Comparative Transcriptomics of Strawberries (*Fragaria* spp.) Provides Insights into Evolutionary Patterns

Qin Qiao[1], Li Xue[2], Qia Wang[3], Hang Sun[3], Yang Zhong[4], Jinling Huang[5], Jiajun Lei[2]* and Ticao Zhang[3]*

[1] School of Agriculture, Yunnan University, Kunming, China, [2] College of Horticulture, Shenyang Agricultural University, Shenyang, China, [3] Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, China, [4] Key Laboratory for Biodiversity Science and Ecological Engineering, School of Life Sciences, Fudan University, Shanghai, China, [5] Department of Biology, East Carolina University, Greenville, NC, USA

Multiple closely related species with genomic sequences provide an ideal system for studies on comparative and evolutionary genomics, as well as the mechanism of speciation. The whole genome sequences of six strawberry species (*Fragaria* spp.) have been released, which provide one of the richest genomic resources of any plant genus. In this study, we first generated seven transcriptome sequences of *Fragaria* species *de novo*, with a total of 48,557–82,537 unigenes per species. Combined with 13 other species genomes in Rosales, we reconstructed a phylogenetic tree at the genomic level. The phylogenic tree shows that *Fragaria* closed grouped with *Rubus* and the *Fragaria* clade is divided into three subclades. East Asian species appeared in every subclade, suggesting that the genus originated in this area at ∼7.99 Mya. Four species found in mountains of Southwest China originated at ∼3.98 Mya, suggesting that rapid speciation occurred to adapt to changing environments following the uplift of the Qinghai–Tibet Plateau. Moreover, we identified 510 very significantly positively selected genes in the cultivated species *F.* × *ananassa* genome. This set of genes was enriched in functions related to specific agronomic traits, such as carbon metabolism and plant hormone signal transduction processes, which are directly related to fruit quality and flavor. These findings illustrate comprehensive evolutionary patterns in *Fragaria* and the genetic basis of fruit domestication of cultivated strawberry at the genomic/transcriptomic level.

Keywords: *Fragaria*, transcriptome, evolutionary pattern, comparative genomics, positive selection

INTRODUCTION

Multiple closely related species with genomic sequences provide an ideal system for studies on comparative and evolutionary genomics, as well as the mechanism of speciation. However, only a few genera in eukaryotes, e.g., *Drosophila* (Clark et al., 2007) and *Oryza* (Zhang et al., 2014), contain multiple genome sequenced relatives because it is expensive to sequence whole genomes at present. Fortunately, transcriptome sequencing (using RNA-seq technology) provides a rapid and effective approach to obtain large numbers of protein-coding genes, which can be used for

understanding evolutionary questions in non-model organisms (Ward et al., 2012; Haas et al., 2013). Recently, comparative transcriptome studies between multiple close relatives have been used to reveal evolutionary patterns in domesticated and wild tomato (Koenig et al., 2013), and for phylogenomic analysis in Caryophyllales (Yang et al., 2015).

The genus *Fragaria* L. (Rosaceae) contains about 25 species, which are most spring-blooming, insect-pollinated herbaceous perennials capable of clonal growth, and have animal-dispersed fleshy fruits (Johnson et al., 2014; Liston et al., 2014). This genus has diversified mating systems and ploidy ($2\times$, $4\times$, $6\times$, $8\times$, and $10\times$; Liston et al., 2014). Cultivated strawberry, *F.* $\times$ *ananassa* ($2n = 8\times = 56$), is an important fruit crop, which originated via octoploid hybridization between *F. virginiana* and *F. chiloensis* in 18th century in Europe. Like natural selection, domestication has influenced the evolutionary pressures on the cultivated strawberry. The geographic distribution of *Fragaria* spans a wide range of Eurasia, North and South America and East Asia (Johnson et al., 2014; Liston et al., 2014). Of these, East Asia is the center of genetic diversity and the most likely original range of *Fragaria* (Njuguna et al., 2013). In addition, dated molecular phylogenies based on whole chloroplast genomes indicated that the genus is quite young (1.0–4.1 Mya; Njuguna et al., 2013). Obviously, this younger genus must have undergone rapid spread and species differentiation to adapt to different environments.

Previous studies have obtained the whole genome sequences of one cultivated (*F.* $\times$ *ananassa*) and five wild strawberries (*F. vesca*, *F. iinumae*, *F. nipponica*, *F. nubicola*, and *F. orientalis*; Shulaev et al., 2011; Hirakawa et al., 2014), which provide one of the richest genomic resources of any plant genus. Most diploid species in *Fragaria* have very small genomes (~200 Mb, **Table 1**) compared to other higher plants. Recently evidence from linkage maps method indicated that four subgenomes of the allo-octoploid cultivated strawberry *F.* $\times$ *ananassa* were originated with diploid donors of *F. vesca*, *F.iinumae* and two *F. iinumae*-like ancestors (Tennessen et al., 2014; Sargent et al., 2016). A virtual diploid reference genome (173.2 Mb, contains 45,377 genes) of the cultivated strawberry was constructed by integrating the sequences of the four homoeologous subgenomes of *F.* $\times$ *ananassa* (in total 697.8 Mb, contains 230,838 genes), which eliminated both heterozygous regions as well as homeologous differences (i.e., differences between subgenomes) using overlap graph method, Hirakawa et al., 2014). Except for *Fragaria*, the genome sequences of other important Rosaceae fruit crops are also currently available, e.g., apple (*Malus domestica*), pear (*Pyrus communis*), peach (*Prunus persica*), Chinese plum (*P. mume*) and black raspberry (*Rubus occidentalis*). These abundant genome resources make it possible to improve our understanding of the genetics of horticultural traits by comparative genomics, including fruit quality, timing of flowering, and pest and disease resistance. In short, *Fragaria* is a very attractive model genus for the study of recent speciation, molecular ecology, evolutionary genomics and fruit development (Liston et al., 2014).

In this study, we performed RNA-seq technology to obtain most transcript sequences of seven diploid and tetraploid *Fragaria* species (**Tables 1** and **2**). Combining these with six *Fragaria* species genomes (including the virtual diploid genome of *F.* $\times$ *ananassa*) and seven other species which also have sequenced genomes in Rosales, we first reconstructed a phylogenetic tree based on single copy genes at the genomic level and estimated the divergence time of each node. Subsequently, the substitution rates were estimated in eight diploid strawberries to identify the rate of evolution in each species. Finally, positively selected genes (PSGs) related to agronomic characters in the cultivated strawberry were identified using evolutionary genomics. We aim to gain a comprehensive understanding of the evolutionary patterns in *Fragaria* at the genomic/transcriptomic level.

