Genome sequence and population genomics provide insights into chromosomal evolution and phytochemical innovation of *Hippophae rhamnoides*

Liang Yu¹,²,†, Songfeng Diao¹,³,†, Guoyun Zhang¹,†, Jigao Yu⁴, Tong Zhang¹, Hongmei Luo⁵, Aiguo Duan¹, Jinfeng Wang⁶,*, Caiyun He¹,*,† and Jianguo Zhang¹,‡,*  

¹State Key Laboratory of Tree Genetics and Breeding & Key Laboratory of Tree Breeding and Cultivation, National Forestry and Grassland Administration, Research Institute of Forestry, Chinese Academy of Forestry, Beijing, China  
²Collaborative Innovation Center of Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing, China  
³Research Institute of Non-Timber Forestry, Chinese Academy of Forestry.Key Laboratory of Non-timber Forest Germplasm Enhancement & Utilization of National and Grassland Administration, Zhengzhou, China  
⁴School of Life Sciences, North China University of Science and Technology, Tangshan, China  
⁵Experimental Center of Desert Forestry, Chinese Academy of Forestry, Dengkou, China

Received 9 December 2021; revised 9 February 2022; accepted 19 February 2022.  
*Correspondence (Tel 13701322109; fax +86 10 62872015; email zhangjg@caf.ac.cn; Tel 13520585643; fax +86 10 62872015; email hecy@caf.ac.cn; Tel 13784677536; fax 0315-8805600; email wangjinpeng1010@vip.sina.com)  
†Equal contribution.

Summary  
Plants of the Elaeagnaceae family are widely used to treat various health disorders owing to their natural phytochemicals. Seabuckthorn (*Hippophae rhamnoides* L.) is an economically and ecologically important species within the family with richness of biologically and pharmacologically active substances. Here, we present a chromosome-level genome assembly of seabuckthorn (http://hipp.shengxin.ren/), the first genome sequence of Elaeagnaceae, which has a total length of 849.04 Mb with scaffold N50 of 69.52 Mb and 30,864 annotated genes. Two sequential tetraploidizations with one occurring ~36-41 million years ago (Mya) and the last ~24-27 Mya were inferred, resulting in expansion of genes related to ascorbate and aldaret metabolism, lipid biosynthesis, and fatty acid elongation. Comparative genomic analysis reconstructed the evolutionary trajectories of the seabuckthorn genome with the predicted ancestral genome of 14 proto-chromosomes. Comparative transcriptomic and metabolomic analyses identified some key genes contributing to high content of polyunsaturated fatty acids and ascorbic acid (AsA). Additionally, we generated and analysed 55 whole-genome sequences of diverse accessions, and identified 9.80 million genetic variants in the seabuckthorn germplasms. Intriguingly, genes in selective sweep regions identified through population genomic analysis appeared to contribute to the richness of AsA and fatty acid in seabuckthorn fruits, among which *GalLDH*, *GMPase* and *ACC, TERR* were the potentially major-effect causative genes controlling AsA and fatty acid content of the fruit, respectively. Our research offers novel insights into the molecular basis underlying phytochemical innovation of seabuckthorn, and provides valuable resources for exploring the evolution of the Elaeagnaceae family and molecular breeding.

Keywords: seabuckthorn assembly, paleopolyploidizations, karyotype reconstruction, fatty acid, ascorbic acid.

Introduction  
Plant-derived natural products have been used for nutritional and medicinal purposes throughout different countries since ancient times (Evans et al., 2020; Ma et al., 2005; Porras et al., 2021; Schenck et al., 2015). Genome sequences of numerous medicinal plants have been decoded to identify the genes involved in specific metabolic innovations (Ma et al., 2021; Tu et al., 2020; Xu et al., 2016). However, due to the remarkable structural diversity and biological activities of plant-derived compounds, additional medicinally relevant molecules and their biosynthetic pathways remain to be unveiled in plants (Hao et al., 2017; Holzmeier et al., 2020). Plants of the Elaeagnaceae family are widely used to treat various health disorders owing to their natural phytochemicals (Bekker and Glushenkova, 2001). However, the vacancy in the genome of Elaeagnaceae plants has limited the understanding of the evolutionary history and diversification of medicinal ingredients.

Seabuckthorn (*Hippophae rhamnoides* L.; 2n = 2x = 24), named from the discovery that the seriously injured horse (‘hippo’ in Greek) had a strong body and a shiny (‘phaos’ in Greek) coat after eating its fruit and leaves (Christaki, 2012), is the most economically and ecologically important species within the family Elaeagnaceae, with a plantation area of ~3.0 million hectares across Asia, Europe, and South and North America (He et al., 2017; Jia et al., 2012). It was widely used in the restoration of degenerated ecosystem due to its ability to fix atmospheric nitrogen, drought and chilling resistance, and tolerance of infertile and bare soil (Ruan et al., 2013). Its berry has long been regarded as a ‘superfruit’ due to its richness in biologically and pharmacologically active substances, such as flavonoids, vitamins, unsaturated fatty acids (UFAs), and others (Nawaz et al., 2019;
Yong et al., 2016). Extracts including UFAs, vitamins, flavonoids, phenolic acids, and tannins derived from the fruits and leaves of seabuckthorn exhibited strong antioxidant and antibacterial activities (Destandau et al., 2012; Kumar et al., 2011), and they were used to treat and prevent various diseases in Europe, Central Asia, and China hundreds of years ago (Olas, 2017). The anti-proliferation and immune-stimulating properties and counteracting-treatment-induced side effects of seabuckthorn extracts may offer a ‘gold mean’ for cancer therapies (Beata et al., 2018), and increasing researches on their clinical usages have been performed due to the antiviral, antibacterial, anti-inflammatory, and antioxidant properties of seabuckthorn (Destandau et al., 2012; Upadhyay et al., 2010).

Vitamin C, also known as ascorbic acid (AsA), is of critical importance to human health, because it is necessary for numerous physiological processes and the exercise of immune functions, such as connective tissue repair, collagen synthesis, and neurotransmitter production (Feng et al., 2020). Recently, it has been taken as a therapeutic agent and immune booster against COVID-19 (coronavirus disease 2019) in clinical trials (Patel et al., 2020). However, humans cannot synthesize AsA by themselves due to the mutation of the \( \text{GLO} \) (glucose-6-phosphate oxidase) gene that occurred 61 million years ago (Mya), which inspires people to take considerable efforts to promote the accumulation of AsA in plants (Feng et al., 2020). Interestingly, seabuckthorn is known as the ‘King of vitamin C’ (He et al., 2017). The AsA content of seabuckthorn is much higher than that of two genome-sequenced vitamin C abundant species, jujube (\( \text{Ziziphus jujuba} \)) and kiwi (\( \text{Actinidia chinesis} \)), and it is nearly threefold more than that of Chinese kiwi (Sytova et al., 2019). In addition, the fruit oil of seabuckthorn is one of the most valuable product of seabuckthorn (Fatima et al., 2012), and two essential UFAs, linoleic acid and \( \alpha \)-linolenic acid, are abundant in seabuckthorn seeds (Yang and Kallio, 2001), which attracted increasing attention due to their important effects on human health (Huu et al., 2015; Wu et al., 2021).

However, comprehensive gene pathways encoding the biosynthesis of fatty acid and AsA have not been characterized in seabuckthorn, and the contribution of chromosomal evolution to the diversification of fatty acid and AsA synthesis remains unclear.

Here, we report a chromosomal-level genome assembly of the seabuckthorn cultivar ‘sunny’ (http://hcpp.shengxin.ren/), the first genome sequence of Elaeagnaceae, which was used to explore the genome trajectories of seabuckthorn chromosomes and two sequential polyploidizations. We also present novel insights into the molecular basis underlying the biosynthesis of fatty acids and AsA through exploration of transcriptomic and metabolomic data generated from different stages of different tissues. Additionally, through population genomic analysis, genomic regions related to the fatty acid and AsA contents in seabuckthorn fruits were identified. This research offered valuable resources for exploring the evolution, phytochemistry, and adaptation of Elaeagnaceae plants.

**Results**

**Genome sequencing and assembly**

The \( H. \ rhamnoides \) subsp. \( m. \) Roussi ‘sunny’ was a ‘Super Fruit’ cultivar (Figure 1a, b) abundant in bioactive components (Shah, 2015). We used several technologies to sequence and assemble its genome (Figure S1; Table S1). We initially analysed the seabuckthorn genome by K-mer=21, and the estimated proportion of heterozygosity and repetitive sequence in the genome was 0.65% and 62.54%, respectively (Figure S2; Tables S2 and S3). The PacBio Sequel platform was then used to produce a total of 88.39 Gb (90x in depth) data, and high-quality long PacBio reads with an N50 length of 15 703 bp and an average read length of 10 083 bp were obtained (Table S4). In addition, Illumina paired-end libraries with a total depth of 265x were sequenced (Table S5). A total of 349.52 Gb (356x in depth) seabuckthorn DNA sequence was \( \text{de novo} \) assembled, with a cumulative scaffold length of 848.92 Mb and scaffold N50 of 9.13 Mb (Table S6). Hi-C libraries were constructed to build pseudo-chromosomes, and 54.42 Gb clean data, corresponding to approximately 55x coverage of our seabuckthorn genome (Figure 2a; Tables S7-S9), was obtained. The final assembly we obtained was 849.04 Mb, with contig N50 reaching 2.15 Mb, and scaffold N50 of 69.52 Mb (Table 1; Table S10).

