

Origin of year-long bean (*Phaseolus dumosus* Macfady, Fabaceae) from reticulated hybridization events between multiple *Phaseolus* species

Angela M. Mina-Vargas¹, Peter C. McKeown^{1*}, Nicola S. Flanagan², Daniel G. Debouck³, Andrzej Kilian⁴, Trevor R. Hodgkinson⁵ and Charles Spillane^{1*}

¹Genetics & Biotechnology Laboratory, Plant and AgriBiosciences Research Centre (PABC), Árás de Brún 2006, University Road, National University of Ireland Galway, Galway H91 REW4, Ireland, ²Pontificia Universidad Javeriana, Cali, Colombia, ³International Center for Tropical Agriculture (CIAT), Apartado Aéreo 6713, Cali, Colombia, ⁴Diversity Arrays Technology Pty Ltd, Building 3, Level D, University of Canberra, Kirinari Street, Bruce, ACT 2617, Australia and ⁵Department of Botany, School of Natural Sciences, Trinity College Dublin, Dublin 2, Ireland

*For correspondence. E-mail charles.spillane@nuigalway.ie or peter.mckeown@nuigalway.ie

Received: 18 February 2016 Returned for revision: 11 May 2016 Accepted: 24 May 2016 Published electronically: 6 August 2016

- **Background and Aims** Improved understanding of the secondary gene pools of crops is essential for advancing genetic gain in breeding programmes. Common bean, *Phaseolus vulgaris*, is a staple crop with several wild relatives in its secondary gene pool. The year-long bean, *P. dumosus*, an important crop in Guatemala, is considered particularly closely related to *P. vulgaris* and a potential source of novel variation. However, the genetic diversity and relationship to other *Phaseolus* species of *P. dumosus* remain unclear.
- **Methods** We conducted the first comprehensive investigation of *P. dumosus* genetic diversity using both nuclear and chloroplast genome markers. Our nuclear marker set included over 700 markers present within the *Phaseolus* DArT (Diversity Arrays Technology) array, which we applied to *P. dumosus* and other relatives of *P. vulgaris* (including every secondary gene pool species: *P. acutifolius*, *P. albescens*, *P. coccineus* and *P. costaricensis*).
- **Key Results** *Phaseolus dumosus* arose from hybridization of *P. vulgaris* and *P. coccineus*, followed by at least two later hybridizations with sympatric congener populations. Existing *P. dumosus* collections have low genetic diversity.
- **Conclusions** The under-utilized crop *P. dumosus* has a complex hybrid origin. Further sampling in the region in which it arose may uncover additional germplasm for introgressing favourable traits into crops within the *P. vulgaris* gene pool.

Key words: *Phaseolus dumosus*, year-long bean, common bean, genetic diversity, secondary gene pool, hybridization, crop wild relative (CWR), reticulate evolution, sympatry, inter-species.

INTRODUCTION

The conservation and sustainable utilization of crop genetic resources is essential for humanity to address food security in the face of a growing global population and climate change challenges (McCouch *et al.*, 2013). Crop genetic resources consist of the genetic diversity of germplasm that can be used in crop breeding programmes (Glaszmann *et al.*, 2010) and can be organized into gene pools in accordance with the gene pool concept of Harlan and de Wet (1971). The primary gene pool of a crop consists of its own cultivars and landraces (and sometimes its wild progenitor), while the secondary gene pool consists of other related species from which agriculturally useful traits can be introgressed (Maxted *et al.*, 2006).

Beans are a globally important component of human diets, food security and livelihoods (Broughton *et al.*, 2003). The common bean, *Phaseolus vulgaris*, is a staple crop throughout South America, Central America and Africa, where it is an important source of dietary protein and a target of ongoing breeding and biofortification programmes (Gepts and Debouck, 1991; Schmutz *et al.*, 2014). To meet current and emerging production challenges, novel genetic variation needs to be

identified within the genetic diversity of germplasm collections and *in situ* populations (Ramírez-Villegas *et al.*, 2010; Beebe *et al.*, 2012; Porch *et al.*, 2013).

While *Phaseolus* is a large genus of ~50 species (Delgado-Salinas *et al.*, 2006), the secondary gene pool of *P. vulgaris* is generally restricted to include only *P. dumosus* Macfady (syn. *Phaseolus polyanthus* Greenm.), *P. coccineus* L., *P. costaricensis* Freytag & Debouck and *P. albescens* McVaugh ex R. Ramirez. & A. Delgado (Singh, 2001). Other *Phaseolus* species, such as *P. acutifolius*, constitute a tertiary gene pool because they are more distantly related, with sexual hybridization barriers (Delgado-Salinas *et al.*, 2006). Greater understanding of the genetic variation within the secondary gene pool can guide breeding via inter-specific crosses, and can identify populations of conservation priority (Porch *et al.*, 2013). Of the species within the secondary gene pool, *P. dumosus* (year-long bean) is closely related to both *P. vulgaris* and to *P. coccineus*, which is also a widespread crop. However, *P. dumosus* is considered to be less genetically diverse than either of these other two species, suggesting a more recent domestication (Freytag and Debouck, 2002). Importantly, *P. dumosus* displays characteristics of agronomic interest for climate change adaptation, including

tolerance to wet and high-altitude conditions (Singh *et al.*, 1991; Schmit, 1992), and field resistance to ascochyta leaf blight, anthracnose and white mould (Baudoin *et al.*, 1997; Mahuku *et al.*, 2002).

Unlike *P. coccineus* and *P. vulgaris*, ancestral wild types of *P. dumosus* are known only from the highlands in Guatemala where it is commonly grown. Wild *P. dumosus* displays an intermediate morphology with characteristics shared with *P. coccineus* (semi-tuberous root, large seeds) and *P. vulgaris* (epigeal cotyledons, introrse anthers). These features led to the suggestion that *P. dumosus* might derive from the hybridization of *P. coccineus* and *P. vulgaris* (Hernandez-Xolocotzi *et al.*, 1959; Miranda Colín, 1967). This hypothesis was contradicted by Smartt (1973), and common ancestry was proposed as a more plausible explanation (Piñero and Eguiarte, 1988). In recent years, the hypothesis of a hybrid origin has received support from discrepancies observed between plastid DNA (cpDNA) and nuclear DNA markers (Delgado-Salinas *et al.*, 1999; Spataro *et al.*, 2011). Typically, *P. dumosus* is resolved to be more closely related to *P. vulgaris* when cpDNA is used, but to *P. coccineus* with nuclear markers. It has also been proposed that a reticulation event may have occurred during the evolution of *P. dumosus*, possibly also involving *P. costaricensis* (Llaca *et al.*, 1994; Debouck, 1999). However, most recent analyses in *Phaseolus* have focused on the relationships of *P. vulgaris* or *P. coccineus*, and only small numbers of *P. dumosus* individuals have been investigated.

In this study, we performed a multilocus population-level study on samples of *P. dumosus* deposited in the germplasm banks of the International Center for Tropical Agriculture (CIAT) and the USDA. We conducted a comparison with the other members of the secondary gene pool of *P. vulgaris* to investigate genetic diversity, hybridization and introgression in the group. Specifically, we analysed genetic markers from the nuclear genomes of *P. dumosus* and the other primary and secondary gene pool species using a DArT array (Diversity Arrays Technology; Jaccoud *et al.*, 2001), augmented with chloroplast genome markers obtained by DNA sequencing. Our study presents the first robust model for the evolution of *P. dumosus* and its relationship to common bean and other members of its gene pool.

MATERIALS AND METHODS

Plant materials

Sixty-three accessions of different *Phaseolus* species were obtained from the CIAT, Cali, Colombia, and the USDA GRIN-NGP Germplasm Collection (Table 1). Seeds of each line were germinated by scarifying and placing them on a layer of water-moistened cotton wool inside a beaker. Germinated seedlings were transferred to a growth chamber (Perceval Scientific Inc., Germany) and grown under a 12-h light/12-h dark (24 h) light cycle at 22 °C/21 °C and harvested at 10 d post-germination for DNA extraction.

