Genomic underpinnings of convergent adaptation to high altitudes for alpine plants

Xu Zhang[1,2], Tianhui Kuang[3], Wenlin Dong[1,2,8], Zhihao Qian[1,2], Huajie Zhang[1,2], Jacob B. Landis[4,5], Tao Feng[1,2], Lijuan Li[1,2,8], Yanxia Sun[1,2], Jinling Huang[3,6,7], Tao Deng[3*], Hengchang Wang[1,2*] & Hang Sun[3*]

[1]CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Chinese Academy of Sciences, Wuhan Botanical Garden, Wuhan, Hubei 430074, China

[2]Center of Conservation Biology, Core Botanical Gardens, Chinese Academy of Sciences, Wuhan, Hubei 430074, China

[3]Yunnan International Joint Laboratory for Biodiversity of Central Asia, Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan 650201, China

[4]School of Integrative Plant Science, Section of Plant Biology and the L. H. Bailey Hortorium, Cornell University, Ithaca, NY 14850, USA

[5]BTI Computational Biology Center, Boyce Thompson Institute, Ithaca, NY 14853, USA

[6]State Key Laboratory of Crop Stress Adaptation and Improvement, School of Life Sciences, Henan University, Kaifeng, Henan 475004, China

[7]Department of Biology, East Carolina University, Greenville, NC 27858, USA

[8]University of Chinese Academy of Sciences, Beijing 100049, China

[*]Correspondence: dengtao@mail.kib.ac.cn; hcwang@wbgcas.cn; sunhang@mail.kib.ac.cn

Abstract

Evolutionary convergence is one of the most striking examples of adaptation driven by natural selection. However, genomic evidence for convergent adaptation to extreme environments remains scarce. The Himalaya-Hengduan Mountains represent the world's most species-rich temperate alpine biota, providing an ideal "natural laboratory" for studying convergent adaptation to high altitudes. Here, we generate reference genomes for two alpine plants, *Saussurea obvallata* (Asteraceae) and *Rheum alexandrae* (Polygonaceae), with 37,938 and 61,463 annotated protein-coding genes. By integrating an additional five alpine genomes, we investigate genomic signatures of convergent adaptation to the hostile environments of high altitudes. We show that alpine genomes tend to mitigate their genetic load by contracting genes functioning in the immune system to survive such harsh environments with few pathogens present. We detect signatures of convergent positive selection on a set of genes involved in reproduction and development and reveal that molecular convergence has acted on genes involved in self-incompatibility, cell wall modification, DNA repair and stress resistance, which underlie adaptation to extremely cold, high UV radiation and hypoxia environments. Using gene expression profiles, we further demonstrate that genes associated with cuticular wax and flavonoid biosynthetic pathways exhibit higher expression levels in leafy bracts, shedding lights on the genetic mechanisms of the adaptive 'greenhouse' morphology. Our integrative data provide genomic insights into the convergent evolution at higher-taxonomic levels, aiding in deep understanding of genetic adaptation to complex environments.

Key words: alpine plants, convergent adaptation, evolutionary rates, genomics, "greenhouse" morphology, molecular convergence

Introduction

Evolutionary biologists have long aimed to understand the extent that evolutionary trajectories are predictable, i.e., the extent to which convergent adaptation in distinct lineages is driven by conserved molecular changes (Zhang and Kumar 1997; Stern and Orgogozo 2009; Zhen et al. 2012; Storz 2016). Evolutionary convergence in settings where different species repeatedly face common selective pressures offer a powerful opportunity to address this issue (Yeaman et al. 2016; Birkeland et al. 2020; Xu et al. 2020). An ideal system for investigating the genetic underpinnings of convergent evolution is the independent adaptation of divergent lineages to high-altitude environments.

The Himalaya-Hengduan Mountains (HHMs) exhibit extraordinarily high species richness, and are believed to be the center and origin of diversity of many organisms in the Northern Hemisphere (Spicer et al. 2020). In the alpine zones of the HHMs (elevation above 4500 m), usually characterized by freezing temperatures, high UV (ultra violet) radiation, and hypoxia, plants typically possess suites of similar morphological and physiological adaptations to allow them to survive and reproduce in the hostile environments (Tsukaya and Tsuge 2001). In comparison to plants of lowland areas, plants living in the HHMs (and other high-altitude areas) have dwarf stems, smaller leaves and higher densities of branches, and often exhibit a specialized morphology such as leafy bracts, woolly coverings and cushion forms (Nagy and Grabherr 2009; Sun et al. 2014). Genome-wide studies have documented some genomic footprints of high-altitude adaptation by testing for positive selection and mining expended gene families, often involving functional pathways such as DNA repair, abiotic stress response, reproductive processes, as well as secondary metabolite biosynthesis (Zeng et al. 2015; Zhang et al. 2019; Wang et al. 2021). However, the limited availability of reference genomes for alpine plants restricts further understanding in the genomic evolution of high-altitude adaptation. Additionally, the genomic convergence underpinning high-altitude adaptation have not been examined.

To gain genomic insights into convergent adaptation to high-altitude environments, we newly assembled and annotated reference genomes of *Saussurea obvallata* (Asteraceae) and *Rheum alexandrae* (Polygonaceae). These two species are mainly found in mountain slopes and alpine

meadows of the HHMs and are renowned for their 'glasshouse' morphology, i.e., the upper leaves of which have developed into large semi-translucent leafy bracts that cover the inflorescences, which have been shown to have significant ecological benefits to the plant (Song et al. 2015). We integrate an additional five available genomes of alpine plants [*Crucihimalaya himalaica* (Brassicaceae), *Eutrema heterophyllum* (Brassicaceae), *Hordeum vulgare* var. *nudum* (Poaceae), *Prunus mira* (Rosaceae) and *Salix brachista* (Salicaceae)] as well as their lowland relatives for comparative genomic analyses (supplementary table S1). These seven alpine species represent major clades of angiosperms that independently colonized high-altitude environments. Different from previous case studies of plant genomes, we take advantage of a comprehensive genomic data set of alpine plants to characterize genome-wide signatures of convergent evolution. Specifically, we intended to address three main questions: (i) Whether expanded or contracted gene families show a convergent pattern in alpine plants and have effects on the high-altitude adaptation? (ii) Which genes have undergone convergent molecular evolution and are involved in adaptation to the extremely cold, high UV radiation and hypoxia environment? (iii) Lastly, what are the genomic bases underlying the adaptive "greenhouse" morphology? In addressing these questions, this study provides novel insights into the genomic convergence that contribute to the adaptation to extreme environments.

Results and Discussion

Assembly and annotation of two reference genomes of alpine plants

Using a *K*-mer analysis method, we first estimated the genome size of *S. obvallata* and *R. alexandrae* to be ~2,251 Mb and 2,137 Mb, respectively (supplementary figs. S1, S2, supplementary table S2). Then we obtained a chromosome-level genome of *S. obvallata* and a contig-level genome of *R. alexandrae* using Illumina, Oxford Nanopore, and high-throughput chromatin conformation contact (Hi-C) sequencing technologies (supplementary table S3). For the *S. obvallata* genome, a total of ~95 Gb Illumina short reads, ~143 Gb Nanopore long reads and ~206 Gb Hi-C data were obtained. For *R. alexandrae* genome, a total of ~104 Gb Illumina short reads and ~144 Gb Nanopore long reads were generated. Additionally, transcriptomic data of *R. alexandrae* (~25 Gb) and *S. obvallata* (~157 Gb) were obtained for transcript-based gene annotation (supplementary table S3). Tissue-

specific transcriptomic data of *S. obvallata* were also used for further gene expression analysis (supplementary table S4).

A *de novo* assembly pipeline allowed us to achieve initial genome assemblies that captured 2,044 Mb and 2,040 Mb in 145 and 129 contigs for *S. obvallata* and *R. alexandrae* genomes, with contig N50 of 36.96 Mb and 36.32 Mb, respectively (table 1; supplementary table S5). Using a Hi-C assisted assembly pipeline, 1,952 Mb which accounted for 95.5% of the assembled *S. obvallata* genome, was anchored on 16 chromosomes (table 1; supplementary figs. S3, S6), in line with previous cytological evidence (Fujikawa et al. 2004). We further evaluated the completeness of the assembled genomes and found high completeness rates (94.6% of *S. obvallata* and 94.1% of *R. alexandrae*) of both assemblies as evidenced by BUSCO (Benchmarking Universal Single-Copy Orthologs) assessments using the Eukaryota_odb10 database (table 1; supplementary tables S7, S8) (Manni et al. 2021). The long terminal repeat (LTR) assembly index (LAI), which evaluates the contiguity of intergenic and repetitive regions of genome assemblies based on the intactness of LTR retrotransposons (Ou et al. 2018), was 19.68 for *S. obvallata* and was 15.01 for *R. alexandrae*, respectively, showing high continuity of both genomes.

Transposable elements (TEs) and other repeat sequences accounted for 81.88% and 81.65% of the *S. obvallata* and *R. alexandrae* assemblies, respectively (table 1). In *S. obvallata*, LTR retrotransposons (43.95%), followed by DNA TEs (2.03%) and LINEs (0.93%), were most abundant, with LTR-*Gypsy* and LTR-*Copia* retrotransposons accounting for 18.19% and 25.76% of the LTRs, respectively (supplementary table S9). LTR retrotransposons (44.31%), LINES (2.73%), and DNA TEs (4.08%) accounted for most of the *R. alexandrae* repeats, with LTR-*Gypsy* (37.98%) and LTR-*Copia* (6.33%) retrotransposons predominant among the LTRs (supplementary table S10). A combination of transcript-based, *de novo* and homology-based prediction methods yielded 37,938 and 61,463 high-confidence protein-coding gene models (table 1; supplementary figs. S4, S5). By comparing to public protein databases, a total of 36,542 (96.32%) and 47,535 (77.34%) predicted genes of *S. obvallata* and *R. alexandrae* were functionally annotated (supplementary table S11). For the annotated genes, 93.5% and 96.1% of the complete BUSCO genes against the Eukaryota_odb10 database could be identified in *S. obvallata* and *R. alexandrae*, respectively. Overall, our newly

assembled and annotated genomes of *S. obvallata* and *R. alexandrae* were of high quality, providing valuable genomic resources for the understanding of the convergent adaptation of alpine plants to high-altitude environments.

Complex adaptive histories of alpine plants

The high-quality genomes of *S. obvallata* and *R. alexandrae* allowed reconstruction of the phylogenomic relationships and the adaptive histories of alpine lineages. We downloaded annotated protein sequences from genomes of five additional alpine species as well as 13 representative sisters living in low elevations (supplementary table S1). Species were phylogenetically widely dispersed with 17 eudicots and three monocots, placed in seven families of angiosperms. Orthologs were inferred using OrthoFinder (supplementary table S12), resulting in 6,711 gene families present in all 20 species, of which 195 single-copy orthogroups were used for phylogeny inference and divergence time estimation. A robustly supported phylogenetic tree obtained through maximum-likelihood (ML) analysis of the concatenated protein sequences was consistent with the known phylogenetic relationships within angiosperms, in which the alpine taxa included in our study occurred in seven independent lineages placed in six families (fig. 1a). The time tree inferred from MCMCTree showed a wide divergence history between alpine species and their sampled sisters, ranging from 2 million years ago (Mya) to 31.81 Mya (fig. 1a). This indicted multiple times of adaptation during a wide range of geological history for alpine plants to the high-altitude environments of the HHMs.

