**Supplementary Tables**

**Table S1 BUSCO assessment of the *P. simonii* genome assembly**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Types of assessment | Complete BUSCOs  （C） | Complete and single-Copy BUSCOs（S） | Complete and Duplicated BUSCOs（D） | Fragmented BUSCOs  （F） | Missing BUSCOs  （M） | Total BUSCO groups searched（N） |
| Genome | 1,347 | 1,036 | 311 | 7 | 21 | 1,375 |
| Gene | 1,305 | 1,033 | 272 | 50 | 20 |

**Table S2 ncRNA genes annotation statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type** | **Copy Number** | **Average Length (bp)** | **Total Length (bp)** | **Percentage (%) of Genome** |
| rRNA | 290 | 1439.87 | 417,561 | 0.000946 |
| 18S | 83 | 1,860.75 | 154,442 | 0.000349 |
| 28S | 58 | 4,205.48 | 243,918 | 0.000552 |
| 5.8S | 42 | 154.00 | 6,468 | 0.000015 |
| 5S | 107 | 119.00 | 12,733 | 0.000029 |
| snRNA | 618 | 115.94 | 71,649 | 0.000162 |
| CD-box | 427 | 105.38 | 44,996 | 0.000102 |
| HACA-box | 88 | 129.01 | 11,353 | 0.000026 |
| splicing | 103 | 148.54 | 15,300 | 0.000035 |
| miRNA | 1,153 | 122.46 | 141,197 | 0.000320 |
| tRNA | 1,177 | 70.94 | 83,501 | 0.000189 |

**Table S3 Statistics of the genome prediction**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gene number | Gene Length (bp) | Average gene length (bp) | CDS length (bp) | Average CDS length (bp) | Exon length (bp) | Average exon length (bp) | Average intron length (bp) | Average intergenic space distance (bp) |
| 45,459 | 144,306,239 | 3,174 | 49,806,522 | 1,096 | 58,745,857 | 257 | 469 | 6,715 |

**Table S4** **Summary of functional annotation for the predicted genes**

|  |  |  |
| --- | --- | --- |
| Type of database | Number of genes | Percent(%) |
| NR | 39,816 | 87.6 |
| Swiss-Prot | 30,104 | 66.2 |
| KEGG | 11,866 | 26.1 |
| KOG | 10,973 | 24.1 |
| Pfam | 28,772 | 63.3 |
| GO | 17,860 | 39.3 |
| Annotated | 39,833 | 87.6 |
| Total | 45,459 | - |

**Table S5 Summary of gene family clustering.** The genes from nine sequenced plant genomes were collected and aligned to each other using the software DIAMOND.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Species | Total  genes | Genes in families | Family | Unclustered genes**#*1*** | Unique families**#*2*** | Single copy gene families**#*3*** | Genes in unique  families |
| *P. trichocarpa* | 42,950 | 40,278 | 23,234 | 2,672 | 9 | 966 | 82 |
| *P. deltoides* | 44,853 | 41,819 | 22,903 | 3,034 | 4 | 966 | 9 |
| *P. euphratica* | 49,760 | 48,040 | 16,641 | 1,720 | 8 | 966 | 24 |
| *P. simonii* | 45,459 | 37,802 | 21,984 | 7,657 | 10 | 966 | 31 |
| *S. suchowensis* | 26,599 | 24,925 | 15,766 | 1,674 | 4 | 966 | 8 |
| *S. purpurea* | 35,125 | 31,937 | 17,006 | 3,188 | 9 | 966 | 35 |
| *R. communis* | 31,221 | 20,514 | 15,783 | 10,707 | 32 | 966 | 159 |
| *M. esculenta* | 32,718 | 27,723 | 15,993 | 4,995 | 25 | 966 | 124 |
| *O. sativa* | 42,101 | 20,302 | 12,032 | 21,799 | 159 | 966 | 1,174 |

***1*** **Unclustered genes:** A gene that has not been put into an orthogroup with any other genes.

***2*** **Unique families:** An orthogroups that consist entirely of genes from one species.

***3*** **Single copy gene families:** A gene family with exactly one gene (and no more) from each species. These gene families are ideal for inferring a species tree.

**Table S6 The total Orthogroups from nine plant species**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Number of genes | Genes in orthogroups | Number of unassigned genes**#*1*** | Number of orthogroups | Number of species-specific orthogroups**#*2*** | G50**#*3*** O50**#*4*** |
| 350,786 | 293,340 | 57,446 | 26,181 | 260 | 14 7,917 |