DArT array

Genomic DNA was isolated for inclusion in the *Phaseolus* DArT array (<http://www.diversityarrays.com/index.html>). DNA

was isolated from single plants following a proteinase protocol (Afanador *et al.*, 1993) and DNA quality tested by digestion with HindIII and PstI (New England Biolabs) according to the manufacturer's instructions. Digested and non-digested DNA was run on 1 % w/v agarose electrophoresis gels and visualized under UV using a Syngene Gel Genius Bio Imaging System. DNA concentrations were quantified using a NanoDrop ND-1000 spectrophotometer (Labtech, UK). Approximately 1 µg of pure DNA from each sample was pooled into skirted V-bottom 96-well plates and used for library development following digestion with PstI/BstNI to reduce complexity. Digested samples were cloned, amplified by PCR, isolated, arrayed in solid-phase slides and hybridized as described previously (Bríñez *et al.*, 2012). A total of 4208 polymorphic markers were detected among the clones of the *P. vulgaris* secondary gene pool. Marker quality was assessed using the call rate, discordance and *P* value; the 742 markers that displayed call rate > 95 %, discordance = 0 and *P* > 0.75 were used for analysis

Analysis of genetic diversity of *P. dumosus*

A presence/absence matrix of DArT markers was used to investigate the genetic diversity and structure of the collection, and also to perform an ancestry analysis of *P. dumosus* for evaluation of the possibility of past hybridization events involving *P. coccineus* and *P. vulgaris*. Genetic structure was characterized by genetic distance- and model-based clustering. Genetic distance-based clustering was undertaken using principal coordinates analysis (PCoA) with a pairwise Euclidian distance matrix as an input. The matrix and PCoA were calculated using GenAlEx 6.5 (Peakall and Smouse, 2006); all available covariance and distance options were tested. The matrix was also used to calculate an analogue of F_{ST} called Φ_{PT} , which allows population genetic differentiation to be estimated from dominant markers. Analysis of molecular variance (AMOVA) among and within all species was also estimated using Arlequin 3.5 (Excoffier and Lischer, 2010). Model-based clustering analysis was performed by applying a Markov chain Monte Carlo (MCMC) algorithm with STRUCTURE 2.3.4 (Pritchard *et al.*, 2000; Falush *et al.*, 2003). The estimated membership or ancestry (*Q*) was also determined for each sample; allelic frequencies were set as correlated among populations. *K* was set from 1 to 10 and the simulation was run with 10^5 iterations for the burn-in period and 10^6 iterations for the MCMC; the Evanno test (Δk ; Evanno *et al.*, 2005) and the probability of the best *K* (Pritchard *et al.*, 2000) were applied to infer the optimal *K* value. Both tests were performed with CLUMPP (Jakobsson and Rosenberg, 2007) and Structure Harvester (Earl, 2012).

Phylogenetic reconstructions

Genetic distances between genotypes were computed using the Jaccard dissimilarity measure (Perrier *et al.*, 2003). A dendrogram was constructed using the weighted neighbour-joining method (Saitou and Nei, 1987) as implemented in DARwin 5 (Perrier and Jacquemoud-Collet, 2006). The significance of each node was evaluated by bootstrapping with 1000 replications. To study conflicting phylogenetic signals potentially caused by reticulation events, the genetic distance matrix for

TABLE 1. Accessions of *Phaseolus* species used in this study. Accessions marked with an asterisk (*) were used only in the DART array

Species	Accession	Biological status	Country of origin	Department/state	Area (where known)	Altitude (m)	Latitude	Longitude	Date of collection	Date of receipt
<i>P. acutifolius</i> var. <i>tenuifolius</i>	G40178	Wild	Mexico	Chihuahua	Balleza	1700	26-90	-106-30	01/09/84	01/06/87
	G40286	Wild	USA	Arizona	Cochise	1788	31-72	-109-33	20/09/90	01/02/95
<i>P. albescens</i>	PL3592	Wild	Mexico	Jalisco	Tepatitlan	2450	19-60	-103-52	17/12/85	21/12/85
	G35837A	Wild	Guatemala	Huehuetenango	Chiamtla	1800	14-75	-91-47	24/01/95	01/02/95
<i>P. coccineus</i>	G36285	Wild	Guatemala	Quezaltenango	Zunil	2010	14-58	-90-08	06/12/87	22/12/87
	G36353	Wild	Guatemala	Jalapa	El Mirador	1570	14-72	-91-52	25/01/95	01/02/95
	G36006	Wild	Guatemala	Quezaltenango	S. M. de Jesus	2040	14-75	-91-48	25/01/95	01/02/95
	G36206	Wild	Guatemala	Quezaltenango	Zunil	1700	0-35	-78-47	30/06/90	18/07/90
	G35650	Wild	Mexico	Morelos	Imbabura	2100	14-47	-90-73	04/12/85	01/08/79
	G35757	Wild	Guatemala	Sacatepéquez	S. M. de Jesus	2080	14-55	-90-85	20/01/95	21/12/85
	G36002	Wild	Guatemala	Sacatepéquez	S. Miguel Dueñas	29-30	23-95	-107-72	01/10/91	01/02/95
	G36334	Wild	Mexico	Chihuahua	Gomez Farias	1830	21-27	-104-28	01/11/78	01/10/91
	G35859	Wild	Mexico	Durango	Nombre de Dios	2113	17-52	-96-93		14/03/79
	PI 449381*	Cultivar	Mexico	Veracruz	Poza Rica	2113	17-52	-96-93		01/03/80
	G35124*	Landrace	Mexico	Oaxaca	Nacaltepec	2113	14-77	-91-18		14/02/68
	PI 433237*	Cultivar	Guatemala	Sololá						01/12/78
PI 451863*	Cultivar	Guatemala	Guatemala City						01/07/80	
PI 451868*	Cultivar	Guatemala	Sololá						01/07/80	
PI 451870*	Cultivar	Guatemala	Sololá						01/07/80	
PI 451872*	Cultivar	Guatemala	Sololá						01/07/80	
G35110*	Landrace	Guatemala	Quiche	Chichicastenango	1219	14-95	-91-12		01/12/76	
G35572*	Landrace	Guatemala	Alta Verapaz	Coban	1600	15-47	-90-37		12/02/53	
G35609*	Landrace	Mexico	Puebla						19/04/66	
G35370*	Cultivar	Mexico							16/12/69	
G35057*	Landrace	Mexico	Hidalgo	Tulancingo		20-10	-98-32		08/05/52	
G40805	Wild	Costa Rica	San Jose	Alajuelita	1620	9-87	-84-12		01/07/87	
G40604	Wild	Costa Rica	Cartago	La Union	1750	9-95	-83-95		01/07/87	
G35946	Wild	Guatemala	Sacatepéquez	San Miguel Dueñas	1730	14-55	-90-85		22/12/87	
G36127	Landrace	Peru	Cajamarca	Conchan	2430	-6-45	-78-63		01/09/86	
G35641	Landrace	Mexico	Puebla	Tetela de Ocampo	1730	20-28	-97-95		01/06/77	
G35758	Wild	Guatemala	Sacatepéquez	S. M. de Jesus	1550	14-48	-90-70		21/12/85	
G35877	Wild	Guatemala	Sololá	Panajachel	1680	14-77	-91-17		04/12/85	
G35908	Wild	Guatemala	Sacatepéquez	S. Miguel Dueñas	1940	14-55	-90-83		21/12/85	
G36045	Wild	Guatemala	Guatemala	Amatitlan	1800	14-43	-90-58		22/12/87	
G36286	Wild	Guatemala	Quezaltenango	S. M. de Jesus	1460	14-70	-91-53		01/02/95	
PI 355423*	Cultivar	Ecuador	Mercado del Sur	Guayaquil	660	-1-47	-77-98		25/01/95	
G35569*	Landrace	Guatemala	Huehuetenango	Jacaltenango	1430	14-63	-89-98		03/07/50	
G35584*	Landrace	Guatemala	Jalapa	Nahuala	2500	14-85	-91-32		19/01/66	
G35082*	Landrace	Guatemala	Sololá	Chuy	2113	15-50	-91-57		22/11/66	
G35611*	Landrace	Guatemala	Huehuetenango		2495	14-92	-91-15		05/03/51	
G35570*	Landrace	Guatemala	Sololá		2330	14-93	-91-12		14/03/51	
G35040*	Landrace	Guatemala	Totonicapan	Chichicastenango	1530	14-43	-91-33		16/04/51	
G35089*	Landrace	Guatemala	Quiche	Antigua	100	10-77	-85-48		19/01/66	
G35042*	Landrace	Guatemala	Zacapa	Cañas	100	10-77	-85-48		16/04/51	
G25224	Wild	Costa Rica	Guacacaste	Cañas	1730	14-75	-91-52		01/05/88	
G25226	Wild	Costa Rica	Guacacaste	Cañas	1730	14-75	-91-52		01/03/73	
G50387	Wild	Guatemala	Quezaltenango	Zunil	1150	18-32	-99-82		01/02/95	
G7479	Wild	Mexico	Guerrero	Teloloapan	1760	14-75	-91-52		20/09/75	
G50388B	Wild	Guatemala	Quezaltenango	Zunil	1760	14-75	-91-52		24/01/95	

(continued)

TABLE 1. Continued

Species	Accession	Biological status	Country of origin	Department/state	Area (where known)	Altitude (m)	Latitude	Longitude	Date of collection	Date of receipt
	G23463	Wild	Mexico	Chihuahua	Yepachic	1530	28.33	-108.50	15/09/84	01/01/88
	G12853	Wild	Guatemala	Jutiapa	Quesada	873	14.28	-89.90		01/20/78
	G19908	Wild	Guatemala	Sacatepéquez	S. Miguel Dueñas	1820	14.55	-90.83	12/06/85	21/12/85
	G19909	Wild	Guatemala	Sacatepéquez	S. Miguel Dueñas	1820	14.55	-90.83	12/06/85	21/12/85
	G21244	Wild	Peru	Cajamarca	San Pablo	2020	-7.18	-78.83	13/08/86	01/09/86
	G24358	Wild	Mexico	Michoacan	Benito Juarez	1300	19.30	-100.42	21/11/84	01/03/92
	G14*	Cultivar	Venezuela	Caracas		1160	10.50	-66.93		28/02/35
	G12*	Cultivar	Venezuela	Caracas		1160	10.50	-66.93		28/02/35
	G13*	Cultivar	Venezuela	Caracas		1160	10.50	-66.93		28/02/35
<i>P. vulgaris</i> × <i>coccineus</i>	G50998C*	Hybrid	Colombia	Antioquia	Andes	1500	5.60	-75.83	03/01/95	12/01/95
<i>P. vulgaris</i> × <i>dtamosus</i>	G35980E*	Hybrid	Colombia	Tolima	Chaparral	2120	3.75	-75.42	04/09/90	15/09/90
	G35980F*	Hybrid	Colombia	Tolima	Chaparral	2120	3.75	-75.42	04/09/90	15/09/90

the wild genotypes was used to generate a phylogenetic network with SplitsTree 4 (Huson and Bryant, 2006) using NeighborNet uncorrected distances and the EqualAngle splits transformation method. The goodness of fit was assessed by the least squares fit value (LS fit) as described below.

