Received: 8 March 2024 | Accepted: 9 April 2024

DOI: 10.1002/ajb2.16357

RESEARCH ARTICLE

Botany

Relationships of the wild peanut species, section *Arachis*: A resource for botanical classification, crop improvement, and germplasm management

Soraya C. M. Leal-Bertioli[1,2] | **Francisco J. de Blas**[1,3] | **M. Carolina Chavarro**[1] | **Charles E. Simpson**[4] | **José F.M. Valls**[5] | **Shyam P. Tallury**[6] | **Márcio C. Moretzsohn**[5] | **Adriana R. Custodio**[5] | **H. Thomas Stalker**[7] | **Guillermo Seijo**[3,8] | **David J. Bertioli**[1,9]

[1]Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA

[2]Department of Plant Pathology, University of Georgia, Athens, GA 30602, USA

[3]Botanical Institute of the Northeast (IBONE), CC 209, Corrientes W3402, Argentina

[4]Texas AgriLife Research, Texas A&M University, Stephenville, TX 76401, USA

[5]Embrapa Genetic Resources and Biotechnology, PqEB W5 Norte Final, Brasília DF 70.770-917, Brazil

[6]USDA-Agricultural Research Service, Plant Genetic Resources Conservation Unit, Griffin, GA 30223, USA

[7]Department of Crop and Soil Sciences North Carolina State University, Raleigh, NC 27695, USA

[8]Faculty of Exact and Natural Sciences, National University of Northeast, Libertad 5470, Corrientes W3402, Argentina

[9]Department of Crop and Soil Science, University of Georgia, Athens, GA 30602, USA

Correspondence

Soraya C. M. Leal-Bertioli, David J. Bertioli, Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA.
Email: sorayab@uga.edu and bertioli@uga.edu

Abstract

Premise: Wild species are strategic sources of valuable traits to be introduced into crops through hybridization. For peanut, the 33 currently described wild species in the section *Arachis* are particularly important because of their sexual compatibility with the domesticated species, *Arachis hypogaea*. Although numerous wild accessions are carefully preserved in seed banks, their morphological similarities pose challenges to routine classification.

Methods: Using a high-density array, we genotyped 272 accessions encompassing all diploid species in section *Arachis*. Detailed relationships between accessions and species were revealed through phylogenetic analyses and interpreted using the expertise of germplasm collectors and curators.

Results: Two main groups were identified: one with A genome species and the other with B, D, F, G, and K genomes. Species groupings generally showed clear boundaries. Structure within groups was informative, for instance, revealing the history of the proto-domesticate *A. stenosperma*. However, some groupings suggested multiple sibling species. Others were polyphyletic, indicating the need for taxonomic revision. Annual species were better defined than perennial ones, revealing limitations in applying classical and phylogenetic species concepts to the genus. We suggest new species assignments for several accessions.

Conclusions: Curated by germplasm collectors and curators, this analysis of species relationships lays the foundation for future species descriptions, classification of unknown accessions, and germplasm use for peanut improvement. It supports the conservation and curation of current germplasm, both critical tasks considering the threats to the genus posed by habitat loss and the current restrictions on new collections and germplasm transfer.

KEYWORDS

crop wild relatives, Fabaceae, germplasm banks, introgression, legume, peanut, phylogeny, pre-breeding, single nucleotide polymorphism (SNP), taxonomy

Wild relatives of crops are strategic resources for crop improvement (Harlan, 1976). Valuable new traits can be introduced to crops from wild species via interspecific hybridization and selective breeding (Dempewolf et al., 2017). Knowledge of the relationships of the species is key to their efficient use in breeding programs. Here we analyzed the taxonomic relationships of a comprehensive collection of over 600 germplasm samples of wild peanut (species of *Arachis* L., Fabaceae). These samples comprise the 32 diploid wild species that are most useful for crop improvement because of their higher degree of sexual compatibility with cultivated peanut (*A. hypogaea* L.). Our analysis of relationships is based on a simple insight from evolutionary biology; the branching and divergence of lineages within the evolutionary tree of life results in more closely related organisms having more similar DNA sequences. Thus, by using appropriate algorithms, we can analyze corresponding DNA data from multiple organisms to create a phylogenetic tree representing their evolutionary relationships. A cost-effective way to generate information-rich DNA data is the use of a highly parallel microarray of oligonucleotide probes that allows the simultaneous determination of thousands of specifically targeted DNA bases within multiple samples. For this study, we used such an array, designed for the genus *Arachis* (Clevenger et al., 2017; Pandey et al., 2017; Korani et al., 2019).

The South American species *A. hypogaea* was initially described by Linnaeus in 1753. In 1841, Bentham gave the genus its first expanded taxonomic treatment. (The early history and origin of peanuts was comprehensively described by Hammons [1994]). However, in the subsequent decades, the classification of the genus fell into chaos. In 1959, Antonio Krapovickas, Walton C. Gregory, and other botanists set out to remedy this situation. After almost three decades of meticulous review of the existing literature, extensive examinations of herbarium specimens, and large-scale new collections, they published the monograph "Taxonomía del género *Arachis* (Leguminosae)" (Krapovickas and Gregory, 1994), which was later translated into English by D. E. Williams and C. E. Simpson (Krapovickas and Gregory, 2007). In this monograph, the genus was subdivided into nine taxonomic sections according to morphology, cytogenetics, geographic distribution, and cross-compatibility relationships. In total, 69 species were recognized and comprehensively described. Subsequently, several new species were identified and described bringing the number of species to 83 (Valls and Simpson, 2005, 2017; Valls et al., 2013; Santana and Valls, 2015; Seijo et al., 2021). The taxonomic section *Arachis* is the subject of the present study because it contains the cultivated peanut and its secondary gene pool. The wild species in this gene pool are crossable with *A. hypogaea* and can be used for peanut crop improvement (Valls and Simpson, 2005; Krapovickas and Gregory, 2007; Seijo et al., 2021). In this group, besides *A. hypogaea*, only one other species is tetraploid (*A. monticola*, not included in this study). The diploid species are grouped in several genome types (A, B, D, F, G, and K) based on karyotypic features, distribution and patterns of heterochromatic bands and rDNA loci, and crossing data (Husted, 1936; Smartt et al., 1978; Stalker and Dalmacio, 1986; Stalker, 1991; Fernández and Krapovickas, 1994; Lavia, 1996, 1998; Peñaloza and Valls, 1997; Seijo et al., 2004; Robledo et al., 2009; Robledo and Seijo, 2010; Silvestri et al., 2015).

The monograph was a landmark for the taxonomy of the genus *Arachis*. However, the biology of the genus makes taxonomy challenging. All *Arachis* species have aerial flowers but subterranean fruit, which severely limits their seed dispersal. Species are mostly highly autogamous, and populations are mostly small and isolated (Krapovickas and Gregory, 1994), limiting gene flow. This isolation likely results in the accentuated accumulation of incompatibilities that are frequently evident even between different accessions of the same species, and speciation rates that are about six times faster than is typical for legumes (Magallón and Sanderson, 2001; Moretzsohn et al., 2013). Furthermore, the identification of *Arachis* species is complicated by the scarcity of distinctive morphological features. Often, these features appear in conjunction with other diagnostic characteristics with varying degrees of expression, and some species exhibit high intraspecific variability, adding to the complexity of their classification (Singh et al., 2004). Conversely, some species are morphologically so similar that they are considered "twin" species. For identifying such cases, observation of key plant parts at different phenological stages is needed (Stalker et al., 1995; Valls et al., 1995; Valls and Simpson, 2005; Seijo et al., 2021). Examples are *A. duranensis* vs. *A. pusilla* Benth. and *A. magna* vs. *A. monticola* Krapov. & Rigoni (Krapovickas and Gregory, 1994). These factors make species identification very challenging even for seasoned botanists.

During the period of intensive *Arachis* botanical expeditions pioneered by Krapovickas and Gregory, seeds from collections were routinely shared and distributed between different germplasm banks, safeguarding their preservation. Currently, the principal collections of wild *Arachis* species are in the Botanical Institute of the Northeast, IBONE (Argentina); Embrapa Genetic Resources and Biotechnology, EMBRAPA (Brazil); Plant Genetic Resources Conservation Unit of the U.S. Department of Agriculture, PGRCU, USDA (USA); Texas AgriLife Research Center of Texas A&M University, TAMU (USA); and the International Crops Research Institute for the Semi-Arid Tropics, ICRISAT (India). Other germplasm banks that hold wild *Arachis* accessions are the Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, OCRI-CAAS (China), National Bureau of Plant Genetic Resources, ICAR-NBPGR (India), National Institute of Agricultural Technology, INTA-Manfredi (Argentina) and North Carolina State University, NCSU (USA). Together, these collections hold all the known *Arachis* species (Williams, 2022). The primary mission of the germplasm banks is to preserve this valuable genetic resource for research and crop improvement.

Germplasm bank curation relies heavily on the accurate identification of species and accessions. Given the challenges of species identification highlighted earlier, some degree of redundancy and misidentification in collections is inevitable.

Reliable species identification is also needed for the efficient use of wild species for peanut improvement. Wild species have alleles that confer valuable traits, such as pathogen and pest resistances, that are absent in the cultigen (Stalker, 2017). Knowledge of their genetic relationships is needed for their efficient hybridization within breeding programs. For these reasons, a series of studies have been conducted to understand the taxonomic and genetic relationships among *Arachis* species, with emphasis on the section *Arachis*. These studies have used phenotypic data, cytogenetic analysis, and biochemical markers, such as isozymes and other proteins, and DNA markers such as RFLPs, RAPDs, AFLPs, microsatellites (short sequence repeats, SSRs), indels, and SNPs (Kochert et al., 1991, 1996; Lu and Pickersgill, 1993; Fernández and Krapovickas, 1994; Stalker et al., 1994, 1995; Hilu and Stalker, 1995; Jung et al., 2003; Moretzsohn et al., 2004; Nóbile et al., 2004; Peanut Crop Germplasm Committee, 2004; Seijo et al., 2004, 2007; Milla et al., 2005; Tallury et al., 2005; Bravo et al., 2006; Fávero et al. 2006; Angelici et al., 2008; Robledo and Seijo, 2008, 2010; Lavia et al., 2009; Robledo et al., 2009; Bechara et al., 2010; Friend et al., 2010; Koppolu et al., 2010; Grabiele et al., 2012; Pandey et al. 2012; Moretzsohn et al., 2013; Wang et al., 2016; Vishwakarma et al., 2017; Saritha et al., 2018; Zheng et al., 2018; Custodio et al., 2023; Ortiz et al., 2023). These studies were important but were limited to a small number of markers or few species, and only one or very few accessions representing each species. Consequently, the variation within and among many species and the larger-scale phylogenetic structure remained unclear, and the assignment of various accessions to a particular species remained ambiguous.

In this work, we used a highly parallel genotyping tool developed in the wake of the sequencing of the peanut genome (Bertioli et al., 2016, 2019; Clevenger et al., 2017; Korani et al., 2019) to perform genome-wide genetic analyses on 32 named and nine unnamed diploid species in section *Arachis*. Accessions were obtained from the main *Arachis* germplasm banks in the United States and South America, which operate under strict curation. The main objectives of this study were to improve our understanding of the phylogenetic structure of section *Arachis* and create a reference database for future classifications.

MATERIALS AND METHODS

Plant materials

In this study, we genotyped 665 samples of which 307 were unique accessions from 32 described diploid species. The

TABLE 1 Species included in this study, genome type, number of accessions attributed to each species (in the final analyses), and country of origin for accessions. Full details on each accession are in Appendix S1.