Convergent changes of gene family number

We determined convergent changes in gene family number when a gene family showed significant expansion or contraction in more than three alpine species. A total of 56 convergently expanded (CoEx) gene families were identified. Biological Process (BP) of Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses of CoEx families found 68 significantly enriched GO terms and seven KEGG pathways (p-adjust < 0.05) (supplementary tables S13, S14). Enriched pathways of CoEx gene families were mainly related to abiotic resistance, such as response to hypoxia, regulation of hormone levels and hormone transports (figs. 1c, 1d; supplementary figs.

S6, S7). Interestingly, we found a greater number of convergently contracted (CoCo) families than CoEx families, with 1,193 gene families convergently contracted, involving 390 significantly enriched GO terms and 27 KEGG pathways (p-adjust < 0.05) (supplementary tables S15, S16). Most of these pathways included genes involved in the response to biotic stresses, such as defense to pathogens and toxicants (supplementary figs. S8, S9).

A unique stress at high altitudes is hypobaric hypoxia (Beall 2014). In plants, an oxygen deficiency dramatically reduces the efficiency of cellular ATP production, which has diverse ramifications for cellular metabolism and developmental processes (Fukao and Bailey-Serres 2004). Various oxygen-sensing mechanisms have been described that are thought to trigger plant response to low-level oxygen and thus adaptation to altitude (Abbas et al. 2022). In our analysis, we found many significantly enriched GO terms of CoEx families were functionally related to the response to oxygen levels, such as response to hypoxia, response to decreased oxygen level and response to hydrogen peroxide (fig. 1c). These pathways included genes encoding the alcohol dehydrogenase (ADH) (Peng et al. 2001), and the HSP20-like chaperones superfamily proteins that are functionally enriched in the GO term, cellular response to hypoxia. In *A. thaliana*, the oxygen-sensing system is mediated by the plant cysteine oxidase (PCO) N-degron pathway substrates group VII ethylene response factors (ERF-VIIs) (Licausi et al. 2011), which are involved in modulating ethylene response activating the expression of *ADH1* (Yang et al. 2011). While we did not discover convergent expansion of ERF-VIIs in alpine plants, we found genes that are significantly enriched GO terms related to response to ethylene, such as *ADH1, ERF1* (*ETHYLENE RESPONSE FACTOR 1*) and *EER1* (*ENHANCED ETHYLENE RESPONSE 1*). Nonetheless, the expansion of genes involved in the response to hypoxia is necessary for the adaptation of alpine plants to low-level oxygen in high-altitude environments.

In addition, multiple CoEx families were found to be significantly enriched in plant hormone pathways. Examples include genes encoding the probable indole-3-pyruvate monooxygenases (YUC) involved in auxin biosynthesis (Cao et al. 2019), and the small auxin up-regulated RNAs (SAUR) and the D6 protein kinase (D6PK) involved in auxin polar transport (van Berkel et al. 2013). Auxin regulates a series of developmental processes such as apical dominance, plant organogenesis,

and reproductive development by affecting cell growth, differentiation, and patterning (Mockaitis and Estelle 2008). Other developmental regulation pathways including leaf senescence, phototropism and plant organ senescence were also detected to be significantly enriched in CoEx families. This result, coupled with the commonly observed morphological divergence between alpine plants and lowland relatives (Sun et al. 2014), indicates that regulation of morphological changes can be a pivotal path for plant adaptation to high-altitude environments.

Many CoCo families in alpine plants were found to be functionally related to response to pathogens or toxicants, involving GO terms of response to toxic substance and oomycetes, xenobiotic transport and detoxification (supplementary tables S15, S16). Related gene families include the cysteine-rich receptor-like kinases (CRKs), a large subfamily of receptor-like protein kinases (RLKs) that play vital roles in defense responses and programmed cell death in plants (Chen et al. 2004), the malectin-like receptor kinases (MLRs) orchestrating the plant immune responses and the accommodation of fungal and bacterial symbionts (Ortiz-Morea et al. 2022), the multidrug and toxic compound extrusion (MATE) proteins involved in xenobiotic detoxification and multidrug resistance (Diener et al. 2001), as well as the ubiquitin-conjugating (E2) enzymes emerging in recent years as an important regulatory factor underlying plant innate immunity (Zhou et al. 2017). Moreover, the results also showed that the genes encoding receptors for tyrosine-sulfated glycopeptide (PSY1Rs) and phytosulfokine (PSKRs), belonging to the leucine-rich repeat receptor kinases (LRR-RKs), have been undergoing contraction in all alpine species. PSKR1 and PSY1R have been shown to involve in plant immune, with antagonistic effects on bacterial and fungal resistances (Mosher et al. 2013).

The largest disease-resistance genes comprise genes encoding nucleotide-binding site and leucine-rich-repeat domain receptors (NBS-LRRs). Three NBS-LRR gene subclasses, TIR-NBS-LRR (TNL), CC-NBS-LRR (CNL), and RPW8-NBS-LRR (RNL), have been characterized based on the N-terminal domains (McHale et al. 2006). We manually annotated NBS-LRR genes in the sampled genomes using the HMMER search with the Pfam database (Wheeler and Eddy 2013). A total of 5,655 NBS-LRR genes, including 4,058 CNLs, 1,024 TNLs and 173 RNLs were identified among all analyzed genomes (supplementary fig. S10). Additionally, we reconciled the NBS-LRR

gene tree to examine the gains and losses of NBS-LRR genes in alpine species. The results showed that, compared to lowland relatives, most alpine species tend to lose more NBS-LRR genes and exhibit reduced copy number, while *E. heterophyllum*, *S. obvallata* and *R. alexandrae* have similar numbers compared to their closest relatives (supplementary fig. S11). A similar phenomenon was also described in the case study of the *C. himalaica* genome, in which the most significantly contracted gene families were functionally enriched in disease and immune responses pathways (Zhang et al. 2019). Due to the harsh environments characterized by freezing temperatures, aridity, and high UV radiation, it is reasonable to hypothesize that gene families involved in pathogen or toxicant defense have undergone contraction in alpine plants, as fewer microorganisms exist. In addition, genes functioning in cellular transport were shown to be contracted in alpine plants. Related pathways include cytoskeleton organization, actin filament-based process, export across plasma membrane and export from the cell (fig. 1d; supplementary table S15). These processes may be a component of plant immune system and possibly have undergone simplification due to the pathogen depauperate environments of high altitudes. These results suggest that contractions immune system genes might be an effective way to mitigate genetic loads while surviving in hostile environments but with few pathogens present.

Tests for convergent positive selection

In harsh environments, positive selection is expected to be common in genes controlling early life history stages, including genes involved in reproduction and development (Cui et al. 2019). Our branch-site tests identified 36 convergently selected genes that show signatures of positive selection in more than three alpine species (supplementary table S17). These genes were functionally related to basal life processes involving reproduction and respiration, such as carpel, gynoecium, ovule and endosperm developments and photorespiration pathways (supplementary fig. S12, supplementary tables S18, S19). Examples included the *SEPALLATA* (*SEP*) MADS-box genes required in floral organ and meristem identity (Pelaz et al. 2000), the *MALE MEIOCYTE DEATH 1* (*MMD1*) gene regulating cell cycle transitions during male meiosis (Yang et al. 2003), the *NIJMEGEN BREAKAGE SYNDROME 1* (*NBS1*) gene involved in double-strand break repair, DNA recombination and

maintenance of telomere integrity in the early stages of meiosis (Zhang et al. 2006), and the *HYDROXYPYRUVATE REDUCTASE* (*HPR*) gene localized in leaf peroxisomes functioning in the glycolate pathway of photorespiration (Mano et al. 1999). Moreover, genes encoding pectin methylesterases (PMEs) were detected to be undergone convergently positive selection (supplementary table S17). PMEs play a central role in the synthesis and metabolism of pectins, which contribute both to the firming and softening of the cell wall that is important for basal developmental processes of higher plants, such as meristematic growth, fruit ripening, programmed cell death, and endosperm rupture upon germination (Louvet et al. 2006; Rodriguez-Gacio Mdel et al. 2012). In addition, the glyoxylate and dicarboxylate metabolism, the most significantly enriched KEGG pathways (supplementary fig. S13, supplementary table S19), is a fundamental biochemical process that ensures a constant supply of energy to living cells. These convergently selected genes detected in our analyses likely contribute to the primary adaptation of alpine plants to similar extreme environments.

Detection of molecular convergence

We investigated signatures of genes undergoing molecular convergence among alpine species using a combination of approaches for the detection of convergent evolutionary rate shifts and site-based estimation of convergent amino acid (AA) evolution. These approaches have been commonly used in previous studies that investigate the genomic signatures of convergent evolution, including convergent adaptation to seasonal habitat desiccation in African killifishes (Cui et al. 2019), convergent regulatory evolution and loss of flight in paleognathous birds (Sackton et al. 2019), and convergent evolution of extreme lifespan in Pacific Ocean rockfishes (Kolora et al. 2021). Convergent shifts in gene evolutionary rates were detected using the RERconverge method (Kowalczyk et al. 2019), which estimates the correlation between relative evolutionary rates (RERs) of protein sequences and the evolution of a convergent binary or continuous trait across a phylogeny. Our analysis focused on positive correlations, representing genes with faster evolutionary rates in alpine species relative to species living in low elevations. An increased RER could arise due to relaxation of a constraint or positive selection, which could be adaptive to habitat-related changes

(Kowalczyk et al. 2020). We identified 69 gene families undergoing convergently accelerated RERs in alpine species, significantly enriched in 93 GO terms and 12 KEGG pathways (p-adjust < 0.05) (supplementary tables S20, S21), involving pathways for self-incompatibility system, cell wall modification, DNA repair and stress resistance (fig. 2). Among them, 26 gene families were found to have convergent AA shifts by a PCOC analysis (supplementary fig. S14). The PCOC method considers shifts in AA preference instead of identical substitutions (Rey et al. 2018). Given the relatively far phylogenetic distances among our analyzed species, selecting only sites that converged to the exact same AA in all species is quite strict and is bound to capture only a subset of the substitutions associated with the convergent trait change.

Self-incompatibility system. Self-incompatibility (SI) in many flowering plants is controlled by the *S* (sterility) locus (Takayama and Isogai 2005). The loss of SI genes in *A. thaliana* is responsible for the evolutionary transition to the self-fertile mating system (Sherman-Broyles et al. 2007). Although self-fertilization is often thought to lead to a decreased fitness of homozygous offspring, this mode ensures reproduction in the absence of pollinators or suitable mates, and therefore can be advantageous for plants to occupy niches in harsh environments (Goodwillie et al. 2005). Our results revealed the biggest orthogroup (OG0000000), which includes the *Arabidopsis* S-receptor kinase (*SRK*) genes and the *S*-locus flanking gene *ARK3* (*RECEPTOR KINASE 3*), undergoing evolutionary rate acceleration and convergent AA shifts in alpine species (fig. 3). GO analysis showed that these genes were functionally enriched in the process of reproduction (recognition of pollen, recognition of pollen, and pollination) and immune system (immune response). Furthermore, we identified the functional domain using NCBI's conserved domain database (CDD) (Lu et al. 2020). The result showed that these proteins contained the *S*-locus glycoprotein domain (Pfam00954), confirming their functions in SI system.