Three strategies were used to assess the quality and completeness of this assembled genome. First, a total of 95.54% Illumina paired-end reads were successfully mapped to the assembly (Table S11). Additionally, core eukaryotic gene mapping approach (CEGMA) (Parra et al., 2007) analysis indicated that 97.38% (446) of core eukaryotic genes were detected in the seabuckthorn genome (Table S12). Finally, BUSCO (Simao et al., 2015) analysis showed that 91.32% of the 1440 core plant genes were present in the seabuckthorn genome, and 89.65% of them were complete (Table S13), and the synteny between physical map and genetic map indicated the accuracy of the assembly (Table S14).

**Genome annotation and gene family expansion**

We used RepeatMasker to identify and classify the repeat sequences in the repeat library employing multiple \( \text{de novo} \) prediction procedures. Approximately, 67.81% (575.72 Mb) of the seabuckthorn sequences were identified as repetitive elements, including retrotransposons (58.03%), DNA transposons (5.33%), potential host genes (0.40%), simple sequence repeats (0.91%), and unclassified elements (8.42%) (Table S15). Long-terminal repeats (LTRs) were the main type of retrotransposons, and the two most frequent LTR types were Copia and Gypsy, accounting for 22.61% and 20.07% of the genome, respectively. In addition, 14 316 full-length LTR-RTs were identified, and we discovered that massive insertion events of LTR-RTs occurred in seabuckthorn within the last six million years (Figure S3).

The gene structure was predicted based on homologous prediction, \( \text{ab initio} \) prediction, and transcript evidence-supported predictions. A total of 30 864 genes were predicted, with an average gene length of 4900 bp and an average coding-sequence length of 1307 bp (Tables S16-S17). A corresponding function for 99.66% (30 760) of the predicted genes was observed in the NR, COG, KEGG, TrEMBL, and GO databases, with 7626 annotated by all 5 databases (Figure S4; Table S18). Additionally, 28 790 (93.28%) of the predicted genes were allocated to 12 chromosomes (Table S10). Furthermore, 108 miRNAs, 699 tRNAs, 211 rRNAs, and 5843 pseudogenes were predicted in the seabuckthorn genome (Table S19).

We then investigated the sizes of gene families in seabuckthorn and 10 other species representing major rosids: \( R. \ \text{occidentalis} \) (V1.0), \( Z. \ jujube \) (V1.0), \( M. \ \text{talis} \) (V1.0), \( V. \ \text{vinifera} \) (V3.0), \( C. \ \text{sinensis} \) (Csi_valencia_1.0), \( P. \ \text{trichocarpa} \) (V3.0), \( E. \ \text{grandis} \) (V1.0), \( C. \ papaya \) (ASGPBv0.4), \( A. \ \text{thaliana} \) (TAIR10), and \( M. \ \text{domestica} \) (ASM211411v1). Compared to other plant genomes, 2570 species-specific single-copy genes and 391 unique gene families were found in seabuckthorn (Figure 2c; Table S20). Notably,
unique gene families in seabuckthorn were mainly related to antioxidant activity, response to stimulus, and immune system (Figure S5). Phylogenomic analysis using single-copy orthologous genes from 11 plant genomes indicated that seabuckthorn diverged from jujube ~62 Mya. We detected 471 expansion and 771 contraction gene families in seabuckthorn (Figure 2d), and the expanded gene families were enriched in KEGG terms including biosynthesis of UFAs and flavonoids, while contracted gene families were mainly related to phenylpropanoid biosynthesis (Tables S21 and S22).

**Sequential polyploidization events in seabuckthorn lineage**

Collinearity analysis among the genomes of seabuckthorn and other species was performed to explore the whole-genome duplication (WGD) history that occurred in the seabuckthorn lineage (Tables S23–S25). The distribution of synonymous substitution rates of synonymous sites (Ks) of homologous genes in the collinearity region and a large number of collinearity regions in the seabuckthorn genome indicated that it experienced two rounds of WGDs (Figures 3a,c,d and S6; Tables S24 and S25). The collinearity between seabuckthorn and other species, such as grape, coffee, peach, barrel medic, and jujube, also indicated that two WGDs occurred in the seabuckthorn genome (Figures 3b and S7–S10). Intrageneric comparison of the seabuckthorn genome showed that the median Ks of each block in one best-matched region was approximately 0.3, corresponding to the collinear regions formed by the most recent WGD, and the median Ks of each block in two secondary regions was approximately 0.45, corresponding to the collinear regions formed by the older WGD (Figure 3a,c,d). Based on the median Ks of the collinearity region within and between species, we distinguished orthologous and paralogous regions in their genomes (Figures S9 and S10; Tables S26–S31). In the seabuckthorn genome, we divided the paralogous collinear blocks into two groups, which corresponded to the younger and older peaks of Ks distribution related to the recent (named α) and an older (named β) tetraploidization events, respectively (Figure S11; Table S32).

We also identified the duplicated chromosomal regions including 4328 anchored gene pairs generated from α and 4102 gene pairs from β events, accounting for 24.53% (7571) and 21.39% (6603) of the total genes, respectively (Table S33).

To infer the time of key evolutionary events for the seabuckthorn genome, we performed comparative analyses of the Ks distribution of collinear blocks within and between the studied genomes. Comparisons of the Ks peaks related to the core eudicot common hexaploidization (ECH) event among the genomes showed that the evolutionary rate of seabuckthorn was faster than that of other selected eudicots (Figure S11). A correction-by-shared-event approach (Song et al., 2020) was employed to date the ages of WGDs, with α inferred to occur ~24–27 and β ~36–41 Mya, respectively (Figure 3e; Table S34). Accordingly, the divergence of seabuckthorn and jujube was inferred to have occurred ~68–77 Mya, similar to the phylogenetically inferred date using the MCMCtree (Figure 2d).

Polyploidization has long been regarded as an important driving force for plant evolution, and it may endow genes with potential sub-functionalization and neo-functionalization (Wang et al., 2021). By performing KEGG enrichment on the α-event-related genes (duplicated by the α event and retained) obtained based on the median Ks value, we found that genes related to ascorbate and aldarate metabolism, lipid biosynthesis, and fatty acid elongation were significantly enriched, which coincided with the fact that seabuckthorn fruits were rich in AsA and fatty acids (Table S35). Interestingly, consistent results were obtained for KEGG enrichment of genes related to β events (Table S36), also enriching pathways such as ascorbate and aldarate metabolism and fatty acid elongation. Therefore, two polyploidizations may contribute to the richness of AsA and fatty acid content in seabuckthorn fruits.

**Genomic fractionation of seabuckthorn**

WGD is usually accompanied by vast gene losses and translocations to reshape plant genomes. Each grape gene would have four orthologs in the seabuckthorn genome if there was no gene loss or translocation, due to the two additional tetraploidizations of seabuckthorn genome. Using grape, jujube, peach, and silver birch as references, the average loss rates of anchored genes in seabuckthorn were 64.68%, 73.83%, 64.38%, and 68.86%, respectively (Figure S12; Tables S37–S40). These results revealed that the seabuckthorn genome experienced large-scale genome fractionation, exhibiting extensive gene deletion or translocation.

In the seabuckthorn genomic regions orthologous to a reference genome, collinear orthologs may exist in one genome but not in another genome. Consecutive gene removal was characterized compared to the reference genomes, which was used to explore the scale and potential mechanisms of the seabuckthorn gene loss after polyploidization. The shorter consecutive gene removal (only one or two consecutive gene deletions) accounted for a large proportion of all removals, which made up 42.63% of
all 7051 removals in seabuckthorn with grape as a reference (Table S41). In addition, 4378, 7097, 5982, and 5905 removals were only 10 or fewer consecutive gene deletions, accounting for 77.34%, 84.23%, 84.84%, and 81.61% of all removals with jujube, peach, grape, and silver birch as references, respectively (Table S41). These results suggested that the initial gene loss tends to be smaller and gradually extends over time. Furthermore, the consecutive gene removals exhibited randomness, and can be modelled by geometric distributions. The extension parameters of geometric distributions were 0.279, 0.278, 0.271, and 0.281 using jujube, peach, grape, and silver birch as references, respectively (Figure S13; Table S42).

Reconstructing the ancestral karyotypes and deducing trajectories of seabuckthorn

To infer the chromosome evolution of seabuckthorn, we identified collinearity blocks across seabuckthorn, jujube, and grape (Vitis vinifera) genomes, using grape as the reference (Figure S14). Grape is a species with a relatively stable structure among core eudicots, and is often taken as a reference to understand the genome of other species (Yu et al., 2022).
sequenced eudicot plants (Sato et al., 2012; Velasco et al., 2010). Jujube has the closest genetic relationship with seabuckthorn among the species with whole genome sequenced.

