Received: 5 October 2025 | Accepted: 12 January 2026

DOI: 10.1002/ajb2.70178

RESEARCH ARTICLE

Botany

Phylotranscriptomics and genome size evolution in *Leucaena* (Fabaceae): Paleotetraploid genomic stability overshadows diploidization and environmental effects

Alex Abair[1] | Ashley N. Egan[2] | Brittany Bugg[2] | Madhugiri Nageswara-Rao[3] | Colin E. Hughes[4] | Kitti Denson[1] | Mike Lopez III[1] | Hailey Sermersheim[2] | Joshua T. Trujillo[5] | Shannon C. K. Straub[6] | Jessica P. Houston[7] | Ya Yang[8] | Susan R. Strickler[9,10] | Richard C. Cronn[11] | Aaron Liston[12] | Carl E. Hjelmen[2] | C. Donovan Bailey[1]

[1]Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA

[2]Department of Biology, Utah Valley University, Orem, UT 84058, USA

[3]Subtropical Horticultural Research Station, USDA-ARS, Coral Gables, FL 33158, USA

[4]Department of Systematic and Evolutionary Botany, University of Zurich, Zurich, Switzerland

[5]NIH Undergrad RISE/MARC Programs, New Mexico State University, Las Cruces, NM 88003, USA

[6]Department of Biology, Hobart and William Smith Colleges, Geneva, NY 14456, USA

[7]Department of Chemical & Materials Engineering, New Mexico State University, Las Cruces, NM 88003, USA

[8]Department of Plant and Microbial Biology, University of Minnesota-Twin Cities, Saint Paul, MN 55108, USA

[9]Program in Plant Biology and Conservation, Northwestern University, Evanston, IL 60201, USA

[10]Negaunee Institute for Plant Conservation Science and Action, Chicago Botanical Garden, Glencoe, IL 60022, USA

[11]USDA Forest Service, Pacific Northwest Research Station, Corvallis, OR 97331, USA

[12]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR 97331, USA

Correspondence C. Donovan Bailey, Department of Biology, New Mexico State University, Las Cruces, NM 88003 USA.
Email: dbailey@nmsu.edu

Ashley N. Egan, Utah Valley University, Department of Biology, Orem, UT 84058 USA.
Email: AEgan@uvu.edu

Funding information
Directorate for Biological Sciences, Grant/Award Number: 1238731

Abstract

Premise: Advances in transcriptomic and reduced representation genomic sequencing are deepening our understanding of how hybridization, reticulation, and environmental variation impact species diversification and genome size. *Leucaena* is a useful system for exploring the genomic basis of allopatric and allopolyploid speciation events and the effect of environmental pressures on genome size across 30° of latitude. We investigate phylogenetic relationships, the roles of polyploidy and hybridization in speciation, and genome size evolution in the genus.

Methods: Using newly generated RNA-sequencing data for *Leucaena*, we applied reference-guided and de novo phylotranscriptomics to reconstruct nuclear and organellar phylogenies. We then used comparative genome sizes from 252 samples and phylogenetic methods to investigate genome size evolution broadly and the impacts of environmental variables specifically.

Results: The phylogenetic results supported cladogenetic, rather than reticulate/hybrid, origins for most of the 19 paleotetraploid species. By contrast, gene tree data supported hybrid origins for octoploid *Leucaena*. The ancestral paleotetraploid genome size (1.52 pg/2 C) is relatively conserved among the paleotetraploids, rejecting our hypothesis associated with environmental variables significantly impacting genome sizes.

Conclusions: The phylogenetic results illustrate the complex interplay of intrinsic and external factors that impact speciation, including ancient whole-genome duplication (WGD), cladogenesis, secondary contact, and allopolyploidy. A weak relationship between genome size and environmental variables suggests that other factors, including paleotetraploid genomic stability, have constrained genome size variation following a WGD 16+ million years ago. The findings are consistent with a small but growing number of studies identifying groups with ancient WGDs that resist diploidization associated with gene and DNA loss.

KEYWORDS

allopolyploidy, balanced gene expression, Caesalpinioideae, diploidization, genomic structural stability, incomplete lineage sorting, Leguminosae, Mimoseae, octoploid

Land plant genomes are known for their exceptional complexity, size variation, flexibility, and genetic variation encoded therein. The 2440-fold range in nuclear genome size across angiosperms (Pellicer et al., 2018) largely results from variation in repetitive elements and polyploidy. Investigations on the impacts of genome size variation in angiosperms focus on diverse topics from the C-value paradox (Thomas, 1971) to cellular phenotype and function (e.g., Freeling et al., 2015; Doyle and Coate, 2019). Previous studies have identified strong correlations between genome size and cell size, photosynthetic potential, and cell cycle times (e.g., Šímová and Herben, 2012; Roddy et al., 2020) as well as more subtle correlations with environmental factors (Bilinski et al., 2018; Qiu et al., 2019; Souza et al., 2019; Bureš et al., 2024). Even the potential for invasiveness and weediness has been associated with genome size variation (e.g., Suda et al., 2015; Cang and Dlugosch, 2022; Guo et al., 2024). Furthermore, phenotypic traits such as lifespan (annual, perennial, etc.), seed mass, flowering time, and other features may also explain some portion of genome size variation (e.g., Bilinski et al., 2018).

Broad perspectives on the importance of polyploidy as a contributor to genome size variation are consistent with a central role of whole-genome duplication (WGD) in ferns and angiosperms, but more limited roles in other major lineages across the tree of life (e.g., Van de Peer et al., 2017; Román-Palacios et al., 2021). By doubling and often shuffling the genome and its genes, WGD and accompanying downstream diploidization can be powerful evolutionary processes that promote adaptation and evolutionary innovation via impacts on mutation rates, gene retention/recruitment for potential novel function(s), chromosomal rearrangements (e.g., Van de Peer et al., 2017) and patterns and rates of speciation (e.g., Werth and Windham, 1991; Otto and Whitton, 2000). However, even within angiosperms, the state of being a polyploid is viewed as transient, with duplicated non-adaptive elements (chromosomes and/or gene copies) being diploidized through various processes (reviewed by, Li et al., 2021). Despite the logic of these ideas and widespread empirical evidence for angiosperm diploidization, recent genomic work has identified exceptions to the rule, with evidence that some genomes and lineages resist diploidization and retain stable polyploid configurations over longer timescales. These include phases of genomic stability and diploidization observed in the salmonid fishes following a WGD event 100 million years ago (Mya) (Gundappa et al., 2021), a prolonged process of diploidization for *Sequoiadendron* J. Buchholz spanning 33 My (Scott et al., 2016), and remarkable genomic stability and gene maintenance following 12 of 14 independent WGD events in the grass tribe Andropogoneae over the last 1–12 My (Snodgrass et al., 2024). These studies suggest that the processes and timeframe underlying diploidization following polyploidy vary considerably among lineages, but the reasons for this variation remain poorly understood.

Within the angiosperms, the legume family (Fabaceae Lindl. or Leguminosae Juss.), with over 22,500 species, represents one of the most diverse, abundant, geographically and ecologically widespread, and economically important plant lineages (Hughes et al., 2025). Polyploidy is thought to have played a significant role in legume diversification, with more than 30 documented WGDs, including multiple events early in the evolutionary history of the family (e.g., Cannon et al., 2015; Koenen et al., 2021; Zhao et al., 2021). However, uncertainties about the precise number and phylogenetic placements of these early legume WGDs mean that the impacts on diversification and evolution of traits have been challenging to assess in detail.

Leucaena (Fabaceae) is a core member of the Dichrostachys clade of tribe Mimoseae, which has been widely impacted by both paleo- and neopolyploidy (Hughes and Luckow, 2024). Specifically, the 24 species of the genus *Leucaena* (Leguminosae: Caesalpinioideae: tribe Mimoseae) diversified following a stem-lineage WGD event that occurred over 16 Mya (Ringelberg et al., 2023; Chen et al., 2024). The majority of the subsequently derived species of *Leucaena*, 19 species typically referred to as "diploids" ($2n = 52$ and 56) (Hughes, 1998a), are thought to have evolved through classic allopatric speciation. Conversely, five additional allotetraploids ($2n = 104$ and 112) (Hughes, 1998a) appear to have resulted from human-mediated translocation and cultivation (Hughes et al., 2002, 2007; Govindarajulu et al., 2011a). Importantly, recent genome sequencing has revealed that the 19 "diploid" species are genomic paleotetraploids that have undergone limited diploidization, leading to the conclusion that five previously defined "tetraploids" are better referred to as octoploids, likely allooctoploids (Bailey et al., 2024; Chen et al., 2024).

Leucaena comprises small to medium-sized tree species concentrated in seasonally dry tropical forests from the southern United States to Peru with the highest species diversity in southern Mexico and Central America (Hughes, 1998a). Within mimosoid legumes, the genus is characterized by the unique synapomorphy of trichome bearing anthers. It is otherwise similar to closely related mimosoid genera in its bipinnate leaves, extrafloral nectaries, flowers aggregated into compact heads, and 10 free stamens. Traits including bark morphology, quantitative leaf traits (number of pairs of pinnae, number of pairs of leaflets, and leaflet size), pollen in polyads vs monads, and anthers apiculate or not have systematic value in defining species and taxonomic groups within *Leucaena* (Hughes, 1998a). For example, *Leucaena* can be divided morphologically into discrete groups based on leaf size and shape, with large leaflets defining a subclade of species and thick corky bark characterizing another clade (Hughes, 1998a). The fruits of *Leucaena* are long, flattened pods whose unripe seeds are an important minor food source, known as guajes, consumed across south-central Mexico now and in the past (e.g.,

Hughes, 1998a). Seed remains from archaeological sites suggest Nahuatl, Mixtec, and Zapotec people have been harvesting and consuming unripe pods and seeds of *Leucaena* as far back as 6000 years ago (Zárate, 2000), with up to 13 species brought into cultivation across south-central Mexico, spawning spontaneous interspecific hybrids associated with artificial sympatry in cultivation (Hughes et al., 2007). Broader cultivation and human translocation have also expanded the range of several *Leucaena* species throughout the tropics, with *L. leucocephala* now displaying a pantropical introduced/invasive distribution. The adoption of this species as a multipurpose crop in combination with its reproductive success has resulted in its listing as one of the world's 100 worst invasive alien species (https://www.iucngisd.org/gisd/100_worst.php). Modern commercial use focuses largely on agroforestry systems and range management (Brewbaker, 1987; Cowley and Roschinsky, 2019; Hopkins et al., 2019; Lemin et al., 2019) where *Leucaena* species are widely used for livestock fodder, green manure, soil stabilization, and as shade plants in commercial and subsistence farming systems (Brewbaker, 1987; Hughes, 1998b).

Chromosomal and genome size variation in *Leucaena* and the scope of its native (spanning >30° of latitude) and introduced distributions (pantropical) (Hughes, 1998a) offer opportunities to investigate the impacts of polyploidy, environmental variation, and other factors on genome size variation within a genus of neotropical woody plants. However, previous phylogenetic studies using a handful of biparentally inherited nuclear markers failed to confidently resolve relationships between major clades of *Leucaena* or within the largest clade of "diploid" taxa (Clade 1) (Govindarajulu et al., 2011a, 2011b).

The complex evolutionary history, effects of polyploidy, geographic distribution, and genome size variation in *Leucaena* provide an excellent system to investigate modes and mechanisms of speciation as well as the genomic constraints and responses to a variety of intrinsic (e.g., diploidization, dosage balance) and extrinsic (e.g., environmental) factors. We hypothesized that observed gene tree conflict in octoploid species of *Leucaena* is mainly associated with hybrid origins, whereas gene tree conflict among the paleotetraploids is primarily related to incomplete lineage sorting (ILS). We also hypothesized that environmental variation has significantly influenced genome sizes among the 19 paleotetraploid species of *Leucaena*. To address these questions, we used cost-effective RNA-sequencing (RNA-seq) data extracted from whole seedlings, which captured thousands of expressed genes per sample across a variety of tissues, and apply both reference-guided and de novo phylotranscriptomic approaches to variously reconstruct nuclear, plastid, and mitochondrial genome-based phylogenies with complete species-level sampling. We then used those results to test for and characterize the relative impact of reticulate evolution and ILS within and among paleotetraploid and octoploid species of *Leucaena*. Lastly, we integrated newly generated comparative genome-size data across species to test for and characterize the effects of environmental variation on genome size.

MATERIALS AND METHODS

Phylogenetics: taxon sampling, sequencing, and locus recovery

Taxon sampling included all currently recognized species of *Leucaena* (19 paleotetraploids and five octoploids) along with previously generated transcriptome data for three mimosoid outgroup taxa, *Albizia julibrissin*, *Entada abyssinica*, and *Microlobius foetidus* (Koenen et al., 2020) (Table 1). Total RNA was extracted from whole seedlings (at the three-leaf stage) for all taxa except *L. pueblana* (see below). RNA quality and quantity were assessed using a Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA), Qubit (Thermo Fisher Scientific), and agarose gel electrophoresis before library preparation. A TruSeq RNA Sample Prep library kit (Illumina, San Diego, CA, USA) was used to generate libraries for each taxon using 4 μg total RNA and the manufacturer's instructions. Each library occupied 20% of a lane on an Illumina HiSeq 2000 (Axeq Technologies, Seoul, South Korea) for 101 cycles. Paired-end (100-bp) reads were trimmed of their Illumina adaptors and filtered for a minimum read length of 65 bp using Trimmomatic v0.34 (Bolger et al., 2014) (parameters ILLUMINACLIP:Trimmomatic-0.32/adapters/TruSeq3-PE-2.fa:2:30:10 MINLEN:65). No fresh tissues could be obtained for *L. pueblana*, so we substituted genomic DNA-seq data in place of RNA-seq (using 111 million 100 bp read pairs [222 M reads] with an average insert size of 250 bp (ca. 30× coverage). Raw reads were also trimmed with Trimmomatic (Trimmomatic-0.32/adapters/TruSeq3-PE-2.fa:2:30:10 MINLEN:65).

Reference-guided phylogenies, including divergent and allopolyploid lineages

We assembled each data set using reference-guided and de novo approaches. We then carried out phylogenomic analyses using maximum likelihood methods on concatenated matrices (sensu Stamatakis, 2014) and coalescent-based methods (sensu Zhang et al., 2018). The sets of taxa used in the construction of trees included (1) putatively divergently related (nonhybrid) paleotetraploid taxa only, (2) those containing the paleotetraploids plus all putative allooctoploids, and (3) sets in which each matrix/tree includes a single allooctoploid species along with all 19 paleotetraploid taxa. The identification of reference-derived loci involved using a chromosomal-scale reference genome for the paleotetraploid species, *L. trichandra* (C. D. Bailey et al., unpublished data) and associated transcript data mapped with STAR v2.7.10b (Dobin et al., 2013). The set of *L. trichandra* expressed loci was reduced to those with one isoform in the genome and a length of 600–5000 bp. We then used STAR to map RNA-seq data for each taxon to this reference transcriptome. The resulting sorted BAM files were used to create a BED file using genomeCoverageBed in BEDTools (Quinlan and Hall, 2010) to mask low coverage bases (<5) before the recovery of bases from each alignment using mpileup (Li et al., 2009). For

TABLE 1 Taxonomic sampling and authorities, SRA accessions, voucher information, and chromosome numbers compiled from Hughes (1998). An asterisk denotes a tentative chromosome number assignment based on the assessment of the original author.

