

Haplotype-resolved genome assembly and resequencing provide insights into the origin and breeding of modern rose

In the format provided by the
authors and unedited

Supplementary Notes

1. Naming of homologous chromosomes of ‘Samantha®’

Modern roses are segmental allopolyploids, with predominantly random pairing between homologous chromosomes during meiosis. Therefore, ‘subgenome’ and ‘haplome’ are not applicable in modern roses. To facilitate the selection of representative chromosome sets for future research, instead of randomly naming the homologous chromosomes of ‘Samantha®’, we named these chromosomes according to their potential relationship with Chinese wild species and old cultivars.

We first selected 17 accessions that are wild species distributed in China or Chinese old roses and possess only the genetic background of Chinese genotypes (Supplementary Table 8). Genome resequencing data of these 17 accessions were then aligned to each chromosome of ‘Samantha®’. Based on the alignments, the proportion of the four-fold synonymous third-codon transversion (4DTv) sites that are identical between each chromosome of ‘Samantha®’ and the Chinese wild/old roses was obtained. A higher proportion indicates a putative greater similarity between the ‘Samantha®’ chromosome and the Chinese wild/old roses.

Chromosomes of ‘Samantha®’ with the highest proportions among their homologous chromosomes were named ChrA, and the ones with the lowest proportions were named ChrB. Chromosomes with the second highest proportions were names ChrC, while those with the third highest proportion were named ChrD. It is important to point out that this naming does not imply that the chromosomes from the same set (e.g., Chr1A to Chr7A) are from the same haplotype.

Proportion of 4DTv sites identical between ‘Samantha®’ and Chinese wild/old roses in each of the 28 ‘Samantha®’ chromosome.

Chromosome	A	B	C	D
Chr1	50.50%	43.02%	49.35%	46.68%
Chr2	55.51%	54.16%	55.07%	54.29%
Chr3	58.76%	53.97%	56.12%	56.05%
Chr4	55.07%	48.02%	53.92%	53.20%
Chr5	54.37%	49.88%	52.48%	50.76%
Chr6	56.64%	48.77%	53.21%	49.16%
Chr7	56.08%	42.66%	55.06%	54.73%

2. Sections within subgenus *Rosa*

According to previous studies¹⁻³, there are ten sections within the subgenus *Rosa*: Chinenses (Indicae), Synstylae, Cinnamomeae, Pimpinellifoliae, Microphyllae, Banksianae, Bracteatae, Laevigatae, Rosa, Caninae.

Section	Description
Chinenses	Include three species, <i>R. chinensis</i> , <i>R. odorata</i> , and <i>R. lucidissima</i> . Include famous old cultivars <i>R. chinensis</i> ‘Old Blush’, <i>R. chinensis</i> ‘Slater’s Crimson

	China', <i>R. odorata</i> 'Hume's Blush Tea-scented China', and <i>R. odorata</i> 'Park's Yellow Tea-scented China' that were introduced into Europe and greatly flourished rose breeding in Europe.
Synstylae	Mainly distributed in Asia. Include 18 species distributed in China, among which <i>R. multiflora</i> and <i>R. wichuraiana</i> are the most commonly used in rose breeding. The most obvious traits are their inflorescence and the combination of style, giving them the name of the section.
Cinnamomeae	Widely distributed in Asia, Europe and America. Include 31 species distributed in China, among which <i>R. fedtschenkoana</i> and <i>R. rugosa</i> were recorded to be involved in rose breeding.
Pimpinellifoliae	Mainly distributed in Asia and Europe. Include 18 species distributed in China. Include four-petal (e.g., <i>R. sericea</i>) or five-petal (e.g., <i>R. foetida</i>) species.
Microphyllae	Include three species, <i>R. roxburghii</i> , <i>R. kweichowensis</i> , and <i>R. praelucens</i> , distributed in China. <i>R. praelucens</i> , endemic to Zhongdian Plateau, Yunnan Province, is the naturally occurring decaploid grown on the highest altitude among all roses.
Banksianae	Include two species, <i>R. banksiae</i> and <i>R. cymosa</i> , distributed in China.
Bracteatae	Include two species in Asia, of which <i>R. bracteata</i> is distributed in China.
Laevigatae	Include only one species, <i>R. laevigata</i> , distributed in China.
Rosa	Originally distributed in Europe and West Asia and then introduced into China. Species including <i>R. gallica</i> , <i>R. damascena</i> , <i>R. centifolia</i> , and <i>R. alba</i> are best known for their strong growth vigor, winter-hardy, and high bushes.
Caninae	Include about 50 species in Europe. Characterized by the Caninae type meiosis with unbalanced heterogamous fully sexual reproduction.