Analysis of the collinearity relationships between them showed that the integrities of grape chromosomes Vv1, Vv5, and Vv17 were essentially preserved in jujube, which were evident from complete correspondence to jujube chromosomes (Figure S14a). For example, almost entire Zj5 shared orthology with grape Vv17, being paralogous to Vv1 and Vv14 due to the ECH event, indicating that Zj5 almost completely retained the chromosomal structure of the core eudicot ancestor, and can be regarded as the proto-chromosome of seabuckthorn and jujube. Although the other jujube chromosomes cannot completely correspond to grape chromosomes, the corresponding orthologous regions of the jujube chromosomes in the grape genome can be inferred based on the collinearity between them. For example, the front and back part of Zj3 correspond to the Vv13 and Vv16, respectively (Figure S14a). In this way, we can speculate on the evolutionary trajectory of jujube chromosomes.

However, the integrities of grape chromosomes were not being preserved in seabuckthorn, indicating that seabuckthorn chromosome evolution was more complicated (Figure S14b). The collinearity relationship of seabuckthorn and jujube genomes was analysed to obtain the possible source of seabuckthorn chromosomes. As mentioned above, the Zj5 almost completely inherited the Vv17. The orthologous regions of Zj5 with seabuckthorn genome were respectively on seabuckthorn chromosomes Hr2, Hr3, Hr6, Hr7, indicating that after seabuckthorn-jujube split, this proto-chromosome (Zj5) was scattered in these seabuckthorn chromosomes (Figure S14c,d). The remaining main part of the Hr2, Hr3, Hr6, Hr7 were merged from other jujube chromosomes.

### Table 1 Summary of seabuckthorn genome assembly and annotation

| Metric                                  | Value   |
|-----------------------------------------|---------|
| Total length of scaffolds (Mb)          | 849.04  |
| Number of scaffolds                     | 3642    |
| Longest scaffolds (Mb)                  | 92.33   |
| N50 of scaffolds (Mb)                   | 69.52   |
| NR90 of scaffolds (Mb)                  | 18.00   |
| Anchored to chromosome (Mb)             | 759.25  |
| Number of predicted protein-coding genes| 30 864  |
| Average gene length (bp)                | 4900    |
| Masked repeat sequence length (Mb)      | 575.72  |
| Percentage of repeat sequences (%)      | 67.81   |

Figure 3 Identification of the WGDs in seabuckthorn genome. (a) Intragenomic collinear blocks for seabuckthorn chromosomes. The collinearity regions formed by the recent or older WGD are highlighted by the red solid box or brown dashed box, respectively. (b) Local collinear blocks of the grape and seabuckthorn genomes. The boxes indicate the collinear regions between the grape and seabuckthorn genomes, in which the dark highlighted boxes indicate the orthologous regions (formed through their split), and the light highlighted boxes indicate the out-paralogous regions (formed through the core eudicot common hexaploidization [ECH] event). Mean $K_s$ of the inferred collinear blocks are shown besides. (c) The intragenomic local collinear blocks for seabuckthorn chromosomes. The boxes indicate the collinear regions within seabuckthorn genome, in which the dark or light highlighted boxes indicate the collinear regions formed by the recent or older WGD, respectively. Mean $K_s$ of inferred collinear blocks are shown besides. (d) $K_s$ distribution of homologous collinearity genes in seabuckthorn genome. ($K_s$ of all homologous gene pairs are replaced by the median $K_s$ of the homologous gene pairs on the block where they are located, due to more conservative.) (e) Correction to the $K_s$ distribution and dating key evolutionary events, and the inferred times. $\alpha$ and $\beta$ represent the two tetraploidization events of seabuckthorn, $\gamma$ represents the core eudicot common hexaploidization (ECH), and L represents the legume-common tetraploid (LCT).
For example, Hr2 had orthologous regions with Zj4, Zj8, Zj11, and Zj12, and their corresponding orthologous regions in grape genome were obtained by using the homologous relationship between jujube and grape. Then, the evolutionary trajectory of the Hr2 chromosome was obtained. The proto-chromosomes of seabuckthorn and jujube were the orthologous regions shared between them. In summary, we inferred the chromosomal evolution trajectories of seabuckthorn and jujube, and showed traces of 14 proto-chromosomes in extant chromosomes (Figure 4).

Fatty acid accumulation and expression of its biosynthetic genes in seabuckthorn

The abundance of fatty acid containing in seabuckthorn fruit (seeds and pulp) (Figure 5a–c), and the UFAs’ biosynthesis pathway significantly enriched in the expanded gene families in seabuckthorn (Table S21) inspired us to investigate the molecular basis underlying fatty acid biosynthesis. Comparative genomics indicated that seabuckthorn had more copies of acetyl-CoA carboxylase (ACC), acyl-ACP thioesterase B (FATB), long-chain acyl-CoA synthetase (LACS), and linoleate desaturase (FAD3) (Figure 5e). Transcriptome sequencing was used to explore key genes for fatty acid synthesis (Figure 5f; Table S43 and S44). ACC is the rate-limiting enzyme in fatty acid biosynthesis, and its overexpression can increase lipid biosynthesis (Chaturvedi et al.; Salie et al., 2016). The significant expansion of the seabuckthorn ACC genes and the continuous expression of more than half of the copies provided the basis for the dramatical accumulation of fatty acids following the fruit ripening. In addition, there were two stearoyl-ACP desaturase (SAD) genes with high expression levels in the pulp (Hr9g0203 and Hr11g2328) and seeds (Hr11g3285 and Hr9g0203), which may contribute to the synthesis of oleic acid (C18:1). FAD2 (oleate desaturase, Hr1g1059) was highly expressed in the pulp and seeds but not in the roots, stems, or leaves, which coincided with the sharp increase in linoleic acid (C18:2) at this stage (Figures 5f and S15a). Notably, lipoxigenases (LOX) catalyse the oxygenation of α-linolenic acid (C18:3), and transcripts of all LOX genes in seeds were almost undetectable at the mature stage, which may contribute to the accumulation of α-linolenic acid in seabuckthorn seeds (Figures 5b,c,f and S15b). In summary, the total fatty acid content in seeds and pulp increased following the maturation of fruit, its content in seeds was higher than that in pulp, and the UFAs in seeds, especially polyunsaturated fatty acids (PUFAs) including linoleic acid and α-linolenic acid, were massively accumulated and become the dominant ingredients (accounts for 79%) in seabuckthorn.

L-Galactose pathway is involved in the high accumulation of ascorbic acid in seabuckthorn

Richness in AsA was another essential feature of seabuckthorn fruit (He et al., 2017), and the AsA content in pulp decreased sharply following the maturation of seabuckthorn fruit (Figure 5d). Comparative genomics and transcriptome analysis were used to explore the molecular basis underlying its biosynthesis (Figures 5e,g and S16). A total of 79 AsA-related genes (Figures 5g and S16) were found in seabuckthorn, and the copy numbers of some of these genes were expanded such as L-
galactose-1-phosphate phosphatase (GalPP) and GDP-L-galactose pyrophosphatase (GGalPP). GalPP has been proven to be a bifunctional enzyme, and has an important influence on the AsA content of plants by participating in the L-Galactose pathway and influencing the inositol pathway (Laing et al., 2004; Torabinejad et al., 2009). GGalPP catalyses the conversion of GDP-L-galactose to L-galactose-1-phosphate, and its activity is crucial for the biosynthesis of L-ascorbate (Sun et al., 2014). Interestingly, GGalPP was continuously expressed in the pulp and gradually decreased as the fruit matures (Figures 5g and S17a), but lowered in the seeds, which was consistent with the abundant accumulation of AsA in the pulp, and the AsA content gradually decreased as the fruit matures (Sytaová et al., 2019). Generally, the majority of ascorbate peroxidase (APX) genes had high expression levels in the pulp and leaves, but the expression level in seeds was lower (Figure 5g). Although APX consumes AsA in the process of reducing hydrogen peroxide (H₂O₂) to water, the appearance of APX is generally accompanied by high levels of AsA, as ascorbate is not only a cofactor of APX but also inactivated by low ascorbate (Gest et al., 2013; Ishikawa and Shigeoka, 2008). In addition, the gene expression profiles of different tissues revealed that transcripts of numerous AsA-related genes, such as glucose-6-phosphate isomerase (PGI), GDP-D-mannose pyrophosphorylase (GMPase), and GDP-D-mannose-3',5'-epimerase (GME), were most abundant in pulp, which may contribute to its high AsA content (Figure 5g).

Genomic variations, population structure, and selective sweeps in the seabuckthorn accessions

We then generated and analysed 645.84 Gb whole-genome sequencing data from 40 wild (wild Hm: wild H. rhamnoides subsp. mongolica Rouss.; wild Hs: H. rhamnoides subsp. sinensis Rouss.) and 15 cultivated (cultivated Hm: cultivated H. rhamnoides subsp. mongolica Rouss.) seabuckthorn accessions to explore the genetic variations in the seabuckthorn populations (Figure 6a–c; Table S45). We identified 9 802 046 high-quality single-nucleotide polymorphisms (SNPs). The phylogenetic and structure analyses classified all 55 seabuckthorn accessions into 3 categories and define their relationship (Figure 6a,c). We calculated the pairwise linkage disequilibrium (LD) between polymorphic sites for all regions in each population, and found that LD decreased more rapidly in wild Hs than in wild Hm and cultivated Hm (Figure 6e).