Species	No. of accessions	Origin of accessions
A genome	**197**	
A. cardenasii Krapov. & W.C. Greg.	19	Bolivia
A. chiquitana Krapov., W.C. Greg. & C.E. Simpson	2	Bolivia
A. correntina (Burkart) Krapov. & W.C. Greg.	9	Argentina, Paraguay
A. diogoi Hoehne	5	Bolivia, Brazil, Paraguay
A. duranensis Krapov. & W.C. Greg.	41	Argentina, Bolivia, Paraguay
A. helodes Mart. ex Krapov. & Rigoni	12	Brazil
A. cf. *helodes*	2	Brazil
A. herzogii Krapov., W.C. Greg. & C.E. Simpson	1	Bolivia
A. kempff-mercadoi Krapov., W.C. Greg. & C.E. Simpson	6	Bolivia
A. aff. kempff mercadoi	1	Bolivia
A. kuhlmannii Krapov. & W.C. Greg.	33	Brazil
A. aff. *kuhlmannii*	3	Brazil
A. cf. *kuhlmannii*	1	Brazil
A. linearifolia Valls, Krapov. & C.E. Simpson	1	Brazil
A. microsperma Krapov., W.C. Greg. & Valls	3	Brazil, Paraguay
A. schininii Krapov., Valls & C.E. Simpson	1	Paraguay
A. simpsonii Krapov. & W.C. Greg.	4	Bolivia, Brazil
Arachis sp. 1	3	Bolivia
Arachis sp. 2	1	Bolivia
Arachis sp. 3	3	Bolivia
Arachis sp. 5	1	Bolivia
Arachis sp. 7	1	Bolivia
Arachis sp. 9	1	Bolivia
Arachis sp. 10	1	Bolivia
A. stenosperma Krapov. & W.C. Greg.	34	Brazil
A. villosa Benth.	8	Argentina, Brazil, Uruguay

(Continues)

TABLE 1 (Continued)

Species	No. of accessions	Origin of accessions
B genome	**23**	
A. gregoryi C.E. Simpson, Krapov. & Valls	2	Bolivia, Brazil
A. inflata Seijo, Atahuachi, C.E. Simpson & Krapov.	2	Bolivia
A. ipaënsis Krapov. & W.C. Greg.	1	Bolivia
A. magna Krapov., W.C. Greg. & C.E. Simpson	10	Bolivia, Brazil
A. sp. 4	2	Bolivia
A. valida Krapov. & W.C. Greg.	5	Brazil
A. williamsii Krapov. & W.C. Greg.	1	Bolivia
D genome	**6**	
A. glandulifera Stalker	6	Bolivia, Brazil
F genome	**5**	
A. benensis Krapov., W.C. Greg. & C.E. Simpson	4	Bolivia
A. trinitensis Krapov. & W.C. Greg.	1	Bolivia
G genome	**10**	
A. decora Krapov., W.C. Greg. & Valls	6	Brazil
A. palustris Krapov., W.C. Greg. & Valls	2	Brazil
A. praecox Krapov., W.C. Greg. & Valls	2	Brazil
K genome	**22**	
A. batizocoi Krapov. & W.C. Greg.	13	Bolivia, Paraguay
A. aff. *batizocoi*	1	Bolivia
A. cruziana Krapov., W.C. Greg. & C.E. Simpson	3	Bolivia
A. krapovickasii C.E. Simpson, D.E. Williams, Valls & I.G. Vargas	4	Bolivia
A. aff. *krapovickasii*	1	Bolivia
Unknown genome	**9**	
Arachis sp. 6	1	Bolivia
A. hoehnei Krapov. & W.C. Greg.	6	Brazil
A. vallsii Krapov. & W.C. Greg.	2	Brazil

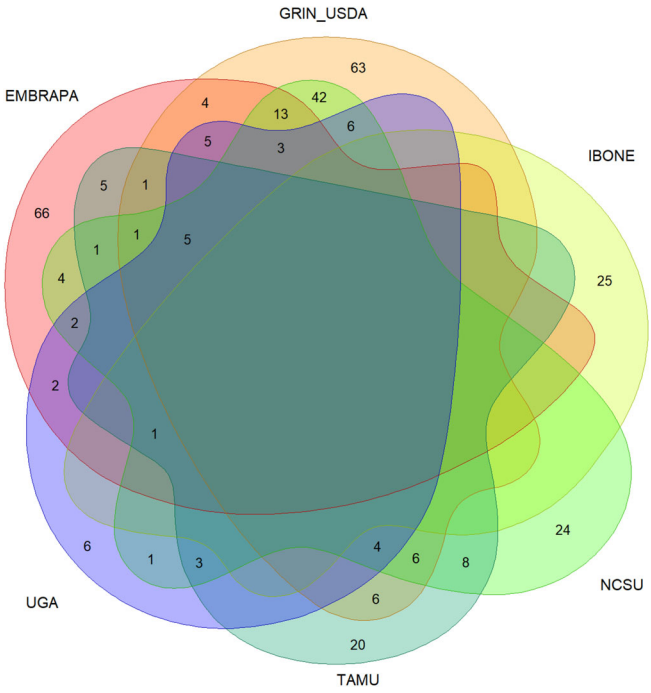

FIGURE 1 Venn diagram representing shared and unique accessions among and within germplasm banks from USDA, TAMU, UGA, EMBRAPA, and NCSU. Note that the samples processed from IBONE are all unique.

other accessions were duplicates present in different seedbanks. Additionally, eight samples had approximate identifications with "aff." or "cf.", and nine samples did not have species descriptions (aff. is short for the Latin *species affinis*, has affinity with, it is not identical to, but it may fall within the variability limits of the species; cf. is short for the Latin *confer*, compare with, and indicates that most of the diagnostic characters correspond to a given species, but some characters are unclear) (Sigovini et al., 2016). Almost all accessions were originally collected from their natural habitat and preserved in different germplasm banks: 183 in the Plant Genetic Resources Conservation Unit from the U.S. Department of Agriculture (PGRCU-USDA, Griffin, GA, USA), 62 from the Texas Agricultural Experiment Station (TAMU, Stephenville, TX, USA), 144 from Embrapa Genetic Resources and Biotechnology (Embrapa, Cenargen, Brasília, Brazil), 30 from the Botanical Institute of the Northeast (IBONE, Corrientes, Argentina), 151 from the North Carolina State University Peanut Breeding Program (NCSU, Raleigh, NC, USA), and 95 from the University of Georgia Wild Peanut Lab (UGA, Athens, GA, USA) (Table 1; Appendix S1: Figure 1).

DNA extraction and SNP genotyping

For samples from USDA, UGA, TAMU, IBONE and NCSU, young leaf tissue from greenhouse-grown plants was

lyophilized and ground using the Geno/Grinder 2010 SPEX (Cole-Parmer, Vernon Hills, IL, USA). DNA was extracted from ~120 mg tissue using the DNeasy Miniprep Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. DNA from Embrapa samples was extracted using a sorbitol-based method (Inglis et al., 2018). DNA quality was checked using a NanoDrop 1000 (Thermo Fisher Scientific, Waltham, MA, USA) and then quantified using a Qubit 4 fluorometer (Thermo Fisher Scientific). For SNP genotyping, 900 ng of DNA and the 47 K Axiom_Arachis v02 array (Thermo Fisher Scientific) and the method described by Korani et al. (2019) were used. Axiom_Arachis was designed to include 21,547 and 22,933 markers from *A. hypogaea* identified relative to the *A. duranensis* and *A. ipaënsis* genomes respectively, plus sequences of *A. duranensis*, *A. stenosperma*, and *A. cardenasii* (A genome species), *A. magna* and *A. ipaënsis* (B genome species) and *A. batizocoi* (K genome species) (Clevenger et al., 2017).

Data processing

The genotyping data were processed using custom UNIX codes and filtered using the quality control (QC) call rate of ≥90% of passing samples in the Axiom Analysis Suite (Applied Biosystems, Waltham, MA, USA) and extracted for analysis. The genotyping information was filtered allowing a minor allele frequency (MAF) > 0.05 and 20% missing calls using TASSEL 5.0 (Bradbury et al., 2007). Data were analyzed using R version 4.3.1 (R Core Team, 2022) using a combination of custom R scripts and available packages (see Appendix S2).

Reproducibility and variability

Analyses were carried out to interrogate the (1) variability of samples between germplasm banks, (2) variability between plants of the same accession originated from the same germplasm bank (to access variability within a seed storage unit), and (3) reproducibility of the same DNA preparation in the same SNP array (technical replicates). We chose seven accessions that were present in at least two germplasm banks. Six accessions were from mostly inbred species (*A. batizocoi* K 9484, *A. duranensis* V 14167, *A. ipaënsis* K 30076, *A. magna* K 30092, *A. stenosperma* V 10309, and *A. valida* K 30011) and one species suspected to have higher rate of outcrossing (*A. correntina* GKP 9530). The pairwise allele matching analyses were calculated using the percentage identity by state (pIBS) according to Singh et al. (2019).

Analysis of genetic variabilities and genotyping reproducibility

Phylogenetic and principal coordinate analyses (PCoAs) were performed to estimate the overall relationship among accessions of the *Arachis* germplasm collections investigated. The PC scores were calculated and plotted using PCAtools in R (R Core Team, 2014; Blighe and Lun, 2022) using the entire set of individuals (665) assessed and the 13,970 SNPs selected (after filtering out the MAF > 0.05). The number of PCs accounting for >80% explained variation was determined by two methods: (1) scree plot (Cattell, 1966), and (2) the cumulative sum for the number of PCs that accumulated 80% of the explained variation using the function which [cumsum(pcaobject$variance) >80] in R (R Core Team, 2014) using eigenvalues based on the Euclidean distance matrix.

A similarity matrix was computed by pairwise comparison of accessions across all SNP sites, using the percentage identity by state (pIBS) and including homozygous and heterozygous genotype calls according to Singh et al. (2019). The resulting matrix (Appendix S3) was used to generate the phylogeny trees calculated by applying a maximum likelihood model in FastTree v2.1.11 (Price et al., 2010), using covarion-autoregressive-transform (CAT) approximation to quickly estimate the phylogenetic relationships. This model accounts for heterogeneity in substitution rates among sites in the sequence alignment. For this analysis, we used 1000 times resampling without optimization of the branch lengths for the resampled alignments. The resulting tree was plotted in FigTree software v1.4.4 (Rambaut, 2016).

An initial analysis was done with all 665 samples. We then performed several iterations of data curation that involved the removal of samples that were clearly mislabeled or misidentified (see in Results) and the removal of duplicated samples. After this data curation, a phylogenetic tree was constructed with a set of 438 accessions that were considered correct but included samples that originated from different gene banks (Appendix S4). After removing duplicated accessions, the final phylogenetic tree was constructed using 272 unique accessions. This phylogenetic tree served as the basis for the discussion and the major conclusions of this article. The accessions used to construct this tree were originally collected in South America from Brazil (126), Bolivia (96), Argentina (32), Paraguay (11), Uruguay (6), and an unknown country of origin for one sample (Appendix S5). They also represent genomes A, B, D, F, G, and K (Figure 1, Table 1; Appendix S5). The places of origin can be visualized on the geographic map generated with a Google Maps layer EPSG:3857 - WGS 84/Pseudo-Mercator using QGIS 3.30's-Hertogenbosch (http://www.qgis.org) (Appendix S6). The coordinates (longitude and latitude information of the collection sites) used to plot the accessions were obtained from the PeanutBase (https://peanutbase.org/), the USDA database GRIN (https://npgsweb.ars-grin.gov/gringlobal/search), EMBRAPA database ALELO (https://alelo.cenargen.embrapa.br/), and the Botanical Institute of the Northeast (IBONE, Corrientes, Argentina) database (not public) (Figure 2).

Plots of geographical distribution of accessions of the species *A. duranensis*, *A. kuhlmannii*, and *A. stenosperma* were generated using a combination of geographical data

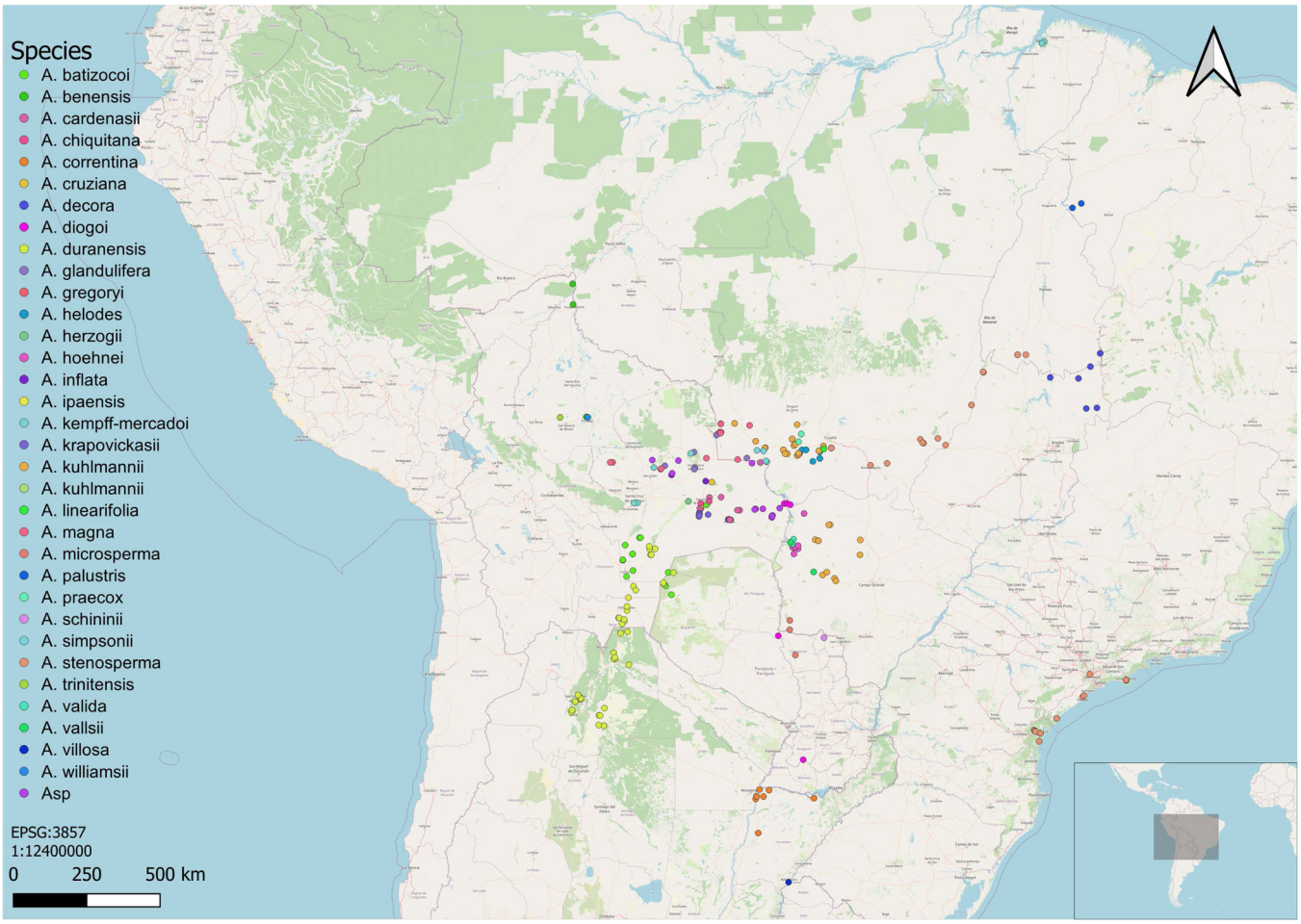

FIGURE 2 Collection location in South America for the 272 accessions used in the phylogenetic analyses. Each dot represents an accession; each color represents a different species.

analysis programs QGIS 3.30's-Hertogenbosch and graphic manipulation software GIMP 2.10.32 (GIMP Development Team, 2019). To test the correlation between the genetic affinities and geographical distances of the sample occurrence site, a Mantel test with 1000 permutations was performed using the genetic distances matrix and geographical distances matrix both calculated with Euclidean method using R package vegan 2.6-4 (Oksanen et al., 2022). All scripts used for data analyses are in Appendix S2.