Loss of function at the *S*-locus in alpine *Crucihimalaya* genomes possibly due to relaxed selection were reported (Zhang et al. 2019; Feng et al. 2022), with a similar phenomenon found in the high-altitude Andes maca (*Lepidium meyenii*) genome (Zhang et al. 2016). Our branch-site tests did not detect any signatures of positive selection on these genes, suggesting that acceleration of

297 evolutionary rates may be the result of relaxed constraints. Therefore, we tested for the relaxation of
298 selection on the *S*-locus genes using the RELAX model (Wertheim et al. 2015). RELAX analysis
299 estimated the relaxation intensity parameter k, $k > 1$ indicates intensified selection (i.e., positive or
300 purifying selection) and $k < 1$ suggests relaxed selection. The results showed that five (*C. himalaica*,
301 *E. heterophyllum*, *H. vulgare* var. *nudum, S. brachista* and *S. obvallata*) of the seven alpine plants
302 exhibited significantly relaxed selection on *S*-locus genes (p-value < 0.05; supplementary table S22).
303 Bingham and Ort (1998) reported that low levels of insect diversity, abundance and activity often
304 occur in alpine ecosystems and hypothesized that these factors may limit the pollination of alpine
305 plants. Species living in isolated habitats like alpine environments or ocean islands are thus less
306 likely to be SI, in line with ‘‘Baker’s Law’’, which assumes that pollen limitation may be an
307 important force driving the transition of mating systems (Cheptou 2012). Pollination biology studies
308 have shown several cases of autonomous selfing in various taxonomically distant species within
309 HHM communities, although the proportion of self-pollinated species has not yet been calculated to
310 test the hypothesis (Sun et al. 2014). We here hypothesize that convergent acceleration of
311 evolutionary rates of *S*-locus genes due to relaxed selection can be evidence of the evolutionary
312 transition from self-incompatibility to the self-compatibility mating system, which is potentially a
313 convergently adaptive process for alpine plants to facilitate their reproduction and the occupation of
314 alpine niches.
315
316 ***Cell wall modification.*** The cuticle membrane lies over and merges into the outer wall of epidermal
317 cells (Martin and Juniper 1970). The primary role of the cuticle, composed of cutin and cuticular
318 waxes, is to mitigate water loss and excessive UV radiation by functioning as a physical barrier
319 between the plant surface and its external environment (Kerstiens 1996). Cuticular waxes are
320 composed of a variety of organic solvent-soluble lipids, consisting of very-long-chain (VLC) fatty
321 acids and their derivatives, as well as secondary metabolites like flavonoids (Pollard et al. 2008). The
322 cell wall modification pathway was found significantly enriched in genes undergoing convergent
323 positive selection (supplementary table S18). Additionally, in the examination of convergent changes
324 in gene family number, we found the most significantly enriched KEGG pathway of co-expanded

genes to be cutin, suberine and wax biosynthesis (supplementary table S14), corresponding to the enriched suberin biosynthetic process in the GO analysis (supplementary table S13), including genes encoding fatty acyl-CoA reductases (FARs). FARs catalyze the formation of fatty alcohols, which are common components of plant surface lipids (i.e., cutin, suberin, and associated waxes). We did not detect convergent acceleration of the evolutionary rate of FARs, suggesting possible modifications of FARs through the increase in gene copy number. The results showed that an aldehyde decarbonylase enzyme CER1 underwent a convergent acceleration of evolutionary rate in alpine plants. Overexpression of the *A. thaliana CER1* gene was reported to promote wax VCL alkane biosynthesis and influences plant response to biotic and abiotic stresses (Bourdenx et al. 2011). The convergent evolution of genes involved in the cutin, suberine and wax biosynthesis implied that cuticular waxes may function as protective screens against UV radiation to protect anatomical structures and dissipate excess light energy for alpine plants.

In addition, we detected positive selection and molecular convergence of three MYB transcription factors, including MYB27, MYB48, and MYB59 (supplementary table S17). Among them, MYB27 was reported to play a role in regulating the accumulation of anthocyanins (Albert et al. 2014), a class of flavonoids. In the STRING database (Szklarczyk et al. 2021), MYB48 was predicted to interact with proteins that are involved in flavonoid biosynthesis, including F3H (naringenin,2-oxoglutarate 3-dioxygenase), catalyzing the 3-beta-hydroxylation of 2S-flavanones to 2R,3R-dihydroflavonols which are intermediates in the biosynthesis of flavonols and anthocyanidins, FLS1 (flavonol synthase/flavanone 3-hydroxylase), catalyzing the formation of flavonols from dihydroflavonols, DFR (dihydroflavonol reductase), catalyzing the conversion of dihydroquercetin to leucocyanidin, and TT5, a member of chalcone-flavanone isomerase family protein (supplementary fig. S15, supplementary table S23). Flavonoids function as antioxidants that reduce DNA damage induced by abiotic stresses such as extreme temperatures, UV radiation and drought, and thus play critical roles in species adapting to high-altitude environments (Agati et al. 2012). Many examples of whole genome studies of alpine plants have shown that expansion and/or positive selection of genes involved in flavonoid biosynthesis constitute an important part of the genomic footprint of alpine adaptation (Zeng et al. 2015; Chen et al. 2019; Wang et al. 2021). Taken together, modifications of

353 the cell wall in alpine plants through evolutionary expansion and adaptive convergence of genes
354 involved in the biosynthesis of cuticular waxes and flavonoids might be vital strategies for adaptation
355 to dramatic weather changes and extensive UV radiation in high altitude.

356 With the newly generated transcriptomic data of *S. obvallata*, we were able to investigate
357 expression patterns of genes related to the biosynthesis of cuticular waxes and flavonoids. Five
358 tissues with three biological replicates, including three from leaves (basal leaves JL, middle leaves
359 ML, and bract leaves BL) and two from flowers and stems, were sampled (supplementary fig. S16,
360 supplementary table S3). After mapping the RNA-seq data to the assembled genome of *S. obvallata*,
361 25,096 genes had expression profiles and were retained for differential gene expression (DEG)
362 analysis. The results showed that, compared to basal and middle leaves, bract leaves exhibit 1,071
363 significant up-regulated genes (fig. 4a). These genes were significantly enriched in cytochrome P450,
364 cutin, suberine and wax biosynthesis and isoflavonoid biosynthesis KEGG pathways (fig. 4b).
365 Furthermore, we analyzed expression profiles of genes involved in the biosynthesis of cuticular
366 waxes and flavonoids. The results showed that many genes had higher expression levels in bract
367 leaves than in other leaf tissues. For example, *CER1, CER3, CER4*, and *MAH1* in the cuticular wax
368 biosynthetic pathway (fig. 4c), and *4CL, CHS, CHI, F3'H, TT7* and *OMT* in the flavonoid
369 biosynthetic pathway (fig. 4d). These results suggest that the accumulation of cuticular waxes and
370 flavonoids is an important genetic pathway from normal leaves to leafy bracts. Our findings provide
371 new insights into the genetic basis of the specialized 'glasshouse' morphology for a better
372 understanding of plant morphological adaptation.
373
374 ***DNA repair and stress resistance pathways.*** The hypoxia and intense UV radiation in alpine
375 environments exert highly abiotic stress that can cause DNA, RNA, and protein damage. DNA repair
376 processes thus play an important role in the high-altitude adaptation of plants, similar to evidence in
377 alpine animals (Li et al. 2018). The *NBS1* gene, involved in DNA repair, cellular response to DNA
378 damage stimulus and double-strand break repair, was found to be under convergent selection in
379 alpine genomes (supplementary table S17). Moreover, several significantly enriched GO terms
380 related to DNA repair and protein ubiquitination were detected undergoing convergent evolution in

alpine plants (fig. 2a, supplementary table S17). Similar KEGG pathways were also identified, including mismatch repair, homologous recombination and nucleotide excision repair (fig. 2b, supplementary table S18). Examples included the *UEV1* genes, enriched in protein K63-linked ubiquitination pathway that reportedly play a role in DNA damage responses and error-free post-replicative DNA repair by participating in lysine-63-based polyubiquitination reactions (Wen et al. 2008), and genes encoding OB-fold proteins which were found to be important for genomic stability including DNA replication, recombination, repair, and telomere homeostasis (Flynn and Zou 2010).

In addition to low oxygen levels and excessive UV radiation, high-altitude environments pose various threats to living organisms from an unpredictable climate. We found many genes with convergent evolutionary rate shifts were significantly enriched in stress resistance pathways, such as genes involved in the hormone signal transduction (cell recognition, intracellular signal transduction, and cytokinin-activated signaling pathway), the cell rhythm system (circadian rhythm, cell cycle phase transition, and programmed cell death), and the regulation of enzyme activity pathways (regulation of kinase activity, regulation of GTPase activity, and regulation of transferase activity) (fig. 2a, supplementary tables S17). Also, some KEGG pathways were found to be significantly enriched in metabolic pathways that may contribute to stress resistance, such as plant hormone signal transduction, phenylpropanoid biosynthesis, and glycine, serine, and threonine metabolism (fig. 2b, supplementary table S18). Plant hormones like cytokinin and ethylene play pivotal regulatory roles in plant growth and development, including cell division, shoot initiation, light responses, and leaf senescence. We found that Type-A *Arabidopsis* response regulator (ARR) protein family, involved in response to cytokinin and cytokinin signal transduction, undergoes molecular convergence in alpine plants. An experimental study revealed that *arr* mutants show altered red-light sensitivity, indicating an important role of type-A ARRs in light adaptation (To et al. 2004). The circadian clock was selected as a mechanism in the control of cell-cycle progression to avoid sunlight-induced DNA damage in ancient unicellular organisms (Hut and Beersma 2011). Circadian rhythm in plants regulates multiple processes such as photosynthesis, flowering, seed germination, and senescence (Srivastava et al. 2019). Genes involved in the regulation of the rhythm system were found to be convergently accelerated evolutionary rates in alpine plants, such as the transcription factor *MYB59*,

409 which participates in the regulation of the cell cycle, mitosis and root growth by controlling the
410 duration of metaphase. The *myb59* mutant was found to have longer roots, smaller leaves and smaller
411 cells than wild-type plants (Fasani et al. 2019), which are commonly observed in alpine plants. Taken
412 together, these results suggested that adaptation to high altitude requires the participation of multiple
413 biological processes.
414

415 ## Conclusions

416 Taking advantage of an integrative genomic data set of alpine plants, our study unraveled the
417 genomic underpinnings of convergent adaptation to high-altitude environments (fig. 5). We identified
418 convergently expanded gene families involved in the oxygen-level response and contracted gene
419 families related to disease resistance in alpine plants, potentially reflecting adaptations to high-
420 altitude environments experiencing hypoxia and depauperate pathogen communities. Our results
421 showed that alpine plants have undergone convergent molecular evolution of genes involved in
422 reproduction and development, self-incompatibility, cell wall modification, as well as DNA repair
423 and stress resistance pathways that may contribute to the survival in extremely cold, high UV
424 radiation and hypoxia environments. Both molecular convergence in changes of gene copy number
425 and accelerated evolutionary rates are consequences of selective pressures posed by the surrounding
426 environment. Thus, genomic signatures of convergent evolution detected here are direct evidence for
427 alpine plants associated with independently colonizing, evolving and adapting to cold, high UV
428 radiation and hypoxia environments of the HHMs. The alpine plants included in this study belong to
429 taxonomically distant plant orders, hence, standing genetic variation and localized introgression of
430 regions of the genome can be ruled out as probable causes of genomic convergence. Identical *de*
431 *novo* mutations must therefore have occurred independently in each taxon during evolution from a
432 low-altitude ancestor. Furthermore, we gain valuable insights into genetic bases underlying the
433 adaptive 'greenhouse' morphology by mining differentially expressed genes involved in the
434 biosynthesis of cuticular waxes and the accumulation of flavonoids that might be vital strategies for
435 adaptation to dramatic weather changes and extensive UV radiation in high altitudes. Taken together,
436 our study reveals genomic convergence underlying adaptation to alpine environments at a high-

taxonomic level of angiosperms, providing a framework for further understanding of high-altitude adaptation. Further genomic study of alpine adaptation should provide detailed evidence related to morphological and physiological specializations using updated data, such as proteome and metabolome, while also referring to discoveries from phylogenetically distant taxon to evaluate the evolutionary convergence in similar environments.

