



OPEN ACCESS

EDITED BY

Adrian Christopher Brennan,
Durham University, United Kingdom

REVIEWED BY

Tao Shi,
Wuhan Botanical Garden, Chinese
Academy of Sciences (CAS), China
Baosheng Wang,
South China Botanical Garden,
Chinese Academy of Sciences (CAS),
China

*CORRESPONDENCE

La Qiong
lhagcho@163.com
Wenju Zhang
wjzhang@fudan.edu.cn

[†]These authors have contributed
equally to this work

SPECIALTY SECTION

This article was submitted to
Plant Systematics and Evolution,
a section of the journal
Frontiers in Plant Science

RECEIVED 23 September 2022

ACCEPTED 22 November 2022

PUBLISHED 14 December 2022

CITATION

Wang R, Wu B, Jian J, Tang Y,
Zhang T, Song Z, Zhang W and
Qiong L (2022) How to survive in the
world's third poplar: Insights from the
genome of the highest altitude woody
plant, *Hippophae tibetana*
(Elaeagnaceae).
Front. Plant Sci. 13:1051587.
doi: 10.3389/fpls.2022.1051587

COPYRIGHT

© 2022 Wang, Wu, Jian, Tang, Zhang,
Song, Zhang and Qiong. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use,
distribution or reproduction is
permitted which does not comply with
these terms.

How to survive in the world's third poplar: Insights from the genome of the highest altitude woody plant, *Hippophae tibetana* (Elaeagnaceae)

Ruoqiu Wang^{1,2†}, Bin Wu^{3†}, Jianbo Jian³, Yiwei Tang²,
Ticao Zhang², Zhiping Song^{1,2}, Wenju Zhang^{1,2*}
and La Qiong^{1,4*}

¹Tibet University-Fudan University Joint Laboratory for Biodiversity and Global Change, School of Life Sciences, Fudan University, Shanghai, China, ²Key Laboratory for Biodiversity Science and Ecological Engineering, Institute of Biodiversity Science, School of Life Sciences, Fudan University, Shanghai, China, ³BGI-Shenzhen, Shenzhen, China, ⁴Research Center for Ecology, College of Science, Tibet University, Lhasa, China

Hippophae tibetana (Tibetan sea-buckthorn) is one of the highest distributed woody plants in the world (3,000–5,200 meters a.s.l.). It is characterized by adaptation to extreme environment and important economic values. Here, we combined PacBio Hifi platform and Hi-C technology to assemble a 1,452.75 Mb genome encoding 33,367 genes with a Contig N50 of 74.31 Mb, and inferred its sexual chromosome. Two *Hippophae*-specific whole-genome duplication events (18.7–21.2 million years ago, Ma; 28.6–32.4 Ma) and long terminal repeats retroelements (LTR-RTs) amplifications were detected. Comparing with related species at lower altitude, *Ziziphus jujuba* (<1, 700 meters a.s.l.), *H. tibetana* had some significantly rapid evolving genes involved in adaptation to high altitude habitats. However, comparing with *Hippophae rhamnoides* (<3, 700 meters a.s.l.), no rapid evolving genes were found except microtubule and microtubule-based process genes, *H. tibetana* has a larger genome, with extra 2, 503 genes (7.5%) and extra 680.46 Mb transposable elements (TEs) (46.84%). These results suggest that the changes in the copy number and regulatory pattern of genes play a more important role for *H. tibetana* adapting to more extreme and variable environments at higher altitude by more TEs and more genes increasing genome variability and expression plasticity. This suggestion was supported by two findings: nitrogen-fixing genes of *H. tibetana* having more copies, and intact TEs being significantly closer genes than fragmentary TEs. This study provided new insights into the evolution of alpine plants.

KEYWORDS

genome sequencing, *Hippophae tibetana*, whole-genome duplication, high-altitude adaptation, complete genome

Introduction

The Qinghai-Tibet Plateau (QTP) is one of the most extreme regions in the world, but it is also one of the regions with the highest biodiversity (Myers et al., 2000; Sun, 2002; Wen et al., 2014). For nearly 20 years, it has become a hot spot in the study of adaptive evolution (Favre et al., 2015; Yang et al., 2019; Zhang et al., 2019). Since 50 Ma (million years ago), this area has undergone long-term and huge geological changes, forming the highest and largest plateau in the world, with an average elevation of more than 4,000 meters (Royden et al., 2008; Deng and Ding, 2015). Especially, since the Quaternary, the QTP and its surrounding areas have been directly affected by drastic geology and climate changes (An et al., 2011; Liu and Dong, 2013; Li et al., 2015; Ren et al., 2018). Although its habitat is extreme and special, many unique taxa (*Rhodiola*, *Rhododendron*, *Berberis*, *Pedicularis*, *Hippophae*, etc) have evolved in the special environment. Organisms growing in the QTP undergo multiple evolutionary pressures than low-altitude species, including low temperature and oxygen, poor soils, and strong ultraviolet (UV) radiation (Sun, 1996). How plants adapt to extreme habitats at high altitudes has always been a fascinating and challenging problem.

With the development of high-throughput sequencing technology, some studies have tried to uncover the adaptation mechanism of plants to high-altitude habitats on the basis of the genome. Previous studies revealed some candidate genes related to high-altitude adaptation, as well as specific whole-genome duplication events (Zhang et al., 2016; Yang et al., 2020), LTR retrotransposons proliferation (Zhang et al., 2019; Geng et al., 2021; Feng et al., 2022), expanding gene families associated with DNA damage repair (Chen et al., 2019; Wang et al., 2021) and mutation of the flowering protein locus (Geng et al., 2021), which may contribute to the high-altitude environment adaption.

However, complete genomes of woody plants are rarely sequenced and published, especially for those at altitudes above 4,000 meters. In addition, existing studies mainly focused on the discovery of the adaptive evolution of specific genes in response to extreme ecological factors (genes involved in altitude adaptation: hypoxia response, eg: *PKLR*, *SRF*; DNA damage repair, eg: *DRT102*, *TFB1*; UV-B tolerance, eg: *SCC3*, *XRCC4*; abiotic stress resistance, eg: *Hsp70*, *MYB*; reproduction pathways and cold tolerance, eg: *FmHd3a*, *FmFT*) (Zhang et al., 2016; Geng et al., 2021; Mao et al., 2021). Although the role of WGD and the proliferation of TE have also been concerned, how they drive organisms to adapt to higher habitats remains unclear. Especially, whether woody plants at higher altitudes have similar or other adaptive mechanisms needs more works.

The genus *Hippophae* (Elaeagnaceae), commonly known as sea-buckthorn, consists of dioecious shrubby or woody species which are typical alpine plants mainly distributed in the Qinghai-Tibet Plateau (Lian et al., 1998). Due to its high drought-tolerance and wind-sand-tolerance, *Hippophae* species

are widely used for soil/water conservation and phytoremediation of mining wasteland. As an important resource plant, large amounts of *Hippophae* plants are planted in northwestern China for desert greening and food ingredients (Lu et al., 2008). As one species of *Hippophae*, *H. tibetana* is a pioneer plant that can live in extreme habitats owing to a remarkable adaptability to cold, drought, extremely low supplies of nitrogen, and powerful UV-B ultraviolet light stress (280-315 nm; Körner, 2003), and is also one of woody plants grown at highest altitude in the world (Figures 1A, B), reaching to ~3,000-5,200 meters above sea level (a.s.l.). The morphological characteristics and high altitude habitat of *H. tibetana* are illustrated in Figures 1A, B. Nowadays, the genome of *Hippophae rhamnoides*, a relative of *H. tibetana*, has been published (Wu et al., 2022; Yu et al., 2022), *H. rhamnoides* is a regular or small tree distributed at lower altitude ranged from 490-3,700 meters a.s.l. (Huang and Yu, 2006). It is possible to reveal how plants adapt to more extreme habitats at higher altitude by comparing the genomes of the two species. In this study, we sequenced and assembled a high-quality genome of *H. tibetana*, and performed a detailed analysis of the genome evolution. The results could provide insights into evolution of the alpine plants.

Materials and methods

Sample collection

One wild *H. tibetana* individual used for *de novo* assembly was collected from Ganzi Tibetan Autonomous Prefecture, Sichuan Province, China (N 100.1056, E 29.1806). Total genomic DNA was isolated from the fresh leaves using the modified cetyltrimethylammonium bromide (CTAB) method (Doyle and Doyle, 1987), then dissolved in 50 μ L of sterilized water and kept at -30°C . Young leaves were collected from another individual seedlings from the same population for Hi-C sequencing. For RNA sequencing, seeds of the same plant used for genome assembly were selected and cultured. After 4 weeks of growth, young leaves, stems and roots were collected. All samples were rinsed with Milli-Q water and immediately stored in liquid nitrogen.