Curation by specialists

The phylogenetic tree constructed without redundant accessions from different sources was analyzed and curated by the specialists involved in this work. This task combined the analysis of the genetic clustering with knowledge or verification of origin, place, and circumstances of collection, knowledge of morphology, and passport information. Some of this information was compiled from interviews and discussions with the collectors. These are transcribed in Appendix S7.

RESULTS

Analysis of reproducibility and variability

After the data filtered for quality, 98% of samples passed the quality thresholds in the Axiom Analysis Suite. After filtering out minor allele frequencies and inconsistent and missing calls, 13,970 SNPs were selected for analysis. Our reproducibility assay investigated the variability between technical replicates (same accession, same DNA preparation assayed in the same array). The average reproducibility rate was 97.98% (range: 96.98–99.26%). Furthermore, two types of biological replicates were tested: (1) The average pIBS (identity by state) between individuals from the same accession deposited in the same bank (UGA) was 97.69% (range: 91.87–99.37), and (2) accessions from different germplasm banks had an average pIBS of 93.94% (range: 86.06–98.79% average). Note that *A. correntina*, the only species thought to have a high rate of outcrossing that was tested here, had the lowest pIBS (Table 2).

TABLE 2 Pairwise allele matching analyses (calculated using the percentage identity by state [pIBS]) between accessions investigating assay reproducibility, variability in the same germplasm bank, and variability between banks.

Accession	Comparison	No. samples compared	Range pIBS (%)	Mean pIBS (%)	Mismatches (%)
A. batizocoi K 9484	**Reproducibility assay**	8	97.23–99.15	98.40	1.60
A. duranensis V 14167		4	98.49–99.26	98.88	1.12
A. ipaënsis K 30076		4	99.16–99.26	99.21	0.79
A. magna K 30092		2	N/T	98.39	1.61
A. stenosperma V 10309		4	96.98–98.14	97.56	2.44
A. valida K 30011		2	N/T	96.49	3.51
Average				**97.98**	**1.92**
A. batizocoi K 9484	**Variability within same bank**	3	97.61–98.06	97.82	2.18
A. correntina GKP 9530		3	91.87–93.94	93.13	6.87
A. duranensis V 14167		3	98.79–99.09	98.97	1.03
A. ipaënsis K 30076		3	98.25–99.37	98.70	1.30
A. magna K 30092		2	NT	98.77	1.23
A. stenosperma V 10309		3	97.49–98.37	97.91	2.09
A. valida K 30011		3	98.27–98.84	98.55	1.45
Average				**97.69**	**2.32**
Average excluding *A. correntina*				**98.45**	**1.55**
A. batizocoi K 9484	**Variability between banks**	10	91.00–95.53	94.49	5.51
A. correntina GKP 9530		6	86.06–91.37	88.14	11.86
A. duranensis V 14167		7	94.96–98.79	96.91	3.09
A. ipaënsis K 30076		7	93.01–97.66	94.94	5.06
A. magna K 30092		5	94.03–96.32	95.38	4.62
A. stenosperma V 10309		13	90.57–98.40	93.49	6.51
A. valida K 30011		4	92.63–95.07	94.21	5.79
Average				**93.94**	**6.06**
Average excluding *A. correntina*				**94.77**	**5.23**

Genetic structure

Principal coordinate (PCo) plots based on the pIBS were generated including the entire set of accessions (Figure 3). Two main groups were identified, one that included all A genome species (Group II) and one with all the other species belonging to the B, D, F, G, and K genomes and some with an unknown genome (Group I). Within this group, accessions of the different genome types tend to cluster together. A notable exception are the K genome species, which were divided into two groups, one with all accessions collected in Bolivia (*A. cruziana*, *A. krapovickasii*, and *A. batizocoi*) and one with all the *A. batizocoi* accessions collected in Paraguay. Accessions of *A. hoehnei* formed a separate group near the A genome species. Accessions of *A. vallsii* and *Arachis* sp. Ma 1469, clustered peripherally to the G genome species. Note that no tetraploid species (*A. hypogaea* or *A. monticola*) were included in this study because the two genome components of the tetraploids would confound analyses. (Tetraploid species cannot be represented as a single point in a PCoA or a nonreticulated tree.)

To investigate the genetic relationships of wild accessions of the section *Arachis*, species assignments, and genetic divergence between accessions, we constructed a phylogenetic tree using maximum likelihood estimation. An initial tree was constructed with all samples of all accessions. The clustering of all samples was a quick method to identify obvious outliers and misclassified accessions. Most replicates of the same accession obtained from different germplasm banks had a similar arrangement with the same or very similar positions in the phylogenetic tree. Different accessions from the same species also tended to cluster together. However, 30 samples had a pIBS much lower than the average for a respective group and were divergent from another sample of the same accession that had clustered as expected or clustered with accessions from different species (which we attribute to the result of mislabeling, seed mix, or cross-pollination). These samples were carefully reanalyzed, and, when identified as outliers or misidentifications, were excluded from the study (Appendices S2, S6, S7). After quality and repeatability checks and the removal of duplicates, the initial 665 samples were narrowed down to 272 unique accessions that were analyzed in a pairwise comparison across all SNP sites using pIBS. With this comparison, an identity matrix was constructed (Appendix S3) and used to generate a tree with an average support value of topology equal to 0.88. This phylogenetic analysis, called here "the curated tree", showed clustering of accessions into two main groups: Group I with all species except those of the A genome, and Group II with all A genome species (Figure 4).

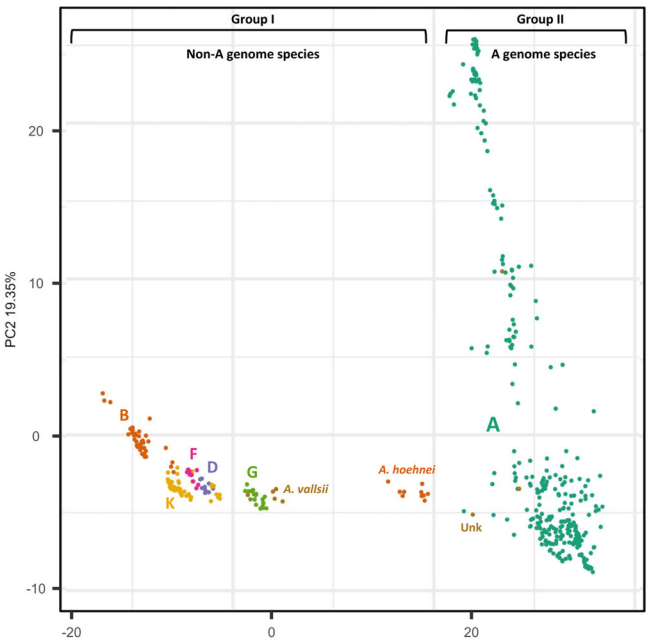

FIGURE 3 First and second principal coordinates for the diploid species from the *Arachis* section. Samples are grouped and colored according to their cytogenetically defined genome type (A, B, D, F, G, K). Note that *A. hoehnei* (unknown genome type) is near the A genome accessions and *A. vallsii* (unknown genome type) is close to G genome accessions. All accessions of genome types tended to form unique clusters, except for those with K genomes, which separated into two groups according to the country of origin.

Group I contained 75 accessions of 17 species plus three accessions not yet formally described. This Group was structured into six clades:

(1) All accessions with 18 chromosomes, G genome species: *A. decora*, *A. palustris*, and *A. praecox*.
(2) All accessions of K genome species: *A batizocoi, A. cruziana* and *A. krapovickasii*. Within the clade, each species is well separated except for *A. batizocoi*, which formed two clades: one with 10 accessions from Bolivia (plus one accession assigned as aff. *batizocoi*), and another clade with the three accessions collected in Paraguay.
(3) All accessions of the D genome species *A. glandulifera.*
(4) All accessions from the two F genome species*:* *A. trinitensis* and *A. benensis.*
(5) Accessions of all six confirmed B genome species: *A. valida*, *A. magna*, *A. williamsii*, *A. inflata*, the only known accession of *A. ipaënsis*, the B genome donor of the cultivated peanut (*A. hypogaea*), *A. gregoryi*, and two accessions of a species yet to be described.
(6) Two species of unknown genomes: *A. hoehnei* and *A. vallsii* and one accession of an undescribed species (Ma 1469).

Group II contained 186 accessions of 15 recognized A genome species, plus 11 accessions waiting for formal description. In this group were two superclades: one with *A. duranensis* and *A. cardenasii* and another with all other species. The annual species *A. duranensis*, the short-lived perennial *A. stenosperma*, and the perennial *A. cardenasii* have a very clear structure (this clade also includes the single accession of *A. herzogii*). However, the other

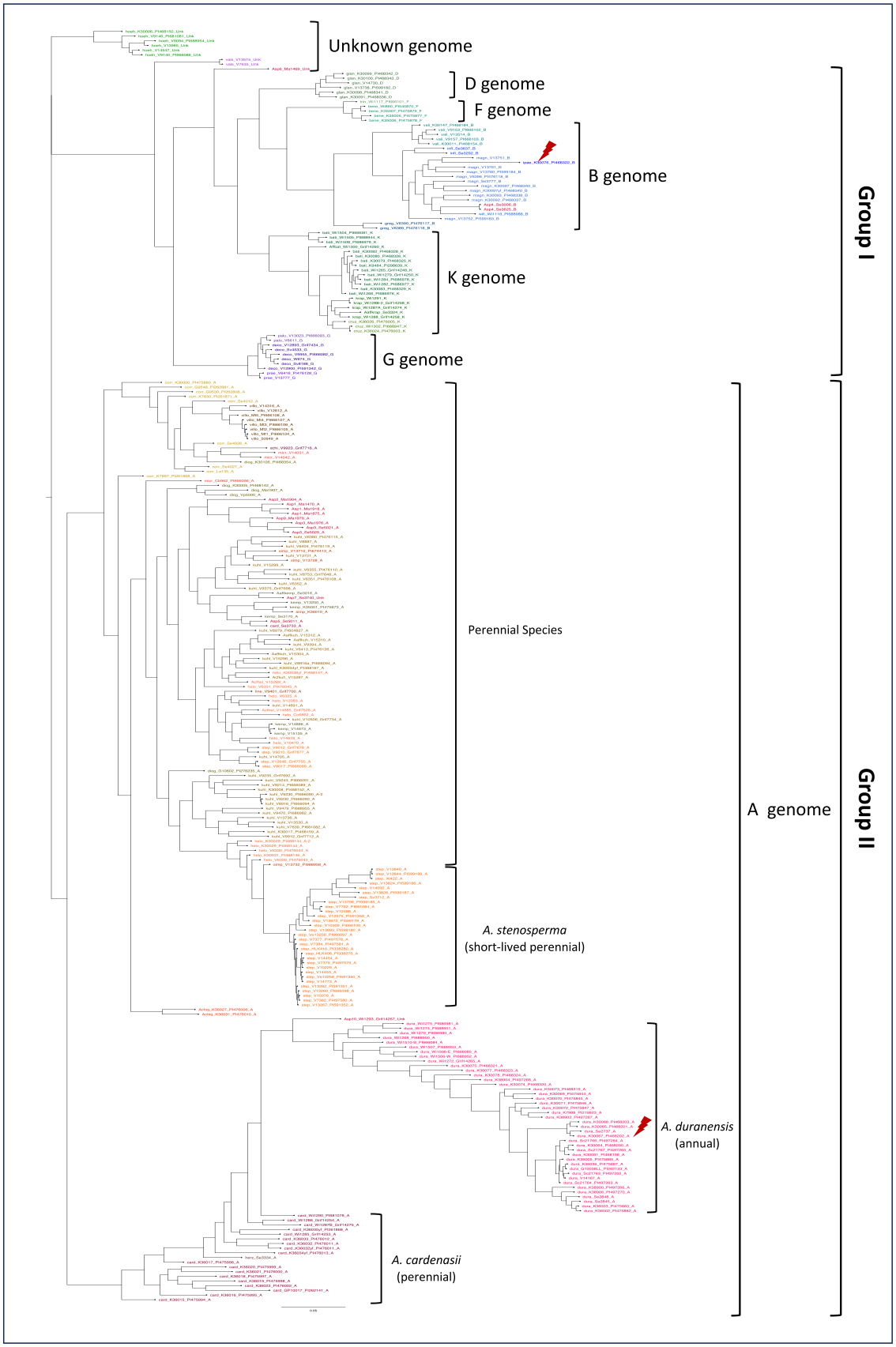

FIGURE 4 The "curated tree": phylogenetic tree of 272 accessions from 32 diploid *Arachis* species of the section *Arachis*, based on 15 K SNP genotyping calls. Accessions of the same species are represented by the same color. The red jagged arrow symbol indicates the accessions that are most likely donors of B and A peanut genomes, respectively: *A. ipaënsis* K 30076 and the *A. duranensis* group of accessions from Río Seco.