Taxon	SRA accession	Voucher (herbarium)	Chromosome number (2*n*)
Leucaena collinsii Britton & Rose	SRX2719653	C.E. Hughes 1137-8. 1187 (FHO, K, MEXU)	52
L. confertiflora var. *adenotheloidea* (S. Zárate) C.E. Hughes	SRX2719652	C.E. Hughes 1796 (FHO, K, MEXU, NY MO)	112
Leucaena cruziana Britton & Rose	SRX2719651	C.E. Hughes 559-579 (FHO, K, MEXU)	52*
Leucaena cuspidata Standl.	SRX2719650	C.E. Hughes 1851 (FHO, K, MEXU, NY)	unknown
Leucaena diversifolia (Schltdl.) Benth.	SRX2719649	C.E. Hughes 1693 (E, FHO, K, MEXU, MO, and NY)	104
Leucaena esculenta (DC.) Benth.	SRX2719648	C.E. Hughes 894, 895, 898 (FHO, K, MEXU)	52
Leucaena greggii S. Watson	SRX2719647	C.E. Hughes 695 (FHO, K, and MEXU)	56
Leucaena involucrata S. Zárate	SRX2719646	C.E. Hughes 1572 (FHO, K, MEXU, E, NY, MO)	unknown
Leucaena lanceolata S. Watson	SRX2719645	C.E. Hughes 603-613 (FHO, K, MEXU)	52
Leucaena lempirana C.E. Hughes	SRX2719644	C.E. Hughes 1411-2, 1414, 1712 (E, FHO, K, MEXU, EAP, NY)	unknown
Leucaena leucocephala subsp. *glabrata* (Rose) S. Zárate	SRX2719643	Stead and Styles 705 (FHO, MEXU)	104
Leucaena leucocephala subsp. *leucocephala* (Lamark) de Wit	SRX2719642	C.E. Hughes 1734 (E, FHO, K, MEXU. MO, NY)	104
Leucaena macrophylla subsp. *istmensis* C.E. Hughes	SRX2719641	C.E. Hughes 580-585, 650-655 (FHO, K, MEXU)	52*
Leucaena macrophylla subsp. *macrophylla* Benth.	SRX2719640	C.E. Hughes 1179 stored in (FHO, MEXU, K)	52*
Leucaena magnifica (C.E. Hughes) C.E. Hughes	SRX2719639	C.E. Hughes 1089, 1090, 1093 (FHO, K, MEXU)	52*
Leucaena matudae (Zarate) C.E. Hughes	SRX2719638	C.E. Hughes 879, 883 (FHO, K, MEXU)	52*
Leucaena multicapitula Schery	SRX2719637	C.E. Hughes 1025, 1029, 1032-1035 (FHO, K, MEXU, PMA)	52
Leucaena pallida Britton & Rose	SRX2719636	C.E. Hughes 1662 stored in the (E, FHO, K, MEXU, NY)	104
Leucaena pueblana Britton & Rose	SRX2719610	C.E. Hughes 2092 (FHO)	unknown
Leucaena pulverulenta (Schltdl.) Benth.	SRX2719635	C.E. Hughes 1058, 1059, 1061 (FHO, K, MEXU)	56
Leucaena retusa Benth.	SRX2719634	Bendeck s.n. with (UANL)	56
Leucaena salvadorensis Standl. ex Britton & Rose	SRX2719633	C.E. Hughes 1407-8, 1431, 1434, 1444 (FHO, EAP, K, MEXU, MO, NY)	56*
Leucaena shannonii Donn. Sm.	SRX2719632	C.E. Hughes 239, 282, 310, 311 (FHO, MEXU)	56
Leucaena trichandra (Zucc.) Urb.	SRX2719631	C.E. Hughes 1421-1424, 1427, 1708-1710 (FHO, EAP, K, MEXU, NY)	52,56*
Leucaena trichodes (Jacq.) Benth.	SRX2719630	C.E. Hughes 992, 997, 1000, 1195, 1197, 1200 (FHO, K, MEXU, QAME)	52
Leucaena zacapana (C.E. Hughes) R. Govind. & C.E. Hughes	SRX2719629	C.E. Hughes, 1096-7 (FHO, MEXU, K)	52
Albizia julibrissin Durazz.	SRX6901075	E. Koenen 601 (Z)	26
Microlobius foetidus (Jacq.) M.Sousa & G. Andrade	SRX6901077	C.E. Hughes 2150 (FHO)	unknown
Entada abyssinica Steud.	SRX6901076	MSB 0133199 (K)	unknown

each accession, the recovered loci were filtered to include only those with <25% missing data across the length of that transcript. Finally, the set of combined matrices was selected to include those loci with data for every species (i.e., no missing terminals). The concatenated matrix of loci was generated with catfasta2phyml.pl (https://github.com/nylander/catfasta2phyml) and the supermatrix species trees were produced with RAxML v8.2.12 (Stamatakis, 2014) (raxmlHPC -f a -T 50 -m GTRCAT -p 12345 -x 12345 -# 100). Gene trees used as input for the ASTRAL (Zhang et al., 2018) element were created using RAxML (raxmlHPC -f a -T 2 -m GTRCAT -p 12345 -x 12345 -# 100). For each gene tree, the weakly supported nodes (<50% BS support) were collapsed. The resulting number of gene trees supporting each node in the associated RAxML best tree topology (above) was identified using Phyparts (https://bitbucket.org/blackrim/phyparts/). A species tree was also estimated for each set of genes using ASTRAL 5.6.2. Scripts associated with all these elements are available on (https://github.com/cdb3ny/Ref-guided_and_Genome_Size).

De novo phylotranscriptomics for paleotetraploids

De novo transcriptome assembly was carried out using Trinity v2.3.1 (Grabherr et al., 2011) with the default parameters for paired end reads (parameters, --seqType fq --JM 20 G --SS_lib_type RF). The procedures used for homology assessment and species tree inference generally followed the protocol of Yang and Smith (2014). Scripts were downloaded from Bitbucket (https://bitbucket.org/yangya/phylogenomic_dataset_construction) in June 2018. Shell scripts were written to parallelize tasks and provide additional functions (https://github.com/AlexanderAbair/Leucaena-de-novo-Transcriptomics). In brief, proteomes from *Arabidopsis thaliana* L. Heynh. (TAIR Araport11 protein lists, GenBank accessions CP002684–CP002688) (Cheng et al., 2017) and *Glycine max* L. Merr. (Uniprot proteome ID: UP000008827) (Schmutz et al., 2010) were selected to build a BLAST database using BLAST 2.2.31 + . Open reading frames (ORFs) ≥300 bp were then extracted from transcriptomes, translated into peptide and coding sequences (CDS) sequences using TransDecoder v3.0.1 (https://github.com/TransDecoder/TransDecoder/issues/113). Redundancy in the resulting CDS files was reduced using cd-hit-est. The remaining CDS for paleotetraploids and the three outgroups were concatenated into a single file for use in building another BLAST database. An all-by-all BLASTn was run on this database file, and the output was filtered by hit fraction of 0.4 using blast_to_mcl.py (Yang and Smith, 2014). Markov clustering of the filtered results produced 48,957 sequence clusters. A FASTA file was then written for each of the 14,381 sequence clusters containing all 22 taxa write_fasta_files_from_mcl.py (Yang and Smith, 2014). Clusters were aligned using MAFFT (Katoh and Standley, 2013) and trimmed by a minimal column occupancy of 10% using Phyutility (Smith and Dunn, 2008), and trees were inferred from each cluster using RAxML v8.2.9 (Stamatakis, 2014). Tips longer than 0.4 were trimmed from these trees. All of the aligning, trimming, and tree building was coordinated by fasta_to_tree.py (Yang and Smith, 2014) with the following parameters: python ~/fasta_to_tree.py ~/clusters 32 dna n. Additionally, tips longer than 0.2 and greater than 10 times the length of their sister were trimmed using trim_tips.py (Yang and Smith, 2014). Monophyletic and paraphyletic tips belonging to one taxon were masked, and only the tip with the highest number of unambiguous characters in each taxon was retained with mask_tips_by_taxonID_transcripts.py (Yang and Smith, 2014). Then the long internal branches connecting distantly related gene copies were cut with cut_long_internal_branches.py (using the default setting of 1) (Yang and Smith, 2014), and FASTA files were written from the remaining sequence ID's in these trees with write_fasta_files_from_trees.py (Yang and Smith, 2014). Sequences were aligned once more, and final homolog trees were inferred from the alignments using RAxML through fasta_to_tree.py.

Filtering for homolog clusters with exactly one representative per taxon was conducted to infer orthologs using filter_1to1_orthologs.py (Yang and Smith, 2014). FASTA files were written from the sequences named in the ortholog trees with write_ortholog_fasta_files.py (Yang and Smith, 2014), and new ortholog cluster alignments were estimated using Prank v150803 through write_alignments_from_orthologs.py (Yang and Smith, 2014). Alignments were trimmed using default parameters in phyutility_wrapper.py (Yang and Smith, 2014). Trimmed alignments with minimal lengths of 300 nucleotides were concatenated into a supermatrix to be used for RAxML species tree estimation with concatenate_matrices.py (Yang and Smith, 2014). RAxML was run with each ortholog as a separate partition with the following parameters: raxml -T 2 -p 12345 -m GTRCAT -q 121_filter300-22.model -# 100 -s 121_filter300-22.phy -n 121_filter300-22.

The same orthologs were used to infer an additional species tree using ASTRAL III V5.6.2 (Zhang et al., 2018). Corresponding gene trees were generated using RAxML and run through ASTRAL (including default parameters plus "-t 3") to infer species trees. Gene tree conflict was explored using PhyParts v1.0.0 (Smith et al., 2015), and the percentages of trees that were concordant and conflicting at each node were displayed respectively in the form of pie charts using phypartspiecharts.py (https://github.com/mossmatters/MJPythonNotebooks).

Since only genomic DNA sequence data were available for *L. pueblana*, the aforementioned steps were repeated to identify the placement of *L. pueblana*. We first generated a draft reference genome (776 Mbp) for *L. pueblana* (https://github.com/cdb3ny/Ref-guided_and_Genome_Size) using SOAPdenovo (Li et al., 2010). Next, we extracted all ORFs longer than 300 nucleotides (using the EMBOSS getorf function) and used those as the input equivalent to the transcriptomes used for the other species.

Exploring phylogenetic incongruence: ILS vs. introgression/hybridization

Paleotetraploid taxa

PhyloNetworks (Solís-Lemus et al., 2017) and the TICR test (Stenz et al., 2015) (both elements following Github version last edited May 29, 2022, https://github.com/crsl4/PhyloNetworks.jl/wiki) were used to investigate whether gene tree conflict among the 19 paleotetraploid taxa is more consistent with ILS or previously undetected hybridization/introgression. For these analyses, each reference-assembled locus (described above) was run through 3SEQ (Lam et al., 2018) with default parameters to identify and remove loci with potential recombination. Sister relationships between terminals of interest were determined with in-house scripts. Bayesian trees were generated for each alignment using the MrBayes script (mb.pl) provided with PhyloNetworks. Concordance factors were estimated from the Bayesian gene trees using BUCKy vers. 1.4.4 script (bucky.pl) (Larget et al., 2010), also provided with PhyloNetworks. The subsequent PhyloNetwork analyses and TICR tests were run in Julia (https://julialang.org/) using the PhyloNetworks, PhyloPlots, RCall, DataFrames, and CSV packages. Finally, we plotted the fit of concordance factors for 0–5 putative hybridization (h0–h5) events and ran the TICR test for the tree versus the networks. We also used the sets of gene trees to calculate the frequency with which each terminal was resolved with any other terminal (https://github.com/cdb3ny/Ref-guided_and_Genome_Size).

Octoploid taxa

The parental origin(s) of the five putatively allooctoploid species were investigated to test both the previous hypotheses of hybrid origins (e.g., Govindarajulu et al., 2011a) and to refine our understanding of their respective parentage. These latter analyses follow those described above for paleotetraploid taxa, including (1) using the individual RAxML gene trees (generated as noted above) to report the number of trees with an octoploid allele/sequence sister to each paleotetraploid terminal, (2) running PhyloNetworks for a tree (h0) and a network (h1) for each octoploid plus all paleotetraploid *Leucaena* (also plotting concordance factors [CFs]), (3) running TICR test, and (4) running plastid and mitochondrial *Leucaena* phylogenomic analyses (both with all octoploid species at once and with single octoploid species to differentiate likely maternal/paternal origin(s). The full workflow is available via GitHub (https://github.com/cdb3ny/Ref-guided_and_Genome_Size).

Investigations of hybrid origins using RNA-seq data are susceptible to potential biased gene tree support for one parent over another as a result of (1) differential paralog retention within the genome and (2) the impacts of the reference sequence being used to recover sequences and/or expression bias between parental genes. To help characterize the potential impact of such expression bias, reference assembly bias, and genomic retention of paralogs, we carried out two additional analyses for the octoploids. In the first reanalysis, we ran a BLASTn for each octoploid set of de novo assembled transcripts (rather than reference assembled transcripts) against the reference transcriptome and added up to four of the closest matches for each octoploid in place of their reference assembled sequence. The sister relationships in gene trees for these results were then compared to the sister relationships in the reference-guided approach using the in-house scripts noted above. Second, we used the CDS from a recently published *L. leucocephala* subsp. *leucocephala* chromosomal-scale genome (Chen et al., 2024) and recovered the four closest matching sequences (BLAST matches) from the *L. leucocephala* genome. We then used these sequences to replace the RNA-seq derived sequence for this octoploid. The sister relationships were then compared to RNA-seq analysis for *L. leucocephala* subsp. *leucocephala* (described above).

Plastid and mitochondrial phylogenies

The RNA-seq data for each accession and gDNA-seq data for *L. pueblana* contained sufficient plastid sequence carryover to permit reconstruction of large portions of plastid genes/genomes and modest portions of the mitochondrial genes/genomes. STAR 2.7.1 was used to map each RNA-seq data set to the *L. trichandra* mitochondrial genome (Kovar et al., 2018) and to a version of the plastid genome (Dugas et al., 2015) containing just one copy of the plastid inverted repeat (the IRb). A consensus sequence for each plastid/mitochondrial genome was generated using the approach used above for reference-guided transcript assembly. Organellar trees were generated using raxmlHPC (-f a -T 50 -m GTRCAT -p 12345 -x 12345 -# 100 -s input_matrix -o Entada_abyssinica). Individual phylogenies were generated for (1) paleotetraploids, (2) the paleotetraploids plus all allooctoploids, and (3) the set of five trees including divergent paleotetraploid taxa and one allooctoploid per tree. Scripts and workflows are available via GitHub (https://github.com/cdb3ny/Ref-guided_and_Genome_Size).