Candidate genomic regions under positive selection (wild Hs vs. wild Hm) and domestication (cultivated Hm vs. wild Hm) were scanned with a combination of three strategies after considering genetic diversity, population differentiation, site frequency spectrum, and the levels of LD along a chromosome. A total of 1322 and 1481 genes were identified under 181 and 161 potential selective regions for wild Hs and cultivated Hm, respectively (Table S46). Gene ontology (GO) enrichment for genes selected in wild Hs were pollination, photosynthesis, and generation of precursor metabolites and energy, indicating that some selected genes in wild Hs may be involved in species differentiation and environmental adaptation (Table S47). Some genes involved in AsA biosynthesis were located within the selective sweep regions, and there were huge differences in AsA content in the pulp of wild Hs and wild Hm (Figure 7a,d). For example, homologs of L-galactono-1,4-lactone dehydrogenase (GalLDH, Hr2g1922) and GDP-D-mannose pyrophosphorylase (GMPase, Hr10g0139) were found in the regions (Chr 2: 67.00–67.80Mb; Chr 10: 2.00–2.60 Mb) with a specific reduction in diversity in wild Hs compared to wild Hm (Figure 7b,c). GalLDH and GMPase are the key genes of the L-galactose pathway, the main AsA biosynthesis pathway in plants, and their overexpression can lead to AsA accumulation (Landi et al., 2015; Lin et al., 2011). Interestingly, Hr2g1922 and Hr10g0139 in wild Hs pulp showed a higher expression level than that in wild Hm, which may also contribute to the higher AsA content of wild Hs pulp (Figure 7e). Other AsA-related genes in the selective sweep regions were also identified, and the transcriptome data supported that some of them may be beneficial to the accumulation of AsA in wild Hs compared to wild Hm (Figure 7e).

Similarly, there were significant difference in the fatty acid content between the pulp of cultivated Hm and wild Hm (Figure 8d), and KEGG enrichment of the genes positively selected in cultivated Hm showed that glycerolipid metabolism was significantly enriched (Figure 8a,f), which indicated that oil content was an indicator of domestication. Trans-2-enoyl-CoA reductase (TER) catalyses reduction of trans-2-enoyl-CoA to acyl-CoA, an important step in the fatty acid biosynthesis pathway (Hoffmeister et al., 2005), and ACC is the rate-limiting enzyme in fatty acid biosynthesis. These two genes were identified in selective sweep regions in cultivated Hm, and they had higher expression levels in cultivated Hm than wild Hm (Figure 8e). Some other genes involved in fatty acid biosynthesis in the selected regions also had higher expression in cultivated Hm than wild Hm, which appeared to contribute to the difference in their fatty acid content (Figure 8e).

Discussion

The Elaeagnaceae family is mainly distributed in Southeast Asia, with approximately 80 species in three genera (Nazir et al., 2020). There are many economically valuable plants in this family, such as species of H. rhamnoides, Elaeagnus latifolia Linn., and Shepherdia argentea, which play important roles in nutritional beverages, pharmaceutical manufacturing, and ecological regulation (Qin et al., 2010). Genomic resources are essential to promote the molecular breeding of seabuckthorn and evolutionary research on the Elaeagnaceae family. Here, we generated a chromosome-level genome assembly of seabuckthorn, the first genome of Elaeagnaceae, with the support of PacBio in association with next-generation sequencing (NGS) and Hi-C mapping. We constructed ‘Hippophae rhamnoides Information Archive’ (http://hipp.shengxin.ren/) for further use of these genome data, which provides a valuable resource for the study of economically valuable traits of this species and exploring the evolution of the Elaeagnaceae family.

Polyploidization is regarded as an important driver for speciation, and the seabuckthorn genome experienced two independent tetraploidization events, similar to celery and kiwi (Song et al., 2020; Wang et al., 2016, 2018b). Enrichment analysis of the genes of the seabuckthorn genome duplicated by WGDs showed that fatty acid synthesis, AsA, and aldonic acid metabolism were all significantly enriched, which was consistent with the fact that seabuckthorn fruits were rich in fatty acid and AsA. Therefore, two polyploidizations may contribute to the fatty acid and AsA contents in seabuckthorn fruits. LTR-RTs play an important role in genome evolution and affect the gene expression (Huang et al., 2021). We estimated the insertion time of the LTR-RTs by analysing the identified full-length LTR-RTs, which was much later than the polyploidization events in seabuckthorn genome, suggesting that WGDs did not contribute
Figure 5  Content of fatty acid and ascorbic acid, and the expression of the related genes. (a) Changes in the content of total fatty acids in pulp and seeds. T1, immature stage, 46 days post-anthesis; T2, semi-mature, 63 days post-anthesis; T3, mature, 76 days post-anthesis. Same below. (b) Changes in the percentage of fatty acids in pulp and seeds. (c) Changes in fatty acid contents in pulp and seeds. (d) Changes in the content of total AsA in pulp. (e) The heat map shows a comparison of the numbers of key genes related to fatty acid biosynthesis and -Galactose ascorbic acid biosynthesis. (f) The expression levels of the genes involved in fatty acid biosynthesis in leaves, roots, stems, pulp (T1, T2, T3), and seeds (T1, T2, T3). ACC, acetyl-CoA carboxylase; MAT, malonyl-CoA:ACP malonyltransferase; KAS III, ketoacyl-ACP synthase III; KAR, ketoacyl-ACP reductase; HAD, hydroxyacyl-ACP dehydrase; EAR, enoyl-acyl carrier protein (ACP) reductase; KAS I, ketoacyl-ACP synthase I; KAS II, ketoacyl-ACP synthase II; SAD, 3-hydroxyacyl-ACP dehydrase; FATA/B, acyl-ACP thioesterase A/B; LACS, long-chain acyl-CoA synthetase; LPCAT, 1-acylglycerol-3-phosphocholine acyltransferase; FAD2, oleate desaturase; FAD3, linoleate desaturase; LOX, lipoxygenases. (g) The expression levels of the genes involved the -Galactose ascorbic acid biosynthesis and recycling pathways in leaves, roots, stems, pulp (T1, T2, T3), and seeds (T1, T2, T3). PGI, glucose-6-phosphate isomerase; PMI, mannose-6-phosphate isomerase; PMM, phosphomannose mutase; GMPPase, GDP-D-mannose pyrophosphorylase; GME, GDP-D-mannose-3',5'-epimerase; GGalPP, GDP-1-lactonate pyrophosphatase; GaPP, GDP-D-mannose 1-phosphate phosphatase; GalDH, L-galactose dehydrogenase; GalDH, L-galactono-1,4-lactone dehydrogenase; AO ascorbate oxidase, APX ascorbate peroxidase; MDAR monodehydroascorbate reductase; DHAR dehydroascorbate reductase.
significantly to the expansion of LTR-RTs. The continuous nearly six million years expansion of the LTR-RTs in seabuckthorn genome explains its widespread distribution on chromosomes and massive accumulation, which may contribute to its large genome size.

Figure 6 Population genetic analyses of wild H. rhamnoides subsp. Mongolica Rousi (wild Hm), wild H. rhamnoides subsp. Sinensis Rousi (wild Hs), and cultivated H. rhamnoides subsp. mongolica Rousi (cultivated Hm). (a) Neighbour-joining phylogenetic tree of 55 seabuckthorn accessions. (b) Some photographs of the fruits of seabuckthorn used in this study. (c) Population structure plots with K = 2/3/4. Individuals were represented as rows partitioned into segments corresponding to the inferred membership as indicated by the colours. (d) Genome-wide distribution of the $F_{ST}$ and $\pi$ in the wild Hs, wild Hm, and cultivated Hm. The innermost three circles show lines representing $\pi$ in the wild Hs, wild Hm, and cultivated Hm, respectively. The next three circles show lines representing $F_{ST}$ between wild Hs and wild Hm, wild Hs and cultivated Hm, cultivated Hm and wild Hm, respectively. (e) Decay of linkage disequilibrium of cultivated Hm, wild Hm, and wild Hs populations measured by $r^2$.

The importance of UFAs and AsA to the human body need not be repeatedly emphasized. Similar to AsA, the human body cannot synthesize n-3 PUFAs, including α-linolenic acid, eicosapentaenoic acid (EPA), and docosahexaenoic acid (DHA), and α-linolenic acid can be converted into EPA and DHA in the human
body, which make ω-3-linolenic acid even more vital to human health (Yuan et al., 2022). ω-3-Linolenic acid is produced from catalytic linoleic acid desaturation at FAD3 and omega-3 fatty acid desaturase. The ectopic expression of the Arabidopsis FAD3 gene in Glycine max and specific expression of the G. max FAD3 gene in Sesamum indicum can significantly increase the ω-3-linolenic acid content (Bhunia et al., 2014). The specific expression of the Jatropha curcas of J. curcas seeds increased the ω-3-linolenic acid content by 25.0–53.2%, while the linoleic acid content decreased by 11.4–23.5% compared with the wild type (Wu et al., 2013). In addition, it has also been reported that other genes also affect the content of linolenic acid, such as FAD2 and LOX. FAD2 catalyses the conversion of oleic acid to linoleic acid, and its low expression will hinder the linoleic acid synthesis, thereby affecting the production of ω-3-linolenic acid (Schmidt et al., 1994). LOX is a key gene in the process of ω-3-linolenic acid metabolism, which regulates the first step of ω-3-linolenic acid oxidation to generate fatty acid hydroperoxide, and its low expression will contribute to the accumulation of ω-3-linolenic acid (Kühn and Borchert, 2002).

