Genomic hitchhiking with moisture adaptation patterns stem prickleless in rose

Mi-Cai Zhong [1,4], Xiao-Dong Jiang [1,4], Guo-Qian Yang [2], Wei-Hua Cui [1,4], Zhi-Quan Suo [1,4], Wei-Jia Wang [3], Yi-Bo Sun [1,4], Dan Wang [1,4], Xin-Chao Cheng [5], Xu-Ming Li [5], Xue Dong [1,2], Kai-Xue Tang [3,*], De-Zhu Li [1,2,*], Jin-Yong Hu [1,*]

1. CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences. Kunming 650201, China.
2. Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, 650201, China.
3. Flower Research Institute, Yunnan Agricultural Academy of Sciences, Kunming 650231, China.
4. University of Chinese Academy of Sciences, Beijing 100049, China.
5. Biomarker Technologies Corporation, Beijing, China.

* For correspondence: hujinyong@mail.kib.ac.cn, dzl@mail.kib.ac.cn, kxtang@hotmail.com

Abstract

"Rose with prickles" is a symbol of adversity and sacrifice but the molecular bases patterning prickle remains poorly explored. Here we generated a high-quality reference genome assembly for prickleless *Rosa wichuraiana* 'Basye's Thornless' (BT), and identified genetic elements relating to stem prickle patterning. BT genome harbors a significant high proportion of *Copia*-type LTR (62.74% of LTRs) and diverges remarkably from that of *R. chinensis* 'Old Blush', a founder genotype for rose domestication. Three negatively selected and fixed QTL were identified underlying BT prickleless. Differentially expressed genes in QTL were enriched for water-use-related genes, suggesting that prickleless might hitchhike the adaptation to moisture environments. While the prickle-related gene-regulatory-network (GRN) was highly conserved, expression variation of key candidate genes associated well with prickle-patterning divergence in roses. Our efforts provide fundamental resources and insights for Rosaceae genome evolution and molecular basis of key traits, thus should benefit rose improvement.

Key words: *Rosa wichuraiana* 'Basye's Thornless', genome assembly, prickleless, gene-regulatory-network (GRN), QTL, purifying selection

Background

Understanding the molecular genetic mechanisms driving biodiversity remains a great challenge. Prickles are acuminate protuberance formed by the deformation of plant trichomes, which act mainly to protect plants against attacks from herbivores, pathogens or mechanical injury and to increase thickness of epidermis, reduce heat and water dissipation (Coverdale, 2020; Szymanski et al., 2000). Many plants feature prickles on the epidermis of leaves, stems, and fruits like in roses and spines and thorns in other members of Rosaceae as well as of other families like Araliaceae, Leguminosae and Rutaceae. Trichome has served as a developmental model to study the control of cell proliferation and growth in both *Arabidopsis* and cotton with the patterning gene-regulatory-network (GRN) being investigated in details (Chopra et al., 2019; Doroshkov et al., 2019; Pattanaik et al., 2014; Pesch and Hülskamp, 2009; Serna and Martin, 2006; Szymanski et al., 2000; Wang et al., 2019c; Yang and Ye, 2013). However, the molecular and genetic mechanisms underlying stem prickle development has not been studied systematically, especially in non-model species.

In *Arabidopsis*, the trichome GRN involving at least three major groups of transcription factors (MYBs, bHLHs, and WD40 repeat proteins) have been proposed to play key roles (Matias-Hernandez et al., 2016a; Pattanaik et al., 2014; Pesch and Hülskamp, 2009; Serna and Martin, 2006; Szymanski et al., 2000). This MYB/bHLH/WD-repeat (MBW) complex involves the physical interaction between GL3 with GL1 and TTG1, respectively, to initiate trichome formation. *TTG2* acts genetically downstream of this complex while sharing partial activity with *GL2* to initiate trichome development (Johnson et al., 2002). Single repeat R3 MYBs (including CPC, ETC1/2/3, TCL1/2, and TRY) act as negative regulators of trichome development via competing with the R2R3 type MYBs to form a repressor complex. The *miR156/SPL*s and *miR172/AP2-like* gene modules, which are key regulators in aging pathway of floral transition, also play important roles in trichome patterning (Matias-Hernandez et al., 2016b; Wang et al., 2019a; Xue et al., 2014; Yu et al., 2010). Hormone signals like gibberellin acids (GA), cytokinin (CK) and jasmonic acids (JA) participate in

the regulation of trichome patterning also via the interaction with the MYB/bHLH/WD-repeat complex (see reviews (Pattanaik et al., 2014; Pesch and Hülskamp, 2009; Serna and Martin, 2006; Szymanski et al., 2000; Yang and Ye, 2013)). This complex seems to be highly conserved with some functional divergence of certain components in *Arabis alpina*, a relative of *Arabidopsis thaliana*, cucumber, and cotton (Che and Zhang, 2019; Chopra et al., 2019; Pattanaik et al., 2014; Wang et al., 2019c). However, it's worthy to note that this complex also participates in the regulation of other developmental processes like anthocyanin formation, root and seed development, thus may be under a strong natural selection in sequences.

Roses have long been revered for their beauty and fragrance, despite the presence of prickles on rachises and stems. Roses represent one of the most important horticultural plants in the world for its economic and cultural values with more than 36,000 cultivated varieties. Rose features diverse morphological novelties, such as continuous flowering (CF), prickles on rachises and stems, doubled petals and scent, that could not be identified and studied in current model species, e.g., *Arabidopsis*, rice and poplar, hence rose is now becoming a woody model plant for studying these important traits (Bendahmane et al., 2013; Dong et al., 2017). However, the genetic makeups of rose are rather complex due to high frequent inter- and intra- species hybridization, polyploidization, and human selection (Fougere-Danezan et al., 2015; Koopman et al., 2008; Li D, 2016; Meng et al., 2011; Smulders et al., 2019; Zhang et al., 2013; Zhu et al., 2015). Determination of a highly heterozygous genome for roses remains challenging, as seen by the very fragmented assembly trials of three species (Lu et al., 2016; Nakamura et al., 2017; Raymond et al., 2018), and currently only two versions of chromosome-level genome sequences for doubled-haploid materials of one cultivated founder genotype (*Rosa chinensis* 'Old Blush', haploOB) in domestication of modern cultivated roses have been described (Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018). Enormous sequence variation including single-nucleotide-polymorphisms (SNPs), insertion/deletions (indels), inversions and translocations can exist between varieties of rose like in maize

and may contribute to key agricultural traits (Li et al., 2019a; Sun et al., 2018; Yang et al., 2019). Therefore, a complete picture of genome assembly for diploid rose genotype is necessary for the purpose of comparative genomics in identifying molecular genetic bases of key morphological traits that have been pursued for decades in roses.

In practice, the glandular-trichome-based prickles on rose stems (Kellogg et al., 2011) can cause strong inconvenience during the handling and management. Most of rose genotypes harbor prickles, while BT, one of the founder species during rose domestication, does not (Figure 1A) (Bendahmane et al., 2013; Byrne et al., 2007; Gu et al., 2006). BT grows in a relatively moisture environment and has at least six pairs of contrasting traits like once flowering behavior, prickleless on stem, prostrate growth during vegetative growth season, white-colored single flowers, high resistance to black spots in comparison to OB genotype (Byrne et al., 2007; Li et al., 2015). These have allowed this species frequently used as parental materials for constructing a set of genetic populations aiming at dissecting the genetic bases of these important traits (Crespel et al., 2002; Hibrand Saint-Oyant et al., 2018; Li et al., 2019c; Li et al., 2015; Shupert et al., 2007; Spiller et al., 2011). Despite of the importance in rose genetics and culture, molecular information of BT is limited with only transcriptomic and molecular markers information available (Crespel et al., 2002; Cui et al., 2020; Hibrand Saint-Oyant et al., 2018; Li et al., 2019c; Li et al., 2018; Li et al., 2015; Shupert et al., 2007; Spiller et al., 2011). In this study, we determined the chromosome-level genome sequences of BT. We systematically analyzed the prickle QTL and GRN, and proposed that prickleless may hitchhike and be selected for moisture adaptation.

Results and discussion

De novo assembly of chromosome level BT genome

The genome size of BT was estimated to be around 525.9 (±5.91) Mb by flow cytometry. The initial genome survey analysis with ~267x Illumina reads predicted a genome size of 525.5 Mb with around 1.03% of heterozygosity (Supplementary Tables S1, S2). We generated ~93x PacBio Sequel reads and used CANU in combination with WTBDG2 and Quickmerge to generate contigs. We obtained 1,847 contigs with N50 at 932 Kb and total contig length of 526.7 Mb (Table 1; Supplementary Table S3, S4). We later generated chromosome level assembly using 34.64 million unique Di-Tags read pairs produced via Hi-C experiment (Supplementary Table S5). Hi-C analysis generated 2711 contigs, among which 857 (52.97% in 1,618 clustered) contigs were ordered and oriented into seven pseudo-molecules spanning 483.89 Mb genome (92.41% of 521.32 Mb total sequences clustered; Supplementary Table S6). There were 1,093 contigs not ordered/oriented (46.18 Mb, 8.71% in length). After manual correction, we obtained the final assembly of BT genome (530.07 Mb) with Chr5 the longest (87.44 Mb) and Chr3 the shortest (46.04 Mb) (Figure 1B; Table 1; Supplementary Figure S1, S2, Supplementary Table S3-S6).

BT genome is in reference-quality with high completeness

We evaluated the quality and completeness of BT genome using five methods. Firstly, this genome were subjected to Benchmarking Universal Single-Copy Orthologs (BUSCO, embryophyta_odb9) analysis (Simão et al., 2015). Overall, 93.9% (2,184; including 1,956 complete and single copy, and 228 complete and duplicated) and 3.0% (69; fragmented) of the 2,326 expected plant gene models were identified in the assembled BT genome. Approximately 73 gene models (3.1%) were considered missing in our assembly (Supplementary Table S7). Secondly, two independent genetic maps based on different populations were used to assess the quality of BT genome. The collinearity between the assembled BT chromosome sequences and the K5 genetic map (Bourke et al., 2018) was very high with most of the association efficiency above 95% (Supplementary Table S8; Supplementary Figure S3). Except

chromosome 4 in the OBxBT genetic map, on which the distribution of molecular markers was obviously skewed (Li et al., 2019c), all the other 6 chromosomes showed a very good collinearity with correlation rate around 90% (Supplementary Figure S4; Supplementary Table S8). We next aligned our BT assembly to the recently reported haploOB genome sequences (Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018) and identified a very high collinearity of BT assembly to both Raymond's (Figure 2A, 2B) and Hibrand Saint-Oyant's (Supplementary Figure S3) genomes, revealing a high synteny level and gene order between these two genotypes. Later, we re-mapped and got the high mapping rates of the Illumina (98.16%) and PacBio (96.39%) reads described above to our assembly (Supplementary Table S9). Finally, a LTR Assembly Index (LAI) assay revealed that genome-wide LAI value of BT was 20.04, a level of golden reference (Supplementary Figure S5) (Ou et al., 2018). Collectively, these data indicate a high quality and completeness of the BT genome, which was annotated further.

BT genome features a high proportion of *Copia* type LTRs

Repeat sequence represented about 65.19% (345.58Mb) of BT genome, in which LTR retrotransposons were about 45.09% (or 69.16% of all repeats; Figure 1B; Supplementary Table S10). In contrast to haploOB (Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018) but similar to *Musa balbisiana* assembly (Wang et al., 2019b), *Copia* elements (28.29% of the genome, or 62.74% of all LTRs) were more frequent than *Gypsy* (16.39%) in BT genome. The distribution of *Copia* type LTRs is antagonistic to that of *Gypsy* type along almost all BT chromosomes (Supplementary Figure S6). The ~0.58 *Gypsy*-to-*Copia* ratio was one of the lowest values in the sequenced species, suggesting that *Copia* amplification contributes much more than other TEs to BT genome evolution (Supplementary Table S11; Supplementary Figure S7) (Bennetzen et al., 2012; Guo et al., 2019; Liu et al., 2019; Shi et al., 2019). In general, TEs displayed an apparently dense distribution in the pericentromeric regions on the chromosomes while inversely correlated with gene density (Figure 1B).

BT genome has 37,634 protein-coding genes and non-coding genes

Gene annotation of the BT genome using *de novo*, homology-based, and transcriptome profiling based predictions predicted a total of 32,674 protein-coding genes, with average gene length about 3.28kb, average exon length about 251bp and mean intron length 390bp (Figure 1B; Table 1; Supplementary Table S12-S14; Supplementary Figure S8). Protein-coding genes showed focused distribution and high expression level on chromosome arms, on which highly variable levels of DNA methylation were also observed (Figure 1B). Furthermore, 15,357 genes were annotated with gene ontology (GO) information, and about 95.97%, 78.88%, and 52.94% of these genes were annotated in the biological process, cellular component and molecular function categories, respectively (Supplementary Figure S9; Supplementary Table S14). Among these 1,703 genes (~5.21%) encode for transcription factors (TFs) in 69 families and 401 genes (~1.23%) encode for transcriptional regulators in 24 families (Supplementary Table S15). We additionally predicted 557 tRNAs in 24 families, 78 rRNAs in 4 families, 71 miRNAs in 20 families, and 4,254 pseudo-genes (mean length 2,928.7 bp, counting for 12,490,896 bp or 2.58%) in the BT genome (Supplementary Table S16).

