the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2025) **121,** e17255 doi: 10.1111/tpj.17255

Genomic insights into drought adaptation of the forage shrub *Caragana korshinskii* (Fabaceae) widely planted in drylands

Fengyuan Mei[1,†], Tianrui Yang[2,†], Haoyu Chao[3], Xiaohui Ma[1], Jingjing Wu[1], Qi Yang[2], Guangpeng Ren[1], Li Song[2], Qian Wang[2], Liwang Qi[4], Xinxing Fu[5], Gegentu[6], Cuiping Gao[7], Ruigang Wang[2,8], Ming Chen[3], Xiangwen Fang[1], Jianquan Liu[1,*], Guojing Li[2,6,7,*] and Shengdan Wu[1,*]

[1]*State Key Laboratory of Herbage Improvement and Grassland Agro-ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China,*

[2]*Inner Mongolia Key Laboratory of Plants Adversity Adaptation and Genetic Improvement in Cold and Arid Regions, Inner Mongolia Agricultural University, Hohhot 010018, China,*

[3]*Department of Bioinformatics, College of Life Sciences, Zhejiang University, Hangzhou 310058, China,*

[4]*Research Institute of Forestry, Chinese Academy of Forestry, Beijing 100091, China,*

[5]*College of Life Sciences, Northwest Normal University, Lanzhou 730070, China,*

[6]*Key Laboratory of Forage Cultivation, Processing and High Efficient Utilization of Ministry of Agriculture, Inner Mongolia Agricultural University, Hohhot 010021, China,*

[7]*Key Laboratory of Grassland Resources of Ministry of Education, Inner Mongolia Agricultural University, Hohhot 010021, China, and*

[8]*Inner Mongolia Enterprise Key Laboratory of Tree Breeding, Mengshu Ecological Construction Group Co., Ltd., Hohhot 011517, China*

Received 29 September 2022; revised 26 December 2024; accepted 30 December 2024.
*For correspondence (e-mail wusd@lzu.edu.cn and liguojing@imau.edu.cn and liujq@nwipb.cas.cn).
†These authors contributed equally to this work.

SUMMARY

The Korshinsk peashrub (*Caragana korshinskii*), known for its exceptional drought tolerance, is widely cultivated in arid and semi-arid regions for vegetation restoration and as a vital forage plant. To elucidate the genomic basis of its drought tolerance, we generated a chromosomal-scale genome sequence of *C. korshinskii*. Our synteny analysis disputes the previously hypothesized genus-specific whole-genome duplication event, as suggested by earlier transcriptome study of this species and its congeners. We identified that tandem duplications were critical for the expansion of gene families, such as early light-induced protein, heat shock protein 100, and Dehydrin, which are involved in cellular protection processes. These expansions are likely pivotal to the superior drought tolerance observed in *C. korshinskii*, as evidenced by the elevated gene expression of these genes under drought conditions. Furthermore, overexpression studies of seven tandemly duplicated *DHN* genes revealed a substantial enhancement in drought survival rates of seedlings, likely attributable to increased gene dosage effects. Conversely, gene silencing via virus-induced gene silencing demonstrated opposing effects. Additionally, we have established the CakorDB, a genomic resource database for *C. korshinskii* (https://bis.zju.edu.cn/cakordb/), accessible freely to the scientific community. Collectively, our study not only provides a valuable genomic resource for the Korshinsk peashrub but also highlights the genetic adaptations that enable *C. korshinskii* to thrive in desert environments, positioning its stress-responsive genes as a valuable genetic reservoir for breeding drought-resistant crops.

Keywords: ***Caragana korshinskii*, drought adaptation, tandem duplication, dehydrin, forage shrub, drylands.**

INTRODUCTION

Arid and semi-arid zones, collectively known as drylands, make up approximately 41% of Earth's land surface (Reynolds et al., 2007). These regions are characterized by limited water availability and often accompanied by extreme temperature fluctuations and degraded, nutrient-poor soil, all of which challenge plant survival (Withfold & Duval, 2019). For example, most domesticated crops

cannot withstand these harsh conditions for widespread cultivation. In contrast, certain indigenous plants display exceptional adaptation, having evolved endurance to these stressful environments. The increasing incidence of droughts (Huang et al., 2016) and the spread of drylands (Walker & Van Loon, 2023), both intensified by global climate change, threaten the sustainability of agriculture, pastoral, and livestock sectors—key for feeding an expanding global population (Gupta et al., 2020). One potential strategy to address these challenges is the discovery of new germplasm from plants that are naturally adapted to aridity and investigation into their genetic basis of drought tolerance. Harnessing such stress-tolerant genes could be pivotal for enhancing crop robustness.

Commonly known as 'Korshinsk peashrub', *Caragana korshinskii* is a member of the genus *Caragana* (Fabaceae) (Liu et al., 2010) and is closely related to commonly cultivated crops like soybean (*Glycine max*), alfalfa (*Medicago* spp.), and the common pea (*Pisum sativum*) (Azani et al., 2017). *C. korshinskii* is native to the arid and semi-arid regions of northern China and Mongolia, where it is a dominant component of sandy grasslands, arid scrublands, and desert ecosystems (Liu et al., 2010). Noted for its extraordinary drought tolerance, *C. korshinskii* has been widely planted in drylands for vegetation rehabilitation (Figure S1). It contributes to soil improvement by fixing atmospheric nitrogen and serves as a valuable forage source for livestock, including sheep, goats, cattle, and camels. Considering *C. korshinskii*'s significant ecological and economic value, a growing number of studies have aimed to elucidate its mechanism of drought tolerance. Ma et al. (2008) suggested that the physiological adaptation of *C. korshinskii* to water-deficit conditions might be due to its low transpiration rate and high water use efficiency. Furthermore, the enhanced stomatal control observed during drought has been attributed to the accumulation of abscisic acid (ABA), which ensures hydraulic safety and sustains water usage in *C. korshinskii* (Yao et al., 2021). Yet, only a few stress-related genes (e.g., *DREB1C* and *NAC3/4*) has been studied to examine their expression levels in response to drought stress at molecular level (Han et al., 2015; Liu et al., 2019). Consequently, a comprehensive understanding of genomic adaptations to drought in this important leguminous forage remains incomplete.

In this study, we aimed to provide a high-quality genome of *C. korshinskii* and comprehensively explore genes relevant to drought tolerance. Previous research integrating genome and transcriptome analysis has identified numerous species-specific genes that confer abiotic stress resistance. These genes have typically arisen from whole-genome duplications (WGDs) or tandem duplications in both trees and herbs (Hu et al., 2021; Ma et al., 2013; Moghaddam et al., 2021; Ren et al., 2022; Wan et al., 2021), as well as in resurrection plants (Costa et al., 2017; Gao et al., 2024; Silva et al., 2021; Xiao et al., 2015; Xu et al., 2018). Based on the genome sequence of *C. korshinskii*, we specifically examined the species-specific genes that originated in this species and assessed their expression changes in response to drought stress using transcriptome data. We further examined the drought-stressed seedlings' phenotype by overexpressing or silencing the species-specific genes, with the previously developed *Agrobacterium*-mediated transient expression system for *C. korshinskii* (Liu et al., 2019). Our study thus generated a high-quality genome of *C. korshinskii* and elucidated the genomic basis for this legume shrub's long-term adaptation to drought in desert regions. Looking ahead, the stress-responsive genes in *C. korshinskii* will serve as a valuable resource for genetically breeding drought-tolerant crops in areas affected by climate change.

RESULTS

Chromosomal-scale genome assembly and annotation

C. korshinskii is a diploid ($2n = 2x = 16$) (Zhou et al., 2002), with an estimated genome size of 1.49 Gb and a heterozygosity of 1.61%, based on the distribution of its *K*-mer frequency (Figure S2). To obtain a high-quality genome for *C. korshinskii*, we generated 137.13 Gb (approximately 92×) of long reads with Oxford Nanopore Technologies (ONT), and 327.68 Gb (approximately 220×) of short reads with the DNBSeq™ sequencing platform (Table S1). We used NECAT2, a *de novo* assembler specific for ONT reads (Chen et al., 2021), to assemble the data for the long reads and further polish the contig with the short reads. After, we produced an assembly of 1.35 Gb with a contig N50 length of 4.33 Mb (Table S2). Then, the contig level assembly was further scaffolded with 131.17 Gb (approximately 88×) of chromosome conformation capture (Hi-C) data (Table S1). Consequently, nearly 89% (1.25 Gb) of the assembled whole-genome sequences were anchored into eight pseudochromosomes (Figure 1; Figure S3), and the final assembly with scaffold N50 of 158.29 Mb (Table S3). The mapping rate of short reads to the assembled genome was approximately 99.32% (Table S4). Additionally, 97.6% of the conserved single-copy genes were captured by BUSCO (Benchmarking Universal Single-Copy Ortholog) analysis using embryophyte dataset obd10 (Table S2), which indicates the assembly of the *C. korshinskii* genome was highly accurate and complete.

Using a combination of *de novo*, homology, and RNA-seq-based prediction methods, we obtained 36 219 protein-coding gene models from the *C. korshinskii* genome (Table S2). Compared with the other selected legume species, *C. korshinskii* has the longest average gene length (5833 bp), largely due to its long average intron length (1022 bp, Figure S4; Table S5). In total, approximately 93.35% of gene models can be

