Horticulture Research, 2026, **13**: uhaf314
https://doi.org/10.1093/hr/uhaf314
Article

Horticulture Research OXFORD

Evolutionary history and genomic consequences of polyploidization in natural populations of *Orychophragmus taibaiensis*

Qiang Lai[1,‡], Zeng Wang[1,‡], Changfu Jia[1], Xiner Qumu[1], Rui Wang[1], Zhipeng Zhao[1], Yao Liu[1], Yukang Hou[1], Jianquan Liu[1], Pär K. Ingvarsson[2] and Jing Wang[1,*]

[1]Key Laboratory for Bio-Resources and Eco-Environment of Ministry of Education, College of Life Sciences, Sichuan University, Chengdu, China
[2]Linnean Centre for Plant Biology, Department of Plant Biology, Uppsala BioCenter, Swedish University of Agricultural Sciences, Uppsala, Sweden
[*]Corresponding author. E-mail: wangjing2019@scu.edu.cn
[‡]These authors contributed equally.

Abstract

Polyploidization has occurred throughout the tree of life and is particularly common in plants. Despite its ubiquity, our understanding of the short- and long-term effects and consequences of genome doubling in natural populations remains incomplete. In this study, we identified a novel ploidy-variable species system within the ornamental and industrial oilseed genus *Orychophragmus* (Brassicaceae), which comprises six species, including diploid and tetraploid cytotypes of *Orychophragmus taibaiensis*. By integrating population-scale genomic and transcriptomic datasets across the species in this genus, we constructed a robust phylogenetic framework and investigated the divergence and demographic history of *O. taibaiensis* in comparison to its relatives. Specifically, we characterized the geographical distribution patterns of diploids and tetraploids in natural populations of *O. taibaiensis*, confirmed the autopolyploid origin of tetraploids, and inferred their origin time relative to diploid counterparts. Our findings further revealed that, following genome doubling, tetraploids accumulated a higher genetic load of deleterious mutations, likely due to relaxed purifying selection facilitated by allelic redundancy. Additionally, genome doubling was associated with pronounced changes in gene expression patterns, with differentially expressed genes evolving under relaxed selective constraints. These results highlight that the initial masking of deleterious mutations, changes in expression regulation, and divergent efficacy of selection likely all contribute to shaping the establishment and evolutionary potential of polyploids.

Introduction

Polyploidization, resulting from whole-genome duplication (WGD), has long been regarded as a key driver of plant speciation, adaptation to novel and extreme environments, and genetic innovation [1–4]. While allopolyploidy has received considerable attention, autopolyploidy remains markedly less studied, despite its prevalence and suitability for exploring the direct impacts of immediate genome doubling without the complications associated with hybridity [5, 6]. Autopolyploidy has been repeatedly shown to arise and establish from diploid populations, often leading to the coexistence of tetraploid and diploid populations within a single species [7, 8]. However, our understanding of the evolutionary dynamics and genomic consequences of genome doubling in these populations remains limited [9, 10]. For example, it is unclear to what extent the additional chromosomal copies resulting from WGD can mask deleterious mutations, thereby influencing the genetic load [11]. Furthermore, the degree to which genome duplication alters the efficiency of selection processes in tetraploid populations compared to their diploid progenitors remains largely unexplored [12]. Polyploidy is also frequently associated with morphological and ecological changes, but the extent to which autopolyploidy reshapes gene expression patterns and the functional roles of differentially expressed genes remains an open and intriguing question [13].

Species diversification within the Brassicaceae family is intricately linked to repeated cycles of WGDs, with polyploidy being a common feature across many lineages [14, 15]. Notably, the ancestors of the *Orychophragmus* genus underwent a unique WGD event [16], followed by rediploidization and speciation, resulting in a genome size of ∼1.3 Gb—larger than that of other Brassicaceae species [17, 18]. Recently, *Orychophragmus* species have garnered attention as early-flowering ornamental plants and as promising industrial oilseed crops, owing to their high dihydroxy fatty acids content. This unique trait provides superior lubrication properties and ensures broad adaptability to diverse environmental conditions [19–22]. Moreover, their close evolutionary relationship with *Brassica* species—renowned for their ease of

Received: 30 March 2025. **Accepted:** 8 November 2025. **Published:** 18 November 2025. **Corrected and Typeset:** 1 February 2026

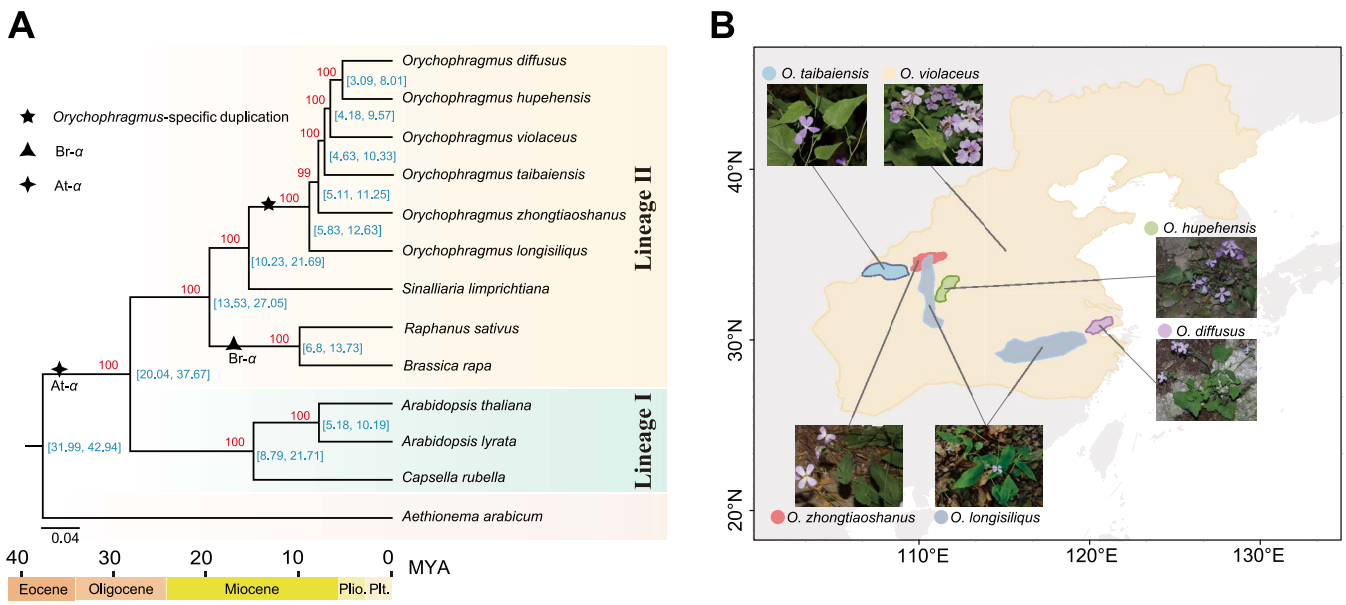

Figure 1 Phylogenetic relationships and geographical distribution of *Orychophragmus*. (A) Phylogenetic inference and divergence time estimation for six *Orychophragmus* species alongside other cruciferous species. Blue numbers (at the nodes) indicate estimated divergence times (95% highest posterior density, in million years ago [MYA]), while red numbers (above the branches) represent bootstrap support values. The black star denotes the *Orychophragmus*-specific WGD event, the black triangle marks the *Brassica*-specific triplication event, and the black tetragon represents the At-α WGD. (B) Geographical distribution and morphological characteristics of the six species within the genus *Orychophragmus*.

hybridization [23, 24]—makes *Orychophragmus* species valuable germplasm resources for *Brassica* genetics and breeding.

Despite their potential importance, the phylogenetic relationships within the *Orychophragmus* genus remain poorly understood. Currently, six species are recognized in this genus: *O. violaceus*, *O. longisiliqus*, *O. zhongtiaoshanus*, *O. taibaiensis*, *O. hupehensis*, and *O. diffusus* [25–27]. These species display significant morphological variation and inhabit diverse environments across East Asia (Fig. 1, Fig. S1). However, previous analyses have relied on a limited number of nuclear and chloroplast genes, resulting in ongoing debates about the genus' evolutionary history and relationships. To address these uncertainties, it is essential to incorporate broader genomic datasets and advanced methodologies to construct a more robust phylogeny, with particular attention to the potential effects of incomplete lineage sorting and introgression.

Notably, the *Orychophragmus* genus includes both widespread and regionally distributed species. *O. violaceus*, the most widely distributed species in the genus, is commonly cultivated as an ornamental plant and is known for its small purple flowers that typically bloom in early spring. In contrast, *O. taibaiensis* is endemic to the mountainous regions of the Taibai Mountains in northwest China (Fig. 2B). The contrasting distribution ranges and ecological divergence between the mountainous endemic species (*O. taibaiensis*) and the widely distributed species (*O. violaceus*) make them ideal candidates for investigating and comparing their evolutionary and demographic histories, particularly in relation to the potential accumulation of genetic load differences between species. Additionally, *O. taibaiensis* has been reported to consist of both diploid ($2n = 2x = 24$) and tetraploid ($2n = 4x = 48$) plants [28], although the mode of origin of the polyploids remains unclear. As such, it also serves as an excellent model for studying the immediate evolutionary and genomic consequences of polyploidization. A direct comparison between tetraploid populations and their diploid progenitors could provide valuable insights into how polyploidization influences both the short-term adaptive responses and long-term evolutionary potential of these populations [29, 30].

In this study, we integrated whole-genome sequencing and transcriptomic datasets to explore the phylogenetic relationships and divergence order of the six species within the genus *Orychophragmus*. We then examined the demographic and divergence history of the widespread species *O. violaceus* and the alpine endemic *O. taibaiensis*, comparing selection efficacy and the deleterious mutation load between the two species. Given that *O. taibaiensis* includes both diploid and tetraploid forms, we characterized the geographical distribution patterns of these cytotypes through extensive field investigations and karyotype analyses, while also investigating the origin of the tetraploid forms. Finally, we explored the genetic and gene expression changes associated with polyploidization in *O. taibaiensis* populations to gain deeper insights into the evolutionary consequences of WGD and its potential short- and long-term selective effects on the ecology and evolution of these populations.