Genome size estimation

Available chromosome numbers for *Leucaena* (Table 1) were obtained through the compilation proved by Hughes (1998a). The recovery of reliable DNA content for *Leucaena* species through standard leaf tissue and flow cytometric approaches proved difficult, presumably due to secondary metabolites leading to viscous extracts using both the Otto and Galbraith buffer systems (for buffers see Loureiro et al., 2006). Extractions from the root and hypocotyl of freshly germinated young seeds have been shown to reduce the mucilaginous portion of the extract and provide much cleaner extracts (Han et al., 2024). As for the transcriptome source material, seed from the University of Oxford Department of Plant

Sciences (Oxford, UK) *Leucaena* seedbank was used (Hughes, 1998b), with sampling from multiple trees per population when material was available (accession numbers and herbarium vouchers in Table 1). We chopped 50–100 mg of plant tissue with 50–100 mg of iceberg lettuce (as our standard) in 1 mL of Galbraith buffer. The material was filtered through 40 μm nylon mesh and treated with 2.5 μL of 10 mg/mL RNase A (ca 10 min. on ice) followed by staining with 8 μL of 10 mg/mL propidium iodide. Samples were run on an Accuri C6 flow cytometer (BD Biosciences, San Jose, CA, USA). The genome size of the lettuce material was standardized by comparison to *Arabidopsis thaliana* ecotype Columbia. Genome size estimates for each species represent the average calculated from all estimates of that species.

Patterns of genome size evolution

To investigate the impact of evolutionary relationships among species on genome size, we employed the R packages ape (Paradis and Schliep, 2018), caper (https://cran.r-project.org/web/packages/caper/index.html), geiger (Harmon et al., 2007), picante (Kembel et al., 2010), and BAMMtools (Rabosky et al., 2014) to apply various phylogenetic comparative methods. We utilized the fitContinuous function from the geiger package to test models of trait evolution and infer Pagel's parameters of evolution. We compared the performance of Brownian motion, Ornstein–Uhlenbeck, early burst, and white noise models of trait evolution using the corrected Akaike information criterion (AICc). This function also provided values for Pagel's parameters of evolution, lambda (λ), kappa (κ), and delta (δ) (Pagel, 1999). This test of signal and trait evolution was also performed across each of the 19 bioclimatic variables (see below).

Further investigation of genome size evolution across the phylogeny involved the use of Bayesian Analysis of Macroevolutionary Mixtures (BAMM) (Rabosky et al., 2013) and StableTraits v1.4 (Elliot and Mooers, 2014). The BAMM analysis was run for 10 million generations with priors generated by the credibleShiftSet function in the BAMMtools R package. The results were investigated using the BAMMtools R Package (Rabosky et al., 2014) to infer the most likely number of trait evolution rate shifts and to generate plots of rate through time across the phylogeny, as well as rates for two distinct clades on the reconstructed phylogeny. The StableTraits analysis was done as described by Qiu et al. (2019), using 10 million generations and sampling every 1000 generations with two independent chains (scripts, https://github.com/cdb3ny/Ref-guided_and_Genome_Size).

Analyses of environmental and geographical correlations with genome size

To investigate environmental correlates of genome size evolution, we utilized a curated set of 2472 *Leucaena* species occurrences compiled from georeferenced herbarium and field observation records (Hughes, 1998a). All allooctoploid specimens and those tetraploid accessions with questionable geographical coordinates were removed, leaving 1747 occurrences representing the 19 paleotetraploid species. Linear regression analyses without phylogenetic independent contrast correction between average genome size and each of minimum, maximum, mean, and range of latitude, minimum, maximum, and range of longitude, and range area estimates of each species from the 1747 occurrences were computed in R version 4.3.0 (R Core Team, 2023).

To investigate the impact of climate on genome size evolution, we utilized the 19 bioclimatic predictors established by Hijmans et al. (2005) as well as mean latitude and elevation, both of which can impact measures of seasonality and climate. To account for phylogenetic relationships, we utilized phylogenetic independent contrasts (PIC) (Felsenstein, 1985), with log-transformed bioclimatic variables as the predictor variable and log-transformed genome size as the response variable. A PIC analysis was completed for each of the environmental variables (alpha set to 0.01). These analyses were completed in R version 4.3.0 using the ape package and visualized using ggplot2 (Wickham, 2016).

Pearson's correlation tests were used for pairwise comparisons between all bioclimatic variables using R and the r.corr function in the R package Hmisc (Harrell and Dupont, 2019) and the cor.test function in the stats package (R Core Team, 2023) to account for potential autocorrelation among bioclimatic variables. When autocorrelations were identified, only one variable was kept to build a multiple regression model. Bioclimatic variables that were significant from previous tests and were not autocorrelated were then included in the multiple regression model to determine the relationships of these variables on genome size, both before and after PIC correction.

RESULTS

RNA-seq and DNA-seq data quantity and quality

We recovered high-quality RNA-seq data with similar assembly statistics across samples (Appendix S1). The *L. pueblana* genomic DNA-seq data assembled into an 803 Mbp draft genome from which we extracted 338,000 CDS for comparison.

Paleotetraploid phylogenomics

Reference-guided assembly

A reference transcriptome comprising 28,924 genes (of 80,000 gene models) was recovered by identifying all expressed genes in the reference *L. trichandra* genome with just one isoform and a length of 600–5000 bp. The subsequent mapping of trimmed transcriptome raw reads from each *Leucaena* paleotetraploid (including the gDNA data for

L. pueblana) and outgroup sample recovered 3287–17,323 loci per taxon. A total of 2265 loci, each with no more than 25% missing data for any sample per locus, were included in the phylogenetic analyses of *Leucaena* paleotetraploids. These loci are broadly distributed across all 28 chromosomes (Appendix S2) in *L. trichandra* (including 43–141 loci per chromosome, $\bar{x} = 80.5$) and have a concatenated length of 3.7 Mbp.

De novo assembled loci

The number of *Leucaena* ORFs (≥300 bp) extracted from the de novo assembled transcriptomes varied from 69,310–92,422 ($\bar{x} = 83{,}011$), while outgroup ORFs ranged from 130,099–192,647. After reducing redundancy in the ORFs, the number of ORFs retained in *Leucaena* paleotetraploids and outgroups ranged from 44,481–62,986. A full summary of initial ORFs, BLAST hits, and retained ORFs is presented in Appendix S3. The BLAST database made from *Leucaena* paleotetraploid and outgroup ORFs includes 1,173,453 sequences. Markov clustering of the filtered all-by-all BLAST output produced 48,957 sequence clusters. Of those, 14,381 contained at least one sequence for each RNA-seq represented taxon. Final filtering for one-to-one orthologs resulted in 761,243 aligned bases from 505 loci.

For the de novo assembly analyses including gDNA for *L. pueblana*, the number of ORFs differed considerably from the other paleotetraploid species. After reducing redundancy (337,749 initial ORFs), we recovered 325,828 ORFs for further analysis. The combined BLAST database included 1,501,607 sequences. Hidden Markov clustering of the filtered all-by-all BLAST output produced 80,701 total clusters with 11,555 representing all 23 taxa. Final filtering for one-to-one orthologs resulted in the recovery of 190 loci.

Species trees for paleotetraploid *Leucaena*

ASTRAL and concatenated RAxML analyses, based on 2265 reference-aligned loci (RNA- and DNA-data from all 23 terminals), produced identical estimates of species relationships (Figure 1) and identified strong support for the three clades (Clades 1–3) previously recovered by Govindarajulu et al. (2011b). With the exception of the *L. pueblana* plus *L. matudae* sister-pair relationship (100% BS and 0.86 PP), all nodes have 100% BS and 1.0 PP in the reference-guided analyses. Similarly, ASTRAL and RAxML analyses of the 505 de novo assembled loci, including 22 of 23 paleotetraploid *Leucaena* and outgroup (22 of 23 taxa), are nearly identical to the reference-guided phylogeny. In the de novo approach, there is low support (65% BS/0.81 PP) for the position of *L. lempirana* as sister to *L. salvadorensis* plus *L. shannonii* (Figure 1).

The 190-locus analyses using de novo assembled data, included to ascertain the position of *L. pueblana* in the de novo approach, provided strong support for the position of

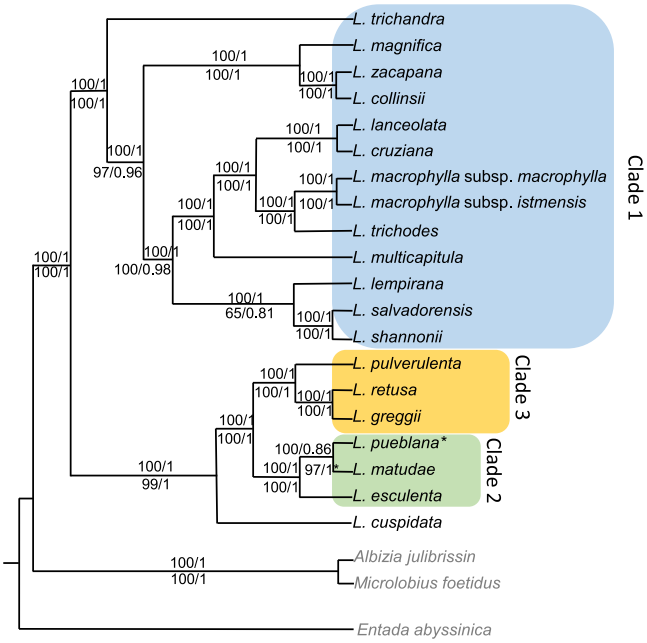

FIGURE 1 Paleotetraploid species tree based on both the reference guided and de novo locus assembly approaches (RAxML and ASTRAL results for each). Support values above the branches correspond to the reference guided RAxML bootstrap and ASTRAL posterior probability, respectively. Support values below the branches are RAxML bootstraps and ASTRAL posterior probabilities for the 505 de novo locus analyses. Clade names 1, 2, and 3 denote the designations developed by Hughes et al. (2002) that are used throughout this manuscript. *Note that *L. pueblana* de novo support is based on the 190-locus analysis (see main text).

L. pueblana as sister to *L. matudae*, consistent with the 2265 referenced-guided locus analyses (see above). However, the reduced gene tree and character space of these analyses resulted in lower support (particularly with ASTRAL PP) and conflicting resolution within Clade 1 (Appendix S4). This set of de novo analyses is not discussed further.

Gene tree conflict among paleotetraploid *Leucaena*

The reference-guided (2265 locus) and de novo (505 locus) data sets generated the same topologies and maximal support across nearly all nodes (Figure 1). Despite the high-degree of traditional support (BS/PP) and topological congruence among paleotetraploid *Leucaena*, we observed short branches and gene tree conflict associated with some nodes, especially within Clade 1 (Appendix S5). We investigated whether these potential conflicts were broadly consistent with the effects of ILS or previously undetected hybridization/introgression events among what have been considered divergently derived taxa (the 19 paleotetraploid species of *Leucaena*) (sensu Govindarajulu et al., 2011b).

Appendix S6 illustrates PhyloNetwork results for 0–4 hybridization events (h0–h4). The sequential plotting of CFs fitted to these increasingly complex networks did not show

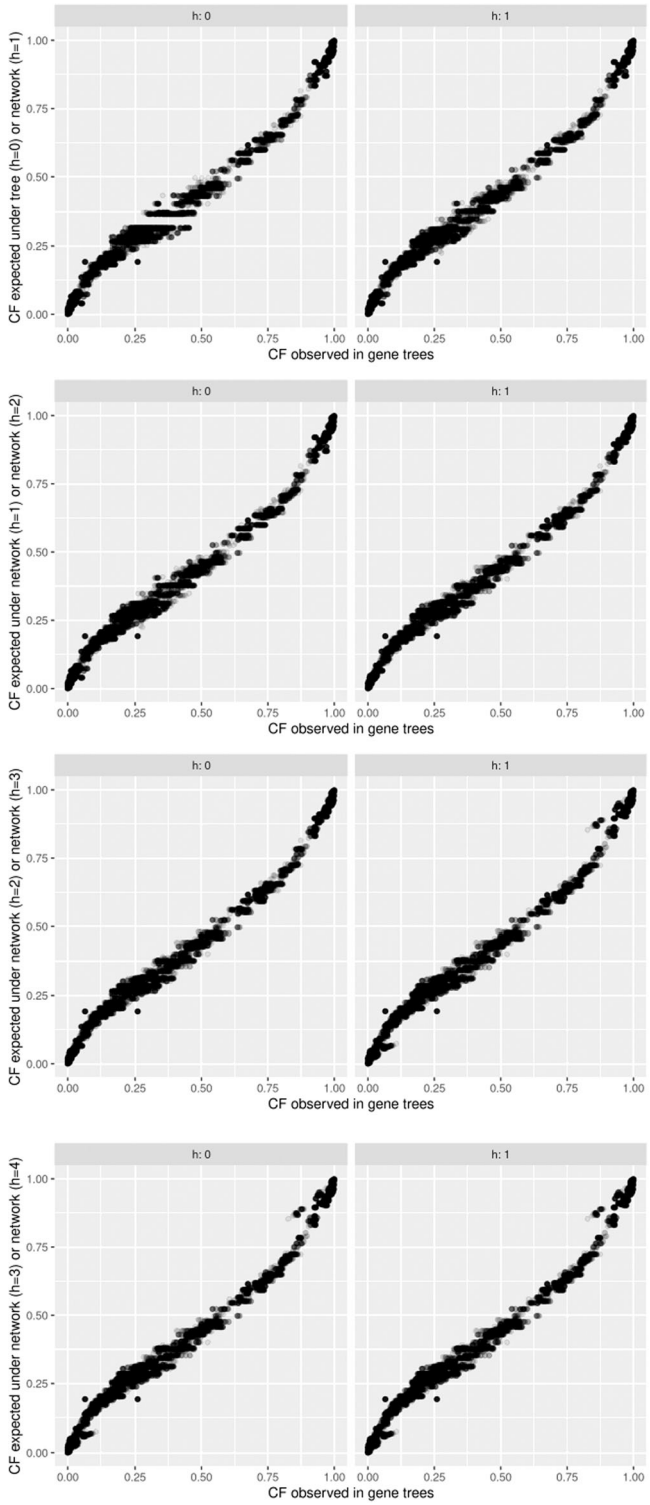

FIGURE 2 Plots for expected versus observed concordance factor fit to Phylonetworks using 0–4 hybridization events (h0–h4). The fit to networks with 1 or more hybridization events is not an improvement over the divergent tree (h0). The TICR test ($P = 0.876$) also rejected the hypothesis that hybridization (h1–h4) better explain the CFs than incomplete lineages sorting (ILS) and h0.

an obvious improvement (Figure 2) and the result of the TICR test ($P = 0.876$) (h0 vs. h1) was not significant (i.e., failed to reject H_0), suggesting that ILS is sufficient to explain the conflict observed along paleotetraploid *Leucaena* gene trees. We further investigated which portions of the species tree are most impacted by ILS. For each species, we identified the number of gene trees that supported the sister relationship between that species and any other species in each of 2265 gene trees (Appendix S7). In most cases, each species shared the greatest number of pairwise relationships with its sister taxon (or clade) identified in the species tree, followed by its next sister taxon or clade with diminishing connections to successively more divergent taxa. These patterns, in stark contrast to those observed for the octoploid taxa (reported and discussed below), are also consistent with general expectations for ILS for paleotetraploid *Leucaena*.