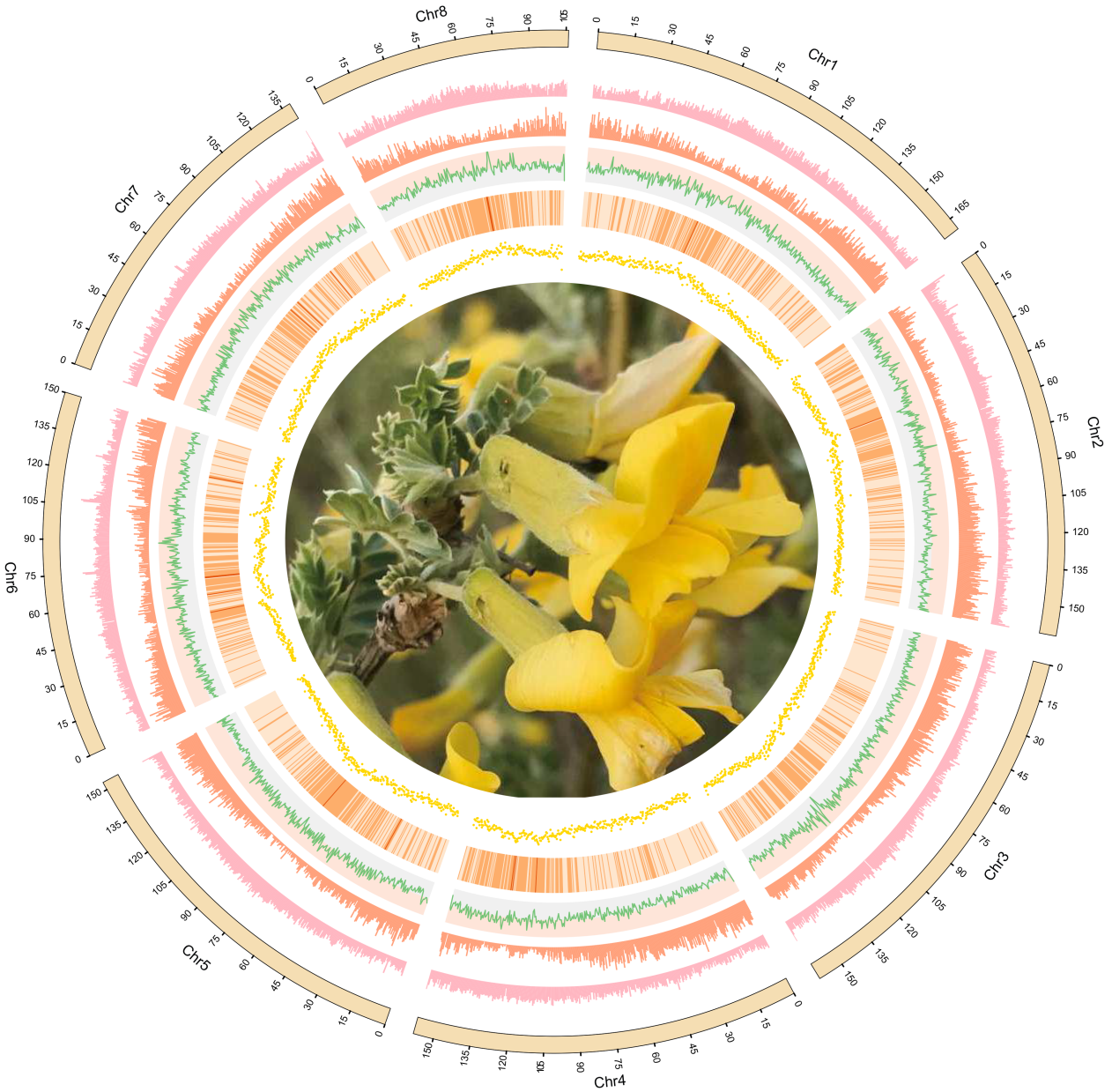

Figure 1. Genomic features of *Caragana korshinskii*.
Tracks from the outside to the inside correspond to pseudochromosomes, GC content, gene density, Gypsy density, Copia density, and density of all LTR retrotransposons, respectively. Density statistics are scaled with a 500-Kb sliding window. At the center of the circle are the flowers of *C. korshinskii*.

functionally annotated according to their homology with plant sequences from Gene Ontology (GO), SWISS-PROT, Eukaryotic Orthologous Groups of proteins (KOG), NCBI non-redundant (NR) protein sequences, and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (see "Materials and Methods" section, Table S6). Furthermore, about 66.23% of the total genome size is composed of transposable elements (TEs), where 50.72% was comprised of long terminal repeats (LTRs) (Table S7). We also predicted 404 ribosomal RNAs, 252 micro RNAs, 893 transfer RNAs, and 2543 small-nuclear RNAs (Table S8).

The lack of genus-specific WGD event

Polyploidy (or WGD) events are prevalent across diverse plant species and have been recognized as a key force in adaptive evolution and diversification (Jiao et al., 2011; Van de Peer et al., 2017; Wu et al., 2020). In addition to an ancient WGD event, known as the Papilionoid WGD,

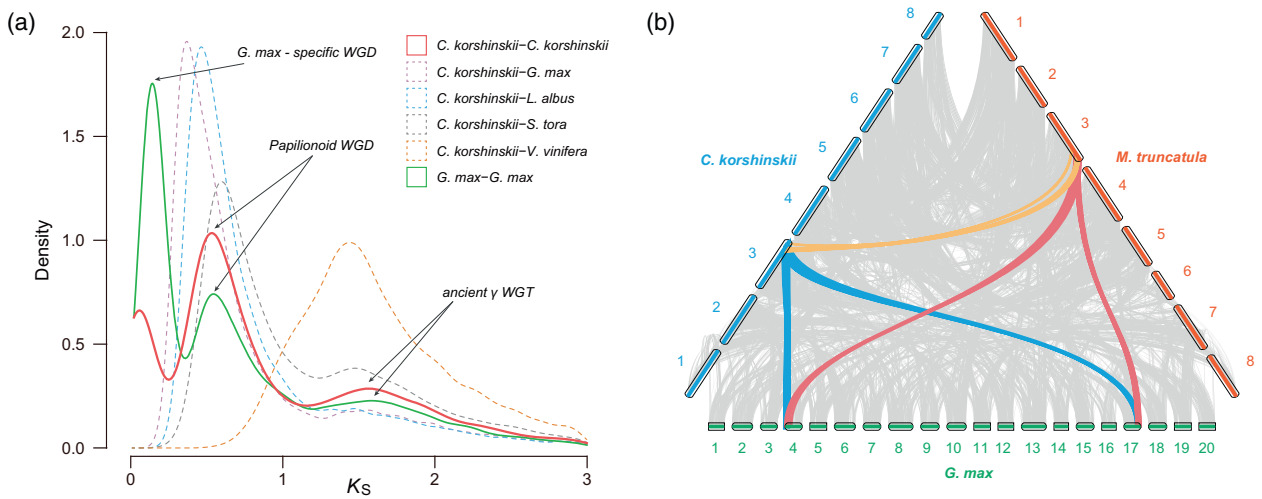

Figure 2. Evidence for the absence of lineage-specific whole-genome duplication (WGD) in *Caragana korshinskii*.
(a) Distribution of K_S values for anchor gene pairs in *C. korshinskii* and *G. max*, and one-to-one orthologs between *C. korshinskii* and four other legume species (dotted lines).
(b) Macro-syntenic comparisons between *C. korshinskii*, *Glycine max*, and *Medicago truncatula*.

shared by all papilionoid plants (Cannon et al., 2015), a recent phylotranscriptomic study of the genus *Caragana*, including *C. korshinskii*, has suggested a genus-specific WGD event (Zhao et al., 2021).

As the first genome in the genus *Caragana*, our chromosomal-scale assembly of *C. korshinskii* provided us with the opportunity to confirm this WGD event, as syntenic blocks are considered the most definitive evidence for such events (Wang et al., 2018). The synonymous substitution per synonymous site (K_S) calculation for anchor gene pairs (paralogs located on syntenic blocks) of *C. korshinskii* showed one small peak around 0.03 and two larger peaks around 0.46 and 1.61, respectively (Figure 2a). The occurrence timing of the second peak around 0.46 was slightly later than the split between *C. korshinskii* (Papilionoideae) and *Senna tora* (Caesalpinioideae, a sister subfamily to Papilionoideae), but earlier than the divergence of three Papilionoideae species (including *C. korshinskii*, *G. max*, and *Lupinus albus*). This confirms that the duplicated genes from this peak were derived from the Papilionoid WGD event (Figure 2a; Schmutz et al., 2010). Furthermore, the old peak around 1.61 corresponded to the well-known *Gamma* (γ) whole-genome triplication (WGT) event shared by core eudicots (Figure 2a; Jiao et al., 2012).

To determine whether the duplicated genes from the most recent peak around 0.03 were derived from a true WGD, we conducted intra- and inter-genomic synteny analyses within and between selected legume species. We used *Medicago truncatula*, which lacks a lineage-specific WGD, and *G. max*, which has undergone an additional lineage-specific WGD, as reference species. Our inter-genomic synteny analysis revealed syntenic block ratios of 1:1 for *C. korshinskii* versus *M. truncatula* and 1:2 for *C. korshinskii* versus *G. max*, respectively (Figure 2b). This indicates that *C. korshinskii*, like *M. truncatula*, did not undergo an additional WGD after the shared Papilionoid WGD. Furthermore, intra-genomic synteny analysis showed that the majority of large syntenic blocks within the *C. korshinskii* genome had average K_S values that aligning with those of the second peak, suggesting their derivation from the Papilionoid WGD event (Figure S5), rather than from a recent WGD corresponding to the K_S peak around 0.03. Consequently, our results provide clear evidence for the absence of a genus-specific WGD event in *C. korshinskii*, contrasting with previous findings from transcriptome data (Zhao et al., 2021).

Expanded and unique gene families are associated with drought adaptation

A robust phylogenetic tree was constructed for 11 selected species, including seven from the Fabaceae family and four outgroup species (Table S9; Figure 3a). This tree was based on 223 strict single-copy genes identified by grouping orthologs using OrthoFinder2 (Emms & Kelly, 2019). Molecular dating analysis indicated that the divergence between *C. korshinskii* and its close relatives (*P. sativum* and *M. truncatula*) occurred 30.67 million years ago (Ma) with a 95% highest posterior density range of 23.83–39.7 Ma (Figure 3a). The timing coincides with the onset of the Central Asian steppe-desert biome (Barbolini et al., 2020). Gene-family expansion and contraction analysis revealed more contracted (3951) than expanded (3021) gene families during the lineage-specific evolution of *C. korshinskii* (Figure 3a). A similar trend was observed in *P. sativum*, but not in *M. truncatula* (Figure 3a). Notably, the lineage-specific expanded gene families in

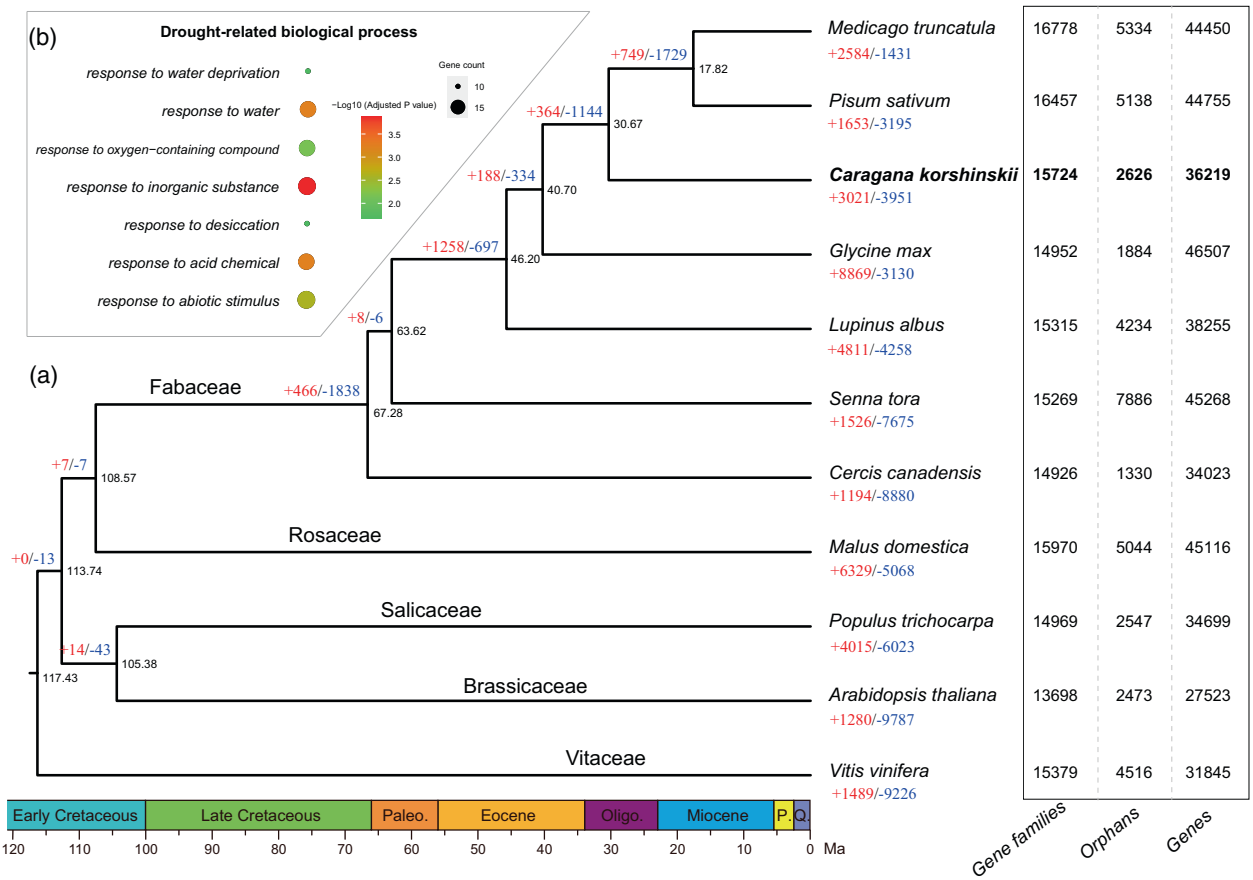

Figure 3. Phylogenetic and evolutionary analyses of the *Caragana korshinskii* genome.
(a) Time-calibrated phylogenetic tree and gene-family expansion and contraction analysis across 11 plant species. Black numerical values near the split nodes indicate estimated divergent times. The numbers of expanded (in red) and contracted (in blue) gene families at each node or within each species are shown above the branches or below the species names, respectively. Counts of gene families, orphan genes, and total genes are displayed next to each species.
(b) Significantly enriched Gene Ontology (GO) terms for drought-related biological processes in the expanded gene families of the *C. korshinskii* genome. Circle size represent the number of genes, and colors indicate the degree of significance, assessed by adjusted *P*-values.

