Origin and diversification of winged bean (*Psophocarpus tetragonolobus* (L.) DC.; Fabaceae) a multi-purpose underutilized legume

Shuyi Yang,2,3 Aurélie Grall4 and Mark A Chapman2,5,6

2 Biological Sciences, University of Southampton, Life Sciences Building 85, Highfield Campus, Southampton, SO17 1BJ, UK
3 College of Forestry and Landscape Architecture, South China Agricultural University, Guangdong 510642, China
4 Africa & Madagascar Team, Identification & Naming Department, Royal Botanic Gardens, Kew TW9 3AE, UK
5 Centre for Underutilised Crops, University of Southampton, Highfield Campus, Southampton, SO17 1BJ, UK
6 Author for correspondence: Tel: +44 (0)2380 594396, Email: m.chapman@soton.ac.uk

ABSTRACT

Premise of the study - For many crops, research into the origin and partitioning of genetic variation is limited and this can slow or prevent crop improvement programs. Many of these underutilized crops have traits which could be of benefit in a changing climate due to stress tolerance or nutritional properties. Winged bean (*Psophocarpus tetragonolobus* (L.) DC.) is one such crop. All parts of the plant can be eaten, from the roots to the seeds, and is high in protein as well as other micronutrients. The goal of our study was to identify the wild progenitor and analyze the partitioning of genetic variation in the crop.

Methods - We used molecular phylogenetic analyses (cpDNA and nuclear ITS sequencing) to resolve relationships between all species in the genus, and population genetics (utilizing microsatellites) to identify genetic clusters of winged bean accessions and compare this to geography.

Key results - We find that winged bean is genetically distinct from all other members of the genus. We also provide support for four groups of species in the genus, largely, but not completely, corresponding to the results of previous morphological analyses. Within winged bean, population genetic analysis using 10 polymorphic microsatellite markers suggests four genetic groups; however there is little correspondence between the genetic variation and the geography of the accessions.

Conclusions - The true wild progenitor of winged bean remains unknown (or is extinct). There has likely been large-scale cross-breeding, trade and transport of winged bean and/or multiple origins of the crop.

1 Manuscript received _______; revision accepted _______.


INTRODUCTION

With climate change and an increasing world’s population, research is needed into nutritious and diverse crops with the ability to grow under stressful or sub-optimal conditions (Schmidhuber and Tubiello, 2007; Campbell et al., 2016). Half of the world’s calorie intake comes from just three crops (wheat, rice and corn), yet these cereals are relatively poor in protein and some micronutrients. Legumes complement cereals in the diet well (Foyer et al., 2016); they supply considerable protein, essential amino acids and micronutrients, and serve as a substitute for meat when this is unavailable or unaffordable (Tharanathan and Mahadevamma, 2003). Several legumes are grown for forage, with the high protein content being of benefit to the livestock.

In addition to the nutrient content, legumes have evolved the ability to fix atmospheric nitrogen meaning that fertilizer input can be lower, legumes can often tolerate poor soils, and they can improve the soil for subsequent crops (Peoples et al., 1995). This poses advantages for resource-poor farmers who cannot obtain or afford fertilizers (Siddique et al., 2012). However legumes remain understudied relative to cereal crops (Foyer et al., 2016).

Several legumes are described as underutilized (or orphan) crops; i.e., crops which are locally important but essentially absent from international trade and having received little research attention (Cullis and Kunert, 2017). These underutilized legumes are typically grown on small scale for feeding families, especially in times of harsh environmental conditions, therefore are vitally important despite their lack of international recognition. Although these crops have received comparatively little research and funding, their importance has been recognized for decades (e.g., NAS, 1975b). In the past decade several genome and transcriptome sequences have been published for underutilized legumes, for example the pigeon pea *Cajanus cajan* (L.) Huth genome (Varshney et al 2012) and lablab *Lablab purpureus* (L.) Sweet, winged bean *Psophocarpus tetragonolobus* (L.) DC., Bambara groundnut *Vigna subterranea* (L.) Verdc. and grasspea *Lathyrus sativus* L. transcriptomes (Chapman, 2015; Vatanparast et al., 2016) with the ability to identify genes and variants associated with adaptive phenotypes (e.g., Varshney et al., 2017). A recent initiative to sequence 100 genomes of 100 orphan crops (http://africanorphancrops.org/) has the potential to increase this amount of genomic information further.

One example of an underutilized legume is the winged bean (*Psophocarpus tetragonolobus* (L.) DC.), which has been recognized by a number of authors and institutes as having much promise for nutritional security in the coming decades. Winged bean is an annual or perennial vine which thrives in hot and humid tropical conditions. It is grown widely, but mainly on a local scale, throughout much of Asia, especially India, Southeast Asia, Indonesia and Papua New Guinea, as well as in some parts of Africa (Verdcourt and Halliday, 1978). Considerable morphological variation exists in winged bean especially in Southeast Asia, Indonesia and Papua New Guinea (Khan, 1976; Harder and Smartt, 1992).