BT genome features a high proportion of allelic sequence polymorphisms

We next identified sequence polymorphisms within BT genome. We found 2.3 million SNPs (0.44% of BT genome), in which 1.45 million were transitions and 0.85 million were transversions (*Ts/Tv* ratio was 1.69) (Supplementary Table S17; Supplementary Figure S10, S11), and 0.51 million indels, of which 44.38% located in intergenic regions, and 520 (0.01%) resulted lost or gain of start/stop codons (Supplementary Table S17, S18; Supplementary Figure S12). About 99.7% of the SNPs were heterozygous with half in intergenic regions. These SNPs caused 98 genes losing start codon, 2,694 genes gaining stop codons, and 2,144 losing stop codons (Supplementary Table S19). These sequence polymorphisms will be valuable for further investigation like allele-specific expression, epigenetic regulation, and genome structure and evolution analysis in roses.

BT diverges from the haploOB genome

BT and OB are two genotypes featuring at least six-pair of contrasting traits like absence/presence of prickles on shoots (Byrne et al., 2007; Li et al., 2015). To understand the genetic variation associated with these phenotypical traits, we identified 95 synteny blocks covering 87.59% of BT (464.31 Mb) or 80.59% of haploOB (415.5 Mb) genomes and detected 20 inversion events (covering 1,464 of BT or 1,718 of haploOB genes) (Figure 2A; Supplementary Table S20, S21). We detected 7.31 million of SNPs and 5.16 millions of one-nucleotide indels as well as 20.27 Mb large structural variants including indels size above 50nt (3,171 events) and tandem (458 events) or repeat (5,146 events) expansion/contractions (Figure 2C, 2D; Supplementary Figure S13-S15; Supplementary Table S22). We additionally found 17,034 genomic rearrangement events (~65.8 Mb) that were not randomly distributed along chromosomes (Figure 2D; Supplementary Table S22a). Interestingly, genes related to DNA integration and metabolic process were significantly enriched in CPG regions (Supplementary Table S22b), a pattern being coincident with the frequent hybridization and genomic rearrangement during domestication of OB (Raymond et al., 2018). Taken these together, a significant sequence variation presents between the two genotypes, which could be used for genetic dissection of important traits in roses.

BT experiences significant gene family expansion/contraction profiles

To investigate further the DNA makeups of BT, we identified 11,593 homologous gene families shared among BT, haploOB, *Prunus*, *Pyrus*, *Fragaria* and *Malus*. We detected 1,320 gene families specific for BT, while 2,593 families unique for haploOB (Figure 2D). The BT specific genes were significantly (FDR<0.05) enriched in GO terms for nucleic acid binding, RNA-dependent DNA biosynthesis, and DNA integration processes, while significantly less of oxidation-reduction genes (Supplementary Table S23). By comparison, the haploOB specific genes were functionally over-represented in biological processes involved in DNA metabolic, organic substance metabolic,

hetero- and organic- cyclic compound binding, suggesting that BT and OB harbored different and specific gene profiles.

To understand the phylogenetic position and identify the expanded/contracted gene families in BT, we searched for the conserved orthologous gene families in genomes of BT, haploOB, *Malus domestica*, *Fragaria vesca*, *Prunus avium*, *Prunus persica*, *Pyrus communis*, and *Rubus occidentalis* (all Rosaceae) with *Vitis vinifera* and *Ziziphus jujube* as outgroups. A total of 1,220 single-copy orthologous groups were identified and used to construct a maximum-likelihood tree (Figure 2E). BT and OB are sister species with *Fragaria* as the closet genus. Following the fossil calibrations of nodes C1 and C3 for Rosaceae family (Zhang et al., 2017), the estimated divergence time of BT and OB was about 6.15 MYA (95% highest posterior density: 2.33-15.43 MYA). A total of 978 and 616 gene families were apparently expanded and contracted, respectively, in BT genome (Figure 2E). This was significantly different from that of OB (*Yates' Chi-square test*, $p<0.0001$), corroborating with that BT and haploOB featured high level of sequence divergence and morphological variation. BT expanded genes were enriched for proteins involved in DNA binding together with macromolecule and primary metabolic processes, while the contracted genes were related to protein phosphorylation, kinase activity, and components in cell wall metabolisms (Supplementary Figure S16; Supplementary Table S24).

Genetic patterning of Stem prickle density in BT x OB populations

To understand the genetic pattern of stem prickle density, we scored stem prickle density in the 99-individual F1 and the 148-individual BC1F1 populations generated via crossing the prickleless BT and prickly OB (Figure 3A; supplementary Figure S17-S19) (Li et al., 2019c; Li et al., 2015). In the F1 population, the number of prickleless vs. prickly individuals is 13:86, a ratio not deviating from 1:7 (three loci hypothesis, *Chi-square test*, $p=0.96$) or 1:15 (four loci hypothesis, *Chi-square test*, $p=0.43$). In the BC1F1 population created by back-crossing a prickleless F1 individual to the prickly OB, plants show a normal distribution of prickle density with 10 plants featuring no prickle, while

the other 138 individuals have prickles at different densities on stem (Figure 3B and 3C; Supplementary Figure S19). These data indicate that, agreeing with previous studies (Bourke et al., 2018; Byrne et al., 2007; Crespel et al., 2002; Hibrand Saint-Oyant et al., 2018; Koning-Boucoiran et al., 2012; Linde et al., 2006; Shupert et al., 2007), stem prickle density is likely a quantitative trait that is regulated by three to four loci with loss of prickle initiation as homozygous recessive in the BTxOB populations.

Sequence divergence in QTL is negatively selected and fixed in prickleless roses

We conducted a QTL analysis using the 2,160 SNP markers for the BC1F1 population (Li et al., 2019c). Three QTL were identified using interval mapping (IM) method, with QTL1 on Chr3 (BT genome, a 2.26 Mb segment from 13.97-16.23 Mb, 153 genes; haploOB genome, a 2.55 Mb region between 14.98 and 17.53 Mb, 254 genes; 117 syntenic), which was next to a genome inversion event, QTL2 (BT genome, 19.03-20.28 Mb, 73 genes; haploOB genome, 21.23-22.34 Mb, 65 genes; ~1.2 Mb; 60 syntenic) and QTL3 (BT genome, 0.98-2.64 Mb, 204 genes; haploOB genome, 1.08-2.73 Mb, 211 genes; ~1.65 Mb; 177 syntenic) on Chr7 (Figure 3D, 3E; Supplementary Figures S20-S21; Supplementary Table S25, S26).

To further understand the genetic pattern of QTL regions in other rose genotypes, we obtained the genomic sequencing data (minimum 15x) for three available prickleless (including one wild species *R. banksiae* and two cultivated genotypes; the only prickleless genotypes in our ~200 rose collection) and eight prickly roses (all wild genotypes), and estimated their weighted-*Fst* values per 100 Kb window (Supplementary Table S27). The prickleless roses were randomly distributed on the phylogenetic tree based on chloroplast genome sequences, suggesting independent origins of prickleless roses (Supplementary Figure S22). The weighted-*Fst* values peaked exactly at the QTL regions on Chr3 and Chr7 indicating that sequence variation in these chromosome segments were fixed between the prickly-prickleless groups of roses (Figure 4A; Supplementary Figure S23; Supplementary Table S28).

Synonymous substitutions (*Ks*) for genes in all QTL regions showed a significant reduction to that of rest of the genome (Figure 4B; Supplementary Table S29). It should be noted that QTL1 region of the prickleless roses had significantly reduced sequence diversity (*Pi*) and non-synonymous to synonymous substitution ratio (*Ka/Ks*) than other parts of Chr3 (Figure 4C; Supplementary Figure S24-S26; Supplementary Table S30). QTL1 region overlapped with one of the two peaks reported previously (Hibrand Saint-Oyant et al., 2018). A detailed examination of the syntenic genes in the QTL regions revealed that genes involved in water usage and desiccation response were significantly differentially expressed in the young shoot tips between BT and OB (Figure 4D; Supplementary Figure S27), suggesting that prickle patterning might hitchhike water adaptation processes, thus prickles could have relatively high water content.

To substantiate this, we measured relative water content in prickle, bark, stele, and leaves at different sections of shoots in two modern rose genotypes (Figure 4E, and Supplementary Figure S29x, S35, S36). Prickles always had relatively higher water content than barks and leaves in the four sections of current season shoots. On the contrary, the relative water content dropped sharply in prickles on old shoots, while maintained more or less similar level in leaves to that of young shoots (Figure 4F, and Supplementary Figure S29x, S35). This provokes us to hypothesize that prickle might serve as water storage place in early developmental stage, while mainly works as defense organ at later development stages. A further examination of the relative water contents in prickles and barks for a collection of 39 rose lines (12 wild genotypes, 6 China roses, and 21 modern rose genotypes) revealed a similar pattern (Figure 4G; Supplementary Figure S37). Taken these together, our results suggested that the prickle patterning might hitchhike the selection on genes involved in water adaptation in roses (Byrne et al., 2007).

Expression variation of candidate genes in QTL and trichome GRN associates with stem prickle patterning

Rose prickles are deformed plant trichomes, for which the responsible GRN has been well studied in *Arabidopsis* (Johnson et al., 2002; Kellogg et al., 2011; Matias-Hernandez et al., 2016a; Pattanaik et al., 2014; Pesch and Hülskamp, 2009; Serna and Martin, 2006; Szymanski et al., 2000; Yang and Ye, 2013). A scanning electronic microscopy (SEM) analysis revealed that, in contrast to the normal initiation of prickles and stomata development on OB shoots, BT maintained the ability to develop stomata but lost its capability for prickle initiation (Figure 5A). Therefore, we focused on potential candidate genes involved in the MBW-complex, especially the MYB-, bHLH- and WRKY-like transcription factors in the QTL regions and the trichome GRN.

Arabidopsis has ~150 trichome-related genes involved in initiation, development and birefringence and other processes (Supplementary Table S31, S32). We identified the potential rose orthologs of these genes based on sequence similarity and collinearity analysis, and mainly focused on the 29 genes known to regulate trichome initiation and development (Figure 5B; Supplementary Table S32, S33). In comparison to trichome GRNs of Arabidopsis, cucumber or cotton, we detected no presence-absence polymorphisms of genes between BT and OB, indicating that the trichome GRN is highly conserved among these species (Che and Zhang, 2019; Chopra et al., 2019; Pattanaik et al., 2014; Wang et al., 2019c). In the QTL regions, we found nine *MYB*s (two in QTL1, two in QTL2, and five in QTL3 including a homolog of *WER*/*GL1/MYB23-like*), three *bHLH*s (all in QTL3), and one *WRKY* (in QTL3) (Supplementary Table S33, S35; Supplementary Figures S28-S30). The previously proposed candidate *TTG2-like* (Rw3G013440/Chr3G0468221) was close to the vicinity of QTL1 and located ~0.6 Mb outside of QTL1 (Figure 3E) (Hibrand Saint-Oyant et al., 2018).

We detected no significant sequence variation in the means of nonsynonymous (*Ka*) vs. synonymous (*Ks*) change between these BT and OB orthologous genes, indicating that likely the regulation of gene expression, not the encoding potential, of these trichome-related genes might underlie the regulation of prickleless in BT. Indeed, we detected five genes featuring at least two-fold

significant expression changes in very young shoots including apical meristems between BT and OB (Figure 5B) (Li et al., 2018). The expression of positive regulators *GIS* and *TTG1* were decreasing, while the negative regulator R3 type MYBs (*TRY*- and *MYBH-like*) and one *MYB8-like* genes increased its expression in BT comparing to OB (Figure 5B, Supplementary Table S33). A further quantitative RT-PCR assay revealed that the expression of *MYB27-like* in QTL1 (Rw3G015190/Chr3G0470621), *TT2-like* in QTL2 (Rw7G018230 /Chr7G0203911), the *GIS-like* (Rw3G019920/Chr3G0479101) and *TTG2-like* was significantly different in the young stem materials of F1 and BC1F1 pools (Figure 5C-F; Supplementary Table S29). Both MYB27-like and TT2-like transcription factors have been proposed to regulate anthocyanin synthesis in petunia and Arabidopsis via modulating the MBW-complex (Albert et al., 2014; Baudry et al., 2004; Gonzalez et al., 2009), though their functional roles in specifying trichome development need further investigation. Acting downstream of the R3 MYBs, GIS was known to positively promote stem trichome development via transcriptional regulation of genes involved in MBW-complex (Gan et al., 2006; Zhang et al., 2018). Featuring strong sequence variations in upstream transcriptional regulatory region of BT and OB (Supplementary Figure S31), the WRKY transcription factor *TTG2* worked downstream of MBW-complex but partially independent of *GL2* to regulate trichome patterning and seed coat anthocyanin biosynthesis (Ishida et al., 2007; Johnson et al., 2002), and had been recently proposed to be a candidate for stem prickle patterning in rose (Hibrand Saint-Oyant et al., 2018). Taking these together, we proposed that gene expression variation of these candidates underlay likely the prickle-development variation between BT and OB (Figure 5G).

Conclusions

In this study, we constructed the high-quality chromosome-level genome assembly for a prickleless diploid rose, and used this assembly together with QTL mapping to study the molecular basis of this trait.

We generated a chromosome-level genome assembly for BT, a rose genotype without stem prickle. BT genome represents a high-quality reference level assembly for a diploid rose, as evidenced by our BUSCO, genetic maps, LTR assembly index, and synteny analyses. Being consistent with the presence of many contrasting pairs of morphological traits, the most important donor genotype to modern rose domestication, BT diverges significantly from OB in the genome sequences including many SNPs, indels, and enormous genomic rearrangements. Furthermore, the heterozygous BT plant has enormous sequence diversity, which has made the high-quality genome assembly difficult for roses (Foucher et al., 2015; Hibrand Saint-Oyant et al., 2018; Raymond et al., 2018).