MATERIALS AND METHODS

Sample Collection and Transcriptome Sequencing

Seven wild strawberries, i.e., *F. nilgerrensis* (FNG), *F. pentaphylla* (FPE), *F. mandshurica* (FMA), *F. viridis* (FVI), *F. corymbosa* (FCO), *F. moupinensis* (FMO), and *F. tibetica* (FTI), which collected from East Asia were cultivated in cultivation base of Shenyang Agricultural University. In order to obtain more expression transcripts, different organs (leaves, stems, flowers, and fruits) of each species were sampled and stored at $-80°C$. Among these species, except the FNG were sequenced using two individuals separately, other six species were sequenced using one individual with pooled different tissues.

High quality total RNA was extracted using the TRIZOL reagent (Sigma-Aldrich) following the manufacturer's instructions. A total amount of 3 µg RNA per species was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext Ultra™ RNA Library Prep Kit for Illumina Inc. (NEB, USA) following manufacturer's recommendations. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina Inc.). After cluster generation, the library preparations were sequenced on an Illumina Hiseq 2500 platform and paired-end reads were generated. The whole step of library construction and Illumina sequencing was performed at Novogene Bioinformatics Technology Co., Ltd (Beijing, China).

De novo Assembly and Functional Annotation

Clean data were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. For FNG, the sequenced left files (read1 files) from two individuals were pooled into one big left.fq file, and right files (read2 files) into one big right.fq file. Transcriptome assembly was accomplished based on the left.fq and right.fq using Trinity program (trinityrnaseq_r20140413) with minimum k-mer coverage set to 2 and all other parameters set by default (Grabherr et al., 2011).

Functional annotations of all assembled unigenes were conducted by searching against the following databases:

TABLE 1 | Statistics of study materials and gene numbers in *Fragaria*.

Species	Abbreviation	Ploidy	Mating system	Genome size (Mb)	Number of genes	Reference
F. vesca	FVE	2x	Hermaphrodite, SC	220.2	34809	Shulaev et al., 2011
F. nipponica	FNI	2x	Hermaphrodite, SI	206.4	87803	Hirakawa et al., 2014
F. iinumae	FII	2x	Hermaphrodite, SC	199.6	76760	Hirakawa et al., 2014
F. nubicola	FNU	2x	Hermaphrodite, SI	203.7	85062	Hirakawa et al., 2014
F. nilgerrensis	FNG	2x	Hermaphrodite, SC	274 ± 3.8*	82537	This study
F. pentaphylla	FPE	2x	Hermaphrodite, SI	268 ± 3.7*	53748	This study
F. mandshurica	FMA	2x	Hermaphrodite, SI	241 ± 4.3*	55911	This study
F. viridis	FVI	2x	Hermaphrodite, SI	230 ± 2.8*	48557	This study
F. orientalis	FOR	4x	Dioecious	214.2	99674	Hirakawa et al., 2014
F. corymbosa	FCO	4x	Dioecious	NA	62587	This study
F. moupinensis	FMO	4x	Dioecious	245 ± 3.3*	52413	This study
F. tibetica	FTI	4x	Dioecious	NA	54022	This study
F. × ananassa	–	8x	Subdioecious	697.8	230838	Hirakawa et al., 2014
FANhybrid	FAN	2x (virtual diploid genome)		173.2	45377	Hirakawa et al., 2014

SC, self-compatible; SI, self-incompatible. *referred from Chen et al. (2015).

NCBI non-redundant protein (Nr), NCBI non-redundant nucleotide (Nt), Protein family (Pfam), Clusters of Orthologous Groups of proteins (KOG), Swiss-Prot protein (Swiss-Prot), KEGG Ortholog (KO), and Gene Ontology (GO) database.

Orthologous Genes Identified and Phylogenetic Analysis

The phylogenetic tree of 18 species in Rosaceae was constructed with *Cannabis sativa* and *Morus notabilis* as outgroups. First, to define a set of conserved genes for cross-taxa comparison, we used OrthoMCL software (Li et al., 2003) to identify homologous gene clusters (orthogroups) among the twenty genomes. Genes with lengths less than 50 amino acids were excluded. OrthoMCL was run with an *e*-value cut-off of 1*e*−15 and an inflation parameter of 2.0. Then, orthogroups with only single copy genes (one-to-one orthologs) that were shared by all 20 genomes were retained for further analysis. Each orthogroups was aligned using MUSCLE v3.8.31 (Edgar, 2004) with default parameters. The poorly aligned regions were further strictly trimmed by using the trimAl v1.4 software (Capella-Gutiérrez et al., 2009) with the parameter "-gt 0.8 -st 0.001." Alignments of all orthogroups were concatenated by our python script (**Supporting Information File S1**). Then maximum likelihood (ML) trees were generated using RAxML v7.0.4 (Stamatakis et al., 2005) with PROTCATJTT model, the ML criteria. The divergence time between all lineages of Rosaceae was estimated using the effectual newly developed MEGA7 software (Kumar et al., 2016). The timetree inferred using the Reltime method (Tamura et al., 2012) and the JTT matrix-based model (Jones et al., 1992). The timetree was computed using two clock calibration constraints. We calibrated with a 53.9–73.9 Mya divergence for Rosoideae vs. (Maloideae and Prunoideae) (Njuguna et al., 2013). The origin time of the common ancestor of Malus, Pyrus and Prunus has been dated to 28.7–56.1 Mya (Liu et al., 2014b).

Positive Selection Analysis

Genome/transcriptome sequences of seven diploid *Fragaria* species and FAN were selected to conduct the positive selection analysis. The orthogroups of eight species were identified using the same methods which described in above phylogenetic analysis. In the positive selection analysis, only single copy genes shared by eight species were considered. We prepared the amino acid and codon-based nucleotide sequences in each orthogroups separately from eight species. The amino acid sequences of each orthogroups was aligned using MUSCLE v3.8.31 (Edgar, 2004) with default parameters. Then, to calculate the nonsynonymous (Ka) and synonymous (Ks) substitution rates between pairs of orthogroups, the corresponding multiple codon-based nucleotide alignment from the above aligned amino acid sequences were constructed by PAL2NAL (Suyama et al., 2006). For each alignment, a gene tree was constructed by RAxML (Stamatakis et al., 2005) using GTR+GAMMA model.

To estimate lineage-specific evolutionary rates for each branch of the eight species, the codeml program in the PAML 4 package (Yang, 2007) with the free-ratio model (model = 1) was run on each orthogroups. We conducted the boxplot analysis using the Ka/Ks ratio derived from free-ratio model results and filtered $Ka/Ks < 0.0001$ or $Ka/Ks > 3$. Significances of the deviations from the median Ka/Ks ratio between eight species branches were detected using Wilcoxon rank sum test. We also established frequency distribution plots of all Ka/Ks ratios of eight species.