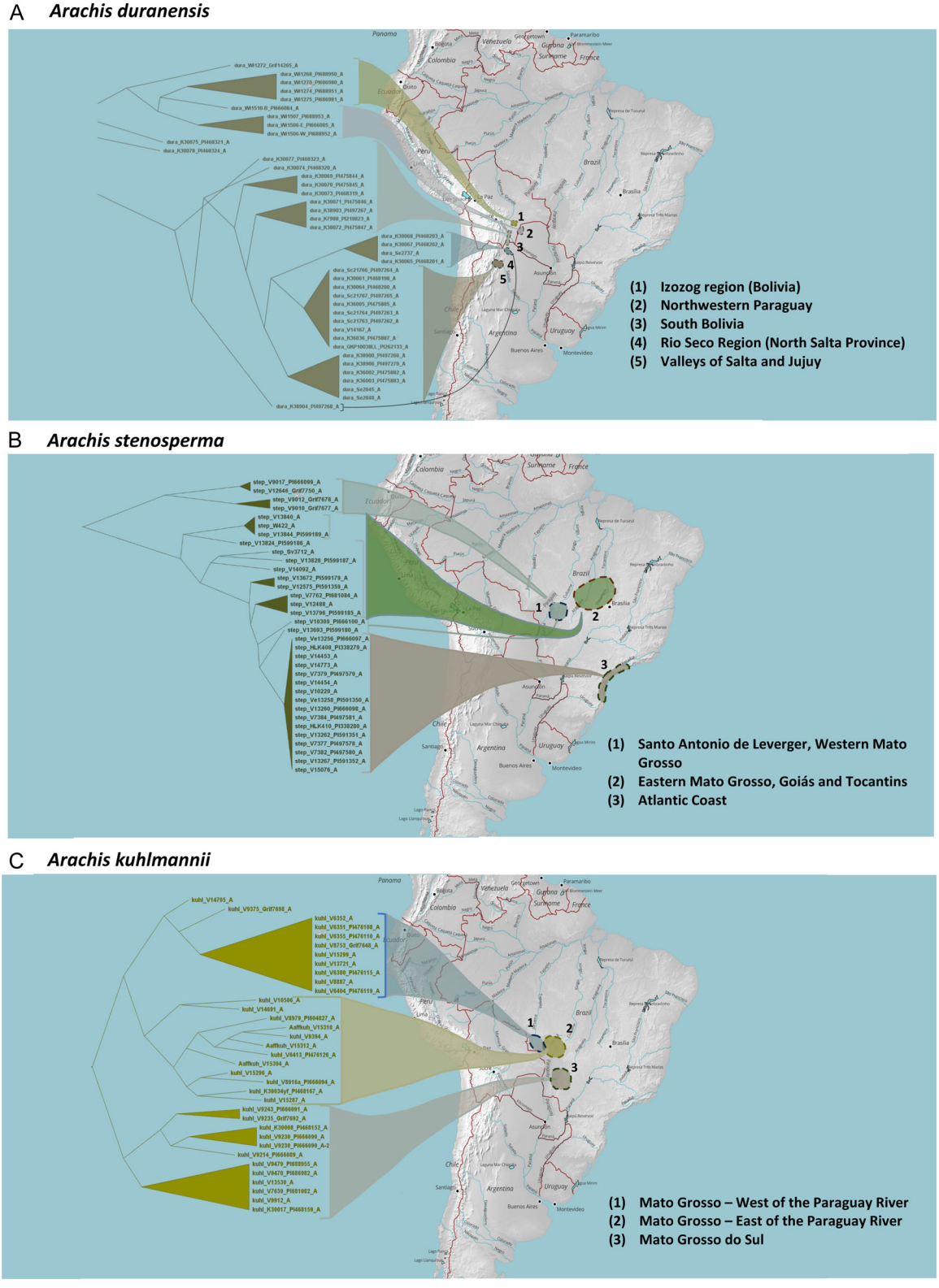

FIGURE 5 Collection sites of accessions of (A) *Arachis duranensis*, (B) *A. stenosperma*, and (C) *A. kuhlmannii* in South America and their correlation with positions on phylogenetic tree. Colors do not have any special meaning.

perennial species, are less structured, with less clear boundaries.

Accessions of *A. duranensis*, the A genome progenitor species of the cultivated species (*A. hypogaea*), group in five clusters that broadly correspond to their original geographic location: valleys of Salta and Jujuy cities, the Rio Seco region (North Salta Province), South Bolivia, northwestern Paraguay, and the Izozog region (Bolivia) (Figures 4 and 5A).

Accessions of *A. cardenasii* separated into two clades, also corresponding to the collection sites: one from Roboré and San Matias (Bolivia) and one clade with accessions from the region of San José de Chiquitos that were more closely related to the *A. duranensis* clade. The only accession of *A. herzogii*, Se 3334, also collected in San José de Chiquitos clustered with this group of *A. cardenasii*.

Arachis stenosperma, a short-lived perennial endemic to Brazil, formed a group with three clades representing the geographic distribution of all accessions previously attributed to the species. The first clade was formed by accessions collected in the westernmost area of the species' supposed distribution (Santo Antônio de Leverger, Mato Grosso), which, in the main tree, were dispersed among other species with the A genome (1); a second clade contained populations distributed mainly along the western drainage of the Araguaia River in the state of Mato Grosso and also present in its eastern drainage in Luiz Alves, state of Goiás, reaching the Pau Seco River, in Araguaçu, Tocantins (2); a third clade had accessions distributed more than 600 miles away in the Atlantic coastal region in the states of Paraná and São Paulo, including the population from which the type specimen was selected (3) (Figure 5B).

The other A genome species had a much less clear clade delimitation and will be described according to their positions in the tree, from top to bottom: Most accessions of *A. correntina* formed a large clade that had a well-defined subclade formed by all the *A. villosa* accessions, two accessions of *A. microsperma* from Brazil, one *A. schininii*, and one accession from Paraguay classified as *A. diogoi* (K 30106). Five accessions assigned to *A. diogoi* were analyzed here: Three accessions collected in adjacent regions of the Pantanal swamps in Brazil and in Bolivia formed a cluster, and the two accessions collected in Paraguay were in different clades. Accession GK 10602, which is resistant to tomato spotted wilt virus (TSWV) and has been used in various crosses with *A. hypogaea* (Milla et al., 2005; Stalker 2017; Hancock et al., 2019) fell outside two clusters of *A. helodes* and *A. kuhlmannii*.

Arachis kuhlmannii was highly heterogeneous, forming three main clusters with some accessions dispersed among some other perennial species (Figures 4 and 5C). When plotted without other species, the 37 *A. kuhlmannii* (and cf. and aff.) accessions clustered into three main groups: (1) an upper clade containing 10 accessions collected in the southwestern region of Mato Grosso state; (2) the middle clade, containing typical *A. kuhlmannii* accessions that were collected in the surroundings of Cáceres to the east of the Paraguay River and including the one representative of the natural population from where its holotype (municipality of Cáceres, state of Mato Grosso, Brazil, voucher Krapovickas & Gregory 30034, CEN) was prepared; (3) a lower clade with the 10 accessions collected in the Mato Grosso do Sul state (Figure 5C). This clustering agrees almost perfectly with the distinction of *A. kuhlmannii* accessions in three subgroups, as described by Fávero et al. (2017). The only two exceptions were V 8916 and V 13736. V 8916 was collected in the municipality of Cáceres, Mato Grosso, Brazil. In previous studies, V 8916 clustered with accessions from east of the Paraguay River (Moretzsohn et al., 2013; Fávero et al., 2017). Here, we tested two samples of the same accession: V 8916 and V 8916a. Sample V 8916a clustered as expected, but V 8916 clustered with the accessions from west of the Paraguay River. Conversely, V 13736 was originally collected west of the Paraguay River, but in our study, it clustered with accessions from east of the Paraguay River (Figures 4 and 5C). Therefore, we believe the plants identified as "V 8916" (but not V 8916a) and "V 13736" used to extract DNA for this study were wrongly identified.

The four accessions of *A. simpsonii* were scattered in different clades: *A. simpsonii* V 13732 was closely associated with the clade consisting exclusively of *A. stenosperma* accessions. *Arachis simpsonii* V 13710 and *A.* aff. *simpsonii* V 13728 were in a *A. kuhlmannii*-predominant clade, and *A. simpsonii* K 36010 was in a *A. kempff-mercadoi*-predominant clade.

The 14 accessions of *A. helodes* separated mostly into two groups. Five accessions grouped outside the clade containing *A. simpsonii* V 13732 and the *A. stenosperma* complex. The other nine, including two *A. cf. helodes* were scattered within a clade that also contained *A. kuhlmannii*, *A. stenosperma* (from the western areas in Mato Grosso), *A. kempff-mercadoi*, and the only known *A. linearifolia* accession.

The two accessions attributed to *A. chiquitana* formed a very distinct clade.

Seven species not formally described fell into specific clades within Group II: Asp1, Asp2, Asp3, and Asp9 clustered together and likely belong to three different new species. These accessions were located in a clade containing mostly *A. kuhlmannii* and *A. kempff-mercadoi* accessions. Asp5 and Asp7 were also located in this clade, but in separate subclades, both closely associated with *A. kempff-mercadoi* accessions. Asp4 clustered in the *A. magna* species complex. Asp10 formed an outgroup of the *A. duranensis* clade. Further confirmation with the morphological, cytogenetic, and crossing information should clarify these species assignment (Figure 4).

All accessions of *A. hoehnei* formed an independent cluster, and those of *A. vallsii* (and Ma 1469) formed a sister clade closely associated with Group I. Additionally, the only accession of a probable new species (Asp6_Ma 1469) was located outside the subgroup with the D, F, B, and K genome species.

The correlation between genetic structure and geographical distribution was clear when analyzing the

collection site for the 272 accessions and the clustering on the phylogenetic tree. The probabilistic Mantel test showed that the genetic distances matrix and the geographical distances matrix were significantly positively correlated ($r = 0.2$, $P = 0.001$).

DISCUSSION

Here we described the results of a much-needed detailed analysis of the phylogenetic relationships between the wild peanut species with closest affinity to *A. hypogaea*. These species in the section *Arachis* of the genus *Arachis* form a secondary gene pool for peanut improvement. Our analysis used unprecedented numbers of accessions, with representatives from multiple germplasm banks, combined with unprecedented numbers of DNA markers. We were thus able to produce a phylogeny of extraordinary detail that was interpreted with the knowledge of the scientists involved in the collection and curation of wild *Arachis*. The analysis revealed the taxonomic relationships of the species and allowed the assignment of accessions to be critically evaluated. This study is of importance for the botanical understanding of the genus and will be an invaluable guide to using the species in interspecific hybridizations to broaden the genetic basis of peanut and introduce new traits such as pathogen and pest resistance and tolerance to drought stress. In the following sections, we discuss the significance of this analysis.

Reproducibility of genotyping assays and genetic variability of accessions

To ensure the robustness of results, in the preliminary stages of this study, we checked (1) the accuracy of the genotyping method (running parallel DNA samples), (2) the homogeneity of the seeds within a seed bag, and (3) the genetic homogeneity of the same accessions between germplasm banks, when available. All these experiments were conducted within the same run of the SNP array. Assays showed that SNP calls (pIBS) were 98.14–99.38% identical when technical repetitions were tested (reproducibility), for an average of 1.15% mismatches. The SNP calls between seeds of the same accession deposited in the same bank ranged from 97.82 to 98.97% identity, an average of 1.58% mismatches—higher than the percentage of mismatches of technical repetitions. This moderately high level of mismatches indicates some genetic variability in the same pool of seeds that compose the accession. However, when comparing the same accessions in different germplasm banks, there are some substantial variations: identity among SNP calls ranged from 90.1% to 95.6%, with an average of 7.31% mismatches. This high percentage of mismatches indicates that, considering the absence of human error (seed mix or mislabeling), biological phenomena such as founder effects and genetic drift play a significant part in quite rapidly changing the genetic makeup of accessions. Additionally, in most cases of the early collections, different seeds from the field went to different seed banks. Thus, the mismatches can also reflect some variation in the original populations. The situation with *A. correntina* was even more pronounced, with only 93.0% identical SNPs between seeds within the same bank and 86.1% similarity between banks. Because this species rarely produces seeds in insect-proof greenhouses (suggesting that it is a pollinator-dependent species), one cannot rule out the possibility of genetic exchange with other accessions or species within the same greenhouse or field. The case of *A. correntina* seems to be an outlier, because for almost all species, our results indicated that the level of unexpected outcrossing, seed mix, or mislabeling is relatively low in the germplasm banks studied here. Because the percentage of SNP identity varied between accessions depending on the species, we cannot ascertain what would be the threshold to determine a species. Therefore, the species definitions/classifications of the different accessions were accepted as they are defined using classical biotaxonomic criteria, unless accessions fell outside their well-defined species clades, as discussed later.

The creation of the curated tree

We initially genotyped 665 samples from 307 unique accessions attributed to the 31 diploid species of the section *Arachis* together with *A. vallsii* (formally included in section *Procumbentes* Krapov. & W.C. Greg.). After doing iterations of phylogenetic analyses, removing wrongly labelled samples and removing duplicates, 272 accessions were used for the final phylogenetic analyses (Figure 4). This work focused on the intrageneric section *Arachis* which is of particular interest because, not only does it contain almost half of all recognized species in the genus, but mainly it contains the cultivated peanut and its secondary gene pool. Here we will make specific observations about species based on the phylogenetic analysis of the 272 unique genotypes (Figure 4), which is called "the curated tree". Accessions clustered, as expected, into two main groups, which perfectly reflected previously known cytogenetic genome types associated in two main groups: Group I with all A genome species and Group II with species of all other genome types.