While those sister patterns are supported by gene trees, three regions of the tree merit further mention. First, the observed high support values for relationships within Clade 2, *L. matudae* plus *L. pueblana* as sister to *L. esculenta*, had considerable underlying gene tree conflict (Appendices S5A, S7). There are 678 gene trees supporting the species tree topology (*L. matudae* sister to *L. pueblana*) while 638 trees support *L. esculenta* sister to *L. matudae* and 574 trees support *L. esculenta* sister to *L. pueblana*. The genomic position of loci supporting these alternative relationships are broadly distributed across chromosomes (Appendix S2). Second, *L. trichandra*, which is resolved as sister to all other members of Clade 1 in the species tree, is sister to 8 of 13 other members of Clade 1 (in a minimum of 20 gene trees each), identifying the potential for broad ILS within Clade 1. Third, ILS in Clade 1 is further highlighted by the position of *L. lempirana*, which is sister to *L. salvadorensis* plus *L. shannonii* in the species tree, but shares a number of supported sister relationships among individual gene trees, most notably with *L. magnifica* and *L. multicapitula* (Appendix S7).

Organellar relationships

As expected, mapping of reads across the plastome resulted in a more complete matrix (30% missing data) than for the comparatively gene-poor but repeat-rich mitochondrial genome (84% missing data). Nonetheless, supported higher order relationships derived from phylogenetic analyses of both are largely congruent (Appendix S8) and reflect most of the core nuclear results for paleotetraploid species (Figure 1). The results include the recovery of well-supported (100% BS) Clades 1–3 and the sister relationship of Clades 2 and 3. While the plastid and nuclear trees resolved *L. cuspidata* as sister to Clades 2 and 3, the mitochondrial tree provides weak support for the species as sister to Clade 2 only (73% BS). Within Clade 3, all three genomes recover the same relationships. For Clade 2, both the organellar genomes recovered *L. esculenta* as sister to *L.*

matudae with modest support (87% plastid and 79% mitochondrial BS), which is in partial contrast to the nuclear result, but consistent with the aforementioned ILS issues for the clade. The resolution and support generated for members of Clade 1 are associated with short branches, providing less decisive or unsupported patterns of relationship for both organellar trees.

Octoploid origins

The results from PhyloNetwork trees/networks (Appendix S9A–F), the phylogenetic analysis of combined paleotetraploid and octoploids (Appendix S10), pairwise gene tree support for origins (Appendix S11), and organellar positions for octoploids (Figure 3; Appendices S12, S13) are summarized in Table 2. Below, we describe those findings in more detail.

L. confertiflora

The TICR test was significant for *L. confertiflora*, supporting the PhyloNetworks identified an h1 hybrid origin (Appendix S9A) between (1) *L. cuspidata* and (2) an ancestor sister to all of Clade 1. The plastid/mitochondrial genome trees (Figure 3; Appendices S12, S13) support a Clade 1 parent as the likely organellar contributor. Gene trees support for the origin of *L. confertiflora* is also consistent with these findings (Appendix S11).

L. diversifolia

The TICR test was significant for *L. diversifolia*, supporting the PhyloNetworks h1 hypothesis (Appendix S9B) with *L. diversifolia* sister to (1) *L. pulverulenta* and (2) an ancestor to Clade 1. The mitochondrial and plastid results (Figure 3; Appendices S12, S13) both resolve the octoploid with *L. pulverulenta* and individual gene tree support is consistent with these findings (Appendix S11).

L. leucocephala

The TICR test was significant for both sampled subspecies of *L. leucocephala*, supporting the PhyloNetwork-based

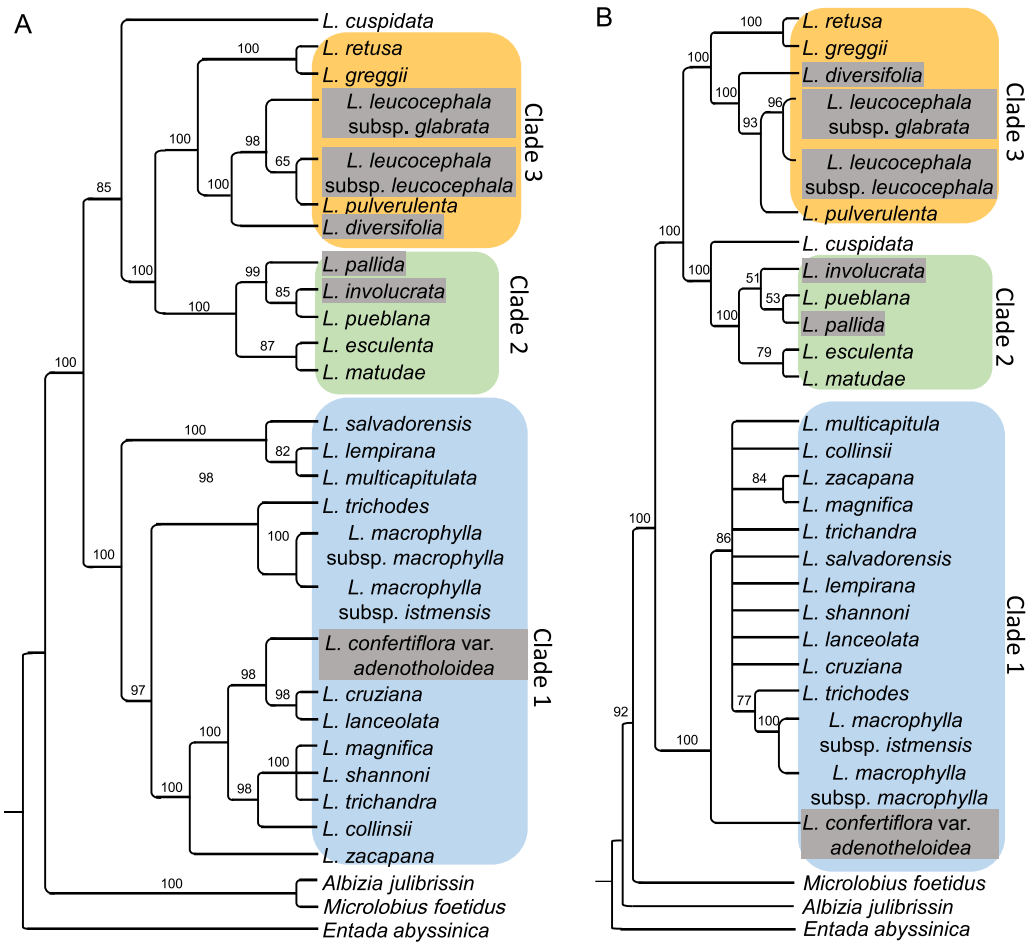

FIGURE 3 Organellar phylogenomic estimates for tetraploid plus alloocotoploid (grey highlight) *Leucaena*. (A) Plastid and (B) mitochondrial consensus trees. Values above each branch represent bootstrap support (BS). Nodes with <50% BS were collapsed.

TABLE 2 Summary of the results used to infer parentage of allooctoploid taxa. Maternal and paternal contributors were inferred based on the majority of data. The terms "singular" vs. "all" for the organellar data refer to phylogenetic analyses with all octoploid species included versus the analysis focusing only on the single octoploid of interest.

Species	TICR *P*, h0 vs h1	PhyloNetworks tree (h0)	PhyloNetworks network (h1)	Plastid all	Plastid singular	Mitochondrial all	Mitochondrial singular	Maternal contributor	Paternal contributor
L. confertiflora var. *adenotheloidea*	5.64E-86	*L. cuspidata*	Sister to *L. cuspidata* and Clade 1	Sister to *L. cruziana* plus *L. lanceolata*	Sister to *L. cruziana* plus *L. lanceolata*	Sister to Clade 1	Sister to Clade 1	*L. cruziana* or *L. lanceolata*	*L. cuspidata*
L. diversifolia	1.17E-23	Sister to *L. cuspidata* plus Clades 2/3	Sister to *L. pulverulenta* and Clade 1	Sister to *L. pulverulenta* plus *L. leucocephala*	Sister to *L. pulverulenta*	Sister to *L. pulverulenta* plus *L. leucocephala*	Sister to *L. pulverulenta*	*L. pulverulenta*	Basal lineage to Clade 1
L. involucrata	0.167	Sister to Clade 2	Sister to *L. pueblana* and Clade 1	Sister to *L. pueblana*	Sister to *L. pueblana*	Sister to *L. pueblana* plus *L. pallida*	Clade 2 unresolved	*L. pueblana*	Basal lineage in Clade 1
L. leucocephala subsp. *glabrata*	3.77E-96	*L. pulverulenta*	Sister to *L. pulverulenta* and *L. cruziana*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. cruziana*
L. leucocephala subsp. *Leucocephala*	5.02E-129	*L. pulverulenta*	Sister to *L. pulverulenta* and *L. cruziana*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. pulverulenta*	*L. cruziana*
L. pallida	0.256	Sister to Clade 2	(1) Sister to Clade 1 and (2) sister to Clade 2	*L. pueblana* plus *L. involucrata*	*L. pueblana*	*L. pueblana*	Sister to *L. esculenta* plus *L. matudae*	*L. pueblana*	Basal lineage in Clade 1

hybrid hypothesis (Appendix S9D, E), which resolved these two taxa as sister to (1) *L. pulverulenta* and (2) *L. cruziana*. The organellar genome trees (Figure 3; Appendices S12, S13) resolved both these taxa as sister to *L. pulverulenta* and the transcriptome gene trees comparisons also support these two species as the putative parents (Appendix S11).

Furthermore, the investigation of *L. leucocephala* subsp. *leucocephala* using loci and alleles derived from the chromosomal-scale genome assembly (Chen et al., 2024), rather than RNA-seq data for this octoploid, supports the hybrid origin of *L. leucocephala* described above and highlights that expression bias (Appendix S14) is being observed in the RNA-seq results for *L. leucocephala* (discussed in more detail below).

L. involucrata and *L. pallida*

The TICR test did not support a hybrid origin for *L. involucrata* or *L. pallida*. These two species were resolved as sister to all of Clade 2, consistent with a Clade 2 or Clade 2 ancestral origin for these species. However, gene tree support (Appendix S11) for these two as potential hybrids is similar to that for the other species noted above and the sum of that support is most consistent with an origin from *L. pueblana* and *L. trichandra* (Table 2), or lineages ancestral to these two taxa.

Genome size variation

We generated 252 genome size estimates, representing an average of 10.8 individuals and 2.5 populations per species. The average genome size for a paleotetraploid species of *Leucaena* ranges from 1.39 to 1.96 pg/2 C (Appendix S15), with *L. zacapana* and *L. retusa* being the smallest and largest, respectively (Figure 4; Appendix S15). The average genome size for these paleotetraploids (1.52 pg/2 C) is concentrated on the lower end of the size distribution (Figure 4) and the clade-based genome size averages are 1.47 pg/2 C, 1.43 pg/2 C, and 1.73 pg/2 C for Clades 1, 2, and 3, respectively.

Intraspecific genome size variation among paleotetraploids is modest within the samples used here (Appendix S15). Two of 20 seedlings from *L. trichodes* (population 61/88) recovered the profiles of octoploid derivatives in a paleotetraploid population. The seedlings in question came from two different mother plants and most seeds from each mother were consistent with paleotetraploid genome sizes. While octoploid seedlings derived from *L. trichodes* in Ecuador are a possibility, a perhaps more likely explanation is a mix-up in seed lots during seed collection or redistribution.

As expected, all the allooctoploid taxa show larger genome sizes, ranging from an average of 2.60 pg/2 C (*L.*

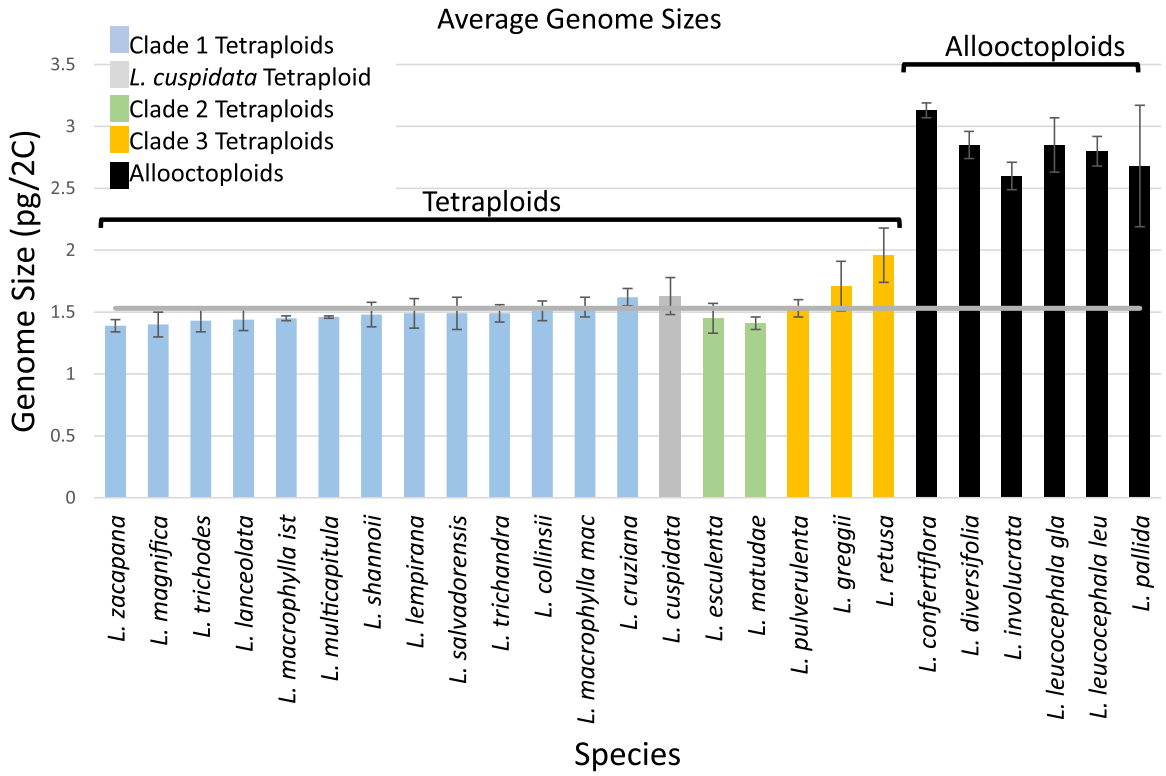

FIGURE 4 Average genome size and standard deviation for 23 of the 24 species of *Leucaena*. The horizontal gray line represents the StableTraits median tetraploid ancestral genome size of 1.53 pg/2 C. Clade 1 species (blue bars), *L. cuspidata* (grey bar), Clade 2 (green bars), Clade 3 (orange bars), and allooctoploids (black bars). *ist*., subsp. *istmensis*; *gla*, subsp. *glabrata*; leu subsp. *leucocephala*; *mac*., subsp. *macrophylla*.

