polarization (FP) of its linked dye. FP is an empirical technique that measures the vertical and horizontal component of the fluorescence emitted after dye excitation by plane polarized light. FP is proportional to the molecules rotational relaxation time (the time it takes to rotate through at an angle of 68.5^{o}) (Chen *et al*., 1999). FP is directly proportional to the molecular volume, which is directly proportional to the molecular weight. For example if a fluorescent molecule is large it rotates and tumbles more slowly in space and FP is preserved the molecule is small so it rotates and tumbles faster and FP is largely lost (Chen *et al*., 1999) (Figure 4-3).

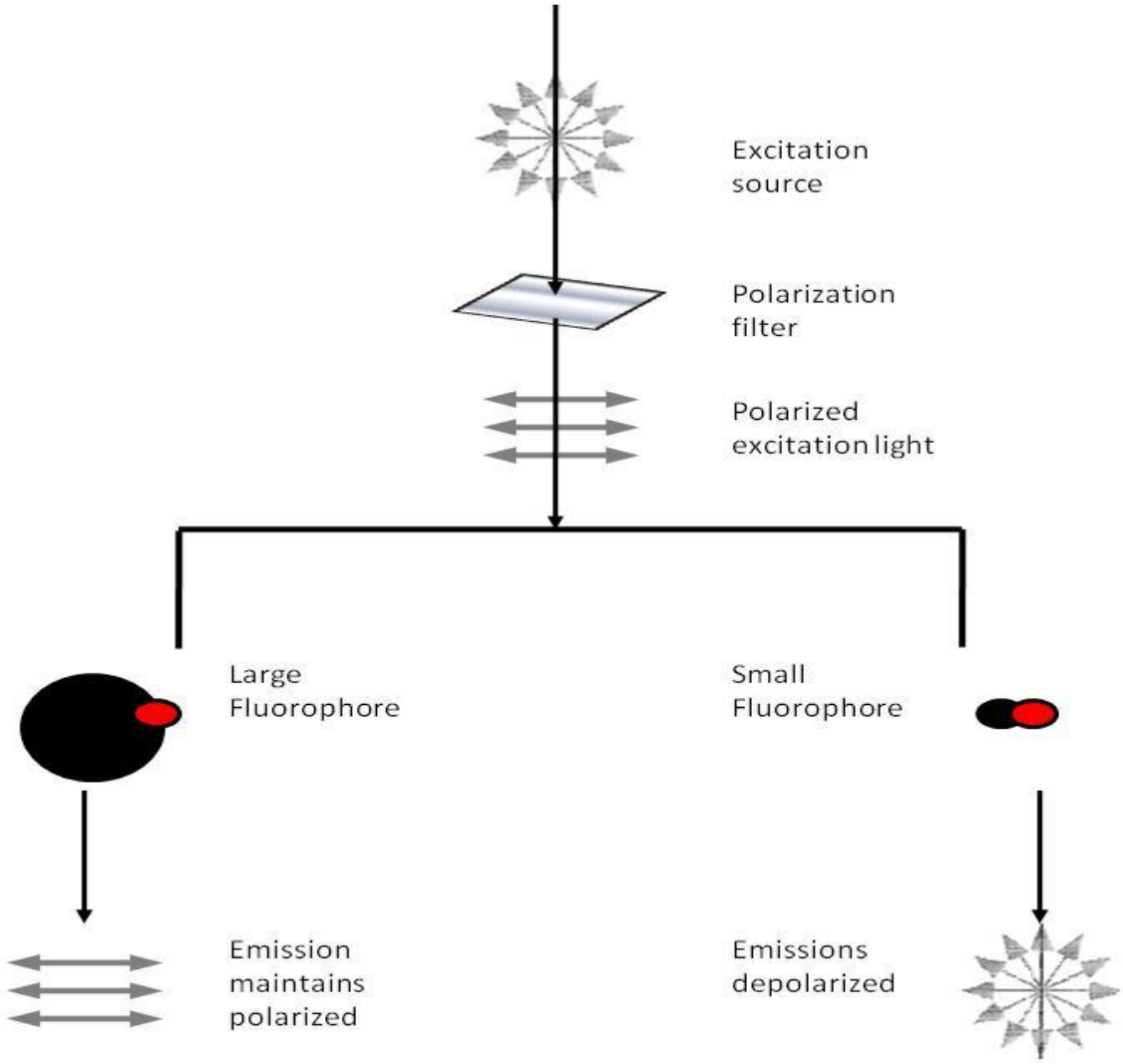


Figure 4-3: Fluorescence polarization (FP) adapted from Chen, Levine *et al*. 1999. Illustrates observations when a large (left hand side) or small (right hand side) molecule (black circle) is attached to a fluorescent dye (red circle) when excited by plane-polarized light (grey arrows) when viscosity and temperature are held constant. Large molecules rotate and slowly and therefore FP remains the same, whereas small molecules rotate faster and therefore FP is lost, therefore a distinction between the molecules can be made.

The incorporation of a fluorescent Acyclo Terminator into an oligonucleotide primer increases its polarization. This increase is used to determine which of the two labelled terminators present have been incorporated. Figure 4-4 shows an example of a candidate gene containing a G/T SNP, were two florescent acyclo terminator dyes; R110 and TAMRA are attached, R110 is incorporated to the base pair G whereas TAMRA is incorporated to T. Total fluorescent signal is then determined by the ratio between the incorporated and unincorporated labelled terminator and SNPs are scored according to the clustering of genotypic groups (Gonzalez-Martinez *et al*., 2007) (Figure 4-4).

In this chapter I compare population differentiation in both putatively neutral and candidate gene SNPs and quantitative traits related to phenology in *P.nigra* using F_{ST}, STRCTURE and Q_{ST}. Eight selected candidate genes involved in leaf development (*AS1, AS2, ERECTA, PHABULOSA, AN, E2Fc, ACC oxidase* and *LEAFY*) were tested for associations with leaf, cell and biomass traits using both structured and non structured association methods.

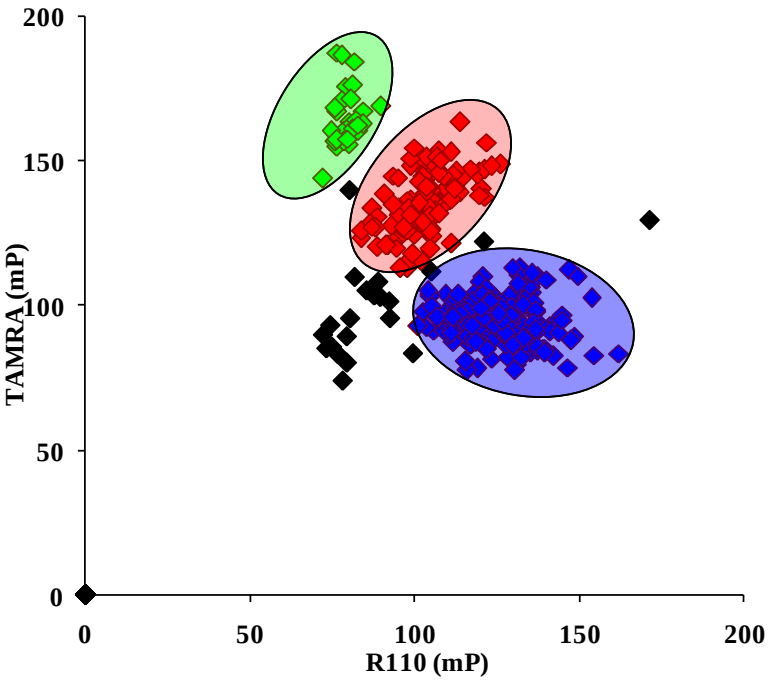


Figure 4-4: A schematic representation of genotyping using florescence polarization technologies. Illustrates schematically the incorporation of a two florescent acyclo terminator dyes; R110 and TAMRA to a candidate gene with the SNP G/T. R110 is incorporated to the base pair G whereas TAMRA is attached to T. Total fluorescent signal is determined by the ratio between the incorporated and unincorporated labeled terminator, shown on the graph on the right hand side. When only TAMRA is detected individuals are homozygous T/T clustering in the top left hand corner of the graph (green), whereas homozygous G/G is when only R110 is incorporated and individuals cluster to the bottom right hand corner (blue), heterozygous G/T individuals incorporate both TAMRA and R110 and cluster in the middle (red).

4.3 Material & Methods

4.3.1 Plant Material

One individual representing each genotype was selected from the experimental plot in Geraardsbergen (2.3.1). Leaves were sampled by the Institute of Forestry & Game management in Geraardsbergen. Several leaves were collected into a pre-labelled plastic envelope and placed immediately on ice, then taken back to the lab where they were flash-frozen into liquid Nitrogen (N_2), vacuum dried and stored at room temperature in the dark.

4.3.2 DNA Extraction

A small piece of lyophilised leaf material was placed in an 1.5ml Eppendorf tube filled with liquid N_2 and ground fast using a small pestle. N_2 was allowed to evaporate off and samples were stored in the -80^{o}C for later use or used immediately. A 2% hexadecyltrimethylammonium bromide (CTAB) extraction buffer was prepared as described by Doyle *et al*. (1990). 900µl of pre-warmed (65^{o}C) CTAB was added to the ground material. Samples were then vortexed and incubated at 65^{o}C for 1 hour. After incubation 900µl of CHISM (chloroform: isoamyl alcohol, 24:1) was added and thoroughly mixed. The aqueous and organic phases were separated following centrifugation at 10000xg (maximum speed) for 10 minutes at room temperature (RT). The upper aqueous phase was recovered and transferred to a new Eppendorf. Sodium acetate (3M, pH 5.2) was added at 1/10 volume aqueous phase and cold isopropanol at 2/3 of the final volume, mixed well and precipitated at -20^{o}C for 30 minutes. After incubation samples were centrifuged for 10 minutes at maximum speed, followed by discarding of the supernatant (isopropanol). The pellet was then washed by adding 500µl cold 70% ethanol, mixing and centrifuging at maximum speed for 10 minutes. Ethanol is then discarded by pipetting out the liquid phase and leaving at RT for 20 minutes. Ribonucleic Acid (RNA) is then digested by adding 50µl of TE solution and 1µl of RNAse 1mg/ml and leaving at room temperature overnight. Following incubation, 200µl of TE solution, 100µl sodium acetate (3M, pH 5.2) and 1ml absolute ethanol is added, mixed and centrifuged for 10 minutes at maximum speed. The liquid phase is discarded and pellet dried at RT for 20 minutes. The pellet is re-suspended in 50µl sterile water and stored at -20^{o}C. DNA quality and quantification checked with a 1µl Ethidium Bromide 1% agarose gel, the protocol yielded 20ng of DNA per 1 µl.

4.3.3 SNP discovery and selection

4.3.4 Candidate gene selection

Eight candidate genes were selected for this study from literature searches, transcriptome analysis and co-location to QTL (Table 4-1). Genomic sequences for each gene were obtained as described previously in Literature Searches. Unique sequences of each family member were obtained using ClustralW alignment (Pearson and Lipman, 1988) and termed fragments, this resulted in eight fragments. Specific primers pairs for PCR were designed using PRIMER3 software (Rozen and Skaletsky, 2000) for each fragment. The expected size of the amplification fragments was chosen to range from 400 to 800 bp (Table 4-1).

Table 4-1: Primers designed for SNP detection. Primers used in PCR reaction, shows both the forward (Sense) and reverse (antisense) primer for each candidate gene (Italics)

Gene	Sense primer (5' to 3')	Antisense primer (5' to 3')
AS1	GCCCTTCTCCAGCATTTGTA	AGAGCAACGACCCAGTGTTT
AS2	CGGTTTTTCATTGCATTTCA	CGAAGGCAAATGAACGTTTT
ERECTA	AGCCAAAGTTGGTTATCCTTCA	TAGAACAATCCCGTAGCTGACA
PHABULOSA	TGGATTTCGTGTCATACCTTTG	TTCCAGGCACTAAACCATTTTT
AN	TCACACCGCCGTAACACTAA	TGCGTGAATAAGTACCTCGTG
E2Fc	GAGTTCAGGCTTGAGGAAGCTA	TTCCAAGCACTTGTATCCATTG
ACC oxidase	TGGTCAAGTCCCAAAGAAAGA	GGTTCCGTAAGACAATTGAAAA
LEAFY	CTGGTCCCTGTCATTTTAGACC	ACCATCCTCCATCAACCAATAC

4.3.5 PCR

DNA amplification was performed in a 20 µl volume for 24 randomly selected individuals per fragment . The reaction contained 1 µl genomic DNA, 10 µl 2xBiomix (Bioline Ltd, London, England), 1 µl of each primer forward and reverse at 10µM concentration (see Table 4-1 for sequences) and 7 µl dH_2O. The reactions were performed in the GeneAmp 9700 PCR system (Applied Biosystems, Foaster City, CA) under the following conditions: 95^{o}C for 2 min, 40 cycles of 30 sec at 95^{o}C, 1 min at 55^{o}C and 2 min at 72^{o}C, followed by a final extension of 7 min at 72^{o}C. PCR products were analysed on an agarose gel and purified using the PCR 96 cleanup plates (Millipore, Bedford, MA) with a vacuum manifold, according to the manufacturer's instructions.

4.3.6 SNP detection and statistical tests

Purified PCR products were sequenced directly on the forward and reverse strand using locus specific primers and the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit v3.1, and then separated on an ABI3730 sequencer. The output sequences of each gene for all

genotypes were aligned and visualised using Phred-Phrap-PolyPhred-Consed programs (Nickerson *et al*., 1997; Gordon *et al*., 1998;Gordon *et al*., 2001). Phred- Phrap- Consed is used to check quality of sequences and number of genotypes successfully sequenced. Additionally, to verify sequences, consensus sequences were Blasted against their corresponding fragment in NCBI ((National Library of Medicine, USA) using tBLASTx and then again in JGI (Tuskan *et al*., 2006a). SNPs and insertion or deletion polymorphisms (indels) were identified directly by visual inspection using PolyPhred detection and results were recorded into an excel spreadsheet. Minor allele frequency was calculated as to determine informative SNPs used in association analysis; by equation 4-2.

$$\text{Minor allele frequency (MAF)} = \frac{\sum minor\ alleles}{\sum alleles} \quad (4\text{-}2)$$

An informative SNP was defined as having a frequency to 0.1. The Linkage disequilibrium (LD) measures LD-r^2 and the Likelihood ratio were calculated with LDA analyzer (Ding *et al*., 2003) for each fragment. SNPs having an LD-r^2 of 1 are in complete LD, therefore convey the same information and therefore only one will be used for genotyping the whole population.

4.3.7 SNP Genotyping using fluorescence polarization

4.3.8 PCR & Primer Design

DNA amplification was performed in a 20 µl volume for the whole population for each fragment. The reaction contained 10 ng genomic DNA, 1 µl AmpliTaq Gold Buffer , 0.2 µl of each primer forward and reverse at 10µM concentration, 0.6 µl $MgCl^2$ (1.5mM), 0.5 µl dNTP (125 µM), 0.2 µl DMSO (2%), 0.1 µl AmpliTaq Gold (Applied Biosystems, Foster City, CA) and 5.2 µl dH_2O. The reactions were performed in the GeneAmp 9700 PCR system (Applied Biosystems, Foaster City, CA) under the following conditions: 95^{o}C for 10 min, 40 cycles of 15 sec at 95^{o}C, 30 sec at 55^{o}C and 2 min at 72^{o}C, followed by a final extension of 7 min at 72 ^{o}C. PCR products were analysed on an agarose gel and purified using the PCR 96 cleanup plates (Millipore, Bedford, MA) with a vacuum manifold, according to the manufacturer's instructions. Specific primers for Single Base extension (SBE) were designed using PRIMER3 software (Rozen and Skaletsky, 2000) for each SNP (Table 4-2). Primers were designed within the immediate 30bps upstream or downstream of the SNP. Operon bioinformatics tools (Operon Biotechnologies, Alabama, USA) were used to avoid primers that formed dimmers and to design primers at a melting temperature >50^{o}C.

Table 4-2: Primers designed for each SNP detected in 4.2.6. Primers used in Single Base Extension (SBE) technologies to genotype individuals within the association population.

Fragment name	Primer Type	SNP	Sequence
ERECTA_471	Sense	C/G	ATTATATGGAAAATGGAAGTCTGT
ACC01_261L	Antisense	C/A	ATCCCATGGTTCACCAACTTTA
ACC01_261R	Sense	G/T	TTGTTGATGTCTACTACTACTC
LEAFY_163	Sense	C/A	CTCCGTCAACTTTGTTATGC
LEAFY_232	Sense	G/T	CCGAGGATGAGCATTGAAAAT
LEAFY_352	Sense	G/A	CTCCTAAGTGCATTGGATGC
AS1_274	Sense	A/G	CTAAGCCTCTCTCCATCAAC
AS1_718	Sense	A/G	GTCTCACTAAGTTTCTCGAACA
AS2_496	Sense	A/G	ATAAACCGGGTCACGGAG
PHAB_609L	Antisense	A/G	GCTTTAAAATAGCAAAATCAAAATCA
PHAB_609R	Sense	T/C	TAGACATCGACTAAAACTACACCA
PHAB_618	Sense	A/C	ACATCGACTAAAACTACACCA
PHAB_665	Sense	C/T	TACTGAAATAGCTTGTTATTGTCC
ANT_175	Sense	C/T	TTAATCGAACACGTCCCTCT
E2Fc_503	Sense	C/A	CATTTTGTATAGTGGAGAACCAAT

4.3.9 Clean-up PCR product & SNP genotyping

To clean-up PCR products 4 µl of PCR product was transferred into a microplate designed for the fluorescence polarization reader (black in colour). Clean-up was performed with AcycloPrime^{TM} II SNP detection kit (Perkin-Elmer, Torrance, CA). For each reaction 0.2 µl PCR clean-up reagent (10x), 1.65 µl PCR clean-up dilution buffer and 0.15 µl PPase reagent was added to make a 6 µl final volume. Incubate at 37^{o}C for 60 minutes to degrade dNTPs, PCR primers and pyrophosphates in the solution. PCR Clean-up reagents enzymes were inactivated by heating at 80^{o}C for 15 minutes using GeneAmp 9700 PCR system (Applied Biosystems, Foaster City, CA). Store plate in 4^{o}C fridge or alternatively start SBE.Prepare SNP primers for SBE by dissolving in 100µM TE buffer (0.1mM EDTA, 10mM Tris-HCL, pH 7.6) for storage. Then dilute stock 10-fold with sterile water to give a concentration of 10µM. Add 3.05 µl AcycloPrime Mix (0.05 µl AcycloPol, 2 µl 10x reaction buffer and 1 µl Acyclo Terminator mix), 0.5 µl primer and 10.25 µl dH_2O to the clean-up plate making a final volume of 20 µl. Perform reactions in a GeneAmp 9700 PCR system (Applied Biosystems, Foaster City, CA) under the following conditions; denature at 95^{o}C for 3 minutes, 25 cycles of 95^{o}C for 15 sec and 55^{o}C for 30 sec. Bring samples to RT and read on a Victor^{2}-Wallac SNP genotyping platform for 25 cycles. Five extra cycles were performed if an insufficient difference was seen between negative controls and samples. Florescence Polarization (FP) was then calculated by instrument software using equation 4-3.

$$mP = 1000\ x\ \frac{[I_{vv} - I_{vh}]}{[I_{vv} + I_{vh}]} \quad (4\text{-}3)$$

Where I_{VV} is the emission intensity, measured when the excitation and emission polarizer's are parallel and I_{vh} is the emission intensity, when the emission and excitation polarizer's are oriented perpendicular to each other. A plot of the polarization of TAMRA versus the polarization of R110 showed four distinct data clusters (two homozygotes, one heterozygote's and one negative control); SNPs were scored according to these clusters. Positive controls from the SNP discovery experiment were used to check scoring.

4.3.10 Population Structure

The level of population structure in *P.nigra* association collections was measured by two separate tools using the same input data. To test population structure, all clones were genotyped by 29 putatively neutral SNPs, that did not show strong associations to leaf, cell or biomass traits. Neutral markers were provided by the University of Udine, who carried out SNP detection and genotyping on the *P.nigra* samples. The first test of population structure was done using FSTAT (version 2.9.3.2 (Goudet *et al*., 2002)to estimate F_{ST} for each SNP independently and overall all loci.

Complementary to F_{ST} , Q_{ST} was estimated for quantitative traits from the common garden experiment described in chapter 2, to estimate variance among clones within a population and among populations using a general linear model formula (4-4), in Minitab. Where Z_{ijkl} is the phenotype of the lth individual in the kth block from the jth clone from the ith population taken from Hall *et al*., (2007). In equation μ is the grand mean and ε_{ijkl} is the residual error term. Q_{ST} can then be calculated as shown in equation 4-5 as population (α_i) and clone (β_{ij}) effects provide estimates of variance between (σ_B^2) and variance within (σ_w^2).

$$Z_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \varepsilon_{ijkl} \quad (4\text{-}4)$$

$$QST = \frac{\sigma_B^2}{\sigma_B^2 + 2\sigma_w^2} \quad (4\text{-}5)$$

In addition population structure was investigated using a model-based clustering algorithm (STRUCTURE software; (Pritchard *et al*., 2000), with populations assigned as in Phenotypic statistical analysis summarised in Table 4-3. Population 7 and population 6 were not included in model due to small sample size. Models were run with a putative number of clusters (K parameter) from one to twenty one, non-correlated allele frequencies, 50,000 burn-in , to

minimize the effect of the starting configuration, and run-length periods of 500 000 interactions.

Table 4-3: *P.nigra* populations. Populations were based on river system over a population gradient from southern Spain to the Netherlands. *P.nigra* genotypes were derived from collections from EUROPOP and INRA and planted in 2004 in a common garden experiment in Belgium.

Population Number	Population Name	Location [a]	Number of genotypes
1	Loire Est	47°09'15.54"N, 2°36'00.00"E	8
2	Loire WO	46°43'05.83"N, 0°09'22.74"E	8
3	Drome 1	44°25'01.34"N, 5°16'20.80"E	47
4	Drome 6	44°27'14.57"N, 4°33'11.46"E	51
5	Individual Clone F	45°12'58.90"N, 2°52'46.39"E	4
6	Durance	43°28'14.41"N, 5°18'01.70"E	1
7	Ebro1	41°33'19.90"N, 1°12'40.27"E	0
8	Ebro2	41°21'09.13"N, 0°26'28.85"E	16
9	Rhine	49°29'34.56"N, 8°17'57.13"E	48
10	Ticino (N)	45°10'18.54"N, 8°34'51.63"E	46
11	Ticino (SN)	45°07'38.05"N, 9°02'08.05"E	39
12	Netherlands	51°48'01.07"N, 5°23'09.29"E	49

a Median value are given if the original data give only the range

4.3.11 Association analysis

Association analysis was performed using the software package TASSEL (Yu *et al*., 2006), on informative SNP loci. Phenotype data used in the association analysis can be seen in Chapter 2.3.1. A separate analysis was performed on each trait using least squares analysis according to the General Linear Model (Σ-restricted model) tool provided by the software. Two tests were implemented; (i) marker (SNP) was considered the only effect on the model (ii) marker (SNP) and population structure were considered. In order to account for multiple tests, 1,000 permutations of the data were run (Wilson *et al*., 2004). A significant association was detected if the P-value of the SNP was <0.001 of the permutations (***), <0.01 of the permutations (**) and <0.05 of the permutations (*) according to Bonferroni's correction method.

4.4 Results

4.4.1 Population Structure

$F_{ST,}$ used as a measure of genetic differentiation among populations was 0.098 for neutral SNPs and 0.13 for candidate gene loci.Q_{ST} estimates for each trait exceeded that of neutral marker loci in all traits, indicating directional selection (Table 4-4), however leaf area 04, abaxial cell number per leaf 04, abaxial stomatal index 06 and height 04 did not show significant variation among subpopulations (Table 4-4). Interestingly, candidate gene loci exhibited a larger F_{ST} over all loci (0.131) indicating moderate genetic differentiation, however Q_{ST} still exceeded F_{ST} (Table 4-4) in all traits.

Table 4-4: Q_{ST} and F_{ST} estimates. A Comparison of overall Q_{ST} and F_{ST} estimates for leaf, cell and biomass traits from neutral and candidate gene markers. All markers were SNPs detected from the association population in Belgium (Chapter 2: plant material and plantation layout). Neutral markers consisted of 30 SNPs detected in 315 genotypes detected by Dr Guisi Zaina from the University of Udine, whereas candidate gene markers consisted of eight SNPs detected in 315 genotypes.