***1*Unassigned gene:** A gene that has not been put into an orthogroup with any other genes.

***2*Species-specific orthogroup:** An orthogroups that consist entirely of genes from one species.

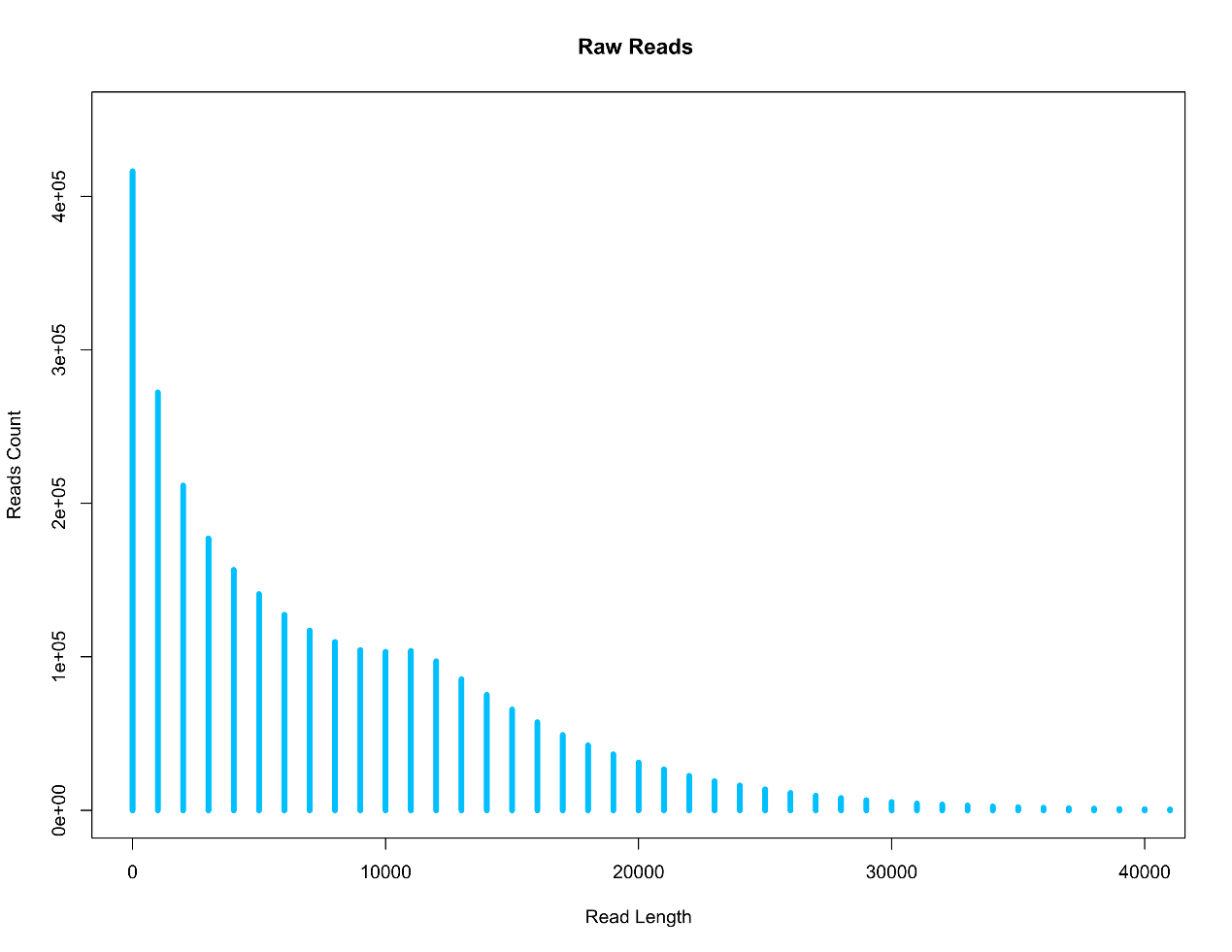
***3*G50:** The number of genes in the orthogroup such that 50% of genes are in orthogroups of that size or larger.

***4*O50:** The smallest number of orthogroups such that 50% of genes are in orthogroups of that size or larger.

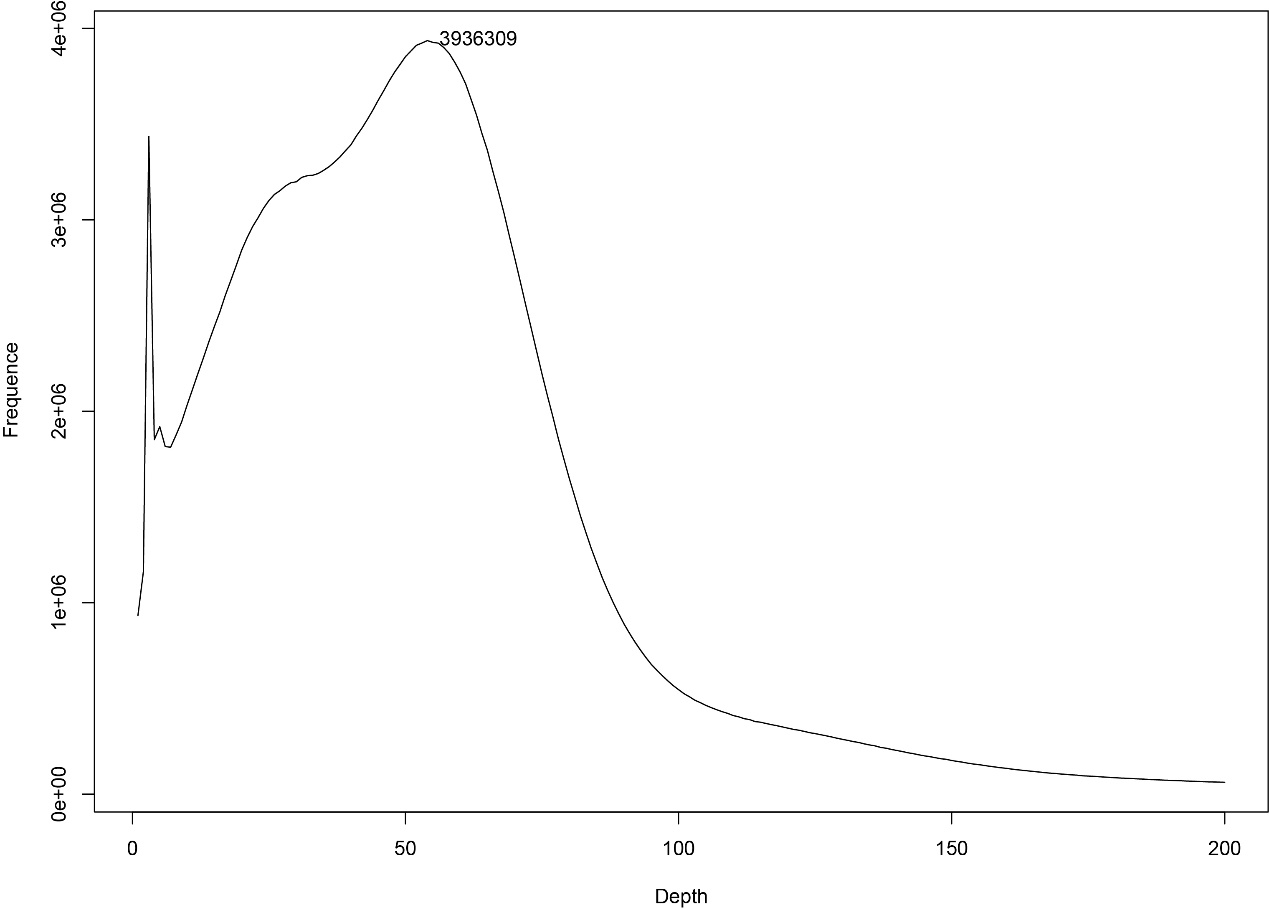
**Table S7 Analysis of gene family expansion and contraction in *P. simonii* and eight other plants species**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Species | Expanded families | Genes/expansion | Contracted families | Genes/contraction | Genes gained | Genes lost |
| *P. deltoids* | 2,790 | 1.54 | 3,132 | 1.09 | 4,305 | 3,401 |
| *P. trichocarpa* | 1,527 | 1.66 | 3,098 | 1.05 | 2,529 | 3,239 |
| *S. suchowensis* | 324 | 1.22 | 3,853 | 1.20 | 395 | 4,608 |
| *P. euphratica* | 7,333 | 2.42 | 9,673 | 1.11 | 17,713 | 10,743 |
| *R. communis* | 915 | 1.53 | 2,703 | 1.10 | 1,401 | 2,986 |
| *O. sativa* | 519 | 2.08 | 14,990 | 1.01 | 1,079 | 15,124 |
| *M. esculenta* | 5,278 | 1.28 | 1,019 | 1.01 | 6,744 | 1,085 |
| *P. simonii* | 2,356 | 1.45 | 5,224 | 1.26 | 3,427 | 6,562 |
| *S. purpurea* | 2,057 | 1.78 | 830 | 1.06 | 3,653 | 881 |