One notable feature of the winged bean is the potential for almost all parts of the plant to be eaten, from the seeds, pods and flowers, to the leaves and tuberous roots (Haq 1982; Khan 1976), with the stems and leaves used as fodder. It also serves well in crop rotation due to the nitrogen fixation capability (Rahman et al., 2014). As with several underutilized crops, however, there are anti-nutrition factors in winged bean (specifically trypsin inhibitors), requiring thorough soaking, rinsing...
and cooking of the dried beans (NAS, 1975a), which may deter more widespread adoption of winged
bean. Protein content of the winged bean seeds is high (30-39%, NAS, 1975a; 38-45%, Prakash et al., 1987)
and comparable to that of soybean. Content of essential amino acids is generally as high as, or
higher than, soybean (Prakash et al., 1987). It is also relatively high in vitamins A and C, calcium and
iron (USDA Food Composition Databases, 2015). In countries where protein deficiency is high, or
access to meat protein is low, winged bean is a candidate for helping to diversify diets and improve
nutrition (NAS, 1975a). Three wild relatives of the winged bean are cultivated, but on much smaller
scales, two in Eastern Africa (P. palustris Desv. and P. grandiflorus R.Wilczek) and one in Southeast
Asia, Brazil and Jamaica (P. scandens (Endl.) Verdc.) (Zeven and de Wet, 1982; Harder and Smartt,
1992). Knowing more about the relationships between the species of Psophocarpus is important to
inform future winged bean breeding and improvement.

The wild progenitor of winged bean has remained somewhat enigmatic given the absence of wild
Psophocarpus in Asia, leading to one suggestion that the true wild progenitor is now extinct
(Verdcourt and Halliday, 1978). Morphological phylogenetic analyses of the nine species in the genus
have come to varied conclusions. In particular, the position of P. grandiflorus, a wild species found
throughout East Africa from the Democratic Republic of the Congo to Ethiopia, is inconsistent. Some
studies suggest this species is closely related to, if not the progenitor of, winged bean (Verdcourt
and Halliday, 1978; Smartt, 1980; Maxted, 1990; Harder and Smartt, 1992), whereas others place P.
grandiflorus more distantly related in the genus (Nur Fatihah et al., 2012). The most recent
morphological analysis places winged bean alongside P. scandens and P. palustris in subgenus
Psophocarpus, with P. scandens the closest wild species to winged bean (Nur Fatihah et al., 2012).

Few attempts to cross winged bean with other members of the genus have been reported, however
one successful cross between winged bean and P. scandens has been made following several
attempts (Nazmul Haq, University of Southampton, unpublished data).

As yet a molecular phylogenetic analysis is lacking. Identifying the true wild progenitor(s) may assist
in the breeding of winged bean and would be necessary to understand the genetic changes
associated with domestication. Relatively few studies have investigated the domestication genetics
of legumes, except for studies of common bean (e.g., Bitocchi et al., 2012; Schmutz et al., 2014), and
therefore little is known about the genes and alleles that were under selection by early farmers. This
may contribute to the observation that genetic enhancement of legumes remains slow relative to
other crops (Foyer et al., 2016). As a first step to remedy this lack of investigation, more work is
needed to identify wild ancestors of legume crops. Breeding with crop wild relatives has the
potential to introduce novel variation for crop improvement (reviewed in Dempewolf et al., 2017).

Crop improvement can also come from selectively breeding different varieties with complementary
qualities, for example morphological characteristics, nutrition, abiotic tolerance and pest resistance.
An understanding of the partitioning of genetic variation in crops is therefore important to begin to
identify and locate genetically distinct varieties and potentially understand the genetic basis of traits
of interest (Hawkes, 1991; Mickelbart et al., 2015).

In this investigation we addressed the following objectives: First, using a molecular phylogenetic
approach we investigate the relationships between members of the genus Psophocarpus, in
particular focusing on identifying the close relative(s) of winged bean. Second, employing a population genetic approach we use microsatellite markers to resolve the partitioning and levels of genetic variation in the crop.

MATERIALS AND METHODS

Seeds from 53 accessions of winged bean (Psophocarpus tetragonolobus) were procured from the USDA (https://www.ars-grin.gov/; 23 accessions), IITA (http://my.iita.org/accession2/; 14 accessions), and NARO (http://www.gene.affrc.go.jp/databases-plant_search_en.php; 16 accessions) (Appendix S1A; see Supplemental Data with this article). Seed from 10 accessions of wild species of Psophocarpus were obtained from the USDA (2 accessions), ILRI (http://192.156.137.110/forage/frgdsearch.asp; 6 accessions) and the Millennium Seed Bank Partnership (http://apps.kew.org/seedlist/; 2 accessions), and commercially for a single accession of Erythrina crista-galli L., the outgroup (based on an rbcL phylogeny; Kajita et al., 2001; Appendix S1B).

Taxonomic status of two of the accessions from ILRI was uncertain. Seed were scarified using a razor blade or sandpaper and placed in a 2:1 mix of Levington’s F2+S: vermiculite. Pots were placed in the greenhouse at the University of Southampton under 16h days (supplemented by fluorescent light) and watered daily. Seed was not available for several species in the genus and so material from herbarium sheets was obtained from Kew (5 accessions) and Paris (3 accessions) (Appendix S1B).