C. korshinskii were functionally enriched in GO categories related to responses to water deprivation, desiccation, inorganic substances, oxygen-containing compound, acid chemical, and other abiotic stimuli (Figure 3b). These processes are closely associated with drought tolerance, suggesting that the expanded genes in *C. korshinskii* were likely selected for long-term survival and adaptation in arid environments.

Further analysis of shared and unique gene families among the seven legume species revealed that 10 973 gene families were commonly shared (Figure S6), while 1217 appeared to be unique to *C. korshinskii* (Figure S6). These unique families are involved in various biosynthetic and secondary metabolic pathways, such as flavonoid and fatty acid biosynthesis and metabolism, as well as DNA repair processes, including mismatch repair and nucleotide excision repair (Table S10). Interestingly, genes involved in DNA replication and repair in these families might play protective roles in ensuring the stable transfer of genetic information under the unfavorable conditions caused by drought.

Exploration of drought-responsive genes and the role of tandem duplications in the expansion of cellular protection genes

To understand the genomic basis of *C. korshinskii*'s remarkable drought tolerance, we investigated its gene members across several vital drought-responsive signaling pathways, including ABA synthesis, stomatal regulation, osmotic stress regulation, cellular protection, and reactive oxygen species (ROS) scavenging (Chen, Li, et al., 2020; Fang & Xiong, 2015; Zhu, 2002).

Phytohormone ABA is well known for its crucial roles in plant responses to drought stress. It accumulates rapidly under water-deficit conditions and activates a series of drought-responsive genes to optimize water use and improve dehydration tolerance (Chen, Li, et al., 2020). We analyzed genes involved in ABA

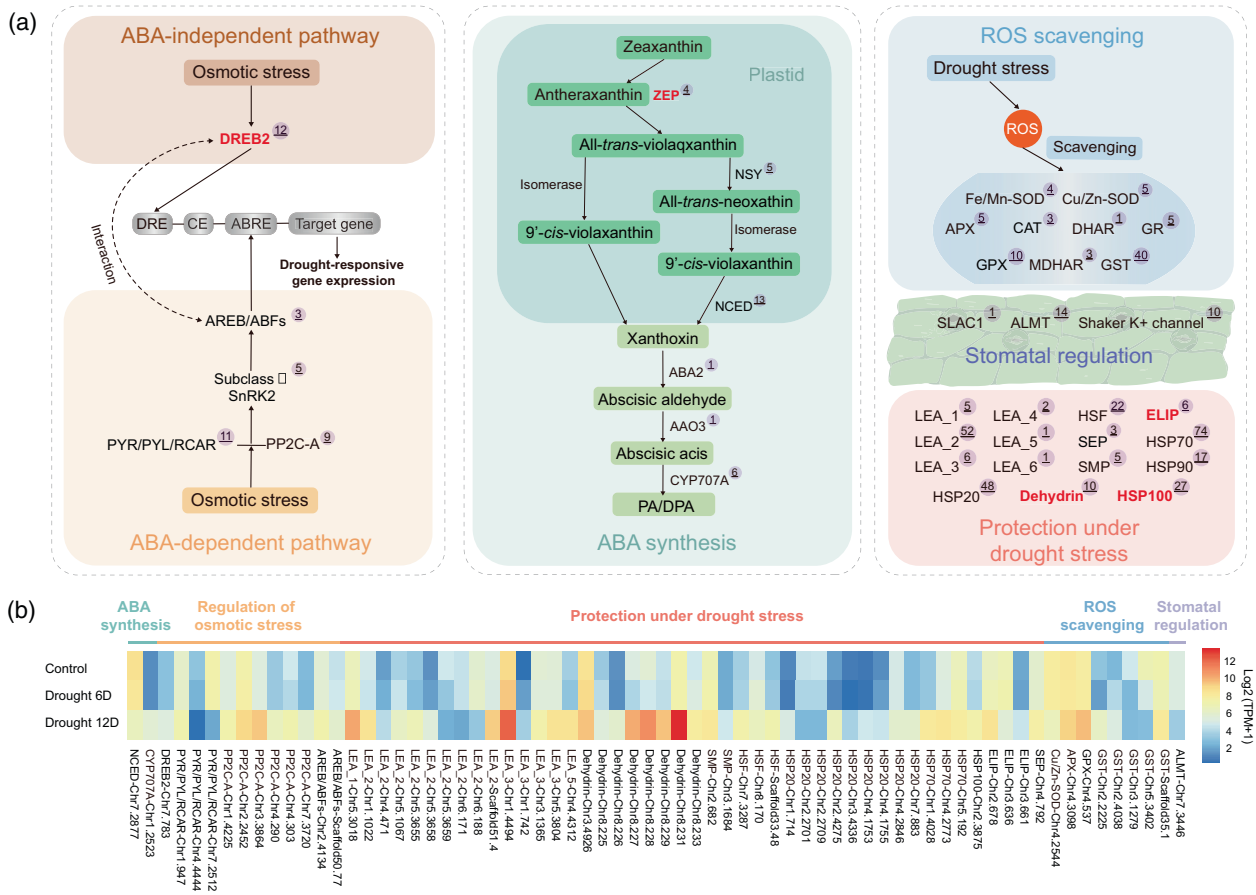

Figure 4. Exploration of drought-responsive genes in *Caragana korshinskii*.
(a) Detailed signaling pathways for each drought-responsive module, differentiated by various colors. Numerical values within circles represent gene copy numbers in *C. korshinskii*. Expanded genes are highlighted in red and bold.
(b) Gene expression heatmap of differentially expressed genes (DEGs) involved in drought-responsive pathways in *C. korshinskii*. Each row represents a specific gene, while columns distinguish between control and drought-treated samples. Expression levels are color-coded: red indicates higher expression and blue indicates lower expression.

biosynthesis, such as zeaxanthin epoxidase (ZEP), neoxanthin synthase (NSY), 9-cis-epoxycarotenoid dioxygenase (NCED), short-chain alcohol dehydrogenase (ABA2), and abscisic aldehyde oxidase 3 (AAO3), and metabolism (specifically cytochrome P450, family 707, subfamily A, CYP707A) in *C. korshinskii* and other species (Figure 4a; Figure S7). ZEP, which catalyzes the conversion of zeaxanthin to all-trans-violaxanthin in plastids (Figure 4a), is a key enzyme in the initial steps of ABA biosynthesis (Chen, Li, et al., 2020). *C. korshinskii* contains four *ZEPs*, more than typically found in other species, which usually have only one to three members (Figure S7). Following the increase in ABA levels, stomatal apertures act as the first guard to reduce water transpiration under drought conditions (Chen, Li, et al., 2020; Li et al., 2024). Gene families, such as shaker K^+ channel, slow anion channel 1 (SLAC1), and aluminum-activated malate transporter (ALMT), play crucial roles in controlling stomatal regulation, showing no significant copy number variation between *C. korshinskii* and other species (Figure 4a; Figure S7).

During drought, osmotic homeostasis is disrupted, impairing normal growth, development, and survival of plants (Gupta et al., 2020; Yoshida et al., 2014; Zhu, 2002). Both ABA-dependent and ABA-independent signaling pathways are pivotal in modulating plant responses to osmotic stress (Yoshida et al., 2014). We compared the copy numbers of ABA receptor Pyrabactin resistance 1/PYR1-like/Regulatory component of ABA receptor (PYR/PYL/RCAR), clade A protein phosphatases of type 2C (PP2C), subclass III SNF1-related protein kinase 2s (SnRK2s), and ABA response element binding (AREB/ABF) transcription factors (TFs) between *C. korshinskii* and other species, finding no significant differences (Figure 4a; Figure S7). In the ABA-independent pathway, dehydration-responsive element/C-repeat (DRE/CRT)-binding protein 2 (DREB2) TFs, which activate the expression of many osmotic response genes, were notably abundant with 12 DREB2s in

C. korshinskii, compared with 5–8 in most species (Figure S7). Phylogenetic and syntenic analyses revealed that four of these members were generated by tandem duplications, contributing to this expansion (Figure S8a,b). However, these expanded genes did not show differential expression (Figure S5c). Among the total 12 DREB2 TFs, only one (*Chr7.783*) exhibited a significant increase in expression following a 12-day drought treatment, while two others (*Chr3.1777* and *Chr3.4819*) also showed increased expression under the same conditions (Figure 4b; Figure S8c).

Drought not only induces osmotic stress but also causes additional stress injuries in plants (Zhu, 2002). Cellular desiccation during drought can lead to protein denaturation or inactivation and the generation of ROS (Gupta et al., 2020; Yoshida et al., 2014; Zhu, 2002). To combat these effects, the expression of protective proteins, including late embryogenesis abundant (LEA) proteins, heat shock proteins (HSPs), early light-induced proteins (ELIPs), and stress-enhanced proteins (SEPs), act as a secondary line of defense against drought stress (Figure 4a). The Dehydrin (DHN) subfamily of LEA proteins, known for their chaperone-like functions preserving membrane stabilization and stabilizing labile proteins (Hundertmark & Hincha, 2008), is notably expanded in *C. korshinskii*, with copy numbers approximately twice as high as those found in other legume species (Figure S7). ELIPs are known to protect chlorophyll under drought conditions (Heddad et al., 2006). *C. korshinskii* contains six ELIP members, compared with only two in *A. thaliana* (Figure S7). HSP proteins play important roles to stabilize protein structure (Swindell et al., 2007). The HSP100 subfamily of the HSP gene family is tremendously expanded in *C. korshinskii*, including 27 genes, compared with other species with 6 to 17 members (Figure S7). Additionally, increased ROS causes oxidative damage and even cell death under drought stress (Tuteja et al., 2012). A range of antioxidant enzymes, including superoxide dismutase (SOD), ascorbate peroxidase (APX), catalase (CAT), dehydroascorbate reductase (DHAR), glutathione S-transferase (GST), and glutathione reductase (GR) are employed to remove ROS (Figure 4a). Among them, no substantial differences of copy numbers between *C. korshinskii* and other species were observed (Figure S7).

