Received: 31 January 2023 | Revised: 19 April 2023 | Accepted: 28 April 2023

DOI: 10.1002/ppp3.10404

RESEARCH ARTICLE

Plants People Planet PPP Open Access

Genome and evolution of *Prosopis alba* Griseb., a drought and salinity tolerant tree legume crop for arid climates

Wenqian Kong[1] | Min Liu[2] | Peter Felker[3,4] | Mauricio Ewens[4] | Cecilia Bessega[5,6] | Carolina Pometti[5,6] | Jinpeng Wang[7,8,9,10] | Peng Xu[9,10] | Jia Teng[7] | Jinyu Wang[7] | Xiyin Wang[7,8] | Yuannian Jiao[9,10] | Magdy S. Alabady[11] | Françoise Thibaud-Nissen[12] | Patrick Masterson[12] | Xin Qiao[13] | Andrew H. Paterson[1,2,7,11,14]

[1]Department of Crop and Soil Science, University of Georgia, Athens, Georgia, USA

[2]Plant Genome Mapping Laboratory, University of Georgia, Athens, Georgia, USA

[3]Altman Plants, Vista, California, USA

[4]Estación Experimental Fernández Universidad Católica Santiago del Estero (UCSE), Santiago, Argentina

[5]Departamento Ecología, Genética y Evolución (EGE), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

[6]Instituto de Ecología, Genética y Evolución (IEGEBA), CONICET-Universidad de Buenos Aires, Buenos Aires, Argentina

[7]Center for Computational Biology and Genomics, and School of Life Sciences, North China University of Science and Technology, Tangshan, China

[8]National Key Laboratory for North China Crop Improvement and Regulation, Agriculture University of Hebei, Baoding, China

[9]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Science, Beijing, China

[10]University of Chinese Academy of Sciences, Beijing, China

[11]Department of Plant Biology, University of Georgia, Athens, Georgia, USA

[12]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland, USA

[13]Centre of Pear Engineering Technology Research, State Key Laboratory of Crop

Societal Impact Statement

Society anticipates a world in which more food and fiber must be produced at warmer temperatures, which, on the contrary, have greater constraints on the use of water and fertilizers. Tree legumes are often the climax vegetation on the semi-arid and arid lands, covering ~25% of the planet, but the knowledge of their genomes is limited. A draft genome sequence for *Prosopis alba*, a salt and heat tolerant tree that is able to fix nitrogen under harsh conditions, yields new clues about its adaptations. Its rich genetic and ecological diversity makes *Prosopis* well-suited to the investigation of gene functions important to its own greater utilization and/or the improvement of climate resilience of other crops.

Summary

- In arid lands that comprise 41% of the Earth's surface and are growing, tree legumes are often the climax vegetation. Now found in much of arid America, *Prosopis alba* is a salt-tolerant nitrogen-fixing tree native to Argentina.
- We present a *Prosopis alba* genome assembly that is 707 Mb in size, comprising of 6087 contigs of up to 2,077,851 bp in length and of ~359.3 Mb (50.8%) being repetitive elements dominated (20.3%) by long terminal repeats (LTR) retrotransposons. Among a total of 57,572 coding sequences (CDS), 42,475 are putative protein coding genes with median length of 2748 bp. The *Prosopis alba* genome shares the legume-common tetraploidy (LCT) but has not reduplicated, evolving 3.5% and 23.1% faster than *Phaseolus vulgaris* and *Glycine max*, respectively, since the LCT.
- The 50 most expanded gene families include many that are involved in ion homeostasis, perhaps related to drought and/or salt adaptation, together with photosynthetic genes carbonic anhydrase (CA), malate dehydrogenase (MDH) and malic enzyme and gene families involved in circadian clock systems, synthesis of

Genetics and Germplasm Enhancement, Nanjing Agricultural University, Nanjing, China
[14]Department of Genetics, University of Georgia, Athens, Georgia, USA

Correspondence
Andrew H Paterson, Department of Genetics, University of Georgia, Athens, GA, USA.
Email: paterson@uga.edu

brassinosteroids, auxin and gibberellin. Some expanded gene families include members showing molecular signatures of positive selection, as do numerous multi-copy orthologous groups with features associated with pathogen resistance and single-copy orthogroups related to drought and salt stress response, root and root hair development, nodulation, heavy metal detoxification and stay-green habit.

- Coupling genomics-based clues about possible causes of its striking physiological adaptations with rich diversity in ecological context offers means to further investigate functional roles of specific *Prosopis* genes/alleles.

KEYWORDS
colinearity, nitrogen fixation, pathways, photosynthesis, stress

1 | INTRODUCTION

Society anticipates a world in which it must produce much more food and fiber at warmer temperatures, with less water and more costly nitrogen. Semi-arid and arid lands comprise about 41% of the earth's surface with many of the poorest soils, greatest environmental variability, and most politically volatile regions—and are expected to expand (https://www.nature.com/articles/d42473-019-00244-y) (Salem, 1989). Tree legumes are often the climax vegetation in arid lands because of (1) very low soil carbon (C) and organic nitrogen (N), (2) low annual N accretion, and (3) susceptibility of N fixation to drought stress among annual legumes (Felker, 1981).

In the most arid ecozones where cultivation of annual crops is marginal, the typical climax vegetation includes nitrogen fixing woody legumes of the genus *Acacia* in Africa and *Prosopis* in the Indo/Pakistan deserts and arid America. For example, *Prosopis glandulosa* (mesquite) grows on the floor of Death Valley, California, the hottest location on earth (Figure 1, note a person on lower right for scale). *Prosopis* (Caesalpinioideae, subfamily Mimosoid) is a genus of five sections divided into 56 species of trees and shrubs native to the arid regions of North and South America, the Caribbean, Africa, the Arabian peninsula, India and Pakistan (Pasiecznik et al., 2001). With the exception of *Prosopis cineraria* native to India and Pakistan, all of the other useful arborescent species belong to the section Algarobia of North and South America and the Caribbean. The center of diversity, with approximately 22 species, is in Argentina, where the wood is valued for flooring and furniture. With pods that can be as high as 40% sucrose by dry weight, a subset of species such as *Prosopis alba* have been important food sources, distinguishing them from other species like *Prosopis articulata* and *Acacia*. The genus is self-incompatible and diploid except for tetraploid *Prosopis juliflora* (Wojtusik & Felker, 1993). Some North and South American species have fertile hybrids.

A proposed reclassification of *Prosopis alba* as *Neltuma alba* (Hughes et al., 2022) appears to be based on the presence or absence of spines, a monogenic trait (table 9.1, Ewens et al., 2022) that segregates both within and among *Prosopis* species. Burkart (1976) concluded that spines and spiny stipules are useful only at the subgeneric level and that the uniformity of the leaf, floral, and fruit characters do not support splitting *Prosopis* (Hilu et al., 1982). As the first herbarium voucher for New World, *Prosopis* was collected along the Canadian river in northern Texas in 1822 (Johnston, 1962), whereas the first voucher for *Neltuma* was collected in Syria in 1838; here, we have retained *Prosopis* as the genus name.

Nitrogen (N) fixation rates of *Prosopis* species compare favorably with the values for other plants, but occur at leaf air temperatures so high and xylem water potentials so low as to totally inhibit N fixation in Papilionoid legumes. *Prosopis* nitrogen fixation rates of 68 μmol per plant per hour have been measured in 3-m long soil columns, whose tops were close to the roof of the greenhouse in California in air temperatures ranging from 43 to 47°C and leaf water potential from −3.8 to −2.9 MPa (Felker, 1981).

The ability of *Prosopis* to fix N under conditions that other plants do not have is accompanied by, and perhaps related to, its ability to tolerate severe drought and salinity. In Death Valley, at air temperature of 45°C and leaf water potential of −4.5 MPa, maximum light saturated photosynthesis rate of 30 mg $CO_2/dm^2/h$ (18 μmol m^{-2} s^{-1}) was measured for *Prosopis* (Mooney et al., 1977), which was among the highest photosynthetic rates for woody plants (Felker & Clark, 1982). Further, hydroponic trials have identified *Prosopis* genotypes that grew in seawater salinities (Felker, 1981; Rhodes & Felker, 1987; Velarde et al., 2003), and some selections grow well at a pH of 9–10 (Ewens et al., 2012; Singh et al., 1989), which is far more alkaline than traditional crops can survive.

Here, we report a draft genome sequence for *Prosopis alba*, comprised largely of assembled Pacific Biosciences single-molecule real-time (SMRT) sequencing long reads. We reveal genomic features of *Prosopis alba*, including repeated sequence and gene annotation. Evolution of *Prosopis alba* is compared with five other species, with four in the Fabaceae family (*Arachis duranensis*, *Medicago truncatula*, *Phaseolus vulgaris*, and *Glycine max*) and one outgroup (*Vitis vinifera*). Expanding gene families underlying traits with important functions, such as photosynthetic rate and nitrogen fixation, have been elucidated. This, to our knowledge, was the first assembled genome in the

FIGURE 1 *Prosopis glandulosa var Torreyana* in situ, in Death Valley, California (USA). Photo taken by Dr Peter Felker.

Mimosiodeae subfamily, which will provide rich resources to studies about many other species in this subfamily and may provide insight into the unique features of *Prosopis alba*.

2 | MATERIALS AND METHODS

2.1 | DNA extraction and library preparation

Leaf samples of an elite clone (B1F8T4) that had a 14 year average pod production of 3.75 ton/ha in a replicated trial (Ewens et al., 2022), were frozen at −80°C and lyophilized for 48 h. Genomic DNA was extracted following cetyltrimethylammonium bromide (CTAB)-based method from the lyophilized leaf sample based on Aljanabi et al. (1999).

2.1.1 | Pacbio long reads libraries and sequencing

The concentration, purity, and size distribution of the DNA were determined using the Qubit, Nanodrop, and Fragment analyzers, respectively. The SMRTbell Libraries were created utilizing the SMRTbell Express Template Preparation Kit 2.0 (Cat #100-939-900) and PacBio's template preparation protocol (PN 101-693-800 Version 01). Briefly, the DNA was cleaned using AMPure beads and sheared to the proper size based (≥30 kbp). Overhangs on single strands were removed, and the DNA was damaged–repaired and A-tailed. The hair-pin adapters were ligated to the repaired and A-tailed DNA. The final library was size-selected (≥30 kbp) using BluePippin. The final SMRTbell library was quality-assessed and sequenced on the PacBio's Sequel system following the manufacturer's procedures (SMRTcell v3, Cat# 101-531-000), Sequencing Kit 3.0 (Cat# 101-597-800), and Sequel binding and internal control kit 3.0 (Cat# 101-626-600). The sequencing movie time was 10 h, and six SMRT cells were used.