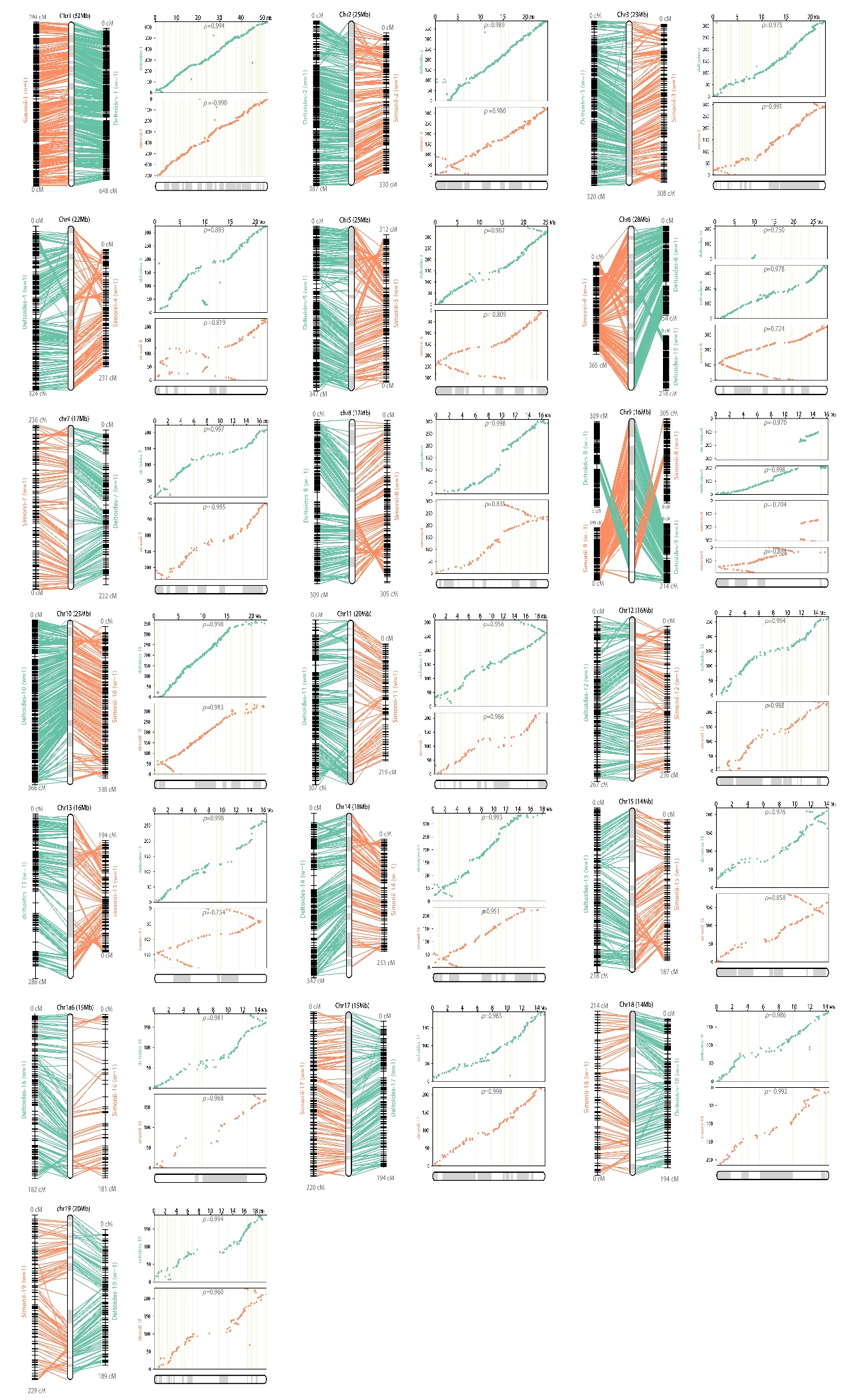
**Supplementary figures**



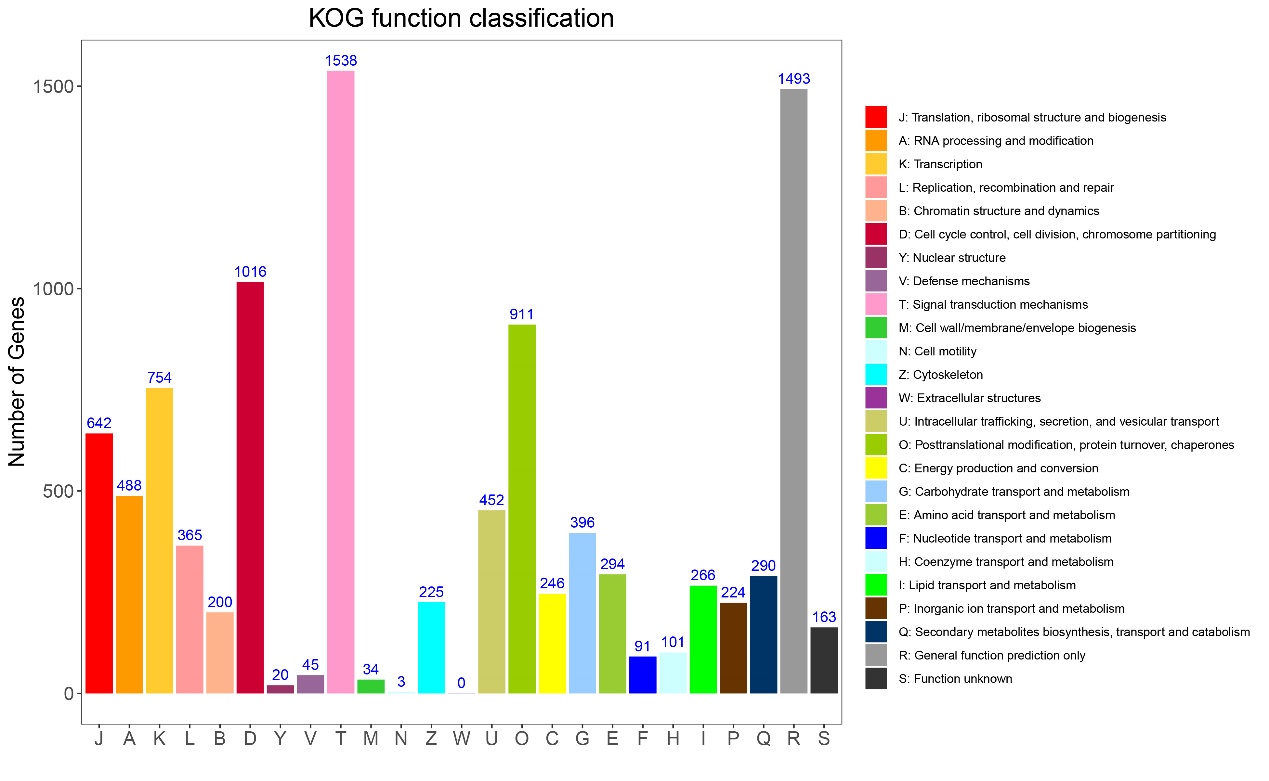
**Figure S1: The length distribution of subreads form PacBio Sequel platform that were used for assembling the *Populus simonii* genome.**