Quantitative Trait	Q_{ST}	Neutral marker comparison Q_{ST} to F_{ST}	Candidate gene marker comparison Q_{ST} to F_{ST}
Leaf Area 04 (mm^2)	$0.40^{(ns)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Leaf Area 05 (mm^2)	$0.98^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Leaf Area 06 (mm^2)	$0.98^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Leaf Ratio 04	$0.56^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Leaf Ratio 05	$0.75^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Leaf Ratio 06	$0.71^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Specific Leaf Area 05 (mm^2g^{-1})	$0.98^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial CA 04 (μm^2)	$0.84^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial CNPL 04	$0.75^{(ns)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial SD 04	$0.91^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial SI 04	$0.92^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial CA 06 (μm^2)	$0.63^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial CNPL 06	$0.96^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial SD 06	$0.83^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Abaxial SI 06	$0.67^{(ns)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Height 04 (cm)	$0.61^{(ns)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Height 05 (August) (cm)	$0.97^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Height 05 (December) (cm)	$0.92^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Diameter 06 (cm)	$0.96^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$
Stem Volume Index 06	$0.89^{(***)}$	$Q_{ST}>F_{ST}$	$Q_{ST}>F_{ST}$

CA = cell area, CNPL = cell number per leaf, SD = stomatal density, SI = stomatal index, 04 = 2004, 05 = 2005 and 06 = 2006. Significance levels are represented as; $<0.001(^{***})$, $<0.01(^{**})$, $<0.05(^{*})$ and not significant (^{ns})

The Bayesian likelihood assignment tests identified 14 clusters. Figure 4-5 indicates that there is admixture between populations as colours are well mixed, however blocks of colour are seen in population 9 and 12, indicating significant structure in these geographic regions.

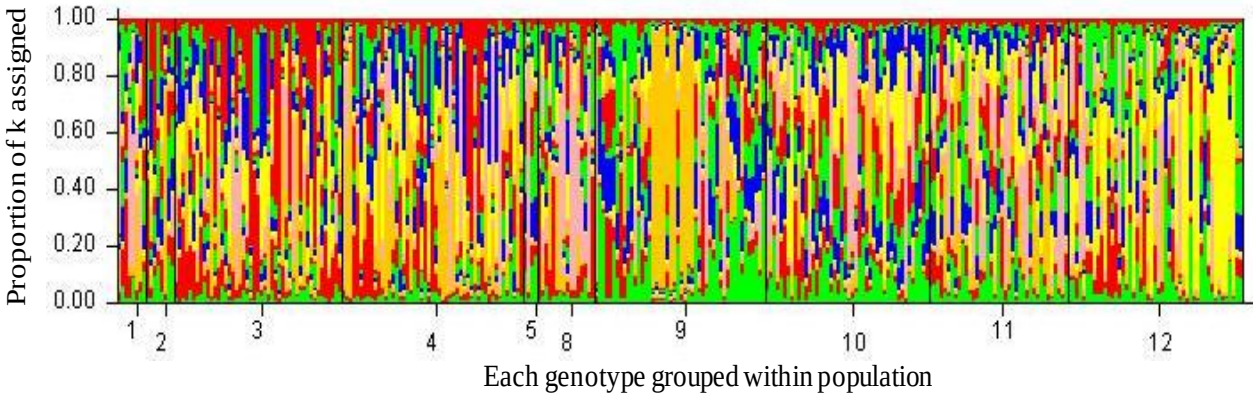


Figure 4-5: Result of population assignment tests.. Each genotype is represented by a single vertical line broken into K coloured segments, with lengths proportional to each of the K inferred clusters, population numbers follow Table 4-3.

4.4.2 Association analysis

Two tests of association analysis were carried out, to determine the importance of using population structure in the analysis. Without population structure fifteen highly significant ($p=<0.001$) associations were seen, whereas when population structure was included a single significant association was detected (Table 4-5). Interestingly the SNP AS1_718 associated with the trait diameter 06 both with and without structure therefore, indicating a robust candidate for biomass.SNP PHAB_618 shows the highest number of associations when population structure is not added to the model, however when structure is added no associations are seen (Table 4-5). AS2_496 is associated with leaf area 05, leaf area 06, SLA 05, cell number per leaf 06, diameter, height August 05 and December 05, without population structure, but when structure is added to the model no associations were found (Table 4-5). SNP E2Fc_503 showed three associations to leaf area 05, SLA05 and stomatal density 04 when structure is not added to the model, and again no associations when structure is included. SNP LEAFY_163 showed no associations either with or without structure. This indicates that without population structure taken into consideration false associations will occur.

Table 4-5: A list of significant associations. (i)considering the marker (SNP) as the only effect in a general linear model (red) and (ii) considering the marker (SNP) and population structure in a general linear model (bold, black). Genotype and phenotype data was collected from the association population located in Belgium. Phenotype data was measured in August 2004 (04), 2005 (05) and 2006 (06). Genotype material was collected in the summer of 2004 and DNA extracted in November 2004 from 315 genotypes.

Leaf Traits

SNP	Gene	Leaf area 04 (mm^2)	Leaf area 05 (mm^2)	Leaf area 06 (mm^2)	Leaf Ratio 04	Leaf Ratio 05	Leaf Ratio 06	SLA 05 (mm^2g^{-1})
ANT_175	*AN*							
AS1_718	*AS1*							
AS2_496	*AS2*		(***)	(***)				(**)
E2Fc_503	*E2Fc*		(*)					(*)
LEAFY_163	*LEAFY*							
LEAFY_232	*LEAFY*							
LEAFY_352	*LEAFY*							
PHAB_618	*PHAB*		(***)	(***)			(**)	(**)

Cell Traits

SNP	Gene	CA04 (μm^2)	CNPL 04	SD 04	SI 04	CA 06 (μm^2)	CNPL 06	SD 06	SI 06
ANT_175	*AN*								
AS1_718	*AS1*								
AS2_496	*AS2*						(**)		
E2Fc_503	*E2Fc*			(*)					
LEAFY_163	*LEAFY*								
LEAFY_232	*LEAFY*								
LEAFY_352	*LEAFY*								
PHAB_618	*PHAB*						(***)		

Biomass Traits

SNP	Gene	Diameter 06	Height 04 (cm)	Height 05 August (cm)	Height 05 December (cm)	Stem volume index 06
ANT_175	*AN*					
AS1_718	*AS1*	**(**)				
AS2_496	*AS2*	(***)		(***)	(***)	
E2Fc_503	*E2Fc*	(*)		(***)		
LEAFY_163	*LEAFY*					
LEAFY_232	*LEAFY*					
LEAFY_352	*LEAFY*					
PHAB_618	*PHAB*	(***)		(***)	(***)	

SLA = Specific Leaf Area, CA = Cell AREA, CNPL = Cell Number Per Leaf, SD = Stomatal Density and SI = Stomatal Index

A significant association was called if the P-value of the SNP was seen in <0.001(***), <0.01(**) and <0.05 of the permutations (**) according to Bonferroni's correction method.

ASYMMETRIC LEAVES 1 (*AS1*) SNP AS1_718 shows one significant associations (P = <0.01) for the biomass traits; diameter (Table 4-5). *Post hoc* one way ANOVA was performed to compare the variation between means of diameter between SNPs (AA, GA & GG).

Figure 4-6 showed a significant difference in the means of diameter between each SNP ($F_{3\ 314}$=6.48, P<0.01**), larger diameters can be seen in both homozygous AA and heterozygous GA, whereas smaller diameters are seen in homozygous GG.

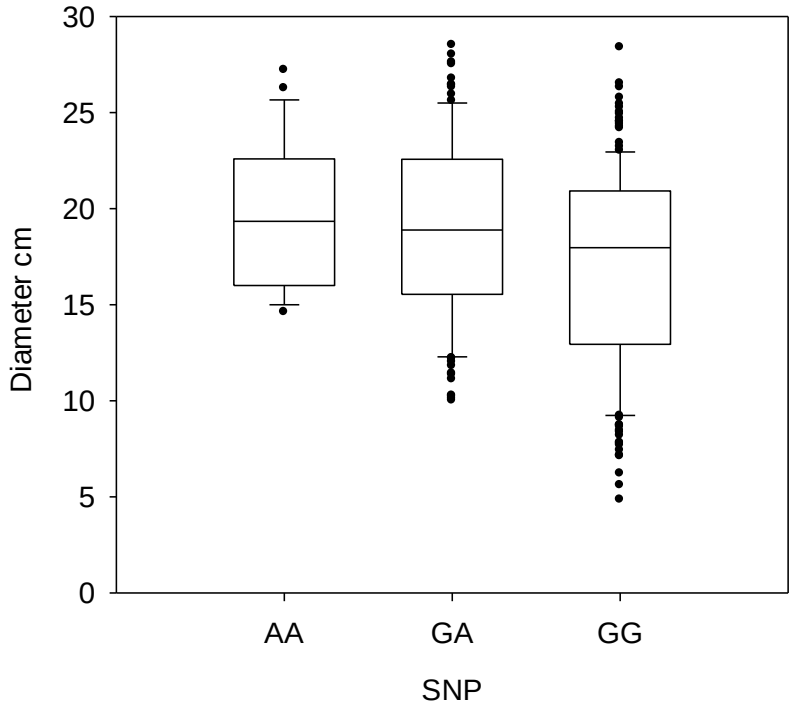


Figure 4-6: Box plot of *ASYMMETRIC LEAVES 1* SNP *AS1_718*.The genotypic effects of *ASYMMETRIC LEAVES 1 SNP AS1_718* on diameter 2006. Genotypic data and phenotypic measurements were collected from the 350 genotypes from the association population planted in Belgium in 2004 (described in chapter 2: plant material and plantation layout).

4.5 Discussion

4.5.1 Population differentiation in structure due to natural selection in *P.nigra*

This study demonstrates that natural selection is the main cause of population differentiation among populations of *P.nigra* in Europe. The main role of selection over genetic drift was supported firstly by strong departures of Q_{ST} from the neutral expectation set by F_{ST} in leaf, cell and biomass traits in both neutral and candidate gene markers. However it is interesting to note that F_{ST} was low in neutral markers (0.098) indicating according to Wrights guidelines moderate genetic differentiation, indicating gene flow between populations. The Q_{ST} value then in all traits is considerably larger indicating fixation of particular genes to latitudinal demes. This is also seen in candidate gene markers, however F_{ST} is much larger but still smaller then estimates of Q_{ST}. The strong genetic differentiation (high Q_{ST}s) estimated in this study are a sharp contrast to the lack of genetic structure seen in neutral and candidate gene markers. Previous studies in poplar have observed similar results, such as Hall *et al*. (2007) within *P.tremula*. It is suggested that low F_{ST}s are observed due to poplar being dioecious, outcrossing and wind pollinators. Dioecious species do not always have free gene flow between populations, as many dioecious plant species including poplar have shown spatial segregation of sexes associated with microhabitat differences (Eppley, 2006). The populations under investigation show a latitudinal difference in traits (2.4) therefore they experience different seasons, photoperiods, precipitation and temperature which result in different growth opportunities. The growing season is shorter in the north compared to the south, this could be one of the possible factors driving adaptive differentiation in this population. Using a similar but slightly different approach studies in the common frog (*Rana temporaria*) found the degree of population differentiation in three heritable quantitative traits (age and size at metamorphosis and growth rate) to exceed that of eight neutral microsatellite markers due to differences in latitudinal cline. As both neutral and candidate markers show larger Q_{ST} values compared to F_{ST} we make the assumption that directional selection is acting on these traits, however another interesting point should be made in terms of genetic diversity. Studies by Cervera *et al*. (2001b) and Storme *et al*. (2004a) have shown that genetic diversity is largest in *P.nigra* collections found in the southern countries in Europe such as France, Italy Spain compared to northern populations in Belgium, Germany and The Netherlands. It is purposed that this difference is due to the topography of the country, variation in climate and soil characteristics and human influence as southern populations promote the survival of natural

populations, whereas human influence within northern populations such as canals would allow gene flow which leads to higher similarity among populations. Interestingly the analysis from SRUCTURE indicates that human influence may have affected genotypes in northern populations such as population 9 found in Germany and population 12 found in the Netherlands as blocks of K is proportional assigned (Figure 4-5). This study combined with Cottrell *et al*. (2005) study indicating that *P.nigra* had refugia in both southwestern and southeastern Europe, one in the Spain were the Pyrenees acts as an effective barrier and one in the Iberian Peninsula and a study by Imbert and Lefevre, (2003) that seed dispersal in *P.nigra* is not effective over a very long distance, indicates that two major populations of *P.nigra* may exist and directional selection is fixing genes within the southern and northern populations. Caution has to be taken when interpreting results obtained within this study as only one environment was observed. A more robust experiment would include a range of biologically relevant conditions (Cano *et al*., 2004). Comparing Q_{ST} estimates derived from a common garden and wild-collected population does however have advantages as Lee and Frost, (2002) found within their study of the copepod, *Eurytrmora affinis* that values were 1.8 times higher in the wild indicating an environmental effect (Palo *et al*., 2003). Also this study is limited to only 30 neutral SNPs and eight candidate SNPs, therefore it is possible that rare alleles could act to increase F_{ST} (Hall *et al*., 2007). .

4.5.2 Association analysis

Before implementing association several factors were taken into consideration such as population structure, nucleotide diversity, LD and phenotype variation. The multiple testing correction used in this study (Bonferroni correction) and population structure was able to reduce the number of false- positive associations due to a moderate level of relatedness in the association population (Gonzalez-Martinez *et al*., 2007). Low F_{ST} values and clustering analysis using K in STRUCTURE have indicated little genetic differentiation; however for a more robust association analysis structure was included in the analysis which again removed many false-positive associations (Table 4-5). A classic example of population structure interfering with associations involved a human study were the occurrence of type 2 diabetes in the Pima and Papago Native American tribes from Southern Arizona. A correlation between a haplotype at the immunoglobin G locus was found with reduced diabetes, however this haplotype was more prevalent in Europeans than in Native Americans, so when different populations were investigated results differed greatly as effects were due to population admixture (Flint-Garcia *et al*., 2003). However some studies have found were sample sizes

were small, adding population structure to the association study results in false negatives (Yu *et al*., 2006). False negatives occur in populations that are highly correlated with population structure, for example flowering time is highly correlated with population structure (R^2 =33-35%) in maize, therefore alleles whose distribution coincides with population structure will not be detected when association models that include structure (Flint-Garcia *et al*., 2005). In this study several SNPs lost significance when structure was added to the model (Table 4-5). Two causes have been proposed to account for this phenomenon; (i) this was a non-functional polymorphism and the association was caused by population structure, or (ii) the polymorphism is functionally related but the polymorphism distribution coincides with population structure (Wilson *et al*., 2004). To determine if (ii) is the case SNPs must be re-evaluated in an alternative population structure, which is an area of future study.

The most important observation in this study was that *AS1_718* SNP showed a strong association to the biomass trait; diameter (Table 4-5). *AS1_718* is found in the *ASYMMETRIC LEAVES 1* (*AS1*) gene. This gene is a member of a small MYB-related gene family also containing *ROUGH SHEATH 2* (*RS2*) and *PHANTASTICA* (*PHAN*) which are involved in the regulation of *KNOX* genes in the meristem, therefore affecting stem cell maintenance which leads to changes in leaf phenotype. Mutational studies have shown that *as1* plants have smaller, more curled, heart-shaped blades with shorter petioles (Zgurski *et al*., 2005). *AS1* has been investigated in great detail to be involved with *ASYMMETRIC LEAVES 2* (*AS2*) a member of the *AS2* gene family (Iwakawa *et al*., 2002), which is also called the *LATERAL ORGAN DOMAIN BOUNDARIES* (*LOB*) domain gene family (Lin *et al*., 2003). *AS2* is important in repressing *KNOX* genes *BP*, *KNAT2* and *KNAT6* in leaves (Chalfun *et al*., 2005) to initiate leaf development , by acting alone or in combination with *ASYMMETRIC LEAVES 1* (*AS1*) (Byrne *et al*., 2002). Previous studies have suggested that *AS1* & *AS2* act together in a common genetic pathway to activate *LOB* which establishes boundaries between the meristem and the differentiated lateral organs (Figure 4-7).

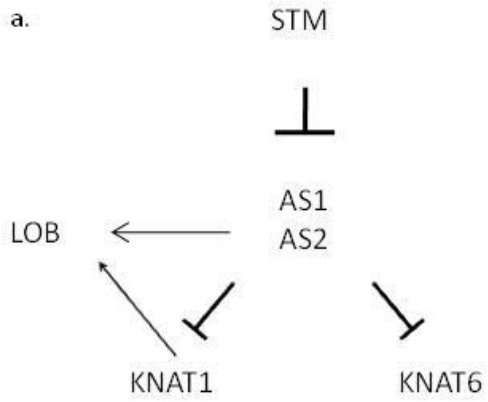


Figure 4-7: Schematic representation of the genetic pathway controlling *LOB* activation adapted from Byrne *et al*., 2002. *STM* (Shoot Meristemless) represses *AS1* and *AS2* in stem cells of the meristem. *AS1* and *AS2* together repress *KNAT1* and *KNAT6* on organ primordial to enable *LOB* is to be expressed in a region between the SAM and organ primordial for leaf development.

AS1 mapped to Linkage group VI in between markers ORPM26 and PMGC 2578, this area was not identified as a 'major' or 'minor' hotspot according to the criteria set in chapter 3 (Figure 4-8), but *AS1* was selected due to its known role with *AS2* (Figure 4-7). However *AS1* is found to co-locate to QTL for leaf production in the short rotation coppice experiment (Figure 4-8), which indicates a role in growth. Interestingly *AS1* did not shown any associations in leaf traits, but only in the biomass trait diameter, which did not have any QTL on this linkage group over the three QTL experiments (Figure 4-8). This indicates that using candidate gene co-location as a technique for candidate gene selection should be taken with some caution. The lack of associations in candidate genes could be firstly because of poor selection of candidates due to QTL having a low resolution, improvements to the genetic map by adding more markers could be a solution, or this could be a poor technique due to the species differences as QTL analysis was run on a controlled cross between *P.deltoides* and *P.trichocarpa* to create family 331, therefore QTL analysis on a *P.nigra* cross would maybe have been more beneficial. An association in *AS1_718* on linkage group VI for biomass traits indicates that this linkage group should be investigated in more detail, especially between markers ORPM26 and PMGC 2578. Combining association studies and QTL analysis within other studies has been shown to be a powerful technique and a strong experimental approach, as candidate genes with an established position under QTL peaks have shown more associations than those not under peaks, such as *sh1* and *sh2* genes in maize co-localized with QTL for kernel traits and showed strong associations (Wilson *et al*., 2004).

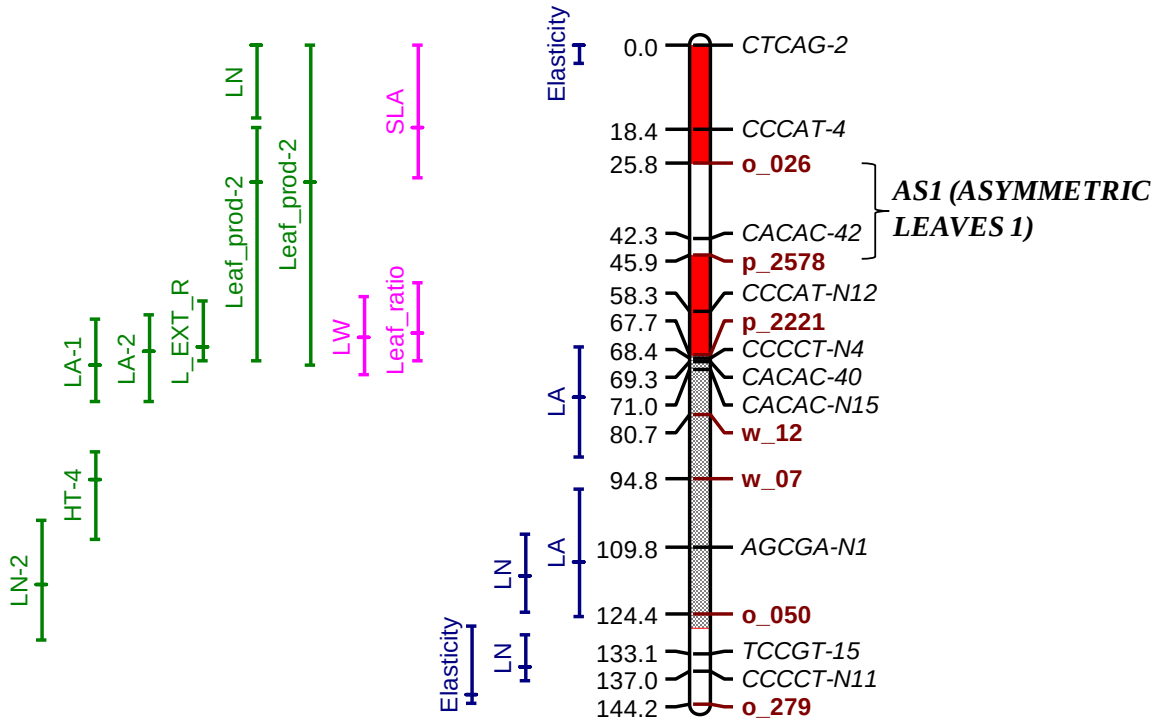


Figure 4-8: Linkage group VI of family 331 molecular map. Quantitative Trait Loci (QTL) for leaf, biomass and cell traits from three different poplar experiments; short rotation coppice experiment (green), drought (red) and CO_2 (blue) described in chapter 3. QTL positions are shown ± confidence internal as determined by F2 drop off, colour represents experiment: drought (pink), CO_2 (blue) and short rotation coppice (green), abbreviations are described in 3.7.2: Table 3-4. Marker names are shown to the right of the linkage group (SSR markers – in bold brown and AFLPs – in italics) and cM distances to the left of the clear bar representing the genetic map. Marker names are shown to the right of the linkage group (SSR markers – in bold brown and AFLPs – in italics) and cM distances to the left of the clear bar representing the genetic map. Linkage group number is shown in Roman numeral above bar. Hotspots are shown on the genetic map; solid fill red is a major hotspot, whereas red criss-crossed is a minor hotspot.

In association studies population size is known to affect the statistical power of analysis. Simulation studies have shown that sufficient power exists to detect SNP-phenotype associations for QTL that account for as little as 5% of the phenotypic variation when approximately 500 individuals are genotyped for approximately 20 SNPs within candidate gene region (Long and Langley, 1999). Within this study a small number of individuals (350) and SNPs were investigated (8), therefore increasing the number of individuals in the population and increasing SNP density is an area of future study. Increasing population size reduces the impact of several factors that limit the power of association studies such as allele class frequency, the number of alleles per locus and the interaction between diverse alleles (Wilson *et al*., 2004). A great advantage in association studies is the analysis of multiple

SNPs or markers in a single experiment; however this comes with its drawbacks. For example more alleles results in more allelic classes to test causing multiple test problems, however permutation analysis determining an experiment-wise significance threshold solves this problem with ease. A second problem results from an increase in possible alleles per locus, such as the number of individuals within each allelic class decreases, therefore decreasing the power of the test. This is where population size can benefit, as it will increase the number of individuals with rare alleles, increasing the power of the test. The third problem that needs to be addressed is the interaction between diverse alleles known as epistasis. No models at present include alleles at different loci interacting. However increasing population sizes will enable the number of individuals in each allelic combination therefore allowing for more powerful tests of epistasis or genes confirmed to play a role in the expression of a trait could be added to the model as a cofactor to test for epistatic effects (Szalma *et al*., 2005).

In this study coding and non-coding regions of the gene were used to detect SNPs, however only a small fragment ~800bp of the gene was investigated. In other tree species within-gene linkage disequilibrium has shown rapid decay (Gonzalez-Martinez *et al*., 2007). Therefore within this study it is possible that some genetic associations involving these genes was not detected. So an area of future work would involve SNP discovery across the whole genomic sequence. In an ideal world conducting an association study with all genes in *P.nigra's* genome would complete our understanding of the genetic architecture of our traits of interest; however at present this is not possible, mostly due to cost, as within the poplar community we are not limited by the whole genome sequence as it was published and is freely available (Tuskan *et al*., 2006a).

4.5.3 Summary

In this study a relatively modest array of genes belonging to leaf development and high yields in poplar were evaluated in the first multigene association genetic study in European *Populus nigra*. We have shown a strong adaptive divergence in several quantitative traits related to phenology, as little population structure and extensive gene flow was revealed in molecular data. In this chapter for the first time association analysis has been used to understand leaf development in *P.nigra*, resulting in *AS1*showing strong genetic association with the biomass trait diameter. This study indicates that a candidate gene approach such as this can be successful in a species with low linkage disequilibrium.