After 3 months a trifoliate leaf from half way up the main stem was removed from all of the winged bean plants and dried flat. Leaves were scanned and the area and perimeter of the terminal leaflet measured using ImageJ (Rasband, 1997-2013). Most plants did not flower and so we are unable to analyze characteristics associated with the flowers or pods. After identifying genetic clusters (see below) we tested for significant differences in leaf morphology (apical leaflet area and perimeter) between the groups (1-way ANOVA with post hoc Tukey test).

DNA extraction

After 2-4 weeks of growth a leaf sample (ca. 100 mg) was removed from each accession for DNA extraction following standard CTAB protocol (Doyle and Doyle, 1990) with minor modification (Chapman and Burke, 2012). For the herbarium samples about 10mg of tissue was extracted. DNA samples were quantified using a Nanodrop (Nanodrop Products, Wilmington, DE, USA) and by running a sample on an agarose gel stained with GelRed (Biotium, Hayward, California, USA).

PCR sequencing and phylogenetic analysis

 Twelve universal primer pairs for amplifying intergenic and non-coding regions of the chloroplast (cp) genome (Appendix S2A) and one pair of primers for the nuclear internal transcribed spacer (ITS) region (White et al., 1990) were tested for amplification on five DNA samples (two P. palustris, two P. scandens and one P. lancifolius Harms). PCR amplification took place using a touchdown program with 55°C final annealing temperature (see Chapman and Burke, 2012 for details). PCR products were checked for amplification on agarose gels stained with GelRed (Biotium, Hayward, California, USA). Successful and clean PCR products were purified using standard protocols and sequenced using BigDye v3 reagents (see Chapman et al., 2008) on an ABI3730xl at the Department of Zoology, University of Oxford.
Three cpDNA regions (the psbC-trnS and psbM2-trnD-GUC intergenic regions and the rpS16 intron; Appendix S2A) and the ITS region were selected for analysis of the complete set of DNAs on the basis of clean amplification and the presence of sequence polymorphisms. For several herbarium DNAs the PCR failed and internal primers were designed to allow amplification of a shorter portion of the locus or to amplify the locus in two portions (Appendix S2A).

Sequences were checked manually using Chromas (Technelysium Pty Ltd, 1998-2001) and aligned using Clustal Omega (Sievers and Higgins, 2014). The cpDNA sequences were analyzed individually as well as concatenated. Maximum likelihood phylogenetic analysis was carried out using PHYML (Guindon et al., 2010) with the model of substitution determined by Akaike Information Criterion and with 1000 bootstrap replicates. Bayesian analysis was carried out using MrBayes 3.2.3 (Ronquist et al., 2011); using the www.phylogeny.fr online server. The model of substitution used was GTR+I. The Markov Chain was run for 10,000 generations, sampling every 10, with the first 250 discarded as burn-in.

**Microsatellite amplification and analysis**

Twenty microsatellite primer pairs were designed from an analysis of three winged bean transcriptome sequences. Reads from two winged bean transcriptomes (Vatanparast et al., 2016) were downloaded from the NCBI short read archive (SRA; accession numbers SRR3039625 and SRR3039626) and assembled together using trinity (Grabherr et al., 2011) and parameters used previously (Chapman, 2015). Microsatellites were identified using misa.pl (http://pgrc.ipk-gatersleben.de/misa/). Reads from a third winged bean transcriptome (Chapman, 2015); SRA accession number SRS826754) were then mapped to the reference to produce bam files which were converted to fasta files using samtools (Li et al., 2009). The sequences were aligned and microsatellite-containing contigs were manually checked for length differences between individuals in the microsatellite region. Twenty contigs with evidence for length variation were used for primer design using primer3 (Rozen and Skaletsky, 2000) (Appendix S2B). A further nine primers from (Wong et al., 2017) were also screened for variation (Appendix S2B).

Initially eight winged bean DNAs were amplified to check for reliable and clean amplification and for the presence of polymorphism. PCR used the ‘3 primer’ method (Schuelke, 2000) and were labelled with the fluorophores 6FAM or TET but were otherwise amplified and resolved on agarose as for the cpDNA loci. Post-PCR, loci were combined such that different sized and labelled loci could be resolved together. Samples were diluted 1:30 and run on an ABI 3730xl at the Department of Zoology, University of Oxford with the GeneScan™ 500 size standard. Primer pairs which produced easy to score polymorphic bands were taken forward for amplification on the full set of DNAs.

Microsatellite alleles were scored using Genemarker V2.4.0 (http://www.softgenetics.com/) and converted into a matrix for further analyses. Cervus (Kalinowski et al., 2007) was used to provide summary statistics for the data. Genetic distances and Principal Coordinates Analysis (PCO) were computed in GenAlEx V6.5 (Peakall and Smouse, 2006) and a minimum evolution (ME) phylogenetic tree was computed with FastME 2.0 (Lefort et al., 2015; http://www.atgc-montpellier.fr/fastme/) using the same distance matrix. STRUCTURE (Falush et al., 2003) was used to infer genetic clusters in the data; the number of clusters (K) was tested from 1 to 8 with five replicate runs (200,000 iterations after 50,000 burn-in iterations) per K. The most likely number of clusters was determined...
using the ΔK method of Evanno et al. (2005) as calculated in STRUCTURE HARVESTER (Earl and von Holdt, 2012). Genetic variation and differentiation statistics were calculated using GenAlEx ver. 6.5 (Peakall and Smouse, 2006).