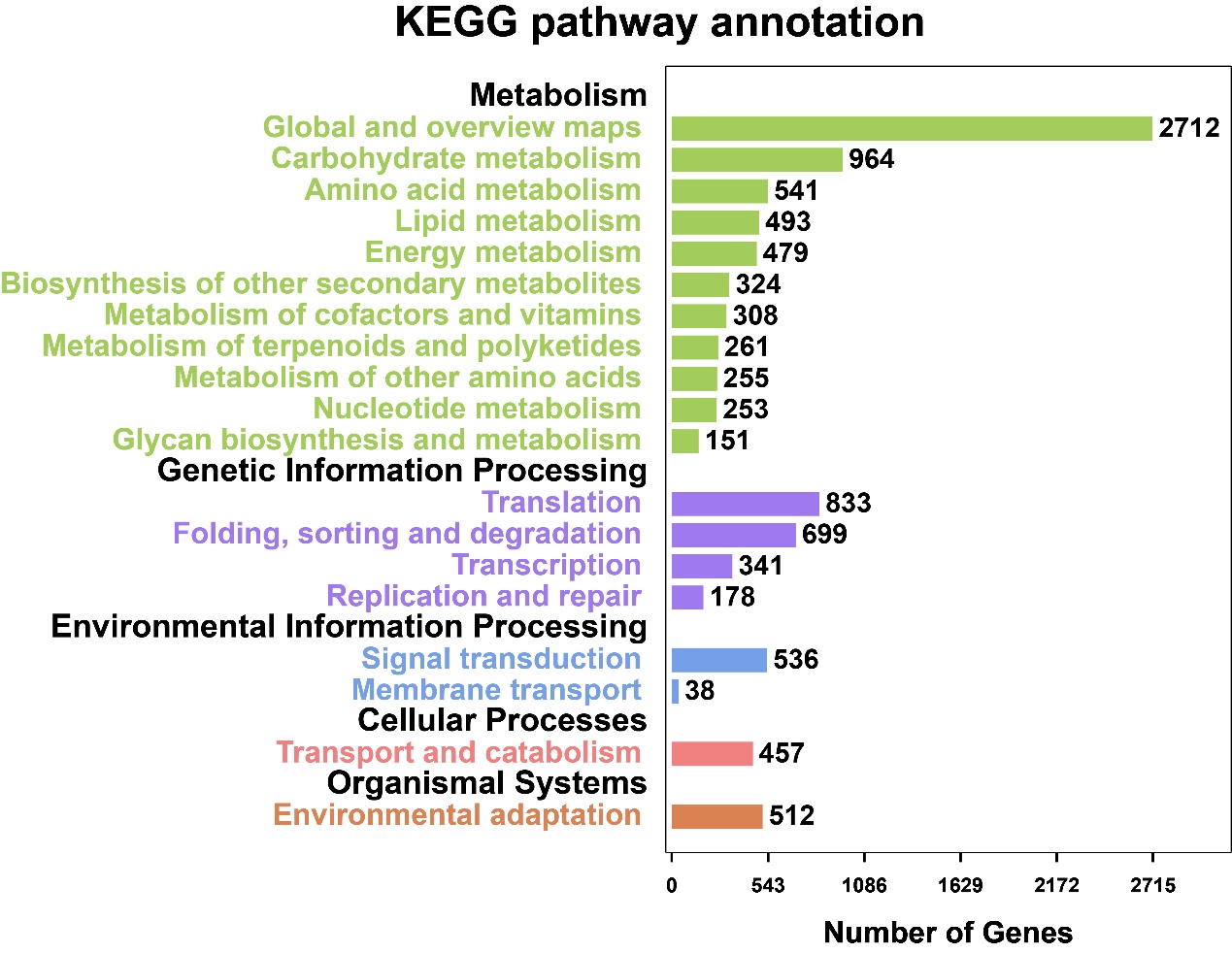
**Figure S2: Kmer frequency distributions from base error corrected reads.** There is a frequency peak value at 53 which is used for genome size estimation. The X-axis is the Kmer depth, and Y-axis represents the frequency of the Kmer for a given depth.



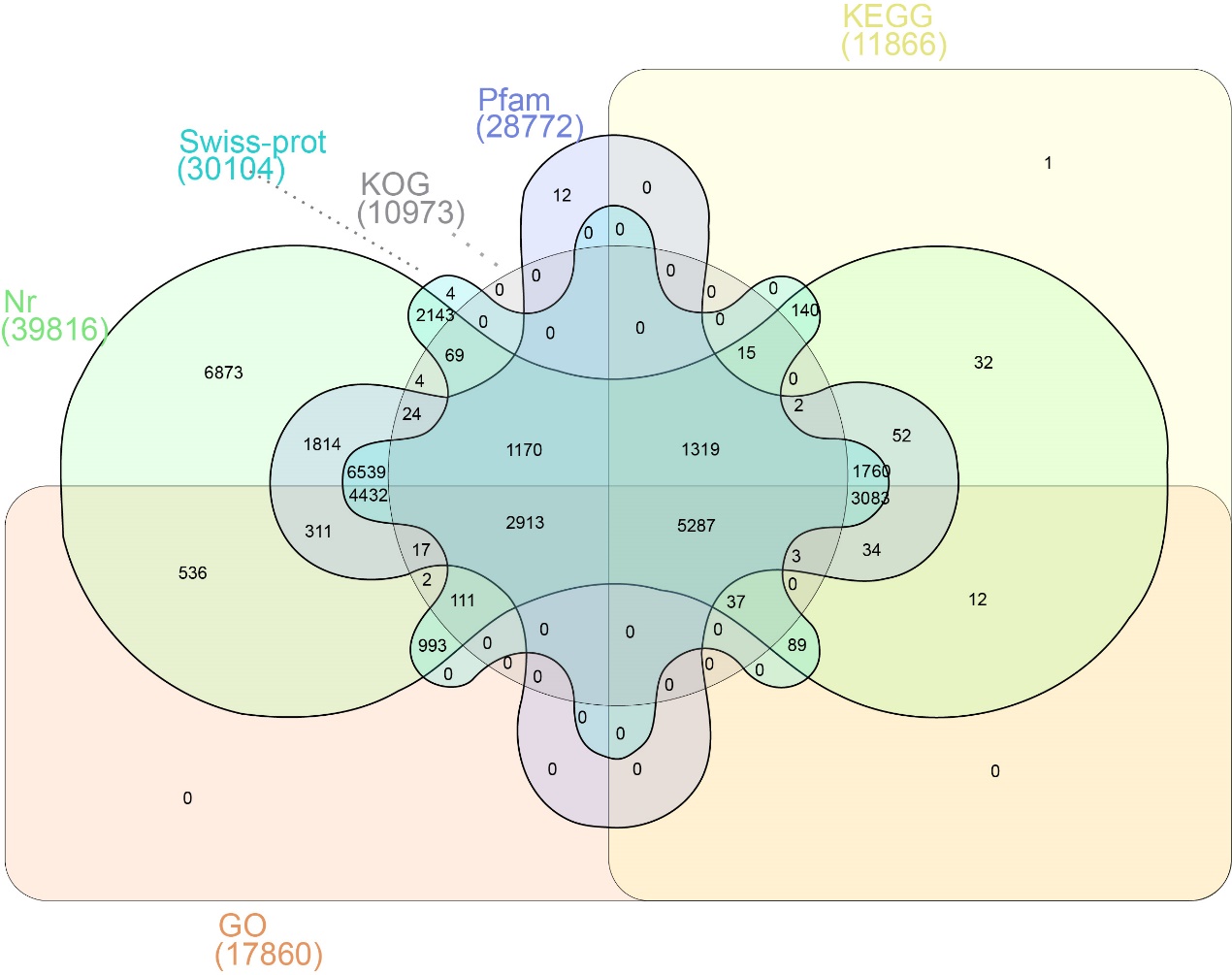
**Figure S3 Collinear analysis between the consensus map and the two individual genetic maps.** The left figure is a collinear comparision of the consensus map with the two individual genetic maps, and the top of the figure is the chromosome number and size. The right figure shows the consistency of the marker position order between the two individual maps and the consensus map. The X axis is the marker position on the pseudo-chromosome, and the Y axis is the marker position on the individual maps. The numbers from the right figure represent the Spearman rank correlation coefficients between the marker order of individual maps and pseudo-chromosome.