To increase the power of our tests for positive selection, we applied the improved branch-site model (Zhang et al., 2005) implemented in codeml program (Yang, 2007) to estimate the Ka/Ks substitution rates (ω value). We also deleted all gaps (clean data = 1) from the alignments to lower the effect of ambiguous bases on the inference of positive selection. A foreground branch was specified as the clade of FAN. A significant likelihood ratio test (LRT) was conducted to determine whether positive selection is operating in the foreground branch. In this study, the extremely significant PSGs were inferred if the *P*-value was less than 0.01.

For PSGs in FAN, functional annotation was inferred by Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). GO enrichment analyses of PSGs were conducted using web-based agriGO[1] (Du et al., 2010) with singular enrichment analysis (SEA) method and TAIR10 database. The KOBAS software (Xie et al., 2011) was also used to test the statistical enrichment of PSGs in KEGG pathways (Kanehisa and Goto, 2000).

RESULTS

Transcriptome Assembly and Annotation

After quality filtering, we generated 59.29 (FVI) – 107.94 (FNG; Abbreviation of species names see **Table 1**) million clean reads for a total of 8.89 (FVI) – 13.48 (FNG) Gb of RNA-seq data from each of seven species (**Table 2**). The clean data were deposited in the NCBI Sequence Reads Archive (SRA) database (no. SRR4030219). *De novo* assembly of the high-quality reads generated 48,557 (FVI) – 82,537 (FNG) unigenes (the longest transcript in one gene; **Tables 1** and **2**). Median unigene lengths were 399 (FCO) – 459 (FMO) bp, with N50 lengths of 1,405 (FNG) – 1,678 (FVI) bp. There were 24,978 (FVI) – 35,166 (FNG) unigenes with homologs in Nr databases on the basis of similarity. A total of 30,825 (FVI) – 45,137 (FNG) unigenes were successfully annotated in at least one database (a total of seven databases, see Materials and Methods) with a significant match (e-value $< 10^{-5}$; Supplementary Table S1). Most top species classification hit in the Nr database is *F. vesca* ($>21,000$ hits; Supplementary Table S2).

Phylogenetic Analysis and Divergence Date Estimation in Rosales

Orthologs are genes that have evolved from a common ancestral gene via speciation (Fitch, 1970). Based on 20 genome and transcriptome sequences of Rosales, a total of 63,122 orthogroups were detected, including 515,610 genes. Among these, 4,690 orthogroups are shared by all species, and 276 orthogroups contain putative 1:1:1 single copy genes. Subsequently, we constructed a phylogenetic tree based on 87,916 amino acid sequences of the trimmed and concatenated 276 single copy gene

---

[1]bioinfo.cau.edu.cn/agriGO

alignments from 20 species (**Figure 1**). Using Reltime method and JTT matrix-based model, the estimated log likelihood value is -487560.5083. All positions with less than 90% site coverage were eliminated. There were a total of 63796 positions in the final dataset. The phylogenic tree shows that *Fragaria* closed grouped with *Rubus* and the *Fragaria* clade is divided into three subclades: an East Asian clade (FTI, FMO, FPE, FCO, FNG, and FVI), a Euro-Asian clade (FVE, FMA, FOR and FNU) and a cultivated strawberry clade (FAN and FII). Divergence time estimation suggests that *Fragaria* originated $\sim$7.99 Mya and split from Rubus $\sim$33.34 Mya, which combined diverged from the clade containing *Malus*, *Pyrus,* and *Prunus* $\sim$63.90 Mya. The species distributed in Southwest China (FTI, FMO, FPE, and FCO) split from the species in Japan (FNI) $\sim$4.66 Mya.

Substitution Rates Estimated in Eight Strawberries

Based on genome and transcriptome sequences of seven diploid *Fragaria* species and FAN, a total of 48,643 orthogroups were detected, including 240,023 genes. Among these, 8,593 orthogroups are shared by all eight species, and 4,487 orthogroups contain putative 1:1:1 single copy genes. Using these genes, we estimated the substitution rates for each orthogroup using the free-ratio model in PAML, which allows an independent Ka/Ks ratio for each branch (Yang, 2007). The boxplot result showed that the FVI has the smallest Ka/Ks median value (0.367), and the FMA has the biggest median (0.577). The median of FAN is 0.520, which adjacent with FII (0.435) and FVE (0.562). All the significance level of Wilcoxon rank sum between two neighboring species are less than 0.0001 (**Figure 2A**). For each species, most Ka/Ks ratios were in the ranges 0.1–0.2 and 0.5–1.0 (**Figure 2B**). The frequency distribution of Ka/Ks ratios shows that FAN has more genes (629, Supplementary Table S5) with elevated Ka/Ks ratios ($1.0 > Ka/Ks > 0.5$) than other species (328–489 genes; **Figure 2B**).

Positive Selection Analysis in Cultivated Strawberry (*F.* × *ananassa*)

For 4,487 single copy orthogroups in eight diploid species, the branch-site model of the PAML 4 package (Yang, 2007) was used to detect genes in FAN with signals of positive

TABLE 2 | Statistics of clean data and unigenes in this study.

Species	Statistics of clean data						Frequency distribution of unigenes' lengths				
	Clean reads	Clean bases	Error (%)	Q20 (%)	Q30 (%)	GC (%)	Total number	Minimum Length	Median Length	Maximum Length	N50
FNG	107941470	13.48G	0.04	94.27	89.13	45.49	82537	201	412	16895	1405
FPE	68356166	10.25G	0.02	96.82	91.66	46.04	53748	201	443	16152	1655
FMA	80770294	12.12G	0.02	96.93	91.87	45.98	55911	201	435	16993	1672
FVI	59286810	8.89G	0.02	96.67	91.31	46.27	48557	201	456	15622	1678
FCO	83949596	12.59G	0.02	96.71	91.35	45.95	62587	201	399	16490	1563
FMO	71829414	10.77G	0.02	96.46	90.85	46.03	52413	201	459	15581	1655
FTI	86674308	13.0G	0.02	96.83	91.61	46.15	54022	201	449	15585	1638