Materials and Methods

Plant material, DNA extraction and sequencing

Fresh leaves of wild *S. obvallata* and *R. alexandrae* individuals were collected from Songpan county, Sichuan Province, China (102°45′E,30°23′N) and Linzhi county, Tibet Province, China (102°45′E,30°23′N), respectively. All samples were sent to Wuhan Benagen Technology Company Limited (Wuhan, China) for genome sequencing. Total genomic DNA was extracted using the Qiagen DNeasy Plant Mini Kit. For the Illumina short reads, DNA libraries with 500 bp insert sizes were constructed and sequenced using an Illumina HiSeq 4000 platform. For Oxford Nanopore sequencing, libraries were prepared using the SQKLSK109 ligation kit using the standard protocol, and the purified library was loaded onto primed R9.4 Spot-On Flow Cells and sequenced using a PromethION sequencer (Oxford Nanopore Technologies, Oxford, UK). The Hi-C sequencing was performed as follows: extracted DNA was first crosslinked by 40ml of 2% formaldehyde solution to capture interacting DNA segments, the chromatin DNA was then digested with the *DpnII* restriction enzyme, and libraries were constructed and sequenced using Illumina HiSeq 4000 instrument with 2×150 bp reads.

For RNA sequencing, fresh tissue samples were collected and immediately frozen in liquid nitrogen. Three biological replicates of five tissues of *S. obvallata* were sampled. Total RNA was extracted using the TRIzol® Reagent (Invitrogen, Shanghai, China). Paired-end cDNA libraries were constructed using TruSeq Stranded mRNA Library Prep Kit (Illumina), and were sequenced using Illumina HiSeq 4000 platform.

Genome assembly and quality control

Before genome assembly, Kmerfreq (https://github.com/fanagislab/kmerfreq) was used for counting *K*-mer frequency with the *K*-mer set to 19, and GCE v1.0 was used for estimating the genome size (Liu et al. 2013). The ONT long reads were corrected and assembled using NextDenovo v2.3.1 (https://github.com/Nextomics/NextDenovo) with default parameters. The contigs were polished using NextPolish v1.4.1 (Hu et al. 2020) with the long reads and Pilon v1.23 (Walker et al. 2014) with Illumina short reads for three rounds. Hi-C data were used to execute chromosome conformation using the 3D-DNA pipeline with default parameters. The accuracy of Hi-C based chromosomal assembly was improved using Juicerbox's chromatin contact matrix (Dudchenko et al. 2017). The completeness and continuity of the assemblies were evaluated by the statistics of BUSCO and LAI, respectively.

Genome annotation

Repeat annotation. TEs were identified based on *de novo* and homology-based strategies. RepeatMasker v4.0.7 was used to run a homology search for known repeat sequences against the Repbase database v22.11 (Tarailo-Graovac and Chen 2009). RepeatModeler v2.0.10 was employed to predict the TEs based on the *de novo* method (Flynn et al. 2020). LTRharvest v1.5.10, LTR_FINDER v1.05 and LTR_retriever v1.8.0 were used to build an LTR library with default parameters (Xu and Wang 2007; Ellinghaus et al. 2008; Ou and Jiang 2018). Finally, RepeatMasker was used to merge these library files of the two methods and to identify the repeat contents.

Gene prediction. A combination of *de novo*-, homology-, and transcript-based methods were used for gene prediction in both genomes. RNA-seq reads were assembled using Trinity v2.1.1 using the *de novo*-based and genome-guided modes, respectively. Coding DNA sequences (CDS) and protein sequences were predicted with TransDecoder (http://transdecoder.github.io). Homologs were predicted by mapping protein sequences using GeMoMa v1.6.1 (Keilwagen et al. 2016). Sequences of *Arabidopsis thaliana, Oryza sativa* and *Solanum tuberosum* were mapped to both genomes. Additionally, sequences of *Helianthus annuus, Lactuca sativa, Cynara cardunculus* and *Mikania*

493 *micrantha* were mapped to *S. obvallata*, and sequences of *Fagopyrum tataricum* and *Rumex hastatus*
494 were mapped to *R. alexandrae*, respectively. A *de novo* gene prediction was performed with Braker2
495 v2.1.5 and GlimmerHMM v3.0.4 (Majoros et al. 2004; Bruna et al. 2021). Assembled transcripts
496 were used for training gene models in Braker2. Gene models from the three main sources were
497 merged to produce consensus models using EVidenceModeler v1.1.1 (Haas et al. 2008).
498
499 ***Gene functional annotation.*** The annotated protein-coding genes were used for a BLAST search
500 against the UniProt and NCBI nonredundant protein databases to predict gene functions. The
501 functional domains of protein sequences were identified by InterProScan v5.51-85.0 using data from
502 Pfam (Jones et al. 2014). KEGG and gene ontology (GO) terms of the gene models were obtained
503 using eggNOG-mapper v2.0.1 (Cantalapiedra et al. 2021).
504
505 **Orthogroup inference and alignment**
506 Protein sequences from the seven alpine genomes as well as 13 relatives living at low altitudes were
507 used for subsequent comparative analyses. OrthoFinder v2.5.4 was used to construct orthogroups for
508 all species with default settings (Emms and Kelly 2019). Because the included species are
509 phylogenetically distant (widely dispersed across the eudicots and monocots), OrthoFinder recovered
510 an extremely low number of strictly single-copy orthogroups. We therefore reduced orthogroups that
511 contained multiple gene copies per species to one copy per species based on the smallest genetic
512 distance following the study of Birkeland et al. (2020). Briefly, all orthogroups were aligned based
513 on protein sequence using MAFFT v7.3 (Katoh and Standley 2013), and genetic distances between
514 all pairs of genes were calculated as Kimura protein distances (Kimura 1980). One gene copy per
515 species was retained based on the smallest genetic distance to the longest protein sequence of *A.*
516 *thaliana* in each orthogroup. The rationale is that the annotation of *A. thaliana* genome is complete
517 and reliable, and our subsequent functional analyses of candidate genes were based on the *A.*
518 *thaliana* orthologs. The resulting orthogroup sequences that contained only one protein sequence per
519 species were realigned using PRANK v170427 (Löytynoja 2014). Coding sequences of genes in
520 orthogroups were extracted based on the same gene identifier of protein sequences and were aligned

using PRANK with the codon mode.

Phylogeny and estimation of divergence time

Protein sequences of 195 single-copy orthogroups were used for phylogenetic inference with RAxML v8.2.12 using the PROTGAMMAAUTO substitution model with 500 bootstrap replicates (Stamatakis 2014). Divergence time was estimated in the MCMCtree program from PAML v4.9j (Yang 2007), using the approximate likelihood method with the tree topology inferred with RAxML, an independent substitution rate (clock = 2), and the HKY85+GAMMA model. Three calibration points were assigned based on the TimeTree database (Kumar et al. 2017) (http://www.timetree.org/, accessed on 1 May 2022): the MRCA (most recent common ancestor) of rosids (95% HPD 105 - 115 Mya), the MRCA of *Pentapetalae* (95% HPD 110 - 124 Mya), and the MRCA of *Mesangiospermae* (95% HPD 148 - 173 Mya). Samples were drawn every 1,000 MCMC steps from a total of 10^6 steps, with a burn-in of 10^5 steps. Convergence was assessed by comparing parameter estimates from two independent runs, with all effective sample sizes >200.

Gene family evolution

The change in gene family number was examined using CAFÉ5 (Mendes et al. 2020). Besides the base model, the number of gamma categories (-*k*) was set to estimate separate lambda values for different lineages in the tree (Gamma model). The highest likelihood was found using $k = 2$ rate categories (-lnL = 286114), with $\lambda = 0.00553$ and $\alpha = 1.58$. Gene family expansions or contractions were identified only when the change in gene count was significant with a p-value < 0.05. Genome-wide NBS-LRR genes were manually identified using HMMER v3.2 with an e-value 1e-05 (Wheeler and Eddy 2013). The NBS-LRR protein domains (NB-ARC: PF00931; RPW8: PF05659; TIR: PF01582; LRR: PF00560, PF07723, PF07725 and PF12799) were retrieved from Pfam (http://pfam.xfam.org) and were used to identify conserved motif of NBS-LRR genes in sampled genomes. The ML phylogenetic tree of NBS-LRR was constructed using RAxML. Then, Notung v2.9 was used to recover gains and losses of NBS-LRR genes by reconciling the NBS-LRR gene tree (Chen et al. 2000). The concatenated tree reconstructed by RAxML was used as input topology.

549

550 **Functional enrichment analysis**

551 GO and KEGG over-representation tests were performed using clusterProfiler v4.3.4 implemented in
552 R to identify significantly enriched pathways (Wu et al. 2021). GO and KEGG terms were assigned
553 according to the orthologous genes of *A. thaliana* genome. In the 'enrichGO' function, we set 'ont =
554 BP' to only search for enriched biological pathways (BP). The resulting *p*-values were corrected for
555 multiple comparisons using a Benjamini–Hochberg FDR correction. A criterion of *p*-adjust < 0.05
556 was used to access the significance of enrichment analyses.

557

558 **Tests for positive selection and relaxation of selection**

559 The branch-site model implemented in CodeML (PAML package) was performed to test for positive
560 selection. In this test, an alternative model allowing sites to be under positive selection on the
561 foreground branch was contrasted to a null model limiting sites to evolve neutrally or under
562 purifying selection using a likelihood-ratio test (LRT). LRT *p*-values were computed based on Chi-
563 squared distribution (df = 2) and were corrected for multiple tests at a *p*-adjust < 0.05 threshold using
564 a Benjamini-Hochberg FDR correction. A gene showing signature of positive selection in more than
565 three alpine species was identified as a convergent selected gene. Additionally, RELAX was
566 employed to test for the relaxation of selection on *S*-locus genes using the likelihood ratio test by
567 comparing the model fixing $k = 1$ and the model allowing k to be estimated (Wertheim et al. 2015).
568 In both analyses, seven tests were conducted separately by setting each alpine species as the
569 foreground branch.

570

571 **Convergent evolutionary rate shifts**

572 To perform the RERconverge analysis, we first used PAML to estimate maximum-likelihood gene
573 trees whose branch lengths represent evolutionary rates using the number of amino acid
574 substitutions. RERs were calculated using 'getAllResiduals' function with weight = T, scale = T and
575 cutoff = 0.001 (Kowalczyk et al. 2019). We set alpine species as foreground branches (branches of
576 the tree with the trait of interest) and ran 'foreground2Paths' function to estimate RERs for all genes

and to correlate them with trait evolution (i.e., alpine habitats in our study). We then used 'correlateWithBinaryPhenotype' function to test for a significant association between RERs and traits across all branches of the tree with a p-value < 0.05.