**Figure S4 The eukaryotic orthologous groups of proteins (KOG) function classification of protein-coding genes in *Populus simonii*.** The X axis indicates the function classification of the protein-coding genes, and the Y axis indicates the number of genes.



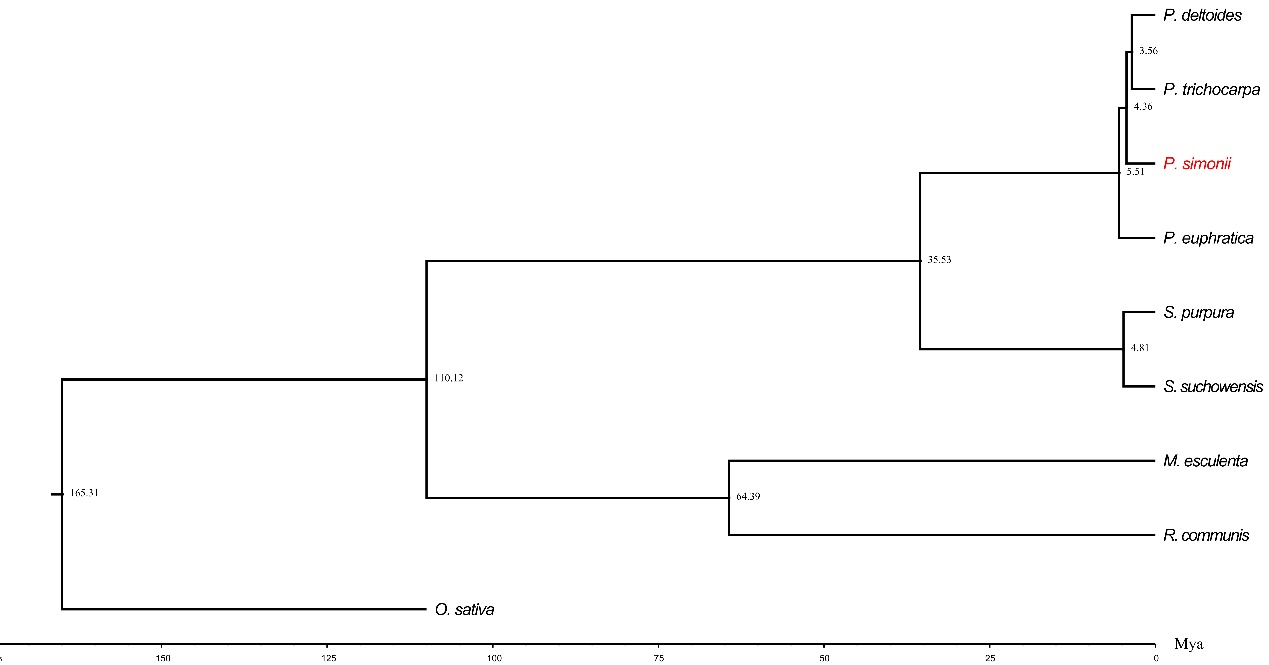
**Figure S5 The Kyoto Encyclopedia of Genes and Genomes (KEGG) function annotation of *Populus simonii*.** The X axis represents the number of genes and the Y axis shows the category of KEGG pathway annotations.