FIGURE 1 | Phylogenetics and divergence time among Rosales.
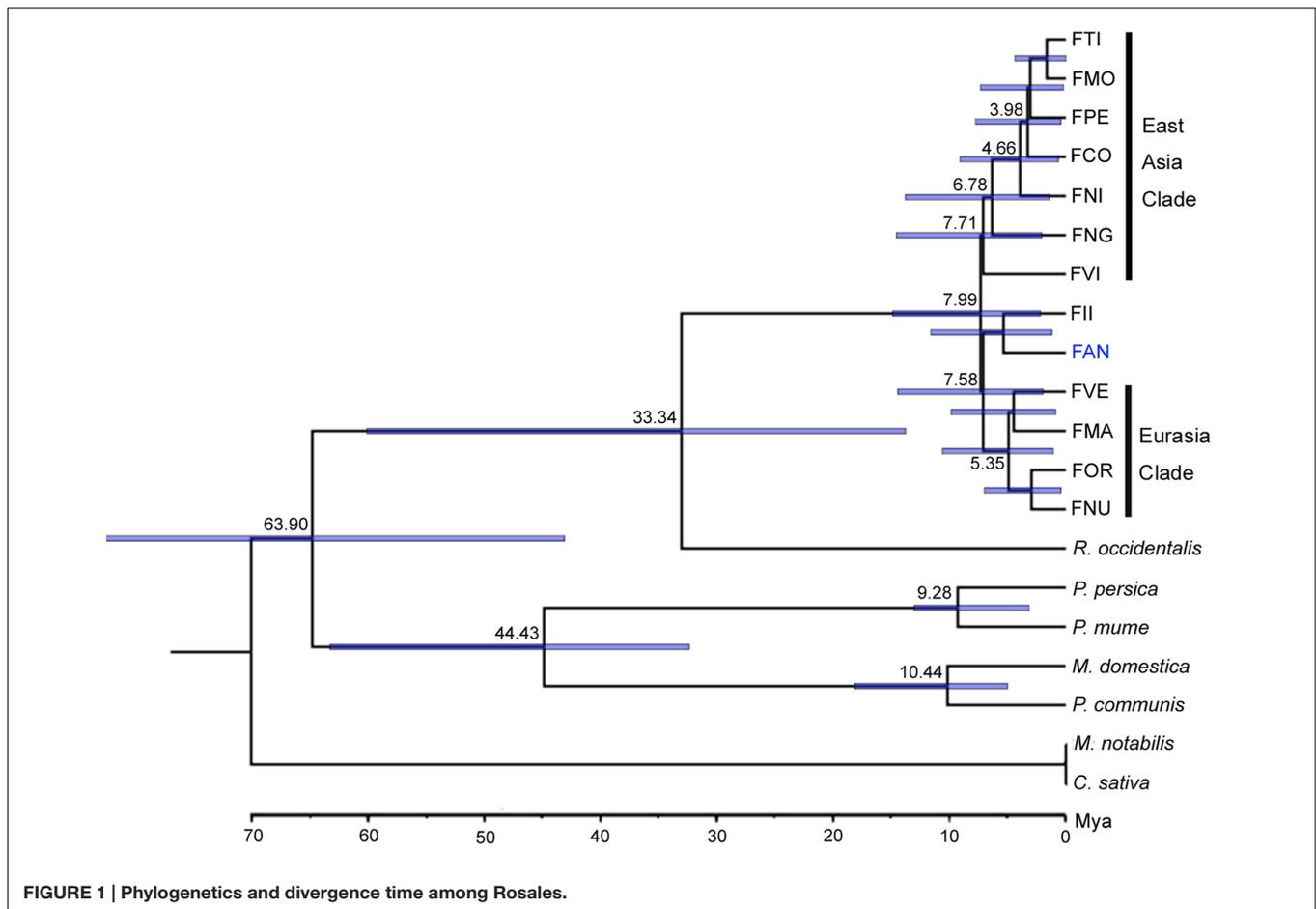
selection. As a result, 1,924 genes possibly under positive selection were identified in the FAN genome ($Ka/Ks > 1$); of these genes, 564 show significant evidence of positive selection (P-value < 0.05), and 510 PSGs show extremely significant evidence of having undergone positive selection (P-value < 0.01; Supplementary Table S3). We also conducted KEGG functional classification (**Figure 3**, Supplementary Table S4) for these PSGs in FAN. The distribution of KEGG classifications of PSGs shows that of all the annotated categories, metabolic processes have the most hits (24 PSGs; **Figure 3**). Carbon fixation in photosynthetic organisms was enriched, with four PSGs encoding ribulose bisphosphate carboxylase (Rubisco), ribulose-phosphate 3-epimerase (RPE), triosephosphate isomerase (TPI), and fructose-1,6-bisphosphatase (FBP). The biosynthesis of secondary metabolites was also enriched, with 13 genes, such as genes encoding 4-coumarate–CoA ligase-like 6, which participates in ubiquinone and flavanone biosynthesis. Moreover, the plant hormone signal transduction process was also enriched, with eight PSGs, including genes encoding auxin-responsive proteins IAA11, IAA17, and IAA31, protein phosphatase 2C (PP2C), abscisic acid-insensitive 5-like (ABI5-like) proteins, transcription factor TGA4-like isoform X1, and coronatine-insensitive protein 1 (COI1).