Convergent AA evolution

Two models were compared in PCOC analysis: the convergent model in which a site on convergent branches evolves under a profile different from that of the nonconvergent branches, and nonconvergent (null) model in which a site evolves under a single amino acid profile throughout the phylogeny (Rey et al. 2018). PCOC then detected convergent sites by identifying the better fit between the two models. To filter for only sites with strong evidence for convergent profile shifts, we set a posterior probability threshold of > 0.9 in the analysis.

Differential gene expression (DEG) analysis

RNA-seq data of *S. obvallata* were mapped to the assembled genome using HISAT2 v2.2.1 (Kim et al. 2019). Only uniquely mapped paired-end reads were retained for read counting by featureCounts v2.0.3 (Liao et al. 2014) to generate the count and Transcripts per Kilobase Million (TPM) tables. DEG analyses among the five tissues were performed in DESeq2 v1.36.0 (Love et al. 2014), with a p-value <0.05 as a cut-off and a log2 fold change cut-off of 1.

Data Availability

The genomic data generated and analyzed in this study including the raw sequencing data of Oxford Nanopore, Illumina, Hi-C and RNA-seq, as well as genome assembly have been deposited in China National GeneBank (CNGB, https://db.cngb.org/) under accession number CNP0003451. All the custom scripts and specific command lines have been deposited at GitHub (https://github.com/ZhangXu-CAS/Alpine_genome).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgements

The authors thank Ticao Zhang and Yingxiong Qiu for insightful comments on our paper, and thank Xinyi Guo for providing genome data. This work was supported by the Second Tibetan Plateau Scientific Expedition and Research program (2019QZKK0502), the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20050203), the Key Projects of the Joint Fund of the National Natural Science Foundation of China (U1802232), the Youth Innovation Promotion Association of Chinese Academy of Sciences (2019382), the Yunnan Young & Elite Talents Project (YNWR-QNBJ-2019-033) and the Ten Thousand Talents Program of Yunnan Province (202005AB160005).

Author contributions

T.D., H.W. and H.S. conceived and led the project. X.Z., T.K., W.D. and Z.Q. processed data, performed analyses and drew figures. X.Z., H.Z., T.F., L.L. and T.D. collected materials. X.Z. wrote the manuscript. J.B.L., Y.S., J.H., T.D., H.W. and H.S. revised the manuscript.

References

Abbas M, Sharma G, Dambire C, Marquez J, Alonso-Blanco C, Proaño K, Holdsworth MJ. 2022. An oxygen-sensing mechanism for angiosperm adaptation to altitude. *Nature* 606:565-569.

Agati G, Azzarello E, Pollastri S, Tattini M. 2012. Flavonoids as antioxidants in plants: location and functional significance. *Plant Sci.* 196:67-76.

Albert NW, Davies KM, Lewis DH, Zhang H, Montefiori M, Brendolise C, Boase MR, Ngo H, Jameson PE, Schwinn KE. 2014. A conserved network of transcriptional activators and repressors regulates anthocyanin pigmentation in eudicots. *Plant Cell* 26:962-980.

Beall CM. 2014. Adaptation to High Altitude: Phenotypes and Genotypes. *Annual Review of Anthropology* 43:251-272.

Bingham RA, Ort AR. 1998. Efficient pollination of alpine plants. *Nature* 391:238-239.

Birkeland S, Gustafsson ALS, Brysting AK, Brochmann C, Nowak MD. 2020. Multiple Genetic

633 Trajectories to Extreme Abiotic Stress Adaptation in Arctic Brassicaceae. *Mol Biol Evol*
634 37:2052-2068.
635 Bourdenx B, Bernard A, Domergue F, Pascal S, Leger A, Roby D, Pervent M, Vile D, Haslam RP,
636 Napier JA, et al. 2011. Overexpression of Arabidopsis ECERIFERUM1 promotes wax very-
637 long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant*
638 *Physiol.* 156:29-45.
639 Bruna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic
640 genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database.
641 *NAR Genom Bioinform* 3:lqaa108.
642 Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2:
643 Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic
644 Scale. *Mol Biol Evol* 38:5825-5829.
645 Cao X, Yang H, Shang C, Ma S, Liu L, Cheng J. 2019. The Roles of Auxin Biosynthesis YUCCA
646 Gene Family in Plants. *Int J Mol Sci* 20:6343.
647 Chen JH, Huang Y, Brachi B, Yun QZ, Zhang W, Lu W, Li HN, Li WQ, Sun XD, Wang GY, et al.
648 2019. Genome-wide analysis of Cushion willow provides insights into alpine plant divergence
649 in a biodiversity hotspot. *Nat Commun* 10:5230.
650 Chen K, Durand D, Farach-Colton M. 2000. NOTUNG: a program for dating gene duplications and
651 optimizing gene family trees. *J Comput Biol* 7:429-447.
652 Chen K, Fan B, Du L, Chen Z. 2004. Activation of hypersensitive cell death by pathogen-induced
653 receptor-like protein kinases from *Arabidopsis*. *Plant Mol Biol* 56:271-283.
654 Cheptou PO. 2012. Clarifying Baker's Law. *Ann. Bot.* 109:633-641.
655 Cui R, Medeiros T, Willemsen D, Iasi LNM, Collier GE, Graef M, Reichard M, Valenzano DR.
656 2019. Relaxed Selection Limits Lifespan by Increasing Mutation Load. *Cell* 178:385-399 e320.
657 Diener AC, Gaxiola RA, Fink GR. 2001. Arabidopsis ALF5, a multidrug efflux transporter gene
658 family member, confers resistance to toxins. *Plant Cell* 13:1625-1638.
659 Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I,
660 Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C

yields chromosome-length scaffolds. *Science* 356:92-95.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* 9:18.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20:238.

Fasani E, DalCorso G, Costa A, Zenoni S, Furini A. 2019. The *Arabidopsis thaliana* transcription factor MYB59 regulates calcium signalling during plant growth and stress response. *Plant Mol Biol* 99:517-534.

Feng L, Lin H, Kang M, Ren Y, Yu X, Xu Z, Wang S, Li T, Yang W, Hu Q. 2022. A chromosome-level genome assembly of an alpine plant *Crucihimalaya lasiocarpa* provides insights into high-altitude adaptation. *DNA Res.* 29:dsac004.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A* 117:9451-9457.

Flynn RL, Zou L. 2010. Oligonucleotide/oligosaccharide-binding fold proteins: a growing family of genome guardians. *Crit. Rev. Biochem. Mol. Biol.* 45:266-275.

Fujikawa K, Ikeda H, Murata K, Kobayashi T, Nakano T, Ohba H, Wu SG. 2004. Chromosome numbers of fifteen species of the genus *Saussurea* DC. (Asteraceae) in the Himalayas and the adjacent regions. *Journal of Japanese Botany* 79:271-280.

Fukao T, Bailey-Serres J. 2004. Plant responses to hypoxia--is survival a balancing act? *Trends Plant Sci.* 9:449-456.

Goodwillie C, Kalisz S, Eckert CG. 2005. The evolutionary enigma of mixed mating systems in plants: Occurrence, theoretical explanations, and empirical evidence. *Annu Rev Ecol Evol Syst* 36:47-79.

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read

689 assembly. *Bioinformatics* 36:2253-2255.
690 Hut RA, Beersma DG. 2011. Evolution of time-keeping mechanisms: early emergence and
691 adaptation to photoperiod. *Philos Trans R Soc Lond B Biol Sci* 366:2141-2154.
692 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
693 Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification.
694 *Bioinformatics* 30:1236-1240.
695 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
696 improvements in performance and usability. *Mol Biol Evol* 30:772-780.
697 Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016. Using intron position
698 conservation for homology-based gene prediction. *Nucleic Acids Res* 44:e89.
699 Kerstiens G. 1996. Cuticular water permeability and its physiological significance. *Journal of*
700 *Experimental Botany* 47:1813-1832.
701 Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and
702 genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37:907-915.
703 Kimura M. 1980. A simple method for estimating evolutionary rates of base substitutions through
704 comparative studies of nucleotide sequences. *J Mol Evol* 16:111-120.
705 Kolora SRR, Owens GL, Vazquez JM, Stubbs A, Chatla K, Jainese C, Seeto K, McCrea M, Sandel
706 MW, Vianna JA, et al. 2021. Origins and evolution of extreme life span in Pacific Ocean
707 rockfishes. *Science* 374:842-847.
708 Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M. 2019. RERconverge: an R
709 package for associating evolutionary rates with convergent traits. *Bioinformatics* 35:4815-4817.
710 Kowalczyk A, Partha R, Clark NL, Chikina M. 2020. Pan-mammalian analysis of molecular
711 constraints underlying extended lifespan. *Elife* 9:e51089.
712 Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees,
713 and Divergence Times. *Mol Biol Evol* 34:1812-1819.
714 Li JT, Gao YD, Xie L, Deng C, Shi P, Guan ML, Huang S, Ren JL, Wu DD, Ding L, et al. 2018.
715 Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proc*
716 *Natl Acad Sci U S A* 115:8406-8411.

717 Liao Y, Smyth GK, Shi W. 2014. featureCounts: an efficient general purpose program for assigning
718 sequence reads to genomic features. *Bioinformatics* 30:923-930.
719 Licausi F, Kosmacz M, Weits DA, Giuntoli B, Giorgi FM, Voesenek LA, Perata P, van Dongen JT.
720 2011. Oxygen sensing in plants is mediated by an N-end rule pathway for protein
721 destabilization. *Nature* 479:419-422.
722 Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W. 2013. Estimation of
723 genomic characteristics by analyzing k-mer frequency in de novo genome projects.
724 Louvet R, Cavel E, Gutierrez L, Guenin S, Roger D, Gillet F, Guerineau F, Pelloux J. 2006.
725 Comprehensive expression profiling of the pectin methylesterase gene family during silique
726 development in *Arabidopsis thaliana*. *Planta* 224:782-791.
727 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-
728 seq data with DESeq2. *Genome Biol* 15:550.
729 Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple
730 Sequence Alignment Methods. Totowa, NJ: Humana Press. p. 155-170.
731 Lu S, Wang J, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Marchler
732 GH, Song JS, et al. 2020. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic*
733 *Acids Res* 48:D265-D268.
734 Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open source ab initio
735 eukaryotic gene-finders. *Bioinformatics* 20:2878-2879.
736 Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO Update: Novel and
737 Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of
738 Eukaryotic, Prokaryotic, and Viral Genomes. *Mol Biol Evol* 38:4647-4654.
739 Mano S, Hayashi M, Nishimura M. 1999. Light regulates alternative splicing of hydroxypyruvate
740 reductase in pumpkin. *Plant J.* 17:309-320.
741 Martin JT, Juniper BE. 1970. The cuticles of plants. New York: St. Martin's Press.
742 McHale L, Tan X, Koehl P, Michelmore RW. 2006. Plant NBS-LRR proteins: adaptable guards.
743 *Genome Biol* 7:212.
744 Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2020. CAFE 5 models variation in evolutionary