3. Group classification used in this study

Chi: wild species from section Chinenses and their derived cultivars.

Syn: wild species from section Synstylae and their derived cultivars.

Cin: wild species from section Cinnamomeae and their derived cultivars.

Rosa: wild species from section Rosa and their derived cultivars.

Can: wild species from section Caninae and their derived cultivars.

Other: wild species from section Pimpinellifoliae, Banksianae, Microphyllae, and Laevigatae, as well as their derived cultivars.

Int: intermediate old cultivars, cultivars produced from hybridization across different sections. Most of them are cultivars bred from Chinese and European germplasm before 1867.

Hyb: cultivars bred after 1867, also known as modern cultivars.

4. Classification of modern rose cultivars

Modern rose cultivars can be classified into following groups:

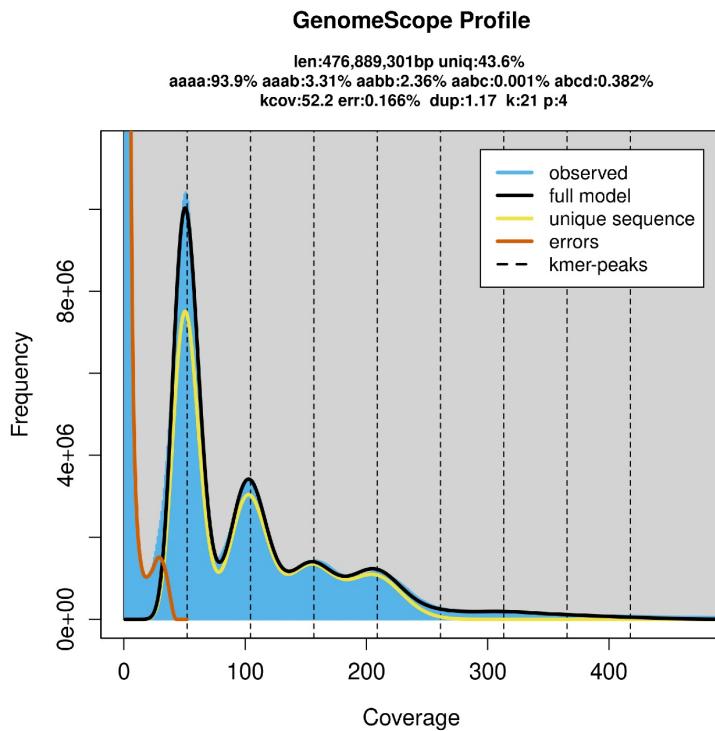
Group	Description
Hybrid Tea	Accepted as the most popular group. Best known for its shapely bloom (high-centered with central cone formed and petal edge could get reflexed at later

	opening stages) and single long stem. Often used as cut roses. Large-flower Hybrid Tea can also be called Grandiflora.
Floribunda	Bears its flowers in cluster/truss. Unrivalled for its large-quantity, long-lasting flowering, though flower shape is inferior to that of Hybrid Tea.
Climber & Rambler	Named after their growth habit of long-arching climbing canes. With a wide range of flower shapes and colors.
Miniature	Average plant height is around 35-75 cm. Flower form and foliage are indeed miniature versions of both Hybrid Tea and Floribundas.
Shrub	Bushy roses that do not fit into the above classes.