RESULTS

DNA was extracted from 53 accessions of winged bean for the microsatellite analysis. Four of these plus a further 18 individuals of other Psophocarpus species (one to four individuals per species) and one Erythrina crista-galli individual were used in the phylogenetic analyses.

Phylogenetic relationships in Psophocarpus

Three out of twelve cpDNA loci were amplified on the full set of DNAs using internal primers for the majority of samples (Table 1). The three alignments were less than 300bp each due to the inability to amplify larger fragments from the herbarium samples. Nonetheless they provided sufficient numbers of informative sites (Table 1) to create supported, if not fully-resolved, phylogenetic trees. Phylogenetic trees of the individual loci are shown in Appendix S3 and the ML and Bayesian phylogenies of the concatenated data are shown in Fig. 1A and B. We were unable to generate a clean sequence for psbc-trnS in Erythrina or for rpS16 in the herbarium sample of P. grandiflorus from Ethiopia (gra ETH); hence these are omitted from the relevant individual locus phylogenies. The ITS alignment was 678 nucleotides in length, of which 308 sites were parsimony informative (Table 1; largely because of the high number of fixed differences between winged bean and the other Psophocarpus accessions). Unfortunately we were unable to amplify the ITS region from any samples of P. monophyllus. ML and Bayesian phylogenetic trees are shown in Fig. 1C, D. Genbank numbers are given in Appendix S1.

ML phylogenetic relationships using each cpDNA locus individually provided consistent support for four main groups: 1, winged bean; 2, P. palustris and P. scandens (along with one of the unknown samples); 3, P. monophyllus Harms, P. lecomtei Tisser. and P. obovalis Tisser., and 4, P. lancifolius and P. lukafaensis (De Wild.) R. Wilczek (plus the second unknown sample). For two of three loci, P. grandiflorus was allied to the P. lancifolius - P. lukafaensis group and for two loci winged bean was sister to the P. palustris - P. scandens group (Appendix S3). Concatenating the cpDNA data added support for these groupings under both ML and Bayesian analysis, however the relationships between these groups differs between these analyses (Fig. 1A, B). The main difference is the placement of winged bean, with ML indicating that winged bean’s most close African relative is either P. palustris or P. scandens whereas in the Bayesian analysis winged bean is sister to all other members of the genus.

The relationships between taxa in the ITS phylogeny recapitulated those in the cpDNA phylogenies, again indicating four clades, and supporting the sister group relationship of winged bean to the rest of the genus. In both the cpDNA and ITS analyses, the phylogenetic distance between winged bean and the other taxa is relatively large compared to that between most other species (Fig. 1).

No sequence diversity was revealed for these three cpDNA fragments or the ITS region amongst the four winged bean samples and divergence within each of the other three groups was absent or minimal. The one exception is the presence of two subgroups in the P. palustris - P. scandens group.
(Fig. 1). This splits *P. palustris* into two clades, one of which contains only two *P. palustris* individuals and the other groups two *P. palustris* individuals with *P. scandens* and one individual of unknown taxonomy (sp1102).

**Population genetics of winged bean**

Five of the 20 primers designed from the combined transcriptomes produced reliable, clean and polymorphic bands. Five further primer pairs from Wong et al. (2017) were also selected based on the same criteria (Appendix S2B). These amplified between two and 11 alleles each (average 5.4) across the panel of winged bean with polymorphism information content (PIC) ranging from 0.370 to 0.724 (average 0.546; Table 2) which is as high as, or higher than, several other investigations of legume population structure (Appendix S4). Observed heterozygosity was low (Table 2) which is not surprising for a crop which is largely inbreeding.

PCO of the winged bean microsatellite genotypes revealed little clustering of geographically proximal accessions (Fig. 2). For some countries the number of accessions genotyped was low and so it is difficult to determine any fine-scale structure between many countries, however even if we focus on countries where several accessions (≥ 7) were genotyped (Indonesia, Malaysia, Papua New Guinea) there was still no clustering by country. Whilst the partitioning of genetic diversity is not expected to observe political borders (Meirmans, 2015), it might be expected that if genetic variation was related to geography then some patterns would be observed. This was however not the case; even at a continental scale accessions from Africa, Asia and the Americas are not separated. Geographical co-ordinates (or even location descriptions) are not available for many accessions and so a Mantel test (Mantel, 1967) was not possible.

In STRUCTURE each individual is assigned proportional membership into each of the *K* clusters. Analysis of the STRUCTURE results suggested that the most likely number of clusters (*K*) was four (Fig. 3), however we also present the STRUCTURE results for *K* values of 2 and 3 (Appendix S5) because other levels of clustering might reveal biologically relevant grouping (Meirmans, 2015). If we take 80% membership into a single cluster to represent an individual which is not admixed, then 44 individuals are non-admixed for *K* = 4, with 6, 14, 16, 8 individuals in clusters 1, 2, 3 and 4, respectively. Genetic variation (Table 3), calculated as unbiased expected heterozygosity, was greatest in cluster 3 (0.503), intermediate in clusters 1 and 2 (0.456 and 0.461, respectively) and lowest in cluster 4 (0.307). Genetic differentiation (*F*<sub>ST</sub>; Table 3), was lowest between clusters 2 and 3 (0.137) and greatest between clusters 1 and 4 (0.307).