The relationships of Group I accessions and species

Group I contained accessions that are annual, self-pollinated, and geographically well-defined. Within it, there is a clade of species that have symmetric karyotypes and chromosomes without centromeric heterochromatic bands that belong to the B genome sensu stricto (Robledo and Seijo, 2010; Seijo et al., 2021). This clade has all accessions of all six accepted B genome species: *A. gregoryi*, *A. valida*,

A. magna, *A. williamsii*, *A. inflata*, and *A. ipaënsis* (the B genome donor of the cultivated peanut) and two accessions yet to be identified (Asp4). All these are annual species (Krapovickas and Gregory, 1994; Valls and Simpson, 2005; Seijo et al., 2021). *Arachis magna* formed the largest, best-defined clade together with the only known accession of *A. ipaënsis*, one accession of *A. williamsii*, two accessions of *A. inflata* and two accessions of an unknown species (Asp4). *Arachis magna* has high genetic variability (Bravo et al. 2006; Moretzsohn et al. 2013), but its distribution is restricted to the border areas between Brazil and Bolivia. *Arachis magna* and *A. ipaënsis* are morphologically very similar and have complete reproductive compatibility (Moretzsohn et al., 2009). Their closeness in the tree reinforces their similarities and further supports the hypothesis that *A. ipaënsis* was transported by prehistoric humans southward from the region of *A. magna* into the range of *A. duranensis*. This human interference likely led to chance hybridization between these two species, resulting in the eventual production of *A. hypogaea*. The recently described species, *A. inflata*, is also very similar to *A. magna* and *A. ipaënsis*, but with a very distinct fruit (with a smooth but bullated epicarp and with air chambers in the mesocarp), which may be the result from the modification of one or very few genes (Seijo et al., 2021). Crosses between *A. inflata* and *A. ipaënsis* yielded progeny with only 61.3% pollen viability, confirming its separation into a new species (Seijo et al., 2021).

All accessions of the only known D genome species, *A. glandulifera*, form a separate clade (number 3) positioned between the F and K genome clades. Genome D is very asymmetric, consisting mainly of submetacentric or subtelocentric chromosomes (Stalker, 1991; Fernández and Krapovickas, 1994). Eight pairs of chromosomes have large blocks of heterochromatin. Most of these bands are centromeric in subtelocentric chromosomes and in one metacentric chromosome, while a few slightly smaller bands were observed interstitially in two pairs of chromosomes (Stalker, 1991; Robledo and Seijo, 2008; Custodio et al., 2013).

All accessions from the two F genome species, *A. trinitensis* and *A. benensis*, formed a separate clade. They have symmetric karyotypes, consisting mainly of metacentric chromosomes (Fernández and Krapovickas, 1994) with small centromeric blocks of heterochromatin in the centromeres of most chromosomes (Robledo and Seijo, 2008). The placement of these species in a separate clade (number 4) from the B genomes strongly supports the validity of the genome assignment made by Robledo and Seijo (2010).

All K genome species (*A. krapovickasii*, *A. cruziana*) and all accessions of *A. batizocoi* from Bolivia clustered together (number 2). However, four accessions of *A. batizocoi* (collected in Paraguay) formed a sister cluster, being more distant to the other *A. batizocoi* accessions than they were from the other species. *Arachis* aff. *batizocoi* Wi 1300 formed a separate branch in the tree between *A. batizocoi* from Paraguay and the clade containing all the accessions of *A. krapovickasii*, *A. cruziana*, and *A. batizocoi* from Bolivia. Although it was collected some distance from the other accessions based on geographic origin and morphology, it probably fits better as a *A. batizocoi* accession. Crossing data suggest that this accession is also very close to *A. cruziana* (C. Simpson, unpublished results). Perhaps, a detailed morphologic treatment and cross-compatibility studies of *A.* aff. *batizocoi* Wi 1300 and the accessions from Paraguay currently classified as *A. batizocoi* would reassign them to new species.

All accessions of the G genome ($2n = 18$) species (*A. decora*, *A. palustris*, and *A. praecox*) formed a unique clade inside a cluster of predominantly $2n = 20$ species. This positioning is coincident with previous molecular analyses (Tallury et al., 2005; Bechara et al., 2010; Moretzsohn et al., 2013) and corroborates the cytogenetic segregation into their G genome (Silvestri et al., 2015).

Arachis hoehnei accessions formed an independent clade outside Group I and Group II clusters, but closer to the A genome species on the PCA (Figures 3 and 4). The genome placement of *A. hoehnei* has been a subject of debate: It is morphologically more like the B genome species (Tallury et al., 2005; Mallikarjuna et al., 2006; Friend et al., 2010), and it does not have the typical "A" chromosome pair (Fernández and Krapovickas, 1994). Later, Robledo and Seijo (2010) and Custodio et al. (2013) showed it has centromeric bands in all chromosomes like an A genome species and a small chromosome that resembles the "A" chromosomes, but it does not have the large heterochromatic band and late euchromatin condensation. Using SSRs and intron-based markers, Moretzsohn et al. (2013) observed that *A. hoehnei* accessions shared greater similarity with A genome species than with the others within section *Arachis*. Based on patterns of telomeric repeat, it has even been suggested that the genome of *A. hoehnei* might be similar to the genome of species in section *Erectoides* Krapov. & W.C. Greg. genome and different from the A and B genomes (Du et al., 2016). However, supporting its greater affinity to the A genome species, our recent crosses of *A. hoehnei* with *A. magna* resulted in sterile hybrids, which, after colchicine treatment, produced some tetraploid hybrids with recovered fertility. These neotetraploid hybrids produced fertile progeny when crossed with *A. hypogaea* (Tallury et al., 2005; D. J. Bertioli et al., unpublished data). However, based on the current analyses, we suggest that *A. hoehnei* is best considered as not having any of the recognized genome types of the section *Arachis*.

The more complex relationships of Group II (A genome) accessions and species

Group II includes all A genome species, which include annuals and perennials. Here, *A. duranensis*, *A. cardenasii*, and *A. stenosperma* are the species with the best delimited genetic structure.

Arachis duranensis

This species is the best represented in this study, with 41 accessions. Their arrangement in the phylogeny mostly corresponds to their geographic origin from south to north along the eastern edge of Los Andes mountains. The *A. duranensis* from Rio Seco (K 30065, K 30067 and Se 2737) are considered the most likely progenitors of the A genome of *A. hypogaea* (Grabiele et al., 2012; Bertioli et al., 2016, 2019). In the monograph (Krapovickas and Gregory, 1994, 2007), all groups of *A. duranensis* accessions were combined into a single species. Previously, because of their geographic distribution and morphological variability, these groups were considered three species: *A. duranensis*, *A. argentinensis* nomen nudum, and *A. spegazzinii* nomen nudum (Krapovickas and Gregory, 1994). The genetic divergence and grouping of *A. duranensis* observed here adds additional support for previous classifications of this group as three species.

Arachis cardenasii

Arachis cardenasii was divided in two different clades corresponding to the collection sites. One has accessions from Roboré, and the other has accessions from around San José de Chiquitos (both in northern Chaco, Bolivia). The sole specimen of *A. herzogii*, Se 3334, falls marginally within the Roboré clade. This accession is morphologically very similar to *A. cardenasii* and only differs from this species by the presence of long hairs on the abaxial side of the leaves. Their populations are intermixed, and they may be derived from a single evolutionary lineage.

The curious case of *A. stenosperma*

Arachis stenosperma is an endemic species of Brazil collected in two main areas: one inland (states of Mato Grosso, Goiás and Tocantins) and one in the Atlantic coastal area (and several islands), separated by over 1200 km and with no collection sites in between these areas. Therefore, the species was likely dispersed by Amerindians who transported it as a food source from the central part of Brazil to the coast (Valls, 1996, 2000; Singh et al., 2004; Custodio et al., 2005), creating new populations on the coast and adjacent islands. Populations can still be found in association with archeological sites (Valls, 1996, 2000), but unfortunately, the populations are rapidly disappearing due to urban growth.

While the bulk of accessions of *A. stenosperma* fell in a well-delineated clade, four collected in Santo Antônio do Leverger, Mato Grosso (V 9010, V 9012, V 9017 and V 12646) were in a different clade (Figures 4 and 5B). They had closer genetic affinity to *A. helodes*, *A. kempff-mercadoi*, and *A. kuhlmannii* collected nearby. Their identification as *A. stenosperma* seems to be incorrect. Further investigation is needed to clarify their relationship with the previously collected accession GKP 9901, also from the eastern region of Cuiabá, which was previously associated with uncertainties about *A. diogoi* in the 1994 monograph.

A. microsperma, A. villosa, and *A. correntina*

These three very morphologically similar species constituted a clade, except for *A. microsperma* accession Cb 962_PI666096 collected 130 km away in San Alfredo, Paraguay. This accession may require further studies to confirm the species assignment. The only very clear difference between *A. villosa* and *A. correntina* is that the fruit of *A. villosa* is reticulate and that of *A. correntina* is smooth. The fruit of *A. microsperma* is similar to that of *A. villosa* but smaller. *Arachis microsperma* and *A. villosa* also have an allopatric distribution. An important feature of *A. correntina* is that it does not easily self-pollinate in the absence of insects. Correspondingly, the average pIBS between individuals of the accession GKP 9530 of *A. correntina* was the lowest of the species analyzed. This result suggests that *A. correntina* conserved ex situ has a high degree of allogamy and may be prone to cross-pollination with other accessions or species, which may pose difficulty in maintaining this species in a germplasm bank and care must be taken to maintain its genetic purity. The only accession of *Arachis schininii* (V 9923) included here fit in this well-resolved cluster.

The perennial species *A. kuhlmannii, A. diogoi, A. helodes, A. kempff-mercadoi*, and *A. simpsonii*

These species were scattered along the tree, forming a highly diverse group. It is possible that they underwent cryptic speciation, resulting in a group of species that are morphologically very similar. Another possibility is gene flow through cross-pollination (Appendix S7). These hypotheses need further testing, for instance, by controlled hybridizations between accessions.

Arachis diogoi

Of the five accessions analyzed, three accessions that were collected in nearby areas of the Pantanal swamps in Brazil and Bolivia formed a cluster. The other two, collected in Paraguay, in regions separated by over 700 km, fell in very different clades. K 30106 was placed within the *A. schininii-A. microsperma-A. correntina* group, and GK 10602 is an outgroup of the *A. kuhlmannii-A. helodes-A. stenosperma* group. *Arachis diogoi* K 30106 was located in a clade close to *A correntina and A. schininii* accessions in the curated tree. This accession was initially classified as *A. diogoi* (Krapovickas and Gregory, 1994) due to its morphological characters and geographic location near the Tebicuary River in Paraguay, but its identification should be revised. GK 10602 was previously classified as *A.* "*chacoense*"

(an inappropriate neutral epithet, never validly published) and has been used for peanut improvement, as a donor for resistance to groundnut rosette virus, tomato spotted wilt virus and leaf spots (Moss et al., 1998; Milla et al., 2005; Okello et al., 2018, 2021; Hancock et al., 2019). *Arachis diogoi*, described by Hoehne in 1919, typically found in areas subject to inundation, is a well-established species from the morphological standpoint, with relatively narrow leaflets. However, the species circumscription in the monograph is too wide and includes accessions subsequently assigned to other species.

Arachis kuhlmannii

Arachis kuhlmannii, as noted above, is a highly heterogenous species (Moretzsohn et al., 2013; Fávero et al., 2017). In this study, *A. kuhlmannii* accessions are grouped into three different clades clearly associated with geographic origin: one clade with accessions collected in the southwestern region of Mato Grosso to the west of the Paraguay River; one clade from accessions collected in the surroundings of Cáceres, in the eastern region of Mato Grosso, east of the Paraguay River; and one clade from Mato Grosso do Sul (Figure 4B). This clustering in the tree, which was reported in previous studies, suggests a polyphyletic origin for this species (Milla et al., 2005; Bechara et al., 2010; Koppolu et al., 2010; Fávero et al., 2017). Different from most accessions from Mato Grosso, accessions collected in Mato Grosso do Sul are all clearly confined to natural sites in the Pantanal that flood periodically. The clade containing accessions from Cáceres to the east, where the holotype (K 30034) was collected, represents the type specimen of *A. kuhlmannii*. The other two clades require taxonomic revision.

Arachis simpsonii

The four accessions of this species were in a different clade, two with *A. kuhlmannii* accessions, one with *A. kempff-mercadoi*, and one with *A. helodes* accessions. The *A. simpsonii* accessions seem to lack a clear species delimitation indicating the need for further study.