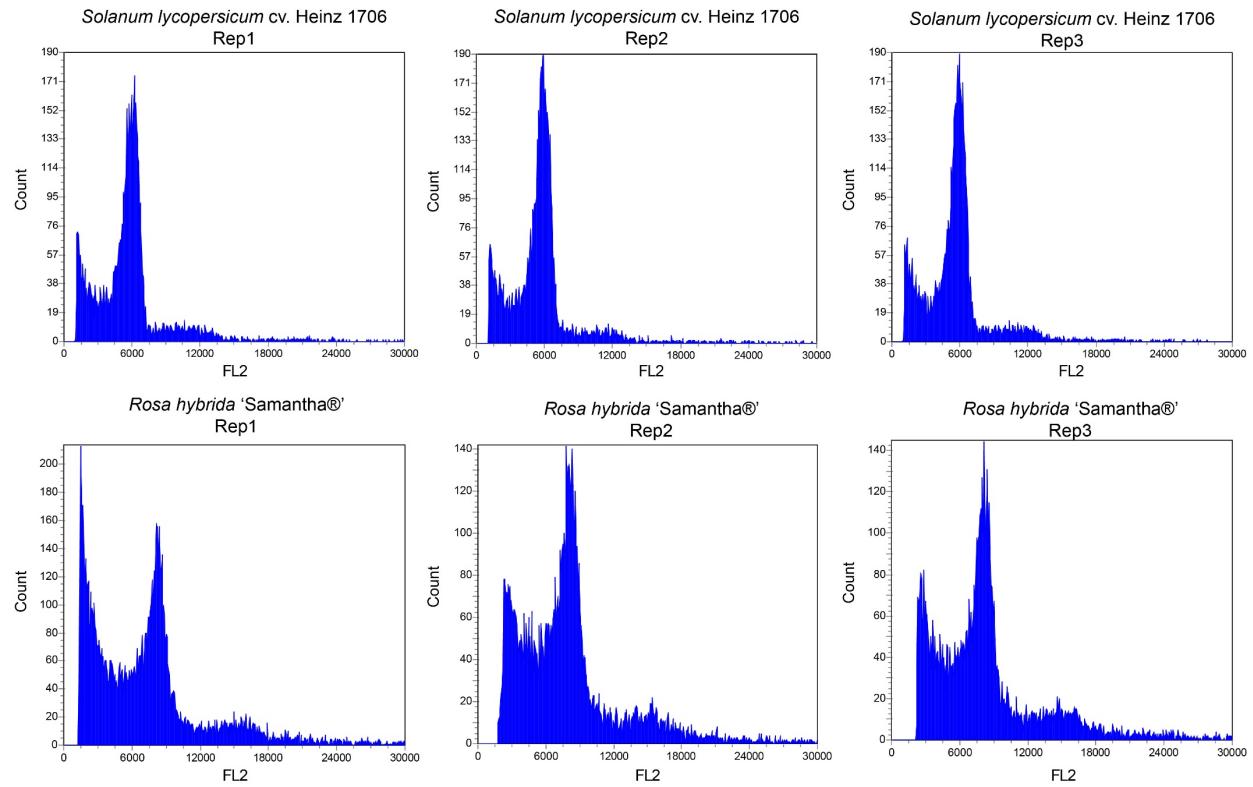
References

1. Wu, Z.-Y. & Al-Shehbaz, I. A. (eds.) *Flora of China. Vol. 9: Pittosporaceae through Connaraceae*. Science Press (2003).
2. Fougère-Danezan, M., Joly, S., Bruneau, A., Gao, X.-F. & Zhang, L.-B. Phylogeny and biogeography of wild roses with specific attention to polyploids. *Ann. Bot.* **115**, 275–291 (2015).
3. Debray, K. *et al.* Unveiling the patterns of reticulated evolutionary processes with phylogenomics: hybridization and polyploidy in the genus Rosa. *Syst. Biol.* **71**, 547–569 (2022).