Each of the four clusters contains individuals from the three main geographic regions (Asia, Africa, the Americas) with the exception of cluster 1 which only contains Asian (two Indonesian, one Papua New Guinean, two Bangladeshi and one Sri Lankan) winged bean individuals (Fig. 3). A χ<sup>2</sup> analysis revealed that the distribution of accessions into the four clusters (plus a fifth category of unclustered) was essentially random in relation to the geographic origin of the accessions (χ<sup>2</sup> = 9.329, d.f. = 8, *P* = 0.315). This indicates that genetic structure in winged bean is not related to geography.

However, there was variation in leaflet size when comparing between plants in the STRUCTURE clusters. Leaflets from plants in cluster 2 were significantly smaller (*F*<sub>3,43</sub> = 3.74, *P* = 0.018, 1-way ANOVA) with a smaller perimeter (*F*<sub>3,43</sub> = 3.02, *P* = 0.040, 1-way ANOVA) than those in group 4, with plants in clusters 1 and 3 intermediate (Fig. 4).
The ME tree (Appendix S6) demonstrates that the accessions from the Americas (blue branches in Appendix S6) and from Africa (black branches) are not monophyletic and are instead interspersed amongst the Asian accessions, in line with the PCO and STRUCTURE results. Because the tree is unrooted it is impossible to confirm the order of the branching events. The accessions grouped into the four clusters by STRUCTURE correspond well, but not perfectly, with the ME tree groups (see shading of the accession names in the ME tree; Appendix S6). For example, five of six accessions assigned to cluster 1 form a monophyletic group in the ME tree and 13 of the 14 accessions assigned to cluster 2 form a monophyletic group (albeit with one cluster 3 and three admixed individuals also found in this group). Admixed individuals (based on STRUCTURE) are found scattered throughout the ME tree.

**DISCUSSION**

*Phylogenetic relationships in Psophocarpus*

The four clades (Fig. 1) identified in the molecular phylogenetic analyses largely group species by geography. Winged bean appears genetically distinct and is the only species found in Asia (with a few reports of *P. scandens* although this is introduced and cultivated). The clade comprising *Psophocarpus grandiflorus*, *P. lancifolius* and *P. lukafaensis* represents the East African species (*P. grandiflorus* and *P. lancifolius* have overlapping distributions in Uganda/Kenya and *P. lancifolius* and *P. lukafaensis* overlap in Zambia/Botswana; Verdcourt and Halliday, 1978). The clade comprising *P. monophyllus*, *P. lecomtei* and *P. obovalis* represents Western (*P. monophyllus*) and Central (Central African Republic, Democratic Republic of Congo; *P. lecomtei* and *P. obovalis*) African species, all with restricted distributions (Verdcourt and Halliday, 1978). The final clade comprises the widespread species *P. palustris* (West to Central Africa) and *P. scandens* (Central and South Eastern Africa), which overlap in their distributions in Benin/Nigeria/Cameroon (Verdcourt and Halliday, 1978).

Although precise identification of the two unknown samples from the Democratic Republic of Congo is not possible yet, they appear to be allied to the *P. lancifolius* group (sp1091) and the *P. palustris* - *P. scandens* group (sp1102). The lack of resolution between some of the species suggests recent divergence and evolution within each group, but relatively more ancient divergence between the groups.

Several aspects of the molecular-based groupings are consistent with previous morphological analyses. For example (Verdcourt and Halliday, 1978; Maxted, 1990; Nur Fatihah et al., 2012) all proposed close relationships between *P. lancifolius* and *P. lukafaensis*, between *P. monophyllus*, *P. lecomtei* and *P. obovalis*, and between *P. palustris* and *P. scandens* (Table 4). The differences between these three studies can be attributed to the placement of *P. grandiflorus* and winged bean, and in fact our study is not completely in line with any of these scenarios.

Verdcourt and Halliday (1978) suggested that *P. grandiflorus* was the progenitor of winged bean, and Maxted (1990) also suggested a close relationship between these taxa, as well as with *P. palustris* and *P. scandens*. The analysis by Nur Fatihah et al. (2012), however, placed *P. grandiflorus* sister to the remainder of the genus (Table 4). Our findings show *P. grandiflorus* is most closely related to, but distinct from, *P. lancifolius* and *P. lukafaensis* (Fig. 1), a different pattern to that suggested by any of the morphological analyses (Table 4).
In addition to the *P. grandiflorus* vs. *P. lancifolius - P. lukafaensis* split, only one other intra-clade divergence is supported; genetic differences were found within *P. palustris*, with the four individuals split into two clades in both the cpDNA and ITS analyses (Fig. 1). One pair of individuals are on a branch of their own, whereas the other pair are genetically identical to two individuals of *P. scandens* and one of the samples of uncertain taxonomic status (sp1102). The first pair of *P. palustris* individuals are from Western Africa (Burkina Faso and The Gambia) where *P. scandens* is absent, whereas the other pair are from Democratic Republic of the Congo and Nigeria where the species distributions overlap, and intermediates have been reported (Verdcourt and Halliday, 1978; Maxted, 1990). This could mean that hybridization has taken place in this region of sympathy causing the introgression of cpDNA and the ITS region from *P. scandens* into *P. palustris*. Alternatively, the morphological similarity (Maxted, 1989) and our finding of a lack of consistent genetic differentiation between *P. palustris* and *P. scandens* could instead mean that the taxonomy of these two species needs revision.