DISCUSSION

Our phylogenetic results based on nuclear genomes mostly agree with previous relationships determined using two nuclear genes or whole chloroplast genomes (Rousseau-Gueutin et al., 2009; Njuguna et al., 2013). The topology of major clades in the *Fragaria* genus is divided into three subclades with high support (100%), a Eurasian clade, an FII-FAN clade and an East Asian clade, which roughly correspond to clades A, B, and C, respectively, in the phylogenetic tree based on chloroplast genomes (Njuguna et al., 2013). East Asian species appear in all clades, indicating that this area is the center of origin of *Fragaria*, which is consistent with previous findings (Njuguna et al., 2013). However, there are some inconsistencies between our results and previous studies. The placement of FVI and FNG is grouped with East Asian clade with moderate support (70%) in our analysis, which has been unresolved in previous studies (Njuguna et al., 2013). In addition, FNU was incomprehensibly clustered with FOR in the Eurasian clade, which conflicts with the accepted idea that FNU is close to the East Asian clade (Rousseau-Gueutin et al., 2009). To be prudent, our natural selection analysis did not include this species' genome sequence. Our analysis supported FAN clustering with FII rather than with Eurasian species. One reason might be that our analysis did not include
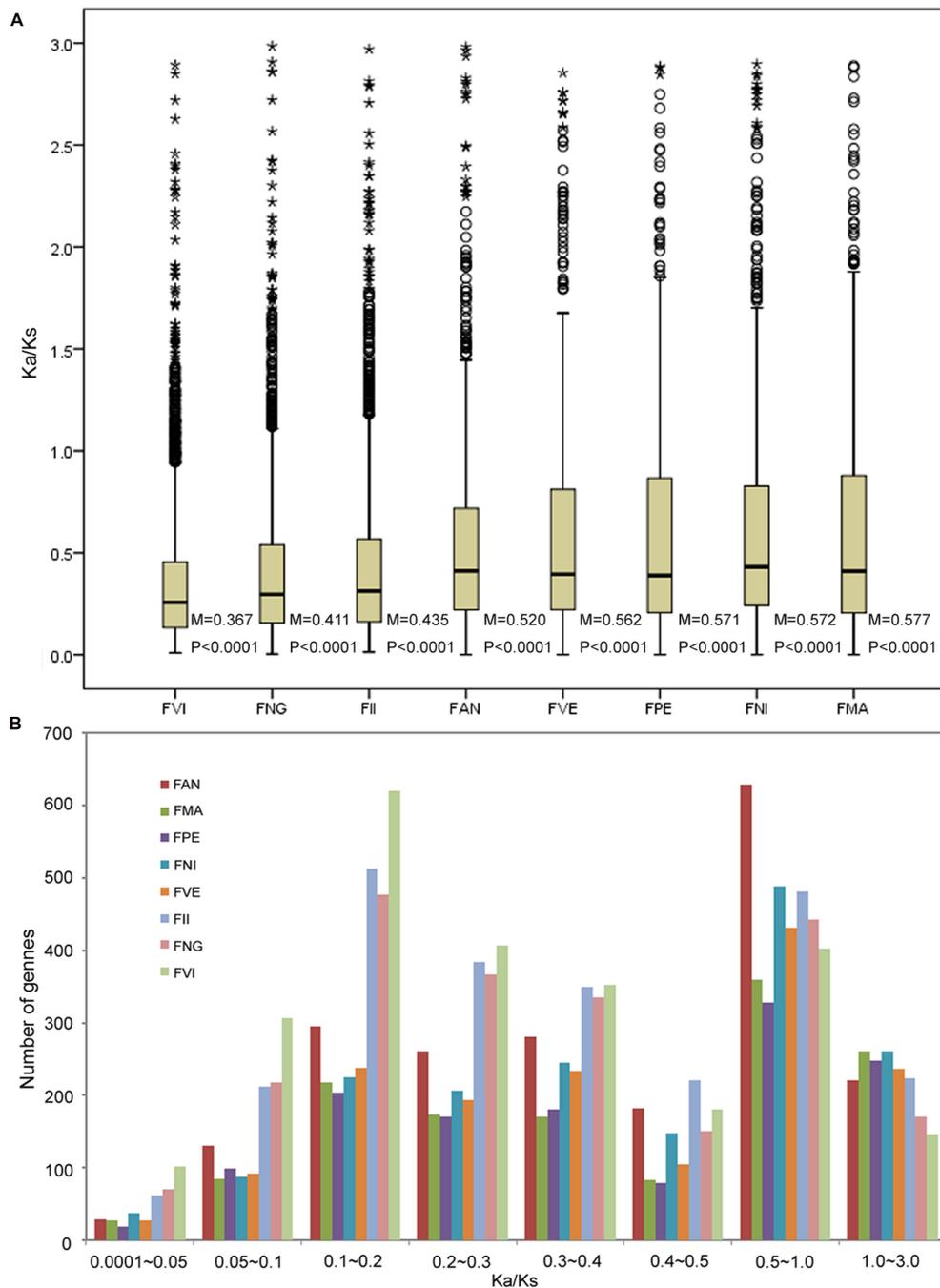
FIGURE 2 | Number of orthologs with given *dN/dS* ratios for the eight species. (A) Boxplots of dN/dS ratios for each species. The median (M) and significances (P) of the deviations using Wilcoxon rank sum test are also showed. **(B)** Frequency distribution of dN/dS ratios of orthologs for the eight species.

the octoploid cross parents of cultivated FAN, i.e., *F. virginiana* and *F. chiloensis*, which most likely derived from diploid FVE or FMA by tetraploidization (Rousseau-Gueutin et al., 2009). Another reason is that FII is an important diploid donor of allopolyploid FAN according to resent studies, which indicated that four subgenomes of allo-octoploid cultivated strawberry derived from FVE, FII, and two with an unknown ancestor

close to FII (Tennessen et al., 2014; Sargent et al., 2016). Since the virtual diploid genome of FAN was created by the overlap graph method, which is essentially a majority-rule consensus process, it should closely resemble the FII clade which provides the majority of the subgenomes. Therefore, it is understandable that our results show a close phylogenetic association between FAN and FII.