cpDNA and nuclear internal transcribed spacer (ITS) sequencing

Markers for 11 regions of the cpDNA were assessed (*trnL-trnF* spacer, *matK-trnK*, *rpoC1-rpoC2* spacer, *petA-psbE* spacer, *rpL16* intron, *atpβ-rbcL* spacer, *ndhA* intron, *accD-psal* spacer, *trnT-trnL* spacer, *trnL* intron and *rps14-psaB* spacer). As the last four markers could be amplified from all species, they were chosen for further use. The internal transcribed spacer (ITS) region of nuclear ribosomal DNA was also sequenced for comparison. PCR was performed with a Dyad Disciple™ Peltier Thermal Cycler (MJ Research) using GoTaq® Flexi DNA polymerase (Promega, WI) under the manufacturer's recommended conditions. PCR fragments were resolved on 1.5 % w/v agarose/Tris-Borate-EDTA (TBE) gels and visualized under UV with 1× SYBR® Safe DNA Gel Stain (Invitrogen, UK). For primer sequences and PCR programmes for both cpDNA and ITS loci see Supplementary Data Table S1. Products were purified with QIAquick PCR purification kits (Qiagen, Germany) according to the manufacturer's instructions and eluted into 30 µL of autoclaved Millipore ultra-pure water. Products obtained from two independent samples of each accession were sequenced in both directions (GATC Biotech, Germany) using single-read 11× ABI 3730xl sequencing technology. It was intended to obtain complete sequences of two samples per accession and those that failed after four reactions were eliminated from the analysis.

Sequence alignment and diversity analysis

Forward and reverse sequences were assembled into contigs using the ContigExpress tool in Vector NTI I® V.11.5 (Life Technology, USA) and the resulting consensus sequence was exported in FASTA file format. Alignments for each locus were constructed with MUSCLE (Edgar, 2004) and manually verified with Jalview 2.7 (Waterhouse et al., 2009) or Gblocks[©] (Castresana, 2000). The four cpDNA loci and one nuclear marker (nrITS) were analysed by independently calculating genetic divergence parameters in DnaSP v5 (Librado and Rozas, 2009). Allelic diversity was calculated as haplotype diversity, $h = n(1 - \sum f_i^2)/(n-1)$, where f is the frequency of the i th allele and n is the number of samples, modified from Nei (1987), replacing $2n$ with n . Nucleotide diversity (π) was calculated according to Nei and Li (1979). Number of segregating sites (S) was calculated according to Nei and Kumar (2000). Apart from π (the significance of which was evaluated by Student's t -test), all tests were applied both between species and between and within accessions, and their significance was assessed by AMOVA in Arlequin 3.5 (Excoffier and Lischer, 2010) using fixation indices (F statistics) calculated according to Wright (1949). Significance was tested using 1000 permutations with $\alpha > 0.05$ (Excoffier et al., 1992).

Determination of pairwise differences

The divergences per locus between individuals of the six species of *Phaseolus* were measured by determining pairwise genetic distance (distance matrix), the total average proportion of nucleotide differences between groups (D_{XY}) and the net nucleotide substitutions per site between populations (D_A), given by $d_A = d_{XY} - (d_X + d_Y)/2$ where XY is the average distance between groups X and Y , and d_X and d_Y are the mean within-group distances. Calculations were performed on the nrITS sequences and a concatenated matrix of the four cpDNA loci. Graphical representations of the divergences were made in R (<http://www.R-project.org>).

Phylogenetic trees and hybridization networks

Gene trees were built using the cpDNA and ITS sequences described above. The cpDNA alignments were analysed as a matrix of concatenated loci or as gene region partitions of this. The concatenated matrix was assembled using Mesquite 2.75 (Maddison and Maddison, 2011). Maximum likelihood (ML) trees were calculated using the GTR model (Tavaré, 1986) and drawn for the cpDNA and ITS, using MEGA5 (Tamura et al., 2011). Matrices were subjected to bootstrap testing (Felsenstein, 1985) calculated from 1000 replicates for the nrITS phylogeny and 10 000 replicates for that of the cpDNA. Branches with bootstrap values <60 % were collapsed. Bayesian MCMC phylogenetic trees were created with MrBayes version 3.2 (Ronquist and Huelsenbeck, 2003; Ronquist et al., 2012). Optimal substitution models were identified for each for each locus using jModelTest and the Akaike information criterion (Akaike, 1973) following Kelchner and Thomas (2007), Posada (2008) and Darriba et al. (2012). For the priors of the model for the chloroplast sequences, three partitions were established: partition 1 corresponds to the *trnT-trnL* spacer and *trnL* intron, analysed under the GTR model; partition 2 corresponds to the *rps14-psaB* spacer, analysed under the F81 model (Felsenstein, 1981); and partition 3 corresponds to the *accD-psaI* spacer, analysed under the GTR model. A discrete γ distribution with six discrete categories was used to model evolutionary rate differences among sites. All parameters were unlinked to allow all partitions to vary under different rates. The nrITS region was analysed under the GTR model, with a discrete γ distribution and four discrete categories. Tree topologies were determined in two parallel runs, each performed with four Markov chain settings (three ‘heated’ and one ‘cold’); searches started from a random tree and 25 % of the sampled trees were used for burn-in. Ten million generations were tested and tree samplings saved every 1000th generation; for the nrITS, 1 000 000 generations were tested and tree samplings were saved every 500th generation. A consensus hybridization network (McBreen and Lockhart, 2006) was generated in Dendroscope 3.1 (Huson and Scornavacca, 2012) using the cpDNA and nrITS ML trees, using only accessions represented in both.

RESULTS

DArT analysis indicates that P. dumosus is intermediate between P. vulgaris and P. coccineus

To investigate the genetic diversity of *P. dumosus* within the *P. vulgaris* secondary gene pool, we analysed the genomic

DNA of individual plants from 45 wild accessions and 28 cultivars using a *Phaseolus* DArT array (Table 1). The wild accessions included plants from all but one of the locations from which *P. dumosus* has been collected (Schmit and Debouck, 1991). Wild and domesticated forms of *P. vulgaris* and the other species of the secondary gene pool were also analysed, together with samples of the tertiary gene pool species *P. acutifolius* as an outgroup (Table 1). For each accession, DNA was taken from at least three plants, generating 152 samples once DNA quality controls were applied. In total, 4208 of the clones that we contributed to the *Phaseolus* DArT array were suitable for use as polymorphic markers in *P. dumosus*: we used these to build a binary (presence versus absence) matrix for each marker in each sample, performed a PCoA and calculated pairwise genetic diversity.

Within *P. dumosus*, three principal coordinates separated the populations into three groups. The wild accession G36286 formed a second group with two landraces (PI195389 and PI317563; top right of Fig. 1A), but all other accessions resolved to a broad group occupying an intermediate position (Fig. 1A). Hence, the three principal coordinates did not resolve accessions, landraces and cultivars into separate categories. This was confirmed by AMOVA, which indicated that variation between categories only accounted for ~15 % of total divergence ($F_{ST} = 0.15385$, $P = 0.0127$), while the majority (~85 %) of the variation corresponded to polymorphisms present within the categories (Supplementary Data Table S2).

To determine the genetic structure of the secondary gene pool we also performed PCoA across all *Phaseolus* species tested. We again discovered that much of the variation (72.52 %) was explained by three coordinates (47.81, 20.85 and 3.86 %; Table S2) which clearly separated accessions by species (Fig. 1B). Certain cultivars of *P. vulgaris* formed an additional group with two cultivated hybrids (*P. vulgaris* × *coccineus*; *P. vulgaris* × *dumosus*). Separation between Mesoamerican and South American genotypes could be observed in this case, but only for *P. vulgaris*. Notably, *P. dumosus* genotypes occupied an intermediate position between *P. coccineus* and *P. vulgaris*. G36286 again diverged from most *P. dumosus* populations, clustering with *P. albescens*. We conclude that *P. dumosus* nuclear markers resolve the species as intermediate between *P. vulgaris* and *P. coccineus*, but provide little differentiation between *P. dumosus* populations.