We identified the genetic inheritance pattern of stem prickles. With a BC1F1 population and its associated molecular markers and transcriptomic data, we showed that stem prickleless was likely regulated by three genetic loci, agreeing with previous findings using F1 populations. We detected three QTL with QTL1 co-localized with previously known QTL on Chr3. All QTL were under strong purifying selection with QTL1 region enriched for differentially-expressed-genes that were involved in water usage and nucleotide metabolism. Furthermore, sequences in these QTL regions seemed fixed in prickleless roses. These data indicate that patterning of stem prickles is likely selected for moisture adaptation, in contrast to the long-believed hypothesis that the generation of shoot prickle is driven by selection pressure against herbivory. Interestingly, relative water content in prickles on young stems was significantly higher than in barks and leaves, while it drops sharply in prickles on relatively old stems, indicating that prickles might have different function at different developmental stages.

Taking advantage of the high-quality genomes of BT and haploOB, we screened for candidate genes participating potentially in stem prickle initiation in roses. We hypothesized the expression diversity in genes like *GIS*-, *MYB27-like*, or R3 type *MYBs* might affect the MBW-complex formation hence alter the expression of *TTG2-like*, which eventually alters the prickle patterning

on BT stems (Figure 5). Though further functional efforts are necessary to justify this, sequence variation between the two genotypes could now be identified for application in marker-assisted-breeding of new rose varieties. We thus believe that the production of BT genome sequence paves an important foundation and serves as novel resources to study rose biology and mine molecular markers associated with important traits, hence will be particularly valuable for targeted design of breeding strategy for new rose varieties and for studying the genome evolution in Rosaceae.

Materials and methods

Plant materials and prickle density scoring

R. wichuraiana 'Basye's Thornless' (BT) plants were grown in a glasshouse without additional light in Kunming. The BC1F1 population between *R. chinensis* 'Old Blush' (OB) and BT was developed as described previously (Li et al., 2015). Briefly, a F1 progeny containing at least 98 individuals was produced via crossing the vegetatively propagated OB and BT plants. One prickleless F1 plant with a simple flower was taken as mother and pollinated by OB pollens. This produced 148 BC1F1 individuals, among which 140 were including in the mapping population. Prickles were scored for 6th-24th nodes on 4-6 main stems from at least 3 individuals for each line. Prickle density was arbitrarily defined as the prickles number on 20cm stem.

Scanning electron microscopy (SEM) analysis of stem prickles

SEM analysis was carried out for BT and OB young stems (shorter than 1cm) according to reference (Hu and Saedler, 2007) with a Zeiss Sigma 300 microscope (Germany). At least five stems from independent individuals for each genotype were examined with one typical photo shown in Figure 5.

Genomic DNA extraction, genome size and heterozygous level estimation

High-quality genomic DNA for sequencing using the Illumina platform (Illumina Inc., San Diego, CA) and PacBio Sequel sequencing (Pacific Biosciences of

California, Menlo Park, CA) was extracted from fresh leaf materials from a single BT plant. Genome size of BT was estimated with C-value measurement using a BD *FACScalibur*. Two short-insert libraries (260bp and 360bp) were constructed and sequenced to produce a total of 140.34 Gb of 150 bp paired-end raw reads using the standard protocol provided by Illumina NextSeq 500 platform. After removal of low-quality and redundant reads, we obtained 135.19 Gb (96.3%) of clean data (Supplementary Table S1). All clean reads were subjected to 17-mer frequency distribution analyses with *JELLYFISH* (v1.1.10) (Supplementary Tables S1, S2; Supplementary Figure S1) (Marcais and Kingsford, 2011). This analysis reveals two clear peaks with one at about 120x (heterozygous peak) and the other one at about 240x (homozygous peak). The heterozygosity level of BT was estimated to be about 1.03%. A pilot genome assembly with *SOAPdenovo* pipeline produced approximately 456.86 Mb but with a very high N rate at 57.98% (Supplementary Table S1, S2), indicating that only Illumina reads were not feasible for assembling genome of BT due to high level of heterozygosity.

Genome assembly using PacBio long reads

Using two 20Kb-genome DNA libraries for PacBio sequencing, 5.69 million reads (48.77 Gb) with a read N50 length of 12.83 kb after removing adaptor and low quality reads were generated and used for genome assembly (Supplementary Table S1). We combined *CANU* (v1.4; the Corrected-Error-Rate parameter was set at 0.025) (Koren et al., 2017) and *WTBDG2* (Ruan and Li, 2020) methods followed by *Quickmerge* analysis (Chakraborty et al., 2016) to generate the initial assembly (Supplementary Table S4). The resulting assembly was polished using Illumina reads that had been used for survey analysis.

Chromosome level assembly using Hi-C approach

To generate physical scaffolds of BT genome, we constructed Hi-C fragmented libraries of ~350bp and generated 19.48Gb clean reads via Illumina NovaSeq-6000 platform (~37x coverage of BT genome; Supplementary Table S1, S5). After mapping of the Hi-C reads and assignment

to restriction fragments (Servant et al., 2015), we used aligner *BWA* (version 0.7.10-r789; *aln*; other parameters, default) and identified the unique reads (~52.28%) and mapped them to above BT assembly. About 81.03% of the uniquely mapped reads were valid interaction pairs (28.07M; Supplementary Table S5). A total of 1,967 contigs (N50 length at 0.855 Mb) were generated after error correction. *LACHESIS* (parameters: *cluster_min_re_sites=52; cluster_max_link_density=2; cluster_ noninformative _ratio=2; order_min_n_res_in_trun=46; order_min_n_res_in_ shreds=42*) (Burton et al., 2013) was next used to assign the order and orientation of each contig. Finally, 521.32 Mb (99.00%) of the assemblies were anchored onto the 7 pseudomolecules, of which 481.76 Mb (92.41%) was ordered, oriented, and evaluated with *HiC-Pro* (v2.10.0) with manual correction for the contigs showing potential error in order/orientation (Supplementary Table S6; Supplementary Figure S2).

Genome quality evaluation

The completeness and quality of the BT genome were evaluated as following. We first subjected the assembly to Benchmarking Universal Single-Copy Orthologs (BUSCO, v4.1.0, eudicots_odb10) (Simão et al., 2015). Overall, 93.9% (2,184; including 1,956 complete and single copy, and 228 complete and duplicated) and 3.0% (69) of the 2,326 expected plant gene models were identified in the assembled BT genome as complete and fragmented BUSCO profiles, respectively. Approximately 73 gene models (3.1%) were considered missing in our assembly (Supplementary Table S7).

The quality of this assembly was next assessed using two independent genetic maps: the integrated K5 map based on 172 tetraploid individuals (Bourke et al., 2018), and a recently developed genetic map for 150 diploid individuals for a BC1F1 population between *R. chinensis* ‘Old Blush’ (OB) and BT (OBxBT map) (Li et al., 2019c). Genetic markers were aligned against the assembled BT genome with *blastn* (*e-value<=1e-5*) and association rates were evaluated with Pearson’s correlation test in *R*. The collinearity between the assembled BT chromosome sequences and the K5 genetic map was tested for the

association efficiency (Supplementary Figure S3; Supplementary Table S8). For the OBxBT population, except Chr4, on which the distribution of molecular markers was obviously skewed (Li et al., 2019c), all the other 6 chromosomes also showed a very good collinearity with correlation rate above 89% (Supplementary Figure S4).

We aligned the BT assembly to the haploOB genome sequences using *Mummer4* pipeline (*nucmer –t 20, -L 12000*) (Marcais et al., 2018b). This analysis identified a very high mapping rate of BT assembly to both Raymond's and Hibrand's (Figure 2; Supplementary Figure S3) reference genomes, indicating a high synteny level and gene order between these two genotypes. Finally, the BT assembly was examined for its *LTR assembly index* (*LAI*) (Ou et al., 2018). On average, the LAI was about 20.04, a level of golden reference. Additionally, we judged the assembly quality by mapping our Illumina and PacBio reads back to the genomes (Supplementary Table S9).

Annotation of repeat sequences in BT genome

To dissect the molecular makeups of the BT genome, we next constructed a de novo repeat database with *LTR FINDER* v1.05 (Xu and Wang, 2007), *RepeatScout* v1.0.5 (Price et al., 2005), *PILER-DF* v2.4 (Edgar and Myers, 2005) programs. We classified these repeats with *PASTEClassifier* (Wicker et al., 2007) and merged them with the *Repbase data* (Jurka et al., 2005). *RepeatMasker* (version 4.0.6) (Chen, 2004) was then applied to identify repeat sequence against the *de novo* repeat library.

Gene annotation

Gene annotation of the BT genome was carried out using *de novo*, homology-based, and transcriptome profiling based predictions (Supplementary Table S13). We employed *Genscan* (Ghosh and Chan, 2016), *Augustus* v2.4 (Trapnell et al., 2009), *Glimmer HMM* v3.0.4 (Majoros et al., 2004), *GeneID* v1.4 (Alioto et al., 2018), *SNAP* (version 2006-07-28) (Korf, 2004) softwares to predict protein coding genes. We used *GeMoMa* v1.3.1 (Keilwagen et al., 2016) to predict genes with homology to *Arabidopsis thaliana*,

Oryza sativa, *Fragaria vesca*, *Malus domestica*, and *Pyrus bretschneideri* gene models. We next mapped Illumina transcriptome short reads onto the genome assembly and predicted gene models with the *PASA* v2.0.2 (Campbell et al., 2006), *transDecoder* (http://transdecoder.github.io) and *GeneMarkS-T* (Tang et al., 2015) pipelines. The transcriptome data was produced by Illumina sequencing of a pool of materials including leaves, shoots, flowers and roots and from publication by Li et al. (Li et al., 2018). All gene models were then integrated using *EVidenceModeler* v1.1.1 (Haas et al., 2008) to generate a consensus gene set.

We carried out functional annotation of protein-coding genes via aligning to the NR, GO, KEGG, KOG, TrEMBL databases with *BLAST* (v2.2.31) with an *e-value cutoff of* 10^{-5} *(Griffiths-Jones et al., 2003)*. We also used *Blast2GO* software (Dimmer et al., 2012) to search the GO and KEGG databases. Ultimately, 97.63% (31,898 genes) of the 32,674 genes were annotated by at least one database (Supplementary Table S12).

Additionally, four types of non-coding RNAs (microRNAs, transfer RNAs, ribosomal RNAs, and small nuclear RNAs) were annotated using the *tRNAscan-SE* (Lowe and Chan, 2016) and *Infenal* v1.1 (Nawrocki and Eddy, 2013) against the *Rfam* (Kalvari et al., 2018) as well as the *miRbase* databases (Kozomara and Griffiths-Jones, 2014). *GenBlastA* (She et al., 2009) was further used for pseudogene predication by scanning the BT genome for sequences homologous to the known protein coding genes it contained, while *GeneWise* (Birney et al., 2004) was employed to search for premature stop codons or frame shift mutations in those sequences.

Gene family identification and phylogenetic relationships of Rosaceae

OrthoMCL (v2.0.9) (Li et al., 2003) was applied to identify the conserved orthologous gene families within genomes of BT, OB, *R. multiflora*, *Malus domestica*, *Fragaria vesca*, *Prunus avium*, *Prunus persica*, *Pyrus communis*, and *Rubus occidentalis* taken *Vitis venifera* and *Ziziphus jujube* as outgroups. A total of 1,220 single-copy orthologous groups was detected and used for

sequence concatenation and alignment with *MUSCLE* (Edgar, 2004). The concatenated sequences were next applied in *PHYML* (v20151210) to construct a maximum-likelihood tree with 500 times of bootstrap supports tested (Guindon et al., 2010). To estimate the divergence time between BT and OB, the C1 and C3 fossil nodes were used for dating (Zhang et al., 2017). Gene family expansion/contraction analysis along lineages on the phylogenetic tree was carried out with *CAFE* (v4.2) (De Bie et al., 2006).

Alignment, sequence variation and synteny analyses between BT and haploOB genomes

Genome alignment between BT and haploOBs was performed using *MUMmer* (version 4)(Marcais et al., 2018a) program with parameters settings *–maxmatch –c 90 –l 40*. The alignments were filtered by delta-filter with parameter *−1*. SNPs and indels in the two species were extracted with *show-snp* in the one-to-one alignment blocks. The clean Illumina reads were mapped to BT genome with *BWA* (version 0.7.10, - *r789*) software to identify SNPs and small indels within BT genome using the *Genome Analysis Toolkit* (*GATK*, version 4.0.9) (McKenna et al., 2010) and *Samtools* (version 0.1.19) (Li et al., 2009). Only the variants supported by at least two tools were used for further analysis. All these variants were annotated using the *ANNOVAR* program (Wang et al., 2010b). An All-vs-All *blastp* (*e-value <1e-10*) was carried out for each genome pair, and then homologous genes were identified using the *MCScanX* with default settings except *gap_penality* as *-3*. Syntenic blocks were defined as the genome region featuring at least five collinearity genes. Linear parallel plots in *SynVisio* (*https://github.com/kiranbandi/synvisio*) were used to visualize the conserved regions in the collinearity data generated with *MCScanX* using default parameters except *e-value* (=1e-5) and *maximum gaps* (=25) with contigs not anchored/oriented on chromosomes ignored.

For genomic rearrangement analysis, we applied *minimap2* (Li, 2018) (*-ax asm5 –eqx*) to generate the local pairwise alignment between BT and haploOB genomes. *SyRI* (Goel et al., 2019) (*-k -F S*) was used to identify the rearrangement events including inversion region (INV), translocated region

(TRANS), and copy gain (CPG) and loss (CPL) in OB in reference to BT genome. Data was plotted as rearrangement length per 100 kb of step size in one Mb window in *R*.