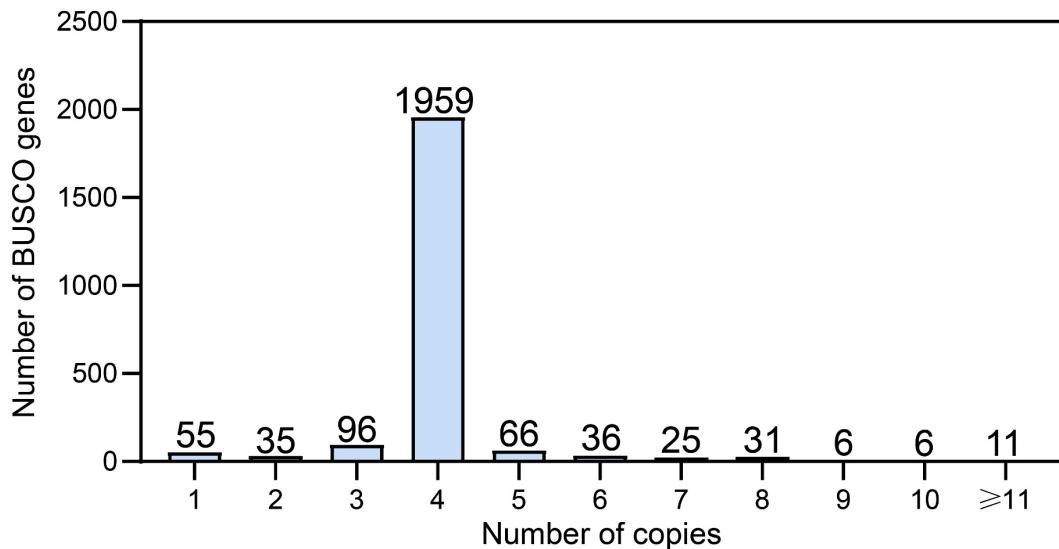
Supplementary Figures



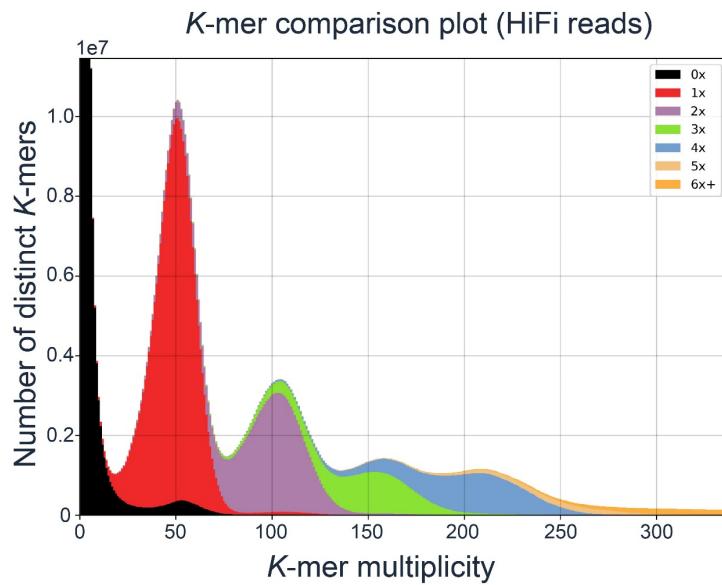
Supplementary Fig. 1 | Genome size estimation of ‘Samantha®’ through k -mer analysis. K -mer abundance in the HiFi reads was calculated with the KMC software (<https://github.com/refresh-bio/KMC>) using a k -mer size of 21, and GenomeScope 2.0 (<http://genomescope.org/genomescope2.0/>) was used to estimate the genome size. The estimated monoploid genome size is 476.9 Mb; therefore, the estimated tetraploid genome size of ‘Samantha®’ is 1907.6 Mb.