involucrata) to 3.13 pg/2 C (*L. confertiflora*) or 1.30–1.56 Gbp/1 C. There is moderate genome size variation within and between these species (Appendix S15). Each of these taxa retains much of the genome size observed for each of their putative modern progenitor lineages (sensu Govindarajulu et al., 2011a), with a tendency toward downsizing in the allooctoploids. Variation in octoploid genome sizes compared to tetraploid ancestors includes *L. confertiflora* ($\bar{x}$ = 3.13 pg/2 C) relative to *L. cuspidata* ($\bar{x}$ = 1.63 pg/2 C) and *L. trichandra* ($\bar{x}$ = 1.49 pg/2 C); *L. diversifolia* ($\bar{x}$ = 2.85 pg/2 C) relative to *L. pulverulenta* ($\bar{x}$ = 1.53 pg/2 C) and *L. trichandra* ($\bar{x}$ = 1.49 pg/2 C); and *L. leucocephala* ($\bar{x}$ = 2.80–2.85 pg/2 C) relative to *L. pulverulenta* ($\bar{x}$ = 1.53 pg/2 C) and *L. cruziana* ($\bar{x}$ = 1.62 pg/2 C). Without seeds available, the comparative flow-cytometric genome size for *L. pueblana* could not be determined directly, so the values for the allooctoploids *L. involucrata* and *L. pallida* were inferred by considering genomes size of Clade 2 tetraploid relatives of *L. pueblana*. Thus, *L. involucrata* (2.60 pg/2 C) and *L. pallida* (2.68 pg/2 C) have genomes that are smaller than the sum of *L. esculenta*/*L. matudae* ($\bar{x}$ = 1.41–1.45) and *L. trichandra* ($\bar{x}$ = 1.49).

Interspecific genome size evolution

StableTraits identified an ancestral genome of 1.53 pg/2 C (95% CI 1.40–1.67 pg/2 C). Members of Clade 2, as well as 11 of 13 taxa in Clade 1, have average genome sizes lower than the ancestral *Leucaena* estimate, while Clade 3 species all have average genome sizes greater than the ancestral estimate. The rates of genome size change inferred across the phylogeny by StableTraits (Appendix S16) are greatest for *L. retusa* (11.8 pg/mutations per site), *L. cruziana* (7.65), the branch uniting *L. greggii* and *L. retusa* (6.50), *L. macrophylla* ssp. *macrophylla* (4.43), the branch uniting *L. esculenta* and *L. matudae* (4.13), *L. zacapana* (3.53), and *L. magnifica* (2.9). Broadly speaking, the rate changes are associated with increased genome sizes in Clade 3 and decreasing genome sizes in Clades 1 (except *L. cruziana* and *L. macrophylla* subsp. *macrophylla*) and Clade 2.

Estimated values of Pagel's parameters of evolution for genome size were 1 for lambda and kappa, and 0.9407 for delta (Appendix S17). These values suggest that genome size is evolving according to their shared evolutionary history outlined in the phylogeny (λ), with rates varying according to a Brownian motion model across individual branches (κ) and evenly distributed across the tree from root to tip (δ). Based on the AICc values (Appendix S17) from our analysis with the fitContinuous function in R, the Brownian Motion model of trait evolution had the highest support. This AICc support, in addition to a λ value, is also consistent with genome size being impacted by phylogenetic relationships in *Leucaena*.

The relative branch lengths recovered from the BAMM analysis (Appendix S18) largely reflect those

recovered via StableTraits. The most likely number of trait evolution rate shifts was 0 (probability = 0.395), followed by one shift (probability = 0.272), and two shifts (probability = 0.16). In terms of core rate shifts in the phylogeny, only two options were proposed as likely: zero shifts (probability = 0.8787) and one shift (probability = 0.1213) (Appendix S19A), the latter occurring along the stem branch to the *L. retusa* plus *L. greggii* clade (Appendix S19B). *Leucaena retusa* represents the largest genomic expansion relative to the ancestral genome size for the paleotetraploid taxa (a 1.28-fold increase). There was a decrease in the rate of change with time across the entire phylogeny (Appendix S20), with the highest decrease found across Clade 1.

Genome size, latitude/longitude, and environmental variables

Linear regression analyses without PIC correction recovered statistical support for positive correlations between minimum, maximum, and mean latitude and genome size in *Leucaena* (Appendix S21), with mean latitude having the most significance ($P = 0.00182$, adj. $R^2 = 0.412$). Indeed, we see that species with the highest genome sizes mostly inhabit the northern latitudes (Figure 5; Appendix S22). By contrast, latitudinal range, minimum longitude, maximum longitude, longitudinal range, and area were not significant (Appendix S21).

Without a phylogenetic correction, eight of the 19 bioclimatic variables tested showed a statistically significant relationship with genome size in *Leucaena* (Appendix S23). Bonferroni correction for multiple tests with a *P*-value threshold of 0.05 did not return different results from initial tests with a threshold of 0.01. The eight variables include isothermality, temperature seasonality, minimum temperature of the coldest, mean temperature of the driest quarter, mean temperature of the coldest quarter, annual precipitation), precipitation of the wettest month, and precipitation of the wettest quarter. Pearson correlation pairwise comparisons did suggest non-independence between various bioclimatic variables (Appendix S24). Testing these eight bioclimatic variables within a linear regression model without phylogenetic correction shows that mean annual temperature and maximum temperature of the warmest month are significant, suggesting that these two variables help explain most of the variance within average genome size, explaining approximately 60% of the variation in genome size (adj. $R^2 = 0.5992$), with the overall model being significant ($P = 0.01$).

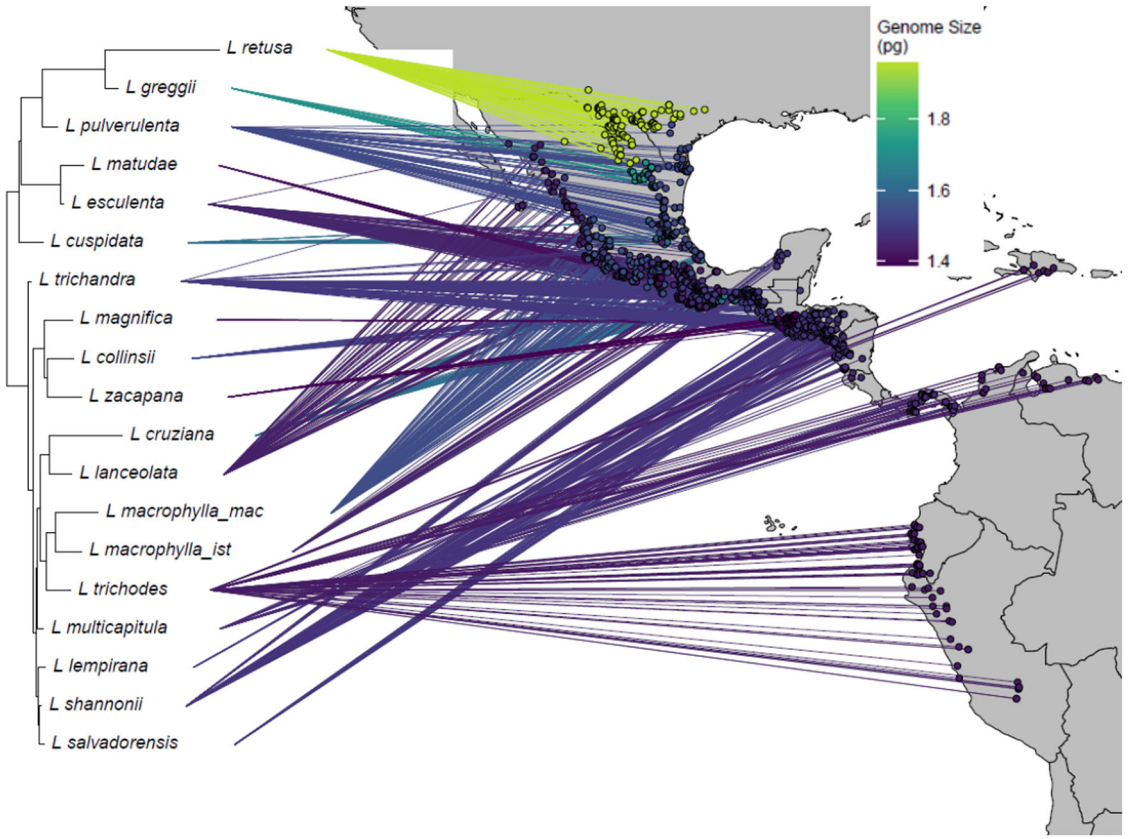

FIGURE 5 Phylogenetic visualization of genome size evolution in *Leucaena*. Genome sizes are in pg/2 C. Species relationships reflect the species-tree presented in Figure 1 with branch lengths scaled to the rates of genome size change along each branch based on the results of StableTraits. *_ist*., subsp. *istmensis*; *_mac*., subsp. *macrophylla*.

However, after accounting for phylogeny via a phylogenetic generalized least squares linear regression model (PGLS), we found that significance was lost for the overall model ($P = 0.15$) and the two variables that were significant are then marginally insignificant (P between 0.05 and 0.1), suggesting that bioclimatic variables do not contribute significantly to an explanation of the overall average genome size across the tree, but rather that evolutionary history drives this change. After phylogenetic independent contrasts (PIC) correction, no bioclimatic variables are significant as contributors to the overall variation in average genome size. The model as a whole, however, is marginally insignificant with a $P = 0.07$, consistent with the results of PGLS.

DISCUSSION

By deploying genome-scale phylotranscriptomic data and a comparative set of new genome size estimates, we have gained insights into divergence, reticulation, and genome size evolution in a complex system of paleotetraploids and octoploids. These findings clarify the evolutionary history of the group and identify factors influencing genome size variation in relation to geographic distribution, environmental conditions, and maintenance of genomic stability in a diverse lineage of neotropical leguminous trees.

Paleotetraploid *Leucaena*: gene tree conflict is mostly attributable to ILS

Prior work on *Leucaena* provided early examples demonstrating the utility of multiple biparentally inherited loci for distinguishing reticulate from cladogenic divergence (Hughes et al., 2007; Govindarajulu et al., 2011a, 2011b; Abair, 2019). The results of those studies were consistent with paleotetraploid species of *Leucaena* being largely derived through allopatric cladogenic speciation events across the neotropics (Govindarajulu et al., 2011a). However, they recovered limited support for core resolution between Clades 1–3 or within Clade 1 and conflicting support for the placement of the potentially early-diverging species *L. cuspidata*.

The species-tree phylotranscriptomic results strongly support the relationships among the major clades and among almost all of the closely related species (Figure 1). Observed gene tree conflict relating to the phylogenetic resolution among 19 paleotetraploid *Leucaena* species was primarily attributable to ILS, a finding consistent with the short branch lengths observed for nodes exhibiting higher levels of conflict (Appendix S5). For the three taxa in Clade 2, the results (Appendices S2, S5, S7) suggest that noise and ILS have effectively eliminated our ability to distinguish patterns of divergence within the group using the available transcriptomic data.

Despite the TICR test rejecting hybridization among paleotetraploid *Leucaena*, in one case, the observed gene tree conflict is harder to explain by ILS. In the pairwise comparison of gene tree support for each paleotetraploid (Appendix S7), *L. magnifica* has >280 gene trees resolving the species with sister taxa *L. collinsii* and *L. zacapana* and >220 gene trees placing it in the clade that includes *L. lempirana*, *L. salvadorensis*, and *L. shannoni*. In the PhyloNetwork trees h1–h5, each hybridization scenario recovered *L. magnifica* as a potential hybrid involving these two clades. Moreover, the plastid data resolved the taxon near *L. shannonii* (Appendix S8), while the mitochondrial data placed it with *L. zacapana*. Furthermore, *L. magnifica* is narrowly endemic in southeastern Guatemala and occurs parapatrically with *L. zacapana* and *L. shannonii* (Hughes, 1998b), a geographic range compatible with possible hybridization between these two species. Despite the TICR test rejecting hybridization as a significant feature on this phylogeny, there are reasons to think that hybridization may have occurred in Clade 1, particularly with regard to the formation of *L. magnifica*. To better address this issue, genetic data across multiple accessions of these species are needed.

Morphological variation tracks phylogenetic history

This is the first phylogenetic study to recover and support the monophyly of the "large-leaflet subgroup" (macrophyllidinous sensu Rupert Barneby for mimosoid legumes) within Clade 1 (*L. macrophylla*, *L. trichodes*, *L. lanceolata*, *L. cruziana*, and *L. multicapitula*), whose leaves have two to six (rarely up to eight) pairs of pinnae and leaflets wider than 1 cm with asymmetric bases. The remainder of the species in Clade 1 have linear to narrow-oblong leaflets, <1 cm wide with strongly asymmetric bases, and more than seven pairs of pinnae. Members of Clade 2 have thin, smooth, metallic bark, while those in Clades 1 and 3 have rough, vertically fissured, thick grayish-brown bark. Furthermore, the Clade 3 sister species pair *L. greggii* and *L. retusa* are the only two species in the genus to have yellow flowers with long-pointed floral bracts (all others have white or pink flowers and round bracts) (Hughes, 1998a).

The phylogenetic results also provide insights relevant to the placement and potential species-level recognition of *L. salvadorensis* and *L. cruziana*. Subsequent to its description, *Leucaena salvadorensis* has been considered conspecific with *L. shannonii* (Zárate, 1984), conspecific with *L. leucocephala* (Sorensson and Brewbaker, 1994), as sister to *L. lempirana* (Harris et al., 1994), and more recently as sister to *L. multicapitula* (Govindarajulu et al., 2011b). Both the species tree and gene tree analyses resolve *L. salvadorensis* as sister to *L. shannonii* with full support (Figure 1). While these two taxa are parapatric (if not sympatric) and have previously been considered the same species, a number of morphological features (Hughes, 1998a) and population-level AFLP analyses (Govindarajulu et al., 2011b) clearly support *L. salvadorensis* as a distinct species in Clade 1.