2.1.2 | Illumina short reads library and sequencing

A short-read library was prepared from the same DNA used in the long-read library preparation using the KAPA Hyper Prep Kit (Cat# KK8504), following the manufacturer protocol. During the library prep, the DNA is fragmented by acoustic shearing with Covaris before end repair, A-tailing, and barcoded-adaptor ligation. The library was cleaned with SPRI beads before being QC'ed by Qubit, qPCR (KAPA Library Quantification Kit Cat# KK4854), and Fragment Analyzer. The final library was sequenced on the Illumina NextSeq 500 sequencer.

2.2 | Assembly

De novo genome assembly used the Celera Assembler Canu v1.7 (Koren et al., 2017, 2018), which is designed for long-reads generated by PacBio or Oxford Nanopore instruments. Default parameters were used, with genome size set to be 700 Mb and corOutCoverate to be 100. Illumina reads were generated to correct for errors in the raw reference sequences. Read alignment used the Bowtie-TopHat2 using the following parameters: --read-mismatch 1, --read-gap-length 1 -p 8, --mate-inner-dist 300, --no-discordant, and --no-mixed (Kim et al., 2013; Langmead & Salzberg, 2012). Draft sequence correction and improvement used pilon using the following parameters: --diploid, --threads 20, --changes, --tracks, --fix bases, and --mindepth 4 (Walker et al., 2014).

2.3 | Annotation

The annotation pipeline followed the NCBI Eukaryotic Genome Annotation Pipeline (Thibaud-Nissen et al., 2013). The de novo annotation of ASM479914v1, NCBI *Prosopis alba* Annotation Release 100, was performed by the NCBI Eukaryotic Genome Annotation Pipeline as previously described for other genomes (Pruitt et al., 2014; Rhie et al., 2021). The annotation of protein-coding and long non-coding genes was derived from the alignments of protein and RNAseq reads to the WindowMasker-masked genome (Morgulis et al., 2006). A total of 488.6 million RNA-Seq reads sequenced from the root and leaf libraries were retrieved from Sequence Read Archive (SRA) and tentatively aligned to the assembly using Splign (Kapustin et al., 2008). Similarly, 48,147 *Arabidopsis thaliana* and 8143 Fabaceae known RefSeq (NP_ prefix) and 41,859 Fabaceae GenBank proteins were retrieved from Entrez and were aligned to the genome using ProSplign. A total of 42,275 protein-coding genes and 4029 long non-coding genes were annotated on ASM479914v1 by Gnomon based on these alignments (see details in Rhie et al. 2021), including 9157 with more than one alternatively-spliced transcript. Short non-coding RNAs, rRNAs, and tRNAs were derived from RFAM models searched with Infernal cmsearch and tRNAscan-SE, respectively.

The vast majority of coding transcripts (78%) are fully supported by alignments. Only 16% lack support for 5% or more of their length and include a span predicted ab initio using a hidden Markov model. The gene set was estimated to be 91.8% complete by BUSCO version 4.0.2 (Seppey et al., 2019) with the fabales_odb10 marker set (single copy: 47.1%, duplicated 44.7%, fragmented: 0.05%, missing: 7.6%).

2.4 | Repeat elements analysis

RepeatModeler (Flynn et al., 2020) was applied to a total of 1584 sequences (>100 kb) to create a database for repeated elements. RepeatMasker (http://www.repeatmasker.org/) was then used to summarize the categories of those elements.

2.5 | Gene colinearity inference

We compared the gene colinearity within and between *Prosopis alba* and four well-assembled genomes (*Arachis duranensis*, *Medicago truncatula*, *Phaseolus vulgaris*, and *G. max*) and an outgroup (*V. vinifera*) using annotated gene protein sequences as input (Table S1). First, potential homologous genes within a genome or between different genomes were identified by performing the BLASTP program (Altschul et al., 1990). Alignment parameters (E-value < 1e-5 and Score > 100) were strictly set to exclude more-diverged homologous genes and remove short-matched gene pairs. Then, homologous pairs and the location information of genes on chromosomes were used as input for ColinearScan to infer homologous gene colinearity (Wang et al., 2006). The key parameter, the maximum gap, was set to be 50 intervening genes, as adopted in previous genomics research (Wang et al., 2015). Gene families with 30 or more copies in a genome were removed from the analysis to avoid false positive alignments.

2.6 | Genome homologous structure analysis

Homologous genome structure was visualized using homologous gene dot plots within and between genome(s). The median Ks values were estimated to infer colinear homologs in each block produced by different events. Sequence divergence between different paralogous and/or orthologous colinear regions was also shown in dot plots. Using the reference/outgroup genome *V. vinifera*, we conducted a two-way comparison and provided homologous dot plots. These dot plots can help distinguish orthologous and outparalogous correspondence between genomes, identify paralogy depth within a genome, and infer whether one or more polyploidization(s) occurred after its split with the reference.

2.7 | Ks calculation, distribution fitting, and correction

Synonymous nucleotide substitutions at synonymous sites (K_s) were estimated using the Nei–Gojobori approach (Nei & Gojobori, 1986), implemented using the Bioperl Statistical module (Stajich et al., 2002). The peak and distribution of Ks values associated with particular genomic features (such as homologous segments) or the genome as a whole provided insight into the antiquity of the corresponding evolutionary event. More detail about the methods can be found in previous publications (Wang et al., 2017, 2018).

2.8 | Positive selection

Two legume genomes (Table S1), *Senna tora* (the most closely related species to *Prosopis alba* with a publically available genome sequence) and *Medicago truncatula* (a well-annotated legume), and one outgroup

V. vinifera were used for inferring positive selection in *Prosopis alba*. Duplicated sequences within orthogroups are removed. As the target species is *Prosopis alba*, orthogroups without *Prosopis alba* sequences were excluded from this analysis. Clustal Omega (Sievers & Higgins, 2018) was used to conduct multiple comparisons for protein sequences in each orthogroup. Then, PAL2NAL (Suyama et al., 2006) was used to construct CDS alignment based on the protein alignment. FastTree was used to construct approximately maximum likelihood phylogenetic tree from the protein alignment. Then, HyPhy (Pond & Muse, 2005) aBSREL (adaptive Branch-Site Random Effects Likelihood) (Smith et al., 2015) was used to detect the positive selection of the entire tree. *Prosopis alba* genes that are positively selected are isolated from the result.

2.9 | Gene family analysis

A total of 11 species with well-annotated genomes were selected, and 33,913 orthogroups (groups, hereafter) were constructed using OrthoMCL with default parameter settings. The orthogroups are different from those derived from the "Positive selection" sections because different taxa best met the experimental criteria. To distinguish the two systems, orthogroups derived from positive selection are named as "OG" followed by a 7-digit number; whereas, the orthogroups derived from the gene family analysis are named as "orthogroup".

Prosopis alba and 10 other species representing major lineages in Viridiplantae, including *Amborella trichopoda*, *Oryza sativa*, *Musa acuminata*, *Aquilegia coerulea*, *Papaver somniferum*, *Solanum tuberosum*, *V. vinifera*, *Populus tristis* Fisch. (syn. *Populus trichocarpa* Torr. & Gray), *G. max* and *Arabidopsis thaliana*, were included for comparative genomics analysis. Orthogroups were constructed using OrthoMCL v2.0.9 (http://orthomcl.org/orthomcl/) (Li et al., 2003) with parameters E-value < 1e-5, alignment coverage >40%, and inflation value 1.5. We compared the shared and specific groups among the *Prosopis alba*, *Populus trichocarpa*, *Arabidopsis thaliana*, and *G. max* species. A Venn diagram was constructed to show the number of shared and specific groups.

Single-copy gene families were selected for constructing the species trees using the maximum likelihood method. We first performed multiple sequence alignment by Muscle (Edgar, 2004) for each single-copy gene orthogroup. Sequences were then forced to fit the amino-acid alignments using PAL2NAL (Suyama et al., 2006) and then trimmed by removing poorly aligned regions using TRIMAL 1.2 (Capella-Gutierrez et al., 2009). Supermatrix analyses were performed on concatenated nuclear gene alignments. Maximum-likelihood analyses were conducted using RAXML (Stamatakis, 2014), searching for the best MLT with the GTRGAMMA model, which represents an acceptable trade-off between speed and accuracy (Stamatakis, 2014).

Dating of the speciation were performed using the r8s program (Sanderson, 2003) with penalized likelihood, setting the differentiation node of eudicots to 125 mya. Then, we used CAFE (De Bie et al., 2006) to analyze the number of expansion/contraction of gene families. For the gene family evolution analysis, one or more gene families with more than 100 copies were filtered out to reduce parameter prediction errors. The modeling of gene family size was performed by CAFE (v4.2) using 17,666 orthogroups, and the gene birth and death rate was estimated. Orthogroups with family-wide $p < 0.01$ were defined as rapidly evolving gene families; whereas, the Viterbi p (<0.01) was used to identify branches with gene families significantly expansions/contractions compared with their last common ancestor.

3 | RESULTS

3.1 | Genome composition

We generated a total of 3,749,426 Pacbio reads and 130,860,979 pair-end Illumina reads that are used for correction. The assembled sequenced *Prosopis alba* genome is 707,161,903 bp in length, containing a total of 6087 contigs with the largest contigs of up to 2,077,851 bp in length (Table 1), with an N50 contig length of 248,703. Past predictions of *Prosopis alba* genome size vary from 392 to 1252 Mb (Anand et al., 2017; Bukhari, 1997; Paterson et al., 2008). Our result is slightly smaller than what was predicted in Bukhari (1997). De novo analysis reveals ~359.3 Mb of repetitive elements (50.81% of the genome), with LTR being the most predominant elements (20.3%, Table 2). A large number of repeated elements were unclassified, possibly unique to the *Prosopis alba* genome. The NCBI assembly number is ASM479914v1, and the Whole Genome Shotgun (WGS) project number is SMJV01 (Accession number SMJV00000000).