P. dumosus is closely related to sympatric populations of P. coccineus

To analyse the population structure within *P. dumosus* accessions in more detail, we performed a Bayesian structure analysis. Using median values of log likelihood value, $L(k)$, the K with the highest probability of best representing the data was $K = 8$ (Fig. 2; Supplementary Data Table S3). The probability distribution was, however, somewhat bimodal, with another peak at $K = 6$, which suggests some secondary structure among genotypes of *P. vulgaris* and *P. coccineus*. When $K = 2$, *P. dumosus* and *P. coccineus* were grouped together, but at $K > 2$ all *P. dumosus* accessions formed a single, separate cluster, in agreement with the

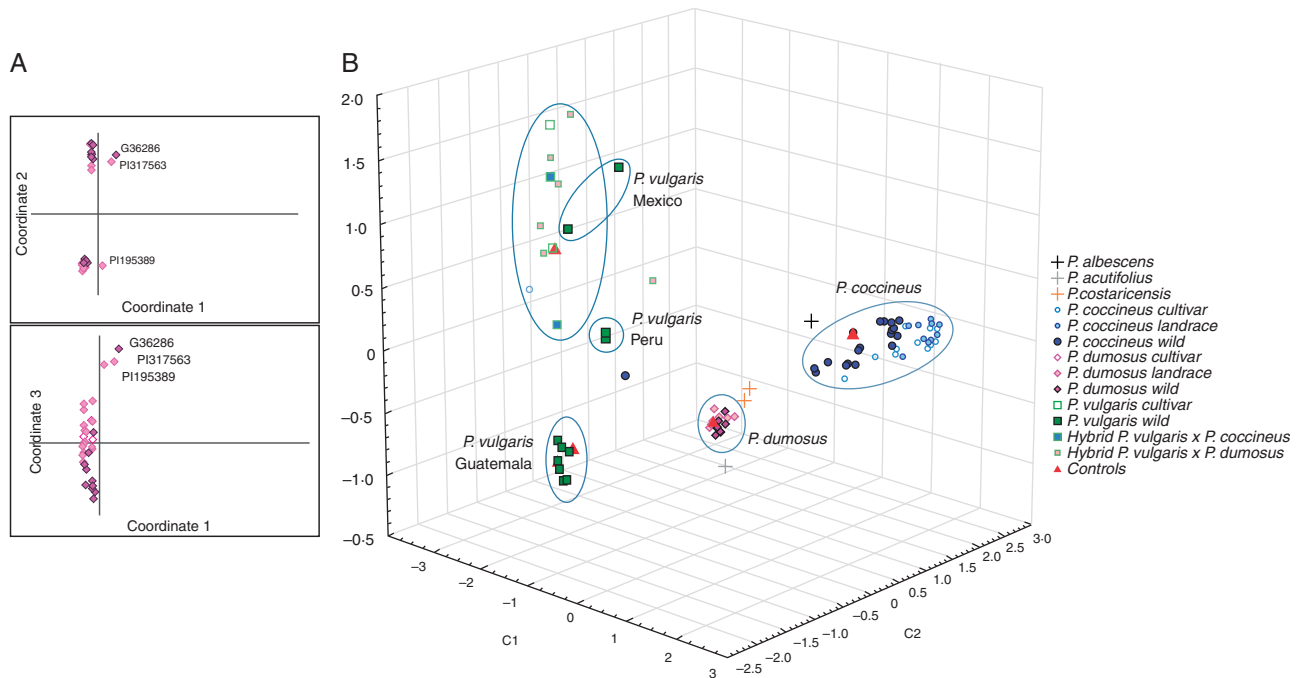


Fig. 1. Genetic distance among species of the *P. vulgaris* secondary gene pool. (A) Principal coordinates analysis of 742 DArT markers for *P. dumosus* indicating that 72.10 % of the total genetic variation is found among the first three components. Pink diamonds, black outline, wild accessions; pink diamonds, pink outline, landraces; open diamonds, cultivars. (B) Principal coordinates analysis accounting for 72.52 % of the molecular variation found among 704 DArT markers, with 85 % of variation in *P. dumosus* accounted for by the first three components. *P. dumosus*: wild (pink diamonds, black outline), landrace (pink diamonds, pink outline), cultivar (open diamonds). *P. coccineus*: wild (solid dark blue circles), landrace (solid light blue circles), cultivar (open circles). *P. vulgaris*: wild (solid green squares), cultivar (open squares). *P. costaricensis* (yellow +), *P. albescens* (black +), *P. acutifolius* (grey +; outgroup). Hybrids: *P. vulgaris* × *P. dumosus* (solid pink squares), *P. vulgaris* × *P. coccineus* (solid blue squares). Controls (red triangles).

PCoA. Among *P. dumosus* samples, admixture was only detected for accession G36286, which was admixed with both *P. dumosus* ($Q = 0.856$) and *P. coccineus* ($Q = 0.144$). This accession was collected in the Guatemalan region of Quetzaltenango, in an area of Santa Maria de Jesus to which *P. coccineus* is also endemic. *Phaseolus coccineus* itself displayed much more widespread admixture; 14 % of samples showed some evidence of admixture. Three percent of *P. vulgaris* samples were also admixed. Differentiation between Guatemalan and Mexican genotypes was again detected for both *P. coccineus* and *P. vulgaris* but not *P. dumosus*. For *P. vulgaris*, Peruvian genotypes also resolved separately.

To further investigate the evolutionary relationships within the secondary gene pool, a weighted neighbour-joining dendrogram was computed using Jaccard dissimilarity (Saitou and Nei, 1987; Perrier et al., 2003), using all 4208 markers to account for any variation not captured by the most polymorphic markers. The dendrogram resolved two main clusters, which we termed A and B, both of which were strongly supported (bootstrap support [BS] = 100 %; Fig. 3). Cluster A comprised all accessions of *P. vulgaris*, while cluster B included representatives of *P. dumosus*, *P. coccineus* and *P. albescens*. The tree clearly resolved *P. dumosus* as monophyletic with a close affinity with *P. coccineus*, and more closely related to *P. coccineus* accessions from Guatemala than with those from Mexico (Fig. 3).

Genetic diversity in wild *P. dumosus* is lower than in other *Phaseolus* species

To determine whether phylogenies based on the chloroplast genome are congruent with those generated from the DArT array, we sequenced DNA from different cpDNA loci and also from the nuclear ITS region (Supplementary Data Fig. S1; Table S1). When applied to seven wild accessions of *P. dumosus*, we noted that the allelic diversity at these markers appeared to be low (Supplementary Data Table S4). Greater diversity was observed between species, which typically only shared a few alleles (Supplementary Data Fig. S2). To investigate this further, we determined the extent of nucleotide diversity for each species in the secondary gene pool and found that diversity in *P. dumosus* was indeed the lowest of any species ($\pi = 0.00035$; Supplementary Data Table S5). *Phaseolus coccineus* was the most polymorphic ($\pi = 0.00903$), with *P. vulgaris* intermediate. For the cpDNA, most molecular variation at the *trnL* intron was due to differences between species (Supplementary Data Table 6A). An AMOVA on the concatenated matrix assessed variation at the other three cpDNA loci and showed a similar result (Supplementary Data Table S6B). The pairwise divergence from *P. vulgaris* was also low, while the *P. dumosus*–*P. coccineus* divergence was somewhat greater (0.6 versus 0.03 %; Supplementary Data Fig. S3, Supplementary Data Table S7). The majority of the variation (76 %) was again concluded to be due to differences between species rather than within them (Fixation index among groups $F_{CT} = 0.76396$, $P = 0.01075$; Table S7).