5 . Analysis of gene expression during leaf growth in *P.nigra* within a controlled glasshouse environment

5.1 Overview

Many leaf developmental genes have been identified and expressional studies carried out in *Arabidopsis*, however there are few studies on *Populus* species. Recently, with the whole genome sequence of *P. trichocarpa* publicly available, homologes between *Arabidopsis* and *Populus* can be identified and therefore individual genes can be studied.

Throughout this project I have used modern genomic approaches to understand leaf development in a natural population of *P.nigra* by; (i) identifying differences in leaf morphology within a common garden experiment (chapter 2), (ii) selecting candidate genes using literature, QTL and microarray analysis (chapter 3), (iii) association mapping using SNPs (chapter 4), resulting in one strong candidate gene for yield *AS1*. In this chapter I will investigate the genes *ASYMMETRIC LEAVES 1* (*AS1*), *ASYMMETRIC LEAVES 2* (*AS2*), *ANGUSTOFOLIA* (*AN*), *PHABULOSA* (*PHAB*), *ACC oxidase* (*ACO*) and *E2Fc* further to see if expression differs with leaf age in a control glasshouse environment using leaf area genotype extremes selected from 2004 and 2005 field data. A detailed morphological study of the extremes will also take place to understand the mechanisms controlling their phenotypic differences.

5.2 Introduction

5.2.1 Leaf development

Leaves play an essential role by providing the site for photosynthesis and transpiration enabling growth and metabolism. To understand the underlying mechanisms controlling leaf development is important to producing high yielding phenotypes for future needs within a fast growing biomass crop industry. Leaf development has been described in three stages; (i) the organogenesis stage, where cells on the flanks on the Shoot Apical Meristem (SAM) are set aside as the founder cells of the initiated leaf, (ii) increased cell division results in the formation of leaf primordium and (iii) cell division, expansion and differentiation occurs to form a leaf consisting many specialized cells such as guard cells, spongy mesophyll and palisade cells.

The SAM contains a number of cells which grow and divide. Growth occurs perpendicular to the surface of the meristem to form leaves. Variations exist to this process, where small groups of cells in the axil between the leaf and the stem termed axillary meristem can become active under suitable conditions to generate new stems and leaves (Fleming, 2003). Initial leaf primordial undergoes growth to generate a proximal-distal axis (base-tip) and a proximal-distal axis (lateral growth) to form the lamina or blade. In most species the lamina will have a flattened side facing the stem (adaxial) and a side facing away from the stem (abaxial) which is termed the dorsiventral axis. Adaxial and abaxial sides meet at the lamina edges forming the blastozone (Fleming, 2003). In nature we see a spectrum of leaf shapes and sizes which are caused by the variation in growth rate. The *P.nigra* leaf would be classified as a simple form of leaf as growth occurs as a smooth continuum along the entire length of the primordium with a maximum rate along the base of the primordium. Many genes have been isolated within leaf development. In this study due to microarray analysis results (3.7.1), co-location within QTL regions (3.7.2) and association analysis results (4.4.2), six strong candidate genes for leaf development have been identified. These six candidate genes will be discuss in this chapter in more detail.

5.2.2 Molecular control of leaf development

The SAM consists of a population of cells termed stem cells which undergo repeated rounds of proliferation causing its growth. Two gene pathways control SAM growth, a positive acting pathway which promotes meristem growth (based on the homeodomain transcription factor *WUSCHEL* (*WUS*) and a negative acting pathway which suppress meristem growth (based on

a series of *CLAVATA* gene products) (Fleming, 2005). The meristem itself is then defined by the expression of STM-like *KNOX* homeobox genes; the so-called class 1 *KNOX* genes (Kerstetter *et al*., 1994). *KNOX* genes are not expressed in the leaf primordial suggesting that suppression of these genes initiates leaf primordial growth. *ASYMMETRIC LEAVES 2* (*AS2*) a member of the *LATERAL ORGAN BOUNDARIES* (*LOB*) domain gene family and *ASYMMETRIC LEAVES 1* (*AS1*) a member of the *MYB* transcription factor gene family are expressed in leaf primordial and is involved in suppressing *KNOX* genes to begin leaf initiation and leaf growth (Figure 5-1).

After leaf initiation stage two begins with the increase in cell division and leaf primordial growth. The retinoblastoma (*RB*)-*E2F* pathway is one of the most important regulatory pathways that control and couple cell division and cell differentiation in both animals and plants (Stevaux and Dyson, 2002). *E2F* and *DP* interact to form active transcription factors that bind to different gene promoters and regulate the expression of cell cycle progression genes. However if *RB* protein binds to *E2F* protein it blocks transcriptional activity (del Pozo *et al*., 2006) and delays cell division. Reduced levels of *E2Fc* show lower levels of DNA endoreplication and more but smaller cells in leaves (Figure 5-1).

The final stage of leaf development involves cell division, expansion and differentiation. There are many genes involved in this process; I will discuss three *ANGUSTOFOLIA* (*AN*), *PHABULOSA* (*PHAB*) and *ACC OXIDASE* (*ACO*). *ANGUSTOFOLIA* (*AN*) contains the conserved *D2-HDH* motif of CtBP (C- terminal binding protein) and regulates the polarized growth of leaf cells by controlling the arrangement of cortical microtubules in epidermal and mesophyll cells (Kim *et al*., 2002b) (Figure 5-1). Mutants display defects in the expansion of leaf blades in the leaf width direction. *PHABULOSA* (*PHAB*) is restricted to the adaxial domain of the leaf, where *PHAB* protein interacts with a small diffusible lipid-based factor in a gradient from the meristem leading to the fixation of adaxial identity (Figure 5-1). Mutations in *PHAB* protein in the *START* domain leads to the formation of radialised leaves lacking lamina growth. It has been suggested that the ethylene pathway is involved in leaf development; the pathway begins with the conversion of S-adensyl-methionine to 1-amnocyclopropane-1-carboxylate (*ACC*), a reaction catalysed by ACC synthase. *ACC* is further converted into ethylene in a reaction catalysed by *ACC oxidase* (*ACO*). It is considered that ethylene is involved in the limitation of cell wall expansion and therefore determines size and shape of leaves (Figure 5-1).

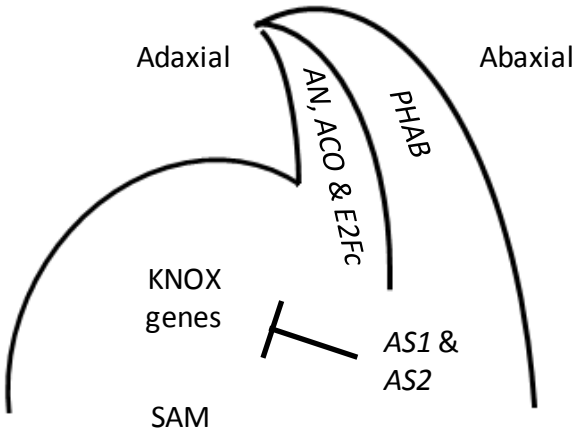


Figure 5-1: A schematic diagram of leaf initiation and growth showing the relationship between genes selected for expressional studies. *AS1* (*ASYMMETRIC LEAVES 1*) and *AS2* (*ASYMMETRIC LEAVES 2*) repress *KNOX* genes in the SAM to initiate leaf growth, where *AN* (*ANGUSTOFOLIA*), *ACO* (*ACC OXIDASE*) and *E2Fc* increase lamina size, whereas *PHAB* (*PHABULOSA*) is involved in cell fate by identifying abaxial leaf surface identity.

5.2.3 Real Time PCR

RT-PCR (reverse transcription-polymerase chain reaction) is an alternative technique to northern blotting and RNase protection assay for the detection of mRNA. Recently RT-PCR has been utilized to quantify changes in gene expression and to validate array analysis. In standard PCR the amplified product is only a qualitative indicator of the template, however in PCR quantitative systems have been developed whereby the buildup of product in a PCR can be monitored during the reaction. RT-PCR requires a thermocycler that can read optical signals coming from the reaction vessels. Currently there are four different chemistries on the market; Taqman® (Applied Biosystems, Foster City, CA, USA), Molecular Beacons, Scorpions® and SYBR® Green, all detect PCR products via the generation of a fluorescent signal. Taqman probes, Molecular Beacons and Scorpions use temple-specific primers that carry two fluorophores, one an emitter and the other a quencher. The fluorescence signal is extinguished by Fluorescence Resonance Energy Transfer (FRET), which is based on the digestion of the FRET primer, as PCR proceeds polymerase activity destroys the primer and reduces the physical proximity between the two dyes allowing the fluorescence of the emitter to become detectable. SYBR Green is a fluorogenic dye that exhibits little fluorescence when in solution but emits a strong fluorescence signal upon binding to a double stranded DNA. A sigmoidal curve is produced when detecting fluorescence dyes using real-time PCR, it usually takes around ten cycles before fluorescence is detected. After detection an exponential phase is entered in which the signal amplifies by a fixed factor with every cycle, until a plateau is

reached. Two parameters are used to describe the curve; cycle threshold and PCR efficiency. Cycle threshold value is defined as the cycle number at which the curve increases above the baseline or error variance (a measure of the amount of template in the original sample). A linear relationship is observed at this point, correlating, the cycle threshold value with the number of template copies. The relationship between the initial number of DNA molecules N0 and the cycle threshold is explained in equation 5-1.

$$N_O = \frac{N_C}{(e+1)^c} \tag{5-1}$$

Where N_c is the number of DNA molecules at the threshold fluorescence, e is the PCR efficiency and c is the cycle threshold (Rutledge and Cote, 2003). PCR efficiency is given as a fraction, so if the efficiency is 100%, e = 1 and products build up by a factor of two with every cycle. Therefore with a rearrangement of equation 5-1, where N_c and e are constants, a plot of $\log N_0$ versus C shows a straight line with slope $-\log(e+1)$ and amplification efficiency can be estimated from this slope (5-2).

$$\log N_O = \log N_c - C \log(e+1) \tag{5-2}$$

Two techniques are commonly used to quantify RT-PCR; (i) relative quantification based on the relative expression of a target gene versus a reference gene and (ii) absolute quantification based either on an internal or an external calibration curve (Pfaffl, 2001). Relative quantification utilizes a reference gene or sometimes termed a house keeping gene, which are assumed not to vary from one sample to the next and therefore are unbiased. Commonly used reference genes include; 28S rRNA, β-actin, elongation factor 1α, albumin, tubulin, actins and glyceraldehydes-3-phosphate dehydrogenase (G3PDH or GAPDH). Relative quantification determines the ratio between the quantity of a target molecule in a sample and in the calibrator (healthy tissue or untreated cells) giving an analysis of gene expression. Target molecule quantity is normalized to the reference gene and it is recommended to use more than one reference gene for a more reliable result. The comparative Ct method is used to determine gene expression which assumes the efficiency to be equal to 2 for the target and the reference amplicon. Within this method the sample and the calibrator data are first normalized (5-3) and then comparative Ct is determined as showed in equation 5-4 and then the expression is determined as show in equation 5-5.

$$\Delta C(t)sample = C(t)target - C(t)reference \quad (5\text{-}3)$$

$$\Delta C(t)calibrator = C(t)target - C(t)reference$$

$$\Delta\Delta C(t) = \Delta C(t)sample - \Delta C(t)calibrator \quad (5\text{-}4)$$

$$Expression = 2^{-\Delta\Delta c(t)} \quad (5\text{-}5)$$

At present, limited studies have used RT-PCR in *P.nigra* to determine changes in gene expression in leaf development. Studies of *Arabidopsis* have shown that *AS1* and *AS2* are expressed in the above ground parts of the plant except internodes and pedicels (Chalfun *et al*., 2005) and *AS1* is expressed throughout the leaf blade (Lin *et al*., 2003). *ACO* is expressed in newly initiated leaves and in induced senescent leaves (Murray and McManus, 2005), whereas *PHABULOSA* expression has been shown to be uniformly expressed throughout the leaf prior to leaf initiation where later it becomes restricted to the abaxial domain of the leaf (Fleming, 2003). Studies reducing the level of *E2Fc* have shown lower levels of endoreplication accompanied by the development of more but smaller cells in leaves and cotyledons (del Pozo *et al*., 2006), but most studies have investigated *E2Fc* at the cellular level. Finally *ANGUSTOFOLIA* has been shown to be expressed throughout the plant including rosette leaves (including SAM), cotyledons, roots and floral buds in *Arabidopsis*, but high levels of expression were seen in leaves and floral buds, indicating a role in leaf development (Kim *et al*., 2002b).

In this study we focus specifically on expression of genes with known function in leaf development; *AS1*, *AS2*, *AN*, *ACO*, *PHAB* and *E2Fc*, in the meristem, young and semi-mature leaves of the 'extreme'subset of *P.nigra* genotypes. Genotypes were selected based on leaf size from the association population (2.4), the selection process is described in Microarray design, preparation, hybridization and analysis (3.5.3), whereby five 'big' leaf genotypes and five 'small' leaf genotypes were identified. In this study an attempt to understand leaf developmental differences and gene expression within *P.nigra* extremes was conducted in a controlled glasshouse environment.

5.3 Materials & Methods

5.3.1 Plant Material

Genotypes were selected for cuttings based on leaf area measurements taken in August 2004 as described in 2.3.1. In brief, average leaf area was calculated for each genotype, values were sorted in ascending order to identify the five lowest leaf area genotypes (the 'small' extremes: B7, C7, C15, FR7 & RIN2) and the five highest leaf area genotypes were selected (the 'big' extremes: N38, N53, N66, NL1682 & SN19). Leaf size measurements were also collected in 2005 from the Belgium field site (Plant material and plantation layout), and these results were again sorted into extremes, resulting in three extra 'big' (N30, N56 & NVHOF-16) and 'small' (71095-1, 71092-36 & CART2) genotypes being added to this study (Figure 5-2).

Plants were propagated from cuttings, soaked overnight and planted in 19-1 pots in a glasshouse on the 17th January 2007 and watered daily. Photoperiod in the glasshouse was maintained at 16h days and the temperature ranged from 19 to 25°C during the day. Experimental design was set out in five randomised blocks, each consisting of 16 experimental cuttings, with five rows of guard trees (stock material) planted around the entire trial at the same spacing to serve as a buffer.

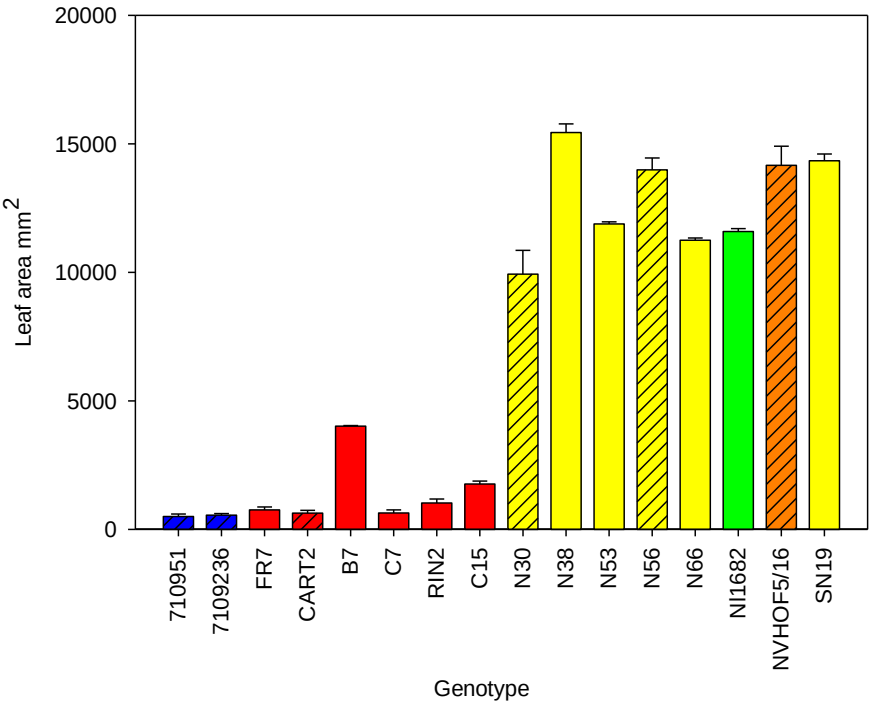
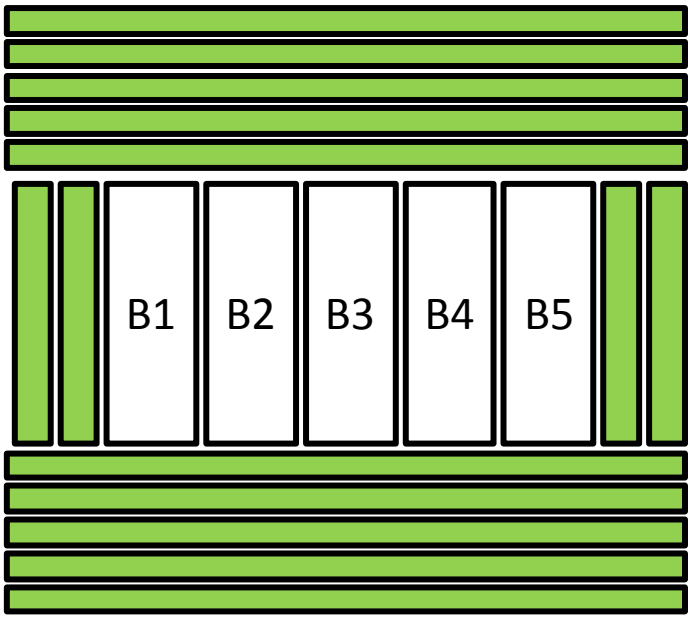
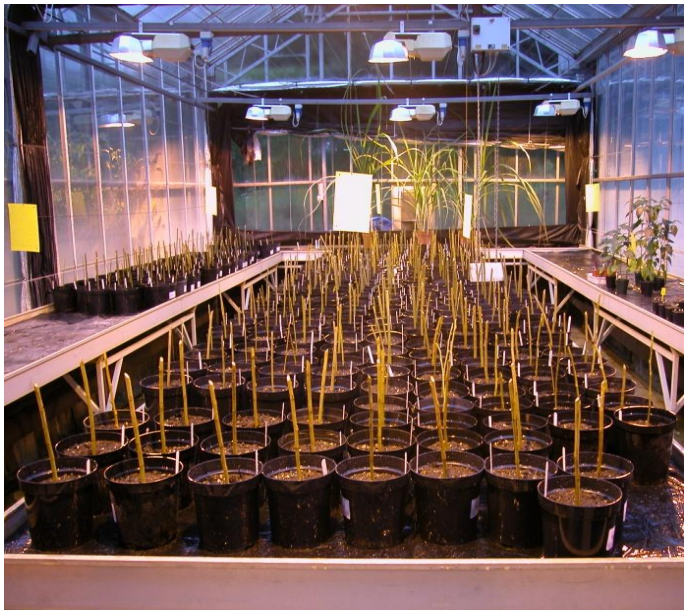


Figure 5-2: Leaf area extremes selected for study.Eight 'small' (710951, 7109236, FR7, CART2, B7,C7, RIN2& C15) and eight 'big' l(N30, N38, N53, N56, N66, NL1682, NVHOF5/16 & SN19) leaf area extreme genotypes selected based on leaf area measurements in 2004 (solid fill) and 2005 (solid slash fill) from the association population described in chapter 2. Colour represents country of origin; blue = France, red= Spain, yellow= Italy, green = the Netherlands and orange = Germany. Data points are the mean leaf area measurement in 2005 for each genotype, error bars represent standard error of the mean.

A.



B.



C.

Figure 5-3: The glasshouse experiment.(A) A schematic representation of the experimental design of the glasshouse gene expression study of *P.nigra* extremes; 'small' leaves and 'big' leaves. The random block design consists of five blocks each containing 16 genotypes (1 rep in each block), represented by clear rectangles and labeled B1, B2, B3, B4, B5. Guard trees are represented by green blocks, whereby 5 rows of 1x 9 trees were placed above and below blocks and 2 rows of 1x16 to the left and right. (B) First day of the experiment, photograph taken on the 17th January 2007, showing cuttings before growth and (C) A photograph taken on the 30th March the last day of the experiment after 73 days of growth.

5.3.2 Leaf growth

On the 13th February 2007 the first fully unfurled young leaf from each genotype was labeled with white cotton and its outline traced with a pencil onto white paper and labeled leaf 1. Growth was followed by repeating this measurement every other day between 13th February and 29th March 2007. Replicates were measured on each genotype by following the same process as the successional leaves developed, and naming each leaf 2, leaf 3, leaf 4 and so on up to leaf 10. The drawn leaves were then scanned using an Umax Astra 6700 scanner, at 200DPI, in black and white, and saved. The scanned images were then processed using Image J (Image J.1.32j, Wayne Rasband, USA). Leaf outlines selected by finding thresholds and measured for leaf area mm^2. Leaf length and leaf width were also measured using Image J. From leaf measurements leaf width to length ratio (Chapter 2:equation (2-1)), leaf length extension (5-6), leaf width expansion (5-7), absolute growth rate(5-8) and relative growth rate(5-9) were calculated. Leaf production was also scored by counting the number of existing visible individual leaves (including those that were unrolled) between the tagged leaf and the tip of the stem on the 30th March. The increase in leaf number was converted to leaf production per day by using equation 5-10. Leaf Length Extension was calculated using equation 5-6:

$$LLE = LL_i - LL_0/d \quad (5\text{-}6)$$

Where LL_i is the last and LL_0 is the first leaf length measurement and d is the number of days between measurements, calculated in millimeters per day. Leaf Width Expansion was calculated using equation 5-7:

$$LWE = LW_i - LW_0/d \quad (5\text{-}7)$$

Where LW_i is the last and LW_0 is the first leaf width measurement and d is the number of days between measurements, calculated in millimeters per day. Absolute growth rate (AGR $mm^2\ d^{-1}$) and Relative Growth Rate (RGR $mm^2\ d^{-1}$) were calculated using the equations 5-8 & equation 5-9 respectively, where A_i is the last and A_0 is the first leaf area measurement and d is the number of days between measurements.

$$AGR = (A_i - A_0)/d \quad (5\text{-}8)$$

$$RGR = (A_i - A_0)/A_0/d \quad (5\text{-}9)$$

Leaf production per day was calculated using equation 5-10.

$$\text{Total number of leaves from tag to stem/d} \quad (5\text{-}10)$$

Mature leaf traits were collected on the 30th March from each genotype, by counting down 10 leaves from the leaf just fully emerged; this was termed leaf age Ln-10. Leaves were drawn around with a pencil onto white paper and scanned as described above using the Umax Astra 6700 scanner and processed in Image J to get leaf area mm^2. Petiole length was measured on Ln-10 to the nearest millimeter.

5.3.3 Epidermal Cell Imprints

Cell imprints were taken at leaf age Ln–10 on each individual tree. Imprints were taken from the abaxial (bottom) and adaxial surface of the leaf on the basal section. An area approximately 1 cm^2 was painted with clear nail varnish and left to dry for 5 minutes. Sellotape was then placed on the nail varnish with a little pressure from the thumb, and then peeled off gently. This left a cell imprint on the sellotape as described by Gardner *et al*. (1995) that was then placed on a glass microscope slide and labeled with the correct block, line, row and genotype name. The slides were placed in a dark container ready for imaging. Slides were viewed on a Zeiss microscope and images captured with a digital camera attached at x 400 magnification at a 100% zoom. Images were then imported for image processing and analysis using ImageJ for windows (Image J.1.32j, Wayne Rasband, USA). Epidermal cell number, stomatal number and 10 epidermal cell areas were measured so that mean epidermal cell area (CA), stomatal density (SD), stomatal index (SI) and number of cells per leaf (CNPL) could be calculated using equations described in chapter 2.3.3 equations 2-3, 2-4 & 2-5.

5.3.4 Biomass Traits

On the 30th March 2007, seventy three days after planting maximum (DAP) stem height was measured to the nearest 0.1 cm with a metre ruler and number of branches counted form each genotype. Stem diameter was measured to the nearest millimetre with digital calipers at 30cm above the soil level and converted into basal stem area (5-11), where π is 3.142 and r^2 is radius. Stem volume index was estimated seventy three DAP using equation 5-12, where l is height (cm) and BA is basal area (cm^2). Huber's formula for stem volume was calculated using equation 5-12.

$$Basal\ area\ (BA) = \pi\ x\ r^2 \quad (5\text{-}11)$$

$$V = l\ x\ BA \quad (5\text{-}12)$$

5.3.5 Data analysis

Data was entered and manipulated in Microsoft excel. Statistical tests were carried out in Minitab release 15(Minitab Inc., State Collage, Penn., USA). Two-way analysis of variance

(ANOVA) were conducted for each trait in the whole population to test for variation between blocks and 'small' and 'big' leaf genotypes. Significant variation are assigned an asterisk (*), $p<0.001$(***), $p<0.01$(**), $p<0.05$ (*) and no significant difference (ns).