We specially investigated the evolutionary history and mechanisms behind the notable expansion of cellular protection genes in *C. korshinskii*. Phylogenetic analysis of the ELIP gene family across nine plant species showed that the six ELIP members in *C. korshinskii* are closely clustered (Figure 5a). Micro-syntenic analysis revealed that five of these six genes are tandemly located on Chromosome 3 (Figure 5b). All of these tandemly expanded genes exhibited increased expression levels under either 6- or 12-day drought treatments (Figure 5c). Further phylogenetic and syntenic analysis of the HSP100 subfamily indicated that 18 out of the 27 genes are tandemly arranged on Chromosomes 4 and 5. In contrast, *M. truncatula* and *G. max* do not display a similar pattern (Figure 5d,e), indicating a species-specific expansion of *HSP100* genes in *C. korshinskii*. Except one (*Chr5.794*), all of these tandemly expanded *HSP100* genes were induced by drought treatment (Figure 5f).

In addition to changes in gene copy numbers, we further investigated the alteration in drought-induced gene expression by comparing normal and drought-treatment conditions in *C. korshinskii* (see "Materials and Methods" section). We identified a total of 1974 differentially expressed genes (DEGs), with 997 upregulated and 977 downregulated (Table S11). We further validated a subset of these DEGs through quantitative real-time PCR (RT-qPCR), which confirmed high consistency with our transcriptome data, as shown in Figure S9. The upregulated DEGs predominantly participated in 'response to water', 'response to oxygen−containing compound', and 'cellular homeostasis' (Figure S10). In contrast, the downregulated DEGs were mainly involved in the 'carbohydrate metabolic process' and 'photosynthesis' (Figure S11). Among these, we identified 68 DEGs as members of the aforementioned drought adaptation processes, including two involved in ABA synthesis and metabolism, one in stomatal regulation, 12 in osmotic stress regulation, 45 in cellular protection, and eight in ROS scavenging (Figure 4b).

Tandem amplification of *DHN* genes enhances drought resistance in *C. korshinskii*

During our investigation into the drought-responsive genes in *C. korshinskii*, we observed significant amplification of *DHN* genes (Figures 4 and 6a; Figure S7). Notably, the orthologous gene *XERO2* (*AT3G50970*) in *A. thaliana* is documented to protect membranes under drought stress by cross-linking with lipids (Gupta et al., 2019). We conducted phylogenetic analysis of the *DHN* gene family across seven legume species and *A. thaliana*, revealing pronounced gene duplication in *C. korshinskii* compared with the orthologous *XERO2* in *A. thaliana* (Figure S12). Micro-syntenic analysis identified seven tandem duplicates of *XERO2* on chromosome 8 in *C. korshinskii*, in contrast to a single copy in *M. truncatula* and two in *G. max* (Figure 6a). Transcriptomic evidence indicated that all the seven members were strongly induced by drought (Figure 6b), suggesting their critical roles in drought tolerance.

To verify the contribution of these *DHNs* to drought tolerance in *C. korshinskii*, we performed *Agrobacterium*-mediated transient transfection experiments with each of the seven *DHNs* (Liu et al., 2019). After 12 days of drought treatment, most control seedlings infiltrated with the pCanG-HA empty vector were wilted and lodged, and their

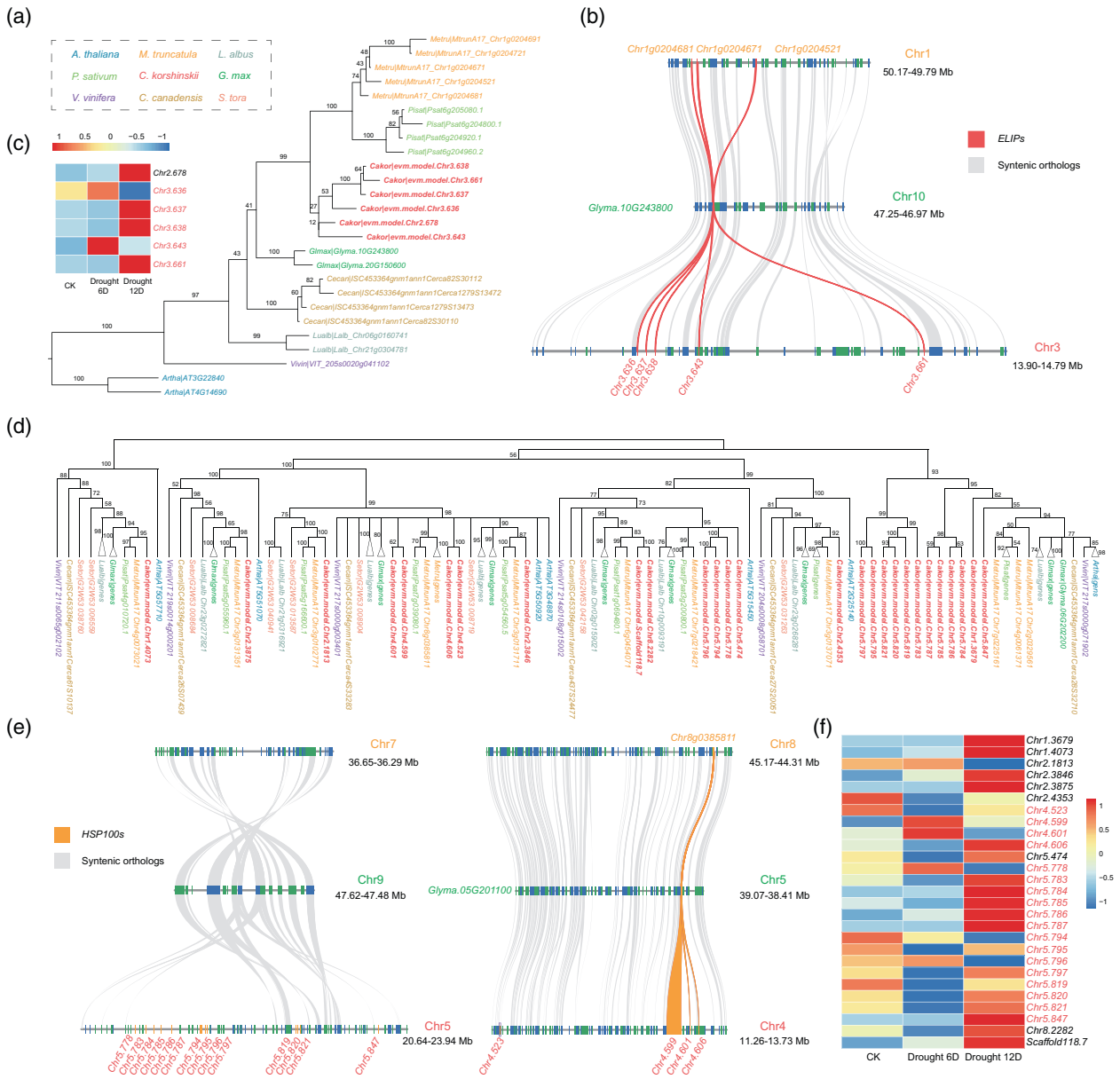

Figure 5. Tandem duplications contribute to the expansion of early light-induced protein (ELIP) and HSP100 gene families in *Caragana korshinskii*.
(a) Phylogenetic tree illustrating ELIP genes across nine plant genomes. The color of each gene corresponds to the background color of each species in the upper left.
(b) Five ELIP genes are tandemly located on chromosome 3 of the *C. korshinskii* genome. Micro-synteny is illustrated with corresponding orthologs in *Medicago truncatula* and *Glycine max*.
(c) Heatmap depicts gene expression levels of ELIP genes between control and drought treatment groups at 6 and 12 days. Warm colors donate high expression levels.
(d) Phylogenetic tree illustrating HSP100 genes across nine plant genomes.
(e) Micro-synteny illustrates the tandem gene clusters of HSP100 genes located on chromosome 4 and 5 of the *C. korshinskii* genome, alongside corresponding orthologs in *M. truncatula* and *G. max*.
(f) Heatmap depicts gene expression levels of HSP100 genes between control and drought-treatment groups at 6 and 12 days.

leaves became yellowish, indicating drought-stressed phenotypes (Figure 6c). However, the damage caused by drought stress was not as severe for the seedlings with transient expressions of each of the seven *DHNs*. The seedlings with transient expressions of the *Chr8.227* and *Chr8.229 DHNs* were especially healthier than the others (Figure 6c). By calculating the survival rates of leaves, we found that the control seedlings displayed a low survival rate (26.7%), whereas the transgenic seedlings of each of the seven *DHNs* showed higher survival rates. In particular,