We did not manage to resolve all interspecific relationships, hence we can only comment on the sectional and subgenus classifications. Based on our results however we propose that the morphological analyses carried out previously (Verdcourt and Halliday, 1978; Maxted, 1990; Nur Fatihah et al., 2012) do not accurately define genetic groupings and therefore may need to be updated. We would like to add more markers before making any commitments to a new classification; however we tentatively suggest four sectional delimitations (Table 4). In particular we highlight that *P. tetragonolobus* is genetically distinct from other members of the genus, in contrast to all morphological analyses which place *P. tetragonolobus* in the same section as *P. palustris* and *P. scandens*. The morphological characters which group these taxa (petiole < 5 cm and pods with a lignified exocarp; Nur Fatihah et al., 2012) may therefore represent the ancestral condition in the group, having been lost in the other taxa, or are products of convergent evolution.

*The origin of winged bean*

Whether the closest African relative of winged bean is a member of the *P. palustris - P. scandens* group (as seen in the cpDNA ML analysis and most morphological analyses) or not (our cpDNA Bayesian analysis and the ITS analysis) is unclear, but given the large genetic distance (Fig. 1) it seems that winged bean should probably be considered in a different section. This suggests that identification the direct progenitor of winged bean still eludes us and could imply that Verdcourt and Halliday (1978)'s conclusion, that the progenitor of winged bean is a now-extinct Asian species, is true. Alternatively winged bean as a crop could be morphologically indistinct from a progenitor and so individuals are all identified as the cultivated taxon.

No sequence differences in the four winged bean samples were observed and it is not unusual for a crop to exhibit very low sequence variation relative to wild relatives (Dane and Lang, 2004; Wills and Burke, 2006; Besnard et al., 2011) due to the genetic bottleneck typically associated with a domestication event. In this study, this is not because the four sequenced individuals are especially closely related to each other; instead these are present in different regions of the microsatellite-based PCO.

*Population genetics of winged bean*
Using ten microsatellites we genotyped 53 accessions of winged bean from throughout the distribution in an effort to elucidate genetic structure, should it be present. Our population genetic analyses suggest that genetic structure in winged bean is roughly partitioned into four groups (Fig. 3); however this is unrelated to geography. Instead the four groups typically comprise accessions from throughout the distribution. This seems to be relatively rare in studies of the partitioning of genetic variation within crop species, however other studies of winged bean have come to similar conclusions (Mohanty et al., 2013; Chen et al., 2015). This can be explained by one or more of a diffuse or multiple origins, significant cross-breeding and/or the transport of seed between locations. Nine individuals showed admixture (i.e., no more than 80% ancestry in one of the four clusters) which could indicate recent cross-breeding or an ancestral admixed ‘stock’ of winged bean from which the other populations are derived. Yet there was no geographic pattern to these nine individuals, coming from all three continents sampled and six different countries. Although we were unable to measure pod and flower characteristics on the winged bean accessions we grew in the greenhouse, we found that leaf size differed between some of the groups found in the STRUCTURE analysis (Fig. 4). If we assume that morphological similarity correlates with genetic similarity then this backs up our finding that the global distribution of winged bean germplasm represents the result of large-scale movement of accessions and/or extensive hybridization with morphologically similar accessions found in disparate geographic locations.

The reasons behind the lack of geographic structure in the genetic data are not clear; however we can speculate some reasons for the successful international trade of winged bean which would erode geographic groupings. First, few legume crops are well-adapted to humid tropics therefore winged bean fits a specific niche, encouraging its international trade to countries with specific environments (Khan, 1976). Second, it has been reported that winged bean needs no specific nodulating bacteria (NAS, 1975b) and so this may have aided its expansion into other regions, where other legumes may be slowed or prevented from establishing. A final possibility is that since the renewed interest in the winged bean in the 1970s there has been much seed-sharing between institutes. We found that the passport data for some accessions was either missing or contradictory (these accessions were not used), but this raises the possibility that some of the geographic data we used is incorrect.

CONCLUSIONS

Using a combination of marker types (from slow-evolving cpDNA loci to fast-evolving nuclear microsatellites) we have been able to investigate the relationship of winged bean to other members of the genus *Psophocarpus* as well as search for population structure in the crop. We show that the wild progenitor of winged bean remains elusive, or that cultivated winged bean is insufficiently diverged from a wild form to be identified as separate taxon. Within the crop there is subtle population structure which is largely unrelated to geography suggesting that winged bean might have rapidly been traded worldwide after a single or multiple origins.