5.3.6 RNA extraction

Leaves were sampled on the 3rd April by removing with scissors at the leaf base and immediately placing in a pre-prepared labeled foil packet and then into liquid nitrogen. Samples from each genotype included; the meristem (the tip of the stem, unfurled leaves and first fully unfurled leaf), Ln 3 and Ln 6, identified by counting down from the meristem. Leaves were stored at -80^{o}C until RNA extraction.

Under liquid nitrogen (N_2) each whole leaf was ground to a fine powder separately with a sterilized pestle and mortar, ground material was the placed into a 50ml falcon tube (Cellstar ® greiner bio-one) and 15 ml pre-warmed (65^{o}C) 2% hexadecyltrimethylammonium bromide (CTAB) extraction buffer and 400µl β-mercapto-ethanol (Sigma-aldrich, Uk) were added as described by Cheng *et al*., 1993. Material was then vortexed and incubated at 65^{o}C for 10 min. Following incubation 15ml of CHISM (chloroform: isoamyl alcohol, 24:1) was added, vortexed lightly, then centrifuged at 4500 rpm (Sorvall® legend RT) for 20 min at room temperature (RT). The upper-phase was transferred into a fresh 50ml labeled falcon tube containing 15ml CHISAM and centrifugation repeated. The upper-phase was then removed and added into a labeled JA-20 tube (Oakridge, USA) containing 3mls of 10M LiCl (1:4 vol LiCl), placed on ice and stored at 4^{o}C over night. Following an over-night precipitation samples were centrifuged (Beckman J2-21 and a JA-20 rotor) at 10 000 rpm for 30min at 4^{o}C. The supernatant was then removed and the pellet re-suspended in 700µl of pre-warmed (60^{o}C) SSTE, transferred into a pre-labeled 2ml Eppendorf (Eppendorf, 5417R, Cambridge, UK) and incubated for a few minutes at 60 ^{o}C to allow re-suspension of the RNA pellet in the buffer. After incubation 700µl of CHISAM was added, samples vortexed and centrifuged at 10 00rpm at RT for 10 min. The upper-phase was then transferred into a fresh 2ml Eppendorf containing 700 µl CHISAM and re-centrifuged at 10 000rpm at RT for 10 minutes. 600 µl of the upper phase was then transferred into a pre-labeled eppendorf containing 1.2ml 99.8% ethanol which was pre-chilled at 20 ^{o}C and the RNA precipitated at -80 ^{o}C for one hour. After period of precipitation samples were centrifuged at 13 000 rpm for 30 min at 4 ^{o}C, then supernatant removed and pellet washed twice with 1 ml of 70% ethanol by centrifugation at 10 000 rpm for 2 min at 4 ^{o}C. Samples were then air dried at RT for 10-20 min to remove all traces of ethanol from the pellet, then re-suspended in 20 µl of DEPC-treated H_2O and stored at -80^{o}C.

RNA quantity was assessed using a spectrophotometer (NanoDrop® ND-1000 Spectrophotometer, Wilmington, DE, USA) and quality assessed using electrophoresis (1x mops agarose gel). Samples tested on the Nanodrop were diluted 1:20 with DEPC-treated H_2O and nucleic acid purity estimated using the ratio absorbance at 260 nm (A_{260}) and 280 nm (A_{280}). A value greater than 1.8 for A_{260}: A_{280} was considered of sufficient purity. RNA quality was assessed using gel images produced by electrophoresis, by seeing the presence of two ribosomal bands (28S and 18S).

5.3.7 cDNA synthesis

Prior to cDNA synthesis samples were treated to remove contaminating DNA, using the TURBO DNA-free kit provided by Ambion (Ambion Inc. Austin, TX), according to manufacturer's instructions. Following treatment, RNA quality was assessed using spectrophotometry (Nanodrop, as described above). Unless stated otherwise, all materials for cDNA synthesis were obtained from Invitrogen (Paisley, UK). 1.5µg of total RNA was transferred into a 0.5ml eppendorf tube, 1µl of anchored oligo $(dt)_{20}$ primers (500ng/ µl), 1 µl 10mM dNTP mix and DEPC-treated H_2O was added to make a 20 µl reaction. The RNA was denatured by heating the sample at 65^{o}C for 5 minutes and quickly transferring onto ice for 1 minute, followed by centrifugation at 10 00rpm for 1 minute. Mean while a reverse transcription master mix was prepared consisting of: 6 µl of 5 x –RT-buffer (first strand buffer), 2 µl of 0.1M DTT, 1 µl RNaseOUT and 1 µl superscript II RT, for a total volume of 30 µl. Samples were then incubated at 50^{o}C for one hour, followed by incubation at 70^{o}C for 15 minute to inactivate the reaction. cDNA was stored at -20^{o}C until real time PCR.

5.3.8 Real Time PCR

Real time RT-PCR was performed on three biological replicate genotypes at three leaf ages; meristem, Ln 3 and Ln 6, with three technical replicates. mRNA was isolated and cDNA synthesis produced as described above (5.3.6 & 5.3.7). Specific primers for *AS1*, *AS2*, AN, *PHAB*, *ACC Oxidase and E2Fc* were designed in Primer 3 (Rozen and Skaletsky, 2000), designed to anneal to sequences in two exons on opposite sides of an intron or at the exon-exon boundary of the mRNA (Table 5-1). PCR reactions were carried out in a 20 µl reaction mixture containing 10 µl SYBR green master mix (Finnzymes, Finland), 5 µl primer mix (0.3 µl forward primer , 0.3 µl reverse primer at a10 µM concentration and 1.9 µl H_2O), 1 µl cDNA and 4 µl H_2O.Amplifications were performed using a DNA Engine Petier thermal cycler named Chromo 4 (Bio Rad), using the following cycle conditions; 95^{o}C for 10 min and 40 cycles at 95 ^{o}C for 10sec, 60 ^{o}C for 15sec, 72 ^{o}C for 15 sec and 75 ^{o}C for 1 sec. A number of

stably expressed house-keeping genes were chosen as internal controls; *PDF* and *UBQ1* (Table 5-1) and melting curves were used to confirm amplications.

Table 5-1: *Populus* gene primers for RT-PCR. Forward (sense) and reverse (antisense) directions shown. Housekeeping genes are shown in bold italics and candidate genes shown in italics.

Gene	**Antisense (5' to 3')**	**Sense (5'to 3')**
ACC_01	*CCAAGCTTTCGGCTTATTCA*	*AGGGGGTGTTCTGTCCCTAC*
AS1	*TGTGTTCATGCGCTGTGATA*	*AGGCAATGGAAAGTGGTGTC*
AS2	*TCGTAGGCCAAGGAATTGAC*	*TGCGTATTTGCCCCTTACTT*
AN	*CCAAGCTTTCGGCTTATTCA*	*AGGGGGTGTTCTGTCCCTAC*
E2Fc	*GGTTTTCGGAAAGAGGAAGG*	*GAACCGGACAATCAATCCAT*
PHAB	*GCTTGTCTCTCTCCTTGCTTTCT*	*TAGAGCAGGAGATAGAGGGG*
UBQ1	***GATCTTGGCCTTCACGTTGT***	***GTTGATTTTTGCTGGGAAGC***
PDF	***GAACCCTCCAATGCCTATCC***	***TCCTGATGTGCGACTGAAC***

Data analysis was carried out in excel using the comparative Ct method, as a constant efficiency equal to 2 was seen in the target and the reference amplicon (Ramakers *et al*., 2003) by running a standard curve using a serial dilution. The amount of target amplicon (T) in sample was normalized to the average of both reference (R) amplicons and related to all samples (5-13), giving a fold difference in expression between the sample and all samples.

$$Expression = 2^{\wedge\left((Ct^{T1} - Ct^{R1}) - (\sum Ct^{T1:T9} - \sum Ct^{R1:R9})\right)} \quad (5\text{-}13)$$

Where ^ denotes to the power, Ct is the amount of amplicon, T is the target amplicon, R is the average of both reference amplicons and numbers represent samples numbers. In this study we have a total of 9 samples per genotype

5.3.9 Expression data analysis

Data was entered and manipulated in Microsoft excel. Statistical tests were carried out in Minitab release 15(Minitab Inc., State Collage, Penn., USA). Two-way analysis of variance (ANOVA) was conducted for each gene to test for variation between 'small' and 'big' genotypes, leaf ages and interaction. Significant variation are assigned an asterisk (*); $p<0.001$(***), $p<0.01$(**), $p<0.05$ (*) and no significant difference (ns).

5.4 Results

5.4.1 Leaf size Characteristics

These experiments were designed to follow key morphological and genomic events as young *P.nigra* leaves progressed through different stages of leaf growth. Leaf growth exhibits a linear relationship to time until it reaches its maximum growth potential, whereby growth plateaus. All *P.nigra* extremes followed a regular pattern of leaf growth, however final leaf size varied considerably between 'small' and 'big' leaf genotypes (

Figure 5-4 & Figure 5-5).Analysis of average final leaf area revealed that leaf area ranged from 58137mm^2 in the 'biggest' genotype to 5747mm^2 in the 'smallest genotype, indicating a 52390 mm^2 difference. Within 'small' and 'big' leaf groups final leaf area also varied. Within 'small' genotypes leaf area ranged from 5747 and 11335mm^2 indicating a 5588mm^2 difference (Figure 5-4), whereas within 'big' genotypes leaf area ranged from 58137 and 32439mm^2 with a 25697mm^2 difference (Figure 5-5). Leaf size extremes also showed considerable differences in rate of leaf development. 'Small' leaf genotypes completed growth at a mean of 33 days, whereas 'big' leaf genotypes finished after an average 35days, giving a 2 day difference (Figure 5-4 & Figure 5-5). The lowest number of days to reach maturity was seen in B7 a member of the 'small' leaf genotypes which took 27 days (Figure 5-4C), compared to 39 days in N38, a member of the 'big' leaf genotypes (Figure 5-5B). These findings confirm these genotypes to be 'small' and 'big' leaf extremes within *P.nigra*.

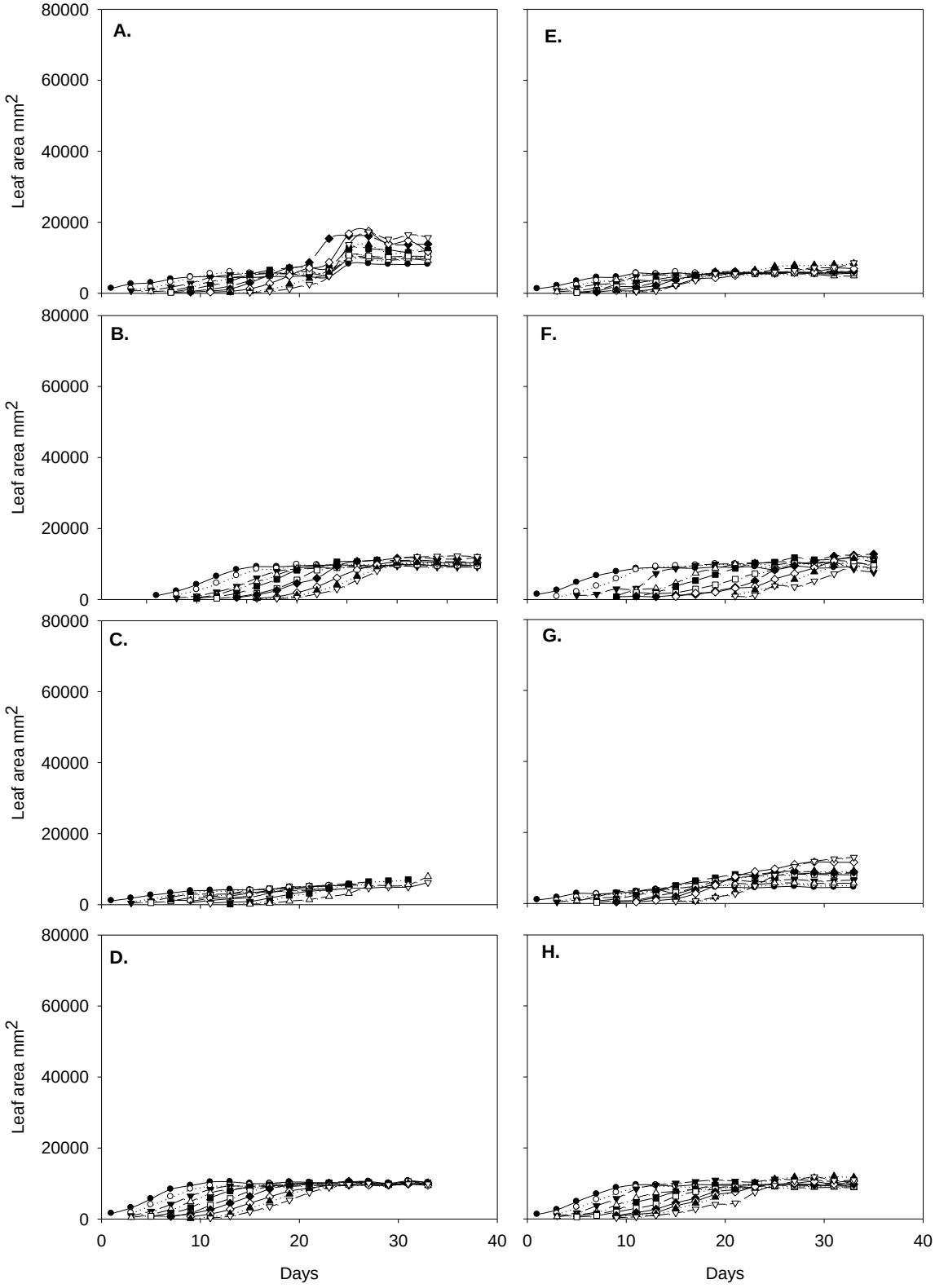


Figure 5-4: Leaf development in 'small' leaf area extremes. Time course - leaf development of 'small' leaf genotypes of *P.nigra* grown in a controlled greenhouse environment. Genotypes include; 71095-1 (A), 71092-36 (B), B7 (C), CART2 (D), C7 (E), C15 (F), FR7 (G) AND RIN2 (H). Leaf (L) number represent each line with the following symbols ;L1-●, L2 - ○, L3 -▼, L4- Δ, L5 - ■, L6 - □, L7 - ♦, L8 - ◊ , L9-▲and L10-○-Δ.

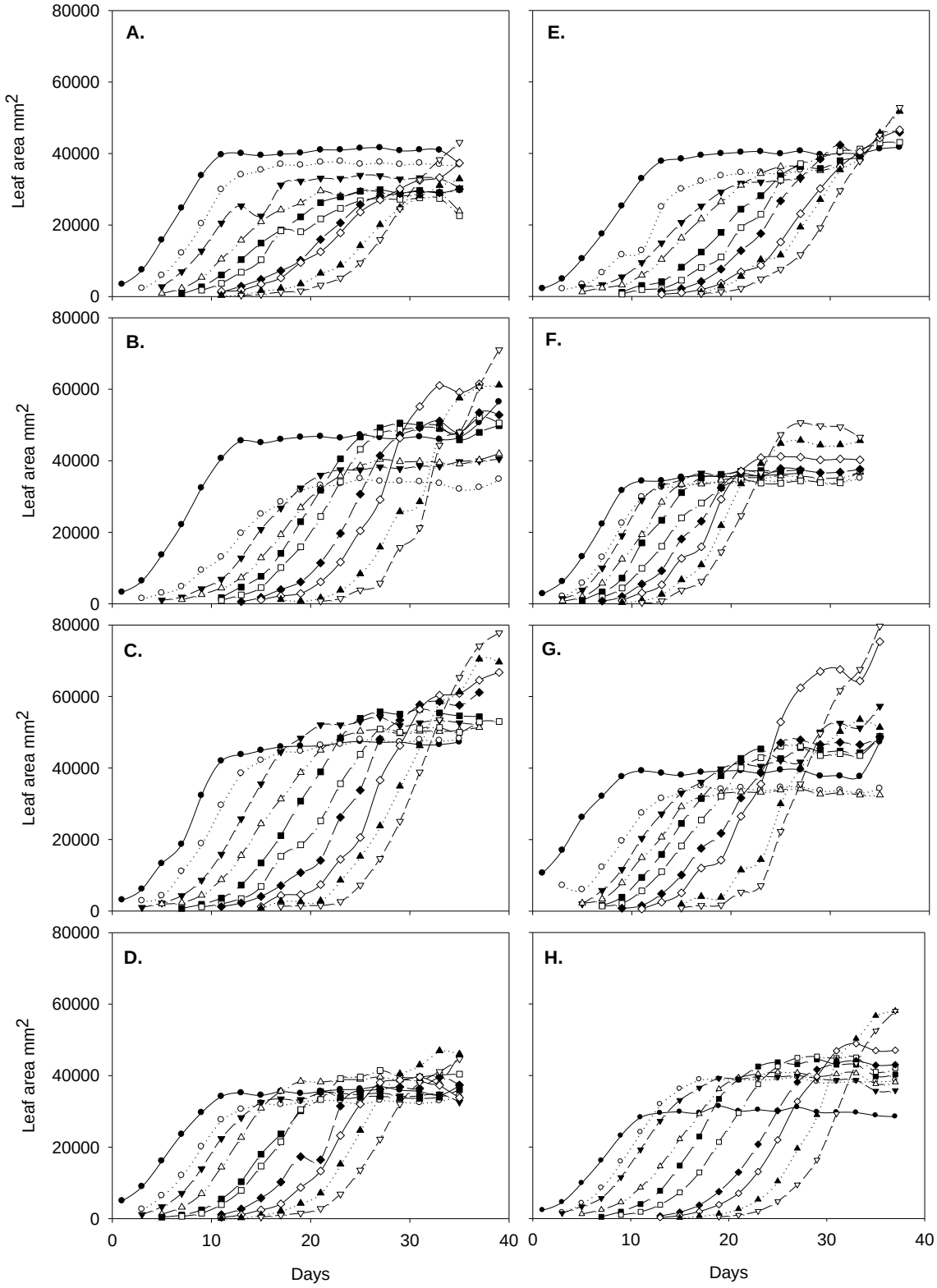


Figure 5-5: Leaf development in 'big' leaf area extremes. Time course - leaf development of 'big' genotypes of *P.nigra* grown in a controlled glasshouse environment. Genotypes include; N30 (A), N38 (B), N53 (C), N56 (D), N66 (E), NL1682 (F), NVHOF5-16 (G) and SN19 (H). Leaf (L) number represent each line with the following symbols ;L1-●, L2 - ○, L3 -▼, L4- Δ, L5 - ■, L6 - □, L7 - ♦, L8 - ◊ , L9-▲and L10 ○-Δ.

5.4.2 Leaf growth Characteristics

Leaf extension, expansion, absolute growth rate and relative growth rate showed a significant difference between genotypes (Figure 5-6A, D, B &E). All traits varied in a consistent pattern between the 'small' and 'big' leaf sample groups, whereby 'big' leaf genotypes had larger leaf extension, expansion and growth rates within the experimental period than those of 'small' leaf genotypes (Figure 5-6A, D, B & E). Interestingly, leaf production rate did not follow this pattern, rather the reverse happened, whereby 'small' leaf genotypes produced more leaves per day then 'big' leaf genotypes (Figure 5-6C). This indicates differences between 'big' and 'small' leaf genotypes in growth and development, with 'small' leaf genotypes having small leaf areas (Figure 5-4) and slow growth (Figure 5-6B& E) compared to 'big' leaf genotypes, though they produce more numerous smaller leaves (Figure 5-6C). Interestingly 'big' leaf genotypes gave significantly higher yields then 'small' leaf genotypes, indicating that leaf size rather then leaf number predicts yields.

5.4.3 Cell Characteristics

Stomatal density and stomatal index measurements showed significant variation between 'big' and 'small' leaf genotypes classes only on the adaxial leaf surface (Figure 5-7E & G). Stomatal density also varied with leaf surface, whereby stomatal density was higher on the abaxial leaf surface (Figure 5-7F). These results are consistent with field measurements taken in 2004 and 2006 (Chapter 2: 2.4) where little difference is seen between stomatal density, stomatal index and place of origin. Cell number per leaf was consistent on both leaf surfaces, displaying more cells per leaf in 'big' leaf genotypes compared to 'small' leaf genotypes (Figure 5-7A &B). Further investigation indicate that cell area is larger in 'small' leaf genotypes compared to 'big' leaf genotypes in both adaxial and abaxial leaf surfaces (Figure 5-7C & D). This observation indicate that 'big' leaf genotypes have a greater number of smaller cells per leaf compared to 'small' leaf genotypes which have fewer larger cells.

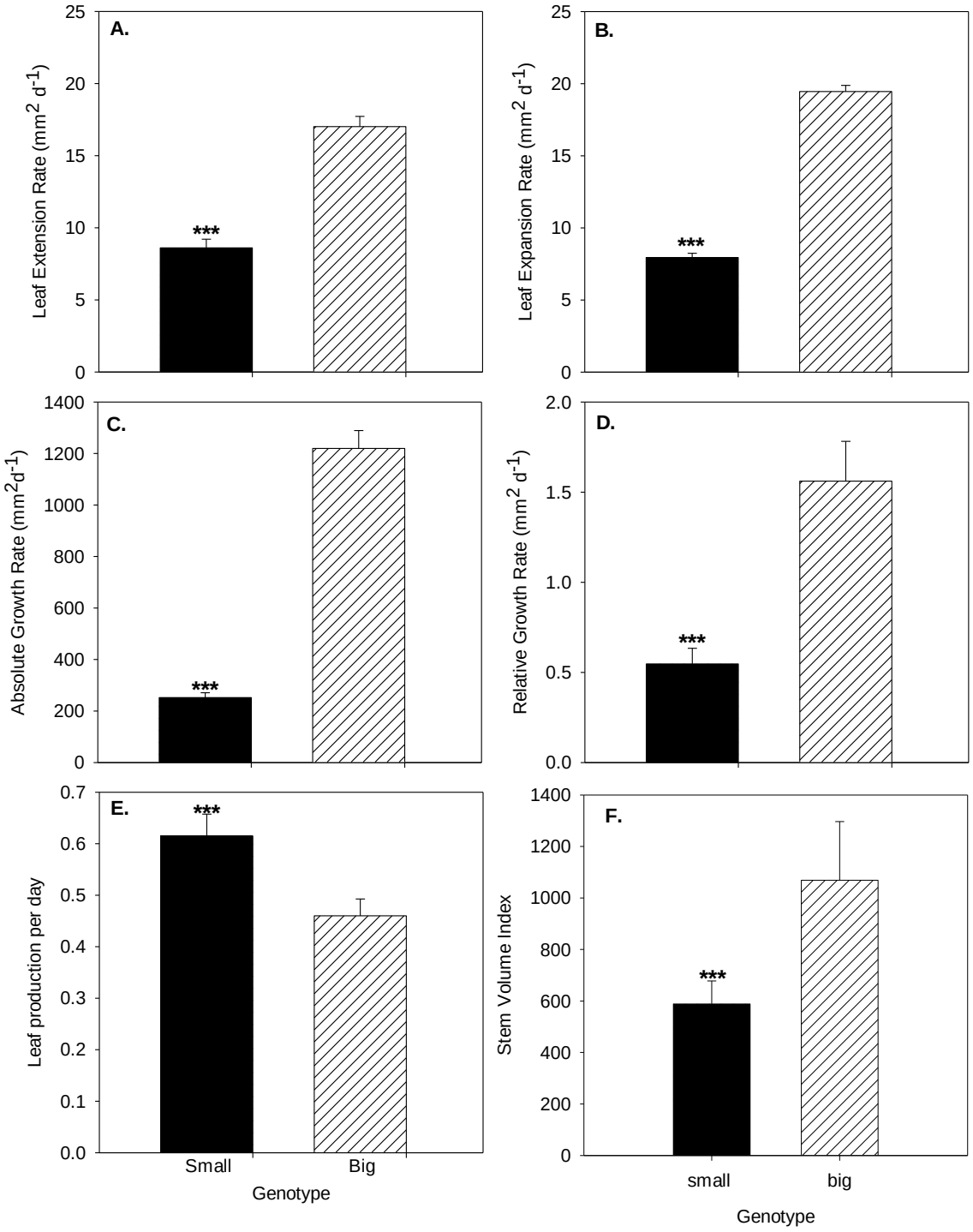


Figure 5-6: Leaf growth characteristics of *P.nigra* extremes. Consists of eight 'small' leaf genotypes (71095-1, 71092-36, B7, CART2, C7, C15, FR7 & RIN2) and eight 'big' leaf genotypes (N30, N38, N53, N56, N66, NL1682, NVHOF5-16 & SN19). (A) Variation in leaf extension, (B) Absolute growth rate, (C) Leaf production per day, (D) Leaf expansion rate, (E) Relative growth rate and (F) stem volume index. Error bars represent standard error of the mean, significant variation between 'small' and 'big' leaf genotype means are assigned an asterisk (*); $p<0.001$ (***), $p<0.01$ (**) and $p<0.05$ (*), no significance differences are blank.