Consistent with previous studies (Yang and Kallio, 2001), we detected a large amount of fatty acids in the pulp and seeds of seabuckthorn, and they continued to increase as the fruit matured. Linoleic acid and ω-3-linolenic acid are two extremely important types of PFA (Yang and Kallio, 2001), and both were abundant in seabuckthorn seeds. Interestingly, we found that some fatty acid biosynthesis genes were significantly expanded in

Figure 7 Genes related to ascorbic acid (AsA) biosynthesis in selective sweep regions of wild H. rhamnoides subsp. sinensis Rouss. (wild Hs). (a) Selection signatures identified by μ statistics of wild Hs are illustrated. GalPP, L-galactose-1-phosphate phosphatase; GalLDH, L-galactono-1,4-lactone dehydrogenase; NAT, ascorbate transporter; GMPase, GDP-D-mannose pyrophosphorylase. (b) Local display of a region (Chr 2 67.0–67.8 Mb) where the diversity of wild Hs was specifically reduced compared with wild Hm, and a homologous gene (Hr2g1922) of GalLDH was located in this region. (c) Local display of a region (Chr 10 2.00–2.60 Mb) where the diversity of wild Hs was specifically reduced compared with wild Hm, and a homologous gene (Hr10g0139) of GMPase was located in this region. (d) AsA content of wild Hs and wild Hm pulp. (e) Differential expression of candidate genes related to AsA content between pulp of wild Hs and wild Hm, at three developmental stages (T1, T2, T3). T1, immature stage, 46 days post-anthesis; T2, semi-mature, 63 days post-anthesis; T3, mature, 76 days post-anthesis. (f) Genotypes of SNPs around putative selective sweeps containing a homologous gene (Hr2g1922) of GalLDH in the wild Hs and wild Hm.
the seabuckthorn genome. For example, ACC, which is the rate-limiting enzyme in fatty acid biosynthesis (Chaturvedi et al., 2016; Salie et al., 2016), plays an important role in fatty acid biosynthesis and accumulation. In addition, at the later stage of fruit ripening, transcriptome sequencing can hardly detect LOX transcripts. Notably, LOX catalyses the oxygenation of PFA, and no or low expression of LOX in the late fruit ripening stage may contribute to the accumulation of PFAs in seabuckthorn seeds. As the fruit matured, the AsA content in the pulp was significantly reduced, but it is an important biologically active substance contained in seabuckthorn fruits. In fact, the AsA redox system is still vital in maintaining all aerobic multicellular eukaryotes (Edgar, 2019). By analysing the resequencing results, we identified several genes related to AsA biosynthesis in the subspecies with extremely different AsA contents in the two types of wild Hs and wild Hm pulp. At the different ripening stages of the fruit, Hr2g1922 and Hr10g0139, homologs of GalLDH and GMPase, respectively, in wild Hs had a high expression level, while they were extremely low in wild Hm. The differential expression patterns of these genes in wild Hs and wild Hm and

Figure 8  Genes related to fatty acid biosynthesis in selective sweep regions of cultivated H. rhamnoides subsp. mongolica Rousi (cultivated Hm). (a) Selection signatures identified by $\mu$ statistics of cultivated Hm are illustrated. TER, Trans-2-enoyl-CoA reductase; LACS, long-chain acyl-CoA synthetase; ACC, acetyl-CoA carboxylase; KCS, 3-ketoacyl-CoA synthase; OPCL1, OPC-8: CoA ligase. (b) Local display of a region (Chr1 90.80–91.80 Mb) where the diversity of cultivated Hm was specifically reduced compared with wild Hm, and a homologous gene (Hr1g3574) of TER was located in this region. (c) Local display of a region (Chr7 0.00–0.40 Mb) where the diversity of cultivated Hm was specifically reduced compared with wild Hm, and a homologous gene (Hr7g0011) of ACC was located in this region. (d) Fatty acid content of cultivated Hm and wild Hm pulp. (e) Differential expression of candidate genes related to fatty acid content between cultivated Hm and wild Hm at three developmental stages (T1, T2, T3). T1, immature stage, 46 days post-anthesis; T2, semi-mature, 63 days post-anthesis. (f) KEGG enrichment results of genes identified in the cultivated Hm genomic regions were scanned by a combination of three strategies.
the extremely different AsA contents in their pulp led us to speculate on the important role of these genes in the synthesis/accumulation of AsA in sea buckthorn.

In summary, we provided high-quality assemblies of *H. rhamnoides*, the first reference genome of the Elaeagnaceae family. Comparative genomics analysis revealed two WGDs and chromosome evolution history of the seabuckthorn genome. The integration of multi-omics data advanced our understanding of fatty acid and AsA biosynthesis in seabuckthorn. Moreover, these data were valuable resources for seabuckthorn research and breeding, and for comparative genomic analysis of the Elaeagnaceae family.

### Methods

#### Genome size estimation, de novo sequencing, and assembly

The genomic DNA of *H. rhamnoides* ‘sunny’ was extracted using a modified cetyltrimethylammonium bromide (CTAB) method and obtained 43.21 Gb reads, which were used to calculate and plot the k-mer frequency distribution (Table S2). Genome size was estimated by the formula \( G = K_{\text{num}}/\text{peak depth} \) (Kmer = 21) (Li et al., 2010) (Table S3).

Ten paired-end (PE) libraries with insert sizes of 180 bp, 300 bp, 500 bp, 3 kb, 4 kb, 5 kb, 8 kb, 10 kb, 15 kb, and 17 kb were constructed following standard Illumina protocols, and sequenced on the Illumina HiSeq 2500 platform (Table S5). For PacBio libraryconstruction, genomic DNA was sheared to 20 kb, and sequenced on the PacBio Sequel system (Table S4). Fresh leaves were used to construct a Hi-C sequencing library, and we digested the cross-linked chromatin with DpnII and ligated in situ after biotinylation. DNA fragments were enriched via the interaction of biotin and blunt-end ligation and then sequenced on the Illumina HiSeq 2500 platform.

ALLPATHS-LG (Gnerre et al., 2011), GapCloser, and SSPACE (Boetzer et al., 2011) were used to assemble the Illumina data, fill the gaps, and link contigs to scaffolds, respectively. The PacBio reads, corrected by CANU v1.7 (Koren et al., 2017), were assembled by WTDGB and DBG2OLC. Finally, Quickmerge software (Mahul et al., 2016) was used to produce a more contiguous assembly. The Hi-C data (Table S6) were aligned to the assembly by BWA v0.7.10-r789 (Li and Durbin, 2009) to detect valid contacts and corrected (Table S7). The preassembled scaffolds were clustered, ordered, and oriented onto chromosomes using LACHESIS software (Burton et al., 2013) (Table S9). The LACHESIS-based assembly was artificially corrected, filled gaps, and deleted duplicate sequences (Burton et al., 2013) (Table 1, Table S10).

#### Genome quality assessment

The CEGMA (Parra et al., 2007) and BUSCO (Simao et al., 2015) pipelines were used to assess the completeness and accuracy of the genome assembly with default parameters (Tables S12 and S13). In addition, we aligned Illumina paired-end reads to the genome assembly and assessed the assembled proportion (Table S11).

#### Gene prediction and annotation

We identified repetitive sequences in seabuckthorn using a combination of *de novo* and homology-based searches with Repbase v19.06 (Jurka et al., 2005). Four software including LTR-FINDER v1.05 (Xu and Wang, 2007), MITE-Hunter (Han and Wessler, 2010), RepeatScout v1.05 (Price et al., 2005), and PILER-DF v2.4 (Edgar and Myers, 2005) were used to construct a repetitive sequence database. PASTEClassify v1.0 (Hoeve et al., 2014) was used to classify the database and merged with the Repbase database. RepeatMasker v4.0.6 (Chen, 2004) was used to predict and classify the repeated elements.

We conducted annotation of protein-coding with homology-based, *de novo*, and RNA-seq based approaches. *De novo* prediction was conducted using Genescan v1.1.0 (Karlin, 1997), Augustus v2.4 (Stanke et al., 2006), GlimmerHMM v3.0.4 (Majors et al., 2004), GenOello 1.4 (Blanco et al., 2007), and SNAP v2006-07-28 software (Korf, 2004) with default parameters and the A. thaliana gene model as the training model. We performed homology-based prediction using GenMoMa v1.3.1 software (Keilwagen et al., 2019) with the protein databases of A. thaliana, *Malus × domestica*, and *Fragaria vesca*. RNA sequencing-based prediction was performed using TransDecoder v2.0. GeneMarkS-T v5.1 (Tang et al., 2014), and PASA v2.0.2 software (Campbell et al., 2006). We combined the outputs from the above three methods into the consensus gene sets with EVM v1.1.1 (Haas et al., 2008). Specifically, all exons in a region were scored based on their weight and length features, and the dynamic programming algorithm was used to find the best path, which was the optimal gene structure. Based on the homologous species information, high-confidence genes were added with GenMoMa v1.3.1 (Keilwagen et al., 2019), and the predicted genes were modified with PASA v2.0.2 (Campbell et al., 2006), including adding UTR, obtaining alternative splicing, and correcting EVM prediction results. Finally, the genes were filtered to obtain the final gene prediction result. We conducted gene function annotation by using the known protein databases, such as KEGG, KOG, TrEMBL, and NR (E-value 1e−5) (Figure S3).