Results

Phylogenetic and evolutionary relationships among species in the genus *Orychophragmus*

To investigate the phylogenetic relationships within the genus *Orychophragmus*, we incorporated seven additional species from the Brassicaceae family and utilized 514 single-copy genes to construct phylogenetic trees using both concatenated

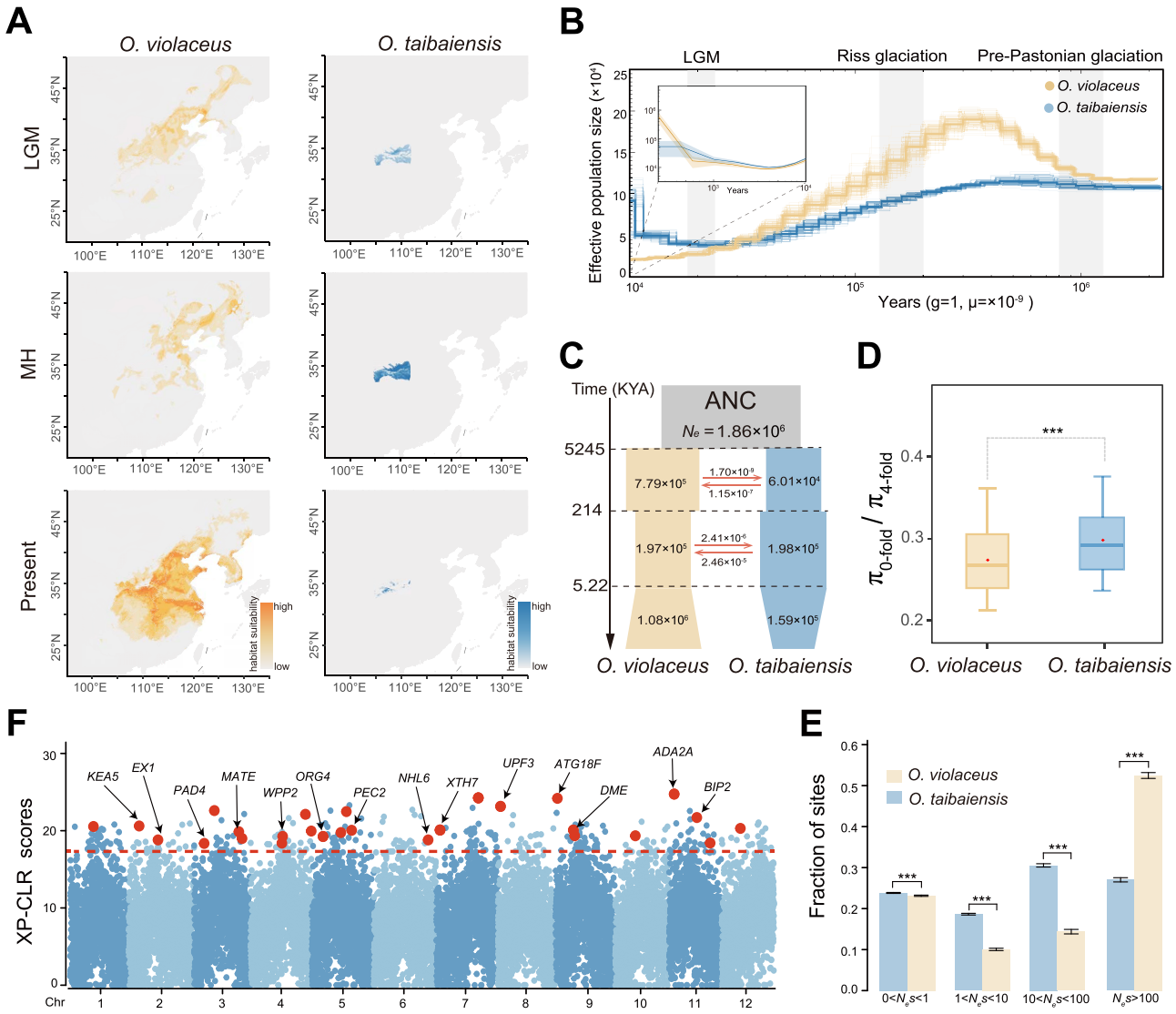

Figure 2 Demographic histories and divergent selection between *O. violaceus* and *O. taibaiensis*. (A) Species distribution modeling under the LGM, MH, and present climate conditions. (B) Demographic history inferred using the PSMC (outer) and MSMC (inner, see Fig. S5 for details) model. Gray vertical bars indicate the LGM, Riss glaciation, and pre-Pastonian glaciation periods. Bold lines represent dynamic changes in effective population size (N_e), while faint lines show 100 bootstrap replicates, ensuring robustness. (C) The best fit demographic model inferred using *fastsimcoal2*. Each block represents a current or ancestral population, with arrows indicating gene flow after divergence (per-generation migration rates). The timing of historical events is shown in KYA. (D) Ratio of nucleotide diversity at 0-fold sites relative to 4-fold sites. (E) DFE in bins of N_es for new 0-fold nonsynonymous mutations for *O. violaceus* and *O. taibaiensis*. Error bars indicate 95% CIs based on 1,000 bootstrap replicates. (F) Selective sweep analysis based on XP-CLR scores along chromosomes. The top 1% of scores, above the red dashed horizontal line (within the coordinate), are considered candidate selective regions. Red circles highlight (large, bold) representative candidate genes located within these regions, with black arrows indicating their names. Asterisks indicate statistical significance from the Wilcoxon test (two-tailed) (NS > 0.05, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

and coalescent approaches. Phylogenetic analysis based on single-copy genes revealed the monophyly of *Orychophragmus*, with its closest relative being the genus *Sinalliaria* (Fig. 1A). According to MCMCTree, the divergence between *Orychophragmus* and its closest relative *Sinalliaria* was estimated to be 15.43 million years ago (MYA, 95% highest posterior density: 10.23–21.69 MYA). Within the clade *Orychophragmus*, the concatenated analysis produced a strongly supported topology of outgroups (*O. longisiliqus*, (*O. zhongtiaoshanus*, (*O. taibaiensis*, (*O. violaceus*, (*O. diffusus*, *O. hupehensis*))))) (Fig. 1A). This topology was also supported by the ASTRAL coalescent approach (Fig. S2A), although some branches exhibited weak support values.

To further explore this, we used DensiTree as a visualization tool to display and quantify the concordance and discordance between individual gene trees and the species tree, which revealed significant inconsistencies between many individual gene trees and the species tree within the *Orychophragmus* clade (Fig. S2B). To further validate the phylogenetic relationship within the genus *Orychophragmus*, we integrated population-level resequencing and transcriptomic datasets, using *Sinalliaria* as the outgroup to construct phylogenies. The phylogenetic relationships within *Orychophragmus*, inferred using both maximum likelihood (ML) and neighbor-joining (NJ) methods (Fig. S3), were consistent with those derived from single-copy gene analyses (Fig. 1A, Fig. S2).

Demographic histories and divergent selection between *O. violaceus* and *O. taibaiensis*

Given the unique features of *O. taibaiensis*—the only species in the genus reported to consist of both diploids and tetraploids as well as being the endemic species found at the highest altitudes within the genus (Fig. S4)—we specifically explored and compared the demographic histories of this species with the closely related widespread species *O. violaceus*. To achieve this, we first applied species distribution models (SDMs) to reconstruct the suitable habitats of *O. violaceus* and *O. taibaiensis* across three evolutionary periods: the present, the Middle Holocene (MH, ∼6 ka), and the Last Glacial Maximum (LGM, ∼21–18 ka) (see Methods for details). Our analysis revealed that the suitable habitat of *O. violaceus* expanded significantly from the LGM and MH to the present. In contrast, the suitable habitat of *O. taibaiensis* remained relatively stable, being primarily confined to the Qinling–Daba mountain regions (Fig. 2A). Furthermore, over time, its range became increasingly restricted to the Taibai Mountain region, from the LGM and MH to the present.

To assess the long-term effective population size (N_e) dynamics of these two species, we applied the pairwise sequential Markovian coalescent (PSMC) method [31]. The results revealed that both species experienced a reduction in N_e from the Riss glaciation to the LGM (Fig. 2B). Consistent with the broader distribution range of *O. violaceus* predicted by the SDMs, the N_e of *O. violaceus* was generally much larger than that of *O. taibaiensis* (Fig. 2B). Since PSMC can only estimate population dynamics up to ∼10,000 years ago, we also employed Multiple Sequentially Markovian Coalescent approach (MSMC2) to focus specifically on the population histories of these two species over the latest 10,000 years. The results indicated that the *O. violaceus* population underwent a more significant recovery following LGM when compared to *O. taibaiensis* (Fig. 2B; Fig. S5).

To further infer the divergence history of the two species, we employed a coalescent simulation-based approach using *fastsimcoal*2 [32, 33]. Twelve models were evaluated (Fig. S6), differing in the presence or absence of postdivergence gene flow and changes in population size following species divergence. The best fitting model (Model 12 in Fig. S6; Table S2) suggested that *O. violaceus* and *O. taibaiensis* diverged ∼5.25 MYA (95% confidence interval (CI) = 1.40–7.33 MYA) (Fig. 2C; Table S3), consistent with the divergence time estimated from phylogenetic analysis (Fig. 1A). Furthermore, the model revealed divergent patterns of changes in effective population sizes and asymmetric gene flow between the species following their divergence, although no gene flow has occurred between the two species within the recent 5.22 thousand years ago (KYA) (95% CI = 5.02–5.79 KYA) (Fig. 2C, Table S2). Additionally, the model inferred that *O. taibaiensis* experienced a recent population reduction ($N_e = 1.59 \times 10^5$, [1.45×10^5–1.99×10^5]), in contrast to *O. violaceus*, which underwent population expansion in both effective population size ($N_e = 1.08 \times 10^6$ [8.45×10^5–1.29×10^6]) and distribution range (Fig. 2A and C).

To further compare the likely influence of different demographic histories on the efficacy of selection and the accumulation of mutational load between the two species, we estimated the ratio of nucleotide diversity at 0- to 4-fold degenerate sites. Consistent with the expectation that species with smaller population sizes have a reduced efficacy of selection to purge deleterious mutations at 0-fold sites [34], we observed that the ratio of 0- to 4-fold nucleotide diversity was significantly elevated in *O. taibaiensis* compared to *O. violaceus* (Fig. 2D). Similarly, the distribution of fitness effects (DFE) analysis revealed that *O. taibaiensis* exhibited weaker purifying selection against strongly deleterious nonsynonymous mutations ($N_eS > 100$) compared to *O. violaceus* (Fig. 2E). Moreover, given that *O. taibaiensis* contains both diploid and tetraploid cytotypes, we reanalyzed and compared the population history divergence, genetic load, and DFE solely between *O. violaceus* and diploid *O. taibaiensis*. These results were highly consistent with the broader analysis, further supporting the conclusion that *O. taibaiensis* exhibits reduced selection efficacy and a higher mutational load compared to *O. violaceus* (Fig. S7B).

Lastly, we applied and calculated XP-CLR statistics to identify potential divergent selection regions between the two species. In total, we detected 2,397 outlier windows in the top 1% of XP-CLR scores (Fig. 2F). Further gene ontology (GO) enrichment analyses of genes within the candidate selective regions revealed significant enrichment in GO terms such as ‘cellular macromolecule metabolic process’ and ‘response to reactive oxygen species’ (Fig. S8A; Table S4), suggesting their association with divergent adaptation to different altitudinal environments between the two species. Many genes known to be involved in stress and defense responses were identified within these regions. For example, the *Arabidopsis* orthologous gene *PAD4*, which plays a critical role in salicylic acid signaling and resistance gene-mediated plant disease resistance [35, 36], was found within these regions.

In addition to the significant signals of XP-CLR, we observed substantially increased interspecies genetic divergence (F_{ST}) and reduced intraspecies Tajima’s *D* values (Figs S8B and S9), providing strong evidence of divergent selection within these genic regions. Similarly, the *Arabidopsis* orthologous gene *PEC2*, which is involved in responding to environmental stimuli such as jasmonic acid, light, and wounding [37], was also identified, highlighting its role in stress responses and adaptive signaling pathways. We found substantially increased F_{ST} and reduced Tajima’s *D* values in both species around these genic regions (Fig. S8B). Similar patterns were observed for other genes related to fundamental processes crucial to plant growth, development, and epigenetic regulation, such as genes *DME* [38] and *XTH7* [39] (Fig. S8D and E). To evaluate the influence of the two cytotypes of *O. taibaiensis* on divergent selection inference, we performed XP-CLR and F_{ST} analyses separately using only diploid *O. taibaiensis* and combined cytotypes. The results showed high correlations (Fig. S10), indicating that ploidy differences likely had minimal impact on the overall findings. However, we acknowledge the limitations of small population sizes in this study. Future studies with broader sampling and sequencing may uncover additional insights into the genomic regions and genes driving speciation and ecological adaptation between the two species.