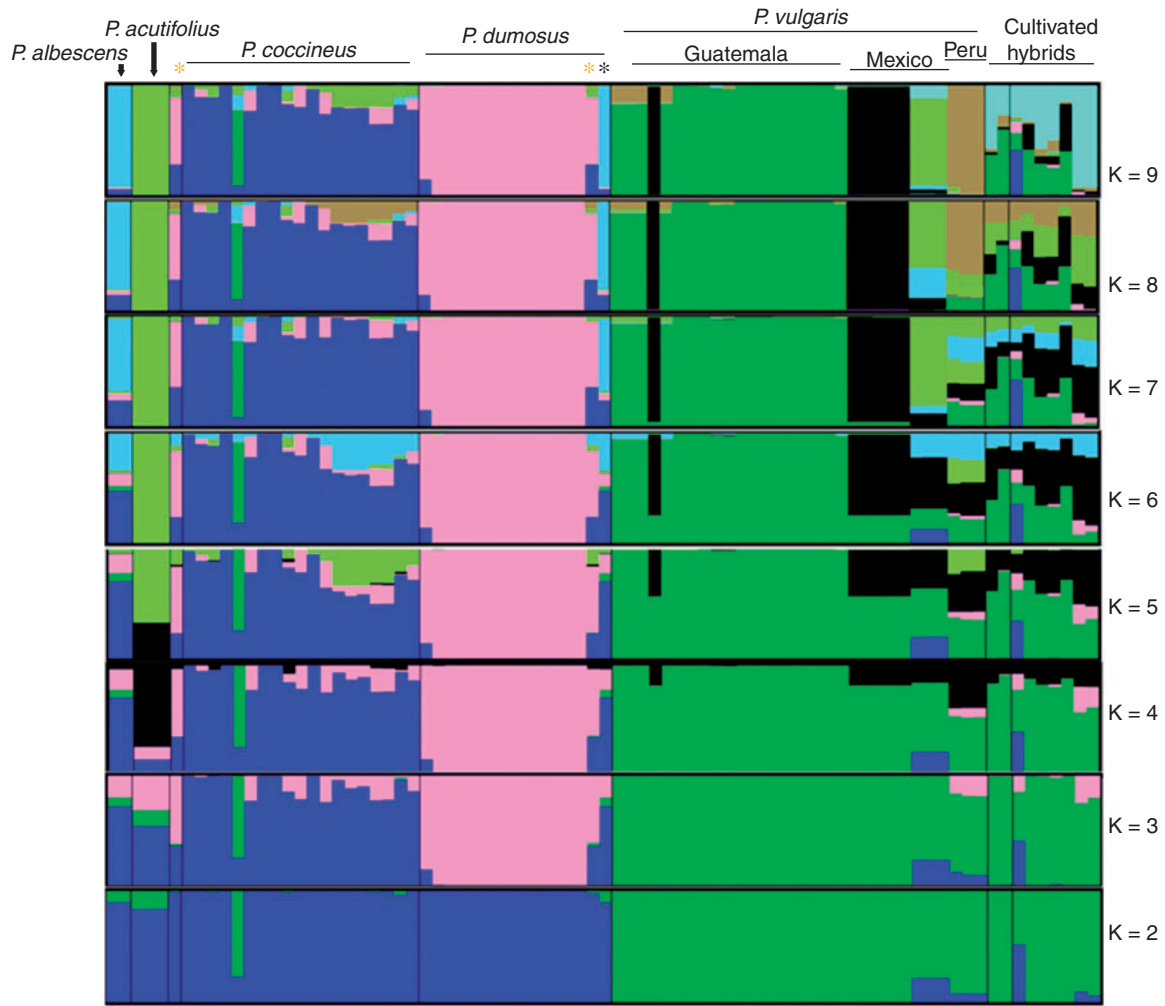


FIG. 2. Structure analysis of wild accessions of *Phaseolus* species in the *P. vulgaris* secondary gene pool. The level of ancestry was inferred by a Bayesian model, with each colour representing one population; each accession is represented by a vertical bar, and the length of each coloured segment in each vertical bar represents the proportion contributed by ancestral populations. Yellow asterisks, *P. costaricensis*, black asterisks, *P. albenscens*.

Conflict between cpDNA and nuclear ribosomal ITS phylogenetic reconstructions

The cpDNA and nrITS sequences were used to generate phylogenetic trees using both ML and Bayesian methods (Fig. 4). The topologies of the two trees for cpDNA data are summarized in an ML bootstrap support tree with added posterior probability values from the Bayesian analyses (Fig. 4A). Two major clades were evident within the secondary gene pool (Fig. 4A): this division was strongly supported by the Bayesian (posterior probability [PP] > 93 %) and the ML tree topology (BS > 82 %). Clade A comprised certain accessions of wild *P. dumosus* and a subclade of *P. vulgaris* (molecular diversity was too low to resolve the *P. dumosus* subclade further). One accession of *P. coccineus* (G36206) and one of *P. costaricensis* (G40604) also grouped in clade A as part of a polytomy with *P. dumosus* and *P. vulgaris*. Clade B comprised a *P. coccineus* subclade and a sister subclade containing the second accession of *P. costaricensis* and the single accession of *P. albenscens*. In the *P. coccineus* subclade, a well-supported branch (BS = 83 %, PP = 100 %) separated accession G35650-Mexico.

Topologies based on nrITS sequences were largely congruent between the Bayesian and ML analyses. Two major clades (C and D) were moderately supported by both methods (BS > 72 %, PP > 83 %); an optimal tree is displayed (Fig. 4B).

Notably, the results of the cpDNA-based phylogeny (Fig. 4A) resolved *P. vulgaris* as the sister species of *P. dumosus*. This was in contrast to the results from the nuclear genomic markers evaluated in the DArT array, in which *P. coccineus* had been resolved as sister to *P. dumosus* (Fig. 3). The nrITS trees instead agreed with the DArT analysis, with the Guatemalan subclade of *P. coccineus* placed as sister to *P. dumosus* (Fig. 5B). Furthermore *P. costaricensis* was resolved as polyphyletic: one accession clustered with *P. albenscens* (G40805), another in an unresolved position close to *P. dumosus* in both trees (G40604).

Probable hybridization events within the secondary gene pool

To test directly whether the conflict between plastid and nuclear DNA markers can be resolved, we developed a

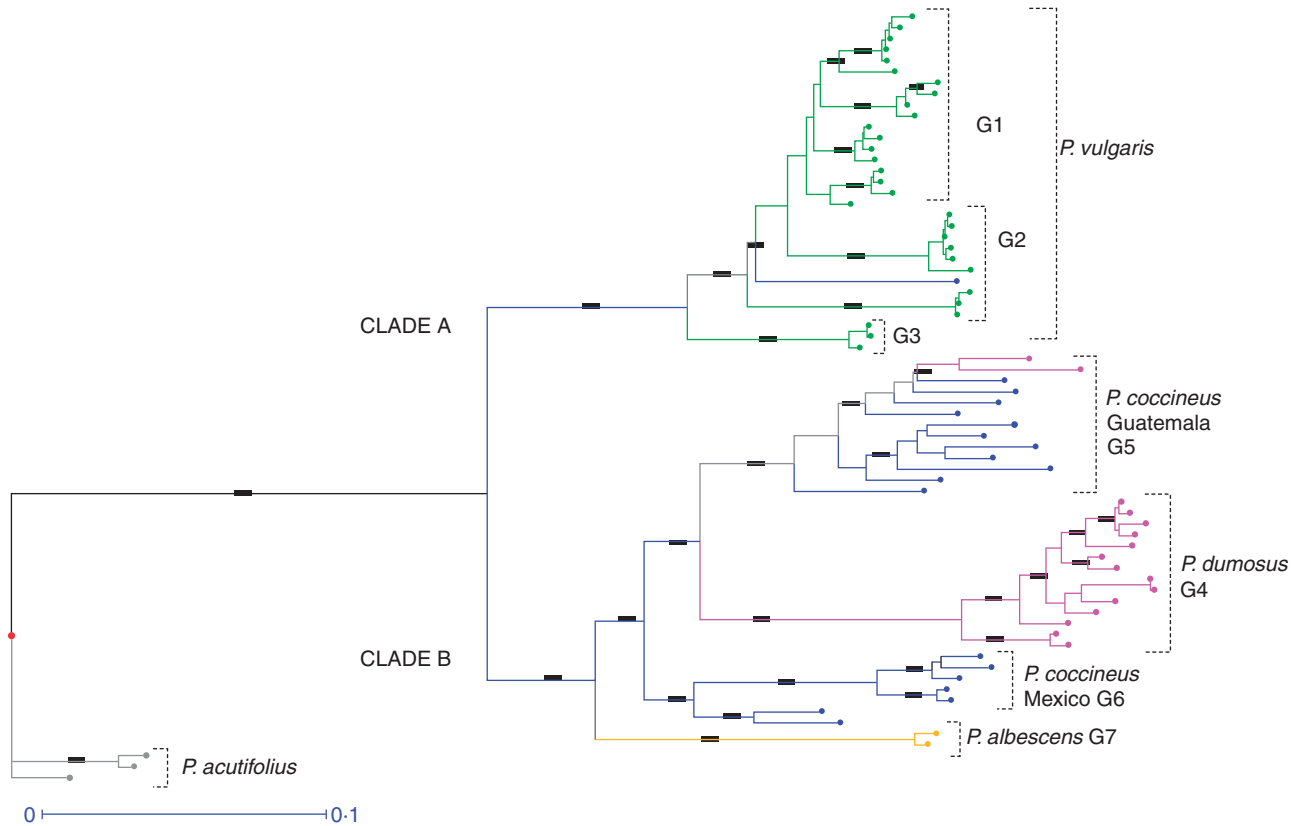


Fig. 3. Weighted neighbour-joining dendrogram for wild germplasm accessions of the *P. vulgaris* secondary gene pool. The tree was outlined from Jaccard's genetic distance calculated for polymorphisms in 4208 DArT markers. The two major clades (A and B) and their internal subclades are shown as follows: *P. vulgaris* (green lines) from Guatemala (G1), Mexico (G2) Peru (G3); *P. dumosus* (pink lines) (G4); *P. coccineus* (blue lines) from Guatemala (G5), from Mexico (G6) and *P. albes-cens* (orange lines) (G7). The tertiary gene pool species *P. acutifolius* was used as an outgroup (grey lines). Dark bars indicate branches supported by bootstrap values >90%. The line segment with the number 0.1 shows the length of branch that represents an amount genetic change of 0.1.

hybridization network to account for possible hybridization in the *P. vulgaris* secondary gene pool (Fig. 5). The network showed three splits, each corresponding to a species group agreeing with the tree topologies identified above. Importantly, this network identified 21 putative hybridization events, comprising 16 reticulation events within species and five putative hybridizations across species. A reticulation node was generated as an explanation of the divergence between the chloroplast and the nuclear genome of *P. dumosus*, strengthening the suggestion that *P. dumosus* arose via such a hybridization event. To test this hypothesis more thoroughly, a phylogenetic analysis was performed on the DArT array data. To identify any possible reticulations, the data were fitted into a NeighborNet (Fig. 6), which robustly supported two main clusters (LS fit = 99.83%), one comprising *P. vulgaris* accessions only and a second containing *P. dumosus*, *P. coccineus*, *P. costaricensis* and *P. albes-cens*. The latter cluster could be further broken down into two levels of sub-cluster, representing each species and each group of accessions. Conflicting signals or reticulations (represented by 'boxes' on the base of the nodes in Fig. 6) were detected both between and within the clusters, confirming that numerous between-population and between-species hybridization events are likely to have occurred.