Of the 52,298 genes and pseudogenes annotated using the NCBI pipeline (Table 1), 42,475 are putative protein coding genes. The median length of the genes is 2748 bp, with an average of five exons per transcript. A total of 57,572 CDS were found with a median length of 1134 bp. A total of 5947 non-coding RNAs were detected.

3.2 | Genome structure and evolution

A total of 458 contigs, each having 20 or more genes (totaling 19,397 genes, 45.89% of those annotated; Table S2), were used in the comparative genomics analysis. Genomic data of *V. vinifera*, *Arachis duranensis*, *Medicago truncatula*, *G. max*, and *Phaseolus vulgaris* was downloaded from either NCBI or Joint Genome Institute (JGI) (Table S1).

Dot-plots show that the colinearity ratio between *V. vinifera* and *Prosopis alba* is 3:6 (Figure 2a), indicating that *Prosopis alba* has experienced additional genome duplication following the eudicot specific hexaploidy (Bowers et al., 2003; Jaillon et al., 2007). For example, regions of *V. vinifera* chromosomes 3, 4, and 18 each correspond to two colinear regions in *Prosopis alba*, with an average Ks value of ~1.1. (Figure 2a). The colinearity ratio of *Prosopis alba* and *Phaseolus*

vulgaris is 2:2, with average Ks of 0.826. Similarly, the ratios of *Prosopis alba* versus the genomes of *Arachis duranensis* and *Medicago truncatula* are also 2:2, collectively indicating that *Prosopis alba* shared the legume-common tetraploidy (LCT) but has not reduplicated since that event. The colinearity ratio of *Prosopis alba* versus *G. max* is 2:4, reflecting that *G. max* experienced lineage specific tetraploidy (abbreviated SST) ~13 mya (Figure 2c), after the LCT (Figure 2d).

TABLE 1 *Prosopis alba* genome assembly and gene annotation information. The assembled *Prosopis alba* genome is 707,161,903 bp in length, containing a total of 6087 contigs, with the largest contig of up to 2,077,851 bp in length, and with an N50 contig length of 248,703. Of the 52,298 genes and pseudogenes annotated using the NCBI pipeline, 42,475 are putative protein coding genes. The median length of the genes is 2748 bp, with an average of five exons per transcript. A total of 57,572 CDS were found, with a median length of 1134 bp. A total of 5947 non-coding RNAs were detected. More details can be found at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Prosopis_alba/100/.

Genome assembly	
Number of contigs	6087
Total length of contigs (bp)	707,161,903
N50 of scaffold	248,703
Longest scaffold	
Guanine or cytosine (GC) content (%)	34.31
Gene annotation	
Number of genes and pseudogenes	52,298
Number of protein coding genes	42,475
Median length of genes	2748
Mean number of exons per transcript	5
Number of transcripts	57,572
Median length	1134
Number of non-coding RNA	5947

Abbreviation: CDS, coding sequence; NCBI, National Center for Biotechnology Information.

Ks distributions were plotted with colinear genes within *Prosopis alba* and colinear genes between *Prosopis alba* and the other five species. Comparative genome analysis has detected at least five syntenic chromosomal regions including a total of 347 colinear regions and 2393 pairs of homologous genes. The number of syntenic genes of *Prosopis alba* is fewer than that in other legumes, perhaps reflecting less complete assembly or because of its early divergence from other legumes. For example, *Medicago truncatula* has a total of 501 colinear regions, including 4474 pairs of homologous genes; *Arachis duranensis* has a total of 511 colinear regions, including 4124 pairs of homologous genes, and *Phaseolus vulgaris* has a total of 564 colinear regions with 5431 pairs of homologous genes (Table S3 and S4). We detected a total of 1064 colinear regions, including 11,238 pairs of homologous genes between *Prosopis alba* and *V. vinifera*. The highest number of colinear regions between *Prosopis alba* and another species (2607 including 27,285 pairs of homologous genes) are with *G. max*, which experienced SST after LCT.

The peak Ks value of homologous genes within *Prosopis alba* is 0.772 (±0.106). The peak Ks of homologous genes related to LCT is 0.746 (±0.082) within *Phaseolus vulgaris* and 0.627 (±0.105) within *G. max*. The difference in the Ks peak values suggested that *Prosopis alba* genes have evolved 3.5% and 23.1% faster than *Phaseolus vulgaris* and *G.max*, respectively, following the LCT (Figure 3a and Table S5). We have corrected the evolutionary rates and infer that *Prosopis alba* branched off from the other legumes approximately 57.9 mya, shortly after LCT (Figure 3b,c).

3.3 | Positive selection

Among all orthogroups comparing *Prosopis alba*, *Senna tora*, *Medicago truncatula*, and *V. vinifera* genes, a total of 18,409 orthogroups containing *Prosopis alba* genes were included in the analysis of signatures of positive selection. For each orthogroup, an approximate maximum-likelihood tree was constructed, with all the genes and nodes tested for potential positive selection using likelihood ratio tests for episodic diversifying positive selection (Smith et al., 2015). Holm–Bonferroni correction was used for multiple testing of genes and nodes.

	Number of elements	Length occupied	Percentage of sequence
SINEs	4908	687,461	0.10%
LINEs	24,892	11,868,513	1.68%
LINE1	14,385	9,327,133	1.32%
LINE2	1,598	135,701	0.02%
LTR elements	120,136	143,692,450	20.32%
DNA elements	90,742	43,448,074	6.14%
hAT-Charlie	1305	130,605	0.02%
Unclassified	659,443	159,611,658	22.57%
Total interspersed repeats		359,308,156	50.81%

Abbreviation: LINE: long interspersed nuclear element; LTR: long terminal repeat; NCBI, National Center for Biotechnology Information; SINE: short interspersed nuclear element; WGS, Whole Genome Shotgun.

TABLE 2 Repeat elements in the *Prosopis alba* genome. De novo analysis reveals ~359.3 Mb of repetitive elements (50.81% of the genome), with LTR being the predominant elements. The NCBI assembly number is ASM479914v1, and the WGS project number is SMJV01 (Accession number SMJV00000000).

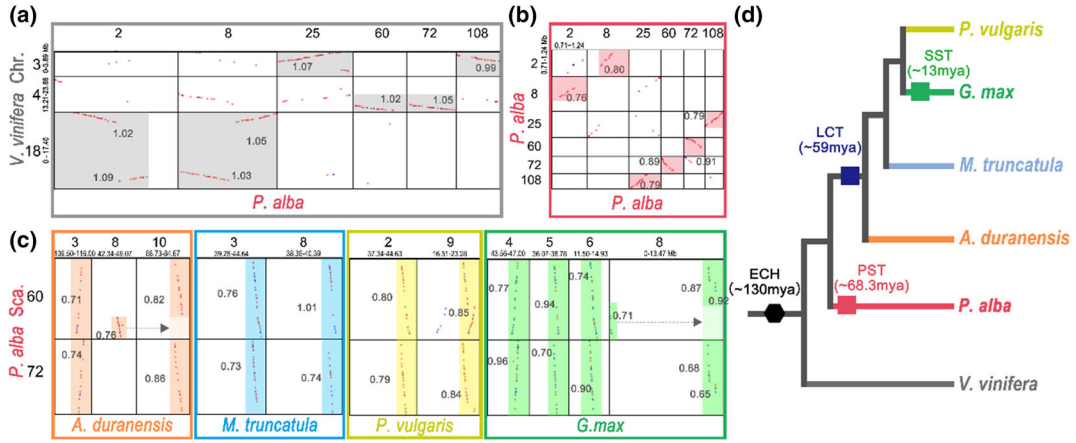

FIGURE 2 Evolution of *Prosopis alba*. (a) Intergenomic colinearity analysis (dot plot) between *Prosopis alba* vs *Vitis vinifera*. (b) Intragenomic colinearity analysis within *Prosopis alba*. (c) Intergenomic colinearity analysis comparing *Prosopis alba* to *Arachis dureanensis*, *Medicato truncatula*, *Phaseolus vulgaris*, and *Glycine max*. (d) Phylogenetic tree of the six species. Colinear analysis suggests that *G. max* experienced lineage specific tetraploidy (SST) ~13 million years ago, after the legume common tetraploidy (LCT) (d). Examples of homologous gene dot-plots of *Prosopis alba* vs *Vitis vinifera* (a), within *Prosopis alba* (b) and four other genomes, i.e., *Arachis duranensis*, *Medicago truncatula*, *Phaseolus vulgaris*, and *G. max* (c). A scaffold of *Prosopis alba* is compared with chromosomes of the other five genomes in Megabase (Mb), and their synonymous mutation rate (Ks) values are illustrated. Best matched genes are shown in red dots; the second best matched genes are shown in blue, and the other matched genes are shown in grey. Best matched chromosomal regions are highlighted. Arrow shows the chromosomal broken regions during evolution. Phylogenetic tree of the six species is shown in panel d, with eudicot hexaploidy shown in black hexagon, LCT in blue square, and SST in green square.

A total of 3553 orthogroups containing at least one *Prosopis alba* gene (totaling 5456 genes) showed molecular evidence of positive selection, with multiple- and single copy orthogroups reflecting distinctive gene functions (Dataset S1). With very few exceptions, orthogroups in which multiple *Prosopis alba* genes showed molecular signatures of positive selection showed features associated with disease resistance. In the most extreme cases, OG0006703 and OG0006733, each of the eight genes in the group(s) showed strong evidence of positive selection, and the group was *Prosopis alba*-specific, lacking any genes from the other test species. Such groups tended to also be comprised of members at proximal locations in the genome—for example, of the eight members of OG0006703, four were on contig NW_021636129.1 and the other four on contig NW_021636085.1. Whereas the *Prosopis alba* assembly was not chromosome-scale, these genes within the orthogroup could all be located to chromosome 2 on *Senna tora*. Likewise, the eight members of OG0006733 were on *Prosopis alba* contigs NW_021636652.1 (5), NW_021633702.1 (2), and NW_021632937.1 (1), all corresponding to the chromosome 5 of *Senna tora*. This suggests that diversifying positive selection (Smith et al., 2015) acting on these genes was coupled to the selection for an increased copy number, which occurred by proximal duplication.