Library construction and sequencing

A 250 bp short insert libraries was constructed and sequenced in 150 bp paired-end mode on the MGISEQ-2000 platform. The adapters and low-quality reads (reads with N bases more than 1% and the number of bases with Q-score ≤ 20 more than 10%) in raw data were removed using SOAPnuke 1.5.6 software (Chen et al., 2018) with the parameters set as follows: -n 0.01, -l 20, -q 0.1, -i, -Q 2, -G, -M 2, -A 0.5, -d. After filtration, 98.05 Gb high-quality data

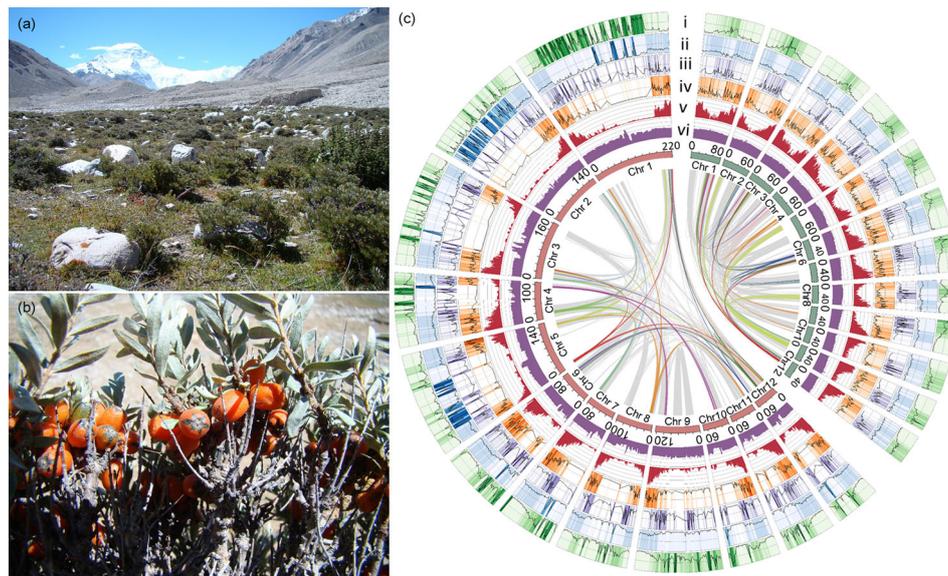


FIGURE 1

Habitat, morphological characteristics, and genomic features of *H. tibetana*. (A) High altitude habitat of *H. tibetana*; (B) Mature *H. tibetana* fruits on the tree; (C) Circos diagram of the chromosome-level *H. tibetana* (red inner circle on the left) and *H. rhamnoides* (green inner circle on the right) genome feature. Chr: chromosome. Window size: 1 Mb with non-overlapping. From outside to inside: i, percent of total *Gypsy* (0–100%); ii, percent of total *Copia* (0–100%); iii, intact *Gypsy* number (0–8); iv, intact *Copia* number (0–20); v, gene count (0–100); vi, GC content (0–100%). The gray line in the circle is the collinearity area between species, and the color line represents the fragments of 4 copies within species.

were retained, representing 68.47× genome coverage. Three 20 kb insert circular consensus sequencing (CCS) libraries were constructed and sequenced using a PacBio Sequel II platform for HiFi (high-fidelity) sequencing. The Hi-C library was constructed according to a published protocol (Durand et al., 2016). In brief, 2 g of young leaves was cross-linked *in situ* in 1% formaldehyde solution. Chromatin was extracted and digested with MboI (New England Biolabs), and the DNA ends were labeled, biotinylated, diluted, and randomly ligated. The DNA fragments were enriched and quality-checked to ensure that they were suitable for library preparation. Three sequencing libraries were constructed and sequenced on the MGISEQ-2000 platform in 150 bp paired-end mode. Transcriptome sequences were generated from a pool of mixed tissues including roots, stems and spear leaves. Raw polymerase reads were produced by PacBio Sequel platform, and then processed by IsoSeq technology through SMRT method (<https://www.pacb.com/smrt-science/smrt-sequencing/>) to obtain full-length non-chimeric reads.

Genome assembly

Before *de novo* assembly, Jellyfish (Marçais and Kingsford, 2011) and Genomescope v.1.0 (Vurture et al., 2017) were employed to calculate the frequency of K-mer ($k = 17$) based on short reads data, and the genome size was estimated using a method based on K-mer distribution (Marçais and Kingsford,

2011). Contigs from CCS clean reads were assembled by Hifiasm (v0.15.1) (Cheng et al., 2021) with default parameters. Furthermore, PurgeDups (v.1.2.3) (Guan et al., 2020; https://github.com/dfguan/purge_dups) was used to filter redundancy and generate the final haploid assembly. Scaffolding was performed using Hi-C based information, in which Hi-C reads were aligned to the assembled draft genome by Juicebox v1.6 (Durand et al., 2016) and contigs were mapped onto chromosome-level scaffolds by 3D-DNA (Dudchenko et al., 2017) under default parameters. Manual checking and refinement of the draft assembly were carried out *via* Juicebox Assembly Tools (<https://github.com/aidenlab/Juicebox>, v1.1108). Aligned with the embryophyta_odb10 database ($n = 1614$), the completeness of the final assembly was evaluated using BUSCO v5.1.2 (Simao et al., 2015) with default parameters.

Annotation of repetitive elements

Repetitive elements within the *H. tibetana* genome were identified using a combination of homology-based and *de novo*-based approaches. For homology-based annotation, RepeatMasker v4.0.7 (-nolow -no_is -norna -engine ncbi -parallel 1) (Chen, 2004) and RepeatProteinMask v4.0.7 (-engine ncbi -noLowSimple -pvalue 0.0001) (Bergman and Quesneville, 2007) were used to align the genome sequences against the Repbase database v21.12 (Jurka et al., 2005) to

identify and classify different repetitive elements. For *de novo*-based prediction, RepeatModeler v1.0.8 (<http://repeatmasker.org/RepeatModeler/>) and LTR Finder (v1.0.6) (Xu & Wang, 2007) with default parameters was firstly used to construct the library and Repeatmasker was used for predicting repetitive elements. In addition, tandem repeats were predicted using Tandem Repeats Finder v4.09 (Price et al., 2005). The intact LTR-RTs in the *H. tibetana* ($\mu=7.06 \times 10^{-9}$ per site per generation), *Hippophae rhamnoides* ($\mu=7.15 \times 10^{-9}$ per site per generation), and *Ziziphus jujuba* ($\mu=6.20 \times 10^{-9}$ per site per generation) genome were identified by TESorter and the parameters are as follows: -db rexdb-plant -p 20 -cov 20 -eval 0.001. We used protein for the alignment and phylogenetic analysis and the timing of their insertion was estimated using LTR_retriever (Ou and Jiang, 2018). Phylogenetic analyses were carried out focusing on proteins of the reverse transcriptase domains of both Ty1-*Copia* and Ty3-*Gypsy* LTR-RTs for both the *Copia* and *Gypsy* family, the three species reverse transcriptase domains were merged and ~1000 paralogs were randomly extracted. The ~1000 multiple sequences alignments were carried out using MUSCLE v3.7 (Edgar, 2004). Phylogeny construction based on the maximum-likelihood (ML) method using IQ-TREE (v1.6.12) (Nguyen et al., 2015), with the best-fit evolutionary substitution model was evaluated using ModelFinder (Kalyaanamoorthy et al., 2017).

Gene annotation

Gene annotations were performed by integrating evidence from homology-, *De novo*- and transcriptome-based information. In the homology-based approach, homologous proteins from *Arabidopsis thaliana*, *Oryza sativa*, *Cannabis sativa*, *Fragaria viridis*, *Z. jujuba*, *Pyrus pyrifolia*, were aligned against the *H. tibetana* genome using TBLASTN (Camacho et al., 2009), and the gene structure was predicted from these alignments by Exonerate v2.2.0 (Slater and Birney, 2005). *De novo* gene prediction was performed using a combination of AUGUSTUS (Stanke et al., 2006) and SNAP (Johnson et al., 2008) with default settings. The transcriptome data was aligned to the genome for predicting gene structure by BLAT. The tool MAKER2 (Holt and Yandell, 2011) was used to integrate the evidence for gene models.

Gene function annotation was done based on sequence similarity and domains conservation. First, the protein coding genes were annotated by homologous searches against public databases using BLAST v2.2.31 ($E\text{-value} < 1e^{-5}$), including SwissProt (Boeckmann et al., 2003), KEGG (Kanehisa et al., 2012), GO (Ashburner et al., 2000), TrEMBL (Bairoch and Apweiler, 2000), InterPro (Zdobnov and Apweiler, 2001) and NR (<https://www.ncbi.nlm.nih.gov/protein/>). Subsequently, the best

match from the alignment was used to represent the gene function by using some custom scripts. Second, we combined with application of InterProScan (51.0-55.0) (Jones et al., 2014) searching against the following databases to identify the motif and domain: Pfam (Bateman et al., 2000), PANTHER (Mi et al., 2017), SUPERFAMILY (Wilson et al., 2009), SMART (Schultz et al., 2000), PRINTS (Attwood et al., 2000) and ProDom (Corpet et al., 1999). The BUSCO was used to assess the completeness of the coding gene prediction for *H. tibetana* with default parameters. We mapped the gene density, GC content, fragmentary/intact *Gypsy* density, fragmentary/intact *Copia* density and chromosome synteny onto 12 chromosomes using the CIRCOS tool (<http://www.circos.ca>) (Krzywinski et al., 2009).

Gene family evolution analysis

The protein-coding genes of *H. tibetana* and 12 additional plants species were used to identify orthologous gene groups, including *Hippophae rhamnoides*, *Amborella trichopoda*, *A. thaliana*, *C. sativa*, *F. viridis*, *M. notabilis*, *O. sativa*, *Populus trichocarpa*, *Prunus persica*, *Rhamnella rubrinervis*, *Vitis vinifera*, and *Z. jujuba* were downloaded. To perform the gene family analysis, orthogroups of the 13 species were identified using OrthoFinder (v2.3.11) with default parameters (Emms and Kelly, 2019). The single-copy genes were used for further phylogenetic analysis. For a certain gene-family in one species (paralog), it specially exists in this species. Other species have no gene clustered in the gene-family. The single-copy orthologs among the 13 species were aligned using MUSCLE v3.7 (Edgar, 2004) with default parameters and then the aligned protein sequences were reversely translated into codon sequences. The alignments were then joined into a super alignment matrix for phylogeny construction based on the maximum-likelihood (ML) method using IQ-TREE (v1.6.12) (Nguyen et al., 2015), with the best-fit evolutionary substitution model was evaluated using ModelFinder (Kalyaanamoorthy et al., 2017). Divergence time for each node in the phylogenetic tree was estimated with using MCMCtree implemented in PAML package v4.8a (Yang, 2007) with the following parameters: -nsample 100000, -burnin 500000. The time correction points were obtained from TimeTree database (<http://www.timetree.org>) (Kumar et al., 2017): 56 ~ 93 million years ago (Ma) for *R. rubrinervis* and *Z. jujuba*, 107 ~ 135 million years ago for *Z. jujuba* and *V. vinifera*, 1.73 ~ 1.99 million years ago for *V. vinifera* and *A. trichopoda*.