Govindarajulu et al. (2011b) used a nonsister relationship and highly divergent AFLP data to justify the recognition of allopatric northern and southern populations of *L. lanceolata* as two distinct species: the northern *L. lanceolata* and the

southern *L. cruziana*. By contrast, our current findings strongly support a sister relationship between these two taxa (Figure 1), raising questions about the justification for species-level segregation. Though *L. cruziana* is morphologically cryptic and phylogenetically sister to *L. lanceolata*, their non-overlapping geographic distributions, highly divergent AFLP clusters (Govindarajulu et al., 2011b), and divergent genome sizes (*L. cruziana* 1.62 ± 0.07; *L. lanceolata* 1.44 ± 0.09) (Appendix S15) suggest genetic isolation and divergence consistent with continued recognition as unique taxonomic units.

Octoploid gene expression bias and gene tree conflict are consistent with allooctoploidy

The previously inferred hybrid and parental origin(s) for the allooctoploid species *L. confertiflora*, *L. diversifolia*, *L. involucrata*, *L. leucocephala*, and *L. pallida* (e.g., Govindarajulu et al., 2011a) were reinvestigated to test both the assertion that each is of hybrid origin and to refine our understanding of their respective parentage with genome-scale data. The results of the PhyloNetworks h1 analyses (Table 2; Appendix S9) for all five octoploids are generally consistent with the outcomes based on the handful of genes cloned and analyzed by Govindarajulu et al. (2011a) (also see Figures 1 and 2 therein). These results are also in line with the sum of gene tree support for sister taxon relationships between the allooctoploid and divergently derived paleotetraploid species (Appendix S11).

However, it is notable that the available data, when analyzed using concordance factors and the TICR test, failed to reject the nonhybrid hypothesis (H_0) for the closely related taxa *L. involucrata* and *L. pallida* (Table 2). When investigating the hybrid origins of lineages, it is important to keep in mind that gene trees constructed from transcriptome data are based on expressed genes, which often doesn't present the full genomic background for any locus or set of loci. While such data are commonly used in phylogenetic investigation of species complexes and hybridization (e.g., Barker et al., 2009; Coate et al., 2012; Bombarely et al., 2014; Yang and Smith, 2014; Wen et al., 2015; Boatwright et al., 2018; Zhao et al., 2021), ancestral expression dominance (e.g., maternal vs. paternal, or other forms) can complicate the interpretation(s) of hybrid origin(s) because of uncertainty surrounding the degree to which expressed presence/absence transcripts faithfully represent genomic presence/absence of homeologous gene copies (e.g., Bombarely et al., 2014).

We were able to assess aspects of this issue in two ways. First, a recently published and annotated octoploid genome of *L. leucocephala* subsp. *leucocephala* (Chen et al., 2024) facilitated the comparison of gene tree support recovered directly from the genome of the octoploid versus that observed in our transcriptome data. When we used paleologs pulled from the genome and included those to represent the octoploid (rather than transcript data), the gene tree support for the origin of *L. leucocephala* went from 87% *L. pulverulenta* plus 10.5% *L. cruziana* (for RNA-seq data) to 51% *L. pulverulenta* and 30.5% *L. cruziana* (genomic data) (Appendix S14).

Based on this analysis, we decided to expand the octoploid transcript/potential paleolog space in our general reference-guided analysis by similarly replacing the reference-guided octoploid sequences with de novo assembled transcripts (from Trinity) for each octoploid. Those replacements are akin to phasing hybrids, as is often done with gDNA Hyb-Seq and other approaches (e.g., Nauheimer et al., 2021). The results identified a greater balance in parental contributions in the RNA-seq data in all taxa (Appendix S14). By parsing conflict in putative homeologous representatives, rather than trying to decipher conflict from within a reference-guided transcript, greater clarity was achieved.

Expression presence/absence bias clearly impacted our results, and there is likely greater genomic support for these hybrid origins than is strictly available in these RNA-seq data for all five octoploid species. While the TICR results did not support hybrid origins for *L. pallida* and *L. involucrata* based on reference-guided transcripts, there is substantial support for allopolyploid origins in all five octoploid species based on the available data (Table 2; Appendix S11).

Patterns of subgenome dominance are mixed among allooctoploid *Leucaena*

These findings also illustrate an interesting pattern of genomic diploidization and/or subgenome expression presence/absence bias among the allooctoploid taxa. For the loci used in this study, four of five of these species show strong expression bias toward the maternal contributor (Appendix S11), and the *L. leucocephala* genome-derived loci also display evidence of diploidization via greater genomic retention of maternal homeologous copies (Appendix S14). However, there is a clear paternal subgenome dominance in *L. confertiflora*.

Early work on expression bias in hybrids and allopolyploids noted the potential importance of maternally favored bias in gene expression/retention as a means of balancing cytonuclear interactions (e.g., Barreto et al., 2014). However, more recent studies have highlighted roles of a wider variety of factors in subgenome dominance, including transposable elements, methylation/chromatin accessibility, as well as cytonuclear interactions (reviewed by Alger and Edger, 2020). An empirical study on *Brassica* found limited evidence for maternally biased pressures, consistent with the idea that a variety of genomic features may influence subgenomic dominance (Chalhoub et al., 2014). Furthermore, some polyploids like *Capsella bursa-pastoris* Medik. (Douglas et al., 2015) and soybean (Schmutz et al., 2010; Zhao

et al., 2017), show more limited evidence of biased subgenome dominance. However, investigations of parental genome dominance largely focus on triploid and paleotetraploid species rather than higher ploidy levels. The mixture of patterns of dominance observed among independently derived *Leucaena* allooctoploids warrants further investigation.

Genome size evolution in paleotetraploid *Leucaena* tracks phylogenetic history and may be constrained by duplicate gene retention and expression

Combinations of higher- and lower-level taxonomic studies have increasingly focused on relationships between phylogeny, geographic location (e.g., latitude), environment, and life history as correlates/drivers of genome size variation in plants (e.g., Bilinski et al., 2018; Pellicer et al., 2018; Qiu et al., 2019; Souza et al., 2019; Bureš et al., 2024). Among the many topics of interest in these studies are the roles played by mechanisms (e.g., polyploidy, repetitive elements, transposable elements, diploidization) underlying observed patterns of variation within and among plant species. Characterizing lineages that include closely related species with genome size variation can provide powerful systems to evaluate and identify underlying mechanisms of action (Qiu et al., 2019; Souza et al., 2019). Like *Helianthus* L. (e.g., Qiu et al., 2019), *Leucaena* offers an opportunity to investigate both the impacts of polyploidy and other drivers/mechanisms of genome size change within a single widely distributed genus.

In contrast to another well-studied neotropical lineage of woody legumes, the Caesalpinia group with its stable karyotypes across 225 species (Souza et al., 2019), most of the significant variation in *Leucaena* genome size is clearly associated with ploidy rather than at the single diploid level (Figure 4). Unlike *Helianthus* (dating to 2.4 Mya, Qiu et al., 2019), the Caesalpinia group (dating to 55 Mya, Souza et al., 2019), or an assessment of 337 species of Zingiberaceae (Záveská et al., 2024) with their nearly, 3.6-, 8-, and 16-fold variation in genome size respectively, the 19 comparable divergently related paleotetraploid species of *Leucaena* have a meager 1.4-fold range in mean genome size. By contrast, variation between species at the allooctoploid level in *Leucaena* is more than 2.2-fold.

We detected weak positive relationships between latitude, mean annual temperature, maximum temperature of the warmest month, and the genome size among species of *Leucaena*. A potential relationship between genome size and latitude (Figure 5) and a more temperate climate is consistent with prior historical work on legumes (Stebbins, 1966; Bennett, 1976; Souza et al., 2019) and recent angiosperm-wide findings that genome size is associated with latitude (Rice et al., 2019; Bureš et al., 2024). However, the weak correlation between environmental variables and genome size for *Leucaena* may be attributable to a variety of factors, not the least of which may be ancestral genome size. Given that angiosperm genome sizes have a mean of 10 pg/2 C and a median of 3.2 pg/2 C (Pellicer et al., 2018), it is notable that the ancestral paleotetraploid genome of *Leucaena* was already small (1.52 pg/2 C), and that its descendant lineages have remained comparatively small. This limited range may offer little variation for environmental pressures to act upon in favor of larger genome sizes and may also constrain further genome size reduction. With both ancestral size and variance among genomes in *Leucaena* considerably smaller than some ancestral genome size estimates for some other well studied groups, like *Helianthus* (8.82 pg/2 C) (Qiu et al., 2019), it is likely that factors other than environment are influencing genome size evolution and keeping genomes smaller than might be expected in a lineage with considerable polyploidy (discussed below).

While WGD is a clear driver of the larger sudden shifts in genome size and allooctoploid speciation in *Leucaena* (Figure 4), the phylogenetic investigation of genome size evolution among the 19 paleotetraploid species sheds new light on other elements of genome size variation. These genome sizes have largely evolved in accordance with phylogenetic history (Pagel's test and fitContinuous function), a finding consistent with some other recent studies (e.g., Chrtek et al., 2009; Wickham, 2016; Zahradníček et al., 2018). Only the branch subtending the sister species pair of *L. greggii* and *L. retusa* is highlighted with a possible change in the rate of genome size evolution, leading to an increase in genome sizes. With the exception of the *L. greggii*/*L. retusa* clade, genome sizes mostly declined slightly across the phylogeny, accompanied by a decreasing rate of change over time (Appendix S20). These combined findings confirm that a lineage of 19 divergently evolving species, which retain much of their genic information from their paleotetraploid ancestor (e.g., 80 k+ gene models in *L. trichandra*, C. D. Bailey et al. unpublished data), has experienced limited diploidization in the form of gene loss (Bailey et al., 2024) or DNA content loss (Figure 4; Appendix S15). The available evidence suggests that this paleotetraploid lineage has retained a remarkably stable genome structure and genome sizes through the last 16+ My.

Limited and/or episodic diploidization through gene/DNA loss among paleotetraploid plants is not unique to *Leucaena*, but it is often thought to be associated with recently derived lineages, such as *Arabidopsis suecica* (Fr.) Norrl. ex O.E. Schulz (16,000 ya; Burns et al., 2024) *Tragopogon* L. allopolyploids (<120 ya, (Ownbey, 1950)), and *Capsella bursa-pastoris* (100,000–300,000 ya, (Douglas et al., 2015)). By contrast, paleotetraploids are more typically expected to have undergone extensive diploidization (e.g, Buggs et al., 2012) with well-documented cases in *Arabidopsis*/Brassiceae (Bowers et al., 2003), *Gossypium hirsutum* L. (1–2 mya; Wang et al., 2020), *Helianthus* (Souza et al., 2019), *Oryza* (Wang et al., 2025), and *Zea mays* L. (5–12 mya; Snodgrass et al., 2024).

The papilionoid legume genus *Glycine* L., with a WGD event ca. 13 Mya (Schmutz et al., 2010) followed by subsequent diversification of 26 diploidized species and more recent allopaleotetraploid derivatives (Harbert et al., 2014), offers an interesting comparison to *Leucaena*. In line with our historical view of *Leucaena*, the paleotetraploid species

of *Glycine* have been considered and referred to as diploids (e.g., Doyle et al., 2004; Harbert et al., 2014). However, genome sequencing results from diploid *Glycine* (reviewed by Yuan and Song, 2023) and *Leucaena* confirm that the groups are similar in retaining extensive genic duplication, but that *Leucaena* differs from *Glycine* by retaining considerable structural conservation/synteny among duplicated chromosomes (Bailey et al., 2024; Chen et al., 2024).

Studies in other plant groups have also uncovered examples of paleotetraploids retaining both genomic content and chromosomal structure over longer periods (reviewed by Li et al., 2021), as well as differences in degrees of downsizing among paleopolyploid lineages derived from the same common ancestor (Mandáková et al., 2017). Perhaps the most striking parallel to what we see in *Leucaena* is the recent finding in the Andropogoneae (Poaceae), which includes 12 independent paleopolyploid events with most of the derivative species retaining the polyploid gene sets, chromosome numbers, and synteny across 1–12 My (Stitzer et al., 2025).

The patterns of genome size evolution observed in salmonid fishes (Gundappa et al., 2021), the Andropogoneae (Stitzer et al., 2025), *Sequoiadendron* (Scott et al., 2016) and paleotetraploid *Leucaena* raise questions about the mechanisms controlling and balancing genome structural stability and gene expression in paleopolyploid lineages full of duplicated genes. The observation that many duplicate gene pairs are largely maintained in synteny and expressed in *Leucaena* (C. D. Bailey et al., unpublished data) is consistent with the gene balance hypothesis, which proposes retention of duplicates after WGD to maintain balance in transcript abundance (reviewed by Coate et al., 2016). Stitzer et al. (2025) found a negative relationship between polyploidy and genomic rearrangement, suggesting that polyploidy can be a filter against genomic rearrangements. A negative relationship between polyploidy and genomic rearrangement, in some polyploids, is explainable by the direct impacts that chromosomal rearrangements can have on patterns of expression (e.g., Otto, 2007; Spoelhof et al., 2017; Peng et al., 2022). However, some polyploids like *Glycine* have undergone extensive chromosomal rearrangements without necessarily losing expression balance (Coate et al., 2016), highlighting the diversity of possible responses and outcomes following WGD.

CONCLUSIONS

The majority of gene tree conflict observed for the diversification of paleotetraploid *Leucaena* is consistent with incomplete lineage sorting and cladogenic origins of species, rather than hybrid-associated speciation. By contrast, we recovered substantial evidence supporting the hybrid origins for the allooctoploid taxa. These findings align with previous assertions that a lack of intrinsic reproductive barriers to interspecific hybridization in paleotetraploid *Leucaena*, combined with secondary sympatry among paleotetraploids, created opportunities for multiple independent origins of neoallooctoploids (Hughes et al., 2007; Govindarajulu et al., 2011a). Such results illustrate the complex interplay of external and intrinsic factors impacting speciation, including ancient whole-genome duplication, cladogenesis, secondary contact, and allopolyploid speciation. In this case, all have contributed to the diversity of just 24 species of neotropical legumes. Furthermore, the development of a robust estimate of evolutionary relationships between species of *Leucaena* enabled the investigation of genome size variation within the group.

The results presented herein cause us to reject our a priori hypothesis focused on the impact of environmental gradients on genome size variation in *Leucaena*. Instead, we observed modest genome size differences among the 19 paleotetraploid species. In contrast, broad-scale studies across diverse taxa have identified patterns of genome size reduction from the temperate zone toward the equator (e.g., Stebbins, 1966; Bennett, 1976; Souza et al., 2019; Bureš et al., 2024). We propose that genomic constraints associated with limited diploidization after a paleotetraploidization event at least 16 Mya in *Leucaena* may be overwhelming other factors imposed by environmental variation among the paleotetraploids. Growing genomic evidence from groups like *Leucaena*, Andropogoneae (Snodgrass et al., 2024; Stitzer et al., 2025), *Sequoiadendron* (Scott et al., 2016), and salmonid fishes (Gundappa et al., 2021) highlights the potential impact of limited diploidization in some groups. The complexity of such intrinsic factors may represent confounding elements that mitigate the actions of extrinsic factors like environmental variation on genome size and structure.