Cytotype diversity and distribution in natural populations of *O. taibaiensis*

Although *O. taibaiensis* was previously reported to have both diploid and tetraploid cytotypes locally distributed in the Taibai Mountains of central China [40, 41], the spatial overlap and

distribution patterns of populations with different ploidy levels remain unclear. To address this, we determined the ploidy levels of *O. taibaiensis* in a relatively large sample set of 94 individuals from various locations (Table S5). Chromosome number determination confirmed the findings of earlier studies [41] and revealed variation in chromosome size. Specifically, we identified two ploidy levels, with chromosome counts of 2n = 24 for diploids (Fig. 3B) and 2*n* = 48 for tetraploids (Fig. 3C). Cytotype distribution analysis indicated that individuals with different ploidy levels were primarily separated by local mountain barriers within relatively limited geographical ranges (Fig. 3A).

Field observations revealed no noticeable differences in landscape or morphology between the two cytotypes (Fig. S11A and B). However, under controlled conditions, tetraploid plants exhibited significantly larger organ sizes, such as increased leaf dimensions, compared to diploids (Fig. S11C and D). To further investigate whether there was environmental differentiation among cytotypes, we performed a principal component analysis (PCA) using five uncorrelated environmental variables derived from 19 bioclimatic factors (bio1-bio19) and elevation (Fig. S12A Tables S6 and S7). The PCA results demonstrated subtle niche differentiation between cytotypes along environmental and climatic gradients (Fig. S12B). Specifically, tetraploids were associated with habitats characterized by higher maximum temperatures during the warmest month (Bio5), greater isothermality (Bio3), and lower precipitation during the driest quarter (Bio17) compared to diploids (Fig. S10C). These findings suggest that tetraploids are more likely to inhabit slightly drier and warmer environments relative to diploids of *O. taibaiensis*.

Next, we selected five diploid and five tetraploid individuals of *O. taibaiensis* for high-depth whole-genome resequencing. We first employed nQuire [42], a statistical method designed to determine the most plausible ploidy model based on the distribution of base frequencies in the sequencing data. The results from nQuire confirmed the diploid and tetraploid forms of *O. taibaiensis* that we had previously identified (Fig. S13). Specifically, the read depth density distribution of the three alleles (alternative, reference, and both) in diploid individuals was close to 1:1:2 (Fig. S14), while in tetraploid individuals, it was ∼1:3:4 (Fig. S14).

To differentiate between autopolyploidization and allopolyploidization, we utilized Smudgeplots [43] to visualize the expected allele ratio patterns. In tetraploids, the AAAB pattern was more prominent than AABB (Fig. S15), consistent with the expectation of an autopolyploid origin. Furthermore, genome-wide heterozygosity analysis using GenomeScope v.2.0 revealed a distribution of allele ratios, with aaab (3.52%) being more prevalent than aabb (1.96%) (Fig. S16A), further supporting an autotetraploid origin. It is important to note that accurate polyploid inference using *k*-mer-based methods such as Smudgeplots and GenomeScope requires relatively high sequencing coverage. Thus, we restricted these analyses to polyploids with coverage exceeding 60×, all of which supported an autopolyploid origin. Lastly, we recalled SNPs based on the tetraploid model for each of the five resequenced tetraploids and calculated the proportions of various genotypes. The results indicated that the AAAa genotype was significantly more prevalent than AAaa across the entire genome (Fig. 3D; Fig. S16B–F), reinforcing the hypothesis of autotetraploidy in the polyploid *O. taibaiensis*.

Origin history of autotetraploids and the evolutionary consequences of short-term polyploidization in *O. taibaiensis*

To investigate the origin and genetic differentiation between diploid and autotetraploid *O. taibaiensis*, we constructed an unrooted tree using 10 highly deep-sequenced individuals from both ploidy levels. The analysis revealed that the individuals formed two distinct clusters, albeit with short genetic distances and low divergence (Fig. 3E). Using *fastsimcoal2*, we identified the gene flow model as the best fit for the divergence history and estimated that diploids and tetraploids diverged ∼328 KYA (95% CI = 326–468 KYA). The contemporary N_e of diploids was estimated to be 2.53×10^5 (95% CI = 2.40×10^5–3.61×10^5), which is slightly larger than that of autotetraploids, estimated at 1.52×10^5 (95% CI = 1.45×10^5–2.06×10^5) (Fig. 3F, Fig. S18A and B; Table S8).

To further explore the effects of ploidy on the efficacy of purifying selection, we evaluated and compared the DFE between diploids and tetraploids. The results showed that diploids exhibited a significantly higher proportion of loci under strong purifying selection ($N_e s > 100$) compared to autotetraploids. This finding suggests relaxed purifying selection acting on nonsynonymous sites in the tetraploid population compared to the diploid population (Fig. 3G). These results remained robust when we randomly subsampled two out of four alleles per site from the genotypes of tetraploid population to account for potential biases introduced by genotype calling (Fig. S17). However, it is important to note that such biases cannot be fully resolved, as a diploid model was assumed for DFE estimation. Additionally, we calculated the ratio of overall 0- to 4-fold nucleotide diversity (Fig. S18C–I) and found that the ratio was significantly higher in autotetraploids compared to diploids. These results further suggest that the tetraploid population has accumulated a higher proportion of deleterious mutations, which aligns with the findings from *fastsimcoal2* and DFE analysis, showing that the tetraploid population, with higher genetic load, exhibits a smaller effective population size and lower purifying selection efficiency.

Gene expression changes and associated signatures of selection upon polyploidization in *O. taibaiensis*

To explore transcriptional changes and responses to WGD, we examined gene expression differences using RNA-seq data from diploids and autotetraploids of *O. taibaiensis* in both leaf and root tissues (Table S9). PCA clearly distinguished samples based on tissue type and cytotype (Fig. S19A). To specifically investigate gene expression changes associated with polyploidization, we identified differentially expressed genes (DEGs) between diploids and autotetraploids. While the majority of genes showed no significant change in expression (NDEs), we identified 3,254 and 3,422 DEGs in leaf and root tissues, respectively (Fig. S19B), with 1,910 DEGs shared between the two tissues.

GO enrichment analysis of these DEGs revealed that genes differentially expressed between diploids and tetraploids are highly enriched in terms related to signal transduction, cell

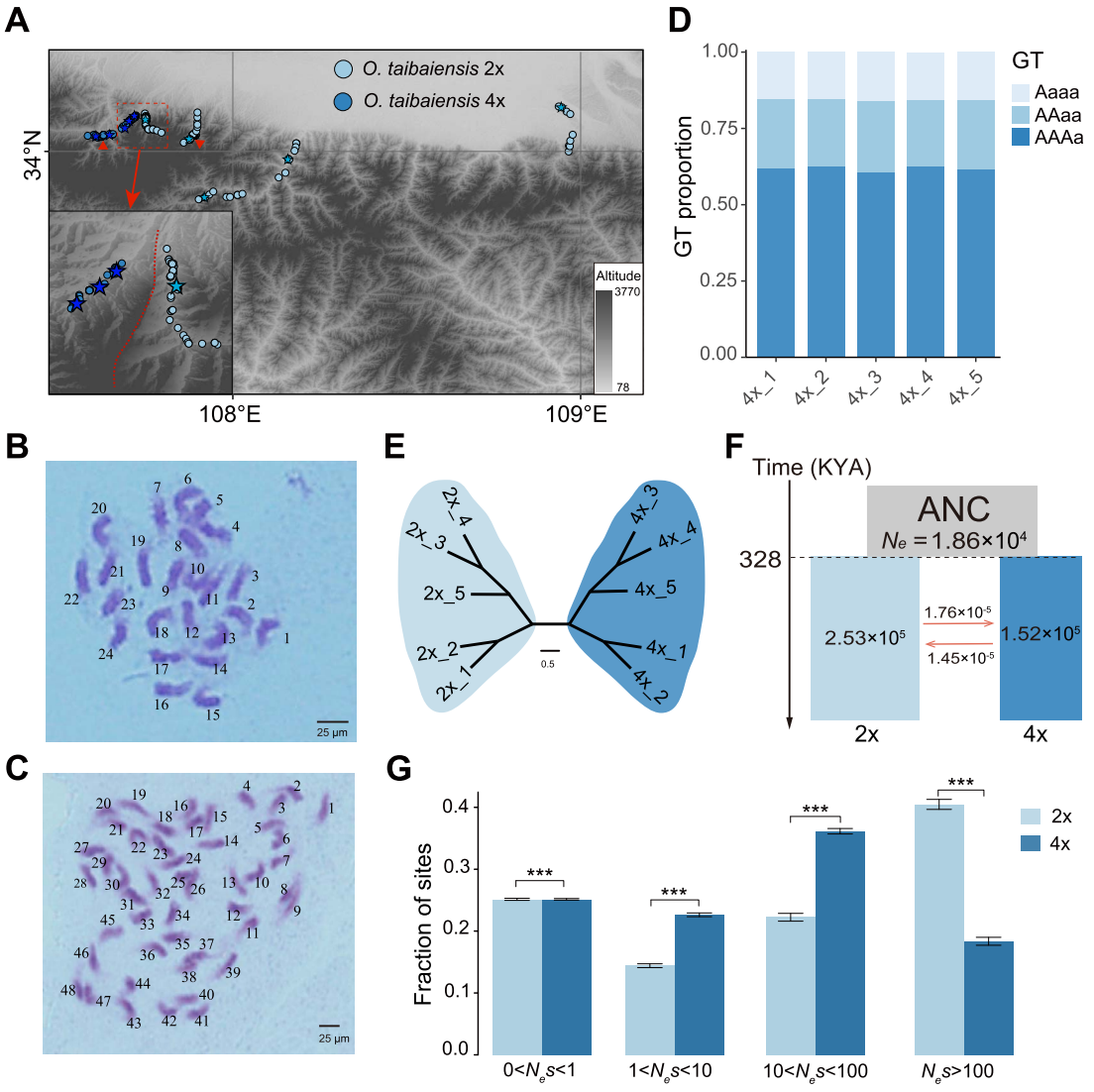

Figure 3 Comparison of the geographic distribution, divergence history, genetic load, and selection efficiency of diploid and tetraploid *O. taibaiensis*. (A) Geographic distribution of diploid and tetraploid *O. taibaiensis*, with stars representing the sequenced individuals mentioned in the text. Red triangles indicate the seed collection sites used for gene expression analysis across ploidy levels. (B) Karyotype of diploid *O. taibaiensis*. (C) Karyotype of tetraploid *O. taibaiensis*. (D) Relative proportions of various genotypes within the five whole-genome resequenced tetraploid *O. taibaiensis* individuals. (E) Phylogenetic relationships among the five resequenced diploid and tetraploid *O. taibaiensis*. (F) Population historical dynamics and divergence time estimation between diploid and tetraploid *O. taibaiensis*, based on *fastsimcoal2*. (G) DFE in bins of N_es for new 0-fold nonsynonymous mutations for diploid and tetraploid *O. taibaiensis*. Error bars represent 95% CIs based on 1,000 bootstrap replicates. Asterisks indicate the level of significance in the Wilcoxon test (two-tailed) (NS > 0.05, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

communication, defense response, regulation of gene expression, and circadian rhythm (Fig. 4A, Table S10). For instance, genes downregulated in tetraploids, such as *SNI1* and *CERK1*, are involved in gene transcription, DNA recombination, and defense signaling (Fig. 4B). Notably, *SNI1* in *Arabidopsis* plays a crucial role in preventing errors during meiotic recombination in plants [44, 45]. Similarly, *α-DOX1* and *ADC* are key players in metabolic processes, with *α-DOX1* participating in fatty acid alpha-oxidation and *ADC* being involved in polyamine biosynthesis, both of which are essential for maintaining cellular integrity and stress responses [46, 47] (Fig. 4B). The significantly reduced expression of these genes in tetraploids may be attributed to genomic instability and the regulatory complexity caused by gene dosage effects in tetraploids. Conversely, in autotetraploid plants, the upregulated genes are primarily involved in functions critical for maintaining genomic stability and responding to environmental stresses. For example, *EXT3* is essential for plant cell wall organization, while *RECA2* is involved in DNA repair [48–50]. Additionally, *LecRK* and *LNK3* have been reported to play roles in cellular responses to salicylic acid, defense against pathogens, and the regulation of circadian rhythms [51, 52]. These findings may reflect a broader theme of adaptation and stress response, highlighting the ability of tetraploids to cope with environmental challenges (Fig. 4A and B).