The time-calibrated phylogenetic tree was used for accessing gene family expansions and contractions by the CAFÉ 4.2.1 program (Han et al., 2013), using a random birth-and-death model with lambda potio. The corresponding *P*-value in each lineage was calculated based on the conditional likelihood method and filtered threshold of *P*-value is 0.05.

Genome collinearity analysis and WGD event identification

MCscan (Python version) (Tang et al., 2008) was used for the genomic analysis between *H. tibetana*, *H. rhamnoides*, *V. vinifera*, *Z. jujuba*, and *A. trichopoda*. The collinearity figure was drawn based on the gene collinear pair information between species by JCVI (<https://github.com/tanghaibao/jcvi>) or by Circos (Krzywinski et al., 2009). We first characterized the synonymous nucleotide substitutions on synonymous substitution sites (K_s) between the collinear genes inferred above. The values of K_s was estimated for each collinearity orthologous using the *WGDI* (Sun et al., 2021) program with Nei-Gojobori method implemented in the YN00 program in the PAML (4.9h).

Divergent evolutionary rates among plants affect dating ancestral events. Here, based on an approach that previous report developed (Wang et al., 2015; Wang et al., 2018), we performed evolutionary rate correction by aligning the peaks in different genomes corresponding to the core-eudicot common hexaploidization (ECH) event which was ~115–130 Ma (Jiao et al., 2012; Vekemans et al., 2012) to the same locations.

For evolutionary rate correction, under the assumption that the *H. tibetana* peak appears at k_H and the *V. vinifera* peak appears at k_V , we can use the equation

$$r = (k_H - k_V)/k_V$$

to describe the relative evolutionary rate of *H. tibetana*. Then, rate correction was performed to discover the corrected rate $k_{H\text{ correction}}$ of *V. vinifera* relative to k_V :

For the K_s between duplicates in *H. tibetana*, we defined the correction coefficient W_H as

$$k_{H\text{ correction}}/k_H = k_V/k_H = W_H$$

thus, we obtained

$$\begin{aligned} k_{H\text{ correction}} &= k_V/k_H \times k_H = 1/(1 + r) \times k_H \text{ and } W_H \\ &= 1/(1 + r) \end{aligned}$$

(2) For the K_s between homologous genes from *H. tibetana* and *V. vinifera*, if the peak was located at K_{H-V} , supposing the correction coefficient W_H in *H. tibetana*, we then calculated a corrected evolutionary rate

$$k_{H-V\text{ correction}} = W_H \times k_{H-V}$$

The different substitution rates of different species were calculated by the K_s analyse above. The equation is:

$$r = K/2T$$

where r is the substitution rate; K is the original observation value in K_s distribution reflecting the core-eudicot common

hexaploidization; and T is the occurrence time of the core-eudicot common hexaploidization even.

Adaptive selection in the genome

The values of K_a and K_s and the K_a/K_s ratio were estimated for each collinearity orthologous genes using the *WGDI* (Sun et al., 2021) program with Nei-Gojobori method implemented in the YN00 program in the PAML (4.9h). We got the K_a/K_s result for the two groups (group 1: *H. tibetana* vs *R. rubrinervis* and *Z. jujuba* vs *R. rubrinervis*; group 2: *H. tibetana* vs *Z. jujuba* and *H. rhamnoides* vs *Z. jujuba*). The GO functions of *H. tibetana*, *H. rhamnoides* and *Z. jujuba* protein-coding genes were annotated by homologous searches against GO (Ashburner et al., 2000) public databases using BLAST v2.2.31 (E-value < $1e^{-5}$). The binomial test (Watanabe and Hattori, 2006; Qiu et al., 2012) was used to identify GO categories with more than 20 orthologs that had an excess of nonsynonymous changes in either *H. tibetana* or *Z. jujuba* (or *Hippophae rhamnoides*) lineages.

Positive selection is biologically important in adaptations to the environment. To identify positive selections in the evolution of *H. tibetana*, orthologous genes from *H. tibetana*, *Hippophae rhamnoides*, *R. rubrinervis*, *Z. jujuba* and *C. sativa* were identified using OrthoFinder (v2.3.11). The single-copy genes (3,218 groups) were used for aligning by using MUSCLE v3.7 (Edgar, 2004) respectively. The genes under positive selection were estimated using the branch-site model of Codeml program (Codon Freq = 2) implemented in the PAML package v4.8a, with *H. tibetana* as the foreground branch and the other four species as background branches. Likelihood ratio test was used to identify positively selected genes (PSGs) (Likelihood ratio test, $P \leq 0.05$).

SVs identification

We detected the SVs of *H. tibetana* and *H. rhamnoides* by using SyRI (Goel et al., 2019). Firstly, the genomes (in multi-fasta format) are aligned using the NUCmer utility v 4.0.0 (-c 500 -l 40 -g 200 -t 10). Secondly, SyRI takes genome alignments coordinates as input to identify SVs (-k -nc 20 -c cords.file -d delta.file -q H. tibetana.fa -r H. rhamnoides.fa -s show-snp).

Nodule inception gene family evolutionary analysis

NODULE INCEPTION (*NIN*) gene family members were identified by BLAST searching homologous sequences in *H. tibetana*, *Z. jujuba* and *H. rhamnoides* genomes by setting the

Medicago NIN (MTR_5g099060) protein as the query sequence (E-value: $1e^{-5}$). Full-length NIN were used to build a ML-based phylogenetic tree using RAxML 7.0.4 software (Stamatakis, 2006) to construct the ML phylogenetic tree (ML-Tree).

To inspect the protein domain differences between NINs and NLPs, NIN/NLP proteins from all species were subjected to perform protein motif analysis with meme suit (Bailey et al., 2009) using the following parameters: *-nmotifs 20*, *-mod zoops*, *-minw 6* and *-maxw 50*, and the results (.xml file) were visualized using TBtools (Chen et al., 2020). For this analysis, only seed plants were used to avoid the extreme sequence variation caused by a long divergence time.

Aligning sex-specific regions to the homogametic sex data sets

A total of 80 *H. tibetana* samples were sampled from the four sites (Mt. Everest, Tibet; Datong, Qinghai; Maizhokunggar, Tibet; Xiahe, Gansu), including 10 male and 10 female plants in each site for RAD seq. The geographic distance between the four sampling sites was between 273 and 1683 km. During the flowering season (June in 2013), female and male individuals were identified by gynoceium and staminate flowers.

A mean of 1.89 Gb raw data was generated for each individual, the sequencing quality of Q20 and Q30 was 95% and 87% respectively. GC content was a little vibrate because of a small amount of chloroplast contamination. After quality control, a total of 110,619 Mb clean sequence reads were acquired from 150 bp paired-end sequencing (Insertsize: 500 bp) on the Illumina Hiseq 2000 platform (Illumina, San Diego, CA, USA), with an average of 1,382 Mb for each individual (minimum = 567,569; maximum = 15,706,831). The SNPs with MAF ≥ 0.05 and missing rate ≤ 0.2 and without imputation were used for GWAS. The Mixed linear model (MLM) analysis was conducted for GWAS using GAPIT software (Lipka et al., 2012). To avoid the false positive, the population structure (Q) and a polygene (K) were applied in GWAS. The total number of bi-allelic single nucleotide polymorphism was 2,486,852, and 2,072,124 SNPs were left after filtering. *H. tibetana* is known to have an XX/XY sex-determining mechanism. All putative sex-specific markers that passed this confirmation step were assembled into Contigs (RAD loci) with paired-end reads using Sequencher software (<http://www.genecodes.com/>). We further validated putative sex-linked RAD loci. First, we searched the female *H. tibetana* genome for putative male-specific RAD loci using BLAT on the UCSC Genome Browser (Kent, 2002; Meyer et al., 2013) and excluded any significant

matches. Then, BLAST was used to search the NCBI nucleotide database for matches to exclude contamination.

Results

Genome assembly and quality assessment

A *H. tibetana* ($2n = 2x = 24$) wild individual in Litang County, Sichuan (China) was used for whole-genome sequencing. Totals of 152 Gb of clean Illumina short-read data ($\sim 100\times$ coverage), 98.05 Gb of PacBio HiFi reads data ($68.47\times$ coverage), and 228 Gb of BGI-sequenced Hi-C data ($\sim 159\times$ coverage) were generated. A total of 531.5 million unique mapped Hi-C PE reads was detected and the Hi-C effective data was accounted for 25.81%. After obtained the assembly from Hifiasm (v0.15.1) and PurgeDups (v.1.2.3), contigs were clustered, sorted, orientated, and finally assembled onto 12 chromosomes with genome size of 1,452.75 Mb (96.04% of the final assembled 1,512.72 Mb genome) and heat maps are drawn accordingly (Figure S2). This genome assembly size was comparable to the size estimated by a K-mer-based method (1,432.13 Mb) (Figure S1). The assembly genome size is consisting with the DNA content of its related species (2.32-3.88 pg/C) that was previously determined with tissue using flow cytometry (Zhou et al., 2012) and is much larger than other Elaeagnaceae species (Mao et al., 2021; Wu et al., 2022; Yu et al., 2022). The Contig N50 and N90 of assembly genome reaches 74.31 Mb and 24.93 Mb and the scaffold N50 and N90 reaches 123.42 Mb and 71 Mb respectively (Table 1). Genome quality assessment was evaluated by Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simao et al., 2015), and revealed that the genome completeness reached 95.5%, a total of 1,542 expected embryophyta genes were identified in *H. tibetana* genome (Table S1), suggesting that the genome assembly is of high quality.