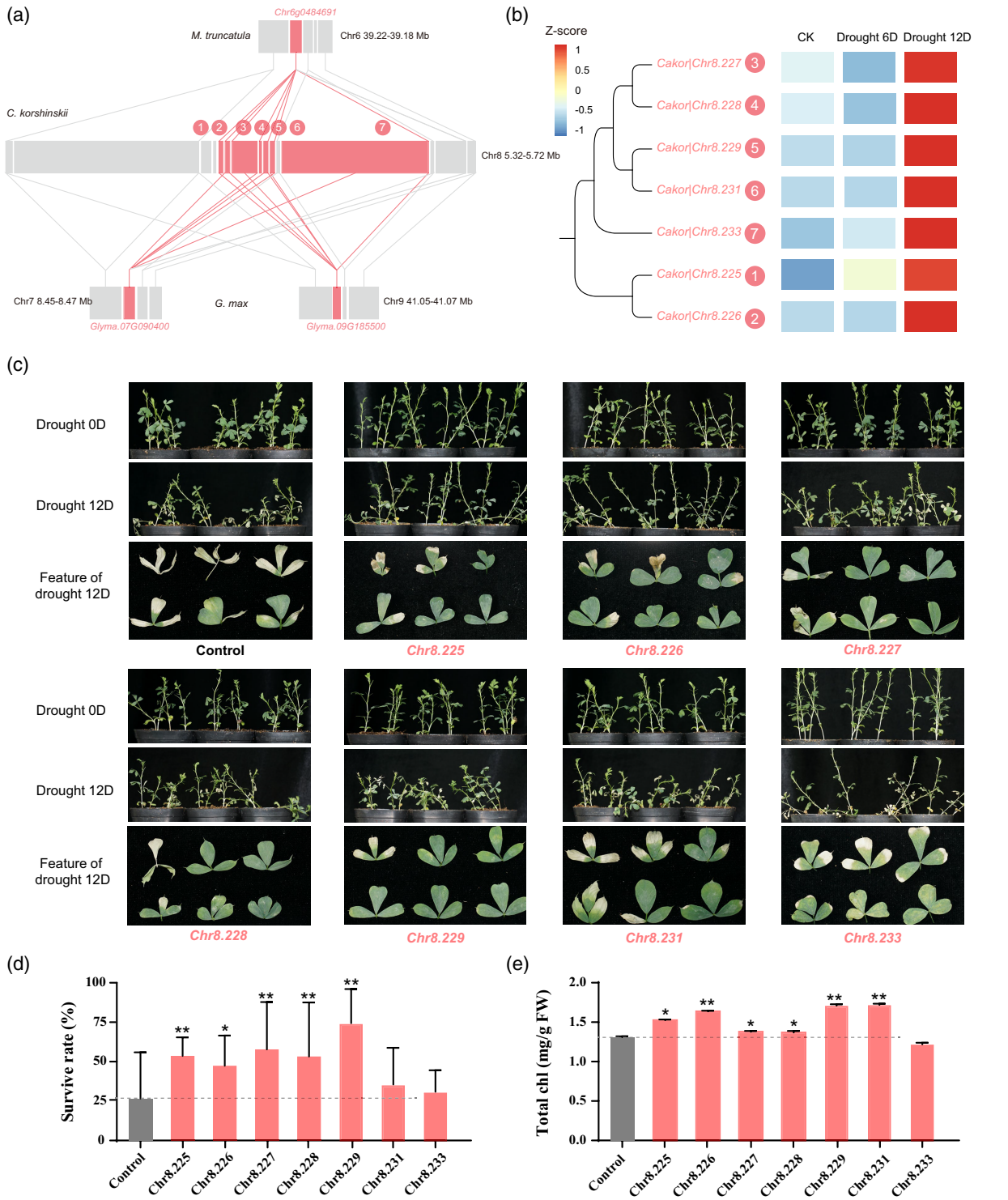

Figure 6. Seven tandemly duplicated *DHN* genes greatly improved drought tolerance in *Caragana korshinskii*.
(a) Seven *DHN* genes are tandemly located on chromosome 8 of the *C. korshinskii* genome. Micro-synteny is shown with corresponding orthologs in *Medicago truncatula* and *Glycine max*.
(b) Phylogenetic tree of the seven *DHN* genes in *C. korshinskii* and their expression differences under control and drought treatments.
(c) Morphological features of control and *C. korshinskii* seedlings transiently expressing each of the *DHN* genes, observed before and after drought treatment.
(d) Survival rates of both control and *DHN*-transgenic *C. korshinskii* seedlings following a 12-day drought treatment.
(e) Total chlorophyll content in both control and *DHN*-transgenic *C. korshinskii* seedlings under identical conditions. Statistical significance is denoted by asterisks: $*P < 0.05$; $**P < 0.01$.

The Plant Journal, (2025), doi: 10.1111/tpj.17255

the survival rates of *Chr8.227* and *Chr8.229* were 57.8 and 74.4%, respectively (Figure 6d). Additionally, the total chlorophyll contents of six *DHNs* transient expression seedlings were higher than that of the controls, and the results were consistent with their phenotypes and survival rates (Figure 6e). These data suggest that the duplicated *DHN* genes have contributed to the enhanced drought tolerance observed in *C. korshinskii*.

Virus-induced gene silencing (VIGS) has been demonstrated as an effective tool for gene expression knockdown without the need for stable transformation (Burch-Smith et al., 2006; Liu et al., 2002; Robertson, 2004; Ruiz et al., 1998). However, the efficiency of VIGS had not previously been tested in *C. korshinskii*. We constructed the *CkPDS* silencing vector and observed that silenced *C. korshinskii* seedlings displayed a spectrum of phenotypes, ranging from green and yellowish-green to partially or fully white leaves (Figure S13a). The distribution of these phenotypes is detailed in Table S12. The overall VIGS efficiency in *C. korshinskii* seedlings was approximately 54.4%. Furthermore, expression analysis confirmed a significant reduction in *CkPDS* expression in the silenced seedlings compared with the control (Figure S13b), validating the effectiveness of the VIGS system in *C. korshinskii*.

To investigate whether silencing *DHNs* could reduce drought tolerance in *C. korshinskii*, we constructed a silencing vector targeting seven *DHNs*. Under normal growth conditions, there were no observable differences between the control and silenced seedlings (Figure S14a). However, after 12 days of drought treatment, silenced seedlings exhibited more pronounced symptoms of drought stress, characterized by extensive drying and yellowing of leaves, compared with the control (Figure S14a). We measured the transcript level of *DHNs* in both groups. As depicted in Figure S14(b), with the exception of *Chr 8.233*, whose expression could not be detected—likely due to a robust silencing effect by VIGS—the expression levels of five out of the six detectable *DHNs* (except for *Chr 8.226*) were significantly reduced in the silenced group relative to the control group. These findings further corroborate the significant role of duplicated *DHN* genes in conferring drought tolerance to *C. korshinskii*.

DISCUSSION

A high-quality genome is crucial for uncovering specific genomic changes that underpin a species' unique adaptability (Ma et al., 2013; Moghaddam et al., 2021; Wan et al., 2021). While genomic studies are prevalent across various crop species, they are less frequently conducted in indigenous plants. In this study, we presented the genome sequence of the Korshinsk peashrub, a species indigenous to and extensively used for vegetation restoration and as forage in the drylands of northern China (Liu et al., 2010). Although the genome assembly of *C. korshinskii* is at the chromosome level and exhibits high completeness (BUSCO score of 97.6), the considerable heterozygosity rate of 1.61% has resulted in approximately 11% of the contigs remaining unanchored to chromosomes. Further improvements could be achieved by employing advanced sequencing and assembly strategies, such as ultra-long read sequencing technologies and haplotype-aware genome assembly. Nevertheless, this genome represents the inaugural genome resource for the genus *Caragana* in the legume family. Furthermore, we have developed CakorDB, a freely accessible genomic database, to support functional genomic research and foster international collaboration on the Korshinsk peashrub.

Based on previous K_S- and phylogenomic-based analyses of transcriptomes from five *Caragana* species, including *C. korshinskii* (Zhao et al., 2021), a WGD event was postulated for the genus. Zhao et al. (2021) reported that a relatively high proportion of gene duplications (5.8%, equivalent 503 gene duplications) occurred in the common ancestor, involving a total of 8673 gene families. However, our inter- and intra-genomic analyses using a chromosomal-scale genome of *C. korshinskii* did not reveal any syntenic signals indicative of a WGD, as shown in Figure 2. In fact, several recent studies have highlighted the limitations of relying solely on transcriptomic data for estimating WGD events. These studies have pointed out that transcriptomes may not cover all genomic genes and often include multiple transcripts per gene, which can introduce biases in the analysis (Nakatani & McLysaght, 2019; Zwaenepoel et al., 2019).

Although our results do not support the occurrence of a WGD event, we identified tandem duplication as a pivotal mechanism for the expansion of gene families involved in cellular protection against drought stress, as illustrated in Figures 5 and 6a. Notably, the substantial amplification of *DHNs* through tandem duplications in *C. korshinskii* has significantly enhance its drought tolerance, as evidenced by transgenic experiments (Figure 6; Figure S14). Similarly, the HSP100 subfamily has undergone a major expansion in *C. korshinskii* (Figure 4a; Figure S7), with over 70% of its members originating from tandem duplications (Figure 5d–f). It is well-documented that HSP100 proteins are critical for addressing protein misfolding and aggregation within cells and organelles (Waters & Vierling, 2020), particular in rescuing stress-damaged proteins under adverse conditions (Lee et al., 2022). The expression of HSP100 genes, typically very low under normal conditions, is highly upregulated in response to drought and heat stress. As expected, almost all of the expanded genes showed high expression levels under drought treatments, indicating their roles as protective molecular chaperones (Figure 5f). In addition to these protective and damage repair proteins, we observed that *C. korshinskii* shares a common feature with most resurrection plants: the expansion of ELIP genes through tandem duplications (Costa et al., 2017; Marks

et al., 2024; Silva et al., 2021; Xiao et al., 2015; Xu et al., 2018). ELIP proteins function as transient pigment-binding proteins that protect chloroplasts from severe desiccation stress. Consequently, the expansion of *ELIPs* likely endows *C. korshinskii* with a similar desiccation-tolerant capability. Therefore, *C. korshinskii* acts as a resilient 'warrior' in arid environments, equipped with enhanced cellular protection capabilities conferred by tandem duplications.

For non-model species, particularly indigenous and woody plants, functional validation of key candidate genes is challenging due to the absence of effective *in vivo* transformation systems. In this study, to assess the function of the expanded *DHN* genes in *C. korshinskii*, we specifically developed a VIGS system. Both overexpression and silencing of these expanded *DHNs* confirmed their critical roles, as illustrated in Figure 6 and Figure S14. These unique genes represent valuable resources for the genetic improvement of crops and other plants, enhancing their tolerance to drought conditions.

MATERIALS AND METHODS

Plant materials and drought treatment

Young, fresh leaves from an individual of *C. korshinskii* plant were collected from the Yuzhong campus of Lanzhou University for genome sequencing. Healthy seeds were selected and sown in a 1:3 mixture of peat soil and vermiculite in a greenhouse (22°C, 16 h light/8 h dark). For drought treatments, the methods previously described in Wan et al. (2018) were used with slight modifications. Briefly, 4-week-old *C. korshinskii* seedlings were subjected to drought treatment by withholding water for 0, 3, 6, 9, 12, and 15 days. Morphologically, the 12 and 15 days' seedlings showed obvious drought-stressed phenotypes (Figure S15). Thus, the shoots taken at 0 and 12 days were used to detect gene expression and for transcriptome sequencing. Each sample contained the shoot mixture of three seedlings.

DNA extraction and library construction and sequencing

Genomic DNA were extracted from young *C. korshinskii* leaves using a QIAGEN® Genomic kits (Qiagen, Hilden, Germany) for genome sequencing, following the manufacturer's instructions. A DNA library of next-generation short reads was constructed and sequenced by using the DNBSEQ™ platform (BGI, Shenzhen, China) in 150-bp paired-end mode. For Oxford Nanopore Technologies (ONT) long read sequencing, nanopore libraries were constructed and sequenced by PromethION sequencer (ONT, Oxford, UK) after fragment size selection. The Hi-C library was constructed and sequenced to assist with genome assembly. The cross-linked DNA was fixed in the cell nucleus and cut with restriction enzyme *Mbol*. The filled ends were marked with biotin and then were ligated with ligase. The DNA was then purified and sheared to pull down the biotin, and sequenced using the DNBSEQ™ platform.

Genome assembly

Genome size was estimated using KmerGenie (http://kmergenie.bx.psu.edu/), which can automatically determine the optimal *K*-mer size for estimating genome size. The ONT long reads were assembled into contigs using NECAT v1.0 (Chen et al., 2021). The assembled contigs were subsequently corrected three times with long reads using Racon v1.4.5 (Vaser et al., 2017) and once with short reads using Pilon v1.23 (Walker et al., 2014). To further enhance assembly contiguity, Hi-C raw reads were used for scaffolding with the 3d-dna pipeline (Dudchenko et al., 2017). The accuracy of the Hi-C-assisted assembly was assessed based on a chromatin contact matrix using Juicerbox Assembly Tools (https://github.com/aidenlab/Juicebox), and some errors were manually adjusted. The final assembly was validated by mapping short reads back to the assembled genome using BWA (Li & Durbin, 2009). The completeness of the assembled genome was accessed using BUSCO v4.05 (Manni et al., 2021).