To further investigate the association between expression divergence and sequence divergence following genome doubling in *O. taibaiensis*, we compared both the relative (F_{ST}) and absolute (d_{xy}) genetic divergence between the two cytotypes using

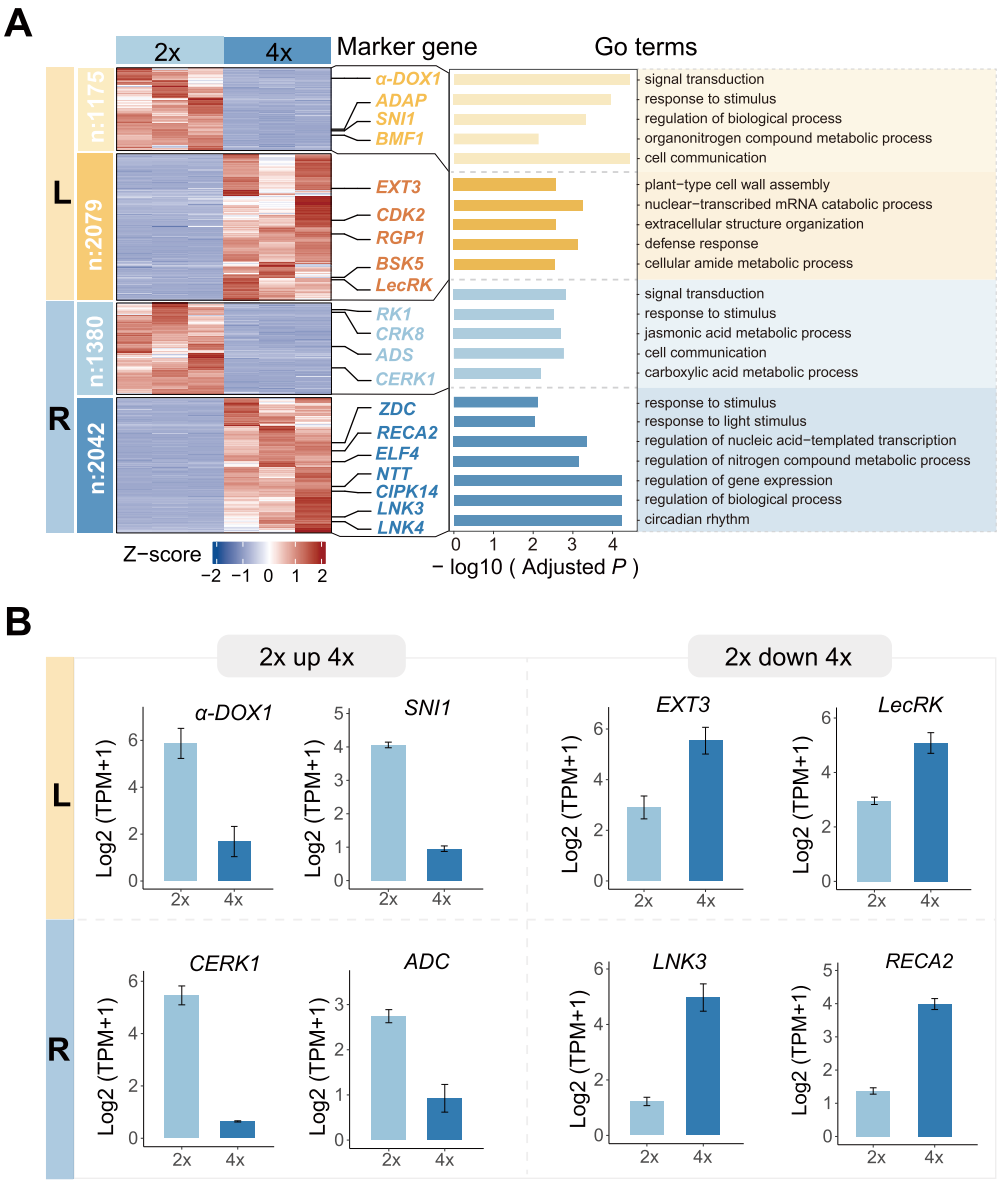

Figure 4 Differential expression analysis between diploid and tetraploid *O. taibaiensis*. (A) Hierarchical clustering of gene expression in diploid and autotetraploid *O. taibaiensis* for leaves (upper, L) and roots (lower, R), highlighting key representative GO terms enriched among DEGs within each cluster. Multiple representative DEGs from each cluster are shown. (B) Examples of expression level comparisons between diploid and tetraploid *O. taibaiensis* for selected representative DEGs identified in (A). Bar heights represent mean expression values from three biological replicates, with error bars indicating ±SD calculated across replicates.

whole-genome resequencing data from five diploids and five tetraploids for DEGs and NDEs. We observed that DEGs exhibited significantly higher d_{xy} compared to NDEs, while no significant differences were observed for F_{ST}. These findings suggest that genes differentially expressed between diploids and tetraploids have accumulated a significantly higher average number of nucleotide differences between ploidies, while nucleotide diversity within each ploidy type remained relatively stable (Fig. 5A and B). Additionally, we compared the ratio of nucleotide diversity at 0- to 4-fold degenerate sites (π_{0fold}/π_{4fold}) between DEGs and NDEs and found a substantially increased π_{0fold}/π_{4fold} ratio within DEGs. This result aligns with expectations of a higher mutation load and reduced purifying selection efficiency for these genes (Fig. 5C). DFE analysis further reinforced these observations, revealing that DEGs showed a significantly lower proportion of novel mutations with likely highly deleterious effects compared to NDEs, implying relaxed purifying selection and reduced functional constraints on these genes (Fig. 5D). Finally, these results were robust to random subsampling of two out of four alleles per site from the genotypes of autotetraploids. In the resampling datasets, F_{ST} values for DEGs were found to be significantly higher than those for NDEs, although the overall trends remained consistent (Figs S20–S22). However, we acknowledge that the sample size of resequenced individuals in this study is relatively small, which limits the robustness of conclusions regarding genetic divergence and selection efficacy. Future studies with larger population-scale datasets will be necessary to provide more robust estimates.

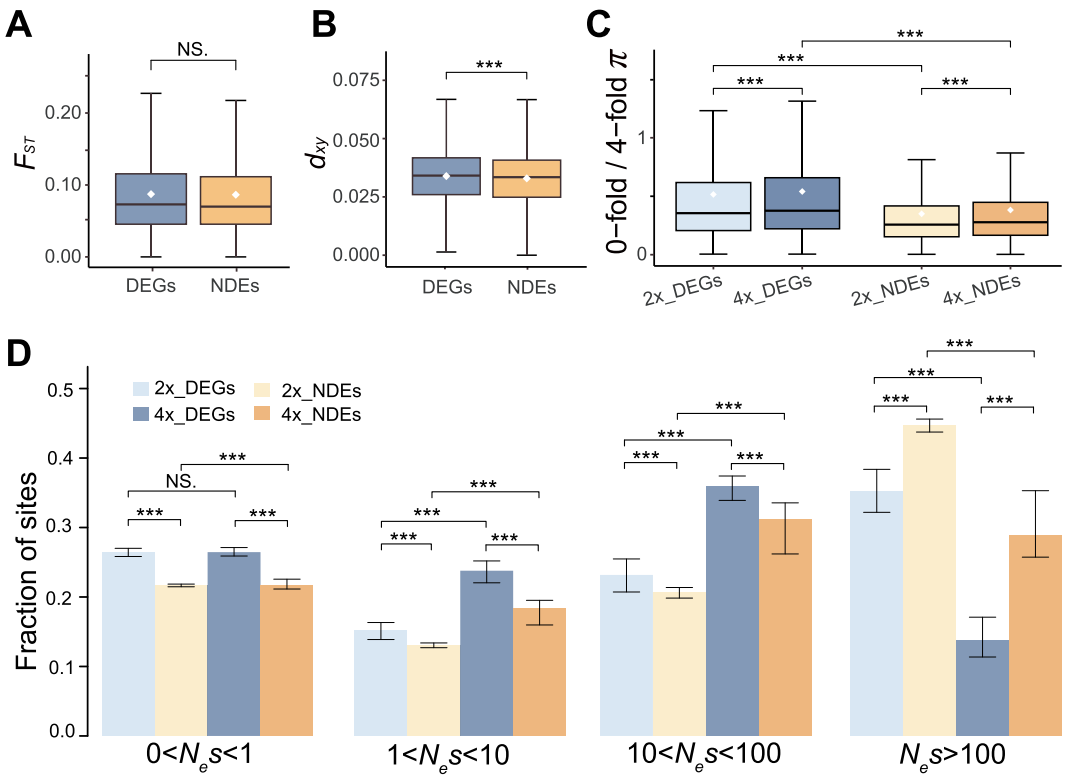

Figure 5 Evolutionary and selection consequences of gene expression changes following autopolyploidization in *O. taibaiensis*. Comparison of genetic divergence of F_{ST} (A), d_{xy} (B) between DEGs (left, in blue) and NDEs (right, in yellow) across samples with different ploidy levels. (C) Comparison of the ratio of 0- to 4-fold genetic diversity, used as a measure of selection efficiency, between DEGs and NDEs in diploid (light color) and tetraploid (dark color) populations in *O. taibaiensis*. (D) DFE for DEGs and NDEs in diploid (light color) and tetraploid (dark color) populations in *O. taibaiensis*. Errors bars represent 95% CI based on 1,000 bootstrap replicates. Asterisks denote significance levels in the Wilcoxon test (two-tailed) (NS > 0.05, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$).

Discussion

Polyploidy resulting from WGD is widespread and has repeatedly occurred throughout the evolution of eukaryotes [1, 53, 54]. However, our understanding of its effects on shaping patterns of genomic variation and influencing the selection process remains limited. In this study, we examined the nonmodel plant genus *Orychophragmus*, which includes both widespread and endemic species, as well as species with variable ploidy levels. By integrating population genomics and transcriptomic datasets, we explored the evolutionary consequences of diverse population demographic histories and the inherent effects of genome doubling on the long-term evolutionary potential of populations.