Linkage map and QTL analysis

A previously published genetic map was used in this study (Li et al., 2019c). The map was constructed with JoinMap 4.0 (Van Ooijen, 2006) with high quality genotyping data (140 individuals, missing data rate < 6%) developed by RAD-seq. QTL analysis was performed using *MapQTL* 6.0 software (Kyazma) by combining linkage map and phenotypic identification data. Permutation Test (1000 repetitions, α= 0.05) was performed to identify 3.4 as the LOD threshold for the presence of a QTL. QTL were identified with both Interval Mapping (IM) and composite interval mapping (MQM) methods. Positions with the highest LOD values on the linkage group were taken as location of the QTL, and interval of 2-LOD is used as the 95% confidence interval. Via *BLAST* search (*blast+, e* ≤ 10^{-10}) of the markers near the QTL intervals within reference genomes, we identified candidate genes and analyzed their function.

Resequencing, SNP identification, and sequence diversity analyses

To investigate sequence variation and diversity of prickleless and prickly populations, we identified three genotypes featuring no prickle on stem in about 200 genotypes growing in Kunming Botanical Garden, KIB, CAS. These include a genotype from species *R. banksiae* (f. *lutea*), and two cultivated genotypes (*R*. cv Emira and Star 'n' Strips) (Supplementary Figure S23). We collected young leaves of these genotypes and obtained about 20 Gb clean reads via Illumina sequencing (Supplementary Table S27). We additionally collected clean reads of eight genotypes featuring various level of prickle density on their shoots (NCBI SRP119986).

For SNP detection, we mapped the clean reads to BT genome using *BWA-MEM* with default parameters (Hassan et al., 2010). SNPs were called with *VCFtools* and *BCFtools* and evaluated with *FreeBayes* (Danecek et al.,

2011; Narasimhan et al., 2016). In order to reduce mapping bias, we discarded the following sites: 1) sites with coverage below 200x in all sample or above 3000x in any sample; 2) sites with more than eight reads but the mapping score equal to zero in any sample; 3) N sites present in more than 50% of the individuals; 4) polymorphic sites present in less than three individuals. We further treated the genotypes as missing when mapping quality score below 20 and coverage lower than 5.

To estimate the sequence diversity and fixation levels in prickleless and prickly sub-populations, we calculated weighted *Fst* and *Pi* values using *VCFtools* with window sizes of 10, 50, 100, 200, and 300 kb using *Weir and Cockerham* as estimator, and *BayeScan* in outlier test. *Dxy* estimation was carried out with custom python scripts based on formula: $Dxy= 1/n \sum Pi_{prickly}(1-Pi_{prickleless}) - Pi_{prickleless}(1-Pi_{prickly})$, in which n=2 (Smith and Kronforst, 2013).

Identification of prickle genes using *Arabidopsis* homologs

We collected 151 genes related to trichome development and formation in Arabidopsis genome (TAIR10). Protein sequences of these genes were extracted and used as seed sequence to *blastp* the BT and haploOB genomes (*e-value* 10^{-7}, *minimum identity 0.35, minimum length coverage 0.6*). Ortholog gene pairs were then identified using OrthoFinder2 (Emms and Kelly, 2019) in combination with the *blastp* results and manual correction based on their genomic synteny. To investigate further the relationship of these gene pairs, we constructed a Neighbor-Joining (NJ) phylogeny with *FastTree* 2.1 (Price et al., 2010) based on sequence alignments using *ClustalW* 2.1. By combining phylogentic position and syntenic information of these gene pairs, we examined the sequence variation between orthologous gene pairs by testing the following types of mutation: a) protein and nucleic sequence completeness, b) small insertions and deletions shifting reading frame, c) large frame-preserving insertions (those creating premature stop codons), d) in-frame stop codons. We additionally validated these mutational events with Illumina reads to minimize sequencing and assembly errors in these gene pairs.

Selection pressure analysis

Non-synonymous (*Ka*) and synonymous (*Ks*) substation rates were estimated for testing the selection pressure acting on orthologous genes based on *MCScan* results as described above. *MUSCLE* was used to generate the nucleotide and protein sequences alignments, which were further used to calculate the pairwise *Ka/Ks* values with *KaKs Calculator2* with Yang-Nielsen (YN) model (Wang et al., 2010a). Only these gene pairs featuring a significant deviation of *Ka/Ks* from 1.0 were treated as positively selected genes. *Ks* values in QTL1 region were plotted in *R* in comparison to other parts of the whole genome (Figure 4).

Gene expression analyses

Leaf and young shoot materials were harvested and sequenced with Illumina platforms as described by Li et al. (Li et al., 2018). Clean reads were mapped to BT and haploOB genomes with *HISAT2* (Kim et al., 2015) and assembled with *StringTie* (Pertea et al., 2015). Uniquely mapped reads (default parameters expect for *--stranded = no*) were used to calculate the relative expression in FPKM with *ballgown* pipeline in *R*.

Relative water content measurement and statistical analysis

Two modern rose genotypes were first used to examine the relative water content in prickle, bark, and stele together with leaves on the same sections of shoots. Normally, four sections with each containing three nodes on young shoot that grew at current seasons, or on old shoot, which was produced in previous year, were measured for more than eight shoots. Prickles were quickly tweezed off to measure the fresh weight. Post removing prickles, bark and stele tissues as well as leaves were quickly sectioned with a blade. After the measuring of fresh weight, all tissues were dried for three days at 80 degree in an oven supplied with a wind circulator. During drying, the weight of stele materials was monitored several times and finally showed steady state. Dry weight was then quickly measured and the relative water content was calculated as (fresh weight – dry weight)/fresh weight. For the relative water

content in a rose collection including 12 wild, 21 modern rose genotypes and 6 China roses, two to three sections (each containing three nodes) on one shoot were cut and analyzed. Statistical comparisons were done with *Wilcoxon* and *Kruskal-Wallis* tests in *R* or *Student's t-test* in excel.

Chloroplast genome assembly and phylogenetic analysis

The chloroplast genome sequences were reconstructed for the three prickleless and eight prickly genotypes (supplementary Table S27) following previously published methods (Bankevich et al., 2012; Camacho et al., 2009; Cui et al., 2020; Langmead and Salzberg, 2012; Li et al., 2019b; Wick et al., 2015). Final genomes were adjusted and annotated using Geneious 9.1.4 (https://www.geneious.com), with the BT chloroplast genome as a reference (Cui et al., 2020). Sequences of other rose genotypes were obtained from NCBI. *Rubus crataegifolius* and *Potentilla freyniana* were used as outgroups. Sequences were aligned using MAFFT (Katoh et al., 2002), and input to RAxML 8.2.11 (Stamatakis, 2014) for phylogeny reconstruction using maximum-likelihood method (*GTRGAMMAI* substitution model, 100 times of bootstrap simulation).

GO enrichment and other statistical analysis

Gene ontology enrichment was analyzed by comparing the interested genes with the genome-wide gene list of BT or haploOB using a custom *perl* script. The p-values generated with a two-sided *Fisher's exact test* were further corrected for multiple comparisons using *Benjamini and Hochberg correction* (Benjamini and Hochberg, 1995). Only GO terms with a false discovery rate of less than 0.05 (FDR <0.05) were retained. Two-sided *Wilcoxon test* was used in Figure 4, *Kruskal-Wallis test* was used in multi-sample comparisons such as Supplementary Table 21.

Availability of supporting data

All data supporting the results of this study are included in the manuscript and its additional files. Genome assembly and annotation were deposited in the NCBI BioProject under the accessions PRJNA542225.

Competing interests

The authors declare no competing interest.

Funding

Research in this work was supported by the following grants: the Strategic Priority Research Program of the Chinese Academy of Sciences to J-Y H and D-Z L (XDB31000000); the CAS Pioneer Hundred Talents Program to J-Y H (292015312D11035); Yunnan Applied Basic Research Projects to X D (2016FB040); CAS Key Laboratory for Plant Diversity and Biogeography of East Asia, and Yunnan Recruitment Program of Experts in Science to J-Y H.

Ethics statement

Not applicable.

Author contributions

J-Y H conceptualized the project; J-Y H, D-Z L and K T coordinated the research. X J, W W, Y S, D W, W C, and Z S collected the samples, extracted the genomic DNA and total RNA, carried out the prickle phenotyping, X J did the quantitative RT-PCR. M Z, X D, G Y, X C and X L generated, analyzed, and visualized the data. J-Y H wrote the paper with the help from D-Z L and K T as well as other authors. All authors have read and approved the final manuscript.

Acknowledgements

We thank Prof. David H. Byrne for BT materials. We appreciate Dr. Shubin Li, Ms. Shulan Chen, Mrs. Hongyuan Yu and Yuanlin Lv for the help in cultivating the plants. We acknowledge Prof. Zhenhua Guo for useful discussion. This work is partially facilitated by the Germplasm Bank of Wild Species and the Kunming Botanical Garden, Kunming Institute of Botany, Chinese Academy of Sciences.

References

Albert, N.W., Davies, K.M., Lewis, D.H., Zhang, H., Montefiori, M., Brendolise, C., Boase, M.R., Ngo, H., Jameson, P.E., and Schwinn, K.E. (2014). A Conserved Network of Transcriptional Activators and Repressors Regulates Anthocyanin Pigmentation in Eudicots. The Plant Cell *26*, 962-980.

Alioto, T., Blanco, E., Parra, G., and Guigo, R. (2018). Using geneid to Identify Genes. Curr Protoc Bioinformatics *64*, e56.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol *19*, 455-477.

Baudry, A., Heim, M.A., Dubreucq, B., Caboche, M., Weisshaar, B., and Lepiniec, L. (2004). TT2, TT8, and TTG1 synergistically specify the expression of BANYULS and proanthocyanidin biosynthesis in Arabidopsis thaliana. Plant Journal *39*, 366-380.

Bendahmane, M., Dubois, A., Raymond, O., and Bris, M.L. (2013). Genetics and genomics of flower initiation and development in roses. Journal of Experimental Botany *64*, 847-857.

Benjamini, Y., and Hochberg, Y. (1995). CONTROLLING THE FALSE DISCOVERY RATE - A PRACTICAL AND POWERFUL APPROACH TO MULTIPLE TESTING. J. R. Stat. Soc. Ser. B-Stat. Methodol. *57*, 289-300.

Bennetzen, J.L., Schmutz, J., Wang, H., Percifield, R., Hawkins, J., Pontaroli, A.C., Estep, M., Feng, L., Vaughn, J.N., Grimwood, J., *et al.* (2012). Reference genome sequence of the model plant Setaria. Nat Biotech *30*, 555-561.

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. Genome Res *14*, 988-995.

Bourke, P.M., Gitonga, V.W., Voorrips, R.E., Visser, R.G.F., Krens, F.A., and Maliepaard, C. (2018). Multi-environment QTL analysis of plant and flower morphological traits in tetraploid rose. Theoret. Appl. Genetics *131*, 2055-2069.

Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R.L., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nature Biotechnology *31*, 1119-+.

Byrne, D.H., Anderson, N., and Pemberton, H.B. (2007). The use of Rosa wichurana in the development of landscape roses adapted to hot humid climates. In Proceedings of the Ivth International Symposium on Rose Research and Cultivation, H.B. Pemberton, ed. (Leuven 1: Int Soc Horticultural Science), pp. 267-+.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST plus : architecture and applications. Bmc Bioinformatics *10*.

Campbell, M.A., Haas, B.J., Hamilton, J.P., Mount, S.M., and Buell, C.R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with Arabidopsis. BMC Genomics *7*, 327.

Chakraborty, M., Baldwin-Brown, J.G., Long, A.D., and Emerson, J.J. (2016). Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. Nucleic Acids Research *44*, 12.

Che, G., and Zhang, X. (2019). Molecular basis of cucumber fruit domestication. Current Opinion in Plant Biology *47*, 38-46.

Chen, N. (2004). Using RepeatMasker to identify repetitive elements in genomic sequences. Current protocols in bioinformatics *Chapter 4*, Unit 4.10.

Chopra, D., Mapar, M., Stephan, L., Albani, M.C., Deneer, A., Coupland, G., Willing, E.M., Schellmann, S., Schneeberger, K., Fleck, C., *et al.* (2019). Genetic and molecular analysis of trichome development in Arabis alpina. P Natl Acad Sci USA *116*, 12078-12083.

Coverdale, T.C. (2020). Defence emergence during early ontogeny reveals important differences between spines, thorns and prickles. Annals of botany *124*, iii-iv.

Crespel, L., Chirollet, M., Durel, C., Zhang, D., Meynet, J., and Gudin, S. (2002). Mapping of qualitative and quantitative phenotypic traits in Rosa using AFLP markers. Theoret. Appl. Genetics *105*, 1207-1214.

Cui, W.-H., Zhong, M.-C., Du, X.-Y., Qu, X.-J., Jiang, X.-D., Sun, Y.-B., Wang, D., Chen, S.-Y., and Hu, J.-Y. (2020). The complete chloroplast genome sequence of a rambler rose, Rosa wichuraiana (Rosaceae). Mitochondrial DNA Part B *5*, 252-253.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., *et al.* (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156-2158.

De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269-1271.

Dimmer, E.C., Huntley, R.P., Alam-Faruque, Y., Sawford, T., O'Donovan, C., Martin, M.J., Bely, B., Browne, P., Mun Chan, W., Eberhardt, R., *et al.* (2012). The UniProt-GO Annotation database in 2011. Nucleic Acids Res *40*, D565-570.

Dong, X., Jiang, X., Kuang, G., Wang, Q., Zhong, M., Jin, D., and Hu, J.-Y. (2017). Genetic control of flowering time in woody plants: roses as an emerging model. Plant Diversity *39*, 104-110.

Doroshkov, A.V., Konstantinov, D.K., Afonnikov, D.A., and Gunbin, K.V. (2019). The evolution of gene regulatory networks controlling Arabidopsis thaliana L. trichome development. BMC Plant Biology *19*, 53.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res *32*, 1792-1797.