Non-coding RNA and pseudogenes were annotated. trRNAscan-SE v1.3.1 (Eddy, 1997) was used to search trRNA. INFERNAL program (Nawrocki and Eddy, 2013) was used to predict miRNAs and rRNAs, based on the RFam and miRBase databases (Kozomara and Griffiths-Jones, 2014). GenBlastA v1.0.4 (She et al., 2009) was used to search for homologous gene sequences on the genome of the completely shielded true locus, and then GeneWise v2.4.1 (Birney et al., 2004) was used to find the immature stop codons and frame shift mutation in the gene sequence as the pseudogenes.

#### Gene family identification, expansion, and contraction

OrthoMCL v2.0.9 (Li et al., 2003) was used for the orthologous gene identification in seabuckthorn and ten species including *Rubus occidentalis*, *Ziziphus jujube*, *Medicago truncatula*, *Vitis vinifera*, *Citrus sinensis*, *Populus trichocarpa*, *Eucalyptus grandis*, *Carica papaya*, *A. thaliana*, and *Malus domestica*. MUSCLE (Edgar, 2004) was used for the sequence alignments on all single-copy orthologous genes. A maximum-likelihood phylogenetic tree was constructed by PHYML v20151210 (Guindon et al., 2010) with the ‘relaxed-clock (clock = 2)’ model and ‘f84’ model, and the species divergence times was inferred by MCMCTREE of PAML v4.7a (Yang, 2007). Divergence times were calibrated using the divergence times for *E. grandis*–*A. thaliana* (<107, >105 Mya) and *V. vinifera–C. papaya* (<117, >110 Mya) from the TimeTree database. The gene family expansion and contraction of 11 species were analysed using CAFE v4.2 (De Bie et al., 2006), and the expanded and contracted gene families were subjected to GO enrichment to analyse their functions.
Inference of gene collinearity, Ks calculation, distribution fitting, and correction

ColinearScan v1.0 was used to infer the colinear genes (Wang et al., 2006). The potentially homologous genes within a genome or between genomes were identified using the BlastP programs, which were used as the input of ColinearScan to locate homologous gene pairs in the collinearity region, and the maximal gap between neighbouring genes was set to be 50 intervening genes (Wang et al., 2017). The gene families with >30 genes were removed before performing ColinearScan (Song et al., 2020). We used the MCScanX toolkit v1.0 (Wang et al., 2012) and the Neo-Gojobori approach to obtain homologous gene dot-plot and synonymous nucleotide substitutions on synonymous sites (Wang et al., 2018a).

We estimated the probability density distribution of Ks using the kernel smoothing density function (ksdensity) in MATLAB R2017a. The Gaussian approximation function in the fitting toolbox cftool was used to perform multi-peak fitting, and the coefficient of determination (R-squared) was at least 0.95. Then, we performed evolutionary rate correction by aligning the peaks of the ECH from different Ks distributions to the corresponding location in the grape Ks distribution (Song et al., 2020).

Reconstructing ancestral karyotypes of seabuckthorn and jujube

With grape as the outer group, we inferred the possible chromosomal evolution process of the seabuckthorn genome through the comparison and analysis of homologous structures between species, and finally obtained karyotype changes and deduced the trajectories of the formation of extant chromosomes (Song et al., 2020). Specifically, when the chromosome rearrangement event occurred in an ancestor, then as species differentiated, the chromosome rearrangement existed in different species. When chromosome rearrangement occurred before WGD, it will be duplicated along with WGD, and the same rearrangement will exist in the subsequent sub-genome. When the chromosomal rearrangement event occurred after the WGD, it will have a different rearrangement in the sub-genome. Using this strategy, we can account for the fact that the complete chromosome corresponding to grape in jujube can be considered as an independent chromosome in the common ancestor of grape and jujube. Based on the comparison of the collinear homology structure of the jujube, seabuckthorn, and grape genomes, the co-occurring chromosomal rearrangements can be considered to have occurred in the ancestral species of jujube and seabuckthorn. Finally, we inferred the proto-chromosomes composition of the shared ancestors of seabuckthorn and jujube, and the general process of the development of the chromosome karyotype to extant plants.

Transcriptome analysis, assay of fatty acid composition and AsA content

Seabuckthorn fruits were collected at 46, 63, and 76 days post-anthesis. Library construction and RNA-seq were performed according to He et al. (2017). TopHat2 (Kim et al., 2013) and Cuffdiff were used to map the clean reads to the reference genome and calculated expression level, respectively. Seeds and pulp at the corresponding time were used to determine the fatty acid composition, and the AsA content of pulp was also determined. Fatty acid and AsA were determined by using GC-MS as previously described (Liu et al., 2011, 2016).

Identification of genomic variations in the seabuckthorn accessions

The genomic DNA of 55 seabuckthorn accessions was extracted, and the paired-end libraries were prepared as previously mentioned (Xie et al., 2019). Sequencing was performed using the Illumina HiSeq 4000 platform. BWA 0.7.17-r1188 (Li and Durbin, 2009) was used to map the clean reads to the genome. SAMtools was used to sort mapping results and remove PCR-duplicated reads (Li et al., 2009). The SNP detection procedure of the Genome Analysis Toolkit (GATK) was used for variant calling.

Population structure analyses

A neighbour-joining phylogenetic tree of the re-sequenced seabuckthorn accessions was constructed by SNPhylo (Lee et al., 2014) using all identified SNPs. All bi-allelic SNPs were used for principal component analysis (PCA) (Price et al., 2006; Teng et al., 2017), and the population structure was analysed using ADMIXTURE (Alexander et al., 2009). LD decay statistics for different populations were calculated by PopLDdecay3.26 (https://github.com/BGI-shenzhen/PopLDdeacy).

Identification of selective sweeps

The population parameters (θ and FST) were calculated for each window by VCFtools (v0.1.13, https://vcftools.github.io) (Danecek et al., 2011) with a window size of 100 kb and a step size of 10 kb. The FST and relative genetic diversity change of wild Hs and cultivated Hm were calculated. RAiSD (Nikolaos and Pavlidis, 2003) with a window size of 20 SNPs was used to calculate the µ statistic (representing signatures of sweeps) across the genomes to detect candidate genomic regions under positive selection for wild Hs and cultivated Hm populations, respectively. We selected the genes within the selective sweep regions identified by the three methods (top 1% for µ statistic; top 10% for FST and pwild Hs/pwild Hm or pcultivated Hm/pwild Hm) as candidate genes under selection.

Acknowledgements

This work was supported by grants from Special Fund for Forest Scientific Research in the Public Welfare (201504103), the National Natural Science Foundation of China (U2003116) and the Fundamental Research Funds of CAF (CAFYBB20205Z2001-2). The genome sequencing, assembly, and annotation were performed with the help of Biomarker Technologies.

Conflicts 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.

Authors contributions

J.Z. and C.H. conceived the project and were responsible for the project initiation. G.Z., H.L., A.D., and T.Z. contributed to sample preparation. G.Z. and C.H. contributed to the database construction. L.Y., S.D., J.Y., and J.W. worked on comparative and population genomic analysis. L.Y., J.W., and C.H. wrote the manuscript. J.Z. revised the manuscript. All authors read and approved the manuscript.
References

Alexander, D.H., Novembre, J. and Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Beata, O., Bartosz, S. and Karolina, U. (2018). The antcancer activity of sea buckthorn [Elaegnus rhamnoides (L.) A. Nelson]. Front. Pharmacol. 9, 232.

Bekker, N.P. and Gleshchenkova, A.I. (2001) Components of certain species of the Elaeagnaceae family. Chem. Nat. Compd. 37, 97–116.

Bunika, R.M., Chakrabarty, A., Karu, R., Gayatyi, T., Bhattacharyya, J., Basu, A., Maiti, M.K. et al. (2014) Seed-specific increased expression of 25 albumin promoter of sesame qualifies it as a useful genetic tool for fatty acid metabolic engineering and related transgenic intervention in sesame and other oil seed crops. Plant Mol. Biol. 86, 351–365.

Binney, E., Clamp, M. and Durbin, R. (2004) GeneWise and genomewise. Genome Res. 14, 988–995.

Blanco, E., Parra, G. and Guigo, R. (2007) Using gencode to identify genes. Curr. Protoc. Bioinformatics, 18, 4.3.1–4.3.26.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D. and Pirovano, W. eds. (2011) Hippophae rhamnoides L. (Sea Buckthorn): a potential source of nutraceuticals. Bioinformatics, 27, 578–579.