Arachis chiquitana

The taxonomy of this species has been the subject of much confusion. Originally described as a member of the section *Procumbentes*, its accession K 36025 (ICG 11560) was reported to have crossed with *A. hypogaea*, setting a large number of seeds that did not reach maturity, so embryo culture was needed to generate viable plants. In contrast, in most crosses of *Procumbentes* species with *A. hypogaea*, embryos were aborted at an early stage (Mallikarjuna, 2005). However, accession K 36025 did not group based on SSR markers with the conspecific K 36028 (ICG 13181) and 36031 (ICG 13241) also available at ICRISAT (Mallikarjuna et al., 2007). Present results indicate that accessions K 36027 (sifted as seeds at the collection site of the herbarium specimens subsequently selected as the holotype and isotypes) and K 36031, available as germplasm at Embrapa, ICRISAT, USDA, IBONE and other institutions, under these identification numbers, belong to a species in section *Arachis*. Previous cytogenetic studies indicated that plants derived from these seeds had chromosome characteristics of section *Arachis* (Lavia, 2000; Lavia et al., 2009; Robledo et al., 2009). Therefore, after years of confusion, this study confirms that seed samples attributed to *A. chiquitana* do belong in section *Arachis*. Further investigation of the holotype of this species should help resolve how this confusion arose.

Suggestions for reassignments of accessions and new species

Based on the traditional/morphological traits, collectors were not able to assign some accessions to an existing species. They were then classified either as aff. or cf. Additionally, some accessions were simply labeled as *Arachis* species (*A.* sp.). Based on our comprehensive analyses, we suggest the following reassignments:

Arachis cf. *helodes* V 14685 is *A. helodes*.

Arachis aff. *krapovickasii* Se 3324 is *A. krapovickasii*.

Arachis aff. *kuhlmannii* V 15304, V 15310 and V 15312 grouped in a clade with V 6413 and V 9394. They can probably be confirmed as *A. kuhlmannii*.

A. helodes V 6330 was originally described as *A. diogoi* in the monograph. Similarly, K 30029 (PI 476044) is identified in GRIN as *A. diogoi* (Appendix S7) despite being included in the monograph within *A. helodes*. This study confirms that both accessions are in fact *A. helodes*.

Arachis sp. Wi 1293 (Grif 14267) is an "in-betweener," geographically flanked to the north by *A. cardenasii*, *A. herzogii*, *A.* aff. *duranensis* and, to the south, by *A. duranensis*. Thus, its taxonomic status remains unresolved.

Arachis aff. *cardenasii* Se 3733 is likely not *A. cardenasii* based on its position in the tree, but it may be a glabrous type of *A. kempff-mercadoi* or belong to the yet-undescribed Asp5.

Arachis aff. *batizocoi* Wi 1300 (Grif 14260) is likely a new species.*Arachis* aff. *kempff-mercadoi* Se 3016 seems to be a new species similar to Asp7.

Arachis diogoi K 30106 might need to be reclassified as a different species, probably as *A. correntina*.

A. stenosperma V 9010, 9012, 9017 and V 12646 might need to be reclassified as a different species. Further experimental hybridizations are needed to confirm this.

Arachis vallsii

The genome type of this species is unknown, and specialists have not reached a consensus on its placement in the

Arachis section. In their monograph, Krapovickas and Gregory (1994) initially described it as belonging to section *Procumbentes*. Because of its cytogenetic traits, Lavia (2001) suggested it should be removed from section *Procumbentes* and placed in the section *Arachis* (Lavia et al., 2009). This proposal was corroborated by Moretzsohn et al. (2013), based on the results obtained using microsatellite and intron-based markers. However, cytogenetically, its genome type is still not defined, and its monotypic karyotype based on CMA and DAPI bands and 5 S and 18S–26 S rDNA loci shows chromosome features that differ from those of the species in section *Procumbentes* (Silvestri et al., 2020). *Arachis vallsii* may be considered a "bridge" species, useful for moving alleles from accessions that belong to sections *Caulorrhizae* Krapov. & W.C. Greg., *Procumbentes*, *Erectoides*, and *Heteranthae* Krapov. & W.C. Greg. due to its cross-compatibility with species belonging to those sections. For instance, successful crosses were performed between *A dardanii* Krapov. & W.C. Greg. and *A. vallsii* in an attempt to introgress alleles for drought tolerance into peanut (Cason et al., 2019; Simpson et al., 2019). Although it crosses with many other species from the *Procumbentes* and *Erectoides* sections with variable degrees of progeny viability, the most successful crosses are with species in section *Arachis* (Simpson et al., 2019; J. F. M. Valls, unpublished data). Arguments can be made for placing this species in section *Arachis*, in a different section, or even in a new section. Here, in the PCA and on the phylogenetic tree, all accessions of *A. vallsii* were nearer the G, D, F, B, and K genome species (Figures 3 and 4). Although additional analyses are needed to define the best placement for this species, current morphological, genomic, and crossing data point to it belonging to section *Arachis*.

Species concepts for genus *Arachis*

The differentiation and species definitions for genus *Arachis* rely on very few morphologically significant characters. Some are underground structures (fruits, rhizomatous stems, root systems, and hypocotyls), while others are more visible aerial traits (flower and leaf morphology, indumentum, and plant architecture). These characters, in conjunction with cross-compatibility and plant life cycle (annual, perennial), are classically used to separate species into sections and distinguish the species within them. Cytological analyses further differentiated genome types within the section *Arachis*. Although the collection site is not formally used for defining a species, the place of origin and the identity of other species growing in the area are valuable pieces of information that can be used for resolving placement of a particular species. Sometimes relatively distant species in different sections (e.g., *A. pusilla* and *A. duranensis*) or even differing in ploidy (*A. magna* and *A. monticola*) are hard to distinguish based on gross morphology (these are referred to as twin species in the monograph [Krapovickas and Gregory, 1994]). At the same time, high levels of genetic variability can be observed within the same species: Intraspecific crosses, that could be expected to generate hybrid progeny with high fertility, can generate hybrids with low fertility (Gregory and Gregory, 1979; Stalker et al., 1991). Karyotypic variation has also been found between accessions of the same species (Stalker et al., 1991, 1995; Custodio et al., 2013). The autogamous reproductive system, underground fruiting habit, and limited means of seed dispersal cause rapid population isolation, leading to restricted gene flow. This isolation may cause drift in chromosomal organization and Dobzhansky-Muller factors leading to infertility barriers between species or even accessions of the same species.

In their monograph, Krapovickas and Gregory (2007, p. 26) wrote "To define or delimit the species one must reconcile or harmonize the exomorphology with all the other information, such as genetic, chromosomal or geographic, because none by itself would suffice in absolute terms. We rarely observed extensive populations; what is usually seen is the formation of small populations due to their adaptation to special, primarily sandy, soil types. In this manner, the species of *Arachis* behave as apomictic species despite their sexuality, due to a gene flow that is restricted within small populations and almost non-existent between populations separated from one another." Because of the complex definition of a species, the high degree of intraspecific variability, the low number of morphological diagnostic characters (that sometimes overlap between defined species), the dispersal mechanisms, reproductive strategies, and the high rate of speciation, some individual *Arachis* accessions cannot be separated into well-defined clades or be grouped with different species from similar geographical locations as observed in the phylogenetic tree.

Analyses of accession identities and genetic diversity

The work described here reconfirmed the existing taxonomy of most species and accessions in different germplasm banks, helped position some accessions of unassigned species, raised doubts about the placement of a few accessions and the delimitation of other species. Our work also raises questions about the appropriate species concept for the A genome species. This study gave us a unique opportunity to compare accessions between germplasm banks. Most samples from different germplasm sources clustered together or were positioned closely in the phylogenetic analysis and PCA, suggesting that the species identities have been properly maintained over the years. Most replicates of the same accession obtained from different banks had the same or very similar positions in the phylogenetic tree. Some heterogeneity was found between seeds of the same accession in the same germplasm bank (Table 2) and is a major challenge for managing ex situ germplasm banks (Richards and Volk, 2010; Barkley et al., 2016). Only a limited number of studies have

investigated the variability within individual wild peanut accessions. Nonetheless, some accessions exhibiting mixed morphological traits, mainly related to flower color, presence or absence of anthocyanin, or the presence and distribution of bristles or their absence, have been observed and duly documented in gene-bank records or catalogs. In a few instances, to safeguard distinct morphological differences, these accessions were either noted as such or separated into distinct accessions (e.g., *A. magna* V 13761 with orange flowers, from which V 13761-y with yellow flowers and V 13761-c with cream flowers were segregated, or *A. stenosperma* V 7762 with yellow flowers, from which plants with orange flowers were isolated as V 7764). Different accessions from the same species also tended to cluster together. Only 30 samples were excluded because they had a pIBS much lower than the average for a particular group, were divergent from another sample of the same accession that had clustered as expected, or had clustered with accessions from different species (probably the result of mislabeling, seed mix, or cross-pollination). The number is small compared to the total number of accessions analyzed, exemplifying the care of the curators to maintain the integrity of the collections for decades. The difficulty in identifying species makes it challenging to detect occasional labeling error or seed mixture that, once they happen, tend to be propagated unnoticed. The repeatability assay allowed us to select a valuable set of informative SNPs for the classification of the accessions comparing different replicates of the samples and different runs of the chip. It also allowed us to identify identical accessions inside and between germplasm banks in an accurate and cost-effective way. This is the first time a genetic diversity analysis in wild relatives of peanut includes the confirmation of markers and accessions by controlling the variability of the data with multiple samples per plant, samples per accession, and accessions from different sources. With the database created by this study, genotyping has become a powerful tool to periodically check the identification of accessions and to detect potential mislabeling.

Arachis germplasm banks

Arachis germplasm banks are invaluable assets. As cities and farmlands grow, ex situ conservation becomes one of the more secure options to preserve the genetic resource of wild *Arachis* species. Ex situ conservation is also a prerequisite for many types of scientific study, such as this one, for accumulating scientific data and improving the knowledge base to incorporate wild germplasm into breeding programs. While numerous populations of *Arachis* remain unrepresented in germplasm banks, new collections have been hindered by national legislations that followed the Cartagena and the Nagoya Protocols of the Convention on Biological Diversity (Secretariat of the Convention on Biological Diversity, 2000, 2011). In the past, germplasm resources were considered the common heritage of all people. Under this legal framework, extensive collection and research based on the free and open exchange of germplasm could be carried out. Accessions collected in any of the countries where *Arachis* is native (Argentina, Bolivia, Brazil, Paraguay, and Uruguay) would be divided between the collectors and stored in different germplasm banks, ensuring the preservation of these materials at multiple sites. Once the Convention entered into force on 29 December 1993, this collaborative framework was undercut by almost universal restrictions on the collecting, use, and exchange of germplasm, including wild species in section *Arachis* (Williams and Williams, 2001; Williams, 2022). National laws, the Andean Pact, and the Nagoya Protocol that followed in the wake of the Convention have added extra layers of restrictions and complexity. Adding to the existing challenges, both peanuts and their wild relatives are absent from Annex 1 of the International Treaty on Plant Genetic Resources for Food and Agriculture (https://www.cbd.int/doc/treaties/agro-pgr-fao-en.pdf), which designates crops for facilitated access to agricultural genetic resources through its multilateral system of access and benefit sharing (MLS). These restrictions created a substantial barrier to the international exchange of peanut genetic resources. It fundamentally hampers the development and eventual implementation of a global strategy for preserving peanut diversity. Currently, most peanut gene banks across the world remain inaccessible to foreign parties (Williams, 2022). At present, as far as we are aware, the only germplasm bank that maintains a policy of international germplasm exchange unhindered by highly restrictive intellectual property rights is the U.S. National Plant Germplasm System, hosted by the U.S. Department of Agriculture (USDA).

Accessions present in germplasm banks are currently irreplaceable, and there is a need for their traceability and precise curation—all of which can be helped by genotypic characterization. It is of note that the *Arachis* scientific community has paid special attention and kept exceptional documentation of *Arachis* germplasm, where practically every accession available in the main germplasm banks has a voucher herbarium specimen (with duplicates deposited in different herbaria in South and North America and Europe) and can be traced to the original field collection author/collector and number. Such original collection vouchers and numbers make it possible to trace the accession's location, origin, travels, and uses and facilitates subsequent re-identifications when needed. This attention to and respect for these original collection numbers makes it possible to trace accessions in such a comprehensive way.

Our phylogenetic analysis reported here is of unprecedented scope and resolution and clarifies the relationships of most of *Arachis* accessions and species and their taxonomic structures. In their natural habitats, *Arachis* species face considerable threats, with rapid declines in their habitats and populations caused by agricultural and cattle ranching expansion, urbanization, and climate change (Jarvis et al., 2003). Given the current restrictions on new

collections and cross-border germplasm transfer, preserving the current ex situ germplasm is crucial. Our analysis serves as a foundation for future germplasm curation, confirmation of accessions, identification of new collections and species, and the optimization of germplasm utilization. We hope that it will help ensure the protection and efficient use of this valuable legacy for future generations.

AUTHOR CONTRIBUTIONS

Conceptualization: S.C.M.L.-B and D.J.B.; genotyping, formal analyses and visualization: F.J.B., M.C.C.; data curation: C.E.S., J.F.M.V., S.P.T., H.T.S., G.S., S.C.M.L.-B, D.J.B.; resources: C.E.S., J.F.M.V., S.P.T., H.T.S., G.S.; DNA extraction and genotyping: M.C.M., A.R.C.; writing original draft: S.C.M.L.-B; review and editing: S.C.M.L.-B, M.C.M., J.F.M.V., S.P.T., H.T.S., G.S., D.J.B.