Edgar, R.C., and Myers, E.W. (2005). PILER: identification and classification of genomic repeats. Bioinformatics *21*, I152-I158.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology *20*, 238.

Foucher, F., Hibrand-Saint Oyant, L., Hamama, L., Sakr, S., Nybom, H., Baudino, S., Caissard, J.P., Hokanson, S.C., Byrne, D.M., Smulder, J.M.S., *et al.* (2015). Towards the Rose Genome Sequence and Its Use in Research and Breeding. In Vi International Symposium on Rose Research and Cultivation, T. Debener, and M. Linde, eds. (Leuven 1: Int Soc Horticultural Science), pp. 167-175.

Fougere-Danezan, M., Joly, S., Bruneau, A., Gao, X.-F., and Zhang, L.-B. (2015). Phylogeny and biogeography of wild roses with specific attention to polyploids. Annals of Botany *115*, 275-291.

Gan, Y., Kumimoto, R., Liu, C., Ratcliffe, O., Yu, H., and Broun, P. (2006). GLABROUS INFLORESCENCE STEMS Modulates the Regulation by Gibberellins of Epidermal Differentiation and Shoot Maturation in Arabidopsis. The Plant Cell *18*, 1383-1395.

Ghosh, S., and Chan, C.-K.K. (2016). Analysis of RNA-Seq Data Using TopHat and Cufflinks. Methods in molecular biology (Clifton, N.J.) *1374*, 339-361.

Goel, M., Sun, H.Q., Jiao, W.B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. Genome Biology *20*, 13.

Gonzalez, A., Mendenhall, J., Huo, Y., and Lloyd, A. (2009). TTG1 complex MYBs, MYB5 and TT2, control outer seed coat differentiation. Developmental Biology *325*, 412-421.

Griffiths-Jones, S., Bateman, A., Marshall, M., Khanna, A., and Eddy, S.R. (2003). Rfam: an RNA family database. Nucleic Acids Res *31*, 439-441.

Gu, C., Li, C., Lu, L., Jiang, S., Crinan, A., Bruce, B., Anthony R., B., David E., B., Hiroshi, I., Hideaki, O., *et al.* (2006). Flora of China, Volume 9: Rosaceae. Edited by Wu Zhengyi, Raven Peter H., Hong Deyuan. Sciense Press and Missouri Botanical Garden Press, Beijing and St. Louis, Vol 9.

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. Systematic biology *59*, 307-321.

Guo, C., Wang, Y., Yang, A., He, J., Xiao, C., Lv, S., Han, F., Yuan, Y., Yuan, Y., Dong, X., *et al.* (2019). The Coix genome provides insights into Panicoideae evolution and papery hull domestication. Molecular Plant.

Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol *9*, R7.

Hassan, M.Q., Gordon, J.A., Beloti, M.M., Croce, C.M., van Wijnen, A.J., Stein, J.L., Stein, G.S., and Lian, J.B. (2010). A network connecting Runx2, SATB2, and the miR-23a~27a~24-2 cluster regulates the osteoblast differentiation program. Proc Natl Acad Sci U S A *107*, 19879-19884.

Hibrand Saint-Oyant, L., Ruttink, T., Hamama, L., Kirov, I., Lakhwani, D., Zhou, N.N., Bourke, P.M., Daccord, N., Leus, L., Schulz, D., *et al.* (2018). A high-quality genome sequence of Rosa chinensis to elucidate ornamental traits. Nature Plants *4*, 473-484.

Hu, J.-Y., and Saedler, H. (2007). Evolution of the Inflated Calyx Syndrome in Solanaceae. Mol Biol Evol *24*, 2443-2453.

Ishida, T., Hattori, S., Sano, R., Inoue, K., Shirano, Y., Hayashi, H., Shibata, D., Sato, S., Kato, T., Tabata, S., *et al.* (2007). Arabidopsis TRANSPARENT TESTA GLABRA2 Is Directly Regulated by R2R3 MYB Transcription Factors and Is Involved in Regulation of GLABRA2 Transcription in Epidermal Differentiation. The Plant Cell *19*, 2531-2543.

Johnson, C.S., Kolevski, B., and Smyth, D.R. (2002). TRANSPARENT TESTA GLABRA2, a trichome and seed coat development gene of Arabidopsis, encodes a WRKY transcription factor. Plant Cell *14*, 1359-1375.

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.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., and Petrov, A.I. (2018). Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res *46*, D335-d342.

Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research *30*, 3059-3066.

Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res *44*, e89.

Kellogg, A.A., Branaman, T.J., Jones, N.M., Little, C.Z., and Swanson, J.D. (2011). Morphological studies of developing Rubus prickles suggest that they are modified glandular trichomes. Botany-Botanique *89*, 217-226.

Kim, D., Landmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nature Methods *12*, 357-U121.

Koning-Boucoiran, C.F.S., Gitonga, V.W., Yan, Z., Dolstra, O., van der Linden, C.G., van der Schoot, J., Uenk, G.E., Verlinden, K., Smulders, M.J.M., Krens, F.A., *et al.* (2012). The mode of inheritance in tetraploid cut roses. Theoret. Appl. Genetics *125*, 591-607.

Koopman, W.J., Volker, W., Katrien, D.C., Johan, V.H., Jan, D.R., JH, S.G., Dirk, V., Ben, V., M, R.C., Bert, M., *et al.* (2008). AFLP markers as a tool to reconstruct complex relationships: A case study in Rosa (Rosaceae). Am. J. Bot. *95*, 353-366.

Koren, S., Walenz, 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 Research *27*, 722-736.

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

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

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods *9*, 357-U354.

Li, C., Song, W., Luo, Y., Gao, S., Zhang, R., Shi, Z., Wang, X., Wang, R., Wang, F., Wang, J., *et al.* (2019a). The HuangZaoSi Maize Genome Provides Insights into Genomic Variation and Improvement History of Maize. Molecular Plant.

Li D, G.S., Zhao H, Wang X (2016). Research advances in the genetic relationship of rose germplasm resources. Journal of Biology *33*, 103-105.

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics *34*, 3094-3100.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Proc, G.P.D. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, L., Stoeckert, C.J., and Roos, D.S. (2003). OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research *13*, 2178-2189.

Li, S., Qu, X., Zhong, M., Jiang, X., Dong, X., Yi, T., Tang, K., Dai, S., and Hu, J.Y. (2019b). Characterization of the complete chloroplast genome of Rosa chinensis 'Old Blush' (Rosaceae), an important cultivated Chinese rose. Acta Hortic. *1232*, 119-124.

Li, S., Yang, G., Yang, S., Just, J., Yan, H., Zhou, N., Jian, H., Wang, Q., Chen, M., Qiu, X., *et al.* (2019c). The development of a high-density genetic map significantly improves the quality of reference genome assemblies for rose. Scientific reports *9*, 5985.

Li, S., Zhong, M., Dong, X., Jiang, X., Xu, Y., Sun, Y., Cheng, F., Li, D.-z., Tang, K., Wang, S., *et al.* (2018). Comparative transcriptomics identifies patterns of selection in roses. BMC Plant Biology *18*, 371.

Li, S., Zhou, N., Zhou, Q., Yan, H., Jian, H., Wang, Q., Chen, M., Qiu, X., Zhang, H., Wang, S., *et al.* (2015). Inheritance of perpetual blooming in *Rosa chinensis* 'Old Blush'. Horticultural Plant Journal *1*, 108-112.

Linde, M., Hattendorf, A., Kaufmann, H., and Debener, T. (2006). Powdery mildew resistance in roses: QTL mapping in different environments using selective genotyping. Theoret. Appl. Genetics *113*, 1081-1092.

Liu, H., Shi, J., Cai, Z., Huang, Y., Lv, M., Du, H., Gao, Q., Zuo, Y., Dong, Z., Huang, W., *et al.* (2019). Evolution and domestication footprints uncovered from the genomes of coix. Molecular Plant.

Lowe, T.M., and Chan, P.P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res *44*, W54-57.

Lu, M., An, H., and Li, L. (2016). Genome Survey Sequencing for the Characterization of the Genetic Background of Rosa roxburghii Tratt and Leaf Ascorbate Metabolism Genes. PLOS ONE *11*, e0147530.

Majoros, W.H., Pertea, M., and Salzberg, S.L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics *20*, 2878-2879.

Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018a). MUMmer4: A fast and versatile genome alignment system. Plos Comput Biol *14*.

Marcais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018b). MUMmer4: A fast and versatile genome alignment system. PLoS Comput. Biol. *14*, 14.

Marcais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764-770.

Matias-Hernandez, L., Aguilar-Jaramillo, A.E., Cigliano, R.A., Sanseverino, W., and Pelaz, S. (2016a). Flowering and trichome development share hormonal and transcription factor regulation. Journal of Experimental Botany *67*, 1209-1219.

Matias-Hernandez, L., Aguilar-Jaramillo, A.E., Osnato, M., Weinstain, R., Shani, E., Suarez-Lopez, P., and Pelaz, S. (2016b). TEMPRANILLO Reveals the Mesophyll as Crucial for Epidermal Trichome Formation. Plant Physiology *170*, 1624-1639.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *et al.* (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res *20*, 1297-1303.

Meng, J., Fougère-Danezan, M., Zhang, L.-B., Li, D.-Z., and Yi, T.-S. (2011). Untangling the hybrid origin of the Chinese tea roses: evidence from DNA sequences of single-copy nuclear and chloroplast genes. Plant Systematics and Evolution *297*, 157-170.

Nakamura, N., Hirakawa, H., Sato, S., Otagaki, S., Matsumoto, S., Tabata, S., and Tanaka, Y. (2017). Genome structure of Rosa multiflora, a wild ancestor of cultivated roses. DNA Research *25*, 113-121.

Narasimhan, V., Danecek, P., Scally, A., Xue, Y.L., Tyler-Smith, C., and Durbin, R. (2016). BCFtools/RoH: a hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics *32*, 1749-1751.

Nawrocki, E.P., and Eddy, S.R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics *29*, 2933-2935.

Ou, S.J., Chen, J.F., and Jiang, N. (2018). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Research *46*, 11.

Pattanaik, S., Patra, B., Singh, S.K., and Yuan, L. (2014). An overview of the gene regulatory network controlling trichome development in the model plant, Arabidopsis. Frontiers in plant science *5*.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nature Biotechnology *33*, 290-+.

Pesch, M., and Hülskamp, M. (2009). One, two, three…models for trichome patterning in Arabidopsis? Current Opinion in Plant Biology *12*, 587-592.

Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21*, I351-I358.

Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2-Approximately Maximum-Likelihood Trees for Large Alignments. Plos One *5*, 10.

Raymond, O., Gouzy, J., Just, J., Badouin, H., Verdenaud, M., Lemainque, A., Vergne, P., Moja, S., Choisne, N., Pont, C., *et al.* (2018). The Rosa genome provides new insights into the domestication of modern roses. Nature Genetics *50*, 772-777.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat Methods *17*, 155-158.

Serna, L., and Martin, C. (2006). Trichomes: different regulatory networks lead to convergent structures. Trends in Plant Science *11*, 274-280.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.J., Vert, J.P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biology *16*, 11.

She, R., Chu, J.S., Wang, K., Pei, J., and Chen, N. (2009). GenBlastA: enabling BLAST to identify homologous gene sequences. Genome Res *19*, 143-149.

Shi, J.P., Ma, X.X., Zhang, J.H., Zhou, Y.S., Liu, M.X., Huang, L.L., Sun, S.L., Zhang, X.B., Gao, X., Zhan, W., *et al.* (2019). Chromosome conformation capture resolved near complete genome assembly of broomcorn millet. Nature communications *10*, 9.

Shupert, D.A., Byme, D.H., and Pemberton, H.B. (2007). Inheritance of flower traits, leaflet number and prickles in roses. In Proceedings of the Ivth International Symposium on Rose Research and Cultivation, H.B. Pemberton, ed. (Leuven 1: Int Soc Horticultural Science), pp. 331-+.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210-3212.

Smith, J., and Kronforst, M.R. (2013). Do Heliconius butterfly species exchange mimicry alleles? Biol Lett *9*, 20130503.

Smulders, M.J.M., Arens, P., Bourke, P.M., Debener, T., Linde, M., Riek, J.D., Leus, L., Ruttink, T., Baudino, S., Hibrant Saint-Oyant, L., *et al.* (2019). In the name of the rose: a roadmap for rose research in the genome era. Horticulture Research *6*, 65.

Spiller, M., Linde, M., Hibrand-Saint Oyant, L., Tsai, C.J., Byrne, D.H., Smulders, M.J.M., Foucher, F., and Debener, T. (2011). Towards a unified genetic map for diploid roses. Theoret. Appl. Genetics *122*, 489-500.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312-1313.

Sun, S., Zhou, Y., Chen, J., Shi, J., Zhao, H., Zhao, H., Song, W., Zhang, M., Cui, Y., Dong, X., *et al.* (2018). Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. Nature Genetics.

Szymanski, D.B., Lloyd, A.M., and Marks, M.D. (2000). Progress in the molecular genetic analysis of trichome initiation and morphogenesis in Arabidopsis. Trends in Plant Science *5*, 214-219.

Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res *43*, e78.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Van Ooijen, J. (2006). JoinMap 4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma BV, Wageningen, The Netherlands.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010a). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics *8*, 77-80.

Wang, K., Li, M., and Hakonarson, H. (2010b). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res *38*, e164.

Wang, L., Zhou, C.M., Mai, Y.X., Li, L.Z., Gao, J., Shang, G.D., Lian, H., Han, L., Zhang, T.Q., Tang, H.B., *et al.* (2019a). A spatiotemporally regulated transcriptional complex underlies heteroblastic development of leaf hairs in Arabidopsis thaliana. Embo Journal *38*, 12.