Among the 1353 single copy orthogroups, a total of 99 contained *Prosopis alba* genes that showed DNA-level signatures of positive selection (Dataset S1). Among these, 22 were classified as uncharacterized proteins, many of which may be undetected retrotransposons, although two (At4g37920 and At5g39865) that match *Arabidopsis* genes exemplify that some are likely to be genes of interest. Another 16 appear to be nuclear-encoded genes functioning in organelles (13 in chloroplast and three in mitochondria). Twelve were transcription factors (one chloroplastic), including genes implicated in multiple stress tolerances such as drought and salt (XP_028780287.1, Xu et al., 2007 and XP_028755164.1, Sadhukhan et al., 2019) as well as root (XP_028774714, Jones et al., 2014; Ohashi-Ito et al., 2013 and XP_028780079.1, Xing et al., 2021) and root hair development (XP_028769850.1, Kim & Dolan, 2016; Menand et al., 2007). Some additional genes of potential relevance to the singular features of *Prosopis alba* are XP_028759743.1, with potential roles in heavy metal detoxification (Sadhukhan et al., 2019) and a putative stay-green protein XP_028802427.1.

3.4 | Gene family analysis

A total of 10,704 gene families (orthogroups) were shared among *Arabidopsis thaliana*, *G. max*, *Prosopis alba*, and *Populus trichocarpa*, and 2031 gene families were unique to *Prosopis alba* (Figure 4a, Dataset S2). *Prosopis alba* shares seven expanding families with *Arabidopsis thaliana* and 17 with *Populus trichocarpa* (Figure 4b), with a striking total of 178 unique expanding gene families. Single-copy gene families were selected for construction of the species tree, and the numbers of expansion/contraction of gene families were calculated (Figure 4c). The results show that expansion/contraction of *Prosopis alba* is most similar to that of *G. max*. Gene ontology (GO) enrichment analysis of the 50 most significantly expanded families found many involved in ion homeostasis (Figure 4d).

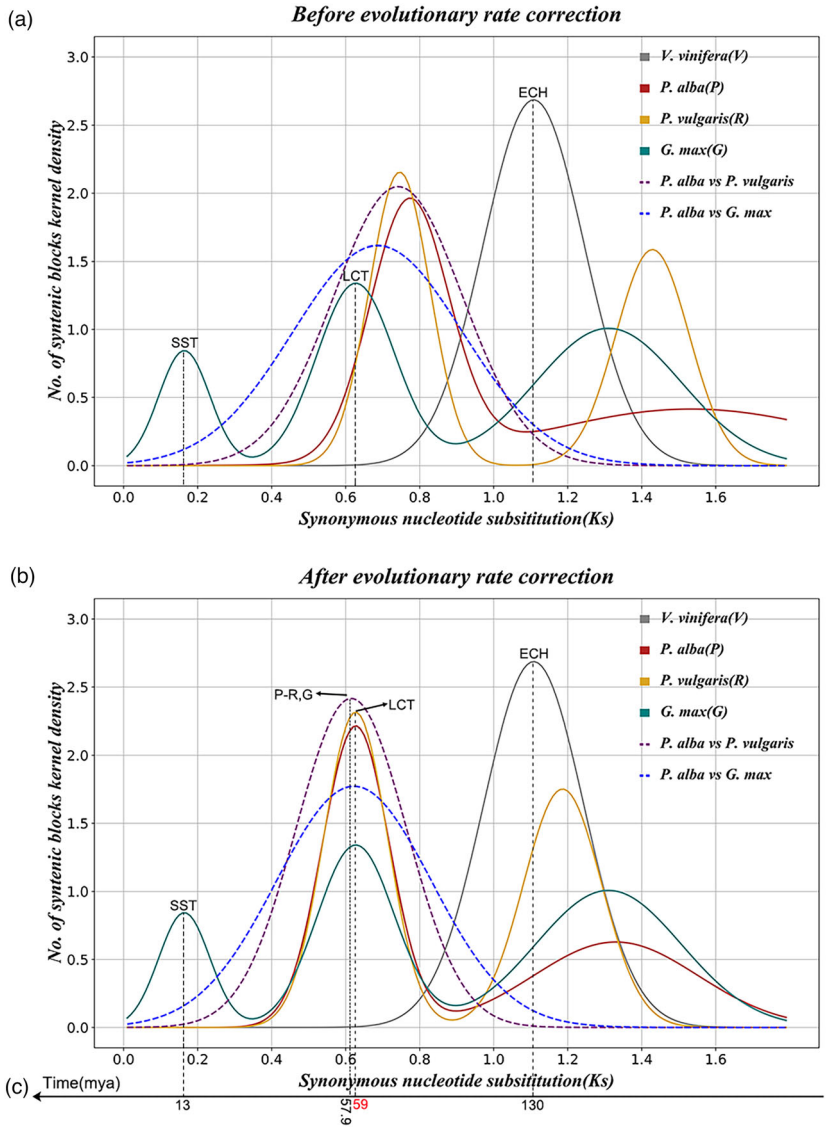

FIGURE 3 Synonymous nucleotide substitution (Ks) distribution comparing *Vitis vinifera*, *Prosopis alba*, *Phaseolus vulgaris*, *and Glycine max*. Original (a) and corrected (b) Ks among colinear genes. (a) Distributions fitted by using original Ks values. (b) Distributions fitted by using corrected Ks values. (c) Inferred evolutionary dates. The difference in the Ks peak values suggested that *Prosopis alba* genes have evolved 3.5% and 23.1% faster than *Phaseolus vulgaris* and *G. max*, respectively, following the legume common tetraploidy (LCT). We have corrected the evolutionary rates and infer that *Prosopis alba* branched off from the other legumes approximately 57.9 mya, shortly after LCT. Continuous lines are used to show Ks distribution in a genome and dashed lines are among genomes.

3.5 | *Prosopis* photosynthesis

Prosopis species have C3 photosynthesis but are highly water efficient and tolerant of salinity, which is uncommon among C3 plants. The photosynthetic rate of *Prosopis* is higher than either winter deciduous trees or desert shrubs (Wan, 1987). *Prosopis* is adapted to areas with the world's highest air temperatures (Death Valley), with a maximum light saturated photosynthesis rate of 30 mg $CO^2/dm^2/h$ (18 μmol m^{-2} s^{-1}) (Mooney et al., 1977).

The *Prosopis alba* genome contains a total of 77 putative genes involved in the carbon fixation module (Table 3), many more than crassulacean acid metabolism (CAM) plants such as pineapple (*Ananas comosus*) and orchid (*Phalaenopsis equestris*). We analyzed seven key photosynthetic enzymes including carbonic anhydrase (CA), phosphoenolpyruvate carboxylase (PEPC), pyruvate phosphate dikinase (PPDK), phosphoenolpyruvate carboxylase kinase (PPCK), phosphoenolpyruvate caboxykinase (PEPCK), malate dehydroenase (MDH) and malic enzyme (ME), comparing their numbers to orchid, pineapple, rice, sorghum, and maize (Table 3). CA, MDH, and ME were enriched in *Prosopis alba*, which may be related to its unique photosynthetic features.

We found two expanded gene families (Orthogroups 1301 and 1093, Dataset S2) involved in circadian clock systems. The clock regulates expression of abiotic stress-response genes as well as downstream signaling of stress-response hormones, although the functions of the clock may be altered in response to abiotic stress, investigated in *Arabidopsis* (Grundy et al., 2015). The former family (Orthogroup 1301) with a total of 16 genes in *Prosopis alba* versus four in *Arabidopsis thaliana*, encodes a *PSEUDO-RESPONSE REGULATOR* (*PRR*) involved in the transmission of light signals to the clock and regulation of the central oscillator (Salome & McClung, 2005). Such family expansion is also observed in another CAM plant, *Sedum album* (Wai et al., 2019). The *PRR* mutants in *Arabidopsis* showed increased tolerance for drought and salinity, properties that distinguish *Prosopis alba* from many other plants (Nakamichi et al., 2009). The latter orthogroup (1093), with a total of 20 genes in *Prosopis alba*, is expanded 10 times compared with *Arabidopsis thaliana* (2) and encodes *NIGHT LIGHT–*