745 rates among gene families. *Bioinformatics* 36:5516-5518.
746 Mockaitis K, Estelle M. 2008. Auxin receptors and plant development: a new signaling paradigm.
747 *Annu Rev Cell Dev Biol* 24:55-80.
748 Mosher S, Seybold H, Rodriguez P, Stahl M, Davies KA, Dayaratne S, Morillo SA, Wierzba M,
749 Favery B, Keller H, et al. 2013. The tyrosine-sulfated peptide receptors PSKR1 and PSY1R
750 modify the immunity of Arabidopsis to biotrophic and necrotrophic pathogens in an antagonistic
751 manner. *Plant J.* 73:469-482.
752 Nagy L, Grabherr G. 2009. The Biology of Alpine Habitats. Oxford: Oxford University Press.
753 Ortiz-Morea FA, Liu J, Shan L, He P. 2022. Malectin-like receptor kinases as protector deities in
754 plant immunity. *Nat Plants* 8:27-37.
755 Ou S, Chen J, Jiang N. 2018. Assessing genome assembly quality using the LTR Assembly Index
756 (LAI). *Nucleic Acids Res* 46:e126.
757 Ou S, Jiang N. 2018. LTR_retriever: A Highly Accurate and Sensitive Program for Identification of
758 Long Terminal Repeat Retrotransposons. *Plant Physiol.* 176:1410-1422.
759 Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF. 2000. B and C floral organ identity
760 functions require *SEPALLATA* MADS-box genes. *Nature* 405:200-203.
761 Peng HP, Chan CS, Shih MC, Yang SF. 2001. Signaling events in the hypoxic induction of alcohol
762 dehydrogenase gene in Arabidopsis. *Plant Physiol.* 126:742-749.
763 Pollard M, Beisson F, Li Y, Ohlrogge JB. 2008. Building lipid barriers: biosynthesis of cutin and
764 suberin. *Trends Plant Sci.* 13:236-246.
765 Rey C, Gueguen L, Semon M, Boussau B. 2018. Accurate Detection of Convergent Amino-Acid
766 Evolution with PCOC. *Mol Biol Evol* 35:2296-2306.
767 Rodriguez-Gacio Mdel C, Iglesias-Fernandez R, Carbonero P, Matilla AJ. 2012. Softening-up
768 mannan-rich cell walls. *J Exp Bot* 63:3976-3988.
769 Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ,
770 Clamp M, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds.
771 *Science* 364:74-78.
772 Sherman-Broyles S, Boggs N, Farkas A, Liu P, Vrebalov J, Nasrallah ME, Nasrallah JB. 2007. S

773 locus genes and the evolution of self-fertility in Arabidopsis thaliana. *Plant Cell* 19:94-106.
774 Song B, Stocklin J, Peng DL, Gao YQ, Sun H. 2015. The bracts of the alpine 'glasshouse' plant
775 Rheum alexandrae (Polygonaceae) enhance reproductive fitness of its pollinating seed-
776 consuming mutualist. *Botanical Journal of the Linnean Society* 179:349-359.
777 Spicer RA, Farnsworth A, Su T. 2020. Cenozoic topography, monsoons and biodiversity
778 conservation within the Tibetan Region: An evolving story. *Plant Divers* 42:229-254.
779 Srivastava D, Shamim M, Kumar M, Mishra A, Maurya R, Sharma D, Pandey P, Singh KN. 2019.
780 Role of circadian rhythm in plant system: An update from development to stress response.
781 *Environmental and Experimental Botany* 162:256-271.
782 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
783 phylogenies. *Bioinformatics* 30:1312-1313.
784 Stern DL, Orgogozo V. 2009. Is genetic evolution predictable? *Science* 323:746-751.
785 Storz JF. 2016. Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet*
786 17:239-250.
787 Sun H, Niu Y, Chen YS, Song B, Liu CQ, Peng DL, Chen JG, Yang Y. 2014. Survival and
788 reproduction of plant species in the Qinghai-Tibet Plateau. *Journal of Systematics and Evolution*
789 52:378-396.
790 Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, Doncheva NT, Legeay M, Fang
791 T, Bork P, et al. 2021. The STRING database in 2021: customizable protein-protein networks,
792 and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res*
793 49:D605-D612.
794 Takayama S, Isogai A. 2005. Self-incompatibility in plants. *Annu. Rev. Plant Biol.* 56:467-489.
795 Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic
796 sequences. *Curr Protoc Bioinformatics* Chapter 4:Unit 4 10.
797 To JP, Haberer G, Ferreira FJ, Deruere J, Mason MG, Schaller GE, Alonso JM, Ecker JR, Kieber JJ.
798 2004. Type-A Arabidopsis response regulators are partially redundant negative regulators of
799 cytokinin signaling. *Plant Cell* 16:658-671.
800 Tsukaya H, Tsuge T. 2001. Morphological adaptation of inflorescences in plants that develop at low

temperatures in early spring: The convergent evolution of "downy plants". *Plant Biol.* 3:536-543.

van Berkel K, de Boer RJ, Scheres B, ten Tusscher K. 2013. Polar auxin transport: models and mechanisms. *Development* 140:2253-2268.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963.

Wang X, Kong L, Zhi P, Chang C. 2020. Update on Cuticular Wax Biosynthesis and Its Roles in Plant Disease Resistance. *Int J Mol Sci* 21.

Wang X, Liu S, Zuo H, Zheng W, Zhang S, Huang Y, Pingcuo G, Ying H, Zhao F, Li Y, et al. 2021. Genomic basis of high-altitude adaptation in Tibetan Prunus fruit trees. *Curr Biol* 31:3848-3860 e3848.

Wen R, Torres-Acosta JA, Pastushok L, Lai X, Pelzer L, Wang H, Xiao W. 2008. Arabidopsis UEV1D promotes Lysine-63-linked polyubiquitination and is involved in DNA damage response. *Plant Cell* 20:213-227.

Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* 32:820-832.

Wheeler TJ, Eddy SR. 2013. nhmmer: DNA homology search with profile HMMs. *Bioinformatics* 29:2487-2489.

Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, et al. 2021. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* 2:100141.

Xu S, Wang J, Guo Z, He Z, Shi S. 2020. Genomic Convergence in the Adaptation to Extreme Environments. *Plant Commun* 1:100117.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* 35:W265-268.

Yang CY, Hsu FC, Li JP, Wang NN, Shih MC. 2011. The AP2/ERF transcription factor AtERF73/HRE1 modulates ethylene responses during hypoxia in Arabidopsis. *Plant Physiol.*

156:202-212.

Yang XH, Makaroff CA, Ma H. 2003. The Arabidopsis MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis. *Plant Cell* 15:1281-1295.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-1591.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, et al. 2016. Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431-1433.

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K. 2008. Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* 20:2160-2176.

Zeng X, Long H, Wang Z, Zhao S, Tang Y, Huang Z, Wang Y, Xu Q, Mao L, Deng G, et al. 2015. The draft genome of Tibetan hulless barley reveals adaptive patterns to the high stressful Tibetan Plateau. *Proc Natl Acad Sci U S A* 112:1095-1100.

Zhang J, Kumar S. 1997. Detection of convergent and parallel evolution at the amino acid sequence level. *Mol Biol Evol* 14:527-536.

Zhang J, Tian Y, Yan L, Zhang G, Wang X, Zeng Y, Zhang J, Ma X, Tan Y, Long N, et al. 2016. Genome of Plant Maca (*Lepidium meyenii*) Illuminates Genomic Basis for High-Altitude Adaptation in the Central Andes. *Mol Plant* 9:1066-1077.

Zhang T, Qiao Q, Novikova PY, Wang Q, Yue J, Guan Y, Ming S, Liu T, De J, Liu Y, et al. 2019. Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *Proc Natl Acad Sci U S A* 116:7137-7146.

Zhang Y, Zhou J, Lim CU. 2006. The role of NBS1 in DNA double strand break repair, telomere stability, and cell cycle checkpoint control. *Cell Res.* 16:45-54.

Zhen Y, Aardema ML, Medina EM, Schumer M, Andolfatto P. 2012. Parallel molecular evolution in an herbivore community. *Science* 337:1634-1637.

Zhou B, Mural RV, Chen X, Oates ME, Connor RA, Martin GB, Gough J, Zeng L. 2017. A Subset of Ubiquitin-Conjugating Enzymes Is Essential for Plant Immunity. *Plant Physiol.* 173:1371-1390.

857 **Figure Legneds**

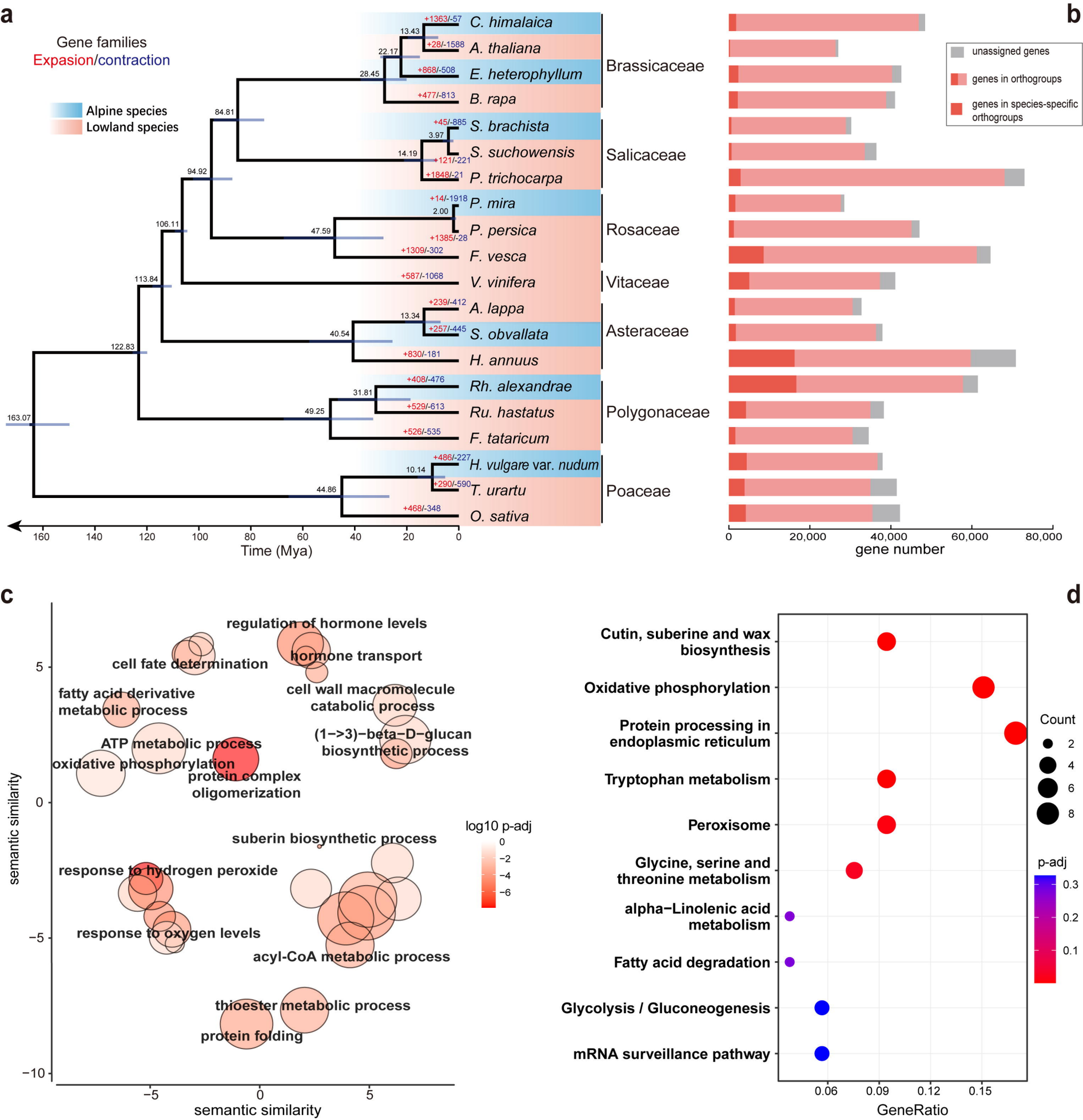
858 **FIG. 1.** Evolutionary history of alpine plants. (a) Chronogram showing divergence times among
859 alpine plants (cyan background) with their lowland relatives (orange background) with node age and
860 95% confidence intervals (blue bars). The red and blue numbers above the branches represent
861 significant expansion and contraction events, respectively. (b) Bar plot showing gene number
862 identified by OrthorFinder. (c) REVIGO clusters of significantly enriched GO terms for convergently
863 expended (CoEx) gene families in alpine plant genomes. Each bubble represents a summarized GO
864 term from the full GO list by reducing functional redundancies, and their closeness on the plot
865 reflects their closeness in the GO graph, i.e., the semantic similarity. (d) KEGG pathways of CoEx
866 gene families in alpine plant genomes. "p-adj" refers to the adjusted *p*-value using the Benjamini–
867 Hochberg method.
868
869 **FIG. 2.** Function enrichment results of gene families undergoing convergently evolutionary rate
870 acceleration in alpine plant genomes. (a) REVIGO clusters of significantly enriched GO terms. Each
871 bubble represents a summarized GO term from the full GO list by reducing functional redundancies,
872 and their closeness on the plot reflects their closeness in the GO graph, i.e., the semantic similarity.
873 (b) Top 20 enriched KEGG pathways. "p-adj" refers to the adjusted *p*-value using the Benjamini–
874 Hochberg method.
875
876 **FIG. 3.** Molecular convergence of *S*-locus genes. (a) Convergent evolutionary rate shifts detected by
877 RERconverge using correlations of relative rates of the gene with traits of interest (here alpine
878 environments) (Kowalczyk et al. 2020). (b) Site-based estimation of convergent amino acid
879 evolution using PCOC method. Only sites with a posterior probability above 0.9 according to the
880 PCOC model are represented. Detailed information for the PCOC, PC, and OC models is described
881 in Rey et al. (2018). Alpine species are in red and their low-land relatives are in blue.
882
883