Combining *De novo*-, transcriptome- and homology-based predictions, 33,367 protein-coding genes were predicted from the genome overall. The number of gene models is comparable to those of other Elaeagnaceae species (Mao et al., 2021; Wu et al., 2022; Yu et al., 2022). Among all the predicted genes, 32,059 genes could be annotated to the functional database, see Table S2. The intersection of functional annotations among the five databases is shown in Figure S3. There are 21,081 genes with functional annotations in these five databases. The BUSCO score indicates that 95.0% of complete core orthologs were identified, suggesting that the gene annotation is of high quality. Otherwise, 15,143 non-coding RNA genes were identified in the genome, including 6,823 rRNAs, 5,221 tRNAs, 164 miRNAs, and 2,935 snRNAs (Table S3). An overview of the distribution of the predicted gene models in the genome is shown in Figure 1C.

TABLE 1 Statistics of the genome assembly and annotation.

Genome Features	<i>H. tibetana</i>	<i>H. rhamnoides</i> ssp. <i>sinensis</i>	<i>H. rhamnoides</i> ssp. <i>mongolia</i>
Assembly strategy	PacBio CCS +Hi-C	PacBio CLR+ Hi-C	PacBio CLR+ Hi-C
Estimated size (Mb)	1,432	749	978
Assembly size (Mb)	1,452.75	730.46	849.04
Number of Contigs	1,409	1,386	3,642
Maximum Contig length (Mb)	123.41	82.44	92.33
Contig N50 (Mb)	74.31	2.81	2.15
Contig N90 (Mb)	26.14	0.373	–
GC content (%)	29.37%	30.07%	30.07%
Repeat content (%)	72.18%	47.48%	67.81%
Number of protein-coding genes	33,367	30,812	30,864
Complete BUSCOs	95.54%	97.60%	89.65%

Genome annotation and recent burst of LTR-RT

After removing the overlapping regions between various methods, the remaining nonredundant repeated sequences accounted for 72.18% of the genome size (Table S4), and transposable elements (TEs) were the dominant components (69.52%, Table S5). The TEs including DNA transposon elements (DNA, 2.77%), long interspersed nuclear elements (LINEs, 0.54%), short interspersed nuclear elements (SINEs, 0.03%), long terminal repeats (LTRs, 64.85%), and elements of unknown classifications (2.05%) (Table S5).

TE contents are highly variable within eukaryotes and generally comprise the important part of plant genome, and the summation of TE show a positive correlation with plant genome size (Wendel et al., 2016). LTR-RTs possessed the most proportion (64.85%) of repetitive elements in *H. tibetana*, and a total of 6,911 intact LTR-RTs in *H. tibetana*, 4,601 in *H. rhamnoides* and 3,004 in *Z. jujuba* genome were identified. Meanwhile, by using the substitution rate values calculated by the Ks analyse, both insertion time of *H. rhamnoides* (~1 Ma) and *H. tibetana* (~1 Ma) were more recent compared with *Z. jujuba* (~1.5 Ma) (Figure 2C).

The Ty3/Gypsy and Ty1/Copia elements were the two main types of LTR in *H. tibetana*, accounting for 25.39% and 19.33%, respectively (Figure 2D; Table S5). Phylogenetic analysis of intact Ty3/Gypsy and Ty1/Copia supergroups showed that the branches of *H. tibetana* were short and clustered distribution, suggesting that recent burst of Ty1/Copia and Ty3/Gypsy, indicating that there was a recent TE amplification of *H. tibetana* (Table S6; Figures 2D, S4). Intact Gypsy/Copia is closer to genes than fragmentary Gypsy/Copia (Figure 2E; chi-square test, $P < 0.05$). Surprisingly, we identified a DNA transposon gene family with as many as 141 members in the *H. tibetana* genome (Figure S5). *H. tibetana* is almost twice the genome size of *H. rhamnoides* (Table 1), but the number of gene models and the total length of exons and introns are comparable to *H. tibetana* (Figure S6). The primary cause is that the length

of intergenic regions in *H. tibetana* genome was significantly greater than that in *H. rhamnoides* genome and the majority of TEs were located in those (Figure 2D). *H. tibetana* is consistent with this pattern, and possessed both large genome size and larger proportion of TEs. The recent proliferation of LTR-RTs in the genome of *H. tibetana* may promote its genome evolution and play an important role in adapting to high-altitude habitats alone after its differentiation with *H. rhamnoides*.

Gene family evolution and phylogenetic analysis

After analysis of the gene family, 241,829 genes from the 13 species were grouped into 22,503 gene families (Figure S7). Venn diagram depicting the number of shared and specific gene families among *H. tibetana* and three representative plants (*Z. jujuba*, *V. vinifera* and *H. rhamnoides*). We identify 10,044 homologous gene families shared by these 4 species, and 1,319 gene families were specific to *H. tibetana* and *H. rhamnoides*, and 713 were found only in *H. tibetana* (Figure S8). *H. tibetana* shows more genes in common with *Z. jujuba*, *R. rubrinervis*, and *C. sativa* than with other species (Table S7). Among the orthologous genes, 581 genes were identified as single-copy genes in these species, which were used for inferring the evolutionary relationships. As shown in the phylogenetic (Figure 2A), *A. thaliana*, *P. trichocarpa*, *V. vinifera*, *O. sativa*, and *A. trichopoda* diverged from one another earlier than *M. notabilis*, *C. sativa*, *P. persica*, and *F. viridis*, diverged from each other, and *Z. jujuba*, together with *R. rubrinervis* cluster as a sister is most closely related to *H. tibetana*. According to the fossil record information, *H. tibetana* separated from the Rhamnaceae approximately 83.7 Ma (Figure 2A). The diverged time of Elaeagnaceae and other Rosales species at the beginning of the Cenozoic era. According to previous studies and reports, the rapid diversification of species in this family only occurred in the Miocene which was less than 23 Ma (Jia and Bartish, 2018).

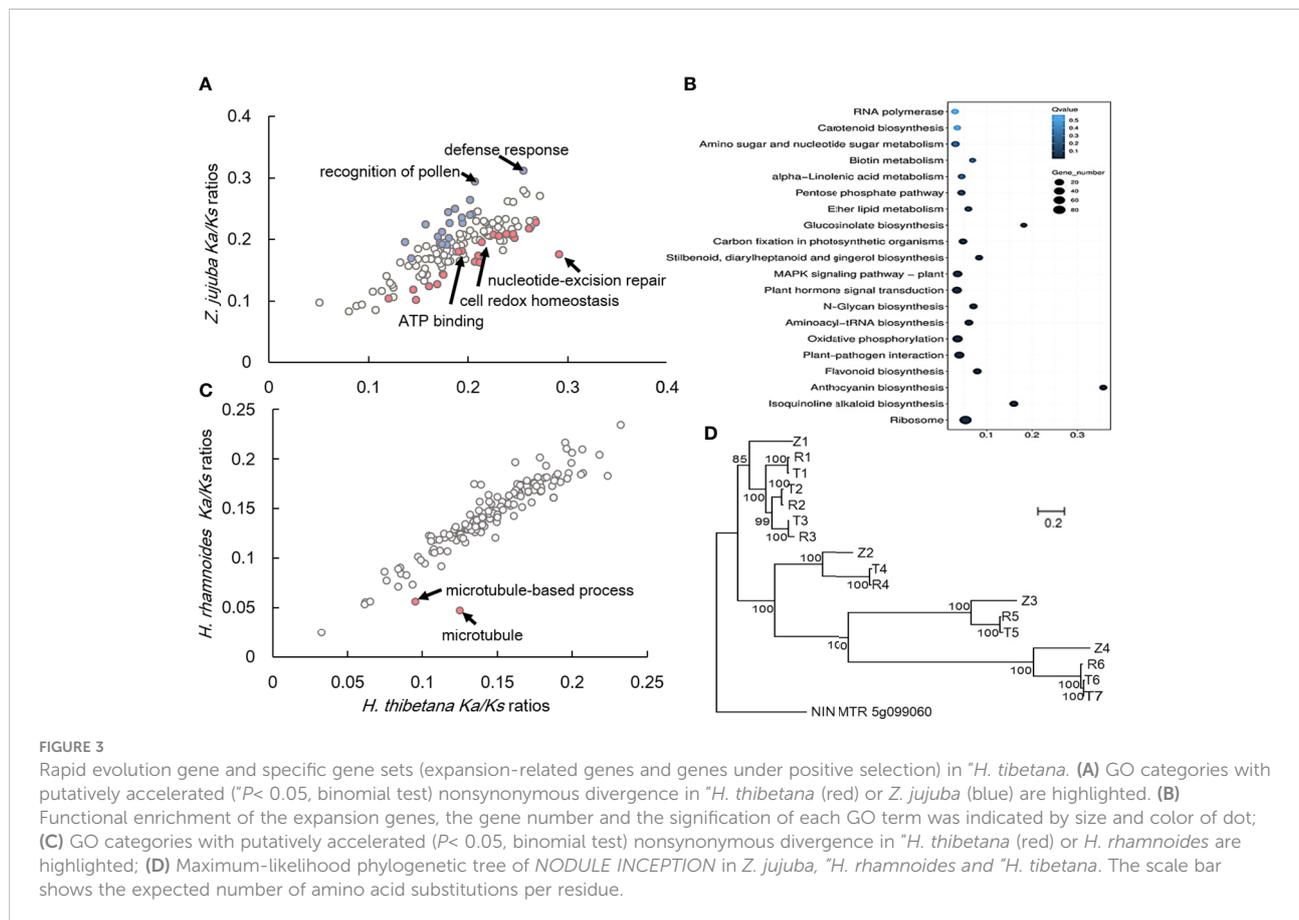
rhamnoides and *H. tibetana*, a total of identified 1,713 structural variants (SVs) was identified (Figure S12), and large chromosomal inversion were detected on Chr1, Chr2, Chr6, Chr11, and Chr12.