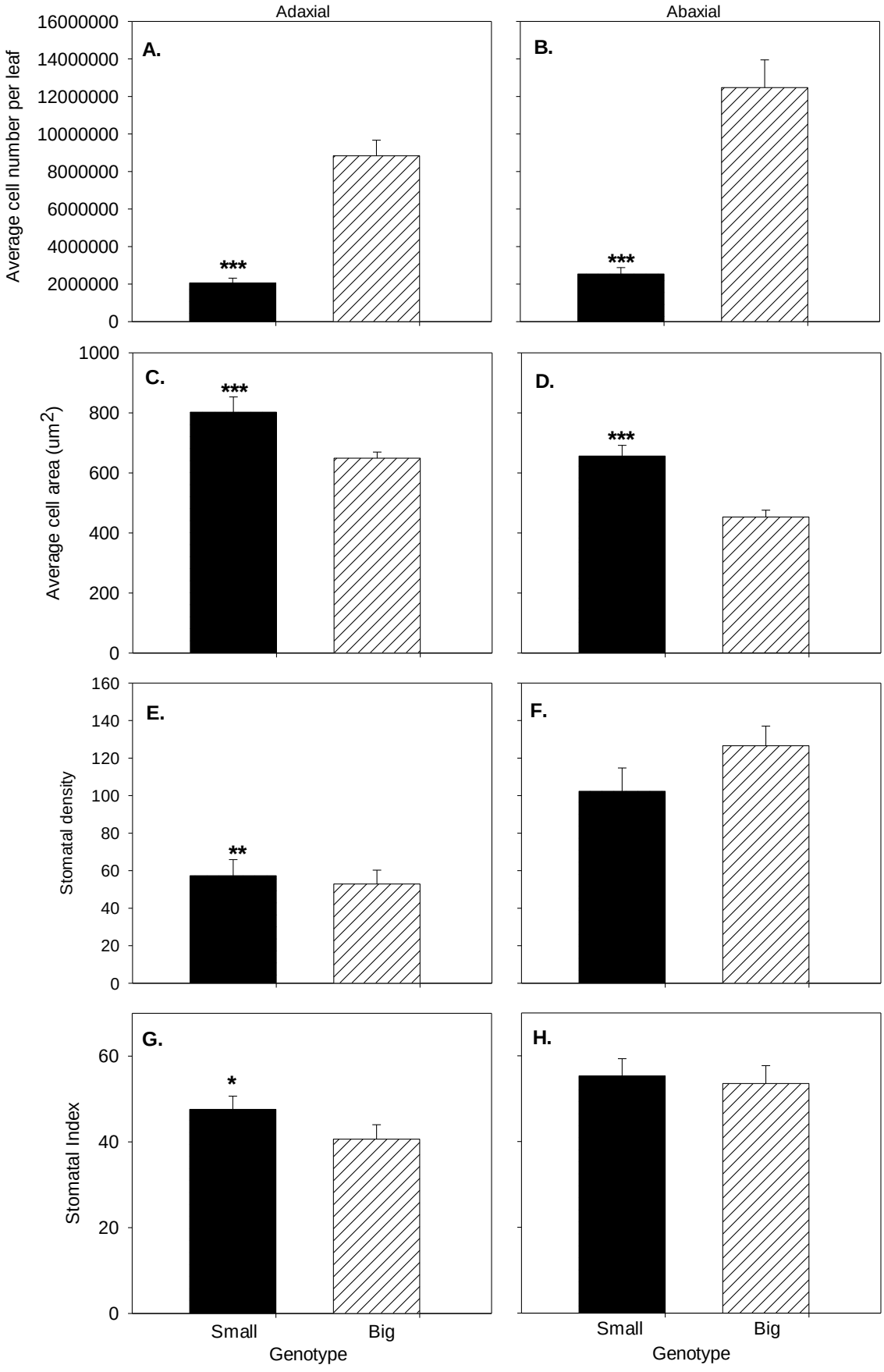

Adaxial
Abaxial
A.
B.
C.
D.
E.
F.
G.
H.
Average cell number per leaf
Average cell area (um2)
Stomatal density
Stomatal Index
***
***
***
***
**
*
Small
Big
Genotype
Small
Big
Genotype

Figure 5-7: Cell characteristics of *P.nigra* extremes Cell number per leaf, cell area, stomatal density and stomatal index in relation to differing *P.nigra* extremes consisting of 8 'small' leaf genotypes (71095-1, 71092-36, B7, CART2, C7, C15, FR7 & RIN2) and 8 'big' leaf genotypes (N30, N38, N53, N56, N66, NL1682, NVHOF5-16 & SN19), comparing upper (adaxial) and lower (abaxial) surfaces. (A) Variation in adaxial cell number per leaf, (B) variation in abaxial cell number per leaf, (C) variation in adaxial cell area, (D) variation in abaxial cell area, (E) variation in adaxial stomatal density, (F) variation in abaxial stomatal density, (G) variation in adaxial stomatal index and (H) variation in abaxial stomatal index. Error bars represent standard error of the mean, significant variation between 'small' and 'big' leaf genotype means are assigned an asterisk (*); $p<0.001$ (***), $p<0.01$ (**) and $p<0.05$ (*), no significance differences are blank.

5.4.4 Expressional analysis

To determine the temporal pattern of expression of developmental genes *ACO, AN, AS1, AS2, E2Fc* and *PHAB*, total RNA was isolated from meristems (leaf primordial), leaf age three (young leaves), leaf age six (semi-mature) from all 'big' leaf and 'small' leaf genotypes. RT-PCR amplification indicated that the expression of *ACO, AN, AS1, AS2* and *PHAB* occured at higher levels in the meristem then decreased as the leaf matures, indicating a role in leaf development (Figure 5-8A,,C,D,E &F). Our study has shown no significant differences between genotypes in leaf developmental gene expression of *ACO, E2Fc* and *PHAB* (Figure 5-8E, B, A), however variation has been seen in *AN* expression (Figure 5-8D). *AN* is up regulated in meristems in 'small' leaf genotypes compared 'big' leaf genotypes whereby *AN* is up regulated in leaf age three (Figure 5-8D). This indicates either differences between 'small' and 'big' leaf genotypes or experimental error as not all 'big' leaf genotypes showed consistency. Previous studies have indicated that *AS1* and *AS2* are found in the same pathway controlling leaf initiation. Within this study we have found that *AS1* is up-regulated in meristems, compared to *AS2*, indicating that *AS2* maybe suppressed (Figure 5-8F & C).

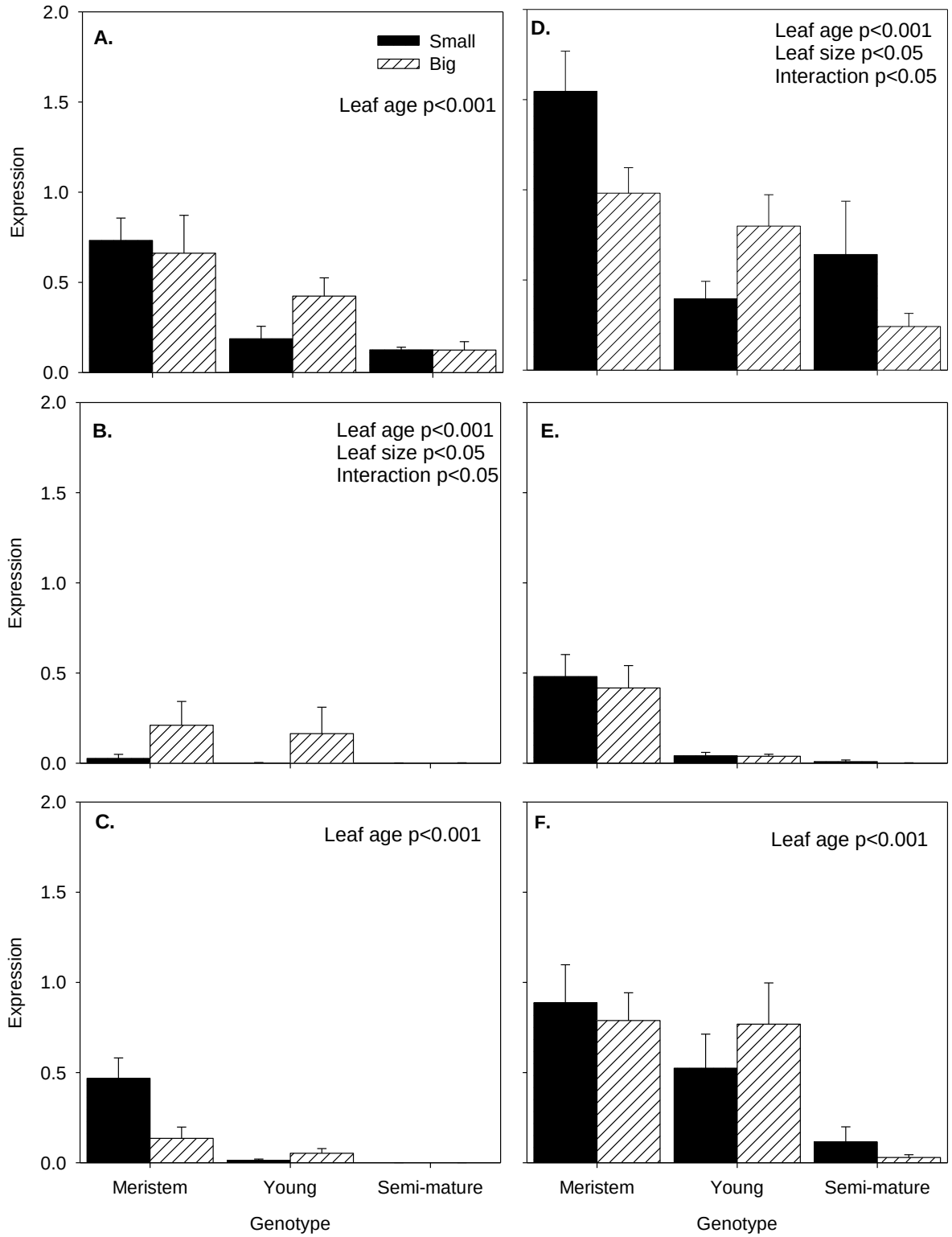


Figure 5-8: Expression of leaf developmental genes. Expression of *PHAB* (A), *E2Fc* (B), *AS2* (C), *AN* (D), *ACO* (E) and *AS1* (F) in 'small' (71095-1, 71092-36, B7, CART2, C7, C15, FR7 & RIN2) and 'large' (N30, N38, N53, N56, N66, NL1682, NVHOF5-16 & SN19) leaf genotypes of *P.nigra* at different developmental stages; merited (primordial), young (leaf three) and semi mature (leaf six) in a controlled glasshouse experiment. Error bars represent standard error of the mean, significant variation between 'small' and 'big' leaf genotype means are assigned an asterisk (*); p<0.001 (***), p<0.01 (**) and p<0.05 (*), no significance differences are blank.

5.5 Discussion

5.5.1 Leaf size, growth, biomass and cell characteristics in extremes

In this study a time course of leaf development verified genotypes selected from field observations represent leaf size extremes, whereby genotypes;71095-1, 71092-36, B7, CART2, C7, C15, FR7 and RIN2 are 'small' leaf genotypes (Figure 5-4) and N30, N38, N53, N56, N66, NL2682, NVHOF5-16 and SN19 are 'big' leaf genotypes (Figure 5-5). Investigations into the growth of 'small' compared to 'big' genotypes found that leaf extension rate, leaf expansion rate, absolute growth rate, relative growth rate and stem volume index were significantly larger in 'big' leaf genotypes (Figure 5-6).

These results indicate that breeders selecting trees for biomass crops that are fast growing with high yield should favor the 'big' leaf genotypes, as results strongly support the hypothesis that larger leaf sizes results in increased surface area for the absorption of radiation for photosynthesis, resulting in increased biomass (Pellis *et al*., 2004). Other studies have suggested leaf area index as a predictor of biomass as well as leaf size (Pellis *et al*., 2004), as trees can be high yielding when characterized by many small leaves, such as the *P.nigra* clone Wolterson, that has shown to be the best performing clone for biomass production yielding 8Mg $ha^{-1}y^{-1}$ (Pellis *et al*., 2004). However ,we found that 'small' leaf genotypes produced more leaves per day therefore characterized by having many small leaves similar to the *P.nigra* clone Wolterson, but 'small' leaf genotypes did not show as high a yield as 'big' leaf genotypes (Figure 5-6 F), suggesting leaf size and growth is a better predictor for biomass in a natural population of *P.nigra*. Leaf expansion also depends on carbon (C) availability, suggesting larger leaf sizes results in more photosynthesis and therefore more carbon for leaf expansion (Tardieu *et al*., 1999). However, studies covering 40% of the plants leaf area with aluminium foil have found no affects on the expansion rates in sunflower leaves (Tardieu *et al*., 1999). It is noteworthy also to point out that poplars native to certain latitude, but transplanted to another location or latitude, generally follow a rhythm of growth in accord with the day length of their place of origin (Ceulemans and Deraedt, 1999). Therefore an assumption can be made that as 'small' leaf genotypes originated from Spain (B7, CART2, C7, C15, FR7 &R RIN2) and France (71095-1 and 71092-36), whereas 'big' leaf genotypes are from northern Italy (N30, N38, N53, N56, N66 and SN19), Germany (NVHOF5-16) and the Netherlands (NL1682), differences in leaf size and growth are due to latitude of origin.

At the cellular level variation between 'small' and 'big' leaf genotypes average cell number per leaf was evident (Figure 5-7A & B). Average cell number per leaf was larger in 'big' leaf genotypes compared to 'small' leaf genotypes on the adaxial and abaxial leaf surface (Figure 5-7A & B). Average cell area was significantly larger in 'small' leaf extremes compared to 'big' leaf extremes on both leaf surfaces, suggesting that 'big' leaf genotypes consist of more cells that are small, whereas 'small' leaf genotypes consist of fewer cells that are big. Supporting the theory that leaf size is made by the sum of the behavior of each cell (Cookson *et al*., 2005), whereby leaf sizes are determined by cell production occurring during the early phase of leaf development, followed by cell expansion in later phases (Trapani *et al*., 1999). 'Big' leaf genotypes support this theory as an assumption can be made that at the early stages of leaf growth many cell were produced, followed by small increases in cell size at the later phases, whereas 'small' leaf genotypes produced less cells in the early stages of leaf development, followed by a greater expansion of cell size in the later stages. (Lecoeur *et al*., 1995), supported this theory in a study of pea, where cell division finished before leaf area was around 20% of its leaf area maximum. Cell expansion has been linked to cell walls via the activity of xyloglucan endotransglycosylase (XET), expansions or peroxidase under the control of Abscisic acid (ABA), were the cell wall is made plastic by the action of these enzymes that break the cellulose cross-linkages. Turgor pressure was first suggested as the driving forces, however turgor pressure is well maintained in plants, so chemicals signals such as ABA and hydraulic signals have been suggested (Tardieu *et al*., 1999). Cell proliferation is controlled by universally conserved molecular machinery known as the cell cycle consisting of one key player; Ser/Thr kinases, known as cyclin dependent kinases (CDKs) (Vandepoele *et al*., 2002). CDK-cyclin complexes phosphorylate a large number of substrates at the G1 to S and G2 to M transitions in the cell cycle, triggering the onset of DNA replication and mitosis (Boudolf *et al*., 2006). All eukaryotes contain a CDK with a PSTAIRE hallmark in their cyclin-binding domain, two CDK PSTAIRE have been discovered in plants and called CDKA. Studies have found that overproduction of a dominant negative CDKA in tobacco results in a reduction of cell division resulting in smaller plants (Boudolf *et al*., 2006), therefore changes in the genes controlling cell cycle could result in the phenotypic differences seen between 'small' and 'big' leaf genotypes.

Stomatal density and stomatal index varied significantly between 'small' and 'big' leaf genotypes on the adaxial surface of the leaf. Stomata are epidermal valves essential for plant survival as they control carbon dioxide assimilation in photosynthesis and optimize water use

efficiency (Bergmann and Sack, 2007). Each stoma is produced by a specialized cell lineage (Bergmann and Sack, 2007), therefore genes controlling stomatal development maybe resulting in changes in stomatal values seen in 'small' and 'big' leaf genotypes. In terms of stomatal number mutations in the gene *ERECTA* (*ER*) have resulted in stomatal over proliferations and spacing defects (Shpak *et al*., 2005) and loss of function mutations in the *MAP kinase kinase kinase gene YODA* results in a similar phenotype to *ER* with a reduction in plant height (Bergmann and Sack, 2007). Stomatal density has shown strong correlations to environmental parameters such as CO_2, humidity and light, therefore differences between 'small' and 'big' genotypes could be a result of their behavior to changes in the environment, due to their latitudinal climate origin . Short-term experiments have shown that *Arabidopsis* plants grown in elevated CO_2 produce fewer stomata per unit area than those grown in ambient CO_2 (Lake *et al*., 2002).

5.5.2 Gene expression in leaf development

Patterns in gene expression varied considerably between each gene studied over each leaf age, between 'small' and 'big' leaf genotypes. *ANGUSTIFOLIA* (*AN*) showed differences in expression between 'small' and 'big' leaf genotypes at all leaf ages. *AN* is involved in leaf morphology, mutational studies have resulted in alterations in microtubule cell pattern in mutant leaves causing altered leaf shape due to decreases in lamina expansion. *AN* expression is larger in 'small' leaf genotypes compared to 'big' leaf genotypes in the meristem and semi-mature leaves (Figure 5-8B), suggesting that *AN* plays a role in leaf expansion rate by reducing expansion in 'small' genotypes (Figure 5-6B). *AS1* and *AS2*s role in leaf development has been described in depth in *Arabidopsis*, whereby they repress *KNOX* gene to promote the expression of *LOB* (*LATERAL ORGAN BOUNDARIES*) to initiate leaf growth (Chalfun *et al*., 2005). Within this study *AS1* and *AS2* showed different patterns of expression with leaf age (Figure 5-8C & F). *AS1* and *AS2* were seen in both 'small' and 'big' leaf genotypes at the beginning of leaf development (in the meristem), however only *AS2* showed significant differences in expression between 'big' and 'small' leaf genotypes (Figure 5-8C). *AS2* expression is up regulated in 'small' leaf genotypes compared to 'big' leaf genotypes suggesting a role in leaf size and development. Mutational studies have found that *as2* mutants show changes in leaf shape, leaf lobbing and leaflet-like structures appear on the leaf (Chalfun *et al*., 2005), decreases in abaxialized vasculature formation (Ha *et al*., 2007) and lack of blade expansion (Zgurski *et al*., 2005). As *AS2* is highly expressed in 'small' leaf genotypes in the meristem an assumption can be made that *AS2* acts to reduce blade expansion, as leaf

expansion is lower in these genotypes (Figure 5-6B). Interestingly studies have found that ectopic *AS2* expression leads to the up-regulation of *PHABULOSA* (*PHAB*) and the repression of *KANADI 1* (*KAN1*), *KAN 2, YABBY* which are involved in leaf polarity, suggesting a these genes are involved in a similar pathway. *PHABULOSA* (PHAB) a member of the class III homeodomian-leucine zipper (*HD-ZIPIII*) protein family expression varied between leaf ages, with the highest levels of expression seen in the meristem, followed by young leaves and semi-mature leaves suggesting a decrease with age (Figure 5-8A), as seen in other studies whereby *PHAB* is expressed in tissue closest to the meristem as it is activated by a small diffusible lipid-based factor and decreases in expression at a distance from the meristem as it is non-activated (Fleming, 2003). *PHAB* controls adaxial/abaxial polarity, which is essential for blade outgrowth, mutational studies have resulted in rod-shaped leaves, lacking adaxial characters and without the SAM (Wenkel *et al*., 2007). Therefore *PHAB* is essential for leaf development and hence its presence in all leaf ages (Figure 5-8A). No significant differences were found between leaf extremes for *PHAB* expression (Figure 5-8A), suggesting *PHAB* is not the only gene controlling differences seen in leaf size. Studies have found that *PHAB* is part of a larger network of regulatory factors that establish adaxial/abaxial leaf fates, involving *AS1, AS2* that promote adaxial fates, *LOB* proteins acting as transcription factors and *KANADI, YABBY* and *AUXIN RESPONSE FACTORS 3* (*ARF3*) and *ARF4* operating on the abaxial side of the leaf, with the addition of microRNAs 165 and 166 controlling HD-ZIP III factors, suggesting that differences in leaf size could be controlled by many of these genes acting together. *ACC oxidase* (*ACO*) is an enzyme involved in the ethylene pathway. Ethylene is a gaseous plant hormone with a known function in fruit ripening, seed germination and organ senescence (Ma and Li, 2006). Ethylene plays a developmental role by mediating development and modifying growth patterns in response to a range of stresses and environmental cues during the plants life cycle (Andersson-Gunneras *et al*., 2003). Studies have found that increased amount of *ACO* results in increased ethylene (Qin *et al*., 2007) and that *ACO* is expressed in vegetative tissue (Qin *et al*., 2007). Studies have suggested that ethylene controls the limitation of cell wall expansion and therefore determines the size and shape of a plant. Within this study *ACO* was expressed highly in the meristems of both 'big' and 'small' leaf genotype, with no significant difference between them suggesting that ethylene is involved in leaf development, but is not the causal factor determining leaf size and shape. Investigation in to other enzymes in the ethylene pathway such *ACC synthase* could help in the understanding of ethylene's role in leaf development. *E2Fc* is involved in the retinoblastoma (*RB*)-*E2F* pathway which is one of the most important regulatory pathways

that control cell division and cell differentiation in both animals and plants (del Pozo *et al*., 2006). Regulation of cell cycle progression involves *E2F* and *DP* proteins interacting with active transcription factors that bind to different gene promoters and regulate expression of genes required in the cell cycle (del Pozo *et al*., 2006). *RB-related* (*RBR*) proteins block transcription activity by binding to *E2F* proteins, therefore reducing cell cycle progression. In this study we investigated *E2Fc*, one of six known *E2Fs* in Arabidopsis, very little is known about their number in poplar. Data indicate that *E2Fc* is up regulated in big leaves in the meristem and young leaves (Figure 5-8B), indicating a role in leaf development. Other study show that over-expression of *E2Fc* delays cell division and represses the expression of S-phase genes forcing cells into the endoreplication program (del Pozo *et al*., 2006); therefore, I conclude that 'big' leaf genotypes have less cell division then 'small' leaf genotypes. However, this is not the case, at the cellular level; average cell number per leaf is larger in 'big' leaf genotypes compared to 'small' leaf genotypes. Other studies have suggested that the activity of *E2Fc* is dependent on the level of transcription and the amount of *CDK* (del Pozo *et al*., 2006); this was not investigated in this study and is a point for further review.

Real time PCR is the technique of choice to analyse mRNA expression, however this technique does come with its own advantages and disadvantages. All currently available expression analysis is based on determining the threshold cycle (C_t), which is the fractional cycle number at which a fixed amount of DNA is formed (Ramakers *et al*., 2003). In this study, we assumed that the PCR efficiency of the interest amplicon is constant over time and has the same value in all studies samples by using the comparative C_t method (Ramakers *et al*., 2003). Some studies however have found that PCR efficiencies of both the target and reference amplicon can vary over a range from 1.8 – 2.0 (Ramakers *et al*., 2003), which effect the fold change in expression. However, within this study efficiencies of the target amplicon and reference amplicon were calculated to be equal, expression was then calculated based on relative expression of a target gene versus the average of two reference genes, therefore standardizing for inter-PCR variations (Pfaffl, 2001).

This study has led us one-step closer to the functional understanding of specific genes in leaf development. We have found that *PHAB*, *AN*, *E2Fc*, *ACO*, *AS2* and *AS1* are all expressed in leaf organs at different leaf development stages, concluding a function in leaf development. However we have described the organ location, but the biological function is still needed for these given genes. Therefore, it is necessary to identify a mutation in *P.nigra* in these genes, to

compare wild-type plants with plants that harbor this mutation (Ostergaard and Yanofsky, 2004).