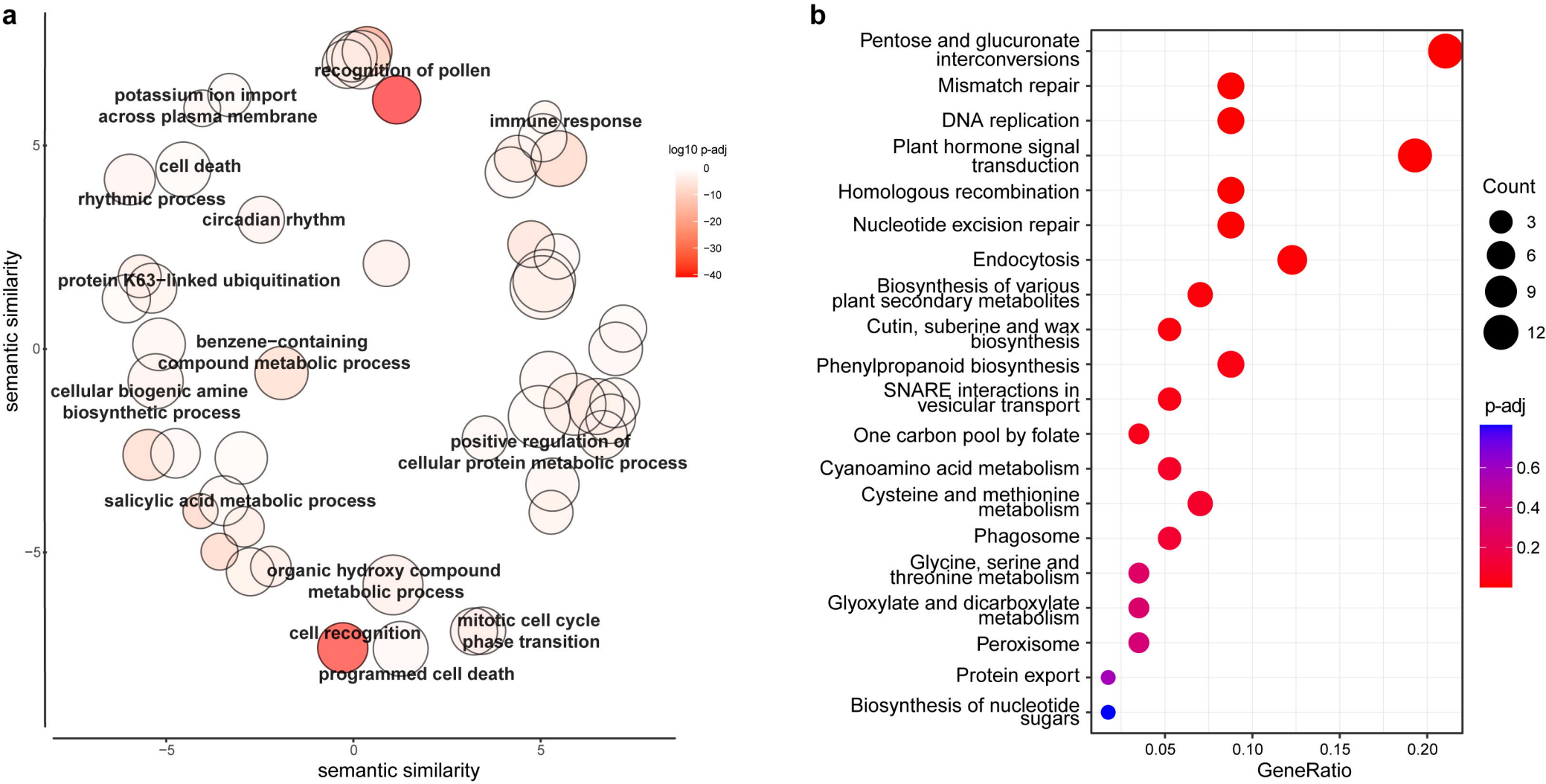
884 **FIG. 4.** Highly expressed genes in bract leaves of *Saussurea obvallata*. (a) Differentially expressed
885 genes between bract leaves with normal leaves. (b) Top 20 KEGG pathways of significantly up-
886 regulated genes in bract leaves. "p-adj" refers to the adjusted *p*-value using the Benjamini–Hochberg
887 method. Expression profiles of genes involved in cuticular wax (c) and flavonoid biosynthetic
888 pathways (d) in different tissues of *S. obvallata* (F: flowers, S: stems, JL: basal leaves, ML: middle
889 leaves and BL: bract leaves). High expressed genes in BL are shown in red in the simplified pathway
890 models. The bar represents the gene expression level of each gene (z-score). Abbreviations: FA-CoA,
891 fatty acyl-coenzyme A; VLC, very-long-chain; CER, protein eceriferum; FAR, fatty acid synthetase;
892 WSD1, diacylglycerol acyltransferase 1; MAH1, midchain alkane hydroxylase 1; 4CL, 4-coumarate:
893 CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase;
894 F3'H, flavonoid 3'-hydroxylase; FLS, flavonol synthase; TT7, transparent testa 7; OMT, O-
895 methyltransferase; DFR, dihydroflavonol 4-reductase; LDOX/ANS, leucoanthocyanidin
896 dioxygenase/anthocyanidin synthase. The pathways for the cuticular wax biosynthesis was adapted
897 from the study of Wang et al. (2020), and for flavonoid biosynthesis was adapted from the study of
898 Yonekura-Sakakibara et al. (2008).
899
900 **FIG. 5.** Summary of convergent adaptation to high-altitude environments for seven alpine species.
901 The outer circle shows examples of enriched GO-terms (biological pathways). The middle circle
902 shows examples of candidate genes named after the *A. thaliana* orthologs. The inner gray circle
903 shows environmental stresses from high altitudes. Species names are provided below the picture.
904 *Genes undergone convergent contraction.
905

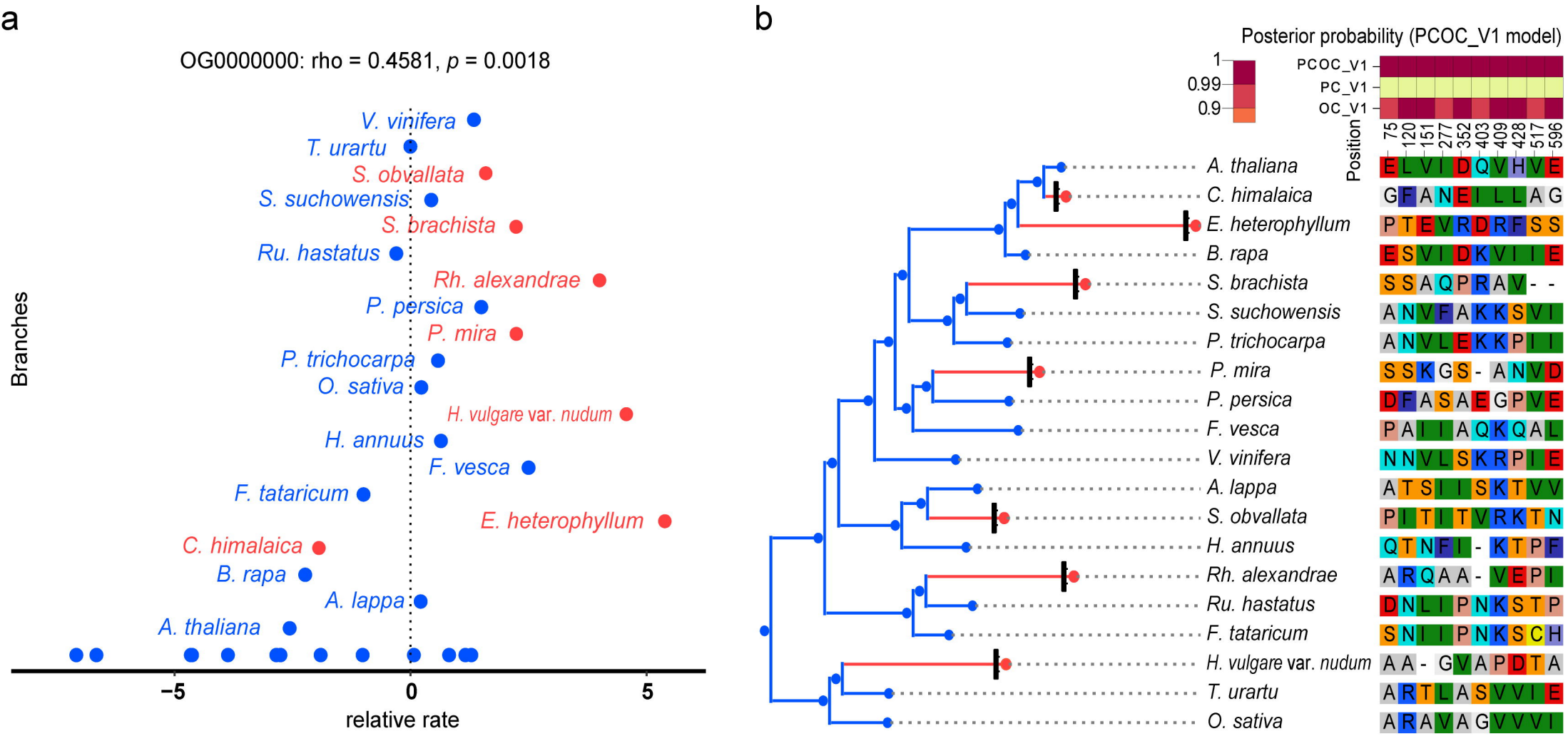
906 **Table 1. Statistics of genome assembly and annontation of *S. obvallata* and *R. alexandrae*.**