FIGURE 3 | Distribution of KEGG classifications of PSGs in FAN.
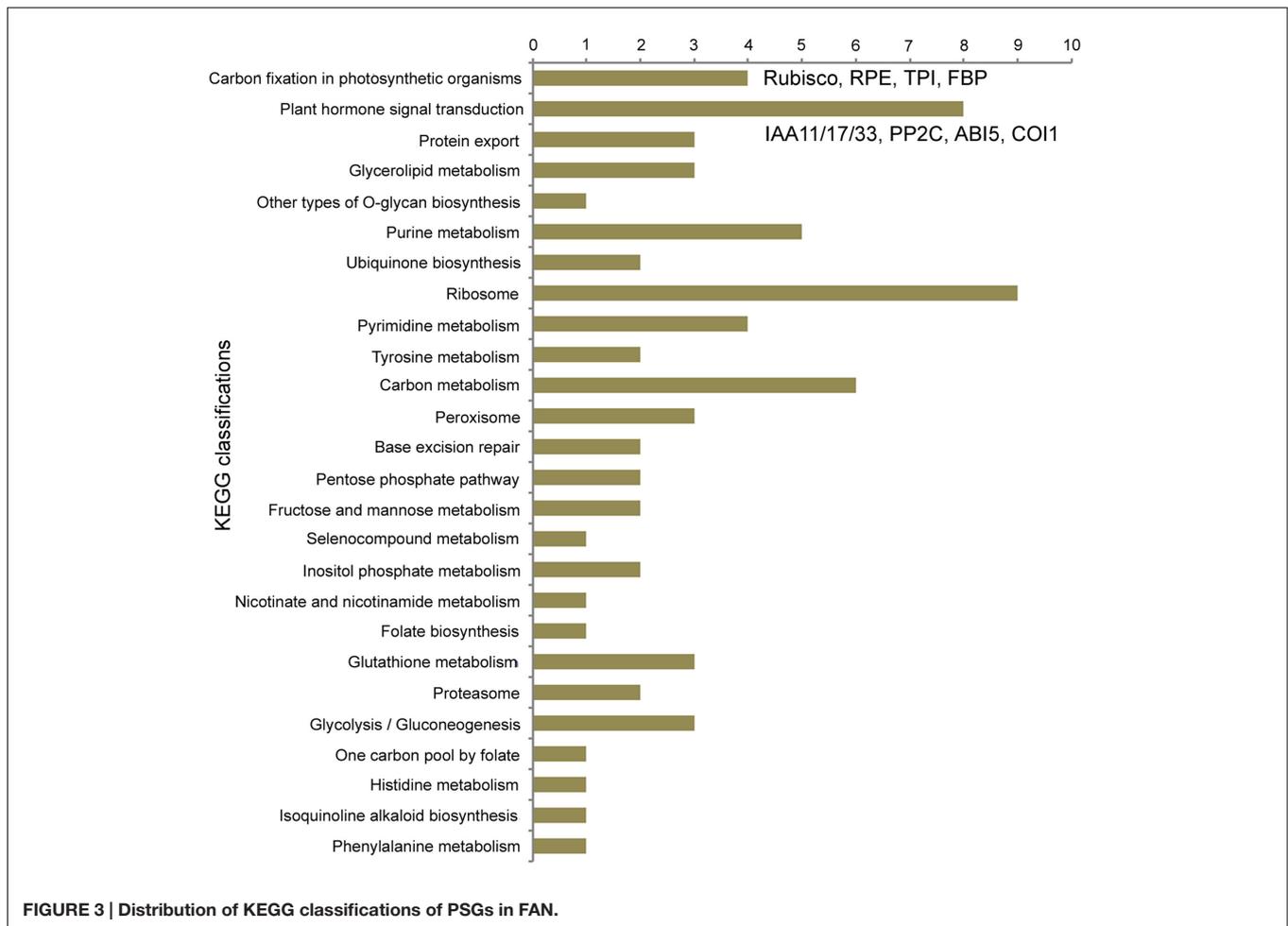
The appropriate calibration is important to the results of phylogenetic dating. In order to reduce the error, we have used two calibration constraints from two credible references (Njuguna et al., 2013; Liu et al., 2014b). From our dating results, we infer that the origin of *Fragaria* dates (∼7.99 Mya) from the Miocene, moderate earlier than the estimate (1.52–4.44 Mya) based on chloroplast genomes (Njuguna et al., 2013). The major reason might be that the genome sequences of the closest genus, *Potentilla*, were unavailable and not included in our analysis. Notably, the clade including four species (FTI, FMO, FPE, and FCO) found in the mountains of southwest China is inferred to have originated ∼3.98 Mya, which is in accordance with the time (3.6 Mya) of the strongest Qinghai–Tibet Plateau (QTP) rapid uplift in the Pliocene (Li and Fang, 1999). It suggests that the uplift of the QTP is a driving force for rapid speciation in this clade of Fragaria. That is, ecological adaptation has occurred within *Fragaria* following climatic and environmental changes after the QTP uplift, which promoted the speciation of these species. As has been reported for many other taxonomic groups, the QTP uplift is usually responsible for the initial isolation and subsequent climatic shifts and has driven the species divergence (Liu et al., 2012, 2014a; Zhou et al., 2012; Wen et al., 2014; Favre et al., 2015).

To investigate selective pressures in eight diploid species, we estimated substitution rates using a free model for each orthogroup. For all species, Ka/Ks ratios of most orthogroups were less than 0.5, indicating that purifying selection is the main driving force in *Fragaria* evolution, acting to remove deleterious mutations. The frequency distribution of Ka/Ks clearly showed that the cultivated species FAN has more genes with elevated interval values ($Ka/Ks = 0.5$–1.0) than the other seven species. The accelerated evolution of genes is often driven by positive selection, relaxed selection or low effective population size (Ne). During cultivation or domestication, both selection pressures for fitness traits important in the wild species and effective population size are reduced (Maceachern et al., 2009). Therefore, reduced selective constraint and reduced Ne could be responsible for the observed elevated evolutionary ratios in cultivated FAN with more elevated interval values ($Ka/Ks = 0.5$–1.0). While the similar number of positive selected genes ($Ka/Ks > 1.0$) between domesticated and wild *Fragaria* may be caused by artificial selection and natural selection respectively. This finding is compatible with branch-site model test for positive selection analysis, which indicated that many genes related to specific agronomic traits in FAN show significant positive selection. Moreover, the median of substitution rates (ka/ks ratio) of FAN

is bigger than that of FII while smaller than FVE in free model (**Figure 2A**). It is noted that the substitution rates median of FAN is closer to FVE than to FII. As described above, the FAN is a virtual genome constructed from an allopolyploid via the overlap layout consensus method; we cannot rule out loci that evolved fast in FAN may actually be loci obtain from FVE rather than FII. Previous study also suggested that extensive unidirectional introgression has converted FII-like subgenomes to be more FVE-like among subgenomes (Tennessen et al., 2014). The reason might be homoploid hybridization in the FII-like diploid ancestors or strong selection spreading FVE-like sequence within the octoploid genome.

For a cultivated strawberry, many domestication-related genes in FAN involving traits desirable to humans should evolve rapidly under artificial selection, alongside natural selection caused by environmental factors. During the domestication process, fruit quality and flavor are the most important agronomic traits for strawberries. Carbohydrates, produced by photosynthesis in plants, are the basic substances of fruit growth and development and are closely related to fruit quality. In our KEGG enrichment analysis for PSGs in FAN, the "carbon fixation in photosynthetic organisms" pathway was the most significant hit. In this pathway, all four detected PSGs encode photosynthesis-related proteins (Rubisco, RPE, TPI and FBP). Rubisco is an essential and rate-limiting enzyme involved in the first step of carbon fixation and subsequently fixed into sugars (Chan et al., 1990). RPE catalyzes the reversible epimerization of D-ribulose 5-phosphate to D-xylulose 5-phosphate. TPI is involved in the gluconeogenesis and glycolysis pathways. Chloroplastic FBPase is a rate-limiting enzyme in the Calvin cycle, which plays a critical role in carbohydrate biosynthesis. A previous study reported that increasing the level of TPI and FBP aldolase in *Anabaena* could enhance photosynthetic activity and increase the rate of cell growth (Ma et al., 2007). We can speculate that as a trait desirable to humans, fruit quality, which can be described as the amount and composition of carbohydrates accumulated in the fruit, could undergo significant artificial selection during the domestication process. These key enzymes involved in carbon fixation and carbohydrate biosynthesis pathways show significantly positive selection further confirms this speculation.