Using single-copy nuclear genes and population-level SNP datasets, our phylogenomic analyses constructed a more robust species-level phylogeny for this genus compared to previous studies [26, 55]. However, potential caveats must be considered regarding the integration of sequencing data generated from various library preparation protocols, sequencing platforms, and sampling processing procedures across species. Additionally, we observed inconsistencies among individual gene trees, likely driven by introgressive hybridization between species [56, 57], which warrants further investigation in future studies.

Notably, *O. taibaiensis* is a locally endemic mountainous species that occupies the highest elevations within the genus and is the only known species to exhibit a mixed-ploidy system. Despite this, little is known about the spatial distribution of different cytotypes across landscapes or the establishment history of polyploids [6]. To address this, we first examined the divergence history of *O. taibaiensis* and the closely related, predominantly widespread species *O. violaceus* within the genus. Ecological modeling and population genomic analyses revealed that the two species have experienced divergent demographic histories following their divergence. Specifically, *O. taibaiensis* underwent a more drastic population contraction and a mild recent population size recovery, whereas *O. violaceus* maintained a more stable demographic history. These divergent demographic trajectories have resulted in differences in selection efficacy between the two species. The lower effective population size of *O. taibaiensis* has likely led to the accumulation of a greater deleterious mutation load due to reduced selection efficacy in this species [58, 59]. Additionally, we explored signals of divergent selection between the two species and identified genomic regions enriched with genes involved in responses to oxidative stress. These genes are likely crucial for *O. taibaiensis* to survive and thrive in high mountain regions, highlighting the need for further in-depth studies to uncover the adaptive mechanisms of *O. taibaiensis* in such harsh environments.

The diploid–tetraploid variation in *O. taibaiensis* provides a unique study system to investigate the origin of polyploids, the niche similarity and differentiation of ploidies across spatial scales, and the genomic and evolutionary consequences of genome doubling [5, 60]. Through the integration of extensive field surveys, karyotype analyses, and bioinformatics approaches, our

results provide strong evidence supporting the autotetraploid origin model of polyploids. Close phylogenetic relationships among individuals of the two cytotypes were observed, indicating a recent divergence history. Population demographic analyses revealed that the two cytotypes diverged ∼328,000 years ago, coinciding with the late Quaternary glaciation periods, during which the Taibai Mountain was identified as a glaciation center within the Qinling Mountain range [61, 62]. These repeated glacial fluctuations, along with associated climatic and environmental changes, may have facilitated the establishment of autotetraploids, as polyploidy is widely recognized to enhance survival and establishment potential under stressful and dynamic environmental conditions [1, 63].

Notably, odd-ploidy cytotypes (e.g. triploids) were not observed in the field, suggesting that ploidy levels likely act as a strong postzygotic reproductive barrier, promoting the establishment and long-term persistence and coexistence of both cytotypes in nature [8, 64]. Across spatial scales, we clearly delineated the geographic distribution of the two cytotypes, which are primarily separated by a mountain barrier. Although generally subtle environmental differentiation between the cytotypes was observed at local spatial scales, tetraploids were found to inhabit slightly drier and warmer environments compared to diploids, suggesting a potentially higher tolerance to abiotic stress. However, further studies are needed to validate this hypothesis.

To gain deeper insights into how autotetraploids evolve and are affected by selection processes, we examined and compared the distribution of fitness effects of new mutations and the genetic load, estimated as the ratio of 0- to 4-fold diversity. Our findings suggest that polysomic masking in autotetraploids, compared to diploids, likely reduces the efficacy of purifying selection. As a result, this relaxed purifying selection leads to a higher accumulation of deleterious mutations in tetraploid populations [7, 65]. In the short term, the masking of deleterious mutations and the relaxed purifying selection that follow genome doubling may facilitate the rapid establishment and niche expansion of autotetraploid populations [1, 66]. However, it remains uncertain whether the accumulation of deleterious mutations will balance against the beneficial effects of masking, ensuring the stable persistence and coexistence of the two cytotypes in the long run. Alternatively, polyploids may represent an evolutionary dead end, with one cytotype eventually replacing the other over time [29]. Moreover, nuclear volume changes arising from polyploidization can increase genome complexity and influence gene expression [67–69]. By analyzing gene expression patterns in leaf and root tissues of the two cytotypes of *O. taibaiensis*, we identified a set of DEGs between diploids and autotetraploids, while the majority of genes were not differentially expressed (NDEs). Consistent with the expected nuclear volume changes, DEGs were significantly enriched in functions related to extracellular structure organization, cell wall assembly, and cell communication [70, 71]. Additionally, genes involved in signal transduction, response to stimuli, defense responses, regulation of primary and secondary metabolic processes, and circadian rhythm were also enriched among DEGs. These findings provide a likely explanation for the frequent association of polyploidization with enhanced stress resistance and local environmental adaptation following divergence from diploid relatives [5, 72]. Strikingly, we found that genes with expression changes between ploidy cytotypes evolve under relaxed selective constraints and accumulate more sequence divergence compared to NDEs. This highlights a strong association between expression and sequence changes in response to genome doubling for these genes. Future studies are needed to uncover the potential evolutionary forces driving this association and to explore how selection and transcriptional mechanisms jointly respond to genome doubling, shaping the subsequent evolution of autopolyploids [73, 74].

Importantly, there are a few caveats that must be acknowledged. First, biases may arise because many population genetic analyses assume a diploid model of allele frequencies at mutation–selection–drift balance, which could affect and bias the estimates for autotetraploids [7, 68]. In this study, we called variants in both diploid and polyploid modes for the autotetraploids and also performed random subsampling of two alleles per site to facilitate comparisons across various datasets. While we found consistent results, we must acknowledge that we cannot completely rule out the possibility of bias. Second, our transcriptome analyses compared the relative expression levels of genes to the total transcriptome between the two cytotypes. However, we did not estimate the true transcriptome sizes due to the lack of normalization of transcripts for cell number and biomass content changes in autotetraploids [75, 76]. This means we cannot entirely exclude the possibility that an overall change in the total number of transcripts occurred following genome doubling. Nevertheless, given that the majority of genes showed balanced expression levels between cytotypes, we believe the likelihood of substantial changes in cell number and transcriptome sizes in autotetraploids is low. Finally, future studies incorporating gene expression data from additional tissues (e.g. flowers, seeds) could offer more comprehensive insights into how transcriptional mechanisms respond to genome doubling and subsequent evolution in natural populations of *O. taibaiensis*. Additionally, the future availability of genome assemblies for both diploid and tetraploid *O. taibaiensis* could enable further exploration of the relative contributions of ancient WGDs shared by all *Orychophragmus* species and the specific WGD unique to *O. taibaiensis* in driving genome evolution and functional innovation.

Altogether, our findings provide a novel empirical study system to explore the genomic consequences of genome doubling and the potential evolutionary drivers underlying the successful establishment of newly formed autotetraploid lineages in the local mountainous endemic species *O. taibaiensis*. Future work could focus on sampling and sequencing a broader range of diploids and autotetraploids to better understand the factors influencing the evolutionary potential and establishment of polyploids, as well as their long-term coexistence with their diploid ancestors.

Materials and methods

Taxon sampling, genomic sequencing, and genetic data collection

We collected samples from representative natural populations of *Orychophragmus* in China, comprising a total of 17 individuals, including 10 *O. taibaiensis* (five diploids and five tetraploids) from the Taibai Mountains, 1 *O. longisiliqus* and 2 *O. zhongtiaoshanus*

from Shanxi and Shaanxi Province, 1 *O. hupehensis* from Hubei Province, and 3 *O. violaceus* from Henan and Shaanxi Provinces. Sampling details, including coordinates and voucher information, are provided in Table S1. Fresh young leaves were immediately dried in silica gel for DNA extraction. For the newly sequenced dataset, genomic DNA was extracted from leaf samples using the Qiagen DNeasy Plant Kit, and whole-genome paired-end reads (PE150) were generated on the BGI T7 platform.

After integrating newly sequenced genomic datasets from this study with transcriptomic datasets from previous research [55], we collected genomic and/or transcriptomic data for 52 individuals, representing all six *Orychophragmus* taxa: 7 *O. longisiliqus*, 13 *O. zhongtiaoshanus*, 10 *O. taibaiensis*, 1 *O. hupehensis*, 6 *O. diffusus*, and 14 *O. violaceus*. Additionally, we included one species of *Sinalliaria limprichtiana* as the outgroup. We integrated both resequencing and transcriptomic data for phylogenetic analysis of the *Orychophragmus* species. Subsequently, genomic data from resequencing were used to study the population evolutionary history of *O. violaceus* and *O. taibaiensis*.

Phylogenomic construction

To construct the phylogenetic relationships among *Orychophragmus* species and their relatives, we utilized RNA-Seq reads from *O. longisiliqus* (SRR6655848), *O. diffusus* (SRR6655832), *O. violaceus* (SRR6655842), and the closely related lineage *S. limprichtiana* (SRR6441722) for *de novo* transcriptome assembly. The RNA-Seq reads were processed using fastp v.0.20.1 [77] for quality control, followed by *de novo* assembly with Trinity v.2.8.5 [78] under default parameters. The resulting assemblies were refined by retaining only the longest isoform for each gene and removing redundant sequences using CD-HIT v4.8.1 [79]. Completeness was assessed with BUSCO v.5.3.0 [80], and protein-coding regions were identified using TransDecoder v.5.26.3 [81]. An OrthoFinder v.2.5.4 [82] analysis of the four transcriptomes identified 3021 single-copy orthologous genes. Using the highest quality assembly (SRR6655848) as the reference, we employed aTRAM v.2.4.0 [83] to extract and reassemble these genes from three additional species with whole-genome resequencing data (*O. zhongtiaoshanus*: LaiQ184P1, *O. taibaiensis*: LaiQ179P10, *O. hupehensis*: LaiQ197P1). After integrating the assembled sequences with protein data from six additional Brassicaceae species (*Aethionema arabicum*: PRJNA202984, *Capsella rubella*: GCA_000375325, *Arabidopsis lyrata*: GCA_000004255, *Arabidopsis thaliana*: GCA_000001735, *Raphanus sativus*: GCA_010725405, and *Brassica rapa*: GCA_000309985), OrthoFinder identified 514 conserved single-copy genes (>300 bp) suitable for phylogenetic reconstruction across all species.

To construct the phylogenetic relationships, protein sequences were aligned with MAFFT v7.475 [84] and converted to codon-based nucleotide alignments using PAL2NAL [85]. After removing spurious sequences and poorly aligned regions with trimAl v1.4.rev22 [86], the remaining alignments were combined into a supergene matrix for phylogenetic construction using RAxML v8.2.8 [87] under the GTRCAT model. To estimate divergence times between species or clades, we employed MCMCTree from the PAML v4.10.0 package [88, 89], incorporating fossil calibration points obtained from TimeTree [90]. Additionally, we performed coalescent-based species tree inference using ASTRAL v.5.15.5 [91] to account for potential discordance among gene trees, with the results visualized through DensiTree v2.2.7 [92]. to examine topological conflicts between individual gene trees and the consensus species tree.