ACKNOWLEDGEMENTS
We thank Dr Myriam Gaudeul at MNHN Paris and Jean Hanson at ILRI for assistance with procuring herbarium samples and seed of the wild taxa, respectively, and the managers of the genebanks for helpfully supplying the seed of winged bean. The authors acknowledge the use of the IRIDIS High Performance Computing Facility, and associated support services at the University of Southampton, in the completion of this work. Alberto Stefano Tanzi, Sean Mayes and Festo Massawe generously provided a pre-print of, and data associated with, the Wong et al. (2017) study. The manuscript was improved though comments from Asia Alrashed, Annabelle Damerum, Nazmul Haq, Rachael Graham, Anna Page and Jasmine Saban. We thank the reviewers for valuable comments which improved our manuscript.
Table 1 - Locus information for the three cpDNA regions studied. Primer sequences and their sources are given in Appendix S2.

<table>
<thead>
<tr>
<th>Locus</th>
<th>NI a</th>
<th>Locus length</th>
<th>#PI sites b</th>
<th>Model c</th>
</tr>
</thead>
<tbody>
<tr>
<td>psbC-trnS</td>
<td>22</td>
<td>290</td>
<td>14</td>
<td>HKY85</td>
</tr>
<tr>
<td>psbM-trnD</td>
<td>23</td>
<td>279</td>
<td>19</td>
<td>GTR</td>
</tr>
<tr>
<td>rpS16</td>
<td>22</td>
<td>232</td>
<td>33</td>
<td>GTR</td>
</tr>
<tr>
<td>Combined cp</td>
<td>n/a</td>
<td>801</td>
<td>68</td>
<td>GTR+I</td>
</tr>
<tr>
<td>ITS</td>
<td>20</td>
<td>678</td>
<td>308</td>
<td>TN93+G</td>
</tr>
</tbody>
</table>

Notes:
- a Number of individuals (out of 23) successfully amplified and sequenced; b Number of parsimony informative sites; c Model of sequence evolution determined by PhyML (HKY, (Hasegawa et al., 1985)); GTR, (Lanave et al., 1984); TN93, (Tamura and Nei, 1993)); +I, with the proportion of invariable sites calculated; + G, modelled by a discrete Gamma distribution).
Table 2 - Polymorphism statistics for the ten SSR markers analyzed. Primer sequences are given in Appendix S2.

<table>
<thead>
<tr>
<th>Locus</th>
<th>NA</th>
<th>Ni</th>
<th>Hobs</th>
<th>Hexp</th>
<th>PIC</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAC4</td>
<td>2</td>
<td>53</td>
<td>0.132</td>
<td>0.498</td>
<td>0.372</td>
<td>310</td>
<td>312</td>
</tr>
<tr>
<td>MAC7</td>
<td>11</td>
<td>52</td>
<td>0.404</td>
<td>0.715</td>
<td>0.670</td>
<td>221</td>
<td>240</td>
</tr>
<tr>
<td>MAC14</td>
<td>4</td>
<td>51</td>
<td>0.196</td>
<td>0.516</td>
<td>0.470</td>
<td>321</td>
<td>330</td>
</tr>
<tr>
<td>MAC15</td>
<td>4</td>
<td>46</td>
<td>0.065</td>
<td>0.623</td>
<td>0.538</td>
<td>235</td>
<td>246</td>
</tr>
<tr>
<td>MAC17</td>
<td>6</td>
<td>46</td>
<td>0.087</td>
<td>0.402</td>
<td>0.370</td>
<td>240</td>
<td>267</td>
</tr>
<tr>
<td>Pt7.2</td>
<td>6</td>
<td>49</td>
<td>0.265</td>
<td>0.770</td>
<td>0.724</td>
<td>424</td>
<td>433</td>
</tr>
<tr>
<td>Pt53</td>
<td>6</td>
<td>47</td>
<td>0.170</td>
<td>0.594</td>
<td>0.553</td>
<td>308</td>
<td>324</td>
</tr>
<tr>
<td>Pt68.1</td>
<td>4</td>
<td>49</td>
<td>0.245</td>
<td>0.630</td>
<td>0.567</td>
<td>225</td>
<td>237</td>
</tr>
<tr>
<td>Pt76.1</td>
<td>5</td>
<td>49</td>
<td>0.306</td>
<td>0.687</td>
<td>0.634</td>
<td>200</td>
<td>209</td>
</tr>
<tr>
<td>Pt93.1</td>
<td>6</td>
<td>51</td>
<td>0.255</td>
<td>0.597</td>
<td>0.562</td>
<td>267</td>
<td>279</td>
</tr>
<tr>
<td>Average</td>
<td>5.4</td>
<td>49.3</td>
<td>0.213</td>
<td>0.603</td>
<td>0.546</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

a MAC primers were designed here, Pt primers are from Wong et al. (2017); b Number of alleles detected; c Number of individuals (out of 53) successfully amplified and scored; d Observed heterozygosity; e Expected heterozygosity; f Polymorphic Information Content; g Minimum allele size; h Maximum allele size
Table 3 – Genetic variation within and differentiation between genetic clusters of winged bean identified using STRUCTURE

<table>
<thead>
<tr>
<th>STRUCTURE cluster</th>
<th>uHe</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>0.456</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C2</td>
<td>0.461</td>
<td>0.194</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C3</td>
<td>0.503</td>
<td>0.210</td>
<td>0.137</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>C4</td>
<td>0.307</td>
<td>0.361</td>
<td>0.229</td>
<td>0.230</td>
<td>X</td>
</tr>
</tbody>
</table>

1 uHe, Unbiased heterozygosity; F_{ST}, Fixation index
Table 4 – Summary of the morphological phylogenetic analyses of the genus *Psophocarpus* alongside tentative taxonomic rearrangements drawn from this study.