Wang, Z., Miao, H., Liu, J., Xu, B., Yao, X., Xu, C., Zhao, S., Fang, X., Jia, C., Wang, J., *et al.* (2019b). Musa balbisiana genome reveals subgenome evolution and functional divergence. Nature Plants *5*, 810-821.

Wang, Z., Yang, Z., and Li, F. (2019c). Updates on molecular mechanisms in the development of branched trichome in Arabidopsis and nonbranched in cotton. Plant Biotechnology Journal *0*.

Wick, R.R., Schultz, M.B., Zobel, J., and Holt, K.E. (2015). Bandage: interactive visualization of de novo genome assemblies. Bioinformatics *31*, 3350-3352.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., *et al.* (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics *8*, 973-982.

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Research *35*, W265-W268.

Xue, X.Y., Zhao, B., Chao, L.M., Chen, D.Y., Cui, W.R., Mao, Y.B., Wang, L.J., and Chen, X.Y. (2014). Interaction between Two Timing MicroRNAs Controls Trichome Distribution in Arabidopsis. Plos Genetics *10*, 13.

Yang, C., and Ye, Z. (2013). Trichomes as models for studying plant cell differentiation. Cellular and Molecular Life Sciences *70*, 1937-1948.

Yang, N., Liu, J., Gao, Q., Gui, S., Chen, L., Yang, L., Huang, J., Deng, T., Luo, J., He, L., *et al.* (2019). Genome assembly of a tropical maize inbred line provides insights into structural variation and crop improvement. Nature Genetics *51*, 1052-1059.

Yu, N., Cai, W.J., Wang, S., Shan, C.M., Wang, L.J., and Chen, X.Y. (2010). Temporal control of trichome distribution by microRNA156-targeted SPL genes in Arabidopsis thaliana. Plant Cell *22*, 2322-2335.

Zhang, J., Esselink, G.D., Che, D., Fougere-Danezan, M., Arens, P., and Smulders, M.J.M. (2013). The diploid origins of allopolyploid rose species studied using single nucleotide polymorphism haplotypes flanking a microsatellite repeat. J. Horticult. Sci. Biotechnol. *88*, 85-92.

Zhang, N., Yang, L., Luo, S., Wang, X., Wang, W., Cheng, Y., Tian, H., Zheng, K., Cai, L., and Wang, S. (2018). Genetic evidence suggests that GIS functions downstream of TCL1 to regulate trichome formation in Arabidopsis. BMC Plant Biology *18*, 63.

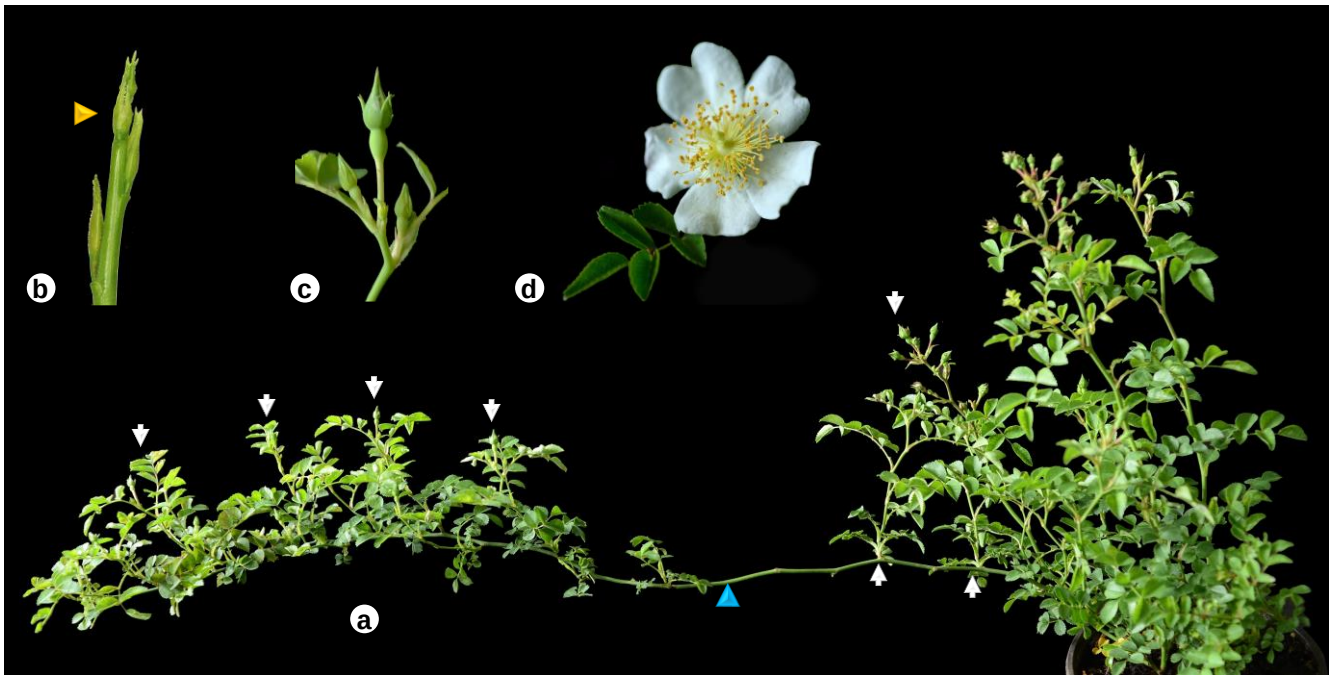
Zhang, S.-D., Jin, J.-J., Chen, S.-Y., Chase, M.W., Soltis, D.E., Li, H.-T., Yang, J.-B., Li, D.-Z., and Yi, T.-S. (2017). Diversification of Rosaceae since the Late Cretaceous based on plastid phylogenomics. New Phytologist *214*, 1355-1367.

Zhu, Z.-M., Gao, X.-F., and Fougère-Danezan, M. (2015). Phylogeny of Rosa sections Chinenses and Synstylae (Rosaceae) based on chloroplast and nuclear markers. Mol Phylogenet Evol *87*, 50-64.

Table 1. Assembly statistics of *R. wichuraiana* ‘Basyes’ Thornless’ (BT).

assembly features	**statistic**
estimated genome size (by K-mer=17 analysis)	525,536,719 bp
number of contigs	1,847
contig N50	932,234 bp
longest contig length	8,250,845 bp
assembled genome size	530,071,794 bp
number of pseudo-chromosomes	7
total assembly on chromosomes	483,888,895 bp (91.29%)
contig number not ordered/oriented	1,093
contig length not ordered/oriented	46,182,899 bp (8.71%)
repeat proportion	345,577,859 bp (65.19%)
number of predicted protein coding gene models	32,674
average length per coding gene	3,281.56 bp
average exons per gene model	5.22
average exon length	250.83 bp

A



B

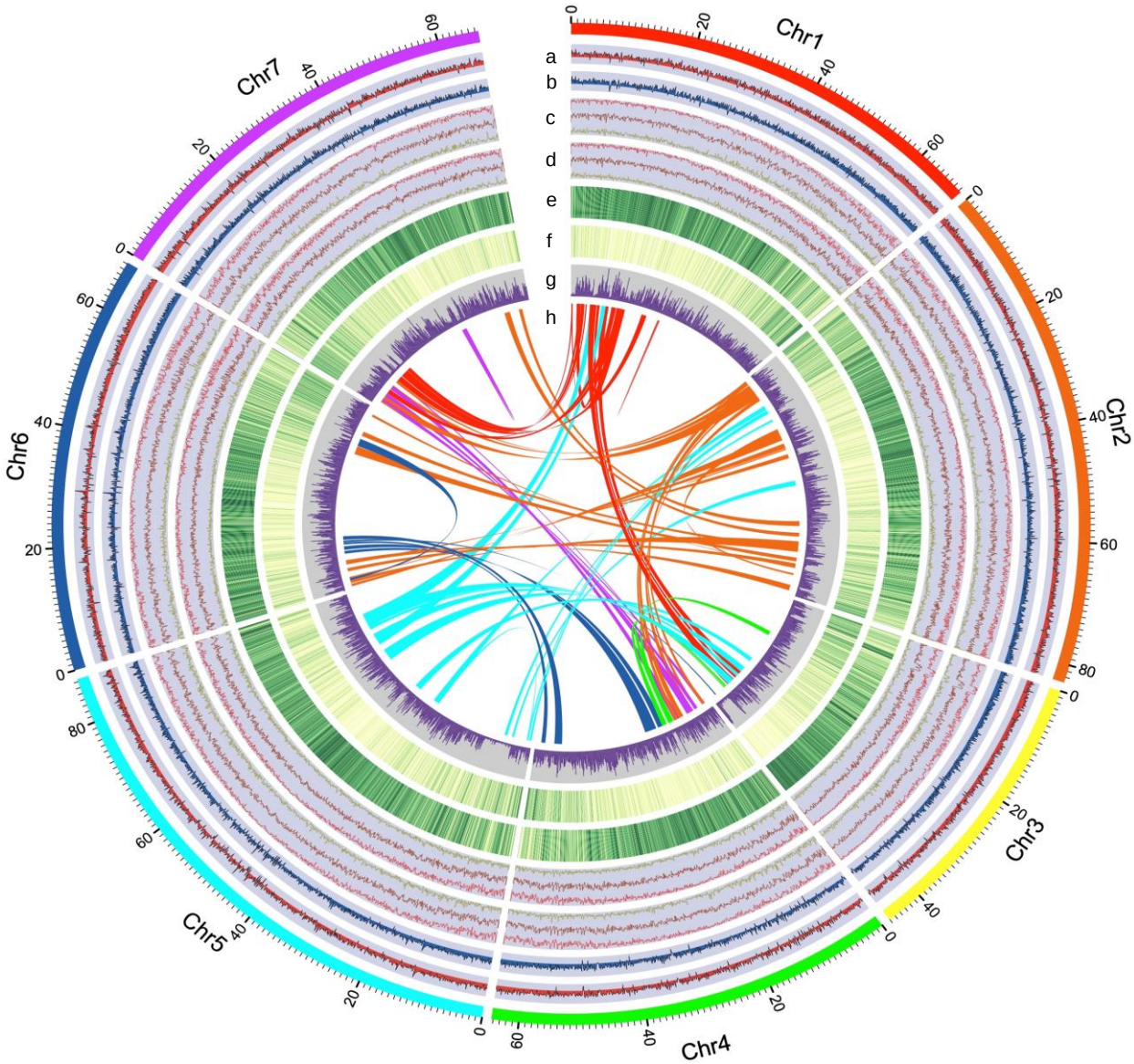


Figure 1. Genome features of *R. wichuraiana* 'Basye's Thornless' (BT).

A. The overall view of flowering BT plant in April of 2020. **a**, At this stage, branches featuring flower buds show a up-right growth pattern (pointed by white arrows) in contrast to its prostrate shoot (blue arrowhead). **b**, A zoom in picture of flowering bud at early stage (orange arrowhead). **c**, A flower bud prior to anthesis. **d**, One open flower.

B. Landscape of genome features for BT genome. **a**, gene expression profiles for leaf materials grown in March 2017. **b**, gene expression profiles for leaf materials grown in November 2016. **c** and **d**, DNA methylation patterns for samples in **a** (**c**) and **b** (**d**). Lines in red, dark red, and brown mark the CG, CHG, and CHH methylation, respectively. **e**, Transposable elements (TEs) content in each chromosome. **f**, gene density in each chromosome. **g**, SNP density along chromosomes in every 100kb bins. **h**, internal syntenic blocks within BT chromosomes.