5.5.3 Summary

In this study, we conclude that genotypes selected from 2004 and 2005 field data for leaf size were extreme; 'small' and 'big' leaf genotypes. I conclude that leaf growth characteristics; leaf extension rate, leaf expansion rate, absolute growth, relative growth, stem volume index and leaf productive per day varied significantly between 'small' and 'big' leaf genotypes. At the cellular level variation also occurred between 'small' and 'big' leaf genotypes, with the most significant difference seen in average cell number per leaf , indicating that 'small' leaf genotypes had fewer cell that were larger compared to 'big' leaf genotypes that had many cells that were smaller.

Real-time PCR using SYBR Green I fluorescence dye has proven to be a rapid and sensitive method to detect low amounts of mRNA molecules in leaf developmental genes, offering an insight into important physiological processes in leaf development. *AN* proved to be a key player in leaf development, varying between 'small' and 'big' leaf genotypes, within the meristem, young and semi-mature leaves. Significantly more *AN* was expressed in 'small' leaf genotypes suggesting that *AN* was acting to reduce blade expansion.

6 . General Discussion

6.1 Overview

Population geneticists and evolutionary biologists are interested in identifying the causal genes of natural variation that will affect fitness, resulting in evolutionary change through natural selection and adaptation. Parallel to this plant breeders have see the benefit of this information to identify causative polymorphisms for important agronomic traits, thus providing a powerful resource for genetic improvement of crops through direct allele selection (Haussmann *et al*., 2004;Gonzalez-Martinez *et al*., 2006).

The leaf is the basic organ for photosynthesis, therefore central to the life strategy of the plant. Leaves size has been strongly correlated to biomass in *Populus* (Taylor, 2002;Bunn *et al*., 2004;Rae *et al*., 2004;Monclus *et al*., 2005), making it a strong indictor of biomass. In this study a detailed investigation of leaf size in a natural population of *Populus nigra* was investigated using a modern genetical genomic approach to understand the key genes involved in leaf size and biomass. The principal finding of this thesis is combining QTL analysis, microarray and bioinformatics is an effective approach for the selection of candidate gene for leaf development and biomass. Genetic association conducted on selected leaf development genes: *ASYMMETRIC LEAVES 1 (AS1), ASYMMETRIC LEAVES 2 (AS2), ACC OXIDASE (ACO), ERECTA, PHABULOSA (PHAB), ANGUSTOFOLIA (AN), E2Fc* and *LEAFY* resulted in a strong association in *AS1* to diameter. Gene expression studies using real time qPCR in leaf primordia, growing and maturing leaves for the extreme leaf size genotypes grown in controlled conditions, revealed significant differences between 'small' and 'big' genotypes in *AN, AS2* and *AS1*. These results suggest that *AS1* is a strong candidate gene for leaf size and biomass.

Leaf development is genetically controlled and variation in leaf shape and size is due to species–specific patterns in leaf development (Kerstetter and Poethig, 1998). With this in mind, this thesis was introduced by a discussion on knowledge of the molecular processes controlling leaf development. A model for leaf development was proposed in *Arabidopsis*, beginning in the shoot apical meristem that contains all proliferative cells which give rise to the leaf, by managing the balance between cell division and cell differentiation. At the molecular level the model begins with; *CLAVATA1 (CLV1),* and *CLV3* genes promoting stem cell differentiation and *WUSCHEL* regulating the size of SAM (Fleming, 2006c). Meristem maintenance and initiation of leaves requires the *1 KNOTTED homeobox* (*KNOX*) gene transcription factors to be repressed, such as *SHOOTMERISTEMLESS* (*STM*), *BP, KNAT2* and *KNAT6* (Canales *et al*., 2005). *KNOX* down-regulation requires the activity of *AS1* a member

of the MYB proteins including *ROUGH SHEATH 2*, *PHANTASTICA* and *AS2* a member of the *LATERAL ORGAN BOUNDARIES* family (Canales *et al*., 2005). The homeodomain – leucine zipper III (HD_ZIPIII) family (including: *PHABULOSA* (*PHAB*) and *PHAVOLUTA (PHV)* then promote meristem and leaf identity (Canales *et al*., 2005). *YABBY* (*YAB3* and *YAB2*) transcription factors and *KANADI1 / KANADI2* (*KAN1*/ *KAN2*) (Eshed *et al*., 2001) promote abaxial cell fate (Canales *et al*., 2005). Followed by cell division and differentiation in the leaf to form the lamina controlled by the genes: *BLADE- ON- PETIOLE 1* (*BOP1*) (Ohno *et al*., 2004), *ASYMMETRIC LEAVES 1, ASYMMETRIC LEAVES 2* (Chalfun *et al*., 2005;Zgurski *et al*., 2005), *AINTEGUMENTA* (*ANT*) (Mizukami and Fischer, 2000), *JAG* (Ohno *et al*., 2004), the *GROWTH-REGULATING FACTOR* family (Horiguchi *et al*., 2005a), *ANGUSTIFOLIA* (*AN3*) and *ROTUNDIFOLIA* (*ROT3*) (Horiguchi *et al*., 2005a).

At the on-set of this work a general consensus was that the genes described above would be involved in leaf development in *P.nigra*, however these results were obtained in *Arabidopsis*. *Populus* and *Arabidopsis* lineages diverged 100 -120 million years ago (Ma) and poplar had a recent genome duplication event 65 Ma (Tuskan *et al*., 2006a), indicating that genes involved in leaf development could have adapted to different needs. Comparisons of the *Populus trichocarpa* sequence to *Arabidopsis* reveal only 9% of predicted gene models did not show similarity (Tuskan *et al*., 2006b). Data indicate that a strong leaf development study to find the genes involved should include a comparison study using all available *Arabidopsis* and poplar genomic resources. Therefore within this study a natural population of *P.nigra* was selected containing genotypes from different latitudes across Europe, to identify variation in leaf size and biomass. A candidate gene selection processes utilized the readily available *Populus trichocarpa* sequence and previous QTL analysis on the *Populus* family 331 genetic map, to co-locate leaf development genes found through literature searches and a microarray experiment comparing 'small' and 'big' leaf genotypes. Genetic association was then carried out on selected candidate genes and resulting associations verified by qPCR.

6.2 Chapter 2

The main findings of this chapter was that leaf traits:- leaf area, specific leaf area (SLA) , cell traits, cell number per leaf and biomass traits; height and diameter showed strong clinal variation across Europe. These finding combined with large heritability (h^2) and Q_{ST} values indicate traits to be adaptive. *P.nigra* leaf and biomass traits have adapted to different seasons, photoperiods, precipitation and temperature across a varying latitudinal gradient, which results in different growth opportunities. The growing season is shorter in the north

compared to the south, this could be one of the possible factors driving adaptive differentiation in this population.

An important consideration in breeding programmes for biomass crops is the selection of high yielding genotypes. Strong positive correlations between leaf area and height conclude that leaf area is a robust indictor of biomass and that genotypes with large leaf sizes will be ideal for biomass crops. Therefore trees located in higher latitudes such as Germany, the Netherlands and Northern Italy would be selected as high yielders for breeding purposes.

The three year field study of agronomic traits in *P.nigra* resulted in a suggestion to biomass crop programme lengths and management. In this study leaf size and height measurements of all genotypes were considerably lower in the establishment year (2004), compared to the second year of growth after coppicing (2005) and third year of growth (2006).Findings support general optimum rotation times suggested for poplar at around four years (Ceulemans and Deraedt, 1999).

6.3 Chapter 3

Combining, microarray analysis and literature searches as a leaf developmental search engine, followed by co-locating genes to the *Populus trichocarpa* sequence and the family 331 molecular map containing previously identified QTL regions proved to be a useful tool. In this study 293 differentially expressed ESTs from microarray analysis and 79 *Arabidopsis* leaf developmental genes from literature, resulted in 372 possible candidates. Utilizing the family 331 genetic map containing previously identified QTL regions and by using the *Populus trichocarpa* sequence to co-locate genes, set criteria for 'minor' and 'major' hotspots were determined resulting in 199 strong candidate genes for leaf development. Nine major hotspot regions were determined for leaf and biomass traits, on linkage groups; I, IV, V, VI, VIII, IX, X, XIII, XIV, XV and XVIII.

The candidate gene list was then further reduced, to eight possible candidates based on above criteria and known function resulting in eight candidate genes: *ASYMMETRIC LEAVES 1 (AS1) , ASYMMETRIC LEAVES 2 (AS2), ERECTA, PHABULOSA (PHAB), ANGUSTOFOLIA (ANT), E2Fc, ACC oxidase* and *LEAFY*. *AS1* is a member of the MYB related gene family, *AS2* is a member of the *LATERAL ORGAN BOUNDARIES* (*LOB*) domain gene family (Chalfun *et al*., 2005). These two genes play a role in repressing *KNOX* genes in the SAM to initiated leaf growth. *ERECTA* is a receptor like kinase (RLK) which regulates inflorescence architecture, mediates cell – cell signals that sense and co-ordinate organ growth (Shpak *et al*.,

2004). *PHABULOSA* (*PHAB*) is a member of the HD-ZIP III protein family involved in adaxial cell fate (Fleming, 2003). *ANGUSTIFOLIA* (*AN*) is a member of the Ct/BARS like protein genes, involved in the regulation of polarized growth of leaf cells (Kim *et al*., 2002a). *E2Fc* plays a functional role in the cell cycle, involved in the *retinoblastoma* (*RB*)-*E2F* pathway controlling cell division and cell differentiation (Stevaux and Dyson, 2002). *ACC OXIDASE* (*ACO*) plays major role in ethylene biosynthesis, is an enzyme in the pathway that converts 1-aminocyclopropane 1- carboxylate into ethylene. It is expressed in newly initiated leaves, indicating a role in leaf development. Finally *LEAFY* is a floral meristem identity gene in *Arabidopsis*. Over expressional studies in *P.trichocarpa* resulted in small deformed leaves (Rottmann *et al*., 2000).

Candidate gene selection combining multiple disciplines should however be used only as a tool, as increasing the number of techniques can increase error. Therefore final candidate genes selected were based on mutational studies displaying changes in leaf morphology, to conclude that they play a role in leaf development.

6.4 Chapter 4

A major finding in this chapter was identifying a strong association between the gene *AS1* and biomass trait diameter. Little is known concerning *AS1* role in diameter, however diameter is strongly correlated to leaf size in phenotypic studies. Several studies have found that *AS1* is involved in the repression of *KNOX* genes to initiate leaf growth (Zgurski *et al*., 2005). Two pathways have resulted from mutational studies in *Arabidopsis* indicating *AS1*s role in leaf development; (i) the first pathway indicates that *STM* represses *AS1* and *AS2* in the stem cells of the meristem reducing the expression of *LOB* for lamina growth, as *AS1* and *AS2* repress *KNAT1* and *KNAT6* which is required for inducing *LOB* (Byrne *et al*., 2001), (ii) the second pathway suggests that *BOP1* and *BOP2* activate *AS1* and *AS2* which then repress *BP*, *KNAT2* and *KNAT6* to induce *LOB* (Ha *et al*., 2007).

Interestingly *AS1* did not co-locate to hotspot QTL regions and was selected due to role with *AS2*. However other genes involved in this pathway did co-locate to major and minor hotspots, such as *BOP1* co-locate to a major hotspot found at the top of Linkage Group VI above *AS1*, *KNAT6* co-located to a major hotspot on Linkage group XV, *AS2* co-located to a major hotspot on Linkage Group VIII and *LOB* co-located to a minor hotspot on Linkage Group XV, inferring that this pathway is a area for further investigation. On the flip side *AS1*s lack of co-

location to QTL regions could be due to low resolution of the molecular map and species differences in technique used.

Before implementing association tests several factors were taken into consideration, including population structure, LD and phenotypic variation. Moderate F_{ST} values and clustering analysis using K in STRUCTURE have indicated moderate genetic differentiation; however, therefore for a more robust association analysis structure was included in the analysis to remove false-positive associations seen.

Q_{ST} was larger then F_{ST}, for all quantitative traits concluding that directional selection is acting on these traits and fixing particular genes (Palo *et al*., 2003).

6.5 Chapter 5

In this chapter I verified that leaf size extremes selected in 2004 and 2005 field trial measurements were indeed extreme 'small' and extreme 'big' leaf genotypes in a controlled glasshouse environment. Growth measurements concluded that 'big' leaf genotypes were characterized by large leaf extension rates, leaf expansion rates, absolute growth rates, relative growth rates and a larger stem volume index compared to 'small' leaf genotypes. At the cellular level 'big' leaf genotypes had more average cells per leaf that were smaller, indicating that 'big' genotypes leaves contained many small cells, whereas 'small' genotypes leaves contained less cells that are bigger.

Expressional studies concluded that *PHAB*, *E2Fc*, *AS2*, *AN*, *ACO* and *AS1* are all expressed in leaves, varying in relative expression over leaf ages. *AN*, *AS1* and *AS2* differed significantly between 'small' and 'big' genotypes, indicating a role in variation found. For example *AN* was more highly expressed in 'small' genotypes in both the meristems and semi-mature leaves. Combining phenotypic analysis and expression data I conclude that as *AN* is involved in lamina expansion in 'small' genotypes by reducing expansion of leaves. *AS2* is expressed in the meristem of developing leaves, showing a significant difference between 'small' and 'big' leaf genotypes, I conclude that *AS2* plays a role in altering leaf size by reducing blade expansion as *AS2* is highly expressed in 'small' leaf genotypes.

Genetic associations concluded that *AS1* plays a role in the variation in diameter found with *P.nigra*. In this study we find that *AS1* is expressed in all leaf ages, however significant differences between 'small' and 'big' leaf genotypes was found only in semi-mature leaves. This concludes that *AS1* plays a role in leaf development, but is not the causal gene, resulting

in leaf size variation and that *AS1*s role in the later stages of development are worthy of further study. Finally to verify finding a mutation developed in *P.nigra* for *AS1* would be of great interest to confirm findings.

6.6 Summary

As far as I am aware no detailed study of leaf development has been carried out in a natural population of *P.nigra*, making this thesis the first of its kind. The subject matter covered in this thesis is topical concerning both European and UK government targets to reduce CO_2 emission by providing electricity from renewable energy sources. Poplar as a bioenergy crop is essential to help governments meet these targets. Strong correlations between leaf size and biomass (Bunn *et al*., 2004;Rae *et al*., 2004;Monclus *et al*., 2005), have indicated that leaf size is a strong predictor of biomass. Therefore understanding the molecular and physiological mechanisms controlling leaf size is of great importance for successful breeding programmes. Data presented in this thesis identifies *P.nigra* genotypes for breeding programmes and identifies the gene *AS1* as a strong candidate for biomass.

7 . Bibliography

Akamatsu, T., Hanzawa, Y., Ohtake, Y., Takahashi, T., Nishitani, K. & Komeda, Y. (1999) *Expression of endoxyloglucan transferase genes in acaulis mutants of Arabidopsis*. *Plant Physiology,* **121,** 715-721.

Al Afas, N., Pellis, A., Niinemets, U. & Ceulemans, R. (2005) *Growth and production of a short rotation coppice culture of poplar. II. Clonal and year-to-year differences in leaf and petiole characteristics and stand leaf area index*. *Biomass & Bioenergy,* **28,** 536-547.

Anderson, L. J., Maherali, H., Johnson, H. B., Polley, H. W. & Jackson, R. B. (2001) *Gas exchange and photosynthetic acclimation over subambient to elevated CO2 in a C-3-C-4 grassland*. *Global Change Biology,* **7,** 693-707.

Andersson-Gunneras, S., Hellgren, J. M., Bjorklund, S., Regan, S., Moritz, T. & Sundberg, B. (2003) *Asymmetric expression of a poplar ACC oxidase controls ethylene production during gravitational induction of tension wood*. *Plant Journal,* **34,** 339-349.

Andersson, A., Keskitalo, J., Sjodin, A., Bhalerao, R., Sterky, F., Wissel, K., Tandre, K., Aspeborg, H., Moyle, R., Ohmiya, Y., Bhalerao, R., Brunner, A., Gustafsson, P., Karlsson, J., Lundeberg, J., Nilsson, O., Sandberg, G., Strauss, S., Sundberg, B., Uhlen, M., Jansson, S. & Nilsson, P. (2004) *A transcriptional timetable of autumn senescence*. *Genome Biology,* **5**.

Aphalo, P. J., Lahti, M., Lehto, T., Repo, T., Rummukainen, A., Mannerkoski, H. & Finer, L. (2006) *Responses of silver birch saplings to low soil temperature*. *Silva Fennica,* **40,** 429-442.

Barker, D., Green, P., Knowlton, R., Schumm, J., Lander, E., Oliphant, A., Willard, H., Akots, G., Brown, V., Gravius, T., Helms, C., Nelson, C., Parker, C., Rediker, K., Rising, M., Watt, D., Weiffenbach, B. & Doniskeller, H. (1987) *GENETIC-LINKAGE MAP OF HUMAN CHROMOSOME-7 WITH 63 DNA MARKERS*. *Proceedings of the National Academy of Sciences of the United States of America,* **84,** 8006-8010.

Beerling, D. J. & Royer, D. L. (2002) *Reading a CO2 signal from fossil stomata*. *New Phytologist,* **153,** 387-397.

Belles-Boix, E., Hamant, O., Witiak, S. M., Morin, H., Traas, J. & Pautot, V. (2006) *KNAT6: an Arabidopsis homeobox gene involved in meristem activity and organ separation*. *Plant Cell,* **18,** 1900-1907.

Bergmann, D. C. & Sack, F. D. (2007) *Stomatal development*. *Annual Review of Plant Biology,* **58,** 163-181.

Bhattramakki, D., Dolan, M., Hanafey, M., Wineland, R., Vaske, D., Register, J. C., 3rd, Tingey, S. V. & Rafalski, A. (2002) *Insertion-deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers*. *Plant Mol Biol,* **48,** 539-547.

Boudolf, V., Inze, D. & De Veylder, L. (2006) *What if higher plants lack a CDC25 phosphatase*. *Trends in Plant Science,* **11,** 474-479.

Bradshaw, H. D., Ceulemans, R., Davis, J. & Stettler, R. (2000) *Emerging model systems in plant biology: Poplar (Populus) as a model forest tree*. *Journal of Plant Growth Regulation,* **19,** 306-313.

Bradshaw, H. D. & Stettler, R. F. (1993) *Molecular-Genetics of Growth and Development in Populus .1. Triploidy in Hybrid Poplars*. *Theoretical and Applied Genetics,* **86,** 301-307.

Bradshaw, H. D., Villar, M., Watson, B. D., Otto, K. G., Stewart, S. & Stettler, R. F. (1994) *MOLECULAR-GENETICS OF GROWTH AND DEVELOPMENT IN POPULUS .3. A GENETIC-LINKAGE MAP OF A HYBRID POPLAR COMPOSED OF RFLP, STS, AND RAPD MARKERS*. *Theoretical and Applied Genetics,* **89,** 167-178.

Brosche, M., Vinocur, B., Alatalo, E. R., Lamminmaki, A., Teichmann, T., Ottow, E. A., Djilianov, D., Afif, D., Bogeat-Triboulot, M. B., Altman, A., Polle, A., Dreyer, E., Rudd, S., Lars, P., Auvinen, P. & Kangasjarvi, J. (2005) *Gene expression and metabolite profiling of Populus euphratica growing in the Negev desert*. *Genome Biology,* **6**.

Brunner, A. M., Busov, V. B. & Strauss, S. H. (2004) *Poplar genome sequence: functional genomics in an ecologically dominant plant species*. *Trends in Plant Science,* **9,** 49-56.

Brunner, A. M., DiFazio, S. P. & Groover, A. T. (2007) *Forest genomics grows up and branches out*. *New Phytologist,* **174,** 710-713.

Buckler, E. S. & Thornsberry, J. M. (2002) *Plant molecular diversity and applications to genomics*. *Current Opinion in Plant Biology,* **5,** 107-111.

Bunn, S. M., Rae, A. M., Herbert, C. S. & Taylor, G. (2004) *Leaf-level productivity traits in Populus grown in short rotation coppice for biomass energy*. *Forestry,* **77,** 307-323.

Byrne, M., Timmermans, M., Kidner, C. & Martienssen, R. (2001) *Development of leaf shape*. *Current Opinion in Plant Biology,* **4,** 38-43.

Byrne, M. E., Simorowski, J. & Martienssen, R. A. (2002) *ASYMMETRIC LEAVES1 reveals knox gene redundancy in Arabidopsis*. *Development,* **129,** 1957-1965.

Canales, C., Grigg, S. & Tsiantis, M. (2005) *The formation and patterning of leaves: recent advances*. *Planta,* **221,** 752-756.

Cano, J. M., Laurila, A., Palo, J. & Merila, J. (2004) *Population differentiation in G matrix structure due to natural selection in Rana temporaria*. *Evolution,* **58,** 2013-2020.

Cervera, M. T., Storme, V., Ivens, B., Gusmao, J., Liu, B. H., Hostyn, V., Van Slycken, J., Van Montagu, M. & Boerjan, W. (2001a) *Dense genetic linkage maps of three Populus species (Populus deltoides, P-nigra and P-trichocarpa) based on AFLP and microsatellite markers*. *Genetics,* **158,** 787-809.

Cervera, M. T., Storme, V., Ivens, B., Gusmao, J., Liu, B. H., Hostyn, V., Van Slycken, J., Van Montagu, M. & Boerjan, W. (2001b) *Dense genetic linkage maps of three Populus species (Populus deltoides, P. nigra and P. trichocarpa) based on AFLP and microsatellite markers*. *Genetics,* **158,** 787-809.

Ceulemans, R. & Deraedt, W. (1999) *Production physiology and growth potential of poplars under short-rotation forestry culture*. *Forest Ecology and Management,* **121,** 9-23.

Chalfun, A., Franken, J., Mes, J. J., Marsch-Martinez, N., Pereira, A. & Angenent, G. C. (2005) *ASYMMETRIC LEAVES2-LIKE1 gene, a member of the AS2/LOB family, controls proximal-distal patterning in Arabidopsis petals*. *Plant Molecular Biology,* **57,** 559-575.

Chen, B. C. & McManus, M. T. (2006) *Expression of 1-aminocyclopropane-1-carboxylate (ACC) oxidase genes during the development of vegetative tissues in white clover (Trifolium repens L.) is regulated by ontological cues*. *Plant Mol Biol,* **60,** 451-467.

Chen, X. N., Levine, L. & Kwok, P. Y. (1999) *Fluorescence polarization in homogeneous nucleic acid analysis*. *Genome Research,* **9,** 492-498.

Chang S, Puryear J, Cairney J. 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**: 114–117.

Cheng, Q., Zhang, B., Zhuge, Q., Zeng, Y. R., Wang, M. X. & Huang, M. R. (2006) *Expression profiles of two novel lipoxygenase genes in Populus deltoides*. *Plant Science,* **170,** 1027-1035.

Chuine, I., Belmonte, J. & Mignot, A. (2000) *A modelling analysis of the genetic variation of phenology between tree populations*. *Journal of Ecology,* **88,** 561-570.

Churchill, G. A. & Doerge, R. W. (1994) *EMPIRICAL THRESHOLD VALUES FOR QUANTITATIVE TRAIT MAPPING*. *Genetics,* **138,** 963-971.

Cnops, G., Neyt, P., Raes, J., Petrarulo, M., Nelissen, H., Malenica, N., Luschnig, C., Tietz, O., Ditengou, F., Palme, K., Azmi, A., Prinsen, E. & Van Lijsebettens, M. (2006) *The TORNADO1 and TORNADO2 genes function in several patterning processes during early leaf development in Arabidopsis thaliana*. *Plant Cell,* **18,** 852-866.

Coen, E. S., Romero, J. M., Doyle, S., Elliott, R., Murphy, G. & Carpenter, R. (1990) *Floricaula - a Homeotic Gene Required for Flower Development in Antirrhinum-Majus*. *Cell,* **63,** 1311-1322.

Cookson, S. J., Van Lijsebettens, M. & Granier, C. (2005) *Correlation between leaf growth variables suggest intrinsic and early controls of leaf size in Arabidopsis thaliana*. *Plant Cell and Environment,* **28,** 1355-1366.

Cosgrove, D. J. (1999) *Enzymes and other agents that enhance cell wall extensibility*. *Annual Review of Plant Physiology and Plant Molecular Biology,* **50,** 391-417.