It is notable that in allooctoploid *Leucaena*, several of which may be recent neoallooctoploids (Govindarajulu et al., 2011a), the degree of quadruplicate gene retention and genome size variation appears to contrast with patterns of duplicate gene retention observed with the paleotetraploids. These findings indicate that the allooctoploid genomes are exhibiting more pronounced signals of gene and DNA loss compared to their more stable paleotetraploid relatives, a finding also supported by Chen et al. (2024). Thus, *Leucaena* is a valuable system for exploring speciation dynamics and the genomic responses to both intrinsic genomic and extrinsic evolutionary pressures within a framework of small to moderately sized genomes across two distinct levels of polyploidization.

Future work on *Leucaena* and similar groups illustrating the retention of duplicated genomes over millions of years should focus on characterizing patterns of gene expression among ancient paleologs and the factors governing gene expression (e.g., stasis or variation in open chromatin and patterns of methylation) (e.g., Kenchanmane Raju et al., 2023). Retention of duplicated copies as a result of gene expression balance may be a driving force behind limited diploidization in these cases. Such future work may help elucidate the interplay of the intrinsic and extrinsic forces of evolution that create the diversity of life on Earth.

AUTHOR CONTRIBUTIONS

C.D.B., A.A., A.N.E., C.E.H., R.C.R., and A.L. conceived the study. All authors contributed to one or more elements of

data acquisition and/or data analysis. C.D.B., A.A., and A.N.E. wrote the initial manuscript, all authors reviewed the manuscript, and C.D.B., C.E.H, and Y.Y. contributed in-depth revisions for the final version.

ACKNOWLEDGMENTS

We thank Sanjuro Jogdeo, Michael Schatz, Diana Dugas, and Kevin Weitemier for experimental design and/or bioinformatics advice, David Galbraith for assistance setting up and troubleshooting early flow cytometric runs, Jeannette Whitton for constructive suggestions, and two anonymous reviewers for comments and guidance. This material is based on work supported by the National Science Foundation under grant 1238731 (CDB), Swiss National Science Foundation grant 31003A_135522 (CEH). We acknowledge the efforts and work of Utah Valley University's Spring 2022 Plant Systematics (BOT 4200) class (including Anna Chan, Olivia Grant, Tyler Hacking, Regina Keyne, Hayden Oyler, Spencer Stephenson, Asherah Swinson, Erin Taylor, Casey Waite, and Tabitha Weir), under the direction of ANE., for initial ideas, analyses, and collaborations on genome size and environmental correlation work done as part of their Course-Based Undergraduate Research Experience.

DATA AVAILABILITY STATEMENT

The transcriptome data are all available from the NCBI SRA (see Table 1 for accession information). The phylotranscriptomic data sets are available via FigShare (https://figshare.com/s/8e78f6a7e676059d7203) Code for the analyses is available via the GitHub sections cited above. Additional supporting information is found online in the Supporting Information section at the end of the article.

ORCID

Ashley N. Egan https://orcid.org/0000-0001-7803-4444
Madhugiri Nageswara-Rao https://orcid.org/0000-0002-4648-2813
Colin E. Hughes https://orcid.org/0000-0002-9701-0699
Kitti Denson https://orcid.org/0000-0002-6884-1207
Joshua T. Trujillo https://orcid.org/0000-0001-9817-4161
Shannon C. K. Straub https://orcid.org/0000-0001-7506-9043
Jessica P. Houston https://orcid.org/0000-0001-8201-6396
Ya Yang https://orcid.org/0000-0001-6221-0984
Aaron Liston https://orcid.org/0000-0002-3020-6400
Carl E. Hjelmen https://orcid.org/0000-0003-3061-6458
C. Donovan Bailey https://orcid.org/0000-0002-3123-4083

REFERENCES

Abair, A. L. 2019. History of *Leucaena* phylogenetics and a *de novo* transcriptomic approach to resolving diploid relationships. M.S. thesis, New Mexico State University, Las Cruces, NM, USA.

Alger, E. I., and P. P. Edger. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Current Opinion in Plant Biology* 54: 108-113.

Bailey, C. D., S. R. Strickler, E. J. M. Koenen, J. J. Ringelberg, B. Kitti, M. Lopez, O. Iloba, et al. 2024. A reference quality genome assembly and investigation of whole genome duplication in the mimosoid legume *Leucaena trichandra*. XX International Botanical Congress, Madrid, Spain [abstract].

Barker, M. S., H. Vogel, and E. Schranz. 2009. Paleopolyploidy in the Brassicales: Analyses of the *Cleome* transcriptome elucidate the history of genome duplications in *Arabidopsis* and other Brassicales. *Genome Biology and Evolution* 1: 391–399.

Barreto, F. S., R. J. Pereira, and R. S. Burton. 2014. Hybrid dysfunction and physiological compensation in gene expression. *Molecular Biology and Evolution* 32: 613-622.

Bennett, M. D. 1976. DNA amount, latitude, and crop plant distribution. *Environmental and Experimental Botany* 16: 93-108.

Bilinski, P., P. S. Albert, J. J. Berg, J. A. Birchler, M. N. Grote, A. Lorant, J. Quezada, et al. 2018. Parallel altitudinal clines reveal trends in adaptive evolution of genome size in *Zea mays*. *PLoS Genetics* 14: e1007162.

Boatwright, J. L., L. M. McIntyre, A. M. Morse, S. Chen, M.-J. Yoo, J. Koh, P. S. Soltis, et al. 2018. A robust methodology for assessing differential homeolog contributions to the transcriptomes of allopolyploids. *Genetics* 210: 883-894.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30: 2114-2120.

Bombarely, A., J. E. Coate, and J. J. Doyle. 2014. Mining transcriptomic data to study the origins and evolution of a plant allopolyploid complex. *PeerJ* 2: e391.

Bowers, J. E., B. A. Chapman, J. Rong, and A. H. Paterson. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433-438.

Brewbaker, J. L. 1987. *Leucaena*: a multipurpose tree genus for tropical agroforestry. *In* H. A. Steppler and P. K. Nair [eds.], Agroforestry: A decade of development. International Council for Research in Agroforestry, Nairobi, Kenya.

Buggs, R. J. A., S. Chamala, W. Wu, J. A. Tate, P. S. Schnable, D. E. Soltis, P. S. Soltis, and W. B. Barbazuk. 2012. Rapid, repeated, and clustered loss of duplicate genes in allopolyploid plant populations of independent origin. *Current Biology* 22: 248-252.

Bureš, P., T. L. Elliott, P. Veselý, P. Šmarda, F. Forest, I. J. Leitch, E. Nic Lughadha, et al. 2024. The global distribution of angiosperm genome size is shaped by climate. *New Phytologist* 242: 744-759.

Burns, R., A. Kulkarni, A. Glushkevich, U. K. Kolesnikova, F. Kolář, A. D. Scott, and P. Y. Novikova. 2024. Diploid origins, adaptation to polyploidy, and the beginning of rediploidization in allotetraploid *Arabidopsis suecica*. *bioRxiv* 1–30. https://doi.org/10.1101/2024.12.06.627142

Cang, F. A., and K. M. Dlugosch. 2022. Ecological effects of genome size in yellow starthistle (*Centaurea solstitialis*) vary between invaded and native ranges. *bioRxiv* https://doi.org/10.1101/2022.10.25.513778 [preprint].

Cannon, S. B., M. R. McKain, A. Harkess, M. N. Nelson, S. Dash, M. K. Deyholos, Y. Peng, et al. 2015. Multiple polyploidy events in the early radiation of nodulating and nonnodulating legumes. *Molecular Biology and Evolution* 32: 193-210.

Chalhoub, B., F. Denoeud, S. Liu, I. A. P. Parkin, H. Tang, X. Wang, J. Chiquet, et al. 2014. Early allopolyploid evolution in the post-Neolithic *Brassica napus* oilseed genome. *Science* 345: 950-953.

Chen, R., S. Meng, A. Wang, F. Jiang, L. Yuan, L. Lei, H. Wang, and W. Fan. 2024. The genomes of seven economic Caesalpinioideae trees provide insights into polyploidization history and secondary metabolite biosynthesis. *Plant Communications* 5: 1-16.

Cheng, C.-Y., V. Krishnakumar, A. P. Chan, F. Thibaud-Nissen, S. Schobel, and C. D. Town. 2017. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant Journal* 89: 789-804.

Chrtek, J., J. Zahradnícek, K. Krak, and J. Fehrer. 2009. Genome size in *Hieracium* subgenus *Hieracium* (Asteraceae) is strongly correlated with major phylogenetic groups. *Annals of Botany* 104: 161-178.

Coate, J., T. Owens, and J. Doyle. 2012. Transcriptome perspectives on the evolution of allopolyploidy in *Glycine* (Leguminosae) [abstract]. Proceedings of Botany 2012, annual meeting of the Botanical Society of America, Columbus, OH, USA.

Coate, J. E., M. J. Song, A. Bombarely, and J. J. Doyle. 2016. Expression-level support for gene dosage sensitivity in three *Glycine* subgenus *Glycine* polyploids and their diploid progenitors. *New Phytologist* 212: 1083-1093.

Cowley, F. C., and R. Roschinsky. 2019. Incorporating leucaena into goat production systems. *Tropical Grasslands-Forrajes Tropicales* 7: 173-181.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, et al. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21.

Douglas, G. M., G. Gos, K. A. Steige, A. Salcedo, K. Holm, E. B. Josephs, R. Arunkumar, et al. 2015. Hybrid origins and the earliest stages of diploidization in the highly successful recent polyploid *Capsella bursa-pastoris*. *Proceedings of the National Academy of Sciences, USA* 112: 2806-2811.

Doyle, J. J., and J. E. Coate. 2019. Polyploidy, the nucleotype, and novelty: the impact of genome doubling on the biology of the cell. *International Journal of Plant Sciences* 180: 1-52.

Doyle, J. J., J. A. Doyle, J. T. Rauscher, and A. H. D. Brown. 2004. Diploid and polyploid reticulate evolution throughout the history of the perennial soybeans (*Glycine* subgenus *Glycine*). *New Phytologist* 161: 121-132.

Dugas, D. V., D. Hernandez, E. J. M. Koenen, E. Schwarz, S. Straub, C. E. Hughes, R. K. Jansen, et al. 2015. Mimosoid legume plastome evolution: IR expansion, tandem repeat expansions, and accelerated rate of evolution in *clpP*. *Scientific Reports* 5: 1-13.

Elliot, M. G., and A. Mooers. 2014. Inferring ancestral states without assuming neutrality or gradualism using a stable model of continuous character evolution. *BMC Evolutionary Biology* 14: 226.

Felsenstein, J. 1985. Phylogenies and the comparative method. *American Naturalist* 125: 1-15.

Freeling, M., M. J. Scanlon, and J. E. Fowler. 2015. Fractionation and subfunctionalization following genome duplications: mechanisms that drive gene content and their consequences. *Current Opinion in Genetics & Development* 35: 110-118.

Govindarajulu, R., C. E. Hughes, P. J. Alexander, and C. D. Bailey. 2011a. The complex evolutionary dynamics of ancient and recent polyploidy in *Leucaena* (Leguminosae; Mimosoideae). *American Journal of Botany* 98: 2064-2076.

Govindarajulu, R., C. E. Hughes, and C. D. Bailey. 2011b. Phylogenetic and population genetic analyses of diploid *Leucaena* (Leguminosae; Mimosoideae) reveal cryptic species diversity and patterns of divergent allopatric speciation. *American Journal of Botany* 98: 2049-2063.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29: 644-652.

Gundappa, M. K., T.-H. To, L. Grønvold, S. A. M. Martin, S. Lien, J. Geist, D. Hazlerigg, et al. 2021. Genome-wide reconstruction of re-diploidization following autopolyploidization across one hundred million years of salmonid evolution. *Molecular Biology and Evolution* 39: msab310.

Guo, K., P. Pyšek, M. van Kleunen, N. L. Kinlock, M. Lučanová, I. J. Leitch, S. Pierce, et al. 2024. Plant invasion and naturalization are influenced by genome size, ecology and economic use globally. *Nature Communications* 15: 1330.

Han, Y., A. Abair, J. van der Zanden, M. Nageswara-Rao, S. P. Vasan, R. Bhoite, M. Castello, et al. 2024. Transcriptome-wide genetic variations in the legume genus *Leucaena* for fingerprinting and breeding. *Agronomy* 14: 1519.

Harbert, R. S., A. H. D. Brown, and J. J. Doyle. 2014. Climate niche modeling in the perennial *Glycine* (Leguminosae) allopolyploid complex. *American Journal of Botany* 101: 710-721.

Harmon, L. J., J. T. Weir, C. D. Brock, R. E. Glor, and W. Challenger. 2007. GEIGER: investigating evolutionary radiations. *Bioinformatics* 24: 129-131.

Harrell, F., and C. Dupont. 2019. Hmisc: Harrell miscellaneous. R package version 4.2-0. Website https://CRAN.R-project.org/package=Hmisc.

Harris, S. A., C. E. Hughes, R. Ingram, and R. J. Abbott. 1994. A phylogenetic analysis of *Leucaena* (Leguminosae Mimosoideae). *Plant Systematics and Evolution* 191: 1-26.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25: 1965-1978.

Hopkins, K., M. Bowen, R. Dixon, and D. Reid. 2019. Evaluating crude protein concentration of leucaena forage and the dietary legume content selected by cattle grazing leucaena and C4 grasses in northern Australia. *Tropical Grasslands-Forrajes Tropicales* 7: 189-192.

Hughes, C. E. 1998a. Monograph of *Leucaena* (Leguminosae-Mimosoideae). *Systematic Botany Monographs* 55: 1-244.

Hughes, C. E. 1998b. Leucaena: a genetic resources handbook. Tropical Forestry Paper. Oxford Forestry Institute, Oxford, UK.

Hughes, C. E., C. D. Bailey, and S. A. Harris. 2002. Divergent and reticulate species relationships in *Leucaena* (Fabaceae) inferred from multiple data sources: insights into polyploid origins and nrDNA polymorphism. *American Journal of Botany* 89: 1057-1073.

Hughes, C. E., R. Govindarajulu, A. Robertson, S. A. Harris, and C. D. Bailey. 2007. Serendipitous backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences, USA* 104: 14389-14394.

Hughes, C. E., and M. Luckow. 2024. Dichrostachys clade. *In* A. Bruneau, L. P. Queiroz, J. J. Ringelberg [eds.], Advances in legume systematics 14. Classification of Caesalpinioideae. Part 2: Higher-level classification. Phytokeys 240: 269–298.

Hughes, C., J. J. Ringelberg, and A. Bruneau. 2025. Legumes. Current Biology 35: R323–R328.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30: 772-780.

Kembel, S. W., P. D. Cowan, M. R. Helmus, W. K. Cornwell, H. Morlon, D. D. Ackerly, S. P. Blomberg, and C. O. Webb. 2010. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* 26: 1463-1464.

Kenchanmane Raju, S. K., M. Ledford, and C. E. Niederhuth. 2023. DNA methylation signatures of duplicate gene evolution in angiosperms. *Plant Physiology* 192: 2883-2901.