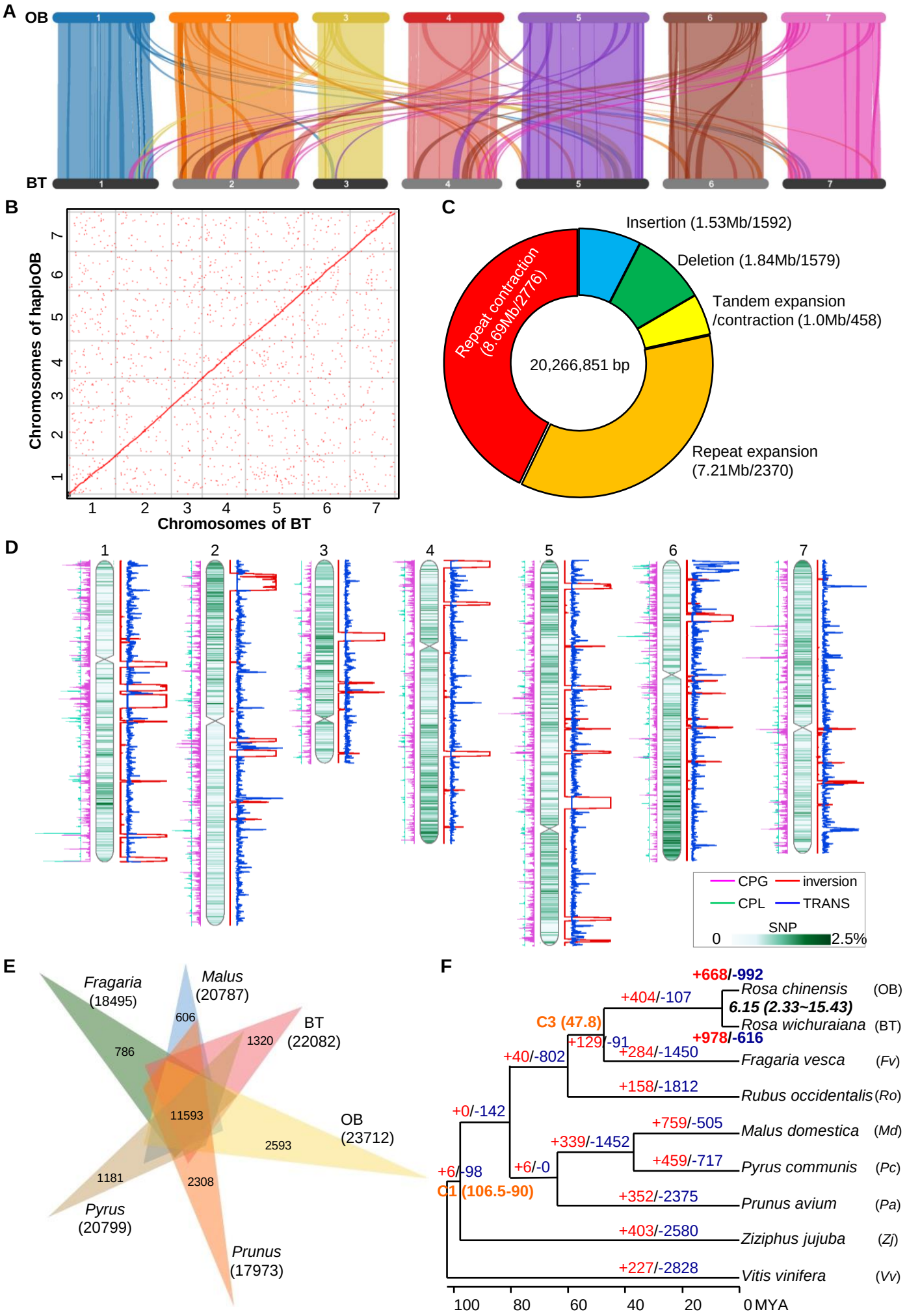

A
OB
BT
B
Chromosomes of haploOB
Chromosomes of BT
C
Insertion (1.53Mb/1592)
Deletion (1.84Mb/1579)
Tandem expansion /contraction (1.0Mb/458)
Repeat expansion (7.21Mb/2370)
Repeat contraction (8.69Mb/2776)
20,266,851 bp
D
CPG
inversion
CPL
TRANS
SNP
0
2.5%
E
Fragaria (18495)
Malus (20787)
BT (22082)
OB (23712)
Pyrus (20799)
Prunus (17973)
606
786
1320
11593
2593
1181
2308
F
+668/-992
Rosa chinensis (OB)
+404/-107
6.15 (2.33~15.41)
Rosa wichuraiana (BT)
C3 (47.8)
+978/-616
+129/-91
+284/-1450
Fragaria vesca (Fv)
+40/-802
+158/-1812
Rubus occidentalis (Ro)
+0/-142
+759/-505
Malus domestica (Md)
+339/-1452
+459/-717
Pyrus communis (Pc)
+6/-98
+6/-0
C1 (106.5-90)
+352/-2375
Prunus avium (Pa)
+403/-2580
Ziziphus jujuba (Zj)
+227/-2828
Vitis vinifera (Vv)
100
80
60
40
20
0 MYA

Figure 2. Comparative genomic analyses between BT and OB as well as other Rosaceae plants.
A. Gene collinearity (minimum ten syntenic gene pairs) between assembled the BT and Raymond's haploOB genome sequences. High level of collinearity between chromosome pairs was indicated by light color lines/blocks. Dark color lines marked a relatively lower collinearity between genome segments that might be a signature of segmental duplication.
B. Macrosynteny patterns between BT and haploOB genome sequences along each chromosome (Raymond et al. 2018).
C. A doughnut chart showing the large-size sequence variation between BT and haploOB. Numbers in brackets indicated the total length of each type of variation (before /) and the number of events (after /).
D. Genome alignment between BT and haploOB identified significant structural variants including genome inversions (red lines), translocation (TRANS; blue lines), copy gain (CPG; cyan lines; in OB with BT as reference) and copy loss (CPL; green lines). Data was plotted as per type of variation in 1 Mb of genomic region with step size of 100 kb by length density. SNP density was plotted on the chromosomes with 2.5 SNPs per 100 bp length at the highest level.
E. Comparison of the numbers of gene families identified via OrthoMCL. Venn diagram showing shared homologous gene families among genomes of BT, haploOB, *Fragraria*, *Malus*, *Pyrus* and *Prunus*. Numbers in brackets indicated total gene families. Only the numbers of genome-specific and shared among six genomes were shown. Numbers in brackets showed the total gene families for each species.
F. Phylogenetic relationships of BT with other Rosaceae plants based on 1230 shared single-copy orthologous genes taking *Ziziphus jujuba* and *Vitis vinifera* as outgroups. Estimated divergence time between BT and OB following the fossil calibrations of nodes C1 (106.5-90 MYA) and C3 (47.8 MYA) (in orange, according to Zhang et al. 2018) as calibration points was shown in bold italic numbers under branches. Numbers on branches indicate the counts of gene families expansion (+, in red) and contraction (-, in blue) along each lineage. Note the about 46.4% expansion and 38% contraction of gene families in BT comparing to haploOB. Abbreviated names in two letters were given in brackets behind each species.

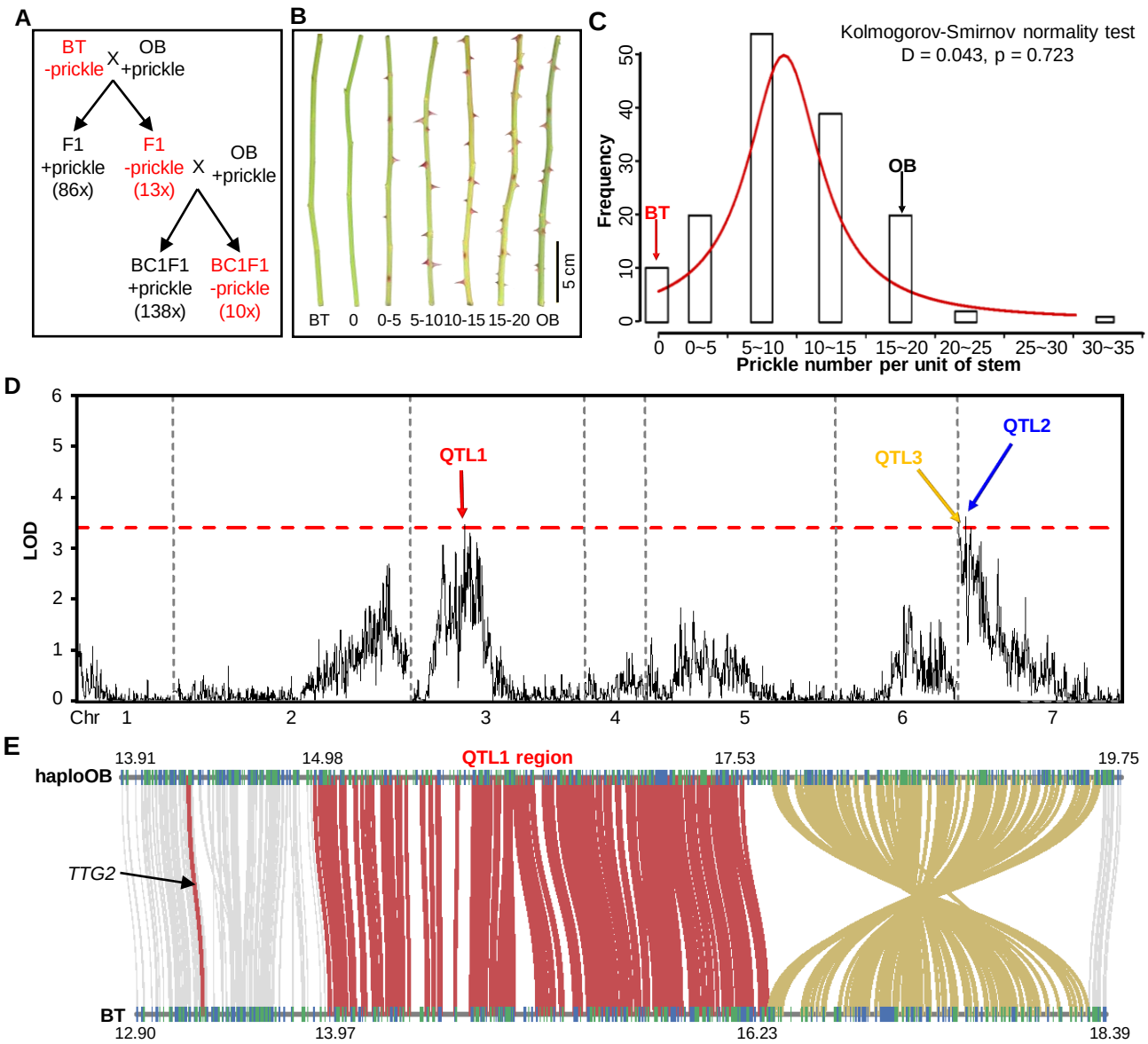

A
BT
-prickle
X
OB
+prickle
F1
+prickle
(86x)
F1
-prickle
(13x)
X
OB
+prickle
BC1F1
+prickle
(138x)
BC1F1
-prickle
(10x)
B
BT
0
0-5
5-10
10-15
15-20
OB
5 cm
C
Kolmogorov-Smirnov normality test
D = 0.043, p = 0.723
Frequency
BT
OB
0
0~5
5~10
10~15
15~20
20~25
25~30
30~35
Prickle number per unit of stem
D
LOD
QTL1
QTL2
QTL3
Chr
1
2
3
4
5
6
7
E
13.91
14.98
QTL1 region
17.53
19.75
haploOB
TTG2
BT
12.90
13.97
16.23
18.39

Figure 3. Genetic analysis of stem prickle distribution in OB x BT populations.

A. Genetic inheritance of prickle distribution on shoots of F1 and BC1F1 populations. Plants without prickle on shoots were indicated with a '-' in red color; a '+' labeled plants with prickles. Numbers in brackets marked the plant number featuring each category of trait.

B. Typical phenotype of prickle density on stems of the parents (BT and OB) and the BC1F1 population. Numbers indicated the range (0, >0~5, >5~10, >10~15, >15~20; the same for **C**) of prickle density per 20cm unit of stem.

C. Normal distribution of stem prickle density in BC1F1 population. X-axis indicated the prickle density per unit of length on stem. Y-axis marked the frequency of BC1F1 individuals showing certain range of prickle numbers. Arrows labelled the prickle features of the two parents (BT in red and OB in dark).

D. QTL analysis for prickle density on shoot in BC1F1 population. Y-axis showed the LOD profiles on each BT chromosome, with the red line indicating the significance threshold (=3.4) determined by permutation test. Red arrows marked the three QTL over threshold (QTL1 on Chr3, QTL2 and QTL3 on Chr7).

E. Syntenic gene pairs between OB and BT surrounding the QTL1 region on Chr3. Red lines indicated the syntenic genes in the QTL1 region. Note that QTL1 region is neighboring to a genome conversion (marked by brown lines) between OB and BT. Potential *TTG2* is labeled by a dark red line out side of the QTL1 peak. Numbers on/under chromosomes showed the physical positions in Mb.