Genome evolution in adaptation and survival

As sessile organisms live in high altitude area, *H. tibetana* have to cope with recurring stress such as frequent low nocturnal temperatures, intense solar ultraviolet, low carbon dioxide concentration. To clarify the evolutionary adaptation of *H. tibetana*, the time-calibrated phylogenetic tree was used for accessing gene family expansions and contractions by the CAFÉ 4.2.1 program (Han et al., 2013). There were 186 and 94 gene families (1,113 genes Gain and 870 genes loss) were significantly expanded and contracted, respectively ($P < 0.05$; Figures 2A, 3B), which indicated that there were more *H. tibetana* gene families experienced expansion rather than contraction during adaptive evolution. After GO enrichment analyses, genes were enriched in “oxidoreductase activity (GO: 0016491),” “cellulose biosynthetic process (GO: 0030244),” and “protein phosphorylation (GO: 0006468)”.

After filtering out organelle genes, 3,101 genes of the *H. tibetana* specific gene families were identified, and Gene Ontology annotation terms mainly containing “response to ultraviolet (UV) radiation,” “Plant-pathogen interaction,” “MAPK signaling pathway – plant,” “Phenylpropanoid biosynthesis,” “Flavonoid biosynthesis,” “Phenylalanine metabolism,” and “Nitrogen metabolism”.

Ka/Ks ratios of nonsynonymous-to-synonymous substitutions for different GO categories revealed an enrichment of elevated pairwise Ka/Ks values in the high-altitude adaptation (Qiu et al., 2012). Analysis of Ka/Ks ratios in the lineages verified that genes with elevated Ka/Ks values in *H. tibetana* were significantly ($P < 0.05$, binomial test) enriched for high altitude adaptation functions (Table S8; Figures 3A, C). Most of genes were enrich in “nucleotide-excision repair,” “cell redox homeostasis,” and “ATP binding”. On the contrary, GO categories with putatively accelerated nonsynonymous divergence in *Z. jujuba* (blue) were enrichment in “recognition of pollen” and “defense response”. After that, we detected genes evolving under elevated Ka/Ks values in either *H. tibetana* or *H. rhamnoides*. Only “microtubule-based process (GO: 0007017, process)” and “microtubule (GO: 0005874, component)” were identified as rapidly evolving (with elevated Ka/Ks values in *H. tibetana*) genes.



Furthermore, the positively selected genes compared with close relatives in the plant genome are usually considered to be related to adaptability (Fitch, 1970). To test the hypothesis that these rapidly evolving genes in *H. tibetana* have been under positive selection, we used the branch-site likelihood ratio test to identify positively selected genes (PSGs) in the *H. tibetana* lineages. In *H. tibetana*, a total of 249 genes were identified as positive selection genes by comparative analysis ($P < 0.05$), after GO enrichment analyses (Table S9; Figure S13), most of genes were enrichment in “protein binding (GO: 0005515),” “cellular response to DNA damage stimulus (GO: 0006974),” “negative regulation of defense response (GO: 0031348),” “DNA repair (GO: 0006281),” “carbohydrate metabolic process (GO: 0030246)”. These processes are related to the adaptation to the extreme environmental conditions of high ultraviolet and low temperature in the high-altitude area of the Qinghai-Tibet Plateau.

Nodule inception gene family evolution of *H. tibetana*

Symbiotic nitrogen fixation provides a large amount of sustainable and environmentally friendly nitrogen for plants. Comparative analysis of NIN-orthologous regions within related species (*Z. jujuba*, *H. rhamnoides*, and *H. tibetana*). Three species exhibited different copy numbers within these regions; *Z. jujuba* exhibited 4, *H. rhamnoides* 6, and *H. tibetana* 7 (Figure 3D). Also, because *H. tibetana* is a pioneer species, it can survive on the bare ground where other plants cannot colonize. We deduced that this additional nitrogen-fixation-related gene was helpful for *H. tibetana* to colonize bare land first.

Identification of sex-linked regions

Screening of male-specific markers by comparative analysis of RAD-seq sequencing of 40 male and 40 female samples of known gender (distinguished according to morphology in reproductive season (Figure 1B; Table S10). RAD-seq paired-end reads was detected SNPs by *Stacks*1.04 (Catchen et al., 2013). Genome-Wide Association Studies (GWAS) among 40 males and 40 females was performed and showed that the regions on Chr2 were related to sex (Figure 4A), which also supported by correlation between the assembly lengths and observed physical lengths of all chromosomes (Figure 4B) (Xing et al., 1989).

Discussion

H. tibetana can live in extreme environments at 5,200 meters a.s.l. and is one of the highest distributed woody plants. Here, a chromosome-level genome of wild *H. tibetana* was assembled

and annotated with contig N50 sizes of 74.31 Mb, which possessed high completeness, accuracy, and complexity. This is also the highest known woody plant genome (up to 5,200 meters above sea level). Although it has the same chromosome number with two related species (*Z. jujuba* and *H. rhamnoides*), the genome size of *H. tibetana* is 1,452.75 Mb and 41.56% larger than that of *H. rhamnoides* (Yu et al., 2022), and 69.87% larger than that of *Z. jujuba* (Huang et al., 2016). The comparison of *H. tibetana* and these related species revealed some new findings.

Studies of high-altitude plant genomes have revealed that some genes with special adaptive significances have been affected by positive selection and undergone rapid evolution. Especially, the genes involved with DNA damage repair, reproductive processes, and UV-B tolerance (Mao et al., 2021). We also found some interesting genes to undergo a rapid/slow evolution by comparing *H. tibetana* with *Z. jujuba*, which altitude is less than 1,700 meters (Chen and Chou, 1982). The expansion or positively selected genes involved in cellular response to DNA damage stimulus, Smc5-Smc6 complex, SUMO ligase complex and nitrogen compound metabolic process. These genes involved in adaptation to high altitude habitats have also been identified in previous studies (Zhang et al., 2016; Geng et al., 2021; Mao et al., 2021).

It is worth noting that compared with jujube, two rapidly evolving genes in jujube were found, which were enrichment in “recognition of pollen” and “defense response”. Since *H. tibetana* is a pioneer species, it can survive on bare land where few other plants survive. Compared with its relatives, it has fewer pollens from other competitors, so “recognition of pollen” genes rapidly evolving should be adaptive in jujube. In addition to the above genes, there are some genes that evolve rapidly relative to jujube were enrich in “nucleotide-excision repair,” “cell redox homeostasis,” and “ATP binding” (Figure 3A), suggesting that high-altitude habitats also have a significant impact on the basic life activities of organisms.

However, when comparing *H. tibetana* with *H. rhamnoides*, none of the genes were subject to significant positive or negative selection except microtubule-based process (GO: 0007017, process) and microtubule (GO: 0005874, component) (Figure 3C). Microtubule and microtubule-based process genes were considered to be related to plant growth and development (Hashimoto, 2003). We guess that they are likely to be the key genes that determine the evolution from tree to shrub to adapt to higher altitudes. We notice that the average altitude of *H. tibetana* is higher ~1,500 meters than that of *H. rhamnoides* (Huang and Yu, 2006). The fact that only two GO categories have undergone fast/slow evolution is interesting. These two species live in such different environments that it is hard to believe that only two GO categories are subject to significant positive selection. This could mean that the forces of natural selection may be acting more on other aspects of the genome than changing the amino acid sequence of proteins, such as gene number, gene regulatory regions, etc.

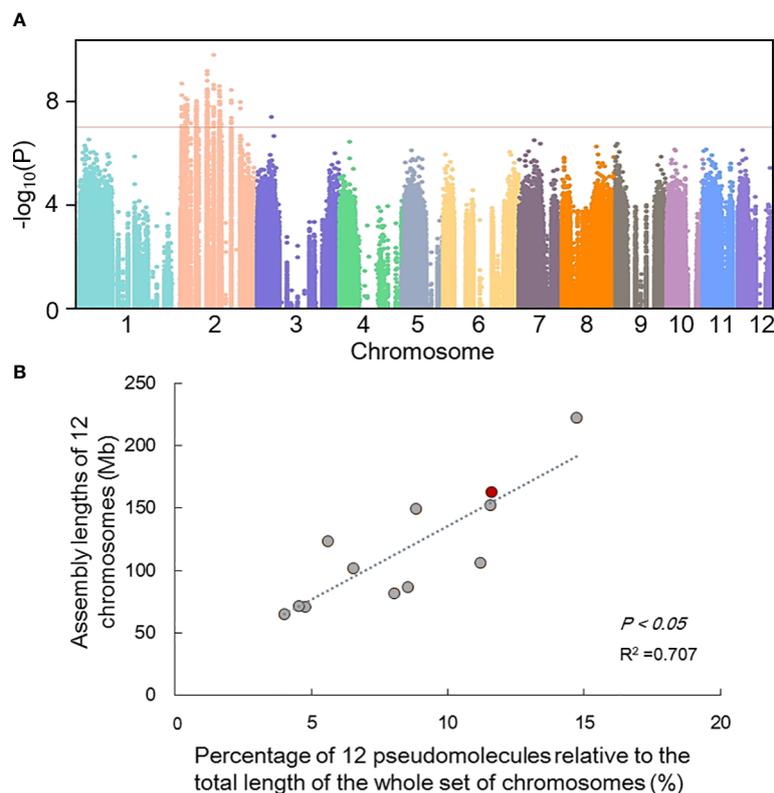


FIGURE 4

Preliminary inference of sex-linked regions. (A) Manhattan depicting significant SNPs identified using MLM model that showed association with gender. The red dotted line represents truncation criterion was set to $-\log_{10} P\text{-value}=7$; (B) The correlation between the assembly lengths and observed physical lengths of all chromosomes. The red dot represents chromosome 2 of *H. tibetana*.