<table>
<thead>
<tr>
<th>Species</th>
<th>(Verdcourt and Halliday, 1978)</th>
<th>(Maxted, 1990)</th>
<th>(Nur Fatihah et al., 2012)</th>
<th>This study</th>
</tr>
</thead>
</table>
FIGURE LEGENDS

**Figure 1** – Phylogenetic trees for winged bean and other *Psophocarpus* species (rooted with *Erythrina*). Species names are followed by the country of origin (ISO 3166-1 alpha-3 code) (see Appendix S1 for further details). (A, B) ML and Bayesian phylogenetic trees for combined cpDNA regions *psbC-trnS*, *psbM-trnD*, and *rpS16*; (C, D) ML and Bayesian phylogenetic trees for the ITS data.

**Figure 2** – Principal Coordinates Analysis of the microsatellite data for cultivated accessions of winged bean.

**Figure 3** – Results of the STRUCTURE analysis of cultivated winged bean based on ten microsatellites. The most highly supported number of clusters was 4 (inset). Each accession is represented by a horizontal bar and the shading of each bar corresponds to percentage membership to each of the four clusters.

**Figure 4** – Apical leaf area (A) and perimeter (B) for winged bean accessions found in each of the four STRUCTURE clusters. Shared letters above the bars indicate no significant difference (1-way ANOVA).

**Appendix S1** – List of accessions of winged bean accessions used in this study. (A) Those analyzed for microsatellite variation, and (B) those used in the cpDNA and ITS sequencing.

**Appendix S2** – Primer sequences. (A) For cpDNA and ITS sequencing, and (B) for microsatellite analysis.

**Appendix S3** – Phylogenetic trees for the analyses of individual cpDNA loci. (A) *psbC-trnS*, (B) *psbM-trnD*, and (C) *rpS16*. * indicates support > 70%, ** indicates support > 90%.

**Appendix S4** – Comparison of Polymorphic Information Content (PIC) of microsatellite from investigations of cultivated legumes.

**Appendix S5** – Results of the STRUCTURE analysis of cultivated winged bean based on ten microsatellites. Results for 2, 3, and 4 clusters are presented. Each accession is represented by a horizontal bar and the colors of each bar correspond to percentage membership to each of the clusters.

**Appendix S6** – Minimum Evolution phylogenetic tree for 53 accessions of winged bean.
LITERATURE CITED


DATA ACCESSIBILITY

cpDNA and ITS sequences have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/genbank) with accessions numbers MF980721 - MF980787 and MG969796 - MG969815, respectively.
Figure 1

A. cpDNA - ML

Erythrina grandiflora UGA
grandiflora ETH
sp1091 COD
lancifolia MWI
lukafuensis ZMB
lancifolia ZWE
obovalis TZA
monophyllus CIV
monophyllus MLI
lecomtei COD
monophyllus BFA
palastris BFA
palastris GMB
scandens MWI
scandens BDI
palastris NGA
palastris COD
sp1102 COD
tetragonolobus LBR
tetragonolobus NGA6
tetragonolobus NGA5
tetragonolobus MYS

0.05

B. cpDNA - Bayesian

Erythrina grandiflora UGA
grandiflora ETH
lancifolia MWI
lukafuensis ZMB
lancifolia ZWE
obovalis TZA
monophyllus CIV
monophyllus MLI
lecomtei COD
monophyllus BFA
palastris BFA
palastris GMB
scandens MWI
scandens BDI
palastris NGA
palastris COD
sp1102 COD
tetragonolobus LBR
tetragonolobus NGA6
tetragonolobus NGA5
tetragonolobus MYS

0.01

C. ITS - ML

Erythrina grandiflora ETH
grandiflora UGA
lancifolia MWI
lukafuensis ZMB
lancifolia ZWE
obovalis TZA
lecomtei COD
palastris BFA
palastris GMB
scandens MWI
scandens BDI
palastris NGA
palastris COD
sp1102 COD
tetragonolobus NGA6
tetragonolobus LBR
tetragonolobus MYS
tetragonolobus NGA5

0.1

D. ITS - Bayesian

Erythrina grandiflora ETH
grandiflora UGA
lancifolia MWI
lukafuensis ZMB
lancifolia ZWE
obovalis TZA
lecomtei COD
palastris BFA
palastris GMB
scandens MWI
scandens BDI
palastris NGA
palastris COD
sp1102 COD
tetragonolobus NGA6
tetragonolobus LBR
tetragonolobus MYS
tetragonolobus NGA5

0.1
Figure 3