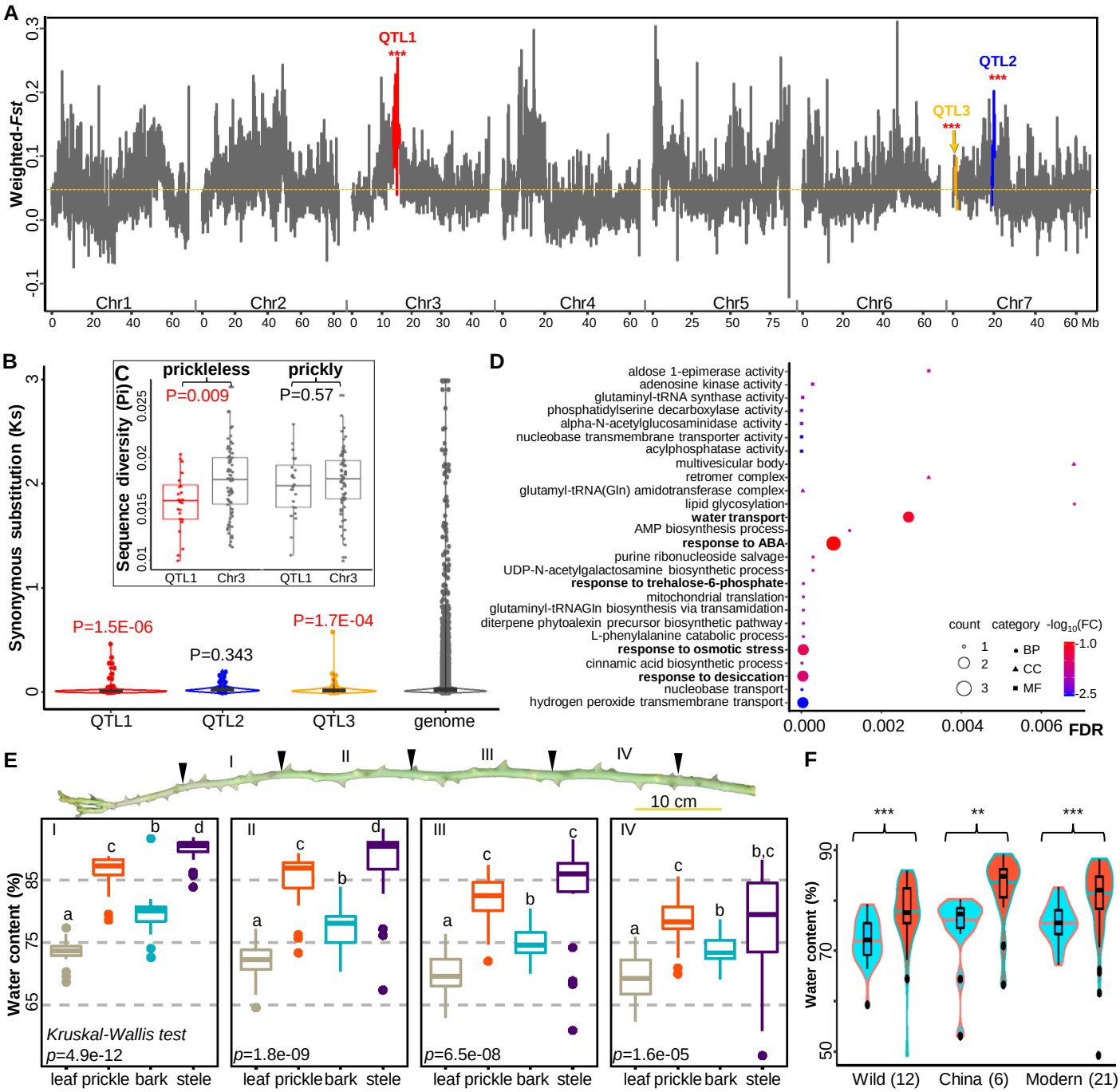

A
Weighted-Fst
QTL1
QTL2
QTL3
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
B
Synonymous substitution (Ks)
P=1.5E-06
P=0.343
P=1.7E-04
QTL1
QTL2
QTL3
genome
C
Sequence diversity (Pi)
prickleless
prickly
P=0.009
P=0.57
QTL1
Chr3
D
aldose 1-epimerase activity
adenosine kinase activity
glutaminyl-tRNA synthase activity
phosphatidylserine decarboxylase activity
alpha-N-acetylglucosaminidase activity
nucleobase transmembrane transporter activity
acylphosphatase activity
multivesicular body
retromer complex
glutamyl-tRNA(Gln) amidotransferase complex
lipid glycosylation
water transport
AMP biosynthesis process
response to ABA
purine ribonucleoside salvage
UDP-N-acetylgalactosamine biosynthetic process
response to trehalose-6-phosphate
mitochondrial translation
glutaminyl-tRNAGln biosynthesis via transamidation
diterpene phytoalexin precursor biosynthetic pathway
L-phenylalanine catabolic process
response to osmotic stress
cinnamic acid biosynthetic process
response to desiccation
nucleobase transport
hydrogen peroxide transmembrane transport
count
category
-log10(FC)
BP
CC
MF
FDR
E
10 cm
Water content (%)
Kruskal-Wallis test
p=4.9e-12
p=1.8e-09
p=6.5e-08
p=1.6e-05
leaf
prickle
bark
stele
F
Water content (%)
Wild (12)
China (6)
Modern (21)

Figure 4. Genetic inheritance pattern of stem prickle QTLs in rose populations.

A. Manhattan plot for the weighted-*Fst* per 100Kb window between three prickleless and seven prickly roses. Dots in red, blue and orange indicated the QTL1, 2 and 3 regions, respectively. *** showed the significant difference between each QTL and rest of the genome, as analyzed by a *Kruskal-Wallis multiple-comparison test*. The brown dotted line marked the mean weighted-*Fst* value (0.054) at the genome wide level.

B. Comparison of synonymous substitutions (*Ks*) between QTL regions and rest of the genome. Note that sequence variation in QTL1 and QTL3 regions was significantly lower than the rest of the genome *(Kruskal-Wallis multiple-comparison test*).

C. Sequence diversities (Pi) in QTL1 region (red) and other parts of chr3 (gray) between prickleless (left) and prickly (right) genotypes (*Wilcoxon rank sum test*).

D. GO enrichment pattern for differentially expressed genes located in three QTL between young shoots of OB and BT. X-axis showed the FDR values; FC, enrichment fold change against genome level; BP, biological process; CC, cellular components; MF, molecular function. GO names in bold marked the items related to water usage.

E. Relative water contents were significantly higher in prickles (red) than in bark (blue) or leaves (brown) in relatively young shoots (**E**) but not in relatively old shoots (**Supplementary Figure S36y**). Data for one modern rose genotype was shown (see **Supplementary Figure 35** for a second genotype). Relative water content was measured as (fresh weight - dry weight)/fresh weight in percentage. Note that water contents decreased following different node sections (I, II, III, and IV, represented every three nodes marked by dark arrow heads), though prickles always had higher contents of water in young shoots. In relatively old shoots, water contents in prickles were significantly lower than in other tissues. Letters (a, b, c, and d) above each boxplot indicated the pairwise significances ($p<0.01$; non-parametric *Wilcoxon.test*) for eight node sections (see **Supplementary Table xxx** for comparison values).

F. Prickles (red) generally contained relatively higher water contents than barks (blue; $p<0.001$, *Student's t-test*) in a collection of randomly selected roses (12 wild rose, 6 China rose, and 21 modern cultivated genotypes; see also **Supplementary Figure S36x**). *** ($P<0.001$; non-parametric *Wilcoxon.test*) and ** ($P<0.01$) labelled the significant difference between prickle and bark for each group. For **E** and **F**, prickle number did not co-vary with the relative water content for either prickle, bark or stele (**Supplementary Figure S36**).

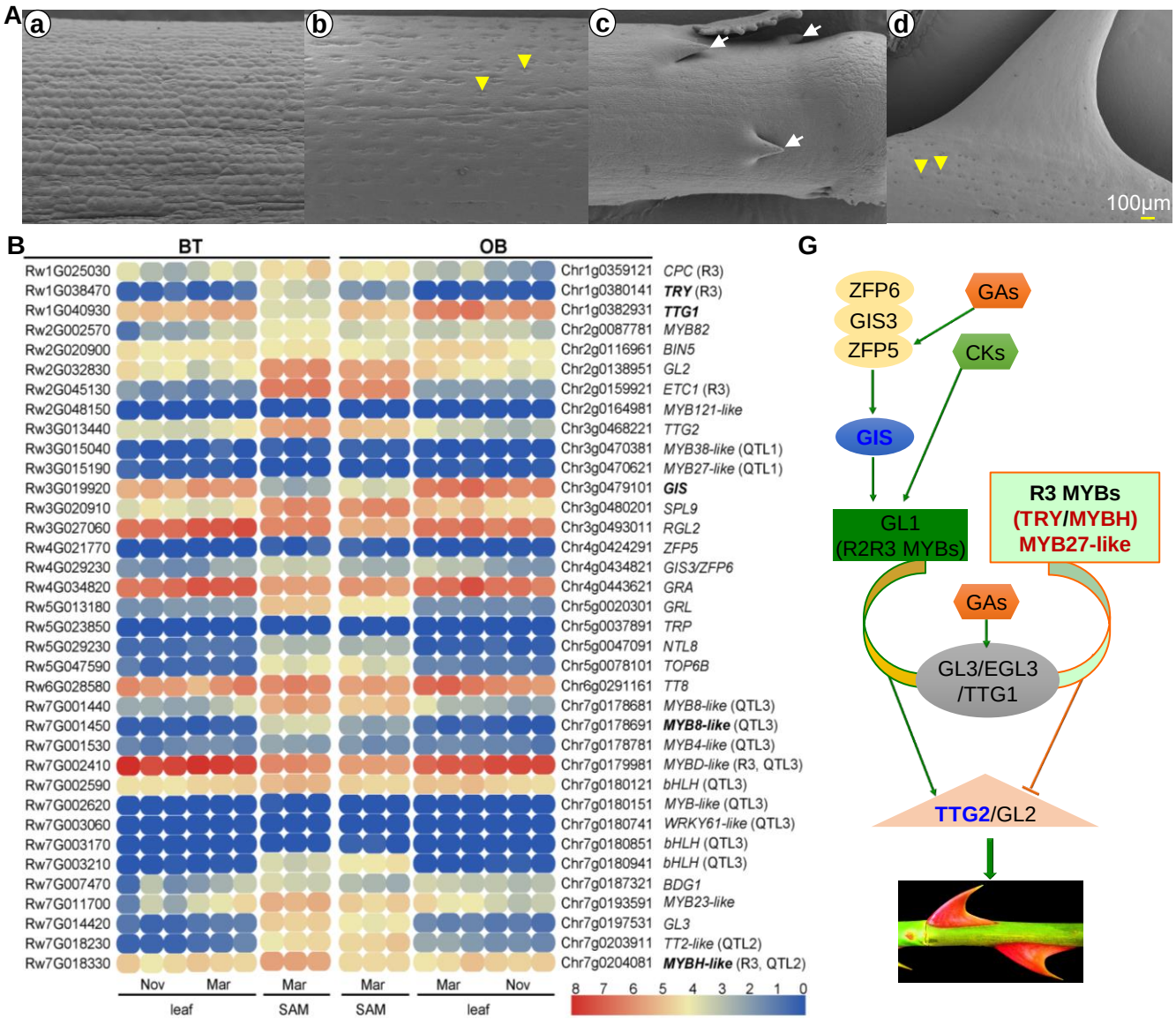

A
a
b
c
d
100μm
B
BT
OB
Rw1G025030
Rw1G038470
Rw1G040930
Rw2G002570
Rw2G020900
Rw2G032830
Rw2G045130
Rw2G048150
Rw3G013440
Rw3G015040
Rw3G015190
Rw3G019920
Rw3G020910
Rw3G027060
Rw4G021770
Rw4G029230
Rw4G034820
Rw5G013180
Rw5G023850
Rw5G029230
Rw5G047590
Rw6G028580
Rw7G001440
Rw7G001450
Rw7G001530
Rw7G002410
Rw7G002590
Rw7G002620
Rw7G003060
Rw7G003170
Rw7G003210
Rw7G007470
Rw7G011700
Rw7G014420
Rw7G018230
Rw7G018330
Chr1g0359121 CPC (R3)
Chr1g0380141 TRY (R3)
Chr1g0382931 TTG1
Chr2g0087781 MYB82
Chr2g0116961 BIN5
Chr2g0138951 GL2
Chr2g0159921 ETC1 (R3)
Chr2g0164981 MYB121-like
Chr3g0468221 TTG2
Chr3g0470381 MYB38-like (QTL1)
Chr3g0470621 MYB27-like (QTL1)
Chr3g0479101 GIS
Chr3g0480201 SPL9
Chr3g0493011 RGL2
Chr4g0424291 ZFP5
Chr4g0434821 GIS3/ZFP6
Chr4g0443621 GRA
Chr5g0020301 GRL
Chr5g0037891 TRP
Chr5g0047091 NTL8
Chr5g0078101 TOP6B
Chr6g0291161 TT8
Chr7g0178681 MYB8-like (QTL3)
Chr7g0178691 MYB8-like (QTL3)
Chr7g0178781 MYB4-like (QTL3)
Chr7g0179981 MYBD-like (R3, QTL3)
Chr7g0180121 bHLH (QTL3)
Chr7g0180151 MYB-like (QTL3)
Chr7g0180741 WRKY61-like (QTL3)
Chr7g0180851 bHLH (QTL3)
Chr7g0180941 bHLH (QTL3)
Chr7g0187321 BDG1
Chr7g0193591 MYB23-like
Chr7g0197531 GL3
Chr7g0203911 TT2-like (QTL2)
Chr7g0204081 MYBH-like (R3, QTL2)
Nov
Mar
Mar
Mar
Mar
Nov
leaf
SAM
SAM
leaf
8 7 6 5 4 3 2 1 0
G
ZFP6
GIS3
ZFP5
GAs
CKs
GIS
GL1
(R2R3 MYBs)
R3 MYBs
(TRY/MYBH)
MYB27-like
GAs
GL3/EGL3
/TTG1
TTG2/GL2
C
MYB27-like (QTL1)
Rw3G015190/Chr3G0470621
0.002
0.001
0.001
0.001
0.001
0.001
0.000
0.000
0.000
F1+
F1-
BC1+
BC1-
D
TT2-like (QTL2)
Rw7G018230/Chr7G0203911
0.120
0.100
0.080
0.060
0.040
0.020
0.000
F1+
F1-
BC1+
BC1-
E
GIS-like
Rw3G019920/Chr3G0479101
0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
F1+
F1-
BC1+
BC1-
F
TTG2-like
Rw3G013440/Chr3G0468221
0.160
0.140
0.120
0.100
0.080
0.060
0.040
0.020
0.000
F1+
F1-
BC1+
BC1-

Figure 5. Identification of candidate genes involved in stem prickle patterning in roses.

A. SEM analysis of early (**a** and **c**) and later (**b** and **d**) stage stems for BT (**a** and **b**) and OB (**c** and **d**). Note that, in contrast to OB, BT lost the prickle initiation capability (white arrows) while the ability of stomata initiation (yellow arrow heads) was similar to OB.

B. Expression heatmap of genes involved in prickle-initiation GRN. RNA-seq data for young shoots in March (SAM Mar) and just opened leaves in March (leaf Mar) and November (leaf Nov) were used for the quantification (in FPKM values). Names in bold marked the genes showing minimum two-fold changes at $p<0.05$ (*Student's two tailed t-test*). R3 type MYBs were indicated with "R3" in brackets, while the genes located in QTLs were also marked.

C-**F**. Expression comparisons for candidate genes (*MYB27-like*, **C**; *TT2-like*, **D**; *GIS-like*, **E**; *TTG2-like*, **F**) in F1 and BC1F1 pools with (F1+ or BC1+) or without (F1- or BC1-) prickles on stem. The mean expression level of minimum three pools (each with at least two individual lines) was shown with standard deviation. A * indicated $p<0.05$ (*Student's two tailed t-test*).

G. A simplified GRN regulating rose prickle initiation and development based on information from Arabidopsis. Comparing to the prickly OB, higher expression of R3 type (*TRY1* and *MYBH-like*) and R2R3 repressor (*MYB27-like*), and lower expression of trichrome initiation promoters *GIS* and *TTG2* might account for the prickleless pattern on BT stems.