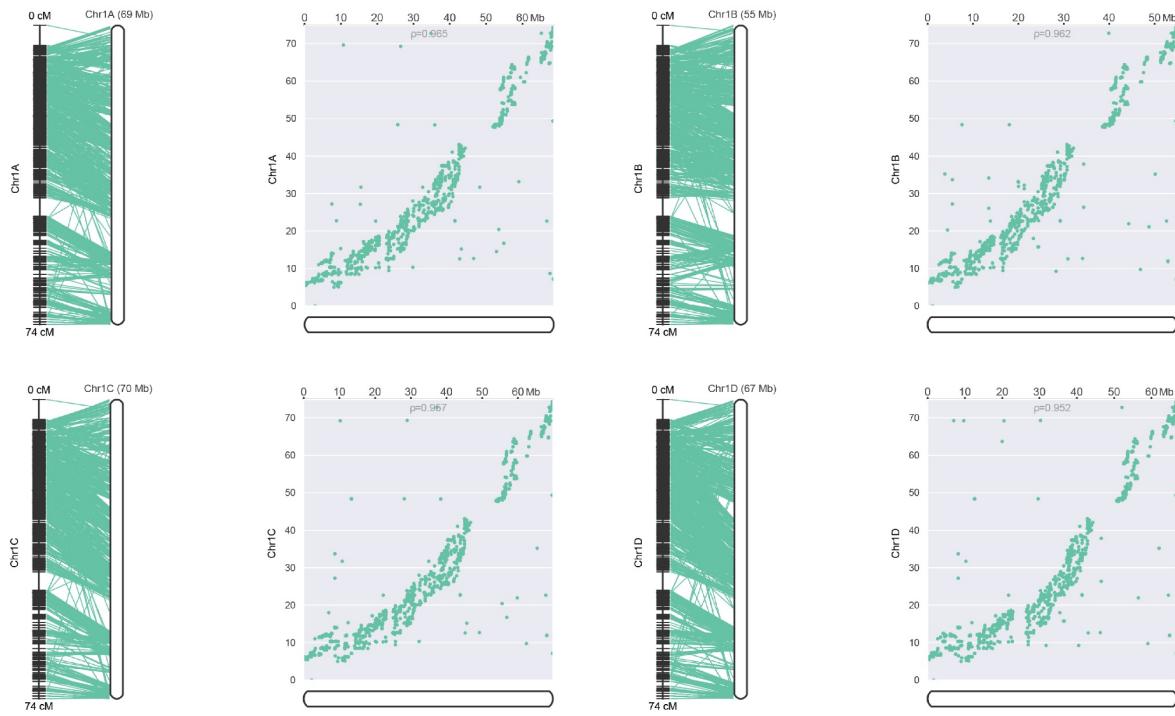
Supplementary Fig. 2 | Genome size estimation of 'Samantha®' genome using flow cytometry. C-value estimation of 'Samantha®' by flow cytometry using *Solanum lycopersicum* cv. Heinz 1706 as the internal reference standards. G1 peak x-values are listed in Supplementary Table 1.