ACKNOWLEDGMENTS

The authors thank Steven Cannon for helpful discussions; Jenny Leverett, Alora Colvin, Nicole Bachleda, Carlos Ruiz, Kennedi Smith, Peter Inglis and Laura Pérez for extracting DNA; and Mark Hopkins for greenhouse assistance. This work was supported by Agriculture and Food Research Initiative Competitive Grant no. 2022-67013-37075; grants from the National Science Foundation (no. MCB-1543922), Georgia Peanut Commission, Peanut Research Foundation, and CNPq-Brazil grants 308661/2022-2 (M.C. Moretzsohn) and 310766/2022-2 (J.F.M. Valls), CONICET and ANPCyt Argentinian fellowship (F. de Blas) and grant (FONCYT 3664G. Seijo). The authors thank the reviewers (Dr. D. E. Williams and anonymous) for their suggestions and thoughtful edits to the manuscript.

DATA AVAILABILITY STATEMENT

All data sets are available in the appendices. Original genotyping files are available upon request from the corresponding author.

ORCID

Soraya C. M. Leal-Bertioli http://orcid.org/0000-0002-9683-5357
Francisco J. de Blas http://orcid.org/0000-0003-4500-3072
M. Carolina Chavarro http://orcid.org/0000-0003-4654-7143
Charles E. Simpson http://orcid.org/0000-0002-4579-7370
José F.M. Valls http://orcid.org/0000-0002-4586-5142
Shyam P. Tallury http://orcid.org/0000-0001-6102-9008
Márcio C. Moretzsohn http://orcid.org/0000-0003-1708-1508
Adriana R. Custodio http://orcid.org/0000-0002-7415-8754
Guillermo Seijo http://orcid.org/0000-0001-8699-2044
David J. Bertioli http://orcid.org/0000-0003-0294-7284

REFERENCES

Angelici, C. M. L. C. D., A. A. Hoshino, P. M. Nobile, D. A. Palmieri, J. F. M. Valls, M. A. Gimenes, and C. R. Lopes. 2008. Genetic diversity in section *Rhizomatosae* of the genus *Arachis* (Fabaceae) based on microsatellite markers. *Genetics and Molecular Biology* 31: 79-88.

Barkley, N. A., H. D. Upadhyaya, B. Liao, and C. C. Holbrook. 2016. Global resources of genetic diversity in peanut. *In* H. T. Stalker, and R. F. Wilson [eds.], Peanuts: genetics, processing, and utilization, 67–109. AOCS Press, Champaign, IL, USA.

Bechara, M. D., M. C. Moretzsohn, D. Palmieri, J. Monteiro, M. Bacci Jr., J. Martins Jr., J. F. M. Valls, et al. 2010. Phylogenetic relationships in genus *Arachis* based on ITS and 5.8S rDNA sequences. *BMC Plant Biology* 10: 255.

Bertioli, D. J., S. B. Cannon, L. Froenicke, G. Huang, A. D. Farmer, E. K. S. Cannon, X. Liu et al. 2016. The genome sequences of *Arachis duranensis* and *Arachis ipaënsis*, the diploid ancestors of cultivated peanut. *Nature Genetics* 48: 438-446.

Bertioli, D. J., J. Jenkins, J. Clevenger, O. Dudchenko, D. Gao, G. Seijo, S. C. M. Leal-Bertioli et al. 2019. The genome sequence of segmental allotetraploid peanut *Arachis hypogaea*. *Nature Genetics* 51: 877-884.

Blighe, K., and A. Lun. 2022. PCAtools: everything principal component analysis. Website: https://github.com/kevinblighe/PCAtools

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss and E. S. Buckler. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633-2635.

Bravo, J. P., A. A. Hoshino, C. M. L. C. D. Angelici, C. R. Lopes and M. A. Gimenes, 2006. Transferability and use of microsatellite markers for the genetic analysis of the germplasm of some *Arachis* section species of the genus *Arachis*. *Genetics and Molecular Biology* 29: 516-524.

Cason, J. M., C. E. Simspon, W. L. Rooney and J. A. Brady. 2019. First report of successful hybridization between *Arachis vallsii* and *Arachis dardani*. *Agrosystems, Geosciences & Environment* 2: 190013.

Cattell, R. B. 1966. The scree test for the number of factors. *Multivariate Behavioral Research* 1: 245-276.

Clevenger J., Y. Chu, C. Chavarro, G. Agarwal, D. J. Bertioli, S. C. M. Leal-Bertioli, M. K. Pandey, et al. 2017. Genome-wide SNP genotyping resolves signatures of selection and tetrasomic recombination in peanut. *Molecular Plant* 10: 309-322.

Custodio, A. R., A. P. S. Peñaloza and J. F. M. Valls. 2005. Further cytogenetic information on *Arachis stenosperma* (Leguminosae). *Cytologia* 70: 331-335.

Custodio, A. R., A. B. Schmidt, M. C. Moretzsohn and J. F. M. Valls. 2023. Genetic relationships of *Arachis* (Fabaceae) accessions based on microsatellite markers. *Genetic Resources and Crop Evolution*. 70: 1879-1893.

Custodio, A. R., G. J. Seijo, and J. F. M. Valls. 2013. Characterization of Brazilian accessions of wild *Arachis* species of section *Arachis* (Fabaceae) using heterochromatin detection and fluorescence in situ hybridization (FISH). *Genetics and Molecular Biology* 36: 364-370.

Dempewolf, H., G. Baute, J. Anderson, B. Kilian, C. Smith and L. Guarino. 2017. Past and future use of wild relatives in crop breeding. *Crop Science* 57: 1070-1082

Du, P., L.-n. Li, Z.-x. Zhang, H. Liu, L. Qin, B.-Y. Huang, W.-Z. Dong, et al. 2016. Chromosome painting of telomeric repeats reveals new evidence for genome evolution in peanut. *Journal of Integrative Agriculture* 15: 2488-2496.

Fávero, A. P., A. P. S. Peñaloza, C. R. Lopes, and J. F. M. Valls. 2017. Morphological, cytogenetic, and molecular characterization of *Arachis kuhlmannii* Krapov. & W.C. Greg. (Leguminosae). *Genetic and Molecular Research* 16: https://doi.org/10.4238/gmr16039733

Fávero, A. P., C. E. Simpson, F. M. J. Valls and N. A. Vello. 2006. Study of evolution of cultivated peanut through crossability studies among *Arachis ipaënsis*, *A duranensis* and *A hypogaea*. *Crop Science* 46: 1546-1552.

Fernández, A., and A. Krapovickas. 1994. Cromosomas y evolución en *Arachis* (Leguminosae). *Bonplandia* 8: 187-220.

Friend, S. A., D. Quandt, S. P. Tallury, H. T. Stalker and K. W. Hilu. 2010. Species, genomes, and section relationships in the genus *Arachis* (Fabaceae): a molecular phylogeny. *Plant Systematics and Evolution* 290: 185–199.

GIMP Development Team. 2019. GIMP. Website: https://www.gimp.org

Grabiele, M., L. Chalup, G. Robledo, and G. Seijo. 2012. Genetic and geographic origin of domesticated peanut as evidenced by 5 S rDNA and chloroplast DNA sequences. *Plant Systematics and Evolution* 298: 1151-1165.

Gregory, M. P., and W. C. Gregory. 1979. Exotic germplasm of *Arachis* L. interspecific hybrids. *Journal of Heredity* 70: 185–193

Hammons, R. O. 1994 The origin and history of the groundnut. *In* J. Smartt [ed.], The groundnut crop: a scientific basis for improvement, 24–42. Springer, Dordrecht, Netherlands.

Hancock, W. G., S. P. Tallury, T. G. Isleib, Y. Chu, P. Ozias-Akins. 2019. Introgression analysis and morphological characterization of an *Arachis hypogaea* × *A. diogoi* interspecific hybrid derived population. *Crop Science* 59: 640-649.

Harlan, J. R. 1976 Genetic resources in wild relatives of crops. *Crop Science* 16: 329–333.

Hilu, K. W., and H. T. Stalker. 1995. Genetic relationships between peanut and wild species of *Arachis* sect. *Arachis* (Fabaceae): evidence from RAPDs. *Plant Systematics and Evolution* 198: 167-178.

Husted L. 1936. Cytological studies on the peanut. *Arachis* II. Chromosome number, morphology and behavior, and their application to the problem of the origin of the cultivated forms. *Cytologia* 7: 396-423.

Inglis, P. W., M. C. R. Pappas, L. V. Resende, and D. Grattapaglia. 2018. Fast and inexpensive protocols for consistent extraction of high-quality DNA and RNA from challenging plant and fungal samples for high-throughput SNP genotyping and sequencing applications. *PLoS One* 13: e0206085.

Jarvis, A., M. E. Ferguson, D. E. Williams, L. Guarino, P. G. Jones, H. T. Stalker, J. F. M. Valls, et al. 2003. Biogeography of wild *Arachis*: assessing conservation status and setting future priorities. *Crop Science* 43: 1100-1108.

Jung, S., P. L. Tate, R. Horn, G. Kochert, K. Moore, and A. G. Abbott. 2003 The phylogenetic relationship of possible progenitors of the cultivated peanut. *Journal of Heredity* 94: 334-340.

Kochert, G., T. Halward, W. D. Branch, and C. E. Simpson. 1991. RFLP variability in peanut (*Arachis hypogaea* L.) cultivars and wild species. *Theoretical and Applied Genetics* 81: 565-570.

Kochert, G., H. T. Stalker, M. Gimenes, L. Galgaro, C. R. Lopes, and K. Moore. 1996. RFLP and cytogenetic evidence on the origin and evolution of allotetraploid domesticated peanut, *Arachis hypogaea* (Leguminosae). *American Journal of Botany* 83: 1282-1291.

Koppolu, R., H. D. Upadhyaya, S. L. Dwivedi, D. A. Hoisington, and R. K. Varshney. 2010. Genetic relationships among seven sections of genus *Arachis* studied by using SSR markers. *BMC Plant Biology* 10: 15.

Korani, W., J. P. Clevenger, Y. Chu and P. Ozias-Akins, 2019 Machine learning as an effective method for identifying true single nucleotide polymorphisms in polyploid plants. *Plant Genome* 12: 180023.

Krapovickas, A., and W. Gregory. 1994. Taxonomía del género *Arachis* (Leguminosae). *Bonplandia* 8: 1-186.

Krapovickas, A., and W. C. Gregory. 2007. Taxonomy of the genus *Arachis* (Leguminosae). *Bonplandia* 16: 1-205. [authorized English translation by D. E. Williams and C. E. Simpson of Krapovickas and Gregory, 1994.]

Lavia, G. I. 1996. Estudios cromosómicos en *Arachis* (Leguminosae). *Bonplandia* 9: 111-120.

Lavia, G. I. 1998. Karyotypes of *Arachis palustris* and *A. praecox* (section *Arachis*), two species with basic chromosome number $x = 9$. *Cytologia* 63: 177-181.

Lavia, G. I. 2000. Chromosome studies of wild *Arachis* (Leguminosae). *Caryologia* 53: 177–181.

Lavia, G. I. 2001. Chromosomal characterization of germplasm of wild species of *Arachis* L. belonging to sections *Trierectoides, Erectoides* and *Procumbentes*. *Caryologia* 54: 115-119.

Lavia, G. I., A. M. Ortiz, and A. Fernández. 2009. Karyotypic studies in wild germplasm of *Arachis* (Leguminosae). *Genetic Resources and Crop Evolution* 56: 755-764.

Lu, J., and B. Pickersgill. 1993. Isozyme variation and species relationships in peanut and its wild relatives (*Arachis* L. — Leguminosae). *Theoretical and Applied Genetics* 85: 550-560.

Magallón, S., and M. S. Sanderson. 2001. Absolute diversification rates in angiosperm clades. *Evolution* 55: 1762–1780.

Mallikarjuna, N. 2005. Production of hybrids between *Arachis hypogaea* and *A. chiquitana* (section *Procumbentes*). *Peanut Science* 32: 144-147.

Mallikarjuna, N., D. Jadhav, S. Chandra, and V.P. Prasanth. 2007. Molecular genetic relationships among *Arachis diogoi* and *A. chiquitana* accessions. *Journal of SAT Agricultural Research* 3: 3.

Mallikarjuna, N., S. K. Tandra, and D. R. Jadhav. 2006. *Arachis hoehnei*, the probable B genome donor of *Arachis hypogaea* based on crossability, cytogenetical and molecular studies. *Journal of SAT Agricultural Research* 2: 1-2.

Milla, S. R., T. G. Isleib, and H. T. Stalker. 2005 Taxonomic relationships among *Arachis* sect. *Arachis* species as revealed by AFLP markers. *Genome* 48: 1-11.

Moretzsohn, M., A. Barbosa, D. Alves-Freitas, C. Teixeira, S. C. M. Leal-Bertioli, P. M. Guimarães, R. W. Pereira, et al. 2009. A linkage map for the B-genome of *Arachis* (Fabaceae) and its synteny to the A-genome. *BMC Plant Biology* 9: 40.

Moretzsohn, M. C., E. G. Gouvea, P. W. Inglis, S. C. M. Leal-Bertioli, J. F. M. Valls, and D. J. Bertioli. 2013. A study of the relationships of cultivated peanut (*Arachis hypogaea*) and its most closely related wild species using intron sequences and microsatellite markers. *Annals of Botany* 111: 113-126.

Moretzsohn, M. C., M. S. Hopkins, S. E. Mitchell, S. Kresovich, J. F. M. Valls, and M. E. R. Ferreira. 2004. Genetic diversity of peanut (*Arachis hypogaea* L.) and its wild relatives based on the analysis of hypervariable regions of the genome. *BMC Plant Biology* 4: 11.

Moss, J. P., A. K. Singh, L. J. Reddy, S. N. Nigam, P. Subrahmanyam D. McDonald, and A. G. S. Reddy. 1998. Registration of ICGV 87165 peanut germplasm line with multiple resistance. *Crop Science* 38: 1028.