The significant fluctuation of gene family copy number is usually related to the adaptive evolution of species (Sudmant et al., 2010). We noticed that the genome of *H. tibetana* contains 2503-2555 more genes than that of *H. rhamnoides*. Some of these genes were from single gene duplication, and most genes were originated from WGD. A special gene doubled in the former way was found in *H. tibetana* (Figure 3D). It was related to nitrogen fixation as the previous study found in the *H. rhamnoides* (Soyano et al., 2013; Liu and Bisseling, 2020). We found that in the *H. tibetana*, the key nitrogen fixation gene increased by one copy through tandem duplication, and that at least at the seedling stage, the transcript level these two genes in *H. tibetana* was 2.43~2.53 times than that in *H. rhamnoides* (*H. tibetana*: T1-T7, *H. rhamnoides*: R1-R6; Figure 3D). As described above, the habitat of *H. tibetana* is very extreme and harsh, especially prefers bare land after glaciers retreat, such habitats are deficient in available nitrogen. We proposed that the genome evolved nitrogen fixation adaptively by increasing the copy number of genes.

Obviously, WGD is the source of gene increase (De Smet and Van de Peer, 2012), and is considered to be an important

way of speciation and adaptation to disturbed habitats (Soltis and Soltis, 1999; Jiao et al., 2011). Plants with known genomes on the Qinghai Tibet Plateau, eg: *Lepidium meyenii* and *Megacarpaea delavayi*, have experienced independent WGD (Zhang et al., 2016; Yang et al., 2020). In this study, we found that the *H. tibetana* genome experienced two WGD events at 18.7-21.2 Ma (HRT) and 28.6-32.4 Ma (HAT), respectively. Interestingly, after HAT and HRT, the genome of *H. tibetana* went through the process of chromosome breakage and fusion, and then the chromosome returned to its ancestral number ($2n=24$). As can be seen from Figure 2A, B, these events occurred before the differentiation of *H. tibetana* and *H. rhamnoides*, that is, they experienced these important processes together, and then went to different evolutionary paths respectively. We noted that the former has 2,503-2,555 genes more than the latter, and the genome size is 41.56%-49.72% larger than *H. rhamnoides*. Through functional enrichment of genes generated by WGD, it is found that these extra genes participate in the basic process of life activities. We speculate that this may be related to the drastic changes in the environment in a short time at a very high altitude, for example, in the area of 5,000 meters a.s.l., the

temperature difference in one day (24 hours) can reach 20°C (Zhang and Yang, 2013), which means that cells need greater biochemistry plasticity to respond. Studies have shown that more gene redundancy can provide greater plasticity (Mattenberger et al., 2017).

Compared with the *H. rhamnoides* genome, one of the most prominent characteristics of the *H. tibetana* genome is to contain more TE elements (Figure 2D), which is the main reason for the difference in size between the two genomes. TE outbreaks have also been widely found in high-altitude plant genomes, and it is believed that rapid proliferation of repetitive elements in *C. lasiocarpa* may play an important role in promoting its genome evolution (Feng et al., 2022). However, the relationship between high-altitude extreme habitats and TE still remains unclear. As a mobile element, the activity of transposable elements increases the variation of genome and can generate mutations rapidly (Kidwell and Lisch, 1997; Wessler, 2006), which is of great significance for alpine organisms to respond to environmental changes (Wos et al., 2021). Compared with the lower altitude area, alpine habitats are more susceptible to climate and other factors. On the Qinghai-Tibet Plateau, the greater impact used to happen in habitats with higher altitude, especially since the Quaternary, it has been affected by the repeated fluctuations of the glacial and interglacial periods (Shi, 2002; An et al., 2011; Yan et al., 2021). More variation in genome often means higher genetic diversity and therefore evolutionary potential. Although since 0.5 Ma, the effective population size of *H. tibetana* has continued to decline (Figure S14), this species has still widely distributed, perhaps due to previously emerging genomic adaptations.

The transposon proliferation time of *H. tibetana*, *H. rhamnoides*, and jujube is around 1–1.5 Ma (Figure 2C), suggesting that this is a parallel response stimulated by the environment, when the QTP enters freezing from this period circumstances and times of environmental upheaval (Shi et al., 1998; Zhou et al., 2011; Zhao et al., 2020). We noticed that the transposon size in the *H. tibetana* genome is the largest, following by *H. rhamnoides*, and then the species at the lowest altitude, *Z. jujuba*, has the smallest transposon size (or is less removal and more fragments remaining). How the activity of the TEs described above is linked to more dramatic environmental variability needs more works, and the activity of TEs can be induced by environmental and population factors and in particular by stresses in various organisms has been confirmed by many studies (Capy et al., 2000; Wos et al., 2021). Also, we detected a burst of DNA transposable elements that are only present within the *H. tibetana* (Figure S6). These results suggest that the extreme habitat of *H. tibetana* is the cause of it has more TEs.

By analyzing the distribution of transposons in the *H. tibetana* genome, we found that the distribution of intact TEs and fragmentary TEs is very different. On average, the former is more significantly closer to the gene than the latter (Figure 2E),

meaning that intact TE is more likely to insert genes or gene regulatory regions, thereby affecting gene function (Niu et al., 2022), which is a way of rapid evolution, and may also be one of the important ways for the *H. tibetana* genome to adapt to rapid variation habitats. Furthermore, as mentioned above, biochemical plasticity or gene expression is of extreme importance for high altitude organisms (Nicotra et al., 2010), and there are studies have shown that TEs can act as translators of phenotypic plasticity (Pimpinelli and Piacentini, 2020). But whether more TEs can result in higher plasticity is unclear. Thus, more TEs in the *H. tibetana* genome were both a cause of it adapting to extreme habitats and a consequence of it growing in extreme habitats.

From the perspective of a longer evolutionary history, the diploidization after tetraploidization is also a process through chromosome breakage, rearrangement and fusion, and studies have shown that these processes can change the expression and function of genes (Shi et al., 2015; Wang et al., 2022). These results suggest that the evolution of proteins is only a small part, and the genome responds more in other ways, including regulating the copy number of genes, changing regulatory patterns, etc.

Data availability statement

The original contributions presented in the study are publicly available. This data can be found here: NCBI, PRJNA796061 and China National GeneBank Database, CNP0003543.

Author contributions

LQ and ZW designed the research. RW, LQ and ZW collected the samples and performed the research. BW, JJ, ZS, TZ, and RW analyzed the data. RW wrote the paper. All authors read and approved the final manuscript.

Funding

This work was supported by National Natural Science Foundation of China (31760127), First-class discipline construction project of ecology (Zangcaijiaozhi [2019]01) and National Natural Science Foundation of China (31670225).

Acknowledgments

We would like to thank the late Prof. Yang Zhong, and dearly cherish the memory of our esteemed mentor, without whom this work would not have been possible.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations,

or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1051587/full#supplementary-material>

References

- An, Z., Clemens, S. C., Shen, J., Qiang, X., Jin, Z., Sun, Y., et al. (2011). Glacial-interglacial Indian summer monsoon dynamics. *Sci.* 333 (6043), 719–723. doi: 10.1126/science.1203752
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.* 25 (1), 25–29. doi: 10.1038/75556
- Attwood, T. K., Croning, M. D., Flower, D. R., Lewis, A. P., Mabey, J. E., Scordis, P., et al. (2000). PRINTS-s: the database formerly known as PRINTS. *Nucleic. Acids Res.* 28 (1), 225–227. doi: 10.1093/nar/28.1.225
- Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., et al. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic. Acids Res.* 37, 202–208. doi: 10.1093/nar/gkp335
- Bairoch, A., and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic. Acids Res.* 28 (1), 45–48. doi: 10.1093/nar/28.1.45
- Bateman, A., Birney, E., Durbin, R., Eddy, S. R., Howe, K. L., and Sonnhammer, E. L. (2000). The pfam protein families database. *Nucleic. Acids Res.* 28 (1), 263–266. doi: 10.1093/nar/28.1.263
- Bergman, C. M., and Quesneville, H. (2007). Discovering and detecting transposable elements in genome sequences. *Brief. Bioinform.* 8 (6), 382–392. doi: 10.1093/bib/bbm048
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic. Acids Res.* 31 (1), 365–370. doi: 10.1093/nar/gkg095
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC. Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421
- Capy, P., Gasperi, G., Biemont, C., and Bazin, C. (2000). Stress and transposable elements: co-evolution or useful parasites? *Heredity.* 85, 101–106. doi: 10.1046/j.1365-2540.2000.00751.x
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., and Cresko, W. A. (2013). Stacks: an analysis tool set for population genomics. *Mol. Ecol.* 22 (11), 3124–3140. doi: 10.1111/mec.12354
- Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform.* 5, 4. doi: 10.1002/0471250953.bi0410s05
- Chen, Y., Chen, Y., Shi, C., Huang, Z., Zhang, Y., Li, S., et al. (2018). SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaSci.* 7 (1), 1–6. doi: 10.1093/gigascience/gix120
- Chen, Y., and Chou, P. (1982). “Rhamnaceae,” in *Flora reipublicae popularis sinicae*. Ed. Y. Chen (Beijing: Science Press) 48(1), 133.
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved *de novo* assembly using phased assembly graphs with hifiasm. *Nat. Methods* 18 (2), 170–175. doi: 10.1038/s41592-020-01056-5
- Chen, J. H., Huang, Y., Brachi, B., Yun, Q. Z., Zhang, W., Lu, W., et al. (2019). Genome-wide analysis of cushion willow provides insights into alpine plant divergence in a biodiversity hotspot. *Nat. Commun.* 10 (1), 5230. doi: 10.1038/s41467-019-13128-y
- Corpet, F., Gouzy, J., and Kahn, D. (1999). Recent improvements of the ProDom database of protein domain families. *Nucleic. Acids Res.* 27 (1), 263–267. doi: 10.1093/nar/27.1.263
- Deng, T., and Ding, L. (2015). Paleotemperature reconstructions of the Tibetan plateau: progress and contradictions. *Natl. Sci. Rev.* 2 (4), 417–437. doi: 10.1093/nsr/nwv062
- De Smet, R., and Van de Peer, Y. (2012). Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* 15 (2), 168–176. doi: 10.1016/j.pbi.2012.01.003
- Doyle, J. J., and Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh tissue. *Phytochemical. Bulletin.* 19, 11–15.
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the aedes aegypti genome using Hi-c yields chromosome-length scaffolds. *Sci.* 356 (6333), 92–95. doi: 10.1126/science.aal3327
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-c experiments. *Cell. Syst.* 3 (1), 95–98. doi: 10.1016/j.cels.2016.07.002
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic. Acids Res.* 32 (5), 1792–1797. doi: 10.1093/nar/gkh340
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome. Biol.* 20 (1), 238. doi: 10.1186/s13059-019-1832-y
- Favre, A., Packert, M., Pauls, S. U., Jahnig, S. C., Uhl, D., Muellner-Riehl, A. N., et al. (2015). The role of the uplift of the qinghai-Tibetan plateau for the evolution of Tibetan biotas. *Biol. Rev.* 90 (1), 236–253. doi: 10.1111/brv.12107
- Feng, L., Lin, H., Kang, M., Ren, Y., Yu, X., Xu, Z., et al. (2022). A chromosome-level genome assembly of an alpine plant crucifera lasiocarpa provides insights into high-altitude adaptation. *DNA. Res.* 29 (1), dsac004. doi: 10.1093/dnares/dsac004
- Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Systematic. Zoology.* 19 (2), 99–113. doi: 10.2307/2412448
- Geng, Y., Guan, Y., Qiong, L., Lu, S., An, M., Crabbe, M. J.C., et al. (2021). Genomic analysis of field pennycress (*Thlaspi arvense*) provides insights into mechanisms of adaptation to high elevation. *BMC. Biol.* 19 (1), 143. doi: 10.1186/s12915-021-01079-0
- Goel, M., Sun, H., Jiao, W. B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome. Biol.* 20 (1), 277. doi: 10.1186/s13059-019-1911-0
- Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 36 (9), 2896–2898. doi: 10.1093/bioinformatics/btaa025
- Han, M. V., Thomas, G. W., Lugo-Martinez, J., and Hahn, M. W. (2013). Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* 30 (8), 1987–1997. doi: 10.1093/molbev/mst100
- Hashimoto, T. (2003). Dynamics and regulation of plant interphase microtubules: a comparative view. *Curr. Opin. Plant Biol.* 6 (6), 568–576. doi: 10.1016/j.pbi.2003.09.011
- Holt, C., and Yandell, M. (2011). MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC. Bioinf.* 12, 491. doi: 10.1186/1471-2105-12-491
- Huang, Q., and Yu, Z. (2006). *Research of seabuckthorn* (Beijing: Science Press).
- Huang, J., Zhang, C., Zhao, X., Fei, Z., Wan, K., Zhang, Z., et al. (2016). The jujube genome provides insights into genome evolution and the domestication of