Burton, J.N., Adey, A., Patwardhan, P.R., Qiu, R., Kitzman, J.O. and Shendure, J. (2013) Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. 31, 1119–1125.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M. and Buell, C.R. (2006) Epigenetic modifiers in aging-associated inflammation and disease. J. Exp. Bot. 57, R15–R15.

Clamp, M. and Durbin, R. (2004) GeneWise and genomewise. Genome Res. 14, 988–995.

De Bie, T., Cristianini, N., Demuth, J.P. and Hahn, M.W. (2006) CAFE: a computational tool for the study of gene family evolution. Plant Biotechnology Journal, 4, 10.1–10.14.

Destandau, E., Floch, G.L., Lucchesi, M.E. and Elfakir, C. (2012) Antimicrobial, antioxidant and phytochemical investigations of sea buckthorn (Hippophae rhamnoides L.) leaf, stem, root and seed. Food Chem., 131(3), 754–760.

Eddy, T.M.L.A.S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res., 25(5), 955–964.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1799–1797.

Edgar, R.C. and Myers, E.W. (2005) PILER: identification and classification of genomic repeats. Bioinformatics, 21(Suppl 1), i152–i158.

Evans, L.W., Stratton, M.S. and Ferguson, B.S. (2020) Dietary natural products and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. Biosci. Biotechnol. Biochem. 72, 1143–1154.

Feng, C., Feng, C., Lin, X., Liu, S., Li, Y. and Kang, M. (2020) A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (Psidium guajava). Plant Biotechnol. J., 19(4), 717–732. https://doi.org/10.1111/pbi.13498

Gest, N., Gautier, H. and Stevens, R. (2013) Ascorbate as seen through plant evolution: the rise of a successful molecule? J. Exp. Bot. 64, 33–53.

Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T. et al. (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl Acad. Sci. USA, 108, 1513–1518.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321.

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

Han, Y. and Wessler, S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 38, e199.

Hao, D.C., Gu, X. and Xiao, P. (2017) Anemone medicinal plants: ethnopharmacology, phytochemistry and biology. Acta Pharm. Sin. B, 7, 146–158.

He, C., Zhang, G., Zhang, J., Zeng, Y. and Liu, J. (2017) Integrated analysis of multomic data reveals the role of the antioxidant network in the quality of sea buckthorn berry. FASEB J. 31, 1929–1938.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamiloux, V. and Quevresse, H. (2014) PASTEC: an automatic transposable element classification tool. PLoS One, 9, e91929.

Honegger, M., Potrovitski, M., Nowitzi, U. and Martin, W. (2005) Mitochondrial trans-2-Enoyl-CoA reductase of wax ester fermentation from Euglena gracilis defines a new family of enzymes involved in lipid synthesis. J. Biol. Chem. 280, 4329–4338.

Holmeyer, L., Hartig, A.K., Franke, K., Brandt, W., Mueller-Riehl, A.N., Wessjohn, L.A. and Schnitzler, J. (2020) Evaluation of plant sources for antifungal lead compound discovery by correlating phylogenetic, spatial, and bioactivity data. Proc. Natl Acad. Sci. USA, 117, 12444–12451.

Huang, H., Liang, J., Tan, Q.I., Ou, L., Li, X., Zhong, C., Huang, H. et al. (2021) Insights into tripterpenoid synthesis and unsaturated fatty-acid accumulation provided by chromosome-level genome analysis of Akebia trifoliata subsp. australis. Hortic. Res. 8, 1–15.

Huu, T., Nguyen, H., Park, K., Koster, L., Rebecca, C.E., Nguyen, H., Shanklin, J. et al. (2015) Redirection of metabolic flux for high levels of omega-7 monounsaturated fatty acid accumulation in camellia seeds. Plant Biotechnol. J. 13, P38–P50.

Ishikawa, T. and Shigeoka, S. (2008) Recent advances in ascorbate biosynthesis and the physiological significance of ascorbate peroxidase in photosynthesizing organisms. Biosci. Biotechnol. Biochem. 72, 1143–1154.

Jia, D., Abbott, R., Liu, T., Mao, K., Bartish, I. and Liu, J. (2012) Out of the Qinghai-Tibet Plateau: evidence for the origin and dispersal of Eurasian temperate plants from a phylogeographic study of Hippophae rhamnoides (Elaeagnaceae). New Phytol. 194, 1123–1133.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Walichiewicz, J. (2005) Repbase update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. 110, 462–467.

Karlín, C.B.A.S. (1997) Predicting complete gene structures in human genomic DNA. J. Mol. Biol. 268, 78–94.

Keilwagen, J., Hartung, F. and Grau, J. (2019). GeMoMa: homology-based gene prediction utilizing intron position conservation and RNA-seq data. In Gene Prediction, Vol 1962, pp. 161–177. New York, NY: Humana. https://doi.org/10.1007/978-1-4399-9173-0_9

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcripts in the presence of insertions, deletions and gene fusions. Genome Biol. 14, R36.

Koren, S., Walen, B.P., Berlin, K., Miller, J.R., Bergman, N.H. and Phillippy, A.M. (2017) Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736.

Korf, I. (2004) Gene finding in novel genomes. BMC Bioinformatics, 5, 59.

Kozomara, A. and Griffiths-Jones, S. (2014) miBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 42, D68–73.

Kühn, H. and Borchert, A. (2002) Regulation of enzymatic lipid peroxidation: the interplay of peroxidizing and peroxide reducing enzymes! This article is part of a series of reviews on “Regulatory and Cytoprotective Aspects of Lipid Hydroperoxide Metabolism”. The full list of papers may be found on the homepage of the journal. Free Radic. Biol. Med. 33, 154–172.
Kumar, M.S.Y., Dutta, R., Prasad, D. and Misra, K. (2011) Subcritical water extraction of antioxidant compounds from Seabuckthorn (Hippophae rhamnoides) leaves for the comparative evaluation of antioxidant activity. Food Chem. 127, 1309–1316.

Li, W., Bulley, S., Wright, M., Cooney, J., Jensen, D., Barraclough, D. and Macrae, E. (2004) A highly specific L-galactose-1-phosphate phosphatase on the path to ascorbate biosynthesis. Proc Natl Acad. Sci. USA, 101, 16976–16981.

Landi, M., Farinacci, M., Basile, A., Salvini, M., Guidi, L. and Pugliesi, C. (2015) Overexpression of L-galactono-1,4-lactone dehydrogenase (L-GalLDH) gene correlates with increased ascorbate concentration and reduced browning in leaves of Prunus salicina L. after cutting. Plant Cell Tissue Organ Culture, 123, 109–120.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMTools. Bioinformatics, 25, 2078–2079.

Li, L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Bioinformatics, 19, 3562–3564.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMTools. Bioinformatics, 25, 2078–2079.

Li, L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, R., Fan, W., Tian, G., Zhu, H., He, L. and Al, E. (2010) The sequence and de novo assembly of the giant panda genome. Nature, 463, 311–317.

Lin, L.L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, R., Fan, W., Tian, G., Zhu, H., He, L. and Al, E. (2010) The sequence and de novo assembly of the giant panda genome. Nature, 463, 311–317.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMTools. Bioinformatics, 25, 2078–2079.

Li, L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, R., Fan, W., Tian, G., Zhu, H., He, L. and Al, E. (2010) The sequence and de novo assembly of the giant panda genome. Nature, 463, 311–317.

Lin, L.L., Shi, Q.H., Wang, H.S., Qin, A.G. and Xian-Chang, Y.U. (2011) Overexpression of L-galactono-1,4-lactone dehydrogenase (L-GalLDH) gene correlates with increased ascorbate concentration and reduced browning in leaves of Prunus salicina L. after cutting. Plant Cell Tissue Organ Culture, 123, 109–120.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMTools. Bioinformatics, 25, 2078–2079.

Li, L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, R., Fan, W., Tian, G., Zhu, H., He, L. and Al, E. (2010) The sequence and de novo assembly of the giant panda genome. Nature, 463, 311–317.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.

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

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G. et al. (2009) The Sequence Alignment/Map format and SAMTools. Bioinformatics, 25, 2078–2079.

Li, L., Stoeckert, C.J. Jr and Roos, D.S. (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, R., Fan, W., Tian, G., Zhu, H., He, L. and Al, E. (2010) The sequence and de novo assembly of the giant panda genome. Nature, 463, 311–317.

Lee, T.-H., Guo, H., Wang, X., Kim, C. and Paterson, A.H. (2014) SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genom. 15, 162.
Wang, J.P., Yu, J.G., Li, J., Sun, P.C., Wang, L., Yuan, J.Q., Meng, F.B. et al. (2018b). Two likely auto-tetraploidization events shaped kiwifruit genome and contributed to establishment of the Actinidiaceae family. *iScience*, **7**, 230–240.

Wang, S., Liang, H., Wang, H., Li, L., Xu, Y., Liu, Y., Liu, M. et al. (2021) The chromosomescale genomes of *Dipterocarpus turbinatus* and *Hopea hainanensis* (*Dipterocarpaceae*) provide insights into fragrant oleoresin biosynthesis and hardwood formation. *Plant Biotechnol. J.* **20**, 538–553.

Wang, X., Guo, H., Wang, J., Lei, T., Liu, T., Wang, Z., Li, Y. et al. (2016) Comparative genomic de-convolution of the cotton genome revealed a decaploid ancestor and widespread chromosomal fractionalization. *New Phytol.* **209**, 1252–1263.