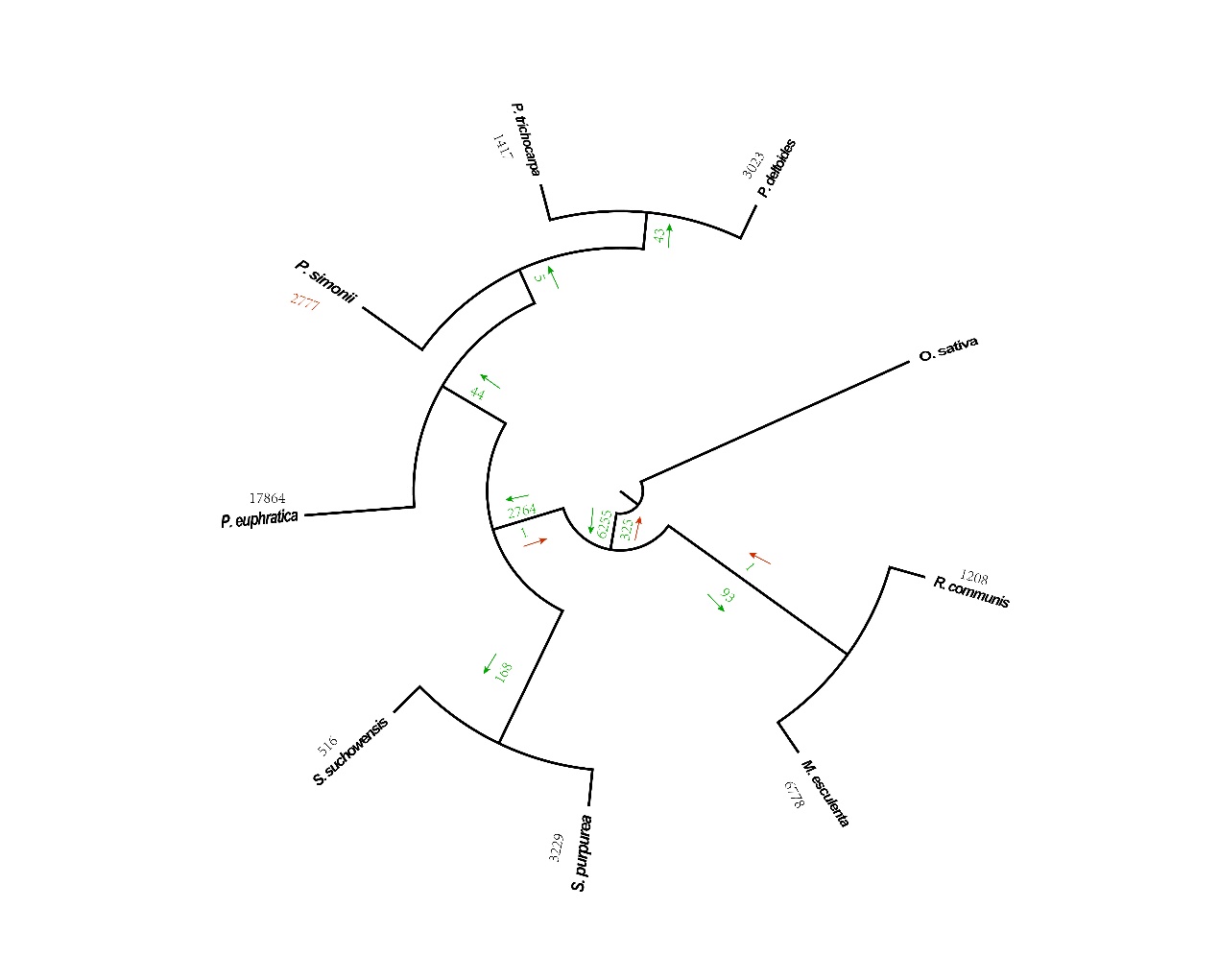
**Figure S6 Venn diagram of the number of protein-coding genes with functional annotation using six databases.** The number of genes annotated in all six databases is 5287.



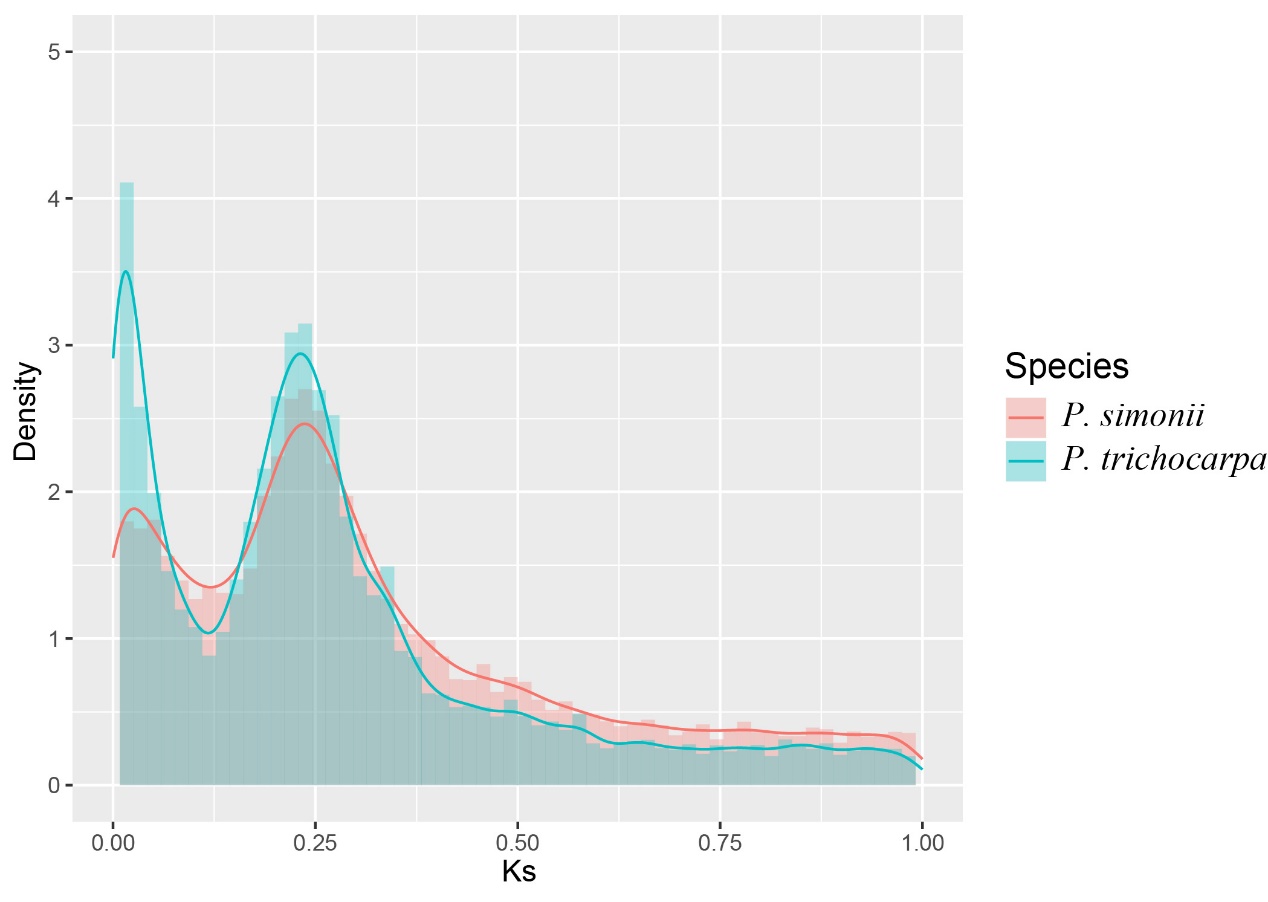
**Figure S7 Comparison of the genomic structure of *P. simonii* with that of *P. trichocarpa*.** The 19 chromosome-anchored sequences of *P. simonii* were ordered based on the alignment with the reference genome sequence of *P. trichocarpa* using MUMmer.