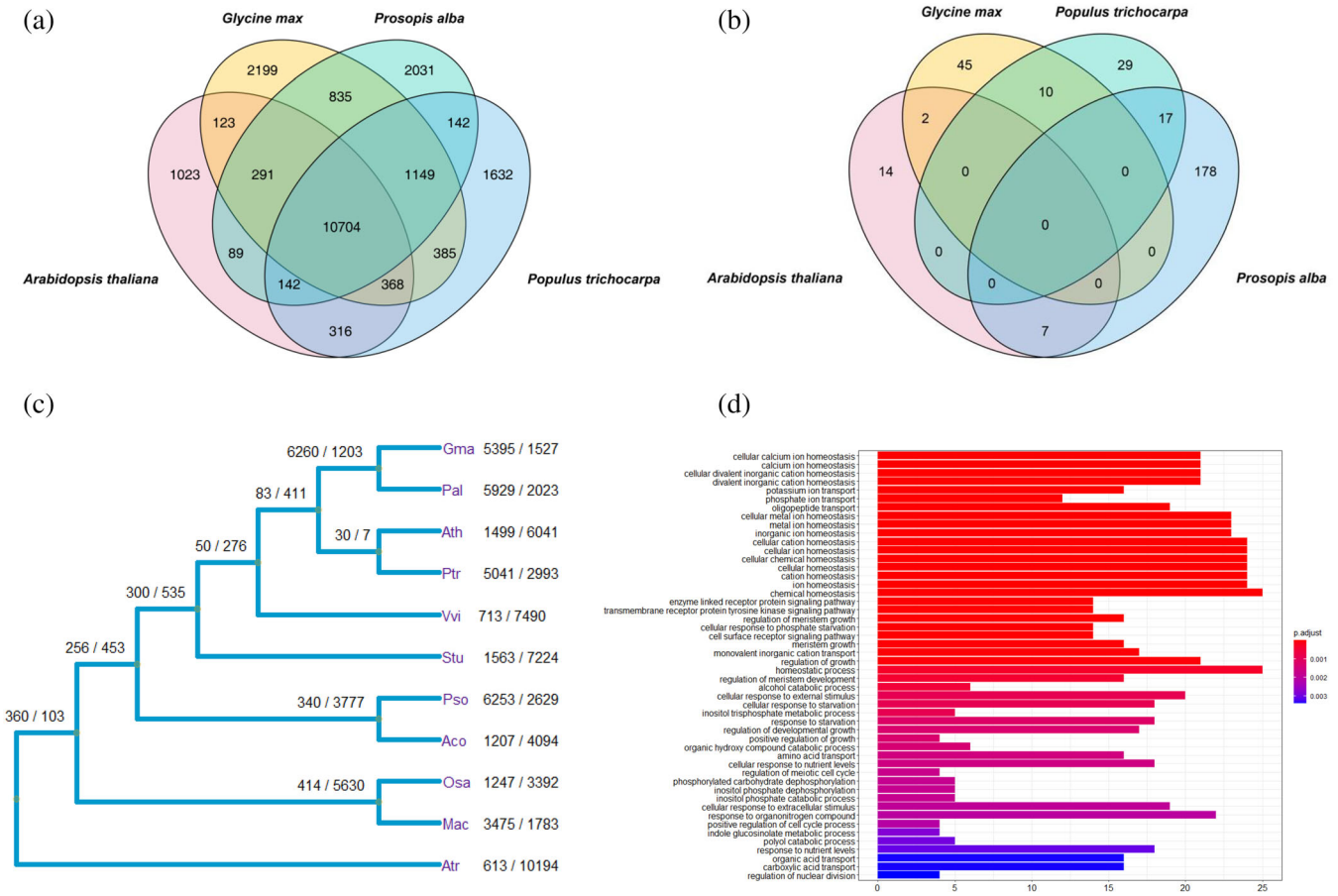

FIGURE 4 Gene family analysis. Shared gene families (orthogroups) and expanding gene families are compared for *Prosopis alba*, *Glycine max*, *Arabidopsis thaliana*, and *Populus trichocarpa*. Expansion/contraction of *Prosopis alba* is most similar to that of *G. max*. Gene ontology (GO) enrichment analysis of the 50 most significantly expanded families found many involved in ion homeostasis. (a) Venn diagram of shared and specific gene families (orthogroups) of *Prosopis alba*, *G. max*, *Arabidopsis thaliana*, and *Populus trichocarpa*. (b) Shared expanding gene family among *Arabidopsis thaliana*, *G. max*, *Populus trichocarpa* and *Prosopis alba*. (c) Gene family expansion/contraction of *Prosopis alba* and 10 other species. (d) *Prosopis alba* rapidly expanding gene families were used for GO enrichment analysis and the first 50 were displayed. Aco, Aquilegia coerulea; Atr, Amborella trichopoda; Ath, *Arabidopsis thaliana; Gma*, *Glycine max*; Mac, Musa acuminata; Osa, Oryza sativa; Pal, *Prosopis alba*; Pso, *Papaver somniferum*; Ptr, *Populus trichocarpa*; Stu, *Solanum tuberosum*; Vvi, *Vitis vinifera*.

TABLE 3 Numbers of genes encoding key photosynthetic enzymes in six species: *Prosopis alba*, *Phalaenopsis equestris* (orchid), *Ananas comosus* (pineapple), *Oryza sativa* (rice), *Sorghum bicolor* (sorghum), and *Zea mays* (maize).

Genes involved in photosynthesis	Mesquite (*Prosopis alba*)	Orchid (*Phalaenopsis equestris*)	Pineapple (*Ananas comosus*)	Rice (*O. sativa*)	Sorghum (*Sorghum bicolor*)	Maize (*Z. mays*)
Carbonic anhydrase (CA)	30	18	9	16	17	16
Phosphoenolpyruvate carboxylase (PEPC)	7	2	3	7	6	6
Pyruvate phosphate dikinase (PPDK)	1	1	1	2	2	2
Phosphoenolpyruvate carboxylase kinase (PPCK)	4	2	1	3	3	6
Phosphoenolpyruvate carboxykinase (PEPCK)	2	1	1	2	3	3
Malate dehydrogenase (MDH)	23	5	14	9	11	13
Malic enzyme	10	5	5	7	8	8

INDUCIBLE AND CLOCK-REGULATED (*LNK*) genes, which functions in integrating early light signals with temporal information integrated with core oscillator components to control the expression of afternoon genes, guiding plants to follow seasonal changes in day length (Rugnone et al., 2013). The expansion of the gene families involved in circadian clock systems may contribute to the water use efficiency in *Prosopis alba*, for example, by regulating the characteristic folding of *Prosopis alba* leaflets at night or during daytime dark periods (P.F., unpublished observations) that probably reduces water loss.

A total of four genes (XP_028758785.1, XP_028758406.1, XP_028783176.1, and XP_028779309) related to photosynthesis showed signatures of positive selection, with the first two genes involved in the light-dependent reactions of oxygenic photosynthesis (Photosystem II), and the last two genes involved in the Calvin cycle (Dataset S2).

3.6 | Nitrogen fixation

Prosopis alba, a nitrogen fixing-tree adapted to semiarid regions of northwestern Argentina (Felker & Clark, 1980), is rich in secondary metabolites, which serve as defense signals and communications to microbes. A total of four expanded gene families in *Prosopis alba* encode secondary metabolites (Orthogroups 79, 127, 135, and 154; Dataset S2), which may serve as signals to rhizobia to initiate nitrogen fixation (Pang et al., 2021). Plant phytohormones are actively involved in nitrogen fixation processes, including nodule development (McGuiness et al., 2019; Suzaki et al., 2013). Four expanded gene families related to the synthesis of brassinosteroid (Orthogroup 50), auxin (Orthogroup 186), and gibberellin (Orthogroup 688) in *Prosopis alba* may be associated with nitrogen fixation, though plant phytohormones have broad functions in regulating other aspects of growth and development.

Prosopis alba has a total of four expanded gene families (Orthogroup 150, 1022, 2839, and 3664) related to immune signaling pathways: pattern trigger immunity (PTI) and effector-triggered immunity (ETI) networks, which are important in plant-pathogen and symbiont interactions. This result agrees with a recent finding of co-evolution of immune receptor genes in plant genomes (Gong et al., 2022; Ngou et al., 2022). There is an increasing understanding that plant innate immunity plays an important role in the establishment and maintenance of symbiosis, especially ETI (Gourion et al., 2015). Those effector receptors recognize microbe-associated molecular patterns (Tóth & Stacey, 2015), leading to the formation of nitrogen fixing root nodules. A total of 20 ETI genes show positive selection in *Prosopis alba*, and one single-copy ETI gene (XP_028780079.1) shows molecular signatures of positive selection, which is involved in ethylene responsive transcription factors required for nodulation.

3.7 | Stress resistance

Using the gene annotations of *Arabidopsis thaliana*, *O. sativa*, and *V. vinifera*, we found that Orthogroups 1212 and 3157 families (Figure 5) putatively function in stress tolerance and are also

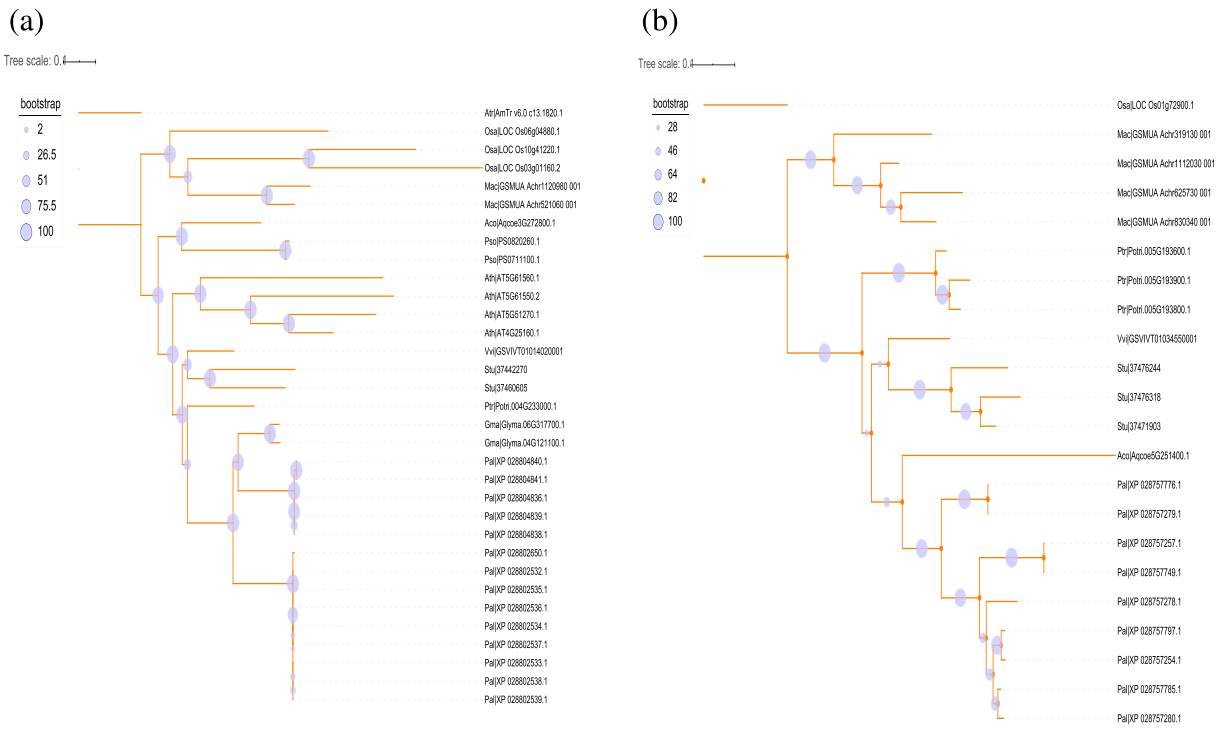

FIGURE 5 Two expanded gene orthogroups of *Prosopis alba* are involved in stress tolerance pathways. Orthogroup 1212 (a) contains genes encoding abscisic stress ripening protein 3 (ASR3); whereas, Orthogroup 3157 (b) contains genes encoding plant U-box domain-containing protein kinase family members.

specifically expanded in *Prosopis alba*. Orthogroup 1212 contains genes encoding orthologs of abscisic stress ripening protein 3 (ASR3), which has been found potentially involved in drought tolerance in several plant species (Cortes et al., 2012; Liang et al., 2019; Philippe et al., 2010). Orthogroup 3157 contains genes encoding plant U-box domain-containing protein kinase family members, which are well known to function in immune and stress response (Trujillo, 2021).