Fruit quality and flavor are also closely associated with plant hormone biosynthesis and signal transduction (Obroucheva, 2014). Notably, the second most significantly enriched KEGG process in our results is plant hormone signal transduction, which includes eight PSGs in FAN. Among these PSGs, three homologs have similar functions in the auxin signaling process. These genes encode auxin-responsive proteins IAA11, IAA17, and IAA31, which function as repressors of early auxin response genes at low auxin concentrations (Liscum and Reed, 2002). A previous study based on an early stage fruit transcriptome in FVE revealed that *IAA31* was significantly up-regulated (6.5-fold) in seeds by comparing post-fertilization and pre-fertilization stages, which suggested that auxin exerts a direct role in stimulating receptacle expansion during fruit development (Kang et al., 2013). We speculate that these PSGs involved in auxin signaling might play

specific functions in early fruit development in the cultivated strawberry.

The strawberry is a non-climacteric fruit, whose ripening is directly stimulated by abscisic acid (ABA) rather than ethylene (Jia et al., 2011; Obroucheva, 2014). ABA signaling via PYR/PYL receptors begins with PP2Cs, which permit activation of SNF1-related protein kinase 2, and then regulate downstream targets such as ABA-response transcription factors. The signaling pathway of ABA involves three PSGs, including two abscisic acid-insensitive 5-like (ABI5-like) proteins, and PP2C 37 (PP2CA). Among these, PP2CA acts as a key negative regulator of ABA responses to plant growth and development, inhibition of vegetative growth, and response to glucose (Yoshida et al., 2006; Park and Cutler, 2009). For example, PP2C is highly expressed during cucumber fruit development and may be involved in regulating fruit ripening (Wang et al., 2012). Two PSGs encode ABI5-like proteins, which participate in positively regulating ABA-mediated inhibition of seed germination and growth to promote fruit maturation (Brocard et al., 2002; Chen et al., 2008; Reeves et al., 2011). Previous study has also shown that exogenous sugar can significantly promote strawberry ripening while stimulating ABA accumulation (Jia et al., 2011). Therefore, all of the above PSGs are apparently related to fruit development and ripening, which attract major attention during domestication.

AUTHOR CONTRIBUTIONS

TZ, JL, YZ, JH, and HS conceived and designed the project. JL, LX, QQ, TZ, YZ, HS, and QW prepared samples. QQ, TZ, QW, and LX contributed to sequencing experiment. TZ and QQ performed data analyses. TZ, QQ, and JH wrote the paper. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.01839/full#supplementary-material

FILE S1 | Python script for concatenate sequences of all orthogroups.

REFERENCES

Brocard, I. M., Lynch, T. J., and Finkelstein, R. R. (2002). Regulation and role of the *Arabidopsis* abscisic acid-insensitive 5 gene in abscisic acid, sugar, and stress response. *Plant Physiol.* 129, 1533–1543. doi: 10.1104/pp.005793

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25, 1972–1973. doi: 10.1093/bioinformatics/btp348

Chan, R. L., Keller, M., Canaday, J., Weil, J. H., and Imbault, P. (1990). Eight small subunits of *Euglena* ribulose 1-5 bisphosphate carboxylase/oxygenase are translated from a large mRNA as a polyprotein. *EMBO J.* 9, 333–338.

Chen, B., Li, J., Huo, Z., Wan, C., Zhang, Z., Qiao, Y., et al. (2015). Estimation of genome size in six wild strawberry species. *J. Fruit Sci.* 32, 51–56.

Chen, H., Zhang, J., Neff, M. M., Hong, S. W., Zhang, H., Deng, X. W., et al. (2008). Integration of light and abscisic acid signaling during seed germination and early seedling development. *Proc. Natl. Acad. Sci. U.S.A.* 105, 4495–4500. doi: 10.1073/pnas.0710778105

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., et al. (2007). Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450, 203–218. doi: 10.1038/nature06341

Du, Z., Zhou, X., Ling, Y., Zhang, Z., and Su, Z. (2010). agriGO: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 38, W64-W70. doi: 10.1093/nar/gkq310

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Favre, A., Päckert, M., Pauls, S. U., Jähnig, S. C., Uhl, D., Michalak, I., et al. (2015). The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. *Biol. Rev.* 90, 236–253. doi: 10.1111/brv.12107

Fitch, W. M. (1970). Distinguishing homologous from analogous proteins. *Syst. Biol.* 19, 99–113.

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotech.* 29, 644–652. doi: 10.1038/nbt.1883

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512. doi: 10.1038/nprot.2013.084

Hirakawa, H., Shirasawa, K., Kosugi, S., Tashiro, K., Nakayama, S., Yamada, M., et al. (2014). Dissection of the octoploid strawberry genome by deep sequencing of the genomes of *Fragaria* species. *DNA Res.* 21, 169–181. doi: 10.1093/dnares/dst049

Jia, H. F., Chai, Y. M., Li, C. L., Lu, D., Luo, J. J., Qin, L., et al. (2011). Abscisic acid plays an important role in the regulation of strawberry fruit ripening. *Plant Physiol.* 157, 188–199. doi: 10.1104/pp.111.177311

Johnson, A. L., Rajanikanth, G., and Tia-Lynn, A. (2014). Bioclimatic evaluation of geographical range in *Fragaria* (Rosaceae): consequences of variation in breeding system, ploidy and species age. *Bot. J. Linnean Soc.* 176, 99–114. doi: 10.1111/boj.12190

Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992). The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* 8, 275–282.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27

Kang, C., Darwish, O., Geretz, A., Shahan, R., Alkharouf, N., and Liu, Z. (2013). Genome-scale transcriptomic insights into early-stage fruit development in woodland strawberry *Fragaria vesca*. *Plant Cell* 25, 1960–1978. doi: 10.1105/tpc.113.111732

Koenig, D., Jiménez-Gómez, J. M., Kimura, S., Fulop, D., Chitwood, D. H., Headland, L. R., et al. (2013). Comparative transcriptomics reveals patterns of selection in domesticated and wild tomato. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2655–2662. doi: 10.1073/pnas.1309606110

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi: 10.1093/molbev/msw054

Li, J., and Fang, X. (1999). Uplift of the Tibetan Plateau and environmental changes. *Chin. Sci. Bull.* 44, 2117–2124. doi: 10.1007/BF03182692

Li, L., Stoeckert, C. J., and Roos, D. S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* 13, 2178–2189. doi: 10.1101/gr.1224503

Liscum, E., and Reed, J. W. (2002). Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol. Biol.* 49, 387–400. doi: 10.1023/A:1015255030047

Liston, A., Cronn, R., and Ashman, T. -L. (2014). *Fragaria*: a genus with deep historical roots and ripe for evolutionary and ecological insights. *Am. J. Bot.* 101, 1686–1699. doi: 10.3732/ajb.1400140

Liu, J.-Q., Duan, Y.-W., Hao, G., Ge, X.-J., and Sun, H. (2014a). Evolutionary history and underlying adaptation of alpine plants on the Qinghai–Tibet Plateau. *J. Syst. Evol.* 52, 241–249. doi: 10.1111/jse.12094