- sweetness/acidity taste in fruit trees. *PLoS. Genet.* 12 (12), e1006433. doi: 10.1371/journal.pgen.1006433
- Jia, D. R., and Bartish, I. V. (2018). Climatic changes and orogeneses in the late miocene of Eurasia: The main triggers of an expansion at a continental scale? *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01400
- Jiao, Y., Leebens-Mack, J., Ayyampalayam, S., Bowers, J. E., McKain, M. R., McNeal, J., et al. (2012). A genome triplication associated with early diversification of the core eudicots. *Genome. Biol.* 13 (1), R3. doi: 10.1186/gb-2012-13-1-r3
- Jiao, Y., Wickett, N. J., Ayyampalayam, S., Chanderbali, A. S., Landherr, L., Ralph, P. E., et al. (2011). Ancestral polyploidy in seed plants and angiosperms. *Nature.* 473 (7345), 97–100. doi: 10.1038/nature09916
- Johnson, A. D., Handsaker, R. E., Pulit, S. L., Nizzari, M. M., O'Donnell, C. J., de Bakker, P. I., et al. (2008). SNP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics.* 24 (24), 2938–2939. doi: 10.1093/bioinformatics/btn564
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome scale protein function classification. *Bioinformatics.* 30 (9), 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., Walichiewicz, J., et al. (2005). Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome. Res.* 110 (1–4), 462–467. doi: 10.1159/000084979
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14 (6), 587–589. doi: 10.1038/nmeth.4285
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic. Acids Res.* 40 (D1), D109–D114. doi: 10.1093/nar/gkr988
- Kent, W. J. (2002). BLAT - the BLAST-like alignment tool. *Genome. Res.* 12 (4), 656–664. doi: 10.1101/gr.229202
- Kidwell, M. G., and Lisch, D. (1997). Transposable elements as sources of variation in animals and plants. *P. Natl. Acad. Sci. U.S.A.* 94 (15), 7704–7711. doi: 10.1073/pnas.94.15.7704
- Körner, C. (2003). *Alpine plant life. 2nd ed* (Berlin, Germany: Springer).
- Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome. Res.* 19 (9), 1639–1645. doi: 10.1101/gr.092759.109
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S. B. (2017). TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* 34 (7), 1812–1819. doi: 10.1093/molbev/msx116
- Lian, Y. S., Chen, X. L., and Lian, H. (1998). Systematic classification of the genus *hippophae* I. *Seabuckthorn. Res.* 1, 13–23.
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., et al. (2012). GAPIT: genome association and prediction integrated tool. *Bioinformatics.* 28 (18), 2397–2399. doi: 10.1093/bioinformatics/bts444
- Liu, J., and Bisseling, T. (2020). Evolution of NIN and NIN-like genes in relation to nodule symbiosis. *Genes* 11 (7). doi: 10.3390/genes11070777
- Liu, X. D., and Dong, B. W. (2013). Influence of the Tibetan plateau uplift on the Asian monsoon-arid environment evolution. *Chinese. Sci. Bull.* 58, 4277–4291. doi: 10.1007/s11434-013-5987-8
- Liu, M. J., Zhao, J., Cai, Q. L., Liu, G. C., Wang, J. R., Zhao, Z. H., et al. (2014). The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* 5, 5315. doi: 10.1038/ncomms6315
- Li, J., Zhou, S., Zhao, Z., and Zhang, J. (2015). The qingzang movement: The major uplift of the qinghai-Tibet plateau. *Sci. China. Earth. Sci.* 58 (11), 2113–2122. doi: 10.1007/s11430-015-5124-4
- Lu, C., Shan, Y., Liu, H., Yang, L., and Ma, G. (2008). Natural vitamin king - seabuckthorn in the application of food ingredients. *China. Food. Addit.* S1, 230–235.
- Mao, Y., Cui, X., Wang, H., Qin, X., Liu, Y., Yin, Y., et al. (2022). De novo assembly provides new insights into the evolution of *Elaeagnus angustifolia* L. *Plant Methods* 18 (1), 84. doi: 10.1186/s13007-022-00915-w
- Mao, K. S., Wang, Y., and Liu, J. Q. (2021). Evolutionary origin of species diversity on the qinghai-Tibet plateau. *J. Syst. Evol.* 59 (6), 1142–1158. doi: 10.1111/jse.12809
- Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 27 (6), 764–770. doi: 10.1093/bioinformatics/btr011
- Mattenberger, F., Sabater-Munoz, B., Toft, C., and Fares, M. A. (2017). The phenotypic plasticity of duplicated genes in *saccharomyces cerevisiae* and the origin of adaptations. *Genes. Genom. Genet.* 7 (1), 63–75. doi: 10.1534/g3.116.035329
- Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., et al. (2013). The UCSC genome browser database: extensions and updates 2013. *Nucleic. Acids Res.* 41 (Database issue), D64–D69. doi: 10.1093/nar/gks1048
- Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., et al. (2017). PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic. Acids Res.* 45 (Database issue), D183–D189. doi: 10.1093/nar/gkx1138
- Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature.* 403 (6772), 853–858. doi: 10.1038/35002501
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32 (1), 268–274. doi: 10.1093/molbev/msu300
- Nicotra, A. B., Atkin, O. K., Bonser, S. P., Davidson, A. M., Finnegan, E. J., Mathesius, U., et al. (2010). Plant phenotypic plasticity in a changing climate. *Trends. Plant Sci.* 15 (12), 684–692. doi: 10.1016/j.tplants.2010.09.008
- Niu, S., Li, J., Bo, W., Yang, W., Zuccolo, A., Giacomello, S., et al. (2022). The Chinese pine genome and methylome unveil key features of conifer evolution. *Cell.* 185 (1), 204–217. doi: 10.1016/j.cell.2021.12.006
- Ou, S., and Jiang, N. (2018). LTR_retriever: A highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant Physiol.* 176 (2), 1410–1422. doi: 10.1104/pp.17.01310
- Pimpinelli, S., and Piacentini, L. (2020). Environmental change and the evolution of genomes: Transposable elements as translators of phenotypic plasticity into genotypic variability. *Funct. Ecol.* 34 (2), 428–441. doi: 10.1111/1365-2435.13497
- Price, A. L., Jones, N. C., and Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics* 21 (Suppl. 1), i351–i358. doi: 10.1093/bioinformatics/bti1018
- Qiu, Q., Zhang, G., Ma, T., Qian, W., Wang, J., Ye, Z., et al. (2012). The yak genome and adaptation to life at high altitude. *Nat. Genet.* 44 (8), 946–949. doi: 10.1038/ng.2343
- Ren, G., Mateo, R. G., Guisan, A., Conti, E., and Salamin, N. (2018). Species divergence and maintenance of species cohesion of three closely related primula species in the qinghai-Tibet plateau. *J. Biogeogr.* 45 (11), 2495–2507. doi: 10.1111/jbi.13415
- Royden, L. H., Burchfiel, B. C., and van der Hilst, R. D. (2008). The geological evolution of the Tibetan plateau. *Sci.* 321 (5892), 1054–1058. doi: 10.1126/science.1155371
- Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000). SMART: a web-based tool for the study of genetically mobile domains. *Nucleic. Acids Res.* 28 (1), 231–234. doi: 10.1093/nar/28.1.231
- Shi, Y. (2002). Characteristics of late quaternary monsoonal glaciation on the Tibetan plateau and in East Asia. *Quat. Int.* 97–98, 79–91. doi: 10.1016/S1040-6182(02)00053-8
- Shi, Y., Li, J., and Li, B. (Eds.). (1998). *The uplift and environmental changes during late cenozoic in the Qinghai-Tibet Plateau* (Guangdong Science & Technology Press).
- Shi, F. X., Li, M. R., Li, Y. L., Jiang, P., Zhang, C., Pan, Y. Z., et al. (2015). The impacts of polyploidy, geographic and ecological isolations on the diversification of panax (Araliaceae). *BMC. Plant Biol.* 15, 297. doi: 10.1186/s12870-015-0669-0
- Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351
- Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC. Bioinf.* 6, 31. doi: 10.1186/1471-2105-6-31
- Soltis, D. E., and Soltis, P. S. (1999). Polyploidy: recurrent formation and genome evolution. *Trends. Ecol. Evol.* 14 (9), 348–352. doi: 10.1016/S0169-5347(99)01638-9
- Soyano, T., Kouchi, H., Hirota, A., and Hayashi, M. (2013). Nodule inception directly targets NF- γ subunit genes to regulate essential processes of root nodule development in lotus japonicus. *PLoS Genet.* 9 (3), e1003352. doi: 10.1371/journal.pgen.1003352
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 22 (21), 2688–2690. doi: 10.1093/bioinformatics/btl446
- Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., et al. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic. Acids Res.* 34 (Web Server issue), W435–W439. doi: 10.1093/nar/gkl200
- Sudmant, P. H., Kitzman, J. O., Antonacci, F., Alkan, C., Malig, M., Tsalenko, A., et al. (2010). Diversity of human copy number variation and multicopy genes. *Sci.* 330 (6004), 641–646. doi: 10.1126/science.1197005
- Sun, H. L. (1996). *Formation and evolution of qinghai-xizang plateau* (Shanghai: Shanghai Scientific & Technical Publishers).
- Sun, H. (2002). Tethys retreat and Himalayas-hengduanshan mountains uplift and their significance on the origin and development of the sino-Himalayan elements and alpine flora. *Acta Botanica. Yunnanica.* 03, 273–288.