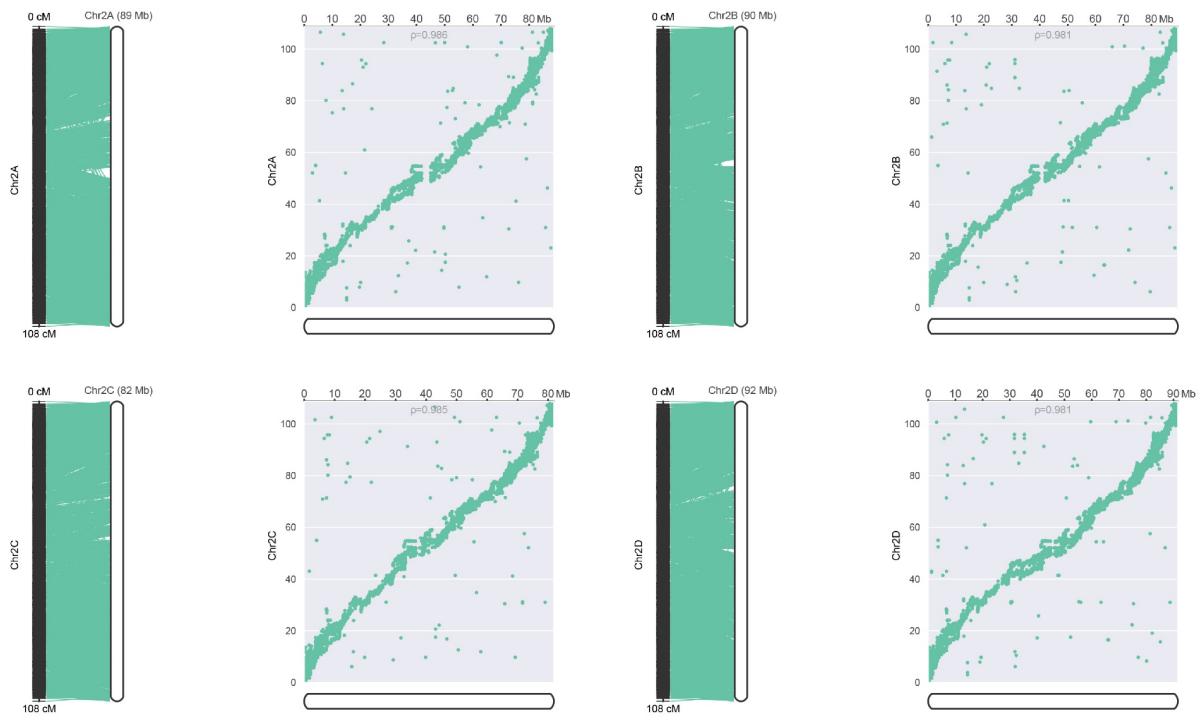
Supplementary Fig. 3 | Distribution of the number of core BUSCO genes completely captured in the ‘Samantha®’ genome assembly. A total of 2,326 eudicots_odb10 BUSCO groups were used in the analysis.



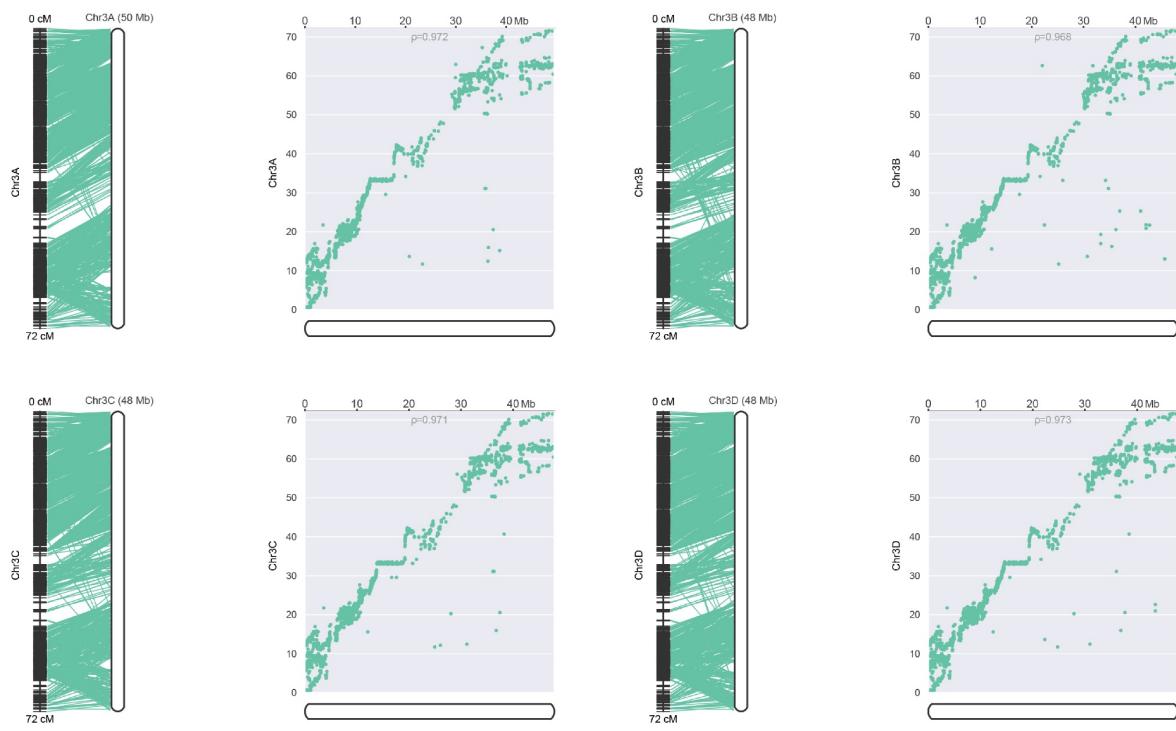
Supplementary Fig. 4 | *K*-mer spectrum analysis. Comparison of 21-mer spectra between HiFi reads and the ‘Samantha®’ assembly. Different colors indicate different copies of the 21-mers in the genome assembly. *K*-mer spectrum analysis was conducted using the KAT program (v.2.4.2; <https://github.com/TGAC/KAT>). The results suggested that the tetraploid genome assembly of ‘Samantha®’ captured the vast majority of the *k*-mers found in HiFi reads (very small black peaks at >1×), indicating the high completeness of the assembly.