Among the total of 178 expanded gene families in *Prosopis alba*, there are 11 related to solute transport (Orthogroups 20, 59, 71, 81, 91, 213, 218, 219, 378, 1011, and 1141) which is an overrepresentation among all annotated expanded gene families (Figure 5). These transporters are involved in key mechanisms that allow plants to respond to abiotic stress (Maathuis, 2017). The expansion of the gene families in *Prosopis alba* may reflect selection for stress resistance in *Prosopis alba* under unfavorable conditions. Indeed, a total of 54 genes under positive selection has been found to be involved in the solute transport pathways (Dataset S3).

Prosopis alba is a highly salt tolerant plant, and its native land, Argentina, is considered the third country most affected by high soil salinity (Roser et al., 2014). We found one expanded gene family (Orthogroup 1938) with six genes encoding putative paralogs of a calcium dependent regulatory protein (SCaBPB/CBL10) involved in the salt overly sensitive (SOS) signaling pathway, while there is only one gene (AT4G33000) in *Arabidopsis*. In addition, the HAK/KUP/KT families were expanded (Orthogroup 20) in *Prosopis alba* and are known to be involved in the maintenance of K^+ homeostasis and salt tolerance (Yang et al., 2020).

4 | DISCUSSION

Although *Prosopis alba* genome organization is generally congruent with existing knowledge of legume genome structure, its gene content and sequences add new "clues" regarding the causes of traits of high and growing importance. For example, identification of genes permitting it to fix nitrogen under conditions that are prohibitive to other plants would be potentially important to improve other taxa and/or *Prosopis* species to produce more food and fiber at warmer temperatures with less inputs.

Particularly strong molecular signatures of positive selection were evident in genes that contain functional domains associated with disease resistance in botanical models, in the clearest cases involving small gene families that were *Prosopis alba*-specific, lacking any members from the other test species. These genes also tended to be clustered in the genome, suggesting that diversifying positive selection was coupled to selection for increased copy number by proximal duplication. Single copy orthogroups containing *Prosopis alba* genes that showed DNA-level signatures of positive selection included 16 nuclear-encoded genes functioning in organelles, 12 transcription factors including genes implicated in multiple stress tolerances such as drought and salt as well as root and root hair development, and genes with potential roles in heavy metal detoxification and a stay-green growth habit.

Expanded copy number of genes involved in ion homeostasis, albeit needing empirical investigation, might contribute to how *Prosopis alba* survived in drought and/or saline conditions in arid or semi-arid regions. Likewise, the *Prosopis alba* genome contains a total of 77 putative genes involved in carbon fixation, many more than CAM plants such as pineapple (*Ananas comosus*) and Orchid (*Phalaenopsis equestris*), which may be related to its unique photosynthetic features.

Four expanded gene families in *Prosopis alba* encoding secondary metabolites (Orthogroups 79, 127, 135, 154, Dataset S2), may serve as signals to rhizobia to initiate nitrogen fixation (Pang et al., 2021); whereas, four expanded gene families (Orthogroup 150, 1022, 2839, and 3664) related to immune signaling pathways (PTI and ETI networks) may be important in plant-pathogen and symbiont interactions, especially those such as the 20 ETI genes that also show positive selection in *Prosopis alba*.

Although the ~30% sucrose pods of *Prosopis* species were thought to be the most important food for indigenous people in the California desert (Barrows, 1900; Bell & Castetter, 1937) and had similar importance in Peru (Beresford-Jones et al., 2009) and Argentina (Capparelli, 2011; Capparelli & Lema, 2011), US ranchers often view *Prosopis* unfavorably because of its spines and aggressive colonization of rangeland (https://www.wideopencountry.com/mesquite-trees-the-most-important-trash-trees-in-texas/; https://www.texasalmanac.com/articles/the-ubiquitous-mesquite). Rangeland colonization by *Prosopis* has been suggested to have its fundamental cause in unsustainable overharvest of ecosystem nitrogen by grazing (Geesing et al., 2000), effectively creating a niche ripe for occupation by a drought-resilient plant able to provide much of its own nitrogen. Indeed, a common feature of many plant invasions is increased soil nitrogen pools and total ecosystem nitrogen stocks (Rout & Callaway, 2009). When *Prosopis* fixes more nitrogen than can be obtained from the soil, as the C/N ratio of legumes is no less than 12 (USDA, 2011), the plant must take up/sequester 12 units of C for each unit of N fixed. Thus, Prosopis is useful in arid lands for C sequestration.

Even in its inferred center of diversity, where it is highly valued for food and lumber, overharvest of the genus without planting or genetic selection (Burkart, 1976) may have sacrificed opportunities to fully benefit from the novel strengths that *Prosopis* species offer as cultigens, perhaps even contributing to disdain for *Prosopis*. For example, a genetically untested, spiny, multi-stemmed tetraploid *Prosopis juliflora* accession with astringent pods was casually introduced into Africa and the Indian subcontinent in the late 1800s and by early 1900s had spread to extensive areas of the very arid regions from Dakar, Senegal to Delhi, India where almost no other species could grow. Spineless, erect, diploid selections of the tropical *Prosopis pallida* from Peru with very palatable pods have been shown in field trials in Cape Verde and the Rajasthan desert to have similar adaptability to *Prosopis juliflora*, but because of the unfavorable view of this genus, genetic improvement efforts to combine the adaptability and favorable characteristics of these species have not been initiated in Africa (to our knowledge). Likewise, the major energy source in Haiti is derived from the charcoal of thorny, multi-stemmed *Prosopis juliflora*

(Tarter, 2022), bearing pods of low human palatability; whereas, erect, single-stemmed, thornless individuals with high human palatable pods and greater growth (Wojtusik et al., 1993) were never put into commercial use.

The properties that enable *Prosopis* species to be productive under extraordinary stresses may also contribute to some of its commercially valuable attributes. For example, possibly because of the adaptation to greatly changing water availability, *Prosopis* lumber has very low volumetric shrinkage values of 4%–5%, with nearly equal radial and tangential shrinkage. The result is that flooring and fine furniture made from *Prosopis* has much less tendency to twist, warp, and cup than other fine hardwoods.

Coupling ecological context with the genomics-based clues reported here offers a powerful new system in which deductions about functional roles of specific genes/alleles in the striking physiological adaptations of *Prosopis* species can be made. For example, SNPs that are shared between the Saline Valley National Park population of *Prosopis glandulosa* and the salt-tolerant *Prosopis pallida*, *Prosopis tamarugo*, and/or *Prosopis alba* would be a logical starting point for empirical tests of gene function. Fertile progeny of an interspecific cross between a North American *Prosopis glandulosa* and a South American *Prosopis alba* have value as freeze hardy ornamental trees (US patent PP 32, 467P2). These parents are only two of the approximately 30 species in the section Algarobia of the genus that range from the hottest location in the world (Death Valley, California) through the hyper-arid coastal regions of Peru, with habits ranging from erect, thornless, sweet podded trees to straight erect, often thornless *Prosopis chilenses* at 32 S latitude in Mendoza, Argentina. This interbreeding gene pool offers almost unlimited potential for hybridization and cloning of elite progeny, with incipient but widely-distributed efforts barely scratching the surface of the opportunity (Felker & Moss, 1996).

AUTHOR CONTRIBUTIONS

Wenqian Kong and Min Liu designed the experiment, conducted the data analysis, and wrote the manuscript. Peter Felker, Mauricio Ewens, Cecilia Bessega, and Carolina Pometti provided the materials for sequencing and designed the research. Jinpeng Wang, Peng Xu, Jia Teng, Jinyu Wang, Xiyin Wang, Yuannian Jiao, Françoise Thibaud-Nissen, Patrick Masterson, and Xin Qiao analyzed the data and wrote the manuscript. Magdy S. Alabady conducted the experiment and wrote the manuscript. Andrew H. Paterson oversaw the project, wrote and edited the manuscript.

ACKNOWLEDGEMENTS

We thank the Georgia Genomics and Bioinformatics Core (GGBC) at the University of Georgia for providing the library preparation and sequencing services. The work of FT-N and PM was supported by the National Center for Biotechnology Information of the National Library of Medicine (NLM), National Institutes of Health.

CONFLICT OF INTEREST STATEMENT

All authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

All sequences and annotation result are available at https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_004799145.1/.

ORCID

Wenqian Kong https://orcid.org/0000-0002-4540-6165
Min Liu https://orcid.org/0000-0001-5143-0420
Cecilia Bessega https://orcid.org/0000-0002-8575-1828
Carolina Pometti https://orcid.org/0000-0003-0122-6338
Peng Xu https://orcid.org/0000-0001-9083-1454
Xiyin Wang https://orcid.org/0000-0003-3454-0374
Yuannian Jiao https://orcid.org/0000-0002-8987-2782
Magdy S. Alabady https://orcid.org/0000-0002-1601-682X
Françoise Thibaud-Nissen https://orcid.org/0000-0003-4957-7807
Patrick Masterson https://orcid.org/0000-0001-8837-3706
Xin Qiao https://orcid.org/0000-0002-2465-4960
Andrew H. Paterson https://orcid.org/0000-0003-2159-0487

REFERENCES

Aljanabi, S. M., Forget, L., & Dookun, A. (1999). An improved and rapid protocol for the isolation of polysaccharide- and polyphenol-free sugarcane DNA. *Plant Molecular Biology Reporter*, *17*(3), 281–281. https://doi.org/10.1023/A:1007692929505

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Anand, S. S., Thakur, S., Gargi, M., Choudhary, S., & Bhardwaj, P. (2017). Development and characterization of genomic microsatellite markers in Prosopis cineraria. *Current Plant Biology*, *9–10*, 37–42.

Barrows, D. P. (1900). The Ethnobotany of the Coahuilla Indians of Southern California. Ph.D., University of Chicago Malki Museum Press, Morongo Indian Reservation, Banning, California.