Genome annotation

Transposable elements in the *C. korshinskii* genome were identified using both *de novo* prediction and evidence-based search approaches, following the procedures outlined by Ma et al. (2023). We used three approaches to predict protein-coding genes in the repeat-masked genome, including *ab initio* prediction, homology search, and transcriptome-guided annotation. Homologous protein sequences from five species (*A. thaliana*, *G. max*, *M. truncatula*, *S. tora*, and *V. vinifera*) were aligned to the assembled *C. korshinskii* genome using GeMoMa v1.9 (Keilwagen et al., 2016). Genetic structural information was inferred from these alignments. Filtered RNA-seq reads were aligned to the reference genome using STAR v2.7.10a (https://github.com/alexdobin/STAR/). The transcripts were then assembled using StringTie (Pertea et al., 2016) and open reading frames (ORFs) were predicted using PASA pipeline (Haas et al., 2008). For *de novo* gene prediction, a training set was built and predictions were made using Augustus v3.4.0 (Stanke et al., 2006). An integrated gene set was produced using EVidenceModeler (EVM) v1.1.1 (Haas et al., 2008). Genes primarily occupied with transposable domain (>60% gene length) were removed using TransposonPSI (http://transposonpsi.sourceforge.net), and miscoded genes were further filtered. Untranslated regions (UTRs) were determined using PASA based on RNA-seq assemblies, retaining the longest transcripts for each locus. For functional annotation of gene models and the annotation of non-coding RNAs (ncRNAs), we utilized the same pipelines as described in our previous work (Ma et al., 2023).

Synteny and WGD identification

For intra- and inter-genomic synteny analyses, we performed all-against-all BLASTP searches for putative homologous genes using an E-value cutoff of 1e-05. We employed MCScanX (Wang et al., 2012) with default parameters to infer syntenic blocks within and between genomes, using the top 10 BLAST matches as inputs. We calculated the synonymous substitution per synonymous site (K_S) for anchor gene pairs using the Nei-Gojobori method (Nei & Gojobori, 1986) implemented in the YN00 program of PAML v4.9 (Yang, 2007). The intra- and inter-genomic syntenic gene pairs were visualized by plotting their locations and K_S values using WGDI software (Sun et al., 2022). Additionally, macro-syntenic ratios among *M. truncatula*, *C. korshinskii*, and *G. max* were analyzed using the Python version of MCScan (Wang et al., 2012), further exploring the extent of genome duplications.

Phylogenomic and gene family evolution analyses

In addition to *C. korshinskii*, we selected six Fabaceae species (*M. truncatula*, *P. sativum*, *G. max*, *L. albus*, *S. tora*, and *C. canadensis*) and four other eudicot species (*M. domestica*, *P. trichocarpa*, *A. thaliana*, and *V. vinifera*) for phylogenomic analysis (Table S9). Protein sequences, represented by the longest

transcript for each gene, underwent BLASTP comparisons and subsequently classified into gene families using Orthofinder2 software with default parameters. For species-tree construction, we identified 223 strict single-copy genes. The amino acid sequences of these genes were aligned using MAFFT v7.490 (Katoh et al., 2005), and the resulting alignments were converted into corresponding codon alignments using PAL2NAL (Suyama et al., 2006). The maximum likelihood phylogenetic trees for each single-copy gene family were inferred using IQ-TREE2 (Minh et al., 2020), and the coalescent species tree was summarized using ASTRAL (Zhang et al., 2018).

To estimate the divergence times, particularly between *C. korshinskii* and the selected legume species, we employed MCMCTREE (Yang, 2007) utilizing three fossil constraints: (1) the divergence between *P. trichocarpa* and *A. thaliana* estimated at 98–117 Ma, and (2) the divergence between *V. vinifera* and other species at 106–125 Ma – both constraints sourced from TimeTree (Kumar et al., 2022); and (3) the crown age of Fabaceae at 67.20–67.36 Ma, as reported by Zhao et al. (2021). Based on the time-calibrated phylogenetic tree, we assessed the dynamics of gene family evolution (i.e., expansion or contraction) using CAFÉ v.5.0 (Mendes et al., 2020). The GO enrichment analyses for the expanded, contracted, and unique gene families of *C. korshinskii* were performed with TBtools (Chen, Chen, et al., 2020.

Copy number analyses of drought-responsive genes

To identify homologous drought-responsive genes, known *A. thaliana* members were used as query sequences to search for similar protein sequences in *C. korshinskii* and other selected species using BLASTP and HMMER search (Johnson et al., 2010). For the BLASTP search, we set the *E*-value to <1e-05 and sequence identity to >30%. The conversed domains of these protein sequences were subsequently searched using HMMER 3.3.2 (Johnson et al., 2010). The domain structure model files for the protein families were downloaded from Pfam (https://pfam.xfam.org/). Following this, the candidate genes along with the known *A. thaliana* members of each gene family underwent phylogenetic tree construction. Each tree was manually examined to exclude any outliers. Gene copy numbers for each species within each gene family were manually counted. The gene IDs corresponding to each family across nine plant genomes are listed in Data S1.

Transcriptome sequencing and DEGs analyses

Transcriptomes sequencing was carried out by the Majorbio Company (Shanghai, China). Total RNA was extracted from *C. korshinskii* seedlings with TRIzol® Reagent (Invitrogen, Waltham, MA, USA) and genomic DNA was removed using DNase I (TaKaRa, Kusatsu, Shiga, Japan). Only high-quality RNA samples (OD260/280 = 1.8–2.2, OD260/230 ≥2.0, RIN ≥6.5, 28S:18S ≥1.0, >1 μg) were used for library construction. The transcriptome library was set up following the TruSeqTM RNA sample preparation Kit from Illumina (San Diego, CA, USA) using 1 μg of total RNA. The RNA was then sequenced with an Illumina NovaSeq 6000 sequencer. Raw paired-end reads were trimmed and quality controlled by FastQC (https://github.com/s-andrews/FastQC). Clean reads were aligned to the reference genome and assembled using HISAT2 and StringTie pipeline (Pertea et al., 2016). Transcriptional expression levels were calculated according to the transcripts per million reads (TPM). DEGs with a differential expression level of FDR ≤0.5 and at least a twofold expression change were identified using DESeq2 (Love et al., 2014). GO and KEGG enrichment analyses were carried out using TBtools (Chen, Chen, et al., 2020) and KOBAS (Bu et al., 2021).

RNA extraction and real-time RT-qPCR assay

RT-qPCR was performed according to Yang et al. (2014). The total RNA extraction kit (Tiangen, Beijing, China) was used to extract total RNA, and RNA quality and quantity was evaluated by agarose gel electrophoresis using a Quawell Q5000 micro volume UV–Vis spectrophotometer (San Jose, CA, USA). Following, 1 μg of total RNA was reverse transcribed into cDNA using PrimeScript™ II 1st Strand cDNA Synthesis Kit (TaKaRa, Japan). RT-qPCR was carried out with a Roche LightCycler 480 Real Time PCR system (Roche, Switzerland) using LightCycler 480 SYBR Green Master Mix (Roche). The cycling program was 95°C for 30 sec, followed by 40 cycles of 95°C for 5 sec, 60°C for 30 sec and 72°C for 15 sec. The melting curves were analyzed at 60–95°C after 40 cycles. All RT-qPCR reactions were carried out with three technical replicates. The relative mRNA transcript levels of the genes were calculated according to the $2^{-\Delta CT}$ and $2^{-\Delta\Delta CT}$ method described in Livak and Schmittgen (2001). *CkEF1α* was chosen as the reference gene for the RT-qPCR analysis (Yang et al., 2014). RT-qPCR primers used are listed in Table S13. Each experiment was independently repeated three times.

DHN genes cloning and plasmids construction

Full-length coding sequences (CDSs) of the *DHN* genes were amplified from the cDNA of *C. korshinskii* cDNA using PrimeSTAR® HS DNA Polymerase (TaKaRa, Japan) and cloned into a pCanG-HA expression vector (a gift from Dr. Qi Xie, Institute of Genetics and Developmental Biology, Chinese Academy of Science) using an In-Fusion® HD Cloning Kit (TaKaRa, Japan). The generated binary plasmids were confirmed by sequencing and enzyme validation. The binary plasmids were then transformed into *Agrobacterium tumefaciens* strain GV3101 by electroporation for transient transformation in the leaves of *C. korshinskii* in next step. Genes cloning primers used are listed in Table S14.

Vector construction for VIGS of *CkPDS* and *CkDHNs*

To silence the *PDS* of *C. korshinskii*, we identified the only one ortholog of the well-known *AtPDS* (*Phytoene desaturase*) (Wang et al., 2005), from the genome data of *C. korshinskii* (https://bis.zju.edu.cn/cakordb/). Because *CkDHNs* are highly conserved, we therefore tried to silence multiple target genes with one insertion fragment, and designed primers according to the *Chr8.229* sequence (Figure S16). The inserted fragments of either *CkPDS* or *CkDHNs* were amplified from the cDNA of *C. korshinskii*, using PrimeSTAR® HS DNA Polymerase (TaKaRa, Japan). The primers' sequence was listed in Table S14. Ligation of the inserted fragments to TRV2 vector (TRV1 and TRV2 vectors are gifts from Dr. Yule Liu, School of Life Sciences, Tsinghua University) was carried out using an In-Fusion® HD Cloning Kit (TaKaRa, Japan). The generated binary plasmids were then confirmed by sequencing and enzyme validation (*Xba*I and *Kpn*I). The binary plasmids were then transformed into *A. tumefaciens* strain GV3101 by electroporation. For VIGS phenotype observation, an *Agrobacterium* mixture containing either TRV_1 and TRV_2 empty vector (the control group) or TRV_1 and $TRV_{(PDS)}$/$TRV_{(DHNs)}$ vector (the silencing group) was injected into leaves of *C. korshinskii* at a 1:1 (v/v) ratio, respectively, as described in Liu et al. (2019). The *CkPDS* silenced phenotype was visible after 1 week, and became very obviously after 2 weeks.

Drought tolerance assay of the transformed *C. korshinskii* seedlings

As described in Liu et al. (2019), seedlings of *C. korshinskii* grown in soil for approximately 3 weeks with sufficient watering were used for transient transformation. The *A. tumefaciens* strain GV3101 containing the above constructed vectors were collected when the OD_{600} reached 1.3–1.5. Then, the *Agrobacterium* was centrifuged at 3214 ***g*** for 10 min at 4°C. The *Agrobacterium* pellet was resuspended and diluted to an OD_{600} of 0.7–0.8 using ½ MS medium that was supplemented with 100 μmol L^{-1} acetosyringone. The bacteria suspension was rested at room temperature for 3–4 h, with inversions every half hour to resuspend the *Agrobacterium*. Silwet L-77 was added to the *Agrobacterium* solution to reach the final concentration of 0.001% (v/v) before injection. The evenly mixed *Agrobacterium* cell suspension was injected into the abaxial side of leaves of *C. korshinskii* seedlings, using a 1 ml disposable syringes without a needle. After being infiltrated for 2 days, the seedlings were subjected to the drought treatment by halting irrigation for *DHNs* over expression seedlings and the control. For VIGS, the seedlings of both groups were stopped watering 2 weeks after the injection with mixed *Agrobacterium* cell suspension, to allow the fully expanded virus infection. When the seedlings showed obvious wilting, yellowing, or lethal phenotypes caused by drought treatment for about 12 days, photographs were taken immediately. The survival rate of each group was then calculated, and the total chlorophyll content was measured. In each experiment, at least 32 *C. korshinskii* seedlings were selected for gene injection, drought treatment, and survival rate statistics, with each pot containing four seedlings. Additionally, three seedlings were randomly selected and pooled for RT-qPCR analysis and chlorophyll content measurement. All experiments were carried out with three biological replicates. Statistical significance differences from the control group were assessed using Student's *t*-test ($*P < 0.05$, $**P < 0.01$) and analyzed with GraphPad Prism8. Survival rate analyses were conducted across three biological replicates, while significance analyses for total chlorophyll content and gene expression levels were performed using three technical replicates.