We further used population genomic data to confirm the interspecific relationship within *Orychophragmus*. Whole-genome resequencing data and transcriptome data were combined to extract variant sites. We processed quality control of raw genomic resequencing and transcriptome sequencing data using fastp v.0.20.1 [77] with stringent quality filtering parameters (-3 20 -5 20 -M 20 -l 36) to ensure data quality. The filtered genomic resequencing reads were then aligned to our newly assembled *O. violaceus* [93] genome using the BWA-MEM algorithm of bwa v.0.7.17 [94] with default settings, whereas the filtered transcriptome sequencing high-quality reads were mapped using HISAT2 v. 2.2.1 [95]. All alignments were processed using SAMtools v.1.9 [96] for sorting and Picard v.2.18.21 for polymerase chain reaction (PCR) duplicate marking (http://broadinstitute.github.io/picard/). Genetic variants (SNP calling) were identified using the Genome Analysis Toolkit (GATK) v.4.2.5.0 [97], including HaplotypeCaller, CombineGVCFs, and GenotypeGVCFs modules with the 'EMIT_ALL_SITES' parameter to retain both variant and invariant sites. After combining the GVCF files from the resequencing and transcriptome datasets, the GenotypeGVCFs tool was applied to genotype the variants, which was followed by applying a strict set of filtering criteria: first, we masked non-uniquely mappable regions using Heng Li's SNPable tool (http://lh3lh3.users.sourceforge.net/snpable.shtml). The reference genome was split into 100-mer sequences and realigned (bwa aln -R 1000000 -O 3 -E 3), retaining only uniquely mapped sites (712,157,434 out of 1,285,736,775 sites). Subsequent filtering removed: (1) multiallelic SNPs (>2 alleles); (2) extreme depth variants (DP <5 or >3× average depth); (3) low-quality SNPs (QD <2.0, FS >60, MQ <20, MQRankSum < −12.5, ReadPosRankSum < −8.0); and (4) sites with >20% missing data. Finally, we identified 4- and 0-fold degenerate sites using the Degeneracy tool (https://github.com/tvkent/Degeneracy) and merged these with transcriptomic data from *S. limprichtiana* (SRR6441722) as outgroup. The final dataset contained 93,969 high-confidence SNPs for population genomic phylogenetic reconstruction.

NJ phylogenetic trees were first constructed using PLINK v.1.90 [98] with the parameter distance 1-ibs to calculate the pairwise identify-by-state (IBS) genetic distance matrix, followed by MEGA X [99] for tree construction. And for ML phylogenetic trees were constructed using IQ-TREE v.2.2.0.3 [100], where ModelFinder [101] was used to select the best fitting substitution model (-B 1000 -m MFP) with 1,000 ultrafast bootstraps.

Demographic history analyses of *O. violaceus* and *O. taibaiensis*

To reconstruct the historical demography of the widespread species *O. violaceus* and the locally endemic species *O. taibaiensis*, SDMs [102] were developed using MAXENT v.3.4.4 [103, 104]. Based on ecological niche theory, SDMs infer a species' potential geographic distribution by combining occurrence records with multiple environmental predictors, thereby estimating habitat suitability under various climatic scenarios. These models were applied to evaluate shifts in suitable habitat ranges across different time periods for both species. To achieve this, current

native distribution records for *O. violaceus* (445 records) and *O. taibaiensis* (91 records, both cytotypes combined) were collected from the Chinese Virtual Herbarium (CVH, https://www.cvh.ac.cn/) and our field investigations.

We utilized 19 bioclimatic variables (BIO1-BIO19) at 2.5-arc-minute resolution from WorldClim (http://www.worldclim.org) [105], examining three temporal scenarios: the present (1970–2000), MH (~6,000 years ago), and LGM (~20,000 years ago). Separate species distribution models were calibrated for each species: one for *O. violaceus* using 445 occurrence records, and one for *O. taibaiensis* using 91 combined records of both cytotypes. Both models were trained under present-day climatic conditions (1970–2000) to characterize the species' current ecological niche. The calibrated models were then projected onto the paleoclimatic reconstructions of the MH and LGM periods to infer potential suitable habitats during these historical epochs. To ensure model robustness, we excluded highly correlated variables (Pearson's $r \geq 0.8$). Model performance was assessed using the area under the receiver operating characteristic (ROC) curve (AUC) metric. [106], which provides reliable evaluation independent of specific thresholds or species prevalence.

To comprehensively reconstruct the demographic histories of these species, we employed a multitiered coalescent-based approach. For long-term effective population size (N_e) dynamics, we implemented the PSMC method [31] with parameters 'N25 -t15 -r5 -p "4+25×2+4+6"', conducting 100 bootstrap replicates to assess estimation robustness. The analysis incorporated a mutation rate of 8.22×10^{-9} per site per generation [107] and an annual generation time assumption for temporal scaling of the results. To gain deeper insights into recent population dynamics, particularly over the past 10,000 years, we employed the MSMC2 [108]. This analysis was performed on phased whole-genome sequences that were generated by Beagle v.4.1 [109] from four individuals (representing eight haplotypes) for each species. For both species, MSMC2 was run on all possible individual configurations, and the medians and SDs of N_e changes were subsequently estimated.

Subsequently, we inferred the divergence history between *O. violaceus* and *O. taibaiensis* using a coalescent simulation-based approach implemented in *fastsimcoal*2 v.27 [32, 33]. The analysis was conducted using 4-fold degenerate sites extracted from whole-genome resequencing data to construct a 2D site frequency spectrum (2D-SFS) through the easySFS pipeline (https://github.com/isaacovercast/easySFS). Twelve distinct demographic models were systematically evaluated, representing various evolutionary scenarios including strict isolation, divergence with or without gene flow, and postdivergence population size changes (Fig. S6). Each model underwent rigorous evaluation through 50 independent optimization runs, with each run performing 100,000 coalescent simulations across 40 iterative cycles to ensure robust parameter estimation. Model selection was based on maximum likelihood optimization, with additional evaluation using Akaike weights and Δlikelihood comparisons. To assess CIs, we performed an extensive bootstrapping procedure consisting of 100 bootstrap replicates, each including 50 independent optimization runs. Throughout the analysis, we maintained consistency with our previous demographic analyses by applying the same mutation rate of 8.22×10^{-9} mutations per site per generation [107] and assuming an annual generation time.

Assessment of selection efficiency and divergent selection signatures between *O. violaceus* and *O. taibaiensis*

Given the differences in demographic history and effective population sizes between *O. violaceus* and *O. taibaiensis*, we compared the efficacy of natural selection and the deleterious genetic load between the two species. First, we calculated and compared the genome-wide ratio of nonsynonymous to synonymous nucleotide diversity ($\pi_{0\text{-fold}}/\pi_{4\text{-fold}}$) using pixy v1.0.4 [110] across 100-kb nonoverlapping window. Next, we estimated the DFE for newly arising nonsynonymous mutations using DFE-alpha v2.16 [111]. This analysis incorporated a demographic model with stepwise population size changes that was first fitted to the neutral SFS. The estimated demographic parameters were then used to infer both the fitness effects of deleterious mutations and the strength of purifying selection ($N_e s$) specific to each species. To evaluate the robustness of our estimates, we performed 1,000 bootstrap replicates by resampling sites within each functional category while excluding the top and bottom 2.5% of values, generating 95% CIs for all parameters. Furthermore, to investigate genomic regions potentially involved in species divergence, we performed genome-wide scans for divergent selection between *O. violaceus* and *O. taibaiensis* using the XP-CLR method [112]. Our analysis calculated composite likelihood ratios (XP-CLR scores) across 5-kb nonoverlapping windows, with the top 1% of windows ($n = 2{,}397$) identified as candidate selection regions. These regions were further validated through complementary analyses of population differentiation (F_{ST}) and Tajima's D statistics to confirm signatures of selection. Finally, we performed GO enrichment analysis on the 340 genes located in these candidate regions using the topGO package (https://bioconductor.org/packages/topGO/) to characterize their functional profiles.

Determination of diploid and tetraploid distribution across *O. taibaiensis* populations

To thoroughly characterize the geographical distribution of diploid and tetraploid cytotypes across the natural populations of *O. taibaiensis*, and based on the descriptions in the original literature of *O. taibaiensis* (Fig. 1B) [40], we conducted extensive sampling across the entire distribution area. Seeds were collected from 94 individuals across seven populations, covering the entire known distribution of *O. taibaiensis* identified in the field (Fig. 3A). Following seed germination on moist filter paper in Petri dishes at 25°C, root meristems were processed through a standardized protocol: (1) pretreatment with 0.1% colchicine solution (25°C, 3 h), (2) fixation in ethanol-acetic acid (3:1) at 4°C for 3 h, (3) hydrolysis in 1 M HCl (37°C, 45 min), and (4) staining with modified carbol-fuchsin solution (room temperature, ≥3 h). Chromosome preparations were made by squashing stained meristems, with counts determined using oil immersion microscopy (1000× magnification) and documented photographically.

To further elucidate and differentiate the origin of polyploidy in *O. taibaiensis*, we conducted a comprehensive genomic analysis to distinguish between autopolyploid and allopolyploid

formation mechanisms. Based on the geographical distribution of diploid and tetraploid individuals, we selected five diploid and five tetraploid individuals (as indicated by the asterisks in Fig. 3A) for high-depth whole-genome resequencing (average depth ~52.48×), representing the key distribution locations of the two cytotypes. Initial ploidy confirmation was achieved through nQuire [42] analysis, which involved noise reduction using the 'denoise' algorithm followed by reference genome alignment to assess read depth distributions and allele frequencies. We also calculated the density distribution of the three allele types (reference, alternative, and both) to validate ploidy levels. Next, we integrated GenomeScope and Smudgeplots [43] to distinguish between autotetraploids and allotetraploids. This analysis involved examining genome-wide heterozygosity patterns in polyploid samples using *k*-mer frequency spectra that were generated from the resequencing data through Jellyfish v2.2.9 [113]. Finally, we reperformed variant calling for the five resequenced individuals of tetraploid *O. taibaiensis* using GATK, setting the parameter '–ploidy 4'. The relative proportions of different genotypes (e.g. Aaaa, AAaa, aaaA) were calculated to further verify whether the tetraploids were of autotetraploid or allotetraploid origin.

Population genomic analysis and assessment of genetic load for diploid and tetraploid individuals of *O. taibaiensis*

To construct the phylogenetic relationships of *O. taibaiensis* at different ploidy levels, we reperformed variant calling and filtering, following the method described in Demographic histories and divergent selection between *O. violaceus* and *O. taibaiensis* section, for the 10 high-depth resequenced individuals. First, we calculated pairwise genetic distances using IBS metrics implemented in PLINK v1.90 [98] with parameter settings for distance calculation (distance 1-ibs). These distance matrices were then used to construct an unrooted NJ phylogenetic tree in MEGA X [99], revealing the genetic clustering patterns among individuals of different ploidy levels. Next, we further utilized *fastsimcoal2* [32] to infer the divergence history of the two cytotypes based on a 2D joint SFS constructed from 4-fold degenerate sites. Whole-genome resequencing data from *O. longisiliqus* and *O. zhongtiaoshanus* were included as outgroups to infer the derived states of alleles using the est-sfs software [114]. The analysis was performed under two models: one accounting for gene flow and one assuming no gene flow (see Fig. S14A and B). The detailed procedures followed those outlined in Cytotype diversity and distribution in natural populations of *O. taibaiensis* section.