Bell, W. H., & Castetter, E. F. (1937). The Utilization of Mesquite and Screwbean by the Aborigines in the American Southwest.

Beresford-Jones, D. G., Arce, S., Oliver, T., Whaley, Q., & Chepstow-Lusty, A. (2009). The role of *Prosopis* in ecological and landscape change in the Samaca Basin, lower Ica Valley, south coast Peru from the earlier horizon to the late intermediate period. *Latin American Antiquity*, *20*, 303–332. https://doi.org/10.1017/S1045663500002650

Bowers, J. E., Chapman, B. A., Rong, J., & Paterson, A. H. (2003). Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature*, *422*(6930), 433–438. https://doi.org/10.1038/nature01521

Bukhari, Y. M. (1997). Nuclear DNA amounts in Acacia and *Prosopis* (Mimosaceae) and their evolutionary implications. *Hereditas*, *126*(1), 45–51. https://doi.org/10.1111/j.1601-5223.1997.00045.x

Burkart, A. (1976). A monograph of the genus Prosopis (Leguminosae subfam. Mimosoideae). *Journal of the Arnold Arboretum*, 450–525.

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Capparelli, A. (2011). Elucidating post-harvest practices involved in the processing of Algarrobo (*Prosopis* spp.) for food at El Shincal Inka site (Northwest Argentina): An experimental approach based on charred remains. *Archaeological and Anthropoogical Science*, *3*, 93–112. https://doi.org/10.1007/s12520-011-0061-4

Capparelli, A., & Lema, V. (2011). Recognition of post-harvest processing of algarrobo (*Prosopis* spp.) as food from two sites of Northwestern Argentina: An ethnobotanical and experimental approach for

desiccated macroremains. *Archaeological and Anthropological Sciences*, *3*, 71–92. https://doi.org/10.1007/s12520-011-0052-5

Cortes, A. J., Chavarro, M. C., Madrinan, S., This, D., & Blair, M. W. (2012). Molecular ecology and selection in the drought-related Asr gene polymorphisms in wild and cultivated common bean (Phaseolus vulgaris L.). *BMC Genetics*, *13*, 58.

De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution. *Bioinformatics*, *22*(10), 1269–1271. https://doi.org/10.1093/bioinformatics/btl097

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Ewens, M., Felker, P., & Paterson, A. (2022). Chapter 9 - Genetic improvement in *Prosopis*. In M. C. Puppo & P. Felker (Eds.), *Prosopis as a heat tolerant nitrogen fixing desert food legume* (pp. 139–156). Academic Press.

Ewens, M., Gezan, S., & Felker, P. (2012). Five year field evaluation of *Prosopis alba* clones on pH 9–10 soils in Argentina selected for growth in the greenhouse at seawater salinities (45 dS m−1). *Forests*, *3*, 95–113. https://doi.org/10.3390/f3010095

Felker, P. (1981). Uses of tree legumes in semiarid regions. *Economic Botany*, *35*(2), 174–186. https://doi.org/10.1007/BF02858684

Felker, P., & Clark, P. R. (1980). Nitrogen fixation (acetylene reduction) and cross inoculation in 12 Prosopis (mesquite) species. *Plant and Soil*, *57*(2), 177–186.

Felker, P., & Clark, P. R. (1982). Position of mesquite (Prosopis spp) nodulation and nitrogen fixation (acetylene reduction) in 3-m long phraetophytically simulated soil columns. *Plant and Soil*, *64*(3), 297–305.

Felker, P., & Moss, J. (1996). *Prosopis*: Semiarid fuelwood and forage tree building consensus for the disenfranchised. *Constitution*, *13*, 15.

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, *117*(17), 9451–9457.

Geesing, D., Felker, P., & Bingham, R. L. (2000). Influence of mesquite (*Prosopis glandulosa*) on soil nitrogen and carbon development: Implications for global C sequestration. *Journal of Arid Environments*, *46*, 157–180. https://doi.org/10.1006/jare.2000.0661

Gong, Z., Shao, Z.-Q., Chen, J.-Q., & Han, G.-Z. (2022). Plant immune receptors evolved hand in hand. *Nature Plants*, *8*, 1–2. https://doi.org/10.1038/s41477-022-01258-z

Gourion, B., Berrabah, F., Ratet, P., & Stacey, G. (2015). Rhizobium-legume symbioses: The crucial role of plant immunity. *Trends in Plant Science*, *20*(3), 186–194. https://doi.org/10.1016/j.tplants.2014.11.008

Grundy, J., Stoker, C., & Carre, I. A. (2015). Circadian regulation of abiotic stress tolerance in plants. *Frontiers in Plant Science*, *6*, 648. https://doi.org/10.3389/fpls.2015.00648

Hilu, K. W., Boyd, S., & Felker, P. (1982). Morphological diversity and taxonomy of California mesquites (*Prosopis*, Leguminosae). *Madrono*, 237–254.

Hughes, C. E., Ringelberg, J. J., Lewis, G. P., & Catalano, S. A. (2022). Disintegration of the genus *Prosopis* L. (Leguminosae, Caesalpinioideae, mimosoid clade). *PhytoKeys*, *205*, 147–189. https://doi.org/10.3897/phytokeys.205.75379

Jaillon, O., Aury, J. M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., … French-Italian Public Consortium for Grapevine Genome Characterization. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, *449*, 463–467. https://doi.org/10.1038/nature06148

Johnston, M. C. (1962). The North American Mesquites *Prosopis* Sect. Algarobia (Leguminosae). Brittonia: 72-90.

Jones, A. M., Xuan, Y., Xu, M., Wang, R. S., Ho, C. H., Lalonde, S., You, C. H., Sardi, M. I., Parsa, S. A., Smith-Valle, E., Su, T., Frazer, K. A., Pilot, G., Pratelli, R., Grossmann, G., Acharya, B. R., Hu, H. C., Engineer, C., Villiers, F., … Frommer, W. B. (2014). Border control—A membrane-linked interactome of *Arabidopsis*. *Science*, *344*(6185), 711–716. https://doi.org/10.1126/science.1251358

Kapustin, Y., Souvorov, A., Tatusova, T., & Lipman, D. (2008). Splign: Algorithms for computing spliced alignments with identification of paralogs. *Biology Direct*, *3*, 20. https://doi.org/10.1186/1745-6150-3-20

Kim, C. M., & Dolan, L. (2016). ROOT HAIR DEFECTIVE SIX-LIKE class I genes promote root hair development in the grass *Brachypodium distachyon*. *PLoS Genetics*, *12*(8), e1006211. https://doi.org/10.1371/journal.pgen.1006211

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36

Koren, S., Rhie, A., Walenz, B. P., Dilthey, A. T., Bickhart, D. M., Kingan, S. B., Hiendleder, S., Williams, J. L., Smith, T. P., & Phillippy, A. M. (2018). Complete assembly of parental haplotypes with trio binning. BioRxiv: 271486.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/gr.215087.116

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Li, L., Stoeckert, C. J., & Roos, D. S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Research*, *13*(9), 2178–2189.

Liang, Y., Jiang, Y., Du, M., Li, B., Chen, L., Chen, M., Jin, D., & Wu, J. (2019). ZmASR3 from the maize ASR gene family positively regulates drought tolerance in transgenic arabidopsis. *International Journal of Molecular Sciences*, *20*(9), 2278.

Maathuis, F. J. M. (2017). Solute transport: Basic concepts. In S. Clemens (Ed.), *Plant physiology and function* (pp. 1–31). New York, NY: Springer.

McGuiness, P. N., Reid, J. B., & Foo, E. (2019). The role of gibberellins and brassinosteroids in nodulation and arbuscular mycorrhizal associations. *Frontiers in Plant Science*, *10*, 269. https://doi.org/10.3389/fpls.2019.00269

Menand, B., Yi, K., Jouannic, S., Hoffmann, L., Ryan, E., Linstead, P., Schaefer, D. G., & Dolan, L. (2007). An ancient mechanism controls the development of cells with a rooting function in land plants. *Science*, *316*(5830), 1477–1480. https://doi.org/10.1126/science.1142618

Mooney, H. A., Simpson, B. B., & Solbrig, O. T. (1977). Phenology, morphology, physiology. In B. B. Simpson (Ed.), *Mesquite: its biology in two desert scrub ecosystems* (pp. 26–43). Dowden, Hutchinson, & Ross, Inc.

Morgulis, A., Gertz, E. M., Schaffer, A. A., & Agarwala, R. (2006). WindowMasker: Window-based masker for sequenced genomes. *Bioinformatics*, *22*(2), 134–141. https://doi.org/10.1093/bioinformatics/bti774

Nakamichi, N., Kusano, M., Fukushima, A., Kita, M., Ito, S., Yamashino, T., Saito, K., Sakakibara, H., & Mizuno, T. (2009). Transcript profiling of an *Arabidopsis* PSEUDO RESPONSE REGULATOR arrhythmic triple mutant reveals a role for the circadian clock in cold stress response. *Plant and Cell Physiology*, *50*(3), 447–462. https://doi.org/10.1093/pcp/pcp004

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, *3*(5), 418–426.

Ngou, B. P. M., Heal, R., Wyler, M., Schmid, M. W., & Jones, J. D. G. (2022). Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nature Plants*, *8*(10), 1146–1152. https://doi.org/10.1038/s41477-022-01260-5

Ohashi-Ito, K., Matsukawa, M., & Fukuda, H. (2013). An atypical bHLH transcription factor regulates early xylem development downstream of auxin. *Plant and Cell Physiology*, *54*(3), 398–405. https://doi.org/10.1093/pcp/pct013

Pang, Z., Chen, J., Wang, T., Gao, C., Li, Z., Guo, L., Xu, J., & Cheng, Y. (2021). Linking plant secondary metabolites and plant microbiomes: A review. *Frontiers in Plant Science*, *12*, 621276. https://doi.org/10.3389/fpls.2021.621276

Paterson, A. H., Felker, P., Hubbell, S. P., & Ming, R. (2008). The fruits of tropical plant genomics. *Tropical Plant Biology*, *1*(1), 3–19.

Pasiecznik, N. M., Felker, P., Harris, P. J., Harsh, L., Cruz, G., Tewari, J., Cadoret, K., & Maldonado, L. J. (2001). *The Prosopis juliflora-Prosopis pallida complex: A monograph*. HDRA.