- Sun, P., Jiao, B., Yang, Y., Shan, L., Li, T., Li, X., et al. (2021). WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. doi: 10.1101/2021.04.29.441969
- Tang, H., Bowers, J. E., Wang, X., Ming, R., Alam, M., Paterson, A. H., et al. (2008). Synteny and collinearity in plant genomes. *Sci.* 320 (5875), 486–488. doi: 10.1126/science.1153917
- Vekemans, D., Proost, S., Vanneste, K., Coenen, H., Viaene, T., Ruelens, P., et al. (2012). Gamma paleohexploidy in the stem lineage of core eudicots: significance for MADS-box gene and species diversification. *Mol. Biol. Evol.* 29 (12), 3793–3806. doi: 10.1093/molbev/mss183
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., et al. (2017). GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.* 33 (14), 2202–2204. doi: 10.1093/bioinformatics/btx153
- Wang, X., Gao, Y., Wu, X., Wen, X., Li, D., et al. (2021). High-quality evergreen azalea genome reveals tandem duplication-facilitated low-altitude adaptability and floral scent evolution. *Plant Biotechnol. J.* 19 (12), 2544–2560. doi: 10.1111/pbi.13680
- Wang, X., Wang, J., Jin, D., Guo, H., Lee, T. H., et al. (2015). Genome alignment spanning major poaceae lineages reveals heterogeneous evolutionary rates and alters inferred dates for key evolutionary events. *Mol. Plant* 8 (6), 885–898. doi: 10.1016/j.molp.2015.04.004
- Wang, Z. H., Wang, X. F., Lu, T., Li, M. R., Jiang, P., et al. (2022). Reshuffling of the ancestral core-eudicot genome shaped chromatin topology and epigenetic modification in panax. *Nat. Commun.* 13 (1), 1902. doi: 10.1038/s41467-022-29561-5
- Wang, J. P., Yu, J. G., Li, J., Sun, P. C., Wang, L., Yuan, J. Q., et al. (2018). Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the actinidiaceae family. *iSci.* 7, 230–240. doi: 10.1016/j.isci.2018.08.003
- Watanabe, H., and Hattori, M. (2006). Chimpanzee genome sequencing and comparative analysis with the human genome. tanpakushitsu. kakusan. koso. *Protein. Nucleic. Acid. Enzyme.* 51 (2), 178–187.
- Wendel, J. F., Jackson, S. A., Meyers, B. C., and Wing, R. A. (2016). Evolution of plant genome architecture. *Genome. Biol.* 17, 37. doi: 10.1186/s13059-016-0908-1
- Wen, J., Zhang, J. Q., Nie, Z. L., Zhong, Y., and Sun, H. (2014). Evolutionary diversifications of plants on the qinghai-Tibetan plateau. *Front. Genet.* 5. doi: 10.3389/fgene.2014.00004
- Wessler, S. R. (2006). Transposable elements and the evolution of eukaryotic genomes. *P. Natl. Acad. Sci. U.S.A.* 103 (47), 17600–17601. doi: 10.1073/pnas.0607612103
- Wilson, D., Pethica, R., Zhou, Y., Talbot, C., Vogel, C., Madera, M., et al. (2009). SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic. Acids Res.* 37 (Database issue), D380–D386. doi: 10.1093/nar/gkn762
- Wos, G., Choudhury, R. R., Kolar, F., and Parisod, C. (2021). Transcriptional activity of transposable elements along an elevational gradient in *Arabidopsis arenosa*. *Mobile. DNA.* 12 (1), 7. doi: 10.1186/s13100-021-00236-0
- Wu, Z., Chen, H., Pan, Y., Feng, H., Fang, D., Yang, J., et al. (2022). Genome of *Hippophae rhamnoides* provides insights into a conserved molecular mechanism in actinorhizal and rhizobial symbioses. *New. Phytol.* 235 (1), 276–291. doi: 10.1111/nph.18017
- Xing, M., Xue, C., Li, R., and Tian, L. (1989). Karyotype analysis of sea buckthorn. *J. Shanxi. Univ. (Natural Sci. Edition).* 12 (3), 323–330.
- Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic. Acids Res.* 35 (Web Server issue), W265–W268. doi: 10.1093/nar/gkm286
- Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24 (8), 1586–1591. doi: 10.1093/molbev/msm088
- Yang, Q., Bi, H., Yang, W., Li, T., Jiang, J., Zhang, L., et al. (2020). The genome sequence of alpine *Megacarpaea delavayi* identifies species-specific whole-genome duplication. *Front. Genet.* 11. doi: 10.3389/fgene.2020.00812
- Yang, Y., Chen, J., Song, B., Niu, Y., Peng, D., Zhang, D., et al. (2019). Advances in the studies of plant diversity and ecological adaptation in the subnival ecosystem of the qinghai-Tibet plateau. *Chinese. Sci. Bull.* 64 (27), 2856–2864.
- Yan, Q., Owen, L. A., Zhang, Z., Wang, H., Wei, T., Jiang, N. X., et al. (2021). Divergent evolution of glaciation across high-mountain Asia during the last four glacial-interglacial cycles. *Geophys. Res. Lett.* 48 (11), e2021GL092411. doi: 10.1029/2021GL092411
- Yu, L., Diao, S., Zhang, G., Yu, J., Zhang, T., Luo, H., et al. (2022). Genome sequence and population genomics provide insights into chromosomal evolution and phytochemical innovation of *Hippophae rhamnoides*. *Plant Biotechnol. J.* 20 (7), 1257–1273. doi: 10.1111/pbi.13802
- Zdobnov, E. M., and Apweiler, R. (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics.* 17 (9), 847–848. doi: 10.1093/bioinformatics/17.9.847
- Zhang, T., Qiao, Q., Novikova, P. Y., Wang, Q., Yue, J., Guan, Y., et al. (2019). Genome of *Crucihimalaya himalaica*, a close relative of *Arabidopsis*, shows ecological adaptation to high altitude. *P. Natl. Acad. Sci. U.S.A.* 116 (14), 7137–7146. doi: 10.1073/pnas.1817580116
- Zhang, J., Tian, Y., Yan, L., Zhang, G., Wang, X., Zeng, Y., et al. (2016). Genome of plant maca (*Lepidium meyenii*) illuminates genomic basis for high-altitude adaptation in the central andes. *Mol. Plant* 9 (7), 1066–1077. doi: 10.1016/j.molp.2016.04.016
- Zhang, H. Z., and Yang, Z. G. (2013). *Climate atlas of Tibet autonomous region* (Beijing: Meteorological Press).
- Zhao, Y., Tzedakis, P. C., Li, Q., Qin, F., Cui, Q., Liang, C., et al. (2020). Evolution of vegetation and climate variability on the Tibetan plateau over the past 1.74 million years. *Sci. Adv.* 6 (19), eaay6193. doi: 10.1126/sciadv.aay6193
- Zhou, S., Li, J., Zhao, J., Wang, J., and Zheng, J. (2011). “Chapter 70 - quaternary glaciations: Extent and chronology in China,” in *Developments in quaternary sciences*, vol. Vol. 15. Eds. J. Ehlers, P. L. Gibbard and P. D. Hughes (Amsterdam, Netherlands: Elsevier), 981–1002. doi: 10.1016/B978-0-444-53447-7.00070-2
- Zhou, X., Zhang, N., Wang, W., Wang, C., and Wu, B. (2012). Comparison of different nuclear isolation buffers for species of *Hippophae* DNA flow cytometry. *J. Gansu. Agric. Univ.* 47 (04), 155–160.