We conducted a comparative analysis to evaluate how polyploidization influences the efficiency of purifying selection and the accumulation of deleterious mutations in *O. taibaiensis.* Our approach involved calculating and comparing the ratio of nonsynonymous to synonymous nucleotide diversity ($\pi_{0\text{-fold}}/\pi_{4\text{-fold}}$) and the DFE for newly arising mutations across both ploidy levels. The intensity of purifying selection ($N_e s$) at nonsynonymous sites was estimated using DFE-alpha v2.16 [111] by fitting a stepwise population size change model to account for differences in the SFS of 0-fold nonsynonymous sites and putatively neutral 4-fold synonymous sites. To evaluate the potential effects of the diploid assumption in these analyses, we reperformed specialized variant calling for tetraploid individuals using GATK, setting the parameter '–ploidy 4'. For each tetraploid individual, we randomly subsampled two alleles from the four alleles per site to generate six independent subsampling datasets. These datasets were then used for associated analyses, including *fastsimcoal*2 modeling, genetic load estimation, and DFE analysis. This approach ensured robust comparison between diploid and tetraploid cytotypes while accounting for the polyploid nature of the tetraploid individuals.

Differential gene expression analysis between the two cytotypes of *O. taibaiensis*

To investigate how polyploidy influences transcriptomic changes, seeds of diploid and tetraploid *O. taibaiensis* collected from two geographic regions (Fig. 3A red triangles) were grown under identical conditions in a growth chamber. The environment was controlled with a 16/8-h light/dark photoperiod, a constant temperature of 25°C, and a light intensity of 100 μmol m^{-2} s^{-1}. After 6 months of growth, leaf and root tissues were collected from plants with similar growth characteristics for both cytotypes. Three biological replicates were collected for each ploidy level to ensure statistical reliability. Immediately after collection, plant tissues were flash-frozen in liquid nitrogen and stored at −80°C to preserve RNA integrity. Total RNA extraction was performed according to the TIANGEN RNA Kit protocol, and subsequent library preparation yielded high-quality sequencing libraries that were processed on the DNBSEQ-T7 platform using paired-end sequencing methodology.

The RNA-seq data analysis pipeline began with quality control of raw sequencing reads using fastp v0.20.1 [77] with stringent quality filtering parameters (-3 20 -5 20 -M 20 -l 36) to ensure data quality. Processed reads were then aligned to the *O. violaceus* reference genome [93] using HISAT2 v2.2.1 [95], followed by transcript quantification with StringTie v1.3.6 [115, 116] to obtain transcripts per million (TPM) values. Of the 52,812 annotated genes in the genome, we detected expression (TPM >0) for 37,607 genes across all examined tissues and ploidy levels.

For differential expression analysis, we employed DESeq2 [117] with stringent significance thresholds of |log2 fold change (FC)| ≥2 and adjusted *P*-value < 0.001 to identify significant polyploidy-associated DEGs. Genes showing minimal expression changes with |log2 fold change (FC)| <1 or lacking statistical significance (adjusted *P*-value > 0.05) were classified as NDEs. The DEGs were subsequently analyzed for expression pattern clustering using the ClusterGVis package's mfuzz algorithm (https://github.com/junjunlab/ClusterGVis) and subjected to functional annotation through GO enrichment analysis implemented in the topGO package.

Selection on genes exhibiting differential expression between cytotypes

To examine and compare differences in the strength and direction of natural selection on DEGs and NDEs associated with polyploidization in diploids and tetraploids of *O. taibaiensis*. We

implemented a three-pronged analytical approach to compare selection patterns: firstly, we quantified pairwise relative genetic divergence (F_{ST}) and absolute genetic divergence (d_{xy}) measures between diploid and tetraploid cytotypes for the DEGs and NDE gene sets using PIXY v1.0.4 [110]. Next, we evaluated mutational load by computing the ratio of nonsynonymous to synonymous nucleotide diversity ($\pi_{0\text{-fold}}/\pi_{4\text{-fold}}$) between the two gene sets. Finally, we assessed the strength of selective constraint on the two gene sets by modeling the DFE. As described earlier, the strength of purifying selection (N_es) on 0-fold nonsynonymous sites was estimated for each gene set in both diploids and tetraploids, using 4-fold synonymous sites as a neutral reference, with DFE-alpha v2.16 [111]. For tetraploid individuals, six additional independent datasets, generated by randomly sampling two alleles from the four alleles, were applied in all analyses in this section.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (32000265) and Fundamental Research Funds for the Central Universities (2023SCUNL105) to J.W.

Author contributions

J.W. conceived and supervised the study. Q.L., C.J., R.W., Y.L., and Y.H. handled the sampling, material collection, and performed experiments. Q.L., Z.W., X. Q., and Z.Z. analyzed the data. Q.L. and J.W. wrote the manuscript with the input from P.K.I and J.L. All authors approved the final version of the manuscript.

Data availability

All data needed to evaluate the conclusions in this study are present in the paper and/or the Supplementary information. The newly generated whole-genome resequencing data and transcriptome data of the samples produced in this study have been deposited in the National Genomics Data Center (https://ngdc.cncb.ac.cn) under the accession number PRJCA035915.

Conflicts of interest statement

The authors declare no competing interests.

Supplementary material

Supplementary material is available at *Horticulture Research* online.

References

1. Van de Peer Y, Mizrachi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017;18:411–24
2. Otto SP, Whitton J. Polyploid incidence and evolution. *Annu Rev Genet*. 2000;34:401–37
3. Wang Z, Xue JY, Hu SY. *et al.* The genome of *Hibiscus hamabo* reveals its adaptation to saline and waterlogged habitat. *Hortic Res*. 2022;9:uhac067
4. Wei T, Wang Y, Liu JH. Comparative transcriptome analysis reveals synergistic and disparate defense pathways in the leaves and roots of trifoliate orange (*Poncirus trifoliata*) autotetraploids with enhanced salt tolerance. *Hortic Res*. 2020;7:88
5. Parisod C, Holderegger R, Brochmann C. Evolutionary consequences of autopolyploidy. *New Phytol*. 2010;186:5–17
6. Barker MS, Arrigo N, Baniaga AE. *et al.* On the relative abundance of autopolyploids and allopolyploids. *New Phytol*. 2016;210:391–8
7. Monnahan P, Kolář F, Baduel P. *et al.* Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat Ecol Evol*. 2019;3:457–68
8. Morgan EJ, Čertner M, Lučanová M. *et al.* Niche similarity in diploid-autotetraploid contact zones of *Arabidopsis arenosa* across spatial scales. *Am J Bot*. 2020;107:1375–88
9. Mortier F, Bafort Q, Milosavljevic S. *et al.* Understanding polyploid establishment: temporary persistence or stable coexistence? *Oikos*. 2024;2024:e09929
10. Kolář F, Čertner M, Suda J. *et al.* Mixed-ploidy species: progress and opportunities in polyploid research. *Trends Plant Sci*. 2017;22:1041–55
11. Griswold CK. The effects of migration load, selfing, inbreeding depression, and the genetics of adaptation on autotetraploid versus diploid establishment in peripheral habitats. *Evolution*. 2021;75:39–55
12. Spoelhof JP, Soltis PS, Soltis DE. Pure polyploidy: closing the gaps in autopolyploid research. *J Syst Evol*. 2017;55:340–52
13. Osborn TC, Pires JC, Birchler JA. *et al.* Understanding mechanisms of novel gene expression in polyploids. *Trends Genet*. 2003;19:141–7
14. Marhold K, Lihová J. Polyploidy, hybridization and reticulate evolution: lessons from the Brassicaceae. *Plant Syst Evol*. 2006;259:143–74
15. Franzke A, Lysak MA, Al-Shehbaz IA. *et al.* Cabbage family affairs: the evolutionary history of Brassicaceae. *Trends Plant Sci*. 2011;16:108–16
16. Zhang K, Yang Y, Zhang X. *et al.* The genome of *Orychophragmus violaceus* provides genomic insights into the evolution of Brassicaceae polyploidization and its distinct traits. *Plant Commun*. 2023;4:100431
17. Lysak MA, Koch MA, Beaulieu JM. *et al.* The dynamic ups and downs of genome size evolution in Brassicaceae. *Mol Biol Evol*. 2008;26:85–98
18. Johnston JS, Pepper AE, Hall AE. *et al.* Evolution of genome size in Brassicaceae. *Ann Bot*. 2005;95:229–35
19. Jia C, Lai Q, Zhu Y. *et al.* Intergrative metabolomic and transcriptomic analyses reveal the potential regulatory mechanism of unique dihydroxy fatty acid biosynthesis in the seeds of an industrial oilseed crop *Orychophragmus violaceus*. *BMC Genomics*. 2024;25:29
20. Li X, Teitgen AM, Shirani A. *et al.* Discontinuous fatty acid elongation yields hydroxylated seed oil with improved function. *Nat Plants*. 2018;4:711–20
21. Huang F, Chen P, Tang X. *et al.* Genome assembly of the Brassicaceae diploid *Orychophragmus violaceus* reveals complex whole-genome duplication and evolution of dihydroxy fatty acid metabolism. *Plant Commun*. 2023;4:100432
22. Wang Z, Zhai L, Xiong S. *et al.* February orchid cover crop improves sustainability of cotton production systems in the Yellow River basin. *Agron Sustain Dev*. 2021;41:67