Cottrell, J. E., Krystufek, V., Tabbener, H. E., Milner, A. D., Connolly, T., Sing, L., Fluch, S., Burg, K., Lefevre, F., Achard, P., Bordacs, S., Gebhardt, K., Vornam, B., Smulders, M. J. M., Broeck, A. H. V., Van Slycken, J., Storme, V., Boerjan, W., Castiglione, S., Fossati, T., Alba, N., Agundez, D., Maestro, C., Notivol, E., Bovenschen, J. & van Dam, B. C. (2005) *Postglacial migration of Populus nigra L.: lessons learrit from chloroplast DNA*. *Forest Ecology and Management,* **206,** 71-90.

De Veylder, L., Beeckman, T., Beemster, G. T. S., Krols, L., Terras, P., Landrieu, I., Van der Schueren, E., Maes, S., Naudts, M. & Inze, D. (2001) *Functional analysis of cyclin-dependent kinase inhibitors of Arabidopsis*. *Plant Cell,* **13,** 1653-1667.

DeBussche, M., Garnier, E. & Thompson, J. D. (2004) *Exploring the causes of variation in phenology and morphology in Mediterranean geophytes: a genus-wide study of Cyclamen*. *Botanical Journal of the Linnean Society,* **145,** 469-484.

Dejardin, A., Leple, J. C., Lesage-Descauses, M. C., Costa, G. & Pilate, G. (2004) *Expressed sequence tags from poplar wood tissues - A comparative analysis from multiple libraries*. *Plant Biology,* **6,** 55-64.

del Pozo, J. C., Diaz-Trivino, S., Cisneros, N. & Gutierrez, C. (2006) *The balance between cell division and endoreplication depends on E2FC-DPB, transcription factors regulated by the ubiquitin-SCFSKP2A pathway in Arabidopsis*. *Plant Cell,* **18,** 2224-2235.

Dewitte, W. & Murray, J. A. (2003) *The plant cell cycle*. *Annu Rev Plant Biol,* **54,** 235-264.

Ding, K., Zhou, K., He, F. & Shen, Y. (2003) *LDA--a java-based linkage disequilibrium analyzer*. *Bioinformatics,* **19,** 2147-2148.

Donnelly, P. M., Bonetta, D., Tsukaya, H., Dengler, R. E. & Dengler, N. G. (1999) *Cell cycling and cell enlargement in developing leaves of Arabidopsis*. *Developmental Biology,* **215,** 407-419.

Douglas, C. J. & Ehlting, J. (2005) *Arabidopsis thaliana full genome longmer microarrays: a powerful gene discovery tool for agriculture and forestry*. *Transgenic Res,* **14,** 551-561.

Doyle, J. J., Doyle, J. L., Brown, A. H. & Grace, J. P. (1990) *Multiple origins of polyploids in the Glycine tabacina complex inferred from chloroplast DNA polymorphism*. *Proc Natl Acad Sci U S A,* **87,** 714-717.

du Cros, E. T., Bilger, I. & Lefevre, F. (2001) *Achievements and ongoing programmes pertaining to conservation of forest gene resources in France and Europe*. *Revue Forestiere Francaise Numero Special (Nancy)***,** 117-124.

Dunlap, J. M., Heilman, P. E. & Stettler, R. F. (1995) *GENETIC-VARIATION AND PRODUCTIVITY OF POPULUS-TRICHOCARPA AND ITS HYBRIDS .8. LEAF AND CROWN MORPHOLOGY OF NATIVE POPULUS-TRICHOCARPA CLONES FROM*

4 RIVER VALLEYS IN WASHINGTON. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **25,** 1710-1724.
Ellsworth, D. S. (1999) *CO2 enrichment in a maturing pine forest: are CO2 exchange and water status in the canopy affected? Plant Cell and Environment,* **22,** 461-472.
Emery, J. F., Floyd, S. K., Alvarez, J., Eshed, Y., Hawker, N. P., Izhaki, A., Baum, S. F. & Bowman, J. L. (2003) *Radial patterning of Arabidopsis shoots by class IIIHD-ZIP and KANADI genes*. *Current Biology,* **13,** 1768-1774.
Eppley, S. M. (2006) *Females make tough neighbors: sex-specific competitive effects in seedlings of a dioecious grass*. *Oecologia,* **146,** 549-554.
Eshed, Y., Baum, S. F., Perea, J. V. & Bowman, J. L. (2001) *Establishment of polarity in lateral organs of plants*. *Current Biology,* **11,** 1251-1260.
Falconer, D.S & Mackay, T.F.C. (1996) Introduction to quantitative genetics. 4th Edition. *Pearson Education Limited*
Farquhar, G. D. (1989) *MODELS OF INTEGRATED PHOTOSYNTHESIS OF CELLS AND LEAVES. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences,* **323,** 357-367.
Fleming, A. (2003) *The molecular regulation of leaf form*. *Plant Biology,* **5,** 341 - 349.
Fleming, A. (2005) *The control of leaf development*. *New Phytologist,* **166,** 9 - 20.
Fleming, A. (2006a) *The co-ordination of cell division, differentiation and morphogenesis in the shoot apical meristem: a perspective*. *Journal of Experimental Botany,* **57,** 25 - 32.
Fleming, A. (2006b) *The integration of cell proliferation and growth in leaf morphogenesis*. *Journal of Plant Research,* **119,** 31 - 36.
Fleming, A. (2006c) *Leaf initiation: The integration of growth and cell division*. *Plant Molecular Biology,* **60,** 905 - 914.
Fleming, A. J., McQueenMason, S., Mandel, T. & Kuhlemeier, C. (1997) *Induction of leaf primordia by the cell wall protein expansion*. *Science,* **276,** 1415-1418.
Flint-Garcia, S. A., Thornsberry, J. M. & Buckler, E. S. (2003) *Structure of linkage disequilibrium in plants*. *Annual Review of Plant Biology,* **54,** 357-374.
Flint-Garcia, S. A., Thuillet, A. C., Yu, J. M., Pressoir, G., Romero, S. M., Mitchell, S. E., Doebley, J., Kresovich, S., Goodman, M. M. & Buckler, E. S. (2005) *Maize association population: a high-resolution platform for quantitative trait locus dissection*. *Plant Journal,* **44,** 1054-1064.
Folkers, U., Kirik, V., Schobinger, U., Falk, S., Krishnakumar, S., Pollock, M. A., Oppenheimer, D. G., Day, I., Reddy, A. S., Jurgens, G. & Hulskamp, M. (2002) *The cell morphogenesis gene ANGUSTIFOLIA encodes a CtBP/BARS-like protein and is involved in the control of the microtubule cytoskeleton*. *EMBO J,* **21,** 1280-1288.
Francis, D. (2007) *The plant cell cycle - 15 years on*. *New Phytologist,* **174,** 261-278.
Frewen, B. E., Chen, T. H. H., Howe, G. T., Davis, J., Rohde, A., Boerjan, W. & Bradshaw, H. D. (2000) *Quantitative trait loci and candidate gene mapping of bud set and bud flush in Populus*. *Genetics,* **154,** 837-845.
Fry, S. C., Smith, R. C., Renwick, K. F., Martin, D. J., Hodge, S. K. & Matthews, K. J. (1992) *XYLOGLUCAN ENDOTRANSGLYCOSYLASE, A NEW WALL-LOOSENING ENZYME-ACTIVITY FROM PLANTS*. *Biochemical Journal,* **282,** 821-828.
Gardner, S. D. L., Taylor, G. & Bosac, C. (1995) *LEAF GROWTH OF HYBRID POPLAR FOLLOWING EXPOSURE TO ELEVATED CO2*. *New Phytologist,* **131,** 81-90.
Gaudet, M., Fara, A. G., Sabatti, M., Kuztninsky, E. & Mugnozza, G. S. (2007) *Single-reaction for SNP genotyping on agarose gel by allele-specific PCR in black poplar (Populus nigra L.)*. *Plant Molecular Biology Reporter,* **25,** 1-9.
Geisler, M. & Sack, F. D. (2002) *Variable timing of developmental progression in the stomatal pathway in Arabidopsis cotyledons*. *New Phytologist,* **153,** 469-476.

Gonzalez-Martinez, S. C., Krutovsky, K. V. & Neale, D. B. (2006) *Forest-tree population genomics and adaptive evolution*. *New Phytologist,* **170,** 227-238.
Gonzalez-Martinez, S. C., Wheeler, N. C., Ersoz, E., Nelson, C. D. & Neale, D. B. (2007) *Association genetics in Pinus taeda L. I. Wood property traits*. *Genetics,* **175,** 399-409.
Goodrich, J., Puangsomlee, P., Martin, M., Long, D., Meyerowitz, E. M. & Coupland, G. (1997) *A Polycomb-group gene regulates homeotic gene expression in Arabidopsis*. *Nature,* **386,** 44-51.
Gordon, D. (2003) *Viewing and editing assembled sequences using Consed*. *Curr Protoc Bioinformatics,* **Chapter 11,** Unit11.12.
Gordon, D., Abajian, C. & Green, P. (1998) *Consed: A graphical tool for sequence finishing*. *Genome Research,* **8,** 195-202.
Gordon, D., Desmarais, C. & Green, P. (2001) *Automated finishing with Autofinish*. *Genome Research,* **11,** 614-625.
Goudet, J., Perrin, N. & Waser, P. (2002) *Tests for sex-biased dispersal using bi-parentally inherited genetic markers*. *Mol Ecol,* **11,** 1103-1114.
Gray, J. E., Holroyd, G. H., van der Lee, F. M., Bahrami, A. R., Sijmons, P. C., Woodward, F. I., Schuch, W. & Heterington, A. M. (2000) *The HIC signalling pathway links CO2 perception to stomatal development*. *Nature,* **408,** 713-716.
Gregis, V., Sessa, A., Colombo, L. & Kater, M. M. (2008) *AGAMOUS-LIKE24 and SHORT VEGETATIVE PHASE determine floral meristem identity in Arabidopsis*. *Plant J*.
Groover, A., Fontana, J. R., Dupper, G., Ma, C. P., Martienssen, R., Strauss, S. & Meilan, R. (2004) *Gene and enhancer trap tagging of vascular-expressed genes in poplar trees*. *Plant Physiology,* **134,** 1742-1751.
Gupta, P. K., Rustgi, S. & Kulwal, P. L. (2005) *Linkage disequilibrium and association studies in higher plants: Present status and future prospects*. *Plant Molecular Biology,* **57,** 461-485.
Ha, C. M., Jun, J. H., Nam, H. G. & Fletcher, J. C. (2007) *BLADE-ON-PETIOLE1 and 2 control Arabidopsis lateral organ fate through regulation of LOB domain and adaxial-abaxial polarity genes*. *Plant Cell,* **19,** 1809-1825.
Haley, C. S. & Knott, S. A. (1992) *A SIMPLE REGRESSION METHOD FOR MAPPING QUANTITATIVE TRAIT LOCI IN LINE CROSSES USING FLANKING MARKERS*. *Heredity,* **69,** 315-324.
Haley, C. S., Knott, S. A. & Elsen, J. M. (1994) *MAPPING QUANTITATIVE TRAIT LOCI IN CROSSES BETWEEN OUTBRED LINES USING LEAST-SQUARES*. *Genetics,* **136,** 1195-1207.
Hall, D., Luquez, V., Garcia, V. M., St Onge, K. R., Jansson, S. & Ingvarsson, P. K. (2007) *Adaptive population differentiation in phenology across a latitudinal gradient in European Aspen (Populus tremula, L.): A comparison of neutral markers, candidate genes and phenotypic traits*. *Evolution,* **61,** 2849-2860.
Hase, Y., Trung, K. H., Matsunaga, T. & Tanaka, A. (2006) *A mutation in UVI4 gene, which promotes the progression of endoreduplication, confers an increased tolerance against UV-B light*. *Plant and Cell Physiology,* **47,** S210-S210.
Haussmann, B. I. G., Parzies, H. K., Presterl, T., Susic, Z. & Miedaner, T. (2004) *Plant genetic resources in crop improvement*. *Plant Genetic Resources Characterization and Utilization,* **2,** 3-21.
Hemerly, A., Engler, J. D., Bergounioux, C., Vanmontagu, M., Engler, G., Inze, D. & Ferreira, P. (1995) *Dominant-Negative Mutants of the Cdc2 Kinase Uncouple Cell-Division from Iterative Plant Development*. *Embo Journal,* **14,** 3925-3936.
Hetherington, A. M. & Woodward, F. I. (2003) *The role of stomata in sensing and driving environmental change*. *Nature,* **424,** 901-908.

Holroyd, G. H., Hetherington, A. M. & Gray, J. E. (2002) *A role for the cuticular waxes in the environmental control of stomatal development*. *New Phytologist,* **153,** 433-439.

Horiguchi, G., Kim, G. T. & Tsukaya, H. (2005a) *The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana*. *Plant Journal,* **43,** 68-78.

Horiguchi, G., Kim, G. T. & Tsukaya, H. (2005b) *The transcription factor AtGRF5 and the transcription coactivator AN3 regulate cell proliferation in leaf primordia of Arabidopsis thaliana*. *Plant J,* **43,** 68-78.

Howe, G. T., Bucciaglia, P. A., Hackett, W. P., Furnier, G. R., Cordonnier-Pratt, M. M. & Gardner, G. (1998) *Evidence that the phytochrome gene family in black cottonwood has one PHYA locus and two PHYB loci but lacks members of the PHYC/F and PHYE subfamilies*. *Molecular Biology and Evolution,* **15,** 160-175.

Howe, G. T., Hackett, W. P., Furnier, G. R. & Klevorn, R. E. (1995) *Photoperiodic Responses of a Northern and Southern Ecotype of Black Cottonwood*. *Physiologia Plantarum,* **93,** 695-708.

Huang, J., Taylor, J. P., Chen, J. G., Uhrig, J. F., Schnell, D. J., Nakagawa, T., Korth, K. L. & Jones, A. M. (2006) *The plastid protein THYLAKOID FORMATION1 and the plasma membrane G-protein GPA1 interact in a novel sugar-signaling mechanism in Arabidopsis*. *Plant Cell,* **18,** 1226-1238.

Imbert, E. & Lefevre, F. (2003) *Dispersal and gene flow of Populus nigra (Salicaceae) along a dynamic river system*. *Journal of Ecology,* **91,** 447-456.

Ingvarsson, P. K., Garcia, M. V., Hall, D., Luquez, V. & Jansson, S. (2006) *Clinal variation in phyB2, a candidate gene for day-length-induced growth cessation and bud set, across a latitudinal gradient in European aspen (Populus tremula)*. *Genetics,* **172,** 1845-1853.

Iwakawa, H., Iwasaki, M., Kojima, S., Ueno, Y., Soma, T., Tanaka, H., Semiarti, E., Machida, Y. & Machida, C. (2007) *Expression of the ASYMMETRIC LEAVES2 gene in the adaxial domain of Arabidopsis leaves represses cell proliferation in this domain and is critical for the development of properly expanded leaves*. *Plant Journal,* **51,** 173-184.

Iwakawa, H., Ueno, Y., Semiarti, E., Onouchi, H., Kojima, S., Tsukaya, H., Hasebe, M., Soma, T., Ikezaki, M., Machida, C. & Machida, Y. (2002) *The ASYMMETRIC LEAVES2 gene of Arabidopsis thaliana, required for formation of a symmetric flat leaf lamina, encodes a member of a novel family of proteins characterized by cysteine repeats and a leucine zipper*. *Plant and Cell Physiology,* **43,** 467-478.

Jansen, R. C. & Nap, J. P. (2001) *Genetical genomics: the added value from segregation*. *Trends in Genetics,* **17,** 388-391.

Jansson, S. & Douglas, C. J. (2007) *Populus: A model system for plant biology*. *Annual Review of Plant Biology,* **58,** 435-458.

Kearsey, M. J. (1998) *The principles of QTL analysis (a minimal mathematics approach)*. *Journal of Experimental Botany,* **49,** 1619-1623.

Kerstetter, R., Vollbrecht, E., Lowe, B., Veit, B., Yamaguchi, J. & Hake, S. (1994) *SEQUENCE-ANALYSIS AND EXPRESSION PATTERNS DIVIDE THE MAIZE KNOTTED1-LIKE HOMEOBOX GENES INTO 2 CLASSES*. *Plant Cell,* **6,** 1877-1887.

Kerstetter, R. A. & Poethig, R. S. (1998) *The specification of leaf identity during shoot development*. *Annual Review of Cell and Developmental Biology,* **14,** 373-398.

Kieffer, M., Stern, Y., Cook, H., Clerici, E., Maulbetsch, C., Laux, T. & Davies, B. (2006) *Analysis of the transcription factor WUSCHEL and its functional homologue in Antirrhinum reveals a potential mechanism for their roles in meristem maintenance*. *Plant Cell,* **18,** 560-573.

Kim, G. T., Ito, T., Seki, M., Shinozaki, K. & Tsukaya, H. (2002a) *Genetic analysis of the factors that regulate leaf differentiation*. *Plant and Cell Physiology,* **43,** S179-S179.

Kim, G. T., Shoda, K., Tsuge, T., Cho, K. H., Uchimiya, H., Yokoyama, R., Nishitani, K. & Tsukaya, H. (2002b) *The ANGUSTIFOLIA gene of Arabidopsis, a plant CtBP gene, regulates leaf-cell expansion, the arrangement of cortical microtubules in leaf cells and expression of a gene involved in cell-wall formation*. *Embo Journal,* **21,** 1267-1279.

Kim, M., Pham, T., Hamidi, A., McCormick, S., Kuzoff, R. K. & Sinha, N. (2003) *Reduced leaf complexity in tomato wiry mutants suggests a role for PHAN and KNOX genes in generating compound leaves*. *Development,* **130,** 4405-4415.

Lake, J. A., Woodward, F. I. & Quick, W. P. (2002) *Long-distance CO2 signalling in plants*. *Journal of Experimental Botany,* **53,** 183-193.

Lander, E., Abrahamson, J., Barlow, A., Daly, M., Lincoln, S., Newburg, L. & Green, P. (1987) *MAPMAKER A COMPUTER PACKAGE FOR CONSTRUCTING GENETIC-LINKAGE MAPS*. *Cytogenetics and Cell Genetics,* **46,** 642-642.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., et al. (2001) *Initial sequencing and analysis of the human genome*. *Nature,* **409,** 860-921.

Larkin, J. C., Brown, M. L. & Schiefelbein, J. (2003) *How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis*. *Annual Review of Plant Biology,* **54,** 403-430.

Laureysens, I., Pellis, A., Willems, J. & Ceulemans, R. (2005) *Growth and production of a short rotation coppice culture of poplar. III. Second rotation results*. *Biomass & Bioenergy,* **29,** 10-21.

Lecoeur, J., Wery, J., Turc, O. & Tardieu, F. (1995) *EXPANSION OF PEA LEAVES SUBJECTED TO SHORT WATER-DEFICIT - CELL NUMBER AND CELL-SIZE ARE SENSITIVE TO STRESS AT DIFFERENT PERIODS OF LEAF DEVELOPMENT*. *Journal of Experimental Botany,* **46,** 1093-1101.

Lee, C. E. & Frost, B. W. (2002) *Morphological stasis in the Eurytemora affinis species complex (Copepoda : Temoridae)*. *Hydrobiologia,* **480,** 111-128.

Li, B., Suzuki, J. I. & Hara, T. (1998) *Latitudinal variation in plant size and relative growth rate in Arabidopsis thaliana*. *Oecologia,* **115,** 293-301.

Li, Y., Jones, L. & McQueen-Mason, S. (2003) *Expansins and cell growth*. *Current Opinion in Plant Biology,* **6,** 603-610.

Lin, W. C., Shuai, B. & Springer, P. S. (2003) *The Arabidopsis LATERAL ORGAN BOUNDARIES-domain gene ASYMMETRIC LEAVES2 functions in the repression of KNOX gene expression and in adaxial-abaxial patterning*. *Plant Cell,* **15,** 2241-2252.

Ma, Q. H. & Li, Y. L. (2006) *Inhibition of ethylene biosynthesis enhances vegetative bud formation without affecting growth and development of transgenic tobacco plants*. *Journal of Plant Growth Regulation,* **25,** 101-109.

Mackay, T. F. C. (2001) *The genetic architecture of quantitative traits*. *Annual Review of Genetics,* **35,** 303-339.

Marron, N. & Ceulemans, R. (2006) *Genetic variation of leaf traits related to productivity in a Populus deltoides x Populus nigra family*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **36,** 390-400.

McCamant, T. & Black, R. A. (2000) *Cold hardiness in coastal, montane, and inland populations of Populus trichocarpa*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **30,** 91-99.

McConnell, J. R., Emery, J., Eshed, Y., Bao, N., Bowman, J. & Barton, M. K. (2001) *Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots*. *Nature,* **411,** 709-713.

Mizukami, Y. & Fischer, R. L. (2000) *Plant organ size control: AINTEGUMENTA regulates growth and cell numbers during organogenesis*. *Proceedings of the National Academy of Sciences of the United States of America,* **97,** 942-947.

Monclus, R., Dreyer, E., Delmotte, F. M., Villar, M., Delay, D., Boudouresque, E., Petit, J. M., Marron, N., Brechet, C. & Brignolas, F. (2005) *Productivity, leaf traits and carbon isotope discrimination in 29 Populus deltoides x P-nigra clones*. *New Phytologist,* **167,** 53-62.

Moreau, C., Aksenov, N., Lorenzo, M. G., Segerman, B., Funk, C., Nilsson, P., Jansson, S. & Tuominen, H. (2005) *A genomic approach to investigate developmental cell death in woody tissues of Populus trees*. *Genome Biology,* **6**.

Morgante, M. & Salamini, F. (2003) *From plant genomics to breeding practice*. *Current Opinion in Biotechnology,* **14,** 214-219.

Murray, T. A. & McManus, M. T. (2005) *Developmental regulation of 1-aminocyclopropane-1-carboxylate synthase gene expression during leaf ontogeny in white clover*. *Physiologia Plantarum,* **124,** 107-120.

Naito, T., Yamashino, T., Kiba, T., Koizumi, N., Kojima, M., Sakakibara, H. & Mizuno, T. (2007) *A link between cytokinin and ASL9 (ASYMMETRIC LEAVES 2 LIKE 9) that belongs to the AS2/LOB (LATERAL ORGAN BOUNDARIES) family genes in Arabidopsis thaliana*. *Biosci Biotechnol Biochem,* **71,** 1269-1278.

Nelissen, H., Clarke, J. H., De Block, M., De Block, S., Vanderhaeghen, R., Zielinski, R. E., Dyer, T., Lusta, S., Inze, D. & Van Lijsebettens, M. (2003) *DRL1, a homolog of the yeast TOT4/KT112 protein, has a function in meristem activity and organ growth in plants*. *Plant Cell,* **15,** 639-654.

Nickerson, D. A., Tobe, V. O. & Taylor, S. L. (1997) *PolyPhred: Automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing*. *Nucleic Acids Research,* **25,** 2745-2751.

Nishitani, K. & Tominaga, R. (1992) *ENDOXYLOGLUCAN TRANSFERASE, A NOVEL CLASS OF GLYCOSYLTRANSFERASE THAT CATALYZES TRANSFER OF A SEGMENT OF XYLOGLUCAN MOLECULE TO ANOTHER XYLOGLUCAN MOLECULE*. *Journal of Biological Chemistry,* **267,** 21058-21064.

Ohno, C. K., Reddy, G. V., Heisler, M. G. B. & Meyerowitz, E. M. (2004) *The Arabidopsis JAGGED gene encodes a zinc finger protein that promotes leaf tissue development*. *Development,* **131,** 1111-1122.

Okada, T., Endo, M., Singh, M. B. & Bhalla, P. L. (2005) *Analysis of the histone H3 gene family in Arabidopsis and identification of the male-gamete-specific variant AtMGH3*. *Plant J,* **44,** 557-568.

Ostergaard, L. & Yanofsky, M. F. (2004) *Establishing gene function by mutagenesis in Arabidopsis thaliana*. *Plant Journal,* **39,** 682-696.

Ouborg, N. J. & Vriezen, W. H. (2007) *An ecologist's guide to ecogenomics*. *Journal of Ecology,* **95,** 8-16.

Palaisa, K. A., Morgante, M., Williams, M. & Rafalski, A. (2003) *Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci*. *Plant Cell,* **15,** 1795-1806.

Palo, J. U., O'Hara, R. B., Laugen, A. T., Laurila, A., Primmer, C. R. & Merila, J. (2003) *Latitudinal divergence of common frog (Rana temporaria) life history traits by natural selection: evidence from a comparison of molecular and quantitative genetic data*. *Molecular Ecology,* **12,** 1963-1978.