DISCUSSION

Wild P. dumosus has low levels of genetic diversity

This study investigated the genetic diversity within germplasm of wild *P. dumosus* and close relatives, to better understand the diversity within the germplasm and to generate data that can inform conservation and breeding activities. Our results, which combine markers from nuclear and plastid genomes, indicate that nucleotide diversity of *P. dumosus* is the lowest of any species in the *P. vulgaris* secondary gene pool. Three further analyses based on the DArT data (PCoA, AMOVA and STRUCTURE; Fig. 2, Table S2) all agreed with this conclusion, and confirmed it across cultivars, landraces and wild accessions. As the wild forms of *P. coccineus* and *P. vulgaris* are considerably more diverse (Table S5), the lack of diversity in *P. dumosus* is not due to the use of inherently conserved loci. Low nucleotide diversity in wild populations can be associated with mating system, geographical distribution or gene flow (Ellstrand and Elam, 1993), although as *P. dumosus* is predominantly allogamous (Schmit *et al.*, 1994), mating system is unlikely to be responsible. Instead *P. dumosus* populations may be subject to a constricted distribution or selection under environmental restrictions. Indeed, wild *P. dumosus* currently has a small geographical range restricted to central and south

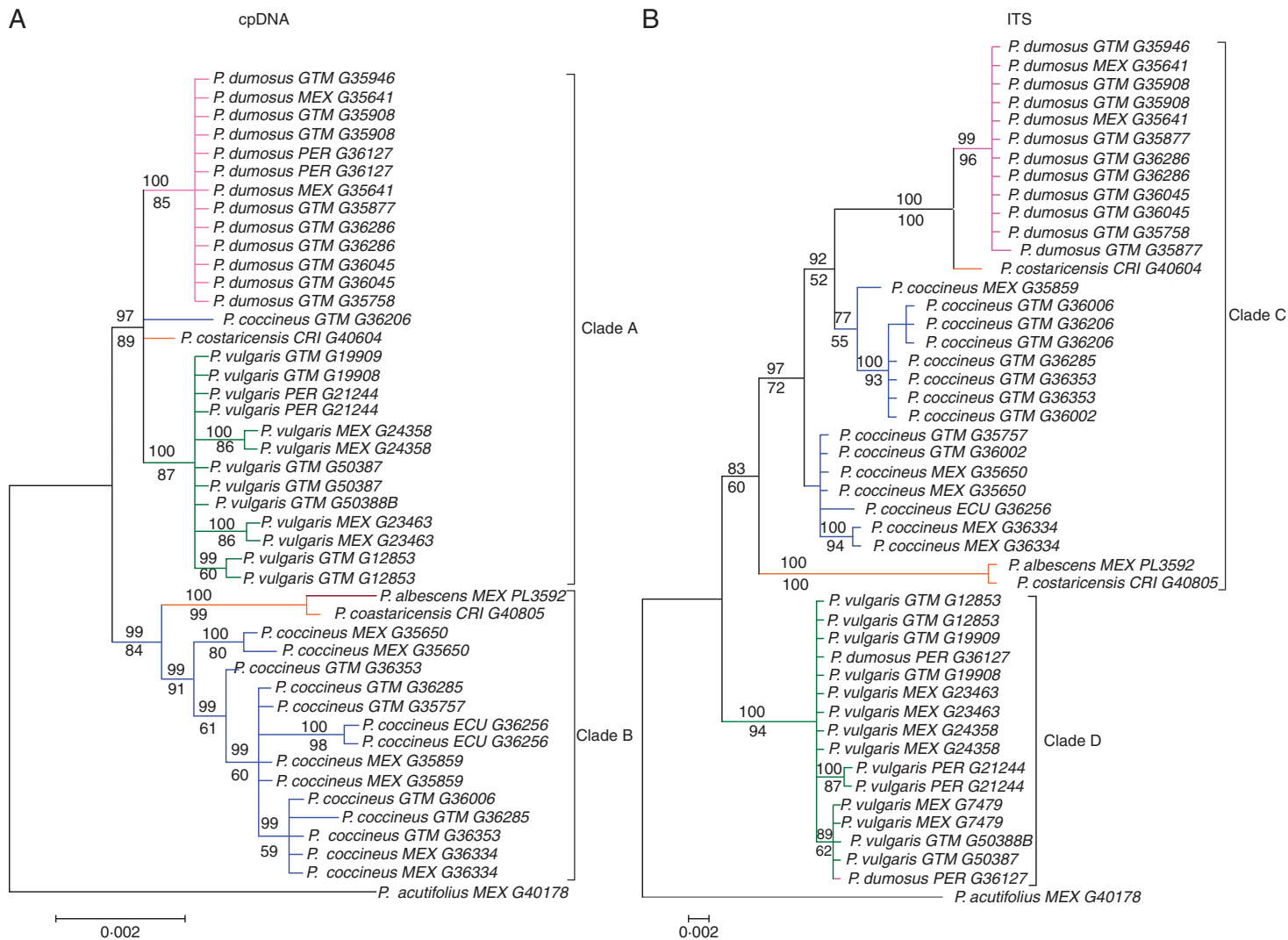


FIG. 4. Bootstrap consensus tree topologies for wild accessions of the *P. vulgaris* secondary gene pool. Trees were generated from sequences of the cpDNA (A) and ITS (B) region. Phylogenetic inference was made by Bayesian MCMC and ML analysis; numbers assigned to branches represent posterior probabilities (number over branch) and bootstrap values >60% (number under branch). Four major clades are designated.

Guatemala (Schmit and Debouck, 1991; Freytag and Debouck, 2002) and also grows under challenging environmental conditions such as high altitude and cool wet climatic conditions. If *P. dumosus* diverged from its sister species relatively recently, this could also contribute to the low genetic diversity. Variation was only observed in G35877-Solola, a Guatemalan accession that was previously identified as unusually divergent by isozyme analysis (Schmit and Debouck, 1991). This accession is included in the core collection in the CIAT germplasm bank (Tohme et al., 1995), which suggests that the core collection captures much of the genetic diversity found among wild *P. dumosus*. Given that the CIAT bean breeding programme has recently succeeded in transferring valuable resistance genes from *P. dumosus* to *P. vulgaris* (J. Tohme, pers. comm.), it is important for ongoing crop improvement that maximal genetic diversity from *P. dumosus* is captured in both the core and base collections.

Discrepancies between the nuclear and plastid genomes of P. dumosus

Our results further highlight the existence of discrepancies between nuclear and plastid genomes: DArT array analysis of *P. dumosus* strongly supports an affinity at nuclear loci with *P. coccineus*, as suggested previously (Piñero and Eguiarte, 1988; Delgado-Salinas et al., 1999, 2006). On the other hand, cpDNA sequences of *P. dumosus* resemble those of *P. vulgaris*. It should be noted that there is a discrepancy between the numbers of markers analysed in each genome, with fewer derived from the cpDNA. However, our results are in agreement with those proposed following analysis of cpSSRs (Angioi et al., 2009; Desiderio et al., 2013), which suggests that our study assessed sufficient plastid markers to draw valid inferences. Interestingly, results based on the nuclear genome reflect sympatric associations: wild *P. dumosus* more closely resembles the Guatemalan accessions of

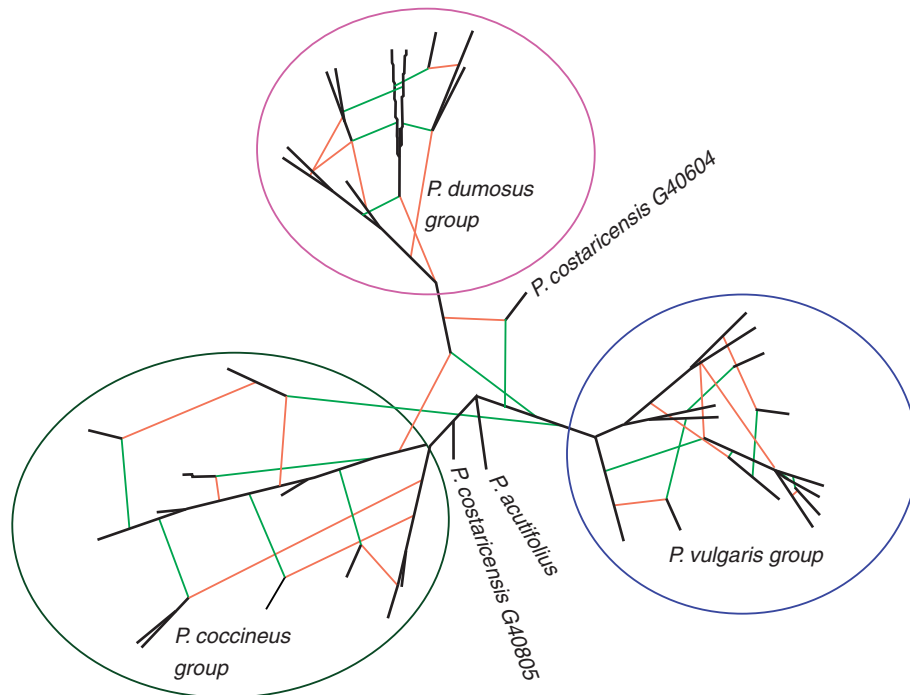


FIG. 5. Hybridization network showing the hypothetical reticulation events among *Phaseolus* species. The network was generated from a 60 % BS consensus tree of ML analyses for sequences of cpDNA and ITS; reticulations derived from the chloroplast genome are indicated by green lines and reticulations from the nuclear genome by orange lines; strong lines indicate reticulations between species and fine lines indicate reticulations within species (accession names have been omitted for clarity).