AUTHOR CONTRIBUTIONS

SW, GL, and JL conceived the study. SW managed the genome-sequencing project. FM, XM, JW, and SW performed data analyses. TY, LS, QW, and GL performed functional validation of target genes. HC and MC constructed database. SW, FM, TY, GL, and JL wrote the paper. GR, QY, XXF, LQ, XWF, G, CG, and RW contributed to revisions. All authors read and approved the manuscript.

ACKNOWLEDGEMENTS

This research was partly supported by the National Natural Science Foundation of China-Gansu Joint Fund (U22A20452), the National Natural Science Foundation of China (32260372, 32471686, 32100170, 32260055, and 32325036), the Major Project of Inner Mongolia Natural Science Foundation (2024ZD17 and 2019ZD05), the Scientific and Technological Projects of Inner Mongolia (2023KJHZ0013 and 2023KYPT0016), and the University Scientific and Technological Innovation Team Project of Inner Mongolia (BR22-13-05 and NMGIRT2222), the Fundamental Research Funds for the Central Universities (lzujbky-2021-ct14), and the Gansu Science and Technology Major Project (22ZD6FA052). We received computational support from Supercomputing Center of Lanzhou University.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The assembled genome and annotation of Korshinsk peashrub can be obtained and viewed at CakorDB (https://bis.zju.edu.cn/cakordb/). Additional data necessary to reproduce the analyses presented in this study are available from the corresponding authors upon reasonable request.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Photographs depicting a mature individual (1), leaves (2), and fruits (3) of *C. korshinskii*, along with restored vegetation in the arid zones of Gansu Province, where *C. korshinskii* has been widely planted (4).

Figure S2. Genome survey of *C. korshinskii* using KmerGenie analysis.

Figure S3. Hi-C heatmaps depicting the interaction frequencies among the eight pseudochromosomes of *C. korshinskii*.

Figure S4. Comparison of intron lengths between *C. korshinskii* and other selected species.

Figure S5. Intra-genomic synteny within the *C. korshinskii* genome.

Figure S6. Upset plot illustrating the shared and unique gene families across seven legume species.

Figure S7. Heatmap displaying the copy numbers of genes involved in ABA synthesis, osmotic stress regulation, protection and repair proteins, ROS scavenging, and stomatal regulation in *C. korshinskii* and eight other plant species.

Figure S8. Phylogenetic tree of DREB2 transcription factors across nine plant genomes, alongside gene expression patterns under drought treatments.

Figure S9. The expression levels of DEGs in *C. korshinskii* were analyzed using both transcriptome data and RT-qPCR validation after 12 days of drought treatment.

Figure S10. Enriched GO terms for upregulated DEGs in *C. korshinskii* under drought treatment.

Figure S11. Enriched GO terms for downregulated DEGs in *C. korshinskii* under drought treatment.

Figure S12. Phylogenetic tree of *DHN* genes across nine plant genomes.

Figure S13. Detection of expression and the phenotypic alterations in *C. korshinskii* after *CkPDS* was silenced by VIGS system.

Figure S14. Reduced drought tolerance in *C. korshinskii* following *DHNs* gene silencing by VIGS.

Figure S15. Morphology of *C. korshinskii* seedlings under different days of drought treatments.

Figure S16. Alignment of nucleic acid sequences showing highly conserved regions across seven *DHN* gene sequences.

Table S1. Summary of DNBseq, ONT, and Hi-C sequencing data.

Table S2. Assembly and annotation features of the *C. korshinskii* genome.

Table S3. Basic statistics of the genome assembly for *C. korshinskii*.

Table S4. Mapping rate of short reads onto the assembled genome.

Table S5. Summary of gene, exon, and intron length information across 11 plant genomes.

Table S6. Functional annotation of gene models in the *C. korshinskii* genome.

Table S7. Summary of repeat sequences identified in the *C. korshinskii* genome.

Table S8. Overview of the annotated non-coding RNAs in the *C. korshinskii* genome.

Table S9. Information on the 11 plant genomes used in this study.

Table S10. Enriched KEGG pathways of the unique gene families in *C. korshinskii* (Corrected *P*-value <0.05).

Table S11. Number of DEGs in *C. korshinskii* under drought treatment.

Table S12. Different phenotypes of *C. korshinskii* leaves observed after VIGS silencing of the *PDS* gene.

Table S13. List of RT-qPCR primers used in this study.

Table S14. List of gene cloning primers used in this study.

Data S1. Gene IDs associated with each gene family across nine plant genomes.

REFERENCES

Azani, N., **Babineau, M.**, **Bailey, C.D.**, **Banks, H.**, **Barbosa, A.R.**, **Pinto, R.B.** ***et al.*** (2017) A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. *Taxon*, **66**, 44–77.

Barbolini, N., **Woutersen, A.**, **Dupont-Nivet, G.**, **Silvestro, D.**, **Tardif, D.**, **Coster, P.M.C.** ***et al.*** (2020) Cenozoic evolution of the steppe-desert biome in Central Asia. *Science Advances*, **6**, eabb8227.

Bu, D., **Luo, H.**, **Huo, P.**, **Wang, Z.**, **Zhang, S.**, **He, Z.** ***et al.*** (2021) KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research*, **49**, W317–W325.

Burch-Smith, T.M., **Schiff, M.**, **Liu, Y.** & **Dinesh-Kumar, S.** (2006) Efficient virus-induced gene silencing in Arabidopsis. *Plant Physiology*, **142**, 21–27.

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

Chen, C.J., **Chen, H.**, **Zhang, Y.**, **Thomas, H.R.**, **Frank, M.H.**, **He, Y.H.** ***et al.*** (2020) TBtools, an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, **13**, 1194–1202.

Chen, K., **Li, G.J.**, **Bressan, R.A.**, **Song, C.P.**, **Zhu, J.K.** & **Zhao, Y.** (2020) Abscisic acid dynamics., signaling., and functions in plants. *Journal of Integrative Plant Biology*, **62**, 25–54.

Chen, Y., **Nie, F.**, **Xie, S.Q.**, **Zheng, Y.F.**, **Dai, Q.**, **Bray, T.** ***et al.*** (2021) Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nature Communications*, **12**, 1–10.

Costa, M.D., **Artur, M.A.**, **Maia, J.**, **Jonkheer, E.**, **Derks, M.F.**, **Nijveen, H.** ***et al.*** (2017) A footprint of desiccation tolerance in the genome of *Xerophyta viscosa*. *Nature Plants*, **3**, 17038.

Dudchenko, O., **Batra, S.S.**, **Omer, A.D.**, **Nyquist, S.K.**, **Hoeger, M.**, **Durand, N.C.** ***et al.*** (2017) De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, **356**, 92–95.

Emms, D.M. & **Kelly, S.** (2019) OrthoFinder, phylogenetic orthology inference for comparative genomics. *Genome Biology*, **20**, 238.

Fang, Y.J. & **Xiong, L.Z.** (2015) General mechanisms of drought response and their application in drought resistance improvement in plants. *Cellular and Molecular Life Sciences*, **72**, 673–689.

Gao, B., **Li, X.**, **Liang, Y.**, **Chen, M.**, **Liu, H.**, **Liu, Y.** ***et al.*** (2024) Drying without dying: a genome database for desiccation-tolerant plants and evolution of desiccation tolerance. *Plant Physiology*, **194**, 2249–2262.

Gupta, A., **Marzinek, J.K.**, **Jefferies, D.**, **Bond, P.J.**, **Harryson, P.** & **Wohland, T.** (2019) The disordered plant dehydrin *Lti30* protects the membrane during water-related stress by cross-linking lipids. *Journal of Biological Chemistry*, **294**, 6468–6482.

Gupta, A., **Rico-Medina, A.** & **Caño-Delgado, A.I.** (2020) The physiology of plant responses to drought. *Science*, **368**, 266–269.

Haas, B.J., **Salzberg, S.L.**, **Zhu, W.**, **Pertea, M.**, **Allen, J.E.**, **Orvis, J.** ***et al.*** (2008) Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biology*, **9**, R7.

Han, X.M., **Feng, Z.Q.**, **Xing, D.D.**, **Yang, Q.**, **Wang, R.G.**, **Qi, L.W.** ***et al.*** (2015) Two NAC transcription factors from *Caragana intermedia* altered salt tolerance of the transgenic *Arabidopsis*. *BMC Plant Biology*, **15**, 208.

Heddad, M., **Norén, H.**, **Reiser, V.**, **Dunaeva, M.**, **Andersson, B.** & **Adamska, I.** (2006) Differential expression and localization of early light-induced proteins in *Arabidopsis*. *Plant Physiology*, **142**, 75–87.

Hu, Q., **Ma, Y.**, **Mándaková, T.**, **Shi, S.**, **Chen, C.**, **Sun, P.** ***et al.*** (2021) Genome evolution of the psammophyte *Pugionium* for desert adaptation and further speciation. *Proceedings of the National Academy of Sciences of the United States of America*, **118**, e2025711118.

Huang, J.P., **Yu, H.P.**, **Guan, X.D.**, **Wang, G.Y.** & **Guo, R.X.** (2016) Accelerated dryland expansion under climate change. *Nature Climate Change*, **6**, 166–171.

Hundertmark, M. & **Hincha, D.K.** (2008) LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics*, **9**, 118.

Jiao, Y., **Leebens-Mack, J.**, **Ayyampalayam, S.**, **Bowers, J.E.**, **McKain, M.R.**, **McNeal, J.** ***et al.*** (2012) A genome triplication associated with early diversification of the core eudicots. *Genome Biology*, **13**, R3.

Jiao, Y., **Wickett, N.J.**, **Ayyampalayam, S.**, **Chanderbali, A.S.**, **Landherr, L.**, **Ralph, P.E.** ***et al.*** (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473**, 97–100.

Johnson, L.S., **Eddy, S.R.** & **Portugaly, E.** (2010) Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, **11**, 431.

Katoh, K., **Kuma, K.**, **Toh, H.** & **Miyata, T.** (2005) MAFFT version 5, improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research*, **33**, 511–518.

Keilwagen, J., **Wenk, M.**, **Erickson, J.L.**, **Schattat, M.H.**, **Grau, J.** & **Hartung, F.** (2016) Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, **44**, e89.

Kumar, S., **Suleski, M.**, **Craig, J.M.**, **Kasprowicz, A.E.**, **Sanderford, M.**, **Li, M.** ***et al.*** (2022) TimeTree 5: an expanded resource for species divergence times. *Molecular Biology and Evolution*, **39**, msac174.

Lee, G., **Kim, R.S.**, **Lee, S.B.**, **Lee, S.** & **Tsai, F.T.** (2022) Deciphering the mechanism and function of Hsp100 unfoldases from protein structure. *Biochemical Society Transactions*, **50**, 1725–1736.

Li, H. & **Durbin, R.** (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, **25**, 1754–1760.

Li, S., **Wei, L.**, **Gao, Q.**, **Xu, M.**, **Wang, Y.**, **Lin, Z.** ***et al.*** (2024) Molecular and phylogenetic evidence of parallel expansion of anion channels in plants. *Plant Physiology*, **194**, 2533–2548.