23. Li Z, Ge X. Unique chromosome behavior and genetic control in *Brassica*×*Orychophragmus* wide hybrids: a review. *Plant Cell Rep*. 2007;26:701–10
24. Li Z, Heneen WK. Production and cytogenetics of intergeneric hybrids between the three cultivated *Brassica* diploids and *Orychophragmus violaceus*. *Theor Appl Genet*. 1999;99:694–704
25. Hu H, Zeng T, Wang Z. *et al.* Species delimitation in the *Orychophragmus violaceus* species complex (Brassicaceae) based on morphological distinction and reproductive isolation. *Bot J Linn Soc*. 2018;188:257–68
26. Hu H, Hu Q, Al-Shehbaz IA. *et al.* Species delimitation and interspecific relationships of the genus *Orychophragmus* (Brassicaceae) inferred from whole chloroplast genomes. *Front Plant Sci*. 2016;7:7
27. Hu H, Al-Shehbaz IA, Sun Y. *et al.* Species delimitation in *Orychophragmus* (Brassicaceae) based on chloroplast and nuclear DNA barcodes. *Taxon*. 2015;64:714–26
28. Zhou L, Yu Y, Song R. *et al.* Phylogenetic relationships within the *Orychophragmus violaceus* complex (Brassicaceae) endemic to China. *Acta Bot Yunnanica*. 2009;31:127–37
29. Comai L. The advantages and disadvantages of being polyploid. *Nat Rev Genet*. 2005;6:836–46
30. Mayrose I, Zhan SH, Rothfels CJ. *et al.* Recently formed polyploid plants diversify at lower rates. *Science*. 2011;333:1257–7
31. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475:493–6
32. Excoffier L, Marchi N, Marques DA. *et al. fastsimcoal*2: demographic inference under complex evolutionary scenarios. *Bioinformatics*. 2021;37:4882–5
33. Excoffier L, Dupanloup I, Huerta-Sánchez E. *et al.* Robust demographic inference from genomic and SNP data. *PLoS Genet*. 2013;9:e1003905
34. Lynch M, Conery J, Burger R. Mutation accumulation and the extinction of small populations. *Am Nat*. 1995;146:489–518
35. Yu H, Xu W, Chen S. *et al.* Activation of a helper NLR by plant and bacterial TIR immune signaling. *Science*. 2024;386:1413–20
36. Zeng Y, Zheng Z, Hessler G. *et al. Arabidopsis* PHYTOALEXIN DEFICIENT 4 promotes the maturation and nuclear accumulation of immune-related cysteine protease RD19. *J Exp Bot*. 2023;75:1530–46
37. Völkner C, Holzner LJ, Day PM. *et al.* Two plastid POLLUX ion channel-like proteins are required for stress-triggered stromal Ca^{2+} release. *Plant Physiol*. 2021;187:2110–25
38. Khouider S, Borges F, LeBlanc C. *et al.* Male fertility in *Arabidopsis* requires active DNA demethylation of genes that control pollen tube function. *Nat Commun*. 2021;12:410
39. Volyanskaya AR, Antropova EA, Zubairova US. *et al.* Reconstruction and analysis of the gene regulatory network for cell wall function in *Arabidopsis thaliana* L. leaves in response to water deficit. *Vavilov J Genet Breed*. 2023;27:1031–41
40. Tan Z, Xu J, Zhao B. *et al.* New taxa of *Orychophragmus* (Cruciferae) from China. *Acta Phytotax Sin*. 1998;36:544–8
41. Zhou L, Völkner C, Wu J. *et al.* Karyotype variation and evolution in populations of the Chinese endemic *Orychophragmus violaceus* complex (Brassicaceae). *Nord J Bot*. 2008;26:375–83
42. Weiß CL, Pais M, Cano LM. *et al.* nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*. 2018;19:122
43. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432
44. Zhu L, Fernández-Jiménez N, Szymanska-Lejman M. *et al.* Natural variation identifies SNI1, the SMC5/6 component, as a modifier of meiotic crossover in *Arabidopsis*. *Proc Natl Acad Sci USA*. 2021;118:e2021970118
45. Liu J, Liu B, Chen S. *et al.* A tyrosine phosphorylation cycle regulates fungal activation of a plant receptor Ser/Thr kinase. *Cell Host Microbe*. 2018;23:241–253.e6
46. Vicente J, Cascón T, Vicedo B. *et al.* Role of 9-lipoxygenase and α-dioxygenase oxylipin pathways as modulators of local and systemic defense. *Mol Plant*. 2012;5:914–28
47. Rossi FR, Marina M, Pieckenstain FL. Role of arginine decarboxylase (ADC) in *Arabidopsis thaliana* defence against the pathogenic bacterium *Pseudomonas viridiflava*. *Plant Biol*. 2015;17:831–9
48. Cannon MC, Terneus K, Hall Q. *et al.* Self-assembly of the plant cell wall requires an extensin scaffold. *Proc Natl Acad Sci USA*. 2008;105:2226–31
49. Saha P, Ray T, Tang Y. *et al.* Self-rescue of an EXTENSIN mutant reveals alternative gene expression programs and candidate proteins for new cell wall assembly in *Arabidopsis*. *Plant J*. 2013;75:104–16
50. Odahara M, Kishita Y, Sekine Y. MSH1 maintains organelle genome stability and genetically interacts with *RECA* and *RECG* in the moss *Physcomitrella patens*. *Plant J*. 2017;91:455–65
51. Bouwmeester K, Govers F. *Arabidopsis* L-type lectin receptor kinases: phylogeny, classification, and expression profiles. *J Exp Bot*. 2009;60:4383–96
52. Kidokoro S, Konoura I, Soma F. *et al.* Clock-regulated coactivators selectively control gene expression in response to different temperature stress conditions in *Arabidopsis*. *Proc Natl Acad Sci USA*. 2023;120:e2216183120
53. Wood TE, Takebayashi N, Barker MS. *et al.* The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci USA*. 2009;106:13875–9
54. Masterson J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science*. 1994;264:421–4
55. Zhong L, Liu H, Ru D. *et al.* Population genomic evidence for radiative divergence of four *Orychophragmus* (Brassicaceae) species in eastern Asia. *Bot J Linn Soc*. 2019;191:18–29
56. Morales-Cruz A, Aguirre-Liguori JA, Zhou Y. *et al.* Introgression among North American wild grapes (*Vitis*) fuels biotic and abiotic adaptation. *Genome Biol*. 2021;22:254
57. Morales-Briones DF, Kadereit G, Tefarikis DT. *et al.* Disentangling sources of gene tree discordance in phylogenomic data sets: testing ancient hybridizations in Amaranthaceae s.l. *Syst Biol*. 2020;70:219–35
58. Harris K, Nielsen R. The genetic cost of Neanderthal introgression. *Genetics*. 2016;203:881–91
59. Liu S, Zhang L, Sang Y. *et al.* Demographic history and natural selection shape patterns of deleterious mutation load and barriers to introgression across *Populus* genome. *Mol Biol Evol*. 2022;39:msac008

60. Padilla-García N, Šrámková G, Záveská E. *et al.* The importance of considering the evolutionary history of polyploids when assessing climatic niche evolution. *J Biogeogr*. 2023;50:86–100
61. Zhang W, Liu L, Chen Y. *et al.* Late glacial ^{10}Be ages for glacial landforms in the upper region of the Taibai glaciation in the Qinling Mountain range, China. *J Asian Earth Sci*. 2016;115:383–92
62. Tilman RK. Pleistocene paleoenvironmental changes in the high mountain ranges of central China and adjacent regions. *Quat Int*. 2000;65-66:147–60
63. Van de Peer Y, Ashman T-L, Soltis PS. *et al.* Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 2020;33:11–26
64. Lin H, Chen L, Cai C. *et al.* Genomic data provides insights into the evolutionary history and adaptive differentiation of two tetraploid strawberries. *Hortic Res*. 2024;11:uhae194
65. Cheng F, Wu J, Cai X. *et al.* Gene retention, fractionation and subgenome differences in polyploid plants. *Nat Plants*. 2018;4:258–68
66. Baduel P, Quadrana L, Hunter B. *et al.* Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun*. 2019;10:5818
67. Burns R, Mandáková T, Gunis J. *et al.* Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat Ecol Evol*. 2021;5:1367–81
68. Yu R-M, Zhang N, Zhang B-W. *et al.* Genomic insights into biased allele loss and increased gene numbers after genome duplication in autotetraploid *Cyclocarya paliurus*. *BMC Biol*. 2023;21:168
69. Jia K, Liu H, Zhang R. *et al.* Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (Lamiaceae) genome. *Hortic Res*. 2021;8:177
70. Zhu Y, Wang X, He Y. *et al.* Chromosome doubling increases *PECTIN METHYLESTERASE 2* expression, biomass, and osmotic stress tolerance in kiwifruit. *Plant Physiol*. 2024;196:2841–55
71. Westermann J. Two is company, but four is a party-challenges of tetraploidization for cell wall dynamics and efficient tip-growth in pollen. *Plants*. 2021;10:2382
72. McDaniel SF. Local adaptation, recombination, and the fate of neopolyploids. *New Phytol*. 2024;244:32–8
73. Bhaskara GB, Haque T, Bonnette JE. *et al.* Evolutionary analyses of gene expression divergence in *Panicum hallii*: exploring constitutive and plastic responses using reciprocal transplants. *Mol Biol Evol*. 2023;40:msad210
74. Louder M, Justen H, Kimmitt AA. *et al.* Gene regulation and speciation in a migratory divide between songbirds. *Nat Commun*. 2024;15:98
75. Coate JE. Beyond transcript concentrations: quantifying polyploid expression responses per biomass, per genome, and per cell with RNA-Seq. In: Van Peer Y, ed. *Polyploidy: Methods and Protocols*. Springer US: New York, NY, 2023,227–50
76. Srikant T, Gonzalo A, Bomblies K. Chromatin accessibility and gene expression vary between a new and evolved autopolyploid of *Arabidopsis arenosa*. *Mol Biol Evol*. 2024;41:msae213
77. Chen S, Zhou Y, Chen Y. *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34:i884–90
78. Grabherr MG, Haas BJ, Yassour M. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29:644–52
79. Fu L, Niu B, Zhu Z. *et al.* CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*. 2012;28:3150–2
80. Simão FA, Waterhouse RM, Ioannidis P. *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31:3210–2
81. Haas BJ, Papanicolaou A, Yassour M. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc*. 2013;8:1494–512
82. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238
83. Allen JM, Boyd B, Nguyen N-P. *et al.* Phylogenomics from whole genome sequences using aTRAM. *Syst Biol*. 2017;66:syw105–98
84. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772–80
85. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 2006;34:W609–12
86. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. TrimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. 2009;25:1972–3
87. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30:1312–3
88. dos Reis M, Inoue J, Hasegawa M. *et al.* Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B Biol Sci*. 2012;279:3491–500
89. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24:1586–91
90. Kumar S, Stecher G, Suleski M. *et al.* TimeTree: a resource for timelines, timetrees, and divergence times. *Mol Biol Evol*. 2017;34:1812–9
91. Yin J, Zhang C, Mirarab S. ASTRAL-MP: scaling ASTRAL to very large datasets using randomization and parallelization. *Bioinformatics*. 2019;35:3961–9
92. Bouckaert RR. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*. 2010;26:1372–3
93. Jia C, Hou Y, Lai Q. *et al.* A reference genome and its epigenetic landscape of potential *Orychophragmus violaceus*, an industrial crop species. *bioRxiv*. 2023. https://doi.org/10.1101/2023.09.21.558835
94. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*:1303.3997v2. 2013. https://doi.org/10.48550/arXiv.1303.3997
95. Kim D, Paggi JM, Park C. *et al.* Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–15
96. Li H, Handsaker B, Wysoker A. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078–9

97. DePristo MA, Banks E, Poplin R. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43:491–8
98. Purcell S, Neale B, Todd-Brown K. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75
99. Kumar S, Stecher G, Li M. *et al.* MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 2018;35:1547–9
100. Nguyen L-T, Schmidt HA, Haeseler A. *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 2015;32:268–74
101. Kalyaanamoorthy S, Minh BQ, Wong T. *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 2017;14:587–9
102. Elith J, Leathwick JR. Species distribution models: ecological explanation and prediction across space and time. *Annu Rev Ecol Evol Syst*. 2009;40:677–97
103. Merow C, Smith MJ, Silander J. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*. 2013;36:1058–69
104. Phillips SJ, Dudík M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*. 2008;31:161–75
105. Hijmans RJ, Cameron SE, Parra JL. *et al.* Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol*. 2005;25:1965–78
106. Fawcett T. An introduction to ROC analysis. *Pattern Recogn Lett*. 2006;27:861–74
107. Beilstein MA, Nagalingum NS, Clements MD. *et al.* Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA*. 2010;107:18724–8
108. Schiffels S, Durbin R. Inferring human population size and separation history from multiple genome sequences. *Nat Genet*. 2014;46:919–25
109. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009;84:210–23
110. Korunes KL, Samuk K. PIXY: unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Mol Ecol Resour*. 2021;21:1359–68
111. Keightley PD, Eyre-Walker A. Joint inference of the distribution of fitness effects of deleterious mutations and population demography based on nucleotide polymorphism frequencies. *Genetics*. 2007;177:2251–61
112. Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps. *Genome Res*. 2010;20:393–402
113. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics*. 2011;27:764–70
114. Keightley PD, Jackson BC. Inferring the probability of the derived vs. the ancestral allelic state at a polymorphic site. *Genetics*. 2018;209:897–906
115. Pertea M, Pertea GM, Antonescu CM. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–5
116. Kovaka S, Zimin AV, Pertea GM. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*. 2019;20:278
117. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550

Horticulture Research, 2026, 13, 1–16
https://doi.org/10.1093/hr/uhaf314
Article