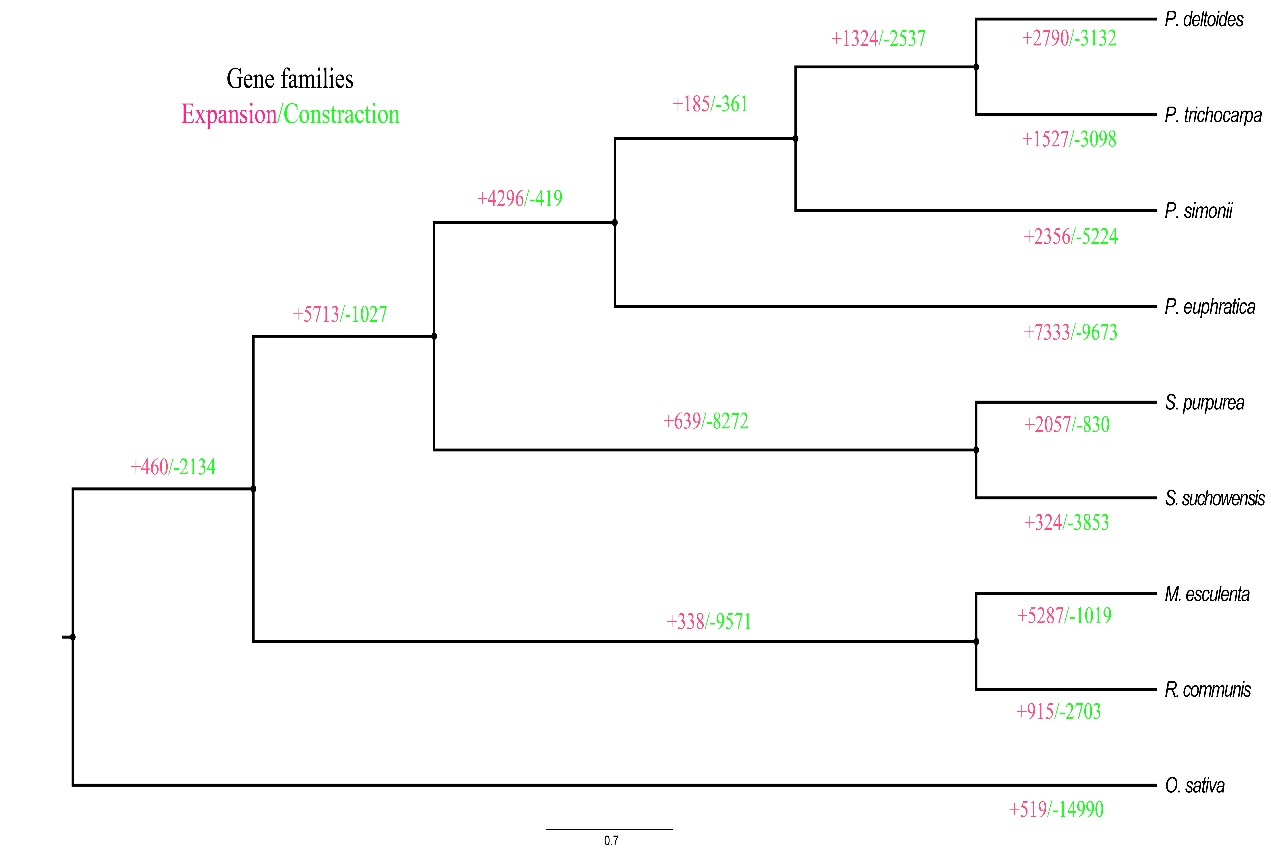
**Figure S8 Estimation of divergence time between *P. simonii*, *P. trichocarpa* and *P. euphratica* through phylogenetic analysis.** The numbers on the nodes represent the estimated divergence time from present(million years ago, Mya). The calibration times for divergence between *O. sativa* and *R. communis* (115-308Mya Mya), *Ricinus communis and S. purpura* (70-86Mya), *S. purpura* and *P. euphratica* (12-48 Mya), wereobtained from the TimeTree database (http://www.timetree.org/).



**Figure S9 Numbers of identified gene duplication events.** The numbers of identified gene duplication events are marked on the branched of the species tree and arrows indicate the direction in which the duplication occurred. The number near the species name represents the number of gene duplication events in this species.



**Figure S10 WGD events of two *Populus* genomes (*P. simonii* and *P. trichocarpa*) inferred by Ks distribution.**



**Figure S11 Dynamic evolution of orthologous gene families.** Numbers of expanded gene families are marked in red, while numbers of contracted gene families are marked in green.