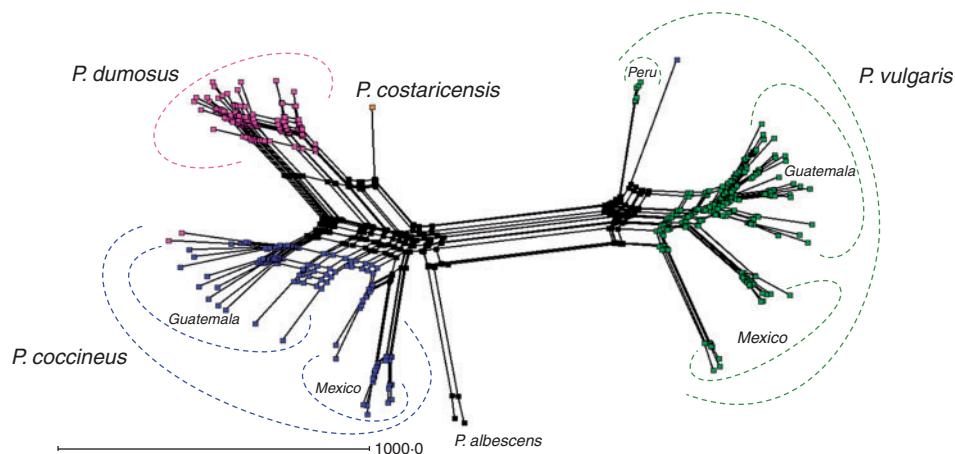


FIG. 6. Unrooted network for wild germplasm accessions of the *P. vulgaris* secondary gene pool. The network was produced from genetic dissimilarity as measured by the DArT array, applying the NeighborNet algorithm.

P. coccineus than those of any other locality. This discrepancy could have arisen if *P. dumosus* originated as a hybrid. We note that there was little evidence for introgressions between the nuclear genomes of the three species. Of the two shared cpDNA alleles, one is shared by all species of the secondary gene pool and is likely an ancestral allele retained in the conserved *rps14-psaB* locus. This allele may be associated with the split between the secondary gene pool and *P. vulgaris* itself, and may be derived from their common

ancestor. A published analysis of cpSSR (Desiderio *et al.*, 2013) also reported evidence for ancestral clusters among accessions of *P. vulgaris* and *P. coccineus* associated with the separation of these lineages. A further allele (at the *accD-psal* spacer locus) was also shared by some accessions, but is again unlikely to indicate hybridization, as it is present only in one accession of *P. coccineus* that occurs in sympatry with *P. dumosus* in Guatemala and in one of the two accessions of *P. costaricensis* from Costa Rica.

Phaseolus dumosus originated from ancient hybridizations

Hybridization has been suggested as a possible explanation for the discrepancy between the chloroplast and nuclear genome results for *P. dumosus*. Confirming that discrepancies between nuclear and cpDNA trees have arisen due to hybridization events is not easy with tree-building methods, which are designed to resolve any conflicting phylogenetic signal (Vriesendorp and Bakker, 2005; Morrison, 2011), so we instead developed a hybridization network to address this question (Figs 5 and 6). This analysis indicates that *P. dumosus* is a hybrid between an ancient lineage of *P. coccineus* (within the Guatemalan clade) and an early diverging lineage of *P. vulgaris*. The location of the split indicates that the hypothetical hybridization event occurred shortly after the split between *P. coccineus* and *P. vulgaris*, which would also contribute to the lack of diversification in the subsequent hybrid lineage. The results of the DArT PCoA analysis also show the intermediate position of *P. dumosus* relative to the two putative parents. A further suggestive finding of the DArT analysis was the presence of admixture and genetic similarity between the representatives of *P. dumosus* and *P. coccineus* from Guatemala, including from the area of Quetzaltenango associated with the Santa Maria de Jesus volcano (Figs 1–3; Table 1). Between closely related species, a higher genetic similarity and higher admixture in regions of sympatry is considered good evidence of hybridization events occurring during speciation (Seehausen, 2004; Grant et al., 2005; McKinnon, 2005). The evidence for such a hybridization event can be summarized as follows: (1) intermediate morphology between *P. coccineus* and *P. vulgaris*; (2) ancient forms of *P. dumosus* are found only in Guatemala and are in sympatry with ancient forms of its putative parental, *P. coccineus* (Schmit and Debouck, 1991); (3) artificial crosses can produce fertile offspring (Freytag and Debouck, 2002); and (4) molecular evidence indicates an incongruence between relationships observed from nuclear and chloroplast genome analyses, now reconfirmed in this study across all available accessions of wild *P. dumosus*. We conclude that the lineages of *P. coccineus* and *P. vulgaris* that hybridized to give rise to the initial line of *P. dumosus*, or closely related populations, have now been identified.

Consequences of the hybrid origin of P. dumosus

Hybrid populations are expected to have high genetic diversity because of the admixture of the parental gene pools. However, this disagrees with the low level of polymorphism identified in the wild *P. dumosus* in this study. The reticulated nature of our hybridization networks suggests this is due to backcrossing events during the evolution of *P. dumosus* (indeed *P. dumosus* groups more closely with *P. coccineus* in the DArT analyses than with *P. vulgaris*). Successive backcrossing to the pollen parent is commonly found after a hybridization event and can dilute the contribution of the maternal parental at the nuclear genome while the cytoplasmic DNA maintains its signal (McKinnon, 2005). In the case of *P. dumosus*, backcrossing may therefore have occurred to *P. coccineus* (pollen donor). In addition, the phylogenetic analysis carried out with the genome-wide DArT markers displays a conflicting signal involving *P. costaricensis*, as its exclusion reduced conflicts in the

tree and increased the bootstrap values of the *P. coccineus* node. When hybrid taxa are included in phylogenetic analysis they often generate conflict and reduce bootstrap values on the nodes of the parental taxa (Rieseberg and Soltis, 1991). This strongly suggests that *P. costaricensis* was involved in the evolution of *P. dumosus*, as proposed by Llaca et al. (1994). We consider that *P. dumosus* has a reticulate origin involving a hybridization event between *P. coccineus* and *P. vulgaris* and several backcrossing events to *P. coccineus*, and in which *P. costaricensis* was also involved. Timing of speciation events in this group are not known but the diversification of the species in the *Phaseolus* gene pool occurred within the last 2 million years and has been argued to be related to the Central American orogenies (Delgado-Salinas et al., 2006), a hypothesis that remains to be tested.

Conclusions

Molecular marker analyses support the hybrid origin theory of *P. dumosus* from *P. coccineus* and *P. vulgaris*. *Phaseolus dumosus* has relatively low diversity compared with its parental species, but not alarmingly low levels, given its restricted geographical distribution. Our results suggest that exploration for new germplasm is needed. Indeed, small stands of potentially wild populations have been identified in the eastern Chiapas and Alta Verapaz in Guatemala (Ramírez-Villegas et al., 2010). In addition, a geographically wider germplasm collecting mission should be undertaken in the localities of Sololá and Sacatepéquez, where our research detected some nucleotide-level variations from these regions.

SUPPLEMENTARY DATA

Supplementary data are available online at www.aob.oxfordjournals.org and consist of the following. Figure S1: divergences between haplotypes observed in the sequences of the (A) nrITS and (B) cpDNA. Figure S2: cpDNA alleles shared between species. Figure S3: summary of pairwise differences in nrITS sequences (A) and cpDNA sequences (B). Table S1: markers and primers used in this study. Table S2: AMOVA derived from DArT-seq. Table S3: tabulated output of STRUCTURE analysis. Table S4: allelic diversity within accessions tested. Table S5: genetic diversity estimated from sequences of the ITS and four cpDNA markers. Table S6: analysis of molecular variance (AMOVA). Table S7: genetic distances (P distances) for the species of the *Phaseolus* secondary gene pool and close congeners as calculated from nrITS and cpDNA markers.

ACKNOWLEDGEMENTS

This work was supported by an IRCSET Government of Ireland Postgraduate PhD Fellowship to A.M.M.V., under the PhD supervision of C.S.

LITERATURE CITED

Afanador LK, Haley SD, Kelly JD, Beebe S. 1993. Adoption of a “mini-prep” DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Annual Report of the Bean Improvement Cooperative* 36: 10–11.