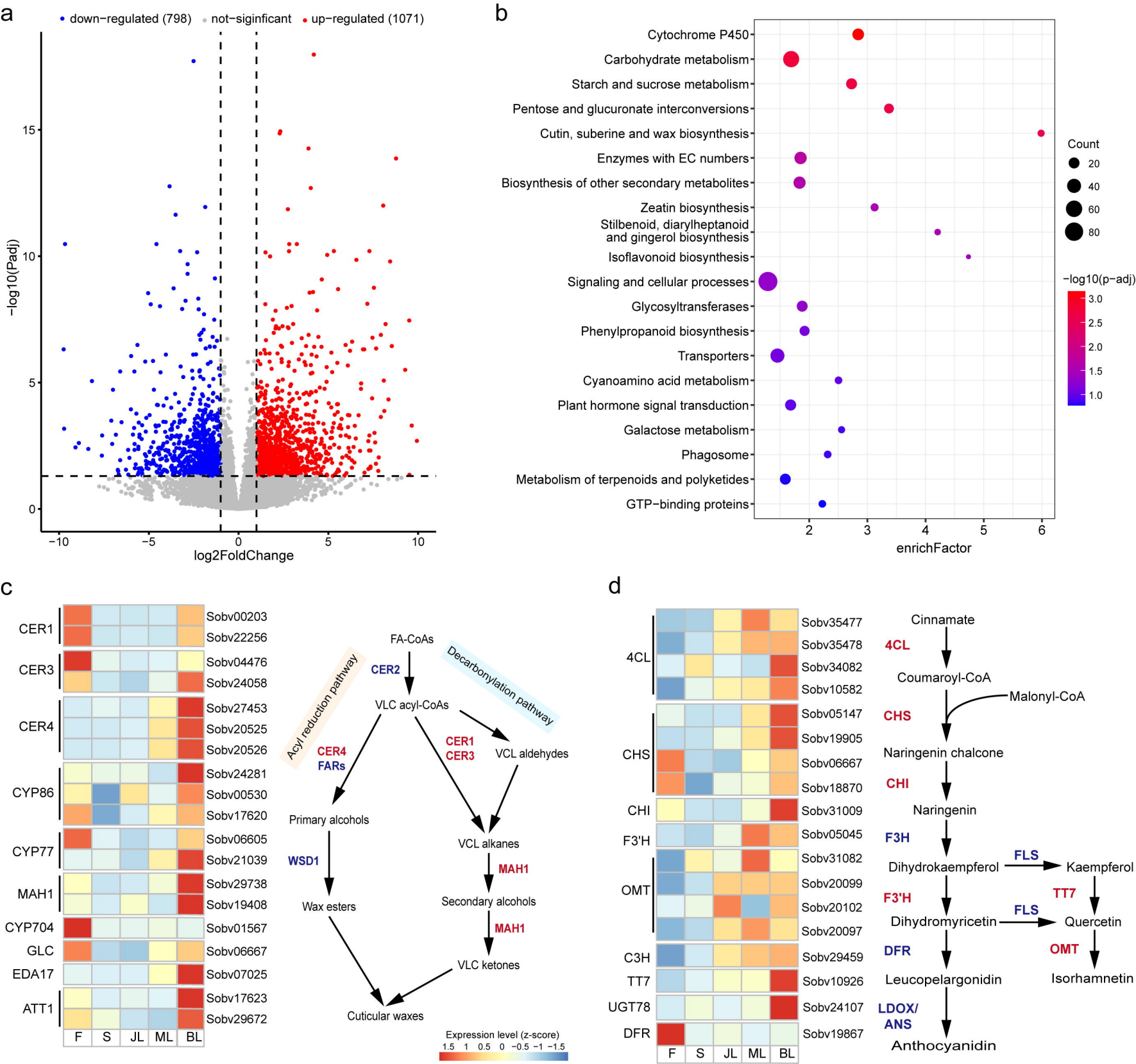
Statistic	***S. obvallata***	***R. alexandrae***
Total length (bp)	2,044,030,733	2,039,881,226
Number of contigs	145	129
Largest contig (bp)	126,457,859	160,815,976
Anchored length (bp)	1,951,503,694	-
GC (%)	37.94	41.41
N50 (bp)	36,958,263	36,323,674
Complete BUSCOs (%)	94.6	94.1
Repeat content (%)	81.88	81.65
Number of genes	37,938	61,463

907


a
Gene families
Expasion/contraction
Alpine species
Lowland species
C. himalaica
A. thaliana
E. heterophyllum
B. rapa
S. brachista
S. suchowensis
P. trichocarpa
P. mira
P. persica
F. vesca
V. vinifera
A. lappa
S. obvallata
H. annuus
Rh. alexandrae
Ru. hastatus
F. tataricum
H. vulgare var. nudum
T. urartu
O. sativa
Brassicaceae
Salicaceae
Rosaceae
Vitaceae
Asteraceae
Polygonaceae
Poaceae
Time (Mya)
b
unassigned genes
genes in orthogroups
genes in species-specific orthogroups
gene number
c
regulation of hormone levels
hormone transport
cell fate determination
cell wall macromolecule catabolic process
fatty acid derivative metabolic process
(1−>3)−beta−D−glucan biosynthetic process
ATP metabolic process
oxidative phosphorylation
protein complex oligomerization
suberin biosynthetic process
response to hydrogen peroxide
response to oxygen levels
acyl−CoA metabolic process
thioester metabolic process
protein folding
log10 p-adj
semantic similarity
d
Cutin, suberine and wax biosynthesis
Oxidative phosphorylation
Protein processing in endoplasmic reticulum
Tryptophan metabolism
Peroxisome
Glycine, serine and threonine metabolism
alpha−Linolenic acid metabolism
Fatty acid degradation
Glycolysis / Gluconeogenesis
mRNA surveillance pathway
Count
p-adj
GeneRatio


a
recognition of pollen
potassium ion import
across plasma membrane
immune response
cell death
rhythmic process
circadian rhythm
protein K63−linked ubiquitination
benzene−containing
compound metabolic process
cellular biogenic amine
biosynthetic process
positive regulation of
cellular protein metabolic process
salicylic acid metabolic process
organic hydroxy compound
metabolic process
mitotic cell cycle
phase transition
cell recognition
programmed cell death
log10 p-adj
0
−10
−20
−30
−40
semantic similarity
semantic similarity
5
0
−5
−5
0
5
b
Pentose and glucuronate
interconversions
Mismatch repair
DNA replication
Plant hormone signal
transduction
Homologous recombination
Nucleotide excision repair
Endocytosis
Biosynthesis of various
plant secondary metabolites
Cutin, suberine and wax
biosynthesis
Phenylpropanoid biosynthesis
SNARE interactions in
vesicular transport
One carbon pool by folate
Cyanoamino acid metabolism
Cysteine and methionine
metabolism
Phagosome
Glycine, serine and
threonine metabolism
Glyoxylate and dicarboxylate
metabolism
Peroxisome
Protein export
Biosynthesis of nucleotide
sugars
0.05
0.10
0.15
0.20
GeneRatio
Count
3
6
9
12
p-adj
0.6
0.4
0.2


a
OG0000000: rho = 0.4581, p = 0.0018
Branches
V. vinifera
T. urartu
S. obvallata
S. suchowensis
S. brachista
Ru. hastatus
Rh. alexandrae
P. persica
P. mira
P. trichocarpa
O. sativa
H. vulgare var. nudum
H. annuus
F. vesca
F. tataricum
E. heterophyllum
C. himalaica
B. rapa
A. lappa
A. thaliana
−5
0
5
relative rate
b
Posterior probability (PCOC_V1 model)
1
0.99
0.9
PCOC_V1
PC_V1
OC_V1
Position
75
120
151
277
352
403
409
428
517
596
A. thaliana
C. himalaica
E. heterophyllum
B. rapa
S. brachista
S. suchowensis
P. trichocarpa
P. mira
P. persica
F. vesca
V. vinifera
A. lappa
S. obvallata
H. annuus
Rh. alexandrae
Ru. hastatus
F. tataricum
H. vulgare var. nudum
T. urartu
O. sativa


a
down−regulated (798)
not−siginficant
up−regulated (1071)
−log10(Padj)
log2FoldChange
b
Cytochrome P450
Carbohydrate metabolism
Starch and sucrose metabolism
Pentose and glucuronate interconversions
Cutin, suberine and wax biosynthesis
Enzymes with EC numbers
Biosynthesis of other secondary metabolites
Zeatin biosynthesis
Stilbenoid, diarylheptanoid and gingerol biosynthesis
Isoflavonoid biosynthesis
Signaling and cellular processes
Glycosyltransferases
Phenylpropanoid biosynthesis
Transporters
Cyanoamino acid metabolism
Plant hormone signal transduction
Galactose metabolism
Phagosome
Metabolism of terpenoids and polyketides
GTP−binding proteins
enrichFactor
Count
−log10(p−adj)
c
CER1
CER3
CER4
CYP86
CYP77
MAH1
CYP704
GLC
EDA17
ATT1
F
S
JL
ML
BL
FA-CoAs
CER2
VLC acyl-CoAs
Acyl reduction pathway
Decarbonylation pathway
CER4
FARs
CER1
CER3
VCL aldehydes
Primary alcohols
VCL alkanes
WSD1
MAH1
Wax esters
Secondary alcohols
MAH1
VLC ketones
Cuticular waxes
Expression level (z-score)
d
4CL
CHS
CHI
F3'H
OMT
C3H
TT7
UGT78
DFR
Cinnamate
4CL
Coumaroyl-CoA
Malonyl-CoA
CHS
Naringenin chalcone
CHI
Naringenin
F3H
Dihydrokaempferol
FLS
Kaempferol
F3'H
TT7
Dihydromyricetin
FLS
Quercetin
DFR
OMT
Leucopelargonidin
Isorhamnetin
LDOX/ ANS
Anthocyanidin

response to hydrogen peroxide

response to decreased oxygen level

response to hypoxia

cutin, suberine and wax biosynthesis

flavonoid biosynthesis

suberin biosynthetic process

BOB1

EER1

ERF1

ADH1

CER1

FARs

MYB27

MYB48

hypobaric hypoxia

high UV radiation

DNA repair

protein ubiquitination

mismatch repair

homologous recombination

UEV1

NBS1

ATELF5A

OB-fold protein

molecular damage

unpredictable climate

hormone signal transduction

circadian rhythm

programmed cell death

regulation of enzyme activity

ARR4

MYB59

LWD2

SHM1

reproduction

photorespiration

leaf senescence

auxin polar transport

SEP1

MMD1

HPR

SAUR

D6PK

general abiotic stress

depauperate pathogens

CRKs*

MLRs*

MATE*

PSY1Rs*

PSKRs*

detoxification

xenobiotic transport

response to oomycetes

response to toxic substance

genes

GO-terms

genes

GO-terms

[1]*S. obvallata*, [2]*R. alexandrae*, [3]*E. heterophyllum*, [4]*C. himalaica*, [5]*P. mira*, [6]*S. brachista*, [7]*H. vulgare* var. *nudum*