Liu, K., **Yang, Q.**, **Yang, T.R.**, **Wu, Y.**, **Wang, G.X.**, **Yang, F.Y.** ***et al.*** (2019) Development of *Agrobacterium*-mediated transient expression system in *Caragana intermedia* and characterization of *CiDREB1C* in stress response. *BMC Plant Biology*, **19**, 237.

Liu, Y., **Schiff, M.** & **Dinesh-Kuma, S.P.** (2002) Virus-induced gene silencing in tomato. *The Plant Journal*, **31**, 777–786.

Liu, Y.X., **Chang, Z.Y.** & **Yakovlev, G.P.** (2010) *Caragana*. In: Wu, Z.Y. & Raven, P.H. (Eds.) *Flora of China*. Beijing: Science Press, pp. 528–545.

Livak, K.J. & **Schmittgen, T.D.** (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(−Delta Delta C(T)) method. *Methods*, **25**, 402–408.

Love, M.I., **Huber, W.** & **Anders, S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.

Ma, C.C., **Gao, Y.B.**, **Guo, H.Y.**, **Wang, J.L.**, **Wu, J.B.** & **Xu, J.S.** (2008) Physiological adaptations of four dominant *Caragana* species in the desert region of the Inner Mongolia plateau. *Journal of Arid Environments*, **72**, 247–254.

Ma, T., **Wang, J.Y.**, **Zhou, G.K.**, **Yue, Z.**, **Hu, Q.J.**, **Chen, Y.** ***et al.*** (2013) Genomic insights into salt adaptation in a desert poplar. *Nature Communications*, **4**, 2797.

Ma, X., **Ru, D.**, **Morales-Briones, D.F.**, **Mei, F.**, **Wu, J.**, **Liu, J.** ***et al.*** (2023) Genome sequence and salinity adaptation of the desert shrub *Nitraria sibirica* (Nitrariaceae, Sapindales). *DNA Research*, **30**, dsad011.

Manni, M., **Berkeley, M.R.**, **Seppey, M.**, **Simão, F.A.** & **Zdobnov, E.M.** (2021) BUSCO update, novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic., prokaryotic., and viral genomes. *Molecular Biology and Evolution*, **38**, 4647–4654.

Marks, R.A., **Van Der Pas, L.**, **Schuster, J.**, **Gilman, I.S.** & **VanBuren, R.** (2024) Convergent evolution of desiccation tolerance in grasses. *Nature Plants*, **10**, 1112–1125.

Mendes, F.K., **Vanderpool, D.**, **Fulton, B.** & **Hahn, M.W.** (2020) CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics*, **16**, btaa1022.

Minh, B.Q., **Schmidt, H.A.**, **Chernomor, O.**, **Schrempf, D.**, **Woodhams, M.D.**, **von Haeseler, A.** ***et al.*** (2020) IQ-TREE 2, new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**, 1530–1534.

Moghaddam, S.M., **Oladzad, A.**, **Koh, C.**, **Ramsay, L.**, **Hart, J.P.**, **Mamidi, S.** ***et al.*** (2021) The tepary bean genome provides insight into evolution and domestication under heat stress. *Nature Communications*, **12**, 2638.

Nakatani, Y. & **McLysaght, A.** (2019) Macrosynteny analysis shows the absence of ancient whole-genome duplication in lepidopteran insects. *Proceedings of the National Academy of Sciences of the United States of America*, **116**, 1816–1818.

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

Pertea, M., **Kim, D.**, **Pertea, G.M.**, **Leek, J.T.** & **Salzberg, S.L.** (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, **11**, 1650–1667.

Ren, G.P., **Jiang, Y.Y.**, **Li, A.**, **Yin, M.**, **Li, M.J.**, **Mu, Y.J.** ***et al.*** (2022) The genome sequence provides insights into salt tolerance ofAchnatherum splendens(Gramineae), a constructive species of alkaline grassland. *Plant Biotechnology Journal*, **20**, 116–128.

Reynolds, J.F., **Smith, D.M.S.**, **Lambin, E.F.**, **Turner, B.L.**, **Mortimore, M.**, **Batterbury, S.P.J.** ***et al.*** (2007) Global desertification, building a science for dryland development. *Science*, **316**, 847–851.

Robertson, D. (2004) VIGS vectors for gene silencing: many targets, many tools. *Annual Review of Plant Biology*, **55**, 495–519.

Ruiz, M.T., **Voinnet, O.** & **Baulcombe, D.C.** (1998) Initiation and maintenance of virus-induced gene silencing. *The Plant Cell*, **10**, 937–946.

Schmutz, J., **Cannon, S.B.**, **Schlueter, J.**, **Ma, J.X.**, **Mitros, T.**, **Nelson, W.** ***et al.*** (2010) Genome sequence of the palaeopolyploid soybean. *Nature*, **463**, 178–183.

Silva, A.T., **Gao, B.**, **Fisher, K.M.**, **Mishler, B.D.**, **Ekwealor, J.T.B.**, **Stark, L.R.** ***et al.*** (2021) To dry perchance to live, insights from the genome of the desiccation-tolerant biocrust moss *Syntrichia caninervis*. *The Plant Journal*, **108**, 1539–1540.

Stanke, M., **Keller, O.**, **Gunduz, I.**, **Hayes, A.**, **Waack, S.** & **Morgenstern, B.** (2006) AUGUSTUS, *ab initio* prediction of alternative transcripts. *Nucleic Acids Research*, **34**, W435–W439.

Sun, P., **Jiao, B.**, **Yang, Y.**, **Shan, L.**, **Li, T.**, **Li, X.** ***et al.*** (2022) WGDI: a user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Molecular Plant*, **15**, 1841–1851.

Suyama, M., **Torrents, D.** & **Bork, P.** (2006) PAL2NAL, robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–W612.

Swindell, W.R., **Huebner, M.** & **Weber, A.** (2007) Transcriptional profiling of *Arabidopsis* heat shock proteins and transcription factors reveal extensive overlap between heat and non-heat stress response pathways. *BMC Genomics*, **8**, 125.

Tuteja, N., **Gill, S.S.**, **Tiburcio, A.F.** & **Tuteja, R.** (2012) Generation and scavenging of reactive oxygen species in plants under stress. In: Tuteja, N., Gill, S.S., Tiburcio, A.F. & Tuteja, R. (Eds.) *Improving Crop Resistance to Abiotic Stress*. Weinheim: Wiley-VCH Verlag, pp. 49–70.

Van de Peer, Y., **Mizrachi, E.** & **Marchal, K.** (2017) The evolutionary significance of polyploidy. *Nature Reviews Genetics*, **18**, 411–424.

Vaser, R., **Sović, I.**, **Nagarajan, N.** & **Šikić, M.** (2017) Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Research*, **27**, 737–746.

Walker, B.J., **Abeel, T.**, **Shea, T.**, **Priest, M.**, **Abouelliel, A.**, **Sakthikumar, S.** ***et al.*** (2014) Pilon, an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, **19**, e112963.

Walker, D.W. & **Van Loon, A.F.** (2023) Droughts are coming on faster. *Science*, **380**, 130–132.

Wan, T., **Liu, Z.M.**, **Leitch, I.J.**, **Xin, H.P.**, **Maggs-Kölling, G.**, **Gong, Y.B.** ***et al.*** (2021) The *Welwitschia* genome reveals a unique biology underpinning extreme longevity in deserts. *Nature Communications*, **12**, 4247.

Wan, Y.Q., **Mao, M.Z.**, **Wan, D.L.**, **Yang, Q.**, **Yang, F.Y.**, **Mandlaa, L.** ***et al.*** (2018) Identification of the WRKY gene family and functional analysis of two genes in *Caragana intermedia*. *BMC Plant Biology*, **18**, 31.

Wang, J.P., **Sun, P.C.**, **Li, Y.X.**, **Liu, Y.Z.**, **Yang, N.S.**, **Yu, J.G.** ***et al.*** (2018) An overlooked Paleotetraploidization in Cucurbitaceae. *Molecular Biology and Evolution*, **35**, 16–26.

Wang, T., **Iyer, L.M.**, **Pancholy, R.**, **Shi, X.** & **Hall, T.C.** (2005) Assessment of penetrance and expressivity of RNAi-mediated silencing of the *Arabidopsis* phytoene desaturase gene. *New Phytologist*, **167**, 751–760.

Wang, Y., **Tang, H.**, **Debarry, J.D.**, **Tan, X.**, **Li, J.**, **Wang, X.** ***et al.*** (2012) MCScanX, a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, **40**, e49.

Waters, E.R. & **Vierling, E.** (2020) Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist*, **227**, 24–37.

Withfold, W.G. & **Duval, B.D.** (2019) *Ecology of desert systems*, 2nd edition. Oxford: Academic Press. Available from: https://doi.org/10.1016/C2017-0-02227-9

Wu, S.D., **Han, B.C.** & **Jiao, Y.N.** (2020) Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Molecular Plant*, **13**, 59–71.

Xiao, L.H., **Yang, G.**, **Zhang, L.C.**, **Yang, X.H.**, **Zhao, S.**, **Ji, Z.Z.** ***et al.*** (2015) The resurrection genome of *Boea hygrometrica*, a blueprint for survival of dehydration. *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 5833–5837.

Xu, Z.C., **Xin, T.Y.**, **Bartels, D.**, **Li, Y.**, **Gu, W.**, **Yao, H.** ***et al.*** (2018) Genome analysis of the ancient tracheophyte *Selaginella tamariscina* reveals evolutionary features relevant to the acquisition of desiccation tolerance. *Molecular Plant*, **11**, 983–994.

Yang, Q., **Yin, J.J.**, **Li, G.**, **Qi, L.W.**, **Yang, F.Y.**, **Wang, R.G.** ***et al.*** (2014) Reference gene selection for qRT-PCR in *Caragana korshinskii* Kom. under different stress conditions. *Molecular Biology Reports*, **41**, 2325–2334.

Yang, Z.H. (2007) PAML 4, phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, **24**, 1586–1591.

Yao, G.Q., **Nie, Z.F.**, **Turner, N.C.**, **Li, F.M.**, **Gao, T.P.**, **Fang, X.W.** ***et al.*** (2021) Combined high leaf hydraulic safety and efficiency provides drought tolerance in *Caragana* species adapted to low mean annual precipitation. *New Phytologist*, **229**, 230–244.

Yoshida, T., **Mogami, J.** & **Yamaguchi-Shinozaki, K.** (2014) ABA-dependent and ABA-independent signaling in response to osmotic stress in plants. *Current Opinion in Plant Biology*, **21**, 133–139.

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

Zhao, Y.Y., **Zhang, R.**, **Jiang, K.W.**, **Qi, J.**, **Hu, Y.**, **Guo, J.** ***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.

Zhou, Q., **Yang, Y.** & **Zhang, M.** (2002) Karyotypes of fourteen species in *Caragana*. *Bulletin of Botanical Research*, **22**, 492–496.

Zhu, J.K. (2002) Salt and drought stress signal transduction in plants. *Annual Review of Plant Biology*, **53**, 247–273.

Zwaenepoel, A., **Li, Z.**, **Lohaus, R.** & **Van de Peer, Y.** (2019) Finding evidence for whole genome duplications, a reappraisal. *Molecular Plant*, **12**, 133–136.