Philippe, R., Courtois, B., McNally, K. L., Mournet, P., El-Malki, R., Le Paslier, M. C., Fabre, D., Billot, C., Brunel, D., & Glaszmann, J.-C. (2010). Structure, allelic diversity and selection of Asr genes, candidate for drought tolerance, in Oryza sativa L. and wild relatives. *Theoretical and Applied Genetics*, *121*(4), 769–787.

Pond, S. L. K., & Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. In *Statistical methods in molecular evolution* (pp. 125–181). Springer. https://doi.org/10.1007/0-387-27733-1_6

Pruitt, K. D., Brown, G. R., Hiatt, S. M., Thibaud-Nissen, F., Astashyn, A., Ermolaeva, O., Farrell, C. M., Hart, J., Landrum, M. J., McGarvey, K. M., & Murphy, M. R. (2014). RefSeq: An update on mammalian reference sequences. *Nucleic Acids Research*, *42*(Database issue), D756–D763. https://doi.org/10.1093/nar/gkt1114

Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., Lee, C., Ko, B. J., Chaisson, M., Gedman, G. L., Cantin, L. J., Thibaud-Nissen, F., Haggerty, L., Bista, I., Smith, M., … Jarvis, E. D. (2021). Towards complete and error-free genome assemblies of all vertebrate species. *Nature*, *592*(7856), 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rhodes, D., & Felker, P. (1987). Mass screening *Prosopis* (mesquite) seedlings for growth at seawater salinity. *Forest Ecology and Management*, *24*, 169–176. https://doi.org/10.1016/0378-1127(88)90041-2

Roser, L. G., Ferreyra, L. I., Ewens, M., Vilardi, J. C., & Saidman, B. O. (2014). Genetic and morphometric characterization of clones of *Prosopis alba*, Algarobia, selected for salt tolerance. *Tree Genetics & Genomes*, *10*(2), 383–397. https://doi.org/10.1007/s11295-013-0693-x

Rout, M. E., & Callaway, R. M. (2009). An invasive plant paradox. *Science*, *324*, 724–725. https://doi.org/10.1126/science.1173651

Rugnone, M. L., Faigón Soverna, A., Sanchez, S. E., Schlaen, R. G., Hernando, C. E., Seymour, D. K., Mancini, E., Chernomoretz, A., Weigel, D., Más, P., & Yanovsky, M. J. (2013). LNK genes integrate light and clock signaling networks at the core of the *Arabidopsis* oscillator. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(29), 12120–12125. https://doi.org/10.1073/pnas.1302170110

Sadhukhan, A., Enomoto, T., Kobayashi, Y., Watanabe, T., Iuchi, S., Kobayashi, M., Sahoo, L., Yamamoto, Y. Y., & Koyama, H. (2019). Sensitive to proton rhizotoxicity1 regulates salt and drought tolerance of *Arabidopsis thaliana* through transcriptional regulation of CIPK23. *Plant and Cell Physiology*, *60*(9), 2113–2126. https://doi.org/10.1093/pcp/pcz120

Salem, B. (1989). Arid zone forestry: a guide for field technicians. FAO Conservation Guide(20).

Salome, P. A., & McClung, C. R. (2005). PSEUDO-RESPONSE REGULATOR 7 and 9 are partially redundant genes essential for the temperature responsiveness of the *Arabidopsis* circadian clock. *Plant Cell*, *17*(3), 791–803. https://doi.org/10.1105/tpc.104.029504

Sanderson, M. J. (2003). r8s: Inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics*, *19*(2), 301–302. https://doi.org/10.1093/bioinformatics/19.2.301

Seppey, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods in Molecular Biology*, *1962*, 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science*, *27*(1), 135–145. https://doi.org/10.1002/pro.3290

Singh, G., Abrol, I. P., & Cheema, S. S. (1989). Effects of gypsum application on mesquite (*Prosopis juliflora*) and soil properties in an abandoned sodic soil. *Forest Ecology and Management*, *29*, 1–14. https://doi.org/10.1016/0378-1127(89)90052-2

Smith, M. D., Wertheim, J. O., Weaver, S., Murrell, B., Scheffler, K., & Kosakovsky Pond, S. L. (2015). Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution*, *32*(5), 1342–1353. https://doi.org/10.1093/molbev/msv022

Stajich, J. E., Block, D., Boulez, K., Brenner, S. E., Chervitz, S. A., Dagdigian, C., Fuellen, G., Gilbert, J. G., Korf, I., Lapp, H., & Lehväslaiho, H. (2002). The Bioperl toolkit: Perl modules for the life sciences. *Genome Research*, *12*(10), 1611–1618. https://doi.org/10.1101/gr.361602

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Suyama, M., Torrents, D., & Bork, P. (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, *34*(suppl_2), W609–W612. https://doi.org/10.1093/nar/gkl315

Suzaki, T., Ito, M., & Kawaguchi, M. (2013). Genetic basis of cytokinin and auxin functions during root nodule development. *Frontiers in Plant Science*, *4*, 42. https://doi.org/10.3389/fpls.2013.00042

Tarter, A. (2022). Chapter 6 - The evolution of *Prosopis* management in Haiti. In M. C. Puppo & P. Felker (Eds.), *Prosopis as a heat tolerant nitrogen fixing desert food legume* (pp. 83–93). Academic Press.

Thibaud-Nissen, F., Souvorov, A., Murphy, T., DiCuccio, M., & Kitts, P. (2013). Eukaryotic genome annotation pipeline. In *The NCBI handbook [Internet]* (2nd ed.). National Center for Biotechnology Information (US).

Tóth, K., & Stacey, G. (2015). Does plant immunity play a critical role during initiation of the legume-rhizobium symbiosis? *Frontiers in Plant Science*, *6*, 401.

Trujillo, M. (2021). Ubiquitin signalling: controlling the message of surface immune receptors. *New Phytologist*, *231*(1), 47–53.

USDA. (2011). Carbon to Nitrogen Ratios in Cropping Systems. In Center NENTS. USDA Natural Resources Conservation Service.

Velarde, M., Felker, P., & Degano, C. (2003). Evaluation of Argentine and Peruvian *Prosopis* germplasm for growth at seawater salinities. *Journal of Arid Environments*, *55*, 515–531. https://doi.org/10.1016/S0140-1963(02)00280-X

Wai, C. M., Weise, S. E., Ozersky, P., Mockler, T. C., Michael, T. P., & VanBuren, R. (2019). Time of day and network reprogramming during drought induced CAM photosynthesis in *Sedum album*. *PLoS Genetics*, *15*(6), e1008209. https://doi.org/10.1371/journal.pgen.1008209

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., & Young, S. K. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *Plos One*, *9*(11), e112963.

Wan, C. (1987). *Photosynthesis and water use in honey mesquite*. Texas Tech University.

Wang, J., Sun, P., Li, Y., Liu, Y., Yang, N., Yu, J., Ma, X., Sun, S., Xia, R., & Liu, X. (2018). An overlooked paleotetraploidization in Cucurbitaceae. *Molecular Biology and Evolution*, *35*(1), 16–26. https://doi.org/10.1093/molbev/msx242

Wang, J., Sun, P., Li, Y., Liu, Y., Yu, J., Ma, X., Sun, S., Yang, N., Xia, R., & Lei, T. (2017). Hierarchically aligning 10 legume genomes establishes a

family-level genomics platform. *Plant Physiology*, *174*(1), 284–300. https://doi.org/10.1104/pp.16.01981

Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W., Ge, S., & Luo, J. (2006). Statistical inference of chromosomal homology based on gene colinearity and applications to *Arabidopsis* and rice. *BMC Bioinformatics*, *7*(1), 1–13. https://doi.org/10.1186/1471-2105-7-447

Wang, X., Wang, J., Jin, D., Guo, H., Lee, T. H., Liu, T., & Paterson, A. H. (2015). Genome alignment spanning major Poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Molecular Plant*, *8*(6), 885–898. https://doi.org/10.1016/j.molp.2015.04.004

Wojtusik, T., & Felker, P. (1993). Interspecific graft incompatibility in *Prosopis*. *Forest Ecology and Management*, *59*(3), 329–340. https://doi.org/10.1016/0378-1127(93)90011-B

Wojtusik, T., Felker, P., Russell, E. J., & Benge, M. D. (1993). Cloning of erect, thornless, non-browsed nitrogen fixing trees of Haiti's principal fuelwood species (*Prosopis juliflora*). *Agroforestry Systems*, *21*(3), 293–300. https://doi.org/10.1007/BF00705248

Xing, Y., Wang, N., Zhang, T., Zhang, Q., Du, D., Chen, X., Lu, X., Zhang, Y., Zhu, M., & Liu, M. (2021). SHORT-ROOT 1 is critical to cell division and tracheary element development in rice roots. *The Plant Journal*, *105*(5), 1179–1191. https://doi.org/10.1111/tpj.15095

Xu, Z.-S., Xia, L.-Q., Chen, M., Cheng, X.-G., Zhang, R.-Y., Li, L.-C., Zhao, Y.-X., Lu, Y., Ni, Z.-Y., & Liu, L. (2007). Isolation and molecular characterization of the *Triticum aestivum* L. ethylene-responsive factor 1 (TaERF1) that increases multiple stress tolerance. *Plant Molecular Biology*, *65*(6), 719–732. https://doi.org/10.1007/s11103-007-9237-9

Yang, T., Lu, X., Wang, Y., Xie, Y., Ma, J., Cheng, X., Xia, E., Wan, X., & Zhang, Z. (2020). HAK/KUP/KT family potassium transporter genes are involved in potassium deficiency and stress responses in tea plants (Camellia sinensis L.): Expression and functional analysis. *BMC Genomics*, *21*(1), 556.

SUPPORTING INFORMATION

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

How to cite this article: Kong, W., Liu, M., Felker, P., Ewens, M., Bessega, C., Pometti, C., Wang, J., Xu, P., Teng, J., Wang, J., Wang, X., Jiao, Y., Alabady, M. S., Thibaud-Nissen, F., Masterson, P., Qiao, X., & Paterson, A. H. (2023). Genome and evolution of *Prosopis alba* Griseb., a drought and salinity tolerant tree legume crop for arid climates. *Plants, People, Planet*, 1–15. https://doi.org/10.1002/ppp3.10404