Liu, J.-Q., Sun, Y.-S., Ge, X.-J., Gao, L.-M., and Qiu, Y.-X. (2012). Phylogeographic studies of plants in China: advances in the past and directions in the future. *J. Syst. Evol.* 50, 267–275. doi: 10.1111/j.1759-6831.2012.00214.x

Liu, M. J., Zhao, J., Cai, Q. L., Liu, G. C., Wang, J. R., Zhao, Z. H., et al. (2014b). The complex jujube genome provides insights into fruit tree biology. *Nat. Commun.* 5:5315. doi: 10.1038/ncomms6315

Ma, W., Wei, L., Wang, Q., Shi, D., and Chen, H. (2007). Increased activity of the non-regulated enzymes fructose-1,6-bisphosphate aldolase and triosephosphate isomerase in *Anabaena* sp. strain PCC 7120 increases photosynthetic yield. *J. Appl. Phycol.* 19, 207–213. doi: 10.1007/s10811-006-9125-8

Maceachern, S., Mcewan, J., Mcculloch, A., Mather, A., Savin, K., and Goddard, M. (2009). Molecular evolution of the Bovini tribe (Bovidae, Bovinae): is there evidence of rapid evolution or reduced selective constraint in Domestic cattle? *BMC Genomics* 10:179. doi: 10.1186/1471-2164-10-179

Njuguna, W., Liston, A., Cronn, R., Ashman, T. L., and Bassil, N. (2013). Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing. *Mol. Phylogenetics Evol.* 66, 17–29. doi: 10.1016/j.ympev.2012.08.026

Obroucheva, N. V. (2014). Hormonal regulation during plant fruit development. *Ontogenez* 45, 11–21.

Park, S. Y., and Cutler, S. R. (2009). Abscisic acid inhibits type 2C protein phosphatases via the PYR/PYL family of START proteins. *Science* 324, 1068–1071. doi: 10.1126/science.1173041

Reeves, W. M., Lynch, T. J., Mobin, R., and Finkelstein, R. R. (2011). Direct targets of the transcription factors ABA-Insensitive(ABI)4 and ABI5 reveal synergistic action by ABI4 and several bZIP ABA response factors. *Plant Mol. Biol.* 75(4-5), 347–363. doi: 10.1007/s11103-011-9733-9

Rousseau-Gueutin, M., Gaston, A., Aïnouche, A., Aïnouche, M. L., Olbricht, K., Staudt, G., et al. (2009). Tracking the evolutionary history of polyploidy in *Fragaria* L. (strawberry): new insights from phylogenetic analyses of low-copy nuclear genes. *Mol. Phylogenetics Evol.* 51, 515–530. doi: 10.1016/j.ympev.2008.12.024

Sargent, D. J., Yang, Y., Šurbanovski, N., Bianco, L., Buti, M., Velasco, R., et al. (2016). HaploSNP affinities and linkage map positions illuminate subgenome composition in the octoploid, cultivated strawberry (*Fragaria × ananassa*). *Plant Sci.* 242, 140–150. doi: 10.1016/j.plantsci.2015.07.004

Shulaev, V., Sargent, D. J., Crowhurst, R. N., Mockler, T. C., Otto, F., Delcher, A. L., et al. (2011). The genome of woodland strawberry (*Fragaria vesca*). *Nat. Genet.* 43, 109–116. doi: 10.1038/ng.740

Stamatakis, A., Ludwig, T., and Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21, 456–463. doi: 10.1093/bioinformatics/bti191

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34(suppl. 2), W609–W612. doi: 10.1093/nar/gkl315

Tamura, K., Battistuzzi, F. U., Billing-Ross, P., Murillo, O., Filipski, A., and Kumar, S. (2012). Estimating divergence times in large molecular phylogenies. *Proc. Natl. Acad. Sci. U.S.A.* 109, 19333–19338. doi: 10.1073/pnas.1213199109

Tennessen, J. A., Govindarajulu, R., Ashman, T. L., and Liston, A. (2014). Evolutionary origins and dynamics of octoploid strawberry subgenomes revealed by dense targeted capture linkage maps. *Genome Biol. Evol.* 6, 3295–3313. doi: 10.1093/gbe/evu261

Wang, Y., Wu, Y., Duan, C., Chen, P., Li, Q., Dai, S., et al. (2012). The expression profiling of the CsPYL, CsPP2C and CsSnRK2 gene families during fruit

development and drought stress in cucumber. *J. Plant Physiol.* 169, 1874–1882. doi: 10.1016/j.jplph.2012.07.017

Ward, J. A., Ponnala, L., and Weber, C. A. (2012). Strategies for transcriptome analysis in nonmodel plants. *Am. J. Bot.* 99, 267–276. doi: 10.3732/ajb.1100334

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*. 5:4. doi: 10.3389/fgene.2014.00004

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., et al. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* 39(suppl. 2), W316-W322. doi: 10.1093/nar/gkr483

Yang, Y., Moore, M. J., Brockington, S. F., Soltis, D. E., Wong, G. K., Carpenter, E. J., et al. (2015). Dissecting molecular evolution in the highly diverse plant clade caryophyllales using transcriptome sequencing. *Mol. Biol. Evol. Adv. Access* 32, 2001–2014. doi: 10.1093/molbev/msv081

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* 24, 1586–1591. doi: 10.1093/molbev/msm088

Yoshida, T., Nishimura, N., Kitahata, N., Kuromori, T., Ito, T., Asami, T., et al. (2006). ABA-hypersensitive germination3 encodes a protein phosphatase 2C (AtPP2CA) that strongly regulates abscisic acid signaling during germination among *Arabidopsis* protein phosphatase 2Cs. *Plant Physiol.* 140, 115–126. doi: 10.1104/pp.105.070128

Zhang, J., Nielsen, R., and Yang, Z. (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol. Biol. Evol.* 22, 2472–2479. doi: 10.1093/molbev/msi237

Zhang, Q.-J., Zhu, T., Xia, E. -H., Shi, C., Liu, Y. -L., Zhang, Y., et al. (2014). Rapid diversification of five *Oryza* AA genomes associated with rice adaptation. *Proc. Natl. Acad. Sci. U.S.A.* 70, E4954–E4962. doi: 10.1073/pnas.1418307111

Zhou, W., Wen, Y., Fu, J., Xu, Y., Jin, J., Ding, L., et al. (2012). Speciation in the Rana chensinensis species complex and its relationship to the uplift of the Qinghai-Tibetan Plateau. *Mol. Ecol.* 21, 960–973. doi: 10.1111/j.1365-294X.2011.05411.x