Wang, X., Shi, X., Li, Z., Zhu, Q., Kong, L., Tang, W., Ge, S. et al. (2006) Statistical inference of chromosomal homology and ancestral gene synteny and collinearity. *BMC Bioinformatics*, **7**, 1–13.

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

Wu, Z., Liu, H., Zhan, W., Yu, Z., Qin, E., Liu, S., Yang, T. et al. (2021) The chromosomescale reference genome of safflower (*Carthamus tinctorius*) provides insights into linolic acid and flavonoid biosynthesis. *Plant Biotechnol. J.* **19**, 1725–1742.

Xie, D., Xu, Y., Wang, J., Liu, W. and Zhang, Z. (2019) The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. *Nat. Commun.* **10**, 5158.

Xu, H., Song, J., Luo, H., Zhang, Y., Li, Q., Zhu, Y., Xu, J. et al. (2016) Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. *Mol. Plant* **9**, 949–952.

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

Yang, B. and Kallio, H.P. (2001) Fatty acid composition of lipids in sea buckthorn (*Hippophae rhamnoides*) berries of different origins. *J. Agric. Food Chem.* **49**, 1939–1947.

Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591.

Yang, W., Liang, Z., Yazhen, H., Feng, Z., Wei, W., Feng, L., Xue, Y. et al. (2016) Protective effect of proanthocyanidins from sea buckthorn (*Hippophae Rhamnoides*) seed against visible light-induced retinal degeneration in vivo. *Nutrients* **8**, 245.

Yuan, Q., Xie, F., Huang, W., Hu, M., Yan, Q., Chen, Z., Zheng, Y. et al. (2022) The review of alpha-linolenic acid: sources, metabolism, and pharmacology. *Phytother Res.* **36**, 164–188.

Supporting information

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

**Figure S1** Seabuckthorn assembly pipeline

**Figure S2** K-mer distribution of the seabuckthorn genome.

**Figure S3** The insertion time of full-length LTR-RTs in seabuckthorn genome.

**Figure S4** Venn diagram of gene function annotations supported by five databases.

**Figure S5** The enrichment analysis of seabuckthorn unique genes.

**Figure S6** Ks distribution of homologous collinearity genes in seabuckthorn genome.

**Figure S7** The phylogenetic tree of seabuckthorn and other genomes orthologous genes.

**Figure S8** Example microsynteny analysis indicating that two whole-genome duplications occurred in seabuckthorn.

**Figure S9** (a) Homologous dot-plot between seabuckthorn and silver birch genomes. (b) Homologous dot-plot between seabuckthorn and peach genomes. (c) Homologous dot-plot between seabuckthorn and grape genomes. (d) Homologous dot-plot between seabuckthorn and jujube genomes.

**Figure S10** Homologous alignment of seabuckthorn and other relative genomes with grape as a reference.

**Figure S11** Distribution of median Ks between anchor gene pairs in intergenomic blocks (solid curves) and intragenomic blocks (dashed curves) before evolutionary rate correction, and the Ks value of the peak corresponding to the key evolutionary events of seabuckthorn and other related species.

**Figure S12** (a) The retention of duplicated genes residing in 4 subgenomes of seabuckthorn using the silver birch as reference. (b) The retention of duplicated genes residing in 4 subgenomes of seabuckthorn using the peach as reference. (c) The retention of duplicated genes residing in 4 subgenomes of seabuckthorn using the jujube as reference. (d) The retention of duplicated genes residing in 4 subgenomes of seabuckthorn using the grape (Vv1-Vv10) as reference. (e) The retention of duplicated genes residing in 4 subgenomes of seabuckthorn using the grape (Vv10-Vv19) as reference.

**Figure S13** (a) Fitting a geometric distribution to gene loss rates in seabuckthorn as to the silver birch. (b) Fitting a geometric distribution to gene loss rates in seabuckthorn as to the peach. (c) Fitting a geometric distribution to gene loss rates in seabuckthorn as to the grape. (d) Fitting a geometric distribution to gene loss rates in seabuckthorn as to the jujube.

**Figure S14** (a) The correspondence between genomes of seabuckthorn and jujube using homologous gene dot-plots. (b) The correspondence between genomes of seabuckthorn and grape using homologous gene dot-plots. (c) The correspondence between genomes of jujube and seabuckthorn using homologous gene dot-plots shows the seven chromosomes of the ancestors of eudicytodes. (d) The correspondence between genomes of seabuckthorn and jujube using homologous gene dot-plots shows 14 chromosomes of the ancestors of jujube and seabuckthorn.

**Figure S15** The relative expression of FAD2 and LOX in seabuckthorn seeds verified by q-PCR.

**Figure S16** The expression levels of the genes involved in ASA biosynthesis in leaf, root, stem, pulp (T1, T2, T3), and seeds (T1, T2, T3).

**Figure S17** The relative expression of GGALPP and GALDH in seabuckthorn pulp verified by q-PCR.

Table S1 Summary of seabuckthorn genome sequencing data.

Table S2 Statistics of sequencing data obtained by illumina Hiseq platform for seabuckthorn genome survey.

Table S3 K-mer statistics of the genomic characteristics of seabuckthorn obtained by genome survey analysis.

Table S4 Statistics of sequencing data of seabuckthorn by Pacbio sequel platform.

Table S5 Detail of Illumina sequencing.

Table S6 The summary of preliminary assembly of the seabuckthorn genome.

Table S7 Statistics of seabuckthorn genome sequencing data used for Hi-C obtained by Illumine Hiseq platform.

Table S8 The quality assessment of the Hi-C construction library.

Table S9 The assembled length of each chromosome of seabuckthorn genome before artificially corrected based on Hi-C heat map.
Table S10 The assembled length/gene number of each chromosome of seabuckthorn genome after artificially corrected based on Hi-C heat map.
Table S11 Seabuckthorn genome sequencing data derived from short-read sequencing.
Table S12 Analysis of the seabuckthorn genome with CEGMA v2.5.
Table S13 Analysis of the seabuckthorn genome with Benchmarking Universal Single-Copy Orthologs.
Table S14 Correlation coefficient between genetic map and assembly.
Table S15 The statistics of repeat sequences in seabuckthorn genome.
Table S16 The result of gene prediction in seabuckthorn genome.
Table S17 The statistical of gene prediction in seabuckthorn genome.
Table S18 Gene annotation details.
Table S19 RNA details.
Table S20 Gene families clustered by OrthoMCL in 11 species.
Table S21 KEGG enrichment analysis of the expanded gene families of seabuckthorn genome.
Table S22 KEGG enrichment analysis of the contracted gene families of seabuckthorn genome.
Table S23 Information of genomic data.
Table S24 Number of homologous blocks and gene pairs within a genome or between genomes.
Table S25 Number of homologous genes residing in inferred collinear gene blocks within a genome or between genomes.
Table S26 Orthologous genomic regions between grape and seabuckthorn.
Table S27 Orthologous genomic regions between jujube and seabuckthorn.
Table S28 Orthologous genomic regions between birch and seabuckthorn.
Table S29 Orthologous genomic regions between peach and seabuckthorn.
Table S30 Number of paralogous gene within genome, orthologous, and out-paralogous gene pairs with related genomes.
Table S31 Protein table listing the homologous gene sets between genomes. A dot (.) is placed where no homolog is identified in the respective genome.

Table S32 Kernel function analysis of $Ks$ distribution related to duplication events within each genome and between genomes (before evolutionary rate correction).
Table S33 The gene pairs related to the three WGD events in seabuckthorn genome.
Table S34 Kernel function analysis of $Ks$ distribution related to duplication events within each genome and between genomes (after evolutionary rate correction).
Table S35 KEGG enrichment analysis of the $a$-event-related gene set of seabuckthorn genome.
Table S36 KEGG enrichment analysis of the $b$-event-related gene set of seabuckthorn genome.
Table S37 *Hippophae rhamnoides* gene loss and gene translocation rates with *Vitis vinifera* as reference genome.
Table S38 *Hippophae rhamnoides* gene loss and gene translocation rates with *Ziziphus jujube* as reference genome.
Table S39 *Hippophae rhamnoides* gene loss and gene translocation rates with *Prunus persica* as reference genome.
Table S40 *Hippophae rhamnoides* gene loss and gene translocation rates with *Betula pendula* as reference genome.
Table S41 Consecutive gene removal in seabuckthorn compared with the reference genomes.
Table S42 The observed distribution of gene loss and translocation numbers fitted by using different density curves of geometry distribution.
Table S43 The summary of data quality for different tissues and stages of RNA-seq in seabuckthorn.
Table S44 Ascorbic acid and fatty acid-related gene expression values of different tissues (leaf, root, stem, fruit, and seed) in different stages (T1, T2, T3) of seabuckthorn by RNA-seq (corresponding to Figure 5e, f).
Table S45 The information of the 55 seabuckthorn accessions.
Table S46 Genes under putatively selected genomic regions for wild Hs and cultivated Hm identified by three selective sweep scan approaches.
Table S47 GO enrichment analysis of genes under putatively selected genomic regions for wild Hs identified by three selective sweep scan approaches.