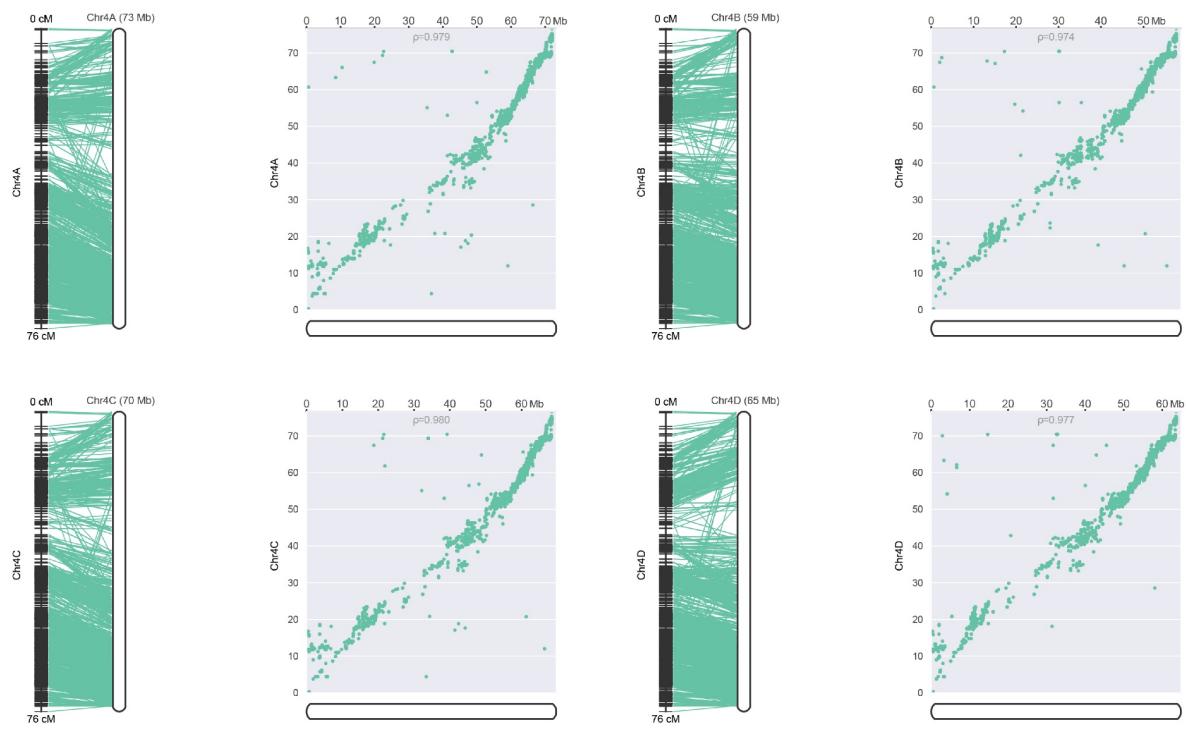
Supplementary Fig. 5 | Collinearity of the assembled ‘Samantha®’ chromosomes with the genetic map. The left panels show the assembled chromosomes and corresponding linkage groups. The right panels show scatterplots of marker distance (cM) and physical distance (Mb). Rho (ρ) in scatterplots represents the Pearson’s correlation coefficient between marker distance and physical distance. The genetic map was constructed using the tetraploid ‘K5’ cut rose mapping population that consists of 172 individuals of the cross between ‘P540’ (mother) and ‘P867’ (father).