Koenen, E. J. M., C. Kidner, É. R. de Souza, M. F. Simon, J. R. Iganci, J. A. Nicholls, G. K. Brown, et al. 2020. Hybrid capture of 964 nuclear genes resolves evolutionary relationships in the mimosoid legumes and reveals the polytomous origins of a large pantropical radiation. *American Journal of Botany* 107: 1710-1735.

Koenen, E. J. M., D. I. Ojeda, F. T. Bakker, J. J. Wieringa, C. Kidner, O. J. Hardy, R. T. Pennington, et al. 2021. The origin of the legumes is a complex paleopolyploid phylogenomic tangle closely associated with the Cretaceous-Paleogene (K-Pg) mass extinction event. *Systematic Biology* 70: 508-526.

Kovar, L., M. Nageswara-Rao, S. Ortega-Rodriguez, D. V. Dugas, S. Straub, R. Cronn, S. R. Strickler, et al. 2018. PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. *Genome Biology and Evolution* 10: 2501-2517.

Lam, H. M., O. Ratmann, and M. F. Boni. 2018. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. *Molecular Biology and Evolution* 35: 247-251.

Larget, B. R., S. K. Kotha, C. N. Dewey, and C. Ané. 2010. BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* 26: 2910-2911.

Lemin, C., J. Rolfe, B. English, R. Caird, E. Black, S. Dayes, K. Cox, et al. 2019. Comparing the grazing productivity of 'Redlands' and 'Wondergraze' leucaena varieties. *Tropical Grasslands-Forrajes Tropicales* 7: 96-99.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)* 25: 2078-2079.

Li, R., H. Zhu, J. Ruan, W. Qian, X. Fang, Z. Shi, Y. Li, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Research* 20: 265-272.

Li, Z., M. T. W. McKibben, G. S. Finch, P. D. Blischak, B. L. Sutherland, and M. S. Barker. 2021. Patterns and processes of diploidization in land plants. *Annual Review of Plant Biology* 72: 387-410.

Loureiro, J., E. Rodriguez, J. Dolezel, and C. Santos. 2006. Comparison of four nuclear isolation buffers for plant DNA flow cytometry. *Annals of Botany* 98: 679-689.

Mandáková, T., M. Pouch, K. Harmanová, S. H. Zhan, I. Mayrose, and M. A. Lysak. 2017. Multispeed genome diploidization and diversification after an ancient allopolyploidization. *Molecular Ecology* 26: 6445-6462.

Nauheimer, L., N. Weigner, E. Joyce, D. Crayn, C. Clarke, and K. Nargar. 2021. HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Applications in Plant Sciences* 9: e11441.

Otto, S. P. 2007. The evolutionary consequences of polyploidy. *Cell* 131: 452-462.

Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34: 401-437.

Ownbey, M. 1950. Natural hybridization and amphploidy in the genus *Tragopogon*. *American Journal of Botany* 37: 487-499.

Pagel, M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401: 877-884.

Paradis, E., and K. Schliep. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.

Pellicer, J., O. Hidalgo, S. Dodsworth, and I. J. Leitch. 2018. Genome size diversity and its impact on the evolution of land plants. *Genes* 9: 1-14.

Peng, R., Y. Xu, S. Tian, T. Unver, Z. Liu, Z. Zhou, X. Cai, et al. 2022. Evolutionary divergence of duplicated genomes in newly described allotetraploid cottons. *Proceedings of the National Academy of Sciences, USA* 119: 1-14.

Qiu, F., E. J. Baack, K. D. Whitney, D. G. Bock, H. M. Tetreault, L. H. Rieseberg, and M. C. Ungerer. 2019. Phylogenetic trends and environmental correlates of nuclear genome size variation in *Helianthus* sunflowers. *New Phytologist* 221: 1609-1618.

Quinlan, A. R., and I. M. Hall. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841-842.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Website https://www.R-project.org/

Rabosky, D. L., M. Grundler, C. Anderson, P. Title, J. J. Shi, J. W. Brown, H. Huang, and J. G. Larson. 2014. BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods in Ecology and Evolution* 5: 701-707.

Rabosky, D. L., F. Santini, J. Eastman, S. A. Smith, B. Sidlauskas, J. Chang, and M. E. Alfaro. 2013. Rates of speciation and morphological evolution are correlated across the largest vertebrate radiation. *Nature Communications* 4: 1958.

Rice, A., P. Šmarda, M. Novosolov, M. Drori, L. Glick, N. Sabath, S. Meiri, et al. 2019. The global biogeography of polyploid plants. *Nature Ecology & Evolution* 3: 265-273.

Ringelberg, J. J., E. J. M. Koenen, B. Sauter, A. Aebli, J. G. Rando, J. R. Iganci, L. P. de Queiroz, et al. 2023. Precipitation is the main axis of tropical plant phylogenetic turnover across space and time. *Science Advances* 9: 1-16.

Roddy, A. B., G. Théroux-Rancourt, T. Abbo, J. W. Benedetti, C. R. Brodersen, M. Castro, S. Castro, et al. 2020. The scaling of genome size and cell size limits maximum rates of photosynthesis with implications for ecological strategies. *International Journal of Plant Sciences* 181: 75-87.

Román-Palacios, C., C. A. Medina, S. H. Zhan, and M. S. Barker. 2021. Animal chromosome counts reveal a similar range of chromosome numbers but with less polyploidy in animals compared to flowering plants. *Journal of Evolutionary Biology* 34: 1333-1339.

Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, et al. 2010. Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178-183.

Scott, A. D., N. W. M. Stenz, P. K. Ingvarsson, and D. A. Baum. 2016. Whole genome duplication in coast redwood (*Sequoia sempervirens*) and its implications for explaining the rarity of polyploidy in conifers. *New Phytologist* 211: 186-193.

Šímová, I., and T. Herben. 2012. Geometrical constraints in the scaling relationships between genome size, cell size and cell cycle length in herbaceous plants. *Proceedings of the Royal Society, B: Biological Sciences* 279: 867-875.

Smith, S. A., and C. W. Dunn. 2008. Phyutility: a phyloinformatics tool for trees, alignments, and molecular data. *Bioinformatics* 24: 715-716.

Smith, S. A., M. J. Moore, J. W. Brown, and Y. Yang. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* 15: 150.

Snodgrass, S. J., M. Woodhouse, A. Seetharam, M. Stitzer, and M. B. Hufford. 2024. Maize and wild relatives show distinct patterns of genome downsizing following polyploidy. *bioRxiv* 1–64. https://doi.org/10.1101/2024.12.24.630189

Solís-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks: a package for phylogenetic networks. *Molecular Biology and Evolution* 34: 3292-3298.

Sorensson, C. T., and J. L. Brewbaker. 1994. Interspecific compatibility among 15 *Leucaena* (Leguminosae Mimosoideae) species via artificial hybridization. *American Journal of Botany* 81: 240-247.

Souza, G., L. Costa, M. S. Guignard, B. Van-Lume, J. Pellicer, E. Gagnon, I. J. Leitch, and G. P. Lewis. 2019. Do tropical plants have smaller genomes? Correlation between genome size and climatic variables in the Caesalpinia Group (Caesalpinioideae, Leguminosae). *Perspectives in Plant Ecology, Evolution and Systematics* 38: 13-23.

Spoelhof, J. P., M. Chester, R. Rodriguez, B. Geraci, K. Heo, E. Mavrodiev, P. S. Soltis, and D. E. Soltis. 2017. Karyotypic variation and pollen stainability in resynthesized allopolyploids *Tragopogon miscellus* and *T. mirus*. *American Journal of Botany* 104: 1484-1492.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.

Stebbins, G. L. 1966. Chromosomal variation and evolution: polyploidy and chromosome size and number shed light on evolutionary processes in higher plants. *Science* 152: 1463-1469.

Stenz, N. W. M., B. Larget, D. A. Baum, and C. Ané. 2015. Exploring tree-like and non-tree-like patterns using genome sequences: an example using the inbreeding plant species *Arabidopsis thaliana* (L.) Heynh. *Systematic Biology* 64: 809-823.

Stitzer, M. C., A. S. Seetharam, A. Scheben, S. K. Hsu, A. J. Schulz, T. M. AuBuchon-Elder, M. El-Walid, et al. 2025. Extensive genome evolution distinguishes maize within a stable tribe of grasses. *bioRxiv* https://doi.org/10.1101/2025.01.22.633974 [preprint].

Suda, J., L. A. Meyerson, I. J. Leitch, and P. Pyšek. 2015. The hidden side of plant invasions: the role of genome size. *New Phytologist* 205: 994-1007.

Thomas Jr., C. A. 1971. The genetic organization of chromosomes. *Annual Review of Genetics* 5: 237-256.

Van de Peer, Y., E. Mizrachi, and K. Marchal. 2017. The evolutionary significance of polyploidy. *Nature Reviews Genetics* 18: 411-424.

Wang, W., D. Chen, D. Liu, Y. Cheng, X. Zhang, L. Song, M. Hu, et al. 2020. Comprehensive analysis of the *Gossypium hirsutum* L. respiratory burst oxidase homolog (*Ghrboh*) gene family. *BMC Genomics* 21: 1-19.

Wang, X., N. Li, Q. Wang, T.-Y. Lei, J. Zhou, F.-M. Zhang, X.-Z. Liao, et al. 2025. Diploidization in a wild rice allopolyploid is both episodic and gradual. *Proceedings of the National Academy of Sciences, USA* 122: 1-12.

Wen, J., A. N. Egan, R. B. Dikow, and E. A. Zimmer. 2015. Utility of transcriptome sequencing for phylogenetic inference and character evolution. *In* E. Hörandl and M. S. Appelhans [eds.], Next-generation sequencing in plant systematics, 5–91. International Association of Plant Taxonomy, Washington, D.C., USA.

Werth, C. R., and M. D. Windham. 1991. A model for divergent, allopatric speciation of Pteridophytes resulting from silencing of duplicate gene expression. *American Midland Naturalist* 137: 515-526.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer-Verlag, NY, NY, USA.

Yang, Y., and S. A. Smith. 2014. Orthology inference in nonmodel organisms using transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy for phylogenomics. *Molecular Biology and Evolution* 31: 3081-3092.

Yuan, J., and Q. Song. 2023. Polyploidy and diploidization in soybean. *Molecular Breeding* 43: 1-9.

Zahradníček, J., J. Chrtek, Z. Ferreira, A. Krahulcová, and J. Fehrer. 2018. Genome size variation in the genus *Andryala* (Hieraciinae, Asteraceae). *Folia Geobotanica* 53: 429–447.

Zárate, P. S. 1984. Taxonomic revision of the genus *Leucaena* from Mexico. *Bulletin of the International Group for the Study of Mimosoideae* 12: 24-34.

Zárate, P. S. 2000. The archaeological remains of *Leucaena* (Fabaceae) revised. *Economic Botany* 54: 477-499.

Záveská, E., O. Šída, J. Leong-Škorničková, Z. Chumová, P. Trávníček, M. F. Newman, A. D. Poulsen, et al. 2024. Testing the large genome constraint hypothesis in tropical rhizomatous herbs: life strategies, plant traits and habitat preferences in gingers. *Plant Journal* 117: 1223-1238.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* 19: 153.

Zhao, M., B. Zhang, D. Lisch, and J. Ma. 2017. Patterns and consequences of subgenome differentiation provide insights into the nature of paleopolyploidy in plants. *Plant Cell* 29: 2974-2994.

Zhao, Y., R. Zhang, K.-W. Jiang, J. Qi, Y. Hu, J. Guo, R. Zhu, et al. 2021. Nuclear phylotranscriptomics and phylogenomics support numerous polyploidization events and hypotheses for the evolution of rhizobial nitrogen-fixing symbiosis in Fabaceae. *Molecular Plant* 14: 748-773.

SUPPORTING INFORMATION

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

Appendix S1. Summary of the de novo transcript assembly metrics, illustrating high RNA-seq data and assembly quality.

Appendix S2. Chromosomal (0–27) distribution of the loci recovered from the reference-guided analyses.

Appendix S3. Quantification of the number of ORFs recovered from each sample in the *de novo* assembly approach.

Appendix S4. Species trees comparison, de novo approaches including *L. pueblana* gDNA for the reduced 190 locus data set.

Appendix S5. Gene tree conflict and branch lengths.

Appendix S6. PhyloNetwork reconstructions for *h*0–*h*4 potential hybridizations among the tetraploid species of *Leucaena*.

Appendix S7. Summary of the number of gene trees supporting observed sister-pair relationships among tetraploid species of *Leucaena*.

Appendix S8. Tetraploid phylogenomic organellar estimates for the (A) plastid and (B) mitochondrial genomic matrices.

Appendix S9. (A–F) PhyloNetwork and concordance factor analysis for each octoploid species of *Leucaena*; h0 and h1.

Appendix S10. RAxML tree including tetraploid and octoploid terminals derived from the 2265 locus reference guided loci.

Appendix S11. A summary of the number of gene trees in which octoploid terminals were resolved with tetraploid species in the reference-guided RAxML gene trees.

Appendix S12. Individual plastid phylogenomic estimates for all tetraploids plus each singular octoploid species (brown highlight).

Appendix S13. Individual mitochondrial phylogenomic estimates for all tetraploids plus each singular octoploid species (brown highlight) of *Leucaena*.

Appendix S14. Comparison of whole genome vs. transcriptome tree support for the inferred parentage of *L. leucocephala*.

Appendix S15. Flow cytometric derived genome size estimates.

Appendix S16. StableTraits assessment of genome size evolution across the *Leucaena*.

Appendix S17. Results of phylogenetic model testing with fitContinuous for *Leucaena* genome size evolution.

Appendix S18. BAMM marginal shift tree. (A) Proportional branch length depiction of rate changes and (B) a colorized depiction of rate changes.

Appendix S19. Rate shift analysis. (A) Zero rate shifts and (B) one rate shift.

Appendix S20. Estimates for rate change across time, which can involve either increase or decrease in genome size.

Appendix S21. Genome size for each species plotted by geography features.

Appendix S22. Mapping of genome sizes (pg/2C) from smallest to largest across the native range for *Leucaena*.

Appendix S23. Results of the linear regression analyses with phylogenetic correction (PIC).

Appendix S24. Correlation plot showing autocorrelations from pairwise comparisons bioclimatic variables.

How to cite this article: Abair, A., A. N. Egan, B. Bugg, M. Nageswara-Rao, C. E. Hughes, K. Denson, M. Lopez III, H. Sermersheim, J. T. Trujillo, S. C. K. Straub, J. P. Houston, Y. Yang, S. R. Strickler, R. C. Cronn, A. Liston, C. E. Hjelmen, and C. D. Bailey. 2026. Phylotranscriptomics and genome size evolution in *Leucaena* (Fabaceae): Paleotetraploid genomic stability overshadows diploidization and environmental effects. *American Journal of Botany* 113: e70178. https://doi.org/10.1002/ajb2.70178