- Akaike H.** 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. *2nd International Symposium on Information Theory*. Tsahkadzor: Armenian SSR, 267–281.
- Angioi SA, Desiderio F, Rau D, Bitocchi E, Attene G, Papa R.** 2009. Development and use of chloroplast microsatellites in *Phaseolus* spp. and other legumes. *Plant Biology* **11**: 598–612.
- Baudoin JP, Camarena F, Lobo M.** 1997. Improving *Phaseolus* genotypes for multiple cropping systems. *Euphytica* **96**: 115–123.
- Beebe SE, Rao I, Mukankusi C, Buruchara R.** 2013. Improving resource use efficiency and reducing risk of common bean production in Africa, Latin America and the Caribbean. In: Hershey C, Neate P, eds. *Eco-efficiency: from vision to reality*. Cali, Colombia: CIAT, 117–134.
- Briñez B, Blair M, Kilian A, Carbonell SAM, Chiorato AF, Rubiano LB.** 2012. A whole genome DArT assay to assess germplasm collection diversity in common beans. *Molecular Breeding* **30**: 181–193.
- Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J.** 2003. Beans (*Phaseolus* spp.) – model food legumes. *Plant and Soil* **252**: 55–128.
- Castresana J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.
- Darriba D, Taboada GL, Doallo R, Posada D.** 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772–772.
- Debouck DG.** 1999. Diversity in *Phaseolus* species in relation to the common bean. In: Singh SP, ed. *Common bean improvement in the twenty-first century*. Dordrecht: Kluwer, 25–52.
- Delgado-Salinas A, Turley T, Richman A, Lavin M.** 1999. Phylogenetic analysis of the cultivated and wild species of *Phaseolus* (Fabaceae). *Systematic Botany* **24**: 438–460.
- Delgado-Salinas A, Bibler R, Lavin M.** 2006. Phylogeny of the genus *Phaseolus* (Leguminosae): a recent diversification in an ancient landscape. *Systematic Botany* **31**: 779–791.
- Desiderio F, Bitocchi E, Bellucci E et al.** 2013. Chloroplast microsatellite diversity in *Phaseolus vulgaris*. *Frontiers in Plant Science* **3**. doi:10.3389/fpls.2012.00312.
- Earl DA.** 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.
- Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.
- Ellstrand NC, Elam DR.** 1993. Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics* **24**: 217–242.
- Evanno G, Regnaut S, Goudet J.** 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.
- Excoffier L, Lischer HEL.** 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–567.
- Excoffier L, Smouse PE, Quattro JM.** 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**: 479–491.
- Falush D, Stephens M, Pritchard JK.** 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.
- Felsenstein J.** 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368–376.
- Felsenstein J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- Freytag GF, Debouck DG.** 2002. *Taxonomy, distribution and ecology of the genus Phaseolus (Leguminosae-Papilionoideae) in North America, Mexico and Central America*. SIDA, Botanical Miscellany, No. 23. Fort Worth, TX: BRIT Press.
- Gepts P, Debouck D.** 1991. Origin, domestication, and evolution of the common bean, *Phaseolus vulgaris*. In: Voysest O, Van Schoonhoven A, eds. *Common beans: research for crop improvement*. Wallingford, UK: CABI, 7–53.
- Glazmann J, Kilian A, Upadhyaya HD, Varshney RK.** 2010. Accessing genetic diversity for crop improvement. *Current Opinion in Plant Biology* **13**: 167–173.
- Grant PR, Grant BR, Petren K.** 2005. Hybridization in the recent past. *American Naturalist* **166**: 56–67.
- Harlan JR, de Wet JM.** 1971. Toward a rational classification of cultivated plants. *Taxon* **20**: 509–517.
- Hernandez-Xolocotzi E, Miranda-Colin S, Prywer C.** 1959. El origen de *Phaseolus coccineus* L. *darwinianus* Hdz.-X. & Miranda-C subespecies nova. *Revista de la Sociedad Mexicana de Historia Natural* **20**: 99–121.
- Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.
- Huson DH, Scornavacca C.** 2012. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Systematic Biology* **61**: 1061–1067.
- Jaccoud D, Peng K, Feinstein D, Kilian A.** 2001. Diversity arrays: a solid state technology for sequence information independent genotyping. *Nucleic Acids Research* **29**: e25–e25.
- Jakobsson M, Rosenberg NA.** 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23**: 1801–1806.
- Kelchner SA, Thomas MA.** 2007. Model use in phylogenetics: nine key questions. *Trends in Ecology and Evolution* **22**: 87–94.
- Librado P, Rozas J.** 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**: 1451–1452.
- Llaca V, Delgado Salinas A, Gepts P.** 1994. Chloroplast DNA as an evolutionary marker in the *Phaseolus vulgaris* complex. *Theoretical and Applied Genetics* **88**: 646–652.
- Maddison WP, Maddison DR.** 2011. *Mesquite: a modular system for evolutionary analysis*. <http://mesquiteproject.org>.
- Mahuku GS, Jara CE, Cajiao C, Beebe S.** 2002. Sources of resistance to *Colletotrichum lindemuthianum* in the secondary gene pool of *Phaseolus vulgaris* and in crosses of primary and secondary gene pools. *Plant Disease* **86**: 1383–1387.
- Maxted N, Ford-Lloyd BV, Jury S, Kell S, Scholten M.** 2006. Towards a definition of a crop wild relative. *Biodiversity and Conservation* **15**: 2673–2685.
- McBreen K, Lockhart PJ.** 2006. Reconstructing reticulate evolutionary histories of plants. *Trends in Plant Science* **11**: 398–404.
- McCouch S, Baute GJ, Bradeen J et al.** 2013. Agriculture: feeding the future. *Nature* **499**: 23–24.
- McKinnon G.** 2005. Reticulate evolution in higher plants. In: Henry R, ed. *Plant diversity and evolution: genotypic and phenotypic variation in higher plants*. Wallingford, UK: CABI, 81–96.
- Miranda Colin S.** 1967. Origen de *Phaseolus vulgaris* L. (Frijol Común). *Agrociencia* **1**: 99–104.
- Morrison DA.** 2011. *Introduction to phylogenetic networks*. Uppsala: RJR Productions.
- Nei M.** 1987. *Molecular evolutionary genetics*. New York: Columbia University Press.
- Nei M, Kumar S.** 2000. *Molecular evolution and phylogenetics*. Oxford: Oxford University Press.
- Nei M, Li WH.** 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Science of the USA* **76**: 5269–5273.
- Peakall ROD, Smouse PE.** 2006. genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**: 288–295.
- Perrier X, Jacquemoud-Collet J.** 2006. *DARwin software*. <http://darwin.cirad.fr/>.
- Perrier X, Flori A, Bonnot F.** 2003. Data analysis methods. In: P Hamon, M Seguin, X Perrier, J Glazmann, eds. *Genetic diversity of cultivated tropical plants*. Enfield: Science Publishers, 43–76.
- Piñero D, Eguiarte L.** 1988. The origin and biosystematic status of *Phaseolus coccineus* ssp. *polyanthus*: electrophoretic evidence. *Euphytica* **37**: 199–203.
- Porch T, Beaver J, Debouck D, Jackson S, Kelly J, Dempewolf H.** 2013. Use of wild relatives and closely related species to adapt common bean to climate change. *Agronomy* **3**: 433–461.
- Posada D.** 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* **25**: 1253–1256.
- Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Ramírez-Villegas J, Khoury C, Jarvis A, Debouck DG, Guarino L.** 2010. A gap analysis methodology for collecting crop gene pools: a case study with *Phaseolus* beans. *PLoS One* **5**: e13497.
- Rieseberg L, Soltis D.** 1991. Phylogenetic consequences of cytoplasmic gene flow in plants. *Evolutionary Trends in Plants* **5**: 65–84.
- Ronquist F, Huelsenbeck J.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

- Ronquist F, Teslenko M, van der Mark P et al. 2012.** MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.
- Saitou N, Nei M. 1987.** The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**: 406–425.
- Schmit V. 1992.** Etude de *Phaseolus polyanthus* Greenman et autres taxons du complexe *Phaseolus coccineus* L. Thesis, Faculté des Sciences Agronomiques, Gembloux, Belgium.
- Schmit V, Debouck D. 1991.** Observations on the origin of *Phaseolus polyanthus* Greenman. *Economic Botany* **45**: 345–364.
- Schmit V, Munoz JE, Du Jardin P, Baudoin, JP, Debouck, DG. 2014.** Phylogenetic studies of some *Phaseolus* taxa on the basis of chloroplast DNA polymorphisms. In: Roca WM, Mayer JE, Pastor Corrales MA, Tohme MJ, eds. *2nd International Scientific Meeting of the Phaseolus Beans Advanced Biotechnology Research Network, Cali, Colombia, 1994*. Cali, Colombia: CIAT.
- Schmutz J, McClean PE, Mamidi S et al. 2014.** A reference genome for common bean and genome-wide analysis of dual domestications. *Nature Genetics* **46**: 707–713.
- Seehausen O. 2004.** Hybridization and adaptive radiation. *Trends in Ecology and Evolution* **19**: 198–207.
- Singh SP. 2001.** Broadening the genetic base of common bean cultivars. *Crop Science* **41**: 1659–1675.
- Singh SP, Gepts P, Debouck DG. 1991.** Races of common bean (*Phaseolus vulgaris*, Fabaceae). *Economic Botany* **45**: 379–396.
- Smartt J. 1973.** The possible status of *Phaseolus coccineus* L. ssp. *darwinianus* Hdz. X. et Miranda C. as a distinct species and cultigen of the genus *Phaseolus*. *Euphytica* **22**: 424–426.
- Spataro G, Tiranti B, Arcaleni P et al. 2011.** Genetic diversity and structure of a worldwide collection of *Phaseolus coccineus* L. *Theoretical and Applied Genetics* **122**: 1281–1291.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011.** MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–2739.
- Tavaré S. 1986.** Some probabilistic and statistical problems in the analysis of DNA sequences. *Lecture Notes on Mathematical Modelling in the Life Sciences* **17**: 57–86.
- Tohme J, Jones P, Beebe S, Iwanaga M. 1995.** The combined use of agroecological and characterization data to establish the CIAT *Phaseolus vulgaris* core collection. In: Hodgkin T, Brown AHC, van Hintum TJL, Morales EAV, eds. *Core collections of plant genetic resources*. Chichester: John Wiley and Sons, 95–107.
- Vriesendorp B, Bakker FT. 2005.** Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon* **54**: 593–604.
- Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. 2009.** Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.
- Wright S. 1949.** The genetical structure of populations. *Annals of Eugenics* **15**: 323–354.