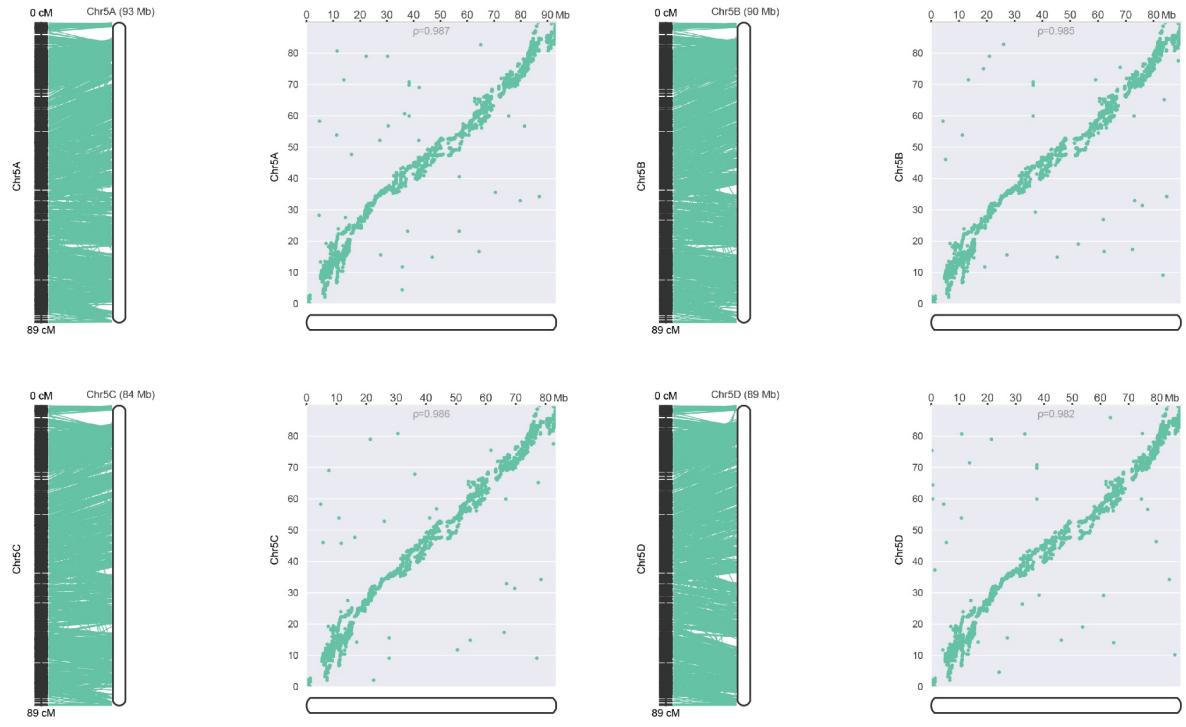
Supplementary Fig. 5 | Continued.