Park, S. O., Zheng, Z., Oppenheimer, D. G. & Hauser, B. A. (2005) *The PRETTY FEW SEEDS2 gene encodes an Arabidopsis homeodomain protein that regulates ovule development*. *Development,* **132,** 841-849.

Pearson, W. R. & Lipman, D. J. (1988) *Improved tools for biological sequence comparison*. *Proc Natl Acad Sci U S A,* **85,** 2444-2448.

Pellis, A., Laureysens, I. & Ceulemans, R. (2004) *Genetic variation of the bud and leaf phenology of seventeen poplar clones in a short rotation coppice culture*. *Plant Biology,* **6,** 38-46.

Pfaffl, M. W. (2001) *A new mathematical model for relative quantification in real-time RT-PCR*. *Nucleic Acids Research,* **29**.

Poole, I., Weyers, J. D. B., Lawson, T. & Raven, J. A. (1996) *Variations in stomatal density and index: Implications for palaeoclimatic reconstructions*. *Plant Cell and Environment,* **19,** 705-712.

Pritchard, J. K. & Rosenberg, N. A. (1999) *Use of unlinked genetic markers to detect population stratification in association studies*. *Am J Hum Genet,* **65,** 220-228.

Pritchard, J. K., Stephens, M., Rosenberg, N. A. & Donnelly, P. (2000) *Association mapping in structured populations*. *American Journal of Human Genetics,* **67,** 170-181.

Qin, Y. M., Hu, C. Y., Pang, Y., Kastaniotis, A. J., Hiltunen, J. K. & Zhu, Y. X. (2007) *Saturated very-long-chain fatty acids promote cotton fiber and Arabidopsis cell elongation by activating ethylene biosynthesis*. *Plant Cell,* **19,** 3692-3704.

Rae, A. M., Ferris, R., Tallis, M. J. & Taylor, G. (2006) *Elucidating genomic regions determining enhanced leaf growth and delayed senescence in elevated CO2*. *Plant Cell and Environment,* **29,** 1730-1741.

Rae, A. M., Robinson, K. M., Street, N. R. & Taylor, G. (2004) *Morphological and physiological traits influencing biomass productivity in short-rotation coppice poplar*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **34,** 1488-1498.

Rae, A. M., Tricker, P. J., Bunn, S. M. & Taylor, G. (2007) *Adaptation of tree growth to elevated CO2: quantitative trait loci for biomass in Populus*. *New Phytologist,* **175,** 59-69.

Rafalski, A. & Morgante, M. (2004) *Corn and humans: recombination and linkage disequilibrium in two genomes of similar size*. *Trends in Genetics,* **20,** 103-111.

Rafalski, J. A. (2002) *Novel genetic mapping tools in plants: SNPs and LD-based approaches*. *Plant Science,* **162,** 329-333.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L. & Moorman, A. F. M. (2003) *Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data*. *Neuroscience Letters,* **339,** 62-66.

Reich, P. B. & Oleksyn, J. (2004) *Global patterns of plant leaf N and P in relation to temperature and latitude*. *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 11001-11006.

Riemenschneider, D. E. & McMahon, B. G. (1993) *GENETIC-VARIATION AMONG LAKE STATES BALSAM POPLAR POPULATIONS IS ASSOCIATED WITH GEOGRAPHIC ORIGIN*. *Forest Science,* **39,** 130-136.

Rinaldi, A. (2007) *Naturally better - Science and technology are looking to nature's successful designs for inspiration*. *Embo Reports,* **8,** 995-U993.

Robinson, K. M., Karp, A. & Taylor, G. (2004) *Defining leaf traits linked to yield in short-rotation coppice Salix*. *Biomass & Bioenergy,* **26,** 417-431.

Roelofs, D., Aarts, M. G. M., Schat, H. & van Straalen, N. M. (2008) *Functional ecological genomics to demonstrate general and specific responses to abiotic stress*. *Functional Ecology,* **22,** 8-18.

Rose, J. K. C., Braam, J., Fry, S. C. & Nishitani, K. (2002) *The XTH family of enzymes involved in xyloglucan endotransglucosylation and endohydrolysis: Current perspectives and a new unifying nomenclature*. *Plant and Cell Physiology,* **43,** 1421-1435.

Rossi, V. & Varotto, S. (2002) *Insights into the G1/S transition in plants*. *Planta,* **215,** 345-356.

Rottmann, W. H., Meilan, R., Sheppard, L. A., Brunner, A. M., Skinner, J. S., Ma, C. P., Cheng, S. P., Jouanin, L., Pilate, G. & Strauss, S. H. (2000) *Diverse effects of overexpression of LEAFY and PTLF, a poplar (Populus) homolog of LEAFY/FLORICAULA, in transgenic poplar and Arabidopsis*. *Plant Journal,* **22,** 235-245.

Rozen, S. & Skaletsky, H. (2000) *Primer3 on the WWW for general users and for biologist programmers*. *Methods Mol Biol,* **132,** 365-386.

Rutledge, R. G. & Cote, C. (2003) *Mathematics of quantitative kinetic PCR and the application of standard curves*. *Nucleic Acids Research,* **31**.

Sabelli, P. A., Dante, R. A., Leiva-Neto, J. T., Jung, R., Gordon-Kamm, W. J. & Larkins, B. A. (2005) *RBR3, a member of the retinoblastoma-related family from maize, is regulated by the RBR1/E2F pathway*. *Proceedings of the National Academy of Sciences of the United States of America,* **102,** 13005-13012.

Scarascia-Mugnozza, G. E., Hinckley, T. M., Stettler, R. F., Heilman, P. E. & Isebrands, J. G. (1999) *Production physiology and morphology of Populus species and their hybrids grown under short rotation. III. Seasonal carbon allocation patterns from branches*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **29,** 1419-1432.

Scarpella, E., Francis, P. & Berleth, T. (2004) *Stage-specific markers define early steps of procambium development in Arabidopsis leaves and correlate termination of vein formation with mesophyll differentiation*. *Development,* **131,** 3445-3455.

Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Quantitative Monitoring of Gene-Expression Patterns with a Complementary-DNA Microarray*. *Science,* **270,** 467-470.

Schluter, U., Muschak, M., Berger, D. & Altmann, T. (2003) *Photosynthetic performance of an Arabidopsis mutant with elevated stomatal density (sdd1-1) under different light regimes*. *Journal of Experimental Botany,* **54,** 867-874.

Schneeberger, R., Tsiantis, M., Freeling, M. & Langdale, J. A. (1998) *The rough sheath2 gene negatively regulates homeobox gene expression during maize leaf development*. *Development,* **125,** 2857-2865.

Scofield, S. & Murray, J. A. (2006) *The evolving concept of the meristem*. *Plant Mol Biol,* **60,** V-VII.

Seaton, G., Haley, C. S., Knott, S. A., Kearsey, M. & Visscher, P. M. (2002) *QTL Express: mapping quantitative trait loci in of simple and complex pedigrees*. *Bioinformatics,* **18,** 339-340.

Shi, C., Thummler, F., Melchinger, A. E., Wenzel, G. & Lubberstedt, T. (2005) *Association between SCMV resistance and macroarray-based expression patterns in maize inbreds*. *Molecular Breeding,* **16,** 173-184.

Shpak, E. D., Berthiaume, C. T., Hill, E. J. & Torii, K. U. (2004) *Synergistic interaction of three ERECTA-family receptor-like kinases controls Arabidopsis organ growth and flower development by promoting cell proliferation*. *Development,* **131,** 1491-1501.

Shpak, E. D., Lakeman, M. B. & Torii, K. U. (2003) *Dominant-negative receptor uncovers redundancy in the Arabidopsis ERECTA Leucine-rich repeat receptor-like kinase signaling pathway that regulates organ shape*. *Plant Cell,* **15,** 1095-1110.

Shpak, E. D., McAbee, J. M., Pillitteri, L. J. & Torii, K. U. (2005) *Stomatal patterning and differentiation by synergistic interactions of receptor kinases*. *Science,* **309,** 290-293.

Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N. & Bowman, J. L. (1999a) *Members of the YABBY gene family specify abaxial cell fate in Arabidopsis*. *Development,* **126,** 4117-4128.

Siegfried, K. R., Eshed, Y., Baum, S. F., Otsuga, D., Drews, G. N. & Bowman, J. L. (1999b) *Members of the YABBY gene family specify abaxial cell fate in Arabidopsis*. *Development,* **126,** 4117-4128.

Sinha, N. (1999) *Leaf development in angiosperms*. *Annual Review of Plant Physiology and Plant Molecular Biology,* **50,** 419-446.

Skroppa, T., Hylen, G. & Dietrichson, J. (1999) *Relationships between wood density components and juvenile height growth and growth rhythm traits for Norway spruce provenances and families*. *Silvae Genetica,* **48,** 235-239.

Slate, J. (2005) *Quantitative trait locus mapping in natural populations: progress, caveats and future directions*. *Molecular Ecology,* **14,** 363-379.

Smith, R. L. & Sytsma, K. J. (1990) *EVOLUTION OF POPULUS-NIGRA (SECT AIGEIROS) - INTROGRESSIVE HYBRIDIZATION AND THE CHLOROPLAST CONTRIBUTION OF POPULUS ALBA (SECT POPULUS)*. *American Journal of Botany,* **77,** 1176-1187.

Smith, W. K. & McClean, T. M. (1989) *ADAPTIVE RELATIONSHIP BETWEEN LEAF WATER REPELLENCY, STOMATAL DISTRIBUTION, AND GAS-EXCHANGE*. *American Journal of Botany,* **76,** 465-469.

Smulders, M. J. M., Cottrell, J. E., Lefevre, F., van der Schoot, J., Arens, P., Vosman, B., Tabbener, H. E., Grassi, F., Fossati, T., Castiglione, S., Krystufek, V., Fluch, S., Burg, K., Vornam, B., Pohl, A., Gebhardt, K., Alba, N., Agundez, D., Maestro, C., Notivol, E., Volosyanchuk, R., Pospiskova, M., Bordacs, S., Bovenschen, J., van Dam, B. C., Koelewijn, H. P., Halfmaerten, D., Ivens, B., van Slycken, J., Broeck, A. V., Storme, V. & Boerjan, W. (2008) *Structure of the genetic diversity in black poplar (Populus nigra L.) populations across European river systems: Consequences for conservation and restoration*. *Forest Ecology and Management,* **255,** 1388-1399.

Song, S. K. & Clark, S. E. (2005a) *POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in Arabidopsis*. *Dev Biol,* **285,** 272-284.

Song, S. K. & Clark, S. E. (2005b) *POL and related phosphatases are dosage-sensitive regulators of meristem and organ development in Arabidopsis*. *Developmental Biology,* **285,** 272-284.

Song, S. K., Lee, M. M. & Clark, S. E. (2006) *POL and PLL1 phosphatases are CLAVATA1 signaling intermediates required for Arabidopsis shoot and floral stem cells*. *Development,* **133,** 4691-4698.

Sozzani, R., Maggio, C., Varotto, S., Canova, S., Bergounioux, C., Albani, D. & Cella, R. (2006) *Interplay between Arabidopsis activating factors E2Fb and E2Fa in cell cycle progression and development*. *Plant Physiology,* **140,** 1355-1366.

Spitze, K. (1993) *Population structure in Daphnia obtusa: quantitative genetic and allozymic variation*. *Genetics,* **135,** 367-374.

Springer, P. S., Holding, D. R., Groover, A., Yordan, C. & Martienssen, R. A. (2000) *The essential Mcm7 protein PROLIFERA is localized to the nucleus of dividing cells during the G(1) phase and is required maternally for early Arabidopsis development*. *Development,* **127,** 1815-1822.

Sterck, L., Rombauts, S., Vandepoele, K., Rouze, P. & Van de Peer, Y. (2007) *How many genes are there in plants (. . . and why are they there)? Current Opinion in Plant Biology,* **10,** 199-203.

Sterky, F., Bhalerao, R. R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A. M., Charbonnel-Campaa, L., Lindvall, J. J., Tandre, K., Strauss, S. H., Sundberg, B., Gustafsson, P., Uhlen, M., Bhalerao, R. P., Nilsson, O., Sandberg, G., Karlsson, J., Lundeberg, J. & Jansson, S. (2004) *A Populus EST resource for plant functional genomics*. *Proceedings of the National Academy of Sciences of the United States of America,* **101,** 13951-13956.

Stevaux, O. & Dyson, N. J. (2002) *A revised picture of the E2F transcriptional network and RB function*. *Current Opinion in Cell Biology,* **14,** 684-691.

Storme, V., Vanden Broeck, A., Ivens, B., Halfmaerten, D., Van Slycken, J., Castiglione, S., Grassi, F., Fossati, T., Cottrell, J. E., Tabbener, H. E., Lefevre, F., Saintagne, C., Fluch, S., Krystufek, V., Burg, K., Bordacs, S., Borovics, A., Gebhardt, K., Vornam, B., Pohl, A., Alba, N., Agundez, D., Maestro, C., Notivol, E., Bovenschen, J., van Dam, B. C., van der Schoot, J., Vosman, B., Boerjan, W. & Smulders, M. J. (2004a) *Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development*. *Theor Appl Genet,* **108,** 969-981.

Storme, V., Vanden Broeck, A., Ivens, B., Halfmaerten, D., Van Slycken, J., Castiglione, S., Grassi, F., Fossati, T., Cottrell, J. E., Tabbener, H. E., Lefevre, F., Saintagne, C., Fluch, S., Krystufek, V., Burg, K., Bordacs, S., Borovics, A., Gebhardt, K., Vornam, B., Pohl, A., Alba, N., Agundez, D., Maestro, C., Notivol, E., Bovenschen, J., van Dam, B. C., van der Schoot, J., Vosman, B., Boerjan, W. & Smulders, M. J. M. (2004b) *Ex-situ conservation of Black poplar in Europe: genetic diversity in nine gene bank collections and their value for nature development*. *Theoretical and Applied Genetics,* **108,** 969-981.

Street, N. R., Skogstrom, O., Sjodin, A., Tucker, J., Rodriguez-Acosta, M., Nilsson, P., Jansson, S. & Taylor, G. (2006) *The genetics and genomics of the drought response in Populus*. *Plant Journal,* **48,** 321-341.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T. Z., Garcia-Hernandez, M., Foerster, H., Li, D., Meyer, T., Muller, R., Ploetz, L., Radenbaugh, A., Singh, S., Swing, V., Tissier, C., Zhang, P. & Huala, E. (2008) *The Arabidopsis Information Resource (TAIR): gene structure and function annotation*. *Nucleic Acids Research,* **36,** D1009-D1014.

Szalma, S. J., Buckler, E. S., Snook, M. E. & McMullen, M. D. (2005) *Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks*. *Theoretical and Applied Genetics,* **110,** 1324-1333.

Takahashi, T., Gasch, A., Nishizawa, N. & Chua, N. H. (1995) *The Diminuto Gene of Arabidopsis Is Involved in Regulating Cell Elongation*. *Genes & Development,* **9,** 97-107.

Tardieu, F., Granier, C. & Muller, B. (1999) *Modelling leaf expansion in a fluctuating environment: are changes in specific leaf area a consequence of changes in expansion rate? New Phytologist,* **143,** 33-44.

Taylor, G. (2002) *Populus: Arabidopsis for forestry. Do we need a model tree? Annals of Botany,* **90,** 681-689.

Taylor, G., Street, N. R., Tricker, P. J., Sjodin, A., Graham, L., Skogstrom, O., Calfapietra, C., Scarascia-Mugnozza, G. & Jansson, S. (2005) *The transcriptome of Populus in elevated CO2*. *New Phytologist,* **167,** 143-154.

Thornsberry, J. M., Goodman, M. M., Doebley, J., Kresovich, S., Nielsen, D. & Buckler, E. S. (2001) *Dwarf8 polymorphisms associate with variation in flowering time*. *Nature Genetics,* **28,** 286-289.

Trapani, N., Hall, A. J. & Weber, M. (1999) *Effects of constant and variable nitrogen supply on sunflower (Helianthus annus L.) leaf cell number and size*. *Annals of Botany,* **84,** 599-606.

Tschaplinski, T. J., Tuskan, G. A., Sewell, M. M., Gebre, G. M., Donald, E. T. I. & Pendleyi, C. (2006) *Phenotypic variation and quantitative trait locus identification for osmotic potential in an interspecific hybrid inbred F-2 poplar pedigree grown in contrasting environments*. *Tree Physiology,* **26,** 595-604.

Tsuge, T., Tsukaya, H. & Uchimiya, H. (1996) *Two independent and polarized processes of cell elongation regulate leaf blade expansion in Arabidopsis thaliana (L) Heynh*. *Development,* **122,** 1589-1600.

Tsukaya, H. (2006) *Mechanism of leaf-shape determination*. *Annual Review of Plant Biology,* **57,** 477-496.

Tuck, G., Glendining, M. J., Smith, P., House, J. I. & Wattenbach, M. (2006) *The potential distribution of bioenergy crops in Europe under present and future climate*. *Biomass & Bioenergy,* **30,** 183-197.

Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E., Unneberg, P., et al. (2006b) *The genome of black cottonwood, Populus trichocarpa (Torr. & Gray)*. *Science,* **313,** 1596-1604.

Tuskan, G. A., DiFazio, S. P. & Teichmann, T. (2004) *Poplar genomics is getting popular: The impact of the poplar genome project on tree research*. *Plant Biology,* **6,** 2-4.

Van Sandt, V. S. T., Suslov, D., Verbelen, J. P. & Vissenberg, K. (2007) *Xyloglucan endotransglucosylase activity loosens a plant cell wall*. *Annals of Botany,* **100,** 1467-1473.

Vandepoele, K., Raes, J., De Veylder, L., Rouze, P., Rombauts, S. & Inze, D. (2002) *Genome-wide analysis of core cell cycle genes in Arabidopsis*. *Plant Cell,* **14,** 903-916.

Verwijst, T. & Telenius, B. (1999) *Biomass estimation procedures in short rotation forestry*. *Forest Ecology and Management,* **121,** 137-146.

Villar, M., Lefevre, F., Augustin, S., Bonduelle, P., Delplanque, A., Duval, H., Faivre-Rampant, P., Goue, M.-C., Legionnet, A., Menard, M., Nesme, X., Pilate, G., Pinon, J., Valadon, A. & Teissier Du Cros, E. (1995) *Perspectives in poplar genetics and breeding in France*. *Comptes Rendus de l'Academie d'Agriculture de France,* **81,** 137-152.

Vissenberg, K., Fry, S. C., Pauly, M., Hofte, H. & Verbelen, J. P. (2005) *XTH acts at the microfibril-matrix interface during cell elongation*. *Journal of Experimental Botany,* **56,** 673-683.

Vlieghe, K., Boudolf, V., Beemster, G. T. S., Maes, S., Magyar, Z., Atanassova, A., Engler, J. D., De Groodt, R., Inze, D. & De Veylder, L. (2005) *The DP-E2F-like gene DEL1 controls the endocycle in Arabidopsis thaliana*. *Current Biology,* **15,** 59-63.

Voorrips, R. E. (2002) *MapChart: software for the graphical presentation of linkage maps and QTLs*. *J Hered,* **93,** 77-78.

Waites, R., Selvadurai, H. R. N., Oliver, I. R. & Hudson, A. (1998) *The PHANTASTICA gene encodes a MYB transcription factor involved in growth and dorsoventrality of lateral organs in Antirrhinum*. *Cell,* **93,** 779-789.

Walle, I. V., Van Camp, N., Van de Casteele, L., Verheyen, K. & Lemeur, R. (2007) *Short-rotation forestry of birch, maple, poplar and willow in Flanders (Belgium) I - Biomass production after 4 years of tree growth*. *Biomass & Bioenergy,* **31,** 267-275.

Wang, D. & Nettleton, D. (2006) *Identifying genes associated with a quantitative trait or quantitative trait locus via selective transcriptional profiling*. *Biometrics,* **62,** 504-514.

Wang, G. X., Chen, Y., Zhao, J. R., Li, L., Korban, S. S., Wang, F. G., Li, J. S., Dai, J. R. & Xu, M. L. (2007) *Mapping of defense response gene homologs and their association with resistance loci in maize*. *Journal of Integrative Plant Biology,* **49,** 1580-1598.

Wang, N. N., Shih, M. C. & Li, N. (2005) *The GUS reporter-aided analysis of the promoter activities of Arabidopsis ACC synthase genes AtACS4, AtACS5, and AtACS7 induced by hormones and stresses*. *J Exp Bot,* **56,** 909-920.

Weir, B. S. & Cockerham, C. C. (1984) *Estimating F-Statistics for the Analysis of Population-Structure*. *Evolution,* **38,** 1358-1370.

Weir, B. S. & Hill, W. G. (2002) *Estimating F-statistics*. *Annual Review of Genetics,* **36,** 721-750.

Wenkel, S., Emery, J., Hou, B. H., Evans, M. M. S. & Barton, M. K. (2007) *A feedback regulatory module formed by LITTLE ZIPPER and HD-ZIPIII genes*. *Plant Cell,* **19,** 3379-3390.

White, D. W. (2006) *PEAPOD regulates lamina size and curvature in Arabidopsis*. *Proc Natl Acad Sci U S A,* **103,** 13238-13243.

Wilson, L. M., Whitt, S. R., Ibanez, A. M., Rocheford, T. R., Goodman, M. M. & Buckler, E. S. (2004) *Dissection of maize kernel composition and starch production by candidate gene association*. *Plant Cell,* **16,** 2719-2733.

Woodward, F. I., Lake, J. A. & Quick, W. P. (2002) *Stomatal development and CO2: ecological consequences*. *New Phytologist,* **153,** 477-484.

Wright, S. (1949) *Population structure in evolution*. *Proc Am Philos Soc,* **93,** 471-478.

Wullschleger, S., Yin, T. M., DiFazio, S. P., Tschaplinski, T. J., Gunter, L. E., Davis, M. F. & Tuskan, G. A. (2005) *Phenotypic variation in growth and biomass distribution for two advanced-generation pedigrees of hybrid poplar*. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere,* **35,** 1779-1789.

Wullschleger, S. D., Tuskan, G. A. & Difazio, S. P. (2002) *Genomics and the tree physiologist*. *Tree Physiology,* **22,** 1273-1276.

Wyrzykowska, J., Pien, S., Shen, W. & Fleming, A. (2002) *Manipulation of leaf shape by modulation of cell division*. *Development,* **129,** 957 - 964.

Xu, Z. Z. & Zhou, G. S. (2006) *Combined effects of water stress and high temperature on photosynthesis, nitrogen metabolism and lipid peroxidation of a perennial grass Leymus chinensis*. *Planta,* **224,** 1080-1090.

Yagil, Y. & Yagil, C. (2006) *Integration - A key to success in the genetic dissection of complex diseases? Trends in Cardiovascular Medicine,* **16,** 35-38.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y. & Sasaki, T. (2000) *Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS*. *Plant Cell,* **12,** 2473-2484.

Yin, C. Y., Duan, B. L., Wang, X. & Li, C. Y. (2004a) *Morphological and physiological responses of two contrasting Poplar species to drought stress and exogenous abscisic acid application*. *Plant Science,* **167,** 1091-1097.

Yin, T. M., DiFazio, S. P., Gunter, L. E., Jawdy, S. S., Boerjan, W. & Tuskan, G. A. (2004b) *Genetic and physical mapping of Melampsora rust resistance genes in Populus and characterization of linkage disequilibrium and flanking genomic sequence*. *New Phytologist,* **164,** 95-105.

Yin, T. M., Zhang, X. Y., Huang, M. R., Wang, M. X., Zhuge, Q., Tu, S. M., Zhu, L. H. & Wu, R. L. (2002) *Molecular linkage maps of the Populus genome*. *Genome,* **45,** 541-555.

Yokoyama, R. & Nishitani, K. (2001) *A comprehensive expression analysis of all members of a gene family encoding cell-wall enzymes allowed us to predict cis-regulatory regions involved in cell-wall construction in specific organs of arabidopsis*. *Plant and Cell Physiology,* **42,** 1025-1033.

Yu, J. M., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., McMullen, M. D., Gaut, B. S., Nielsen, D. M., Holland, J. B., Kresovich, S. & Buckler, E. S. (2006) *A unified mixed-model method for association mapping that accounts for multiple levels of relatedness*. *Nature Genetics,* **38,** 203-208.

Zgurski, J. M., Sharma, R., Bolokoski, D. A. & Schultz, E. A. (2005) *Asymmetric auxin response precedes asymmetric growth and differentiation of asymmetric leaf1 and asymmetric leaf2 Arabidopsis leaves*. *Plant Cell,* **17,** 77-91.