Nóbile, P. M., M. A. Gimenes, J. F. M. Valls, and C. R. Lopes. 2004. Genetic variation within and among species of genus *Arachis*, section Rhizomatosae. *Genetic Resources and Crop Evolution* 51: 299–307.

Okello, D. K., C. M. Deom, and N. Puppala. 2021. Registration of 'Naronut 2 T' groundnut. *Journal of Plant Registrations* 15: 62-67.

Okello, D. K., C. M. Deom, N. Puppala, E. Monyo, and B. Bravo-Ureta. 2018. Registration of 'Serenut 6 T' groundnut. *Journal of Plant Registrations* 12: 43-47.

Oksanen, J., G. L. Simpson, F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R.B. O'Hara, et al. 2022. vegan: Community ecology package. Website: https://github.com/vegandevs/vegan

Ortiz, A. M., L. Chalup, M. C. Silvestri, G. Seijo, and G. I. Lavia. 2023. Genomic relationships of the polyploid rhizoma peanut (*Arachis glabrata* Benth.) inferred by genomic in situ hybridization (GISH). *Anais da Academia Brasileira de Ciências* 95: e20210162.

Pandey, M. K., G. Agarwal, S. M. Kale, J. Clevenger, S. N. Nayak, M. Sriswathi, A. Chitikineni, et al. 2017. Development and evaluation of a high density genotyping 'Axiom_*Arachis*' array with 58 K SNPs for accelerating genetics and breeding in groundnut. *Scientific Reports* 7: 40577.

Pandey, M. K., B. Gautami, T. Jayakumar, M. Sriswathi, H. D. Upadhyaya, M. V. C. Gowda, T. Radhakrishnan, et al. 2012. Highly informative genic and genomic SSR markers to facilitate molecular breeding in cultivated groundnut (*Arachis hypogaea*). *Plant Breeding* 131: 139-147.

Peanut Crop Germplasm Committee. 2004. Report on the status of *Arachis* germplasm in the United States. Available at https://www.ars-grin.gov/npgs/cgc_reports/Status11.pdf. USDA-ARS, Washington, D.C., USA.

Peñaloza, A., and J. Valls. 1997. Contagem do número cromossômico em acessos de *Arachis decora* (Leguminosae). *In* R. F. de A. Veiga, M. L. A. Bovi, J. A, Betti, and R. B. Q. Voltan [eds.], Simpósio latino-

americano de recursos genéticos vegetais, 39. IAC/EMBRAPA-CENARGEN, Campinas, Brazil.

Price, A. L., N. A. Zaitlen, D. Reich, and N. Patterson. 2010. New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11: 459-463.

R Core Team. 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Website: https://www.R-project.org

Rambaut A. 2016. FigTree v1.4.3: tree figure drawing tool. Website: http://tree.bio.ed.ac.uk/software/figtree/ [accessed 15 November 2022].

Richards, C. M., and G. M. Volk. 2010. New challenges for data management in genebanks. *Acta Horticulturae* 859: 333–336.

Robledo, G., G. I. Lavia, and G. Seijo. 2009. Species relations among wild *Arachis* species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. *Theoretical and Applied Genetics* 118: 1295-1307.

Robledo, G., and G. Seijo. 2008. Characterization of the *Arachis* (Leguminosae) D genome using fluorescence in situ hybridization (FISH) chromosome markers and total genome DNA hybridization. *Genetics and Molecular Biology* 31: 717-724.

Robledo, G., and G. Seijo. 2010. Species relationships among the wild B genome of *Arachis* species (section *Arachis*) based on FISH mapping of rDNA loci and heterochromatin detection: a new proposal for genome arrangement. *Theoretical and Applied Genetics* 121: 1033-1046.

Santana, S. H., and J. F. M. Valls. 2015. *Arachis veigae* (Fabaceae), the most dispersed wild species of the genus, and yet taxonomically overlooked. *Bonplandia* 24: 139-150.

Saritha, K., R. P. Vasanthi, M. Shanthi Priya, and P. Latha. 2018. Genetic divergence analysis in groundnut *Arachis hypogaea* L. *Electronic Journal of Plant Breeding* 9: 1355-1361.

Secretariat of the Convention on Biological Diversity. 2000. Cartagena protocol on biosafety to the Convention on Biological Diversity: text and annexes. Secretariat of the Convention on Biological Diversity, Montreal, Quebec, Canada.

Secretariat of the Convention on Biological Diversity. 2011. Nagoya protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity: text and annex. Secretariat of the Convention on Biological Diversity, Montreal, Quebec, Canada. Available at: https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf

Seijo, G., G. I. Lavia, A. Fernandez, A. Krapovickas, D. A. Ducasse, D. J. Bertioli, and E. A. Moscone. 2007. Genomic relationships between the cultivated peanut (*Arachis hypogaea*, Leguminosae) and its close relatives revealed by double GISH. *American Journal of Botany* 94: 1963-1971.

Seijo, G. J., M. Atahuachi, C. E. Simpson, and A. Krapovickas. 2021. *Arachis inflata*: a new B genome species of *Arachis* (Fabaceae). *Bonplandia* 30: 169-174.

Seijo, J. G., G. I. Lavia, A. Fernández, A. Krapovickas, D. Ducasse, and E. A. Moscone. 2004. Physical mapping of the 5 S and 18S-25S rRNA genes by FISH as evidence that *Arachis duranensis* and *A. ipaënsis* are the wild diploid progenitors of *A. hypogaea* (Leguminosae). *American Journal of Botany* 91: 1294-1303.

Sigovini, M., E. Keppel, and D. Tagliapietra. 2016. Open nomenclature in the biodiversity era. *Methods in Ecology and Evolution* 7: 1217-1225.

Silvestri, M. C., A. M. Ortiz, and G. I. Lavia. 2015. rDNA loci and heterochromatin positions support a distinct genome type for 'x = 9 species' of section *Arachis* (*Arachis*, Leguminosae). *Plant Systematics and Evolution* 301: 555-562.

Silvestri, M. C., A. M. Ortiz, G. A. Robledo Dobladez, and G. I. Lavia. 2020. Chromosome diversity in species of the genus *Arachis*, revealed by FISH and CMA/DAPI banding, and inferences about their karyotype differentiation. *Anais da Academia Brasileira de Ciências* 92: e20191364.

Simpson, C. E., A. R. Custodio, L. S. Rodrigues, A. P. Penaloza, J. F. M. Valls. and J. M. Cason. 2019. Using *Arachis vallsiiArachisIn* Krapov. & W.C. Greg. as a bridge species for introgression in [abstract]. 50th Proceedings of the American Peanut Research and Education Society, Inc., annual meeting 2018, Williamsburg, VA, USA, 124. American Peanut Research and Education Society, Alexandria, VA, USA. Website: https://apresinc.com/images/2019/04/2018-Proceedings-50th.pdf

Singh, A. K., J. Smartt, and R. Singh. 2004. Variation studies in a wild groundnut species. *Arachis stenosperma* Krapov. & W.C. Gregory nov. sp. *Plant Genetic Resources* 2: 99-106.

Singh, N., S. Wu, W. J. Raupp, S. Sehgal, S. Arora, V. Tiwari, P. Vikram et al. 2019. Efficient curation of genebanks using next generation sequencing reveals substantial duplication of germplasm accessions. *Scientific Reports* 9: 650.

Smartt, J., W. C. Gregory and M. P. Gregory. 1978. The genomes of *Arachis hypogaea*. 1. Cytogenetic studies of putative genome donors. *Euphytica* 27: 665-675.

Stalker, H. T. 1991. A new species in section *Arachis* of peanuts with a D genome. *American Journal of Botany* 78: 630-637.

Stalker, H. T. 2017. Utilizing wild species for peanut improvement. *Crop Science* 57: 1102-1120.

Stalker, H. T., J. S. Dhesi, and D. C. Parry. 1991. An analysis of the B genome species *Arachis batizocoi* (Fabaceae). *Plant Systematics and Evolution* 174: 159-169.

Stalker, H. T., and R. D. Dalmacio. 1986. Karyotype relations and analysis among varieties of *Arachis hypogaea* L. *Cytologia* 51: 167-629.

Stalker, H. T., J. S. Dhesi, and G. Kochert. 1995. Genetic diversity within the species *Arachis duranensis* Krapov. & W.C. Gregory, a possible progenitor of cultivated peanut. *Genome* 38: 1201-1212.

Stalker, H. T., T. D. Phillips, J. P. Murphy, and T. M. Jones. 1994. Variation of isozyme patterns among *Arachis* species. *Theoretical and Applied Genetics* 87: 746-755.

Tallury, S. P., K. W. Hilu, S. R. Milla, S. A. Friend, M. Alsaghir, H. T. Stalker, and D. Quandt. 2005. Genomic affinities in *Arachis* section *Arachis* (Fabaceae): molecular and cytogenetic evidence. *Theoretical and Applied Genetics* 111: 1229-1237.

Valls, J. F. M. 1996. O gênero *ArachisIn* (Leguminosae): Importante fonte de proteinas na préhistoria sul-americana? Anais da Reunião Científica da Sociedade de Arqueologia Brasileira, 8th meeting, 1995, 265–280. Edipucrs, Porto Alegre, Brazil.

Valls, J. F. M. 2000. Diversidade genética no gênero *Arachis* e a origem do amendoim. *In* Anais do Encontro sobre Temas de Genética e Melhoramento, 17th meeting, 2000, 19–33. ESALQ/Departamento de Genética, Piracicaba, Brazil.

Valls, J. F. M., and C. E. Simpson. 2005. New species of *Arachis* L. (Leguminosae) from Brazil, *Paraguay and Bolivia. Bonplandia* 14: 35-63.

Valls, J. F. M., and C. E. Simpson. 2017. A new species of *Arachis* (Fabaceae) from Mato Grosso, Brazil, related to *Arachis matiensis*. *Bonplandia* 26: 143-149.

Valls, J. F. M., L. C. Costa, and A.R. Custodio. 2013. A novel trifoliolate species of *Arachis* (Fabaceae) and further comments on the taxonomic section *Trierectoides*. *Bonplandia* 22: 91-97.

Valls, J. F. M., C. E. Simpson and V. R. Rao. 1995. Collecting wild species of *Arachis*. In L. Guarino, V. R. Rao, and R. Reid [eds.], Collecting plant genetic diversity: Technical guidelines, 677–684. CAB International, Wallingford, UK.

Vishwakarma, M. K., S. M. Kale, M. Sriswathi, T. Naresh, Y. Shasidhar, V. Garg, M. K. Pandey, and R. K. Varshney. 2017. Genome-wide discovery and deployment of insertions and deletions markers provided greater insights on species, genomes, and sections relationships in the genus *Arachis*. Frontiers in Plant Science 8: 2064.

Wang, H., P. Khera, B. Huang, M. Yuan, R. Katam, W. Zhuang, K. Harris-Shultz et al. 2016. Analysis of genetic diversity and population structure of peanut cultivars and breeding lines from China, India and the US using simple sequence repeat markers. *Journal of Integrated Plant Biology* 58: 452-465.

Williams, D. E. 2022. Global strategy for the conservation and use of peanut genetic resources. Global Crop Diversity Trust, Bonn, Germany. Website: https://doi.org/10.5281/zenodo.7545106

Williams, K. A., and D. E. Williams. 2001. Evolving political issues affecting international exchange of *Arachis* genetic resources. *Peanut Science* 28: 132-135.

Zheng, Z., Z. Sun, Y. Fang, F. Qi, H. Liu, L. Miao, P. Du, et al. 2018. Genetic diversity, population structure, and botanical variety of 320 global peanut accessions revealed through tunable genotyping-by-sequencing. *Scientific Reports* 8: 14500.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Complete set of samples genotyped in this work with species, accession, collector, original collection code, NPGS Plant Introduction (PI) number (when available), ploidy, genome type, coordinates of the collection site, and germplasm bank where the sample was obtained.

Appendix S2. All R scripts used in this work.

Appendix S3. Identity matrix of 272 accessions of 32 diploid species. This matrix was used to construct the phylogenetic tree (Figure 4).

Appendix S4. Phylogenetic tree using 438 samples of 284 accessions of 32 diploid species of section *Arachis* based on 15 K SNPs genotyping calls. Accessions of the same species are represented by the same color.

Appendix S5. Set of 272 accessions included in the curated phylogenetic tree (Figure 4) with species, accession, collector's name, original collection code, NPGS Plant Introduction (PI) number (when available), ploidy, genome type, and coordinates of the collection site.

Appendix S6. Geographical distribution of all accessions genotyped and analyzed for genetic distance. Each dot represents an accession. Genome types are represented by different colors.

Appendix S7. Compilation of notes and observations by the three *Arachis* collectors and coauthors Guillermo Seijo, Charles Simpson, and José Valls, given in writing or as excerpts from our transcriptions of our meetings (in Brazil, Argentina, the United States or virtually) during the analyses of these data.

How to cite this article: Leal-Bertioli, S. C. M., F. J. de Blas, M. C. Chavarro, C. E. Simpson, J. F. M. Valls, S. P. Tallury, M. C. Moretzsohn, A. R. Custodio, H. T. Stalker, G. Seijo, and D. J. Bertioli. 2024. Relationships of the wild peanut species, section *Arachis*: A resource for botanical classification, crop improvement, and germplasm management. *American Journal of Botany* 111(6): e16357. https://doi.org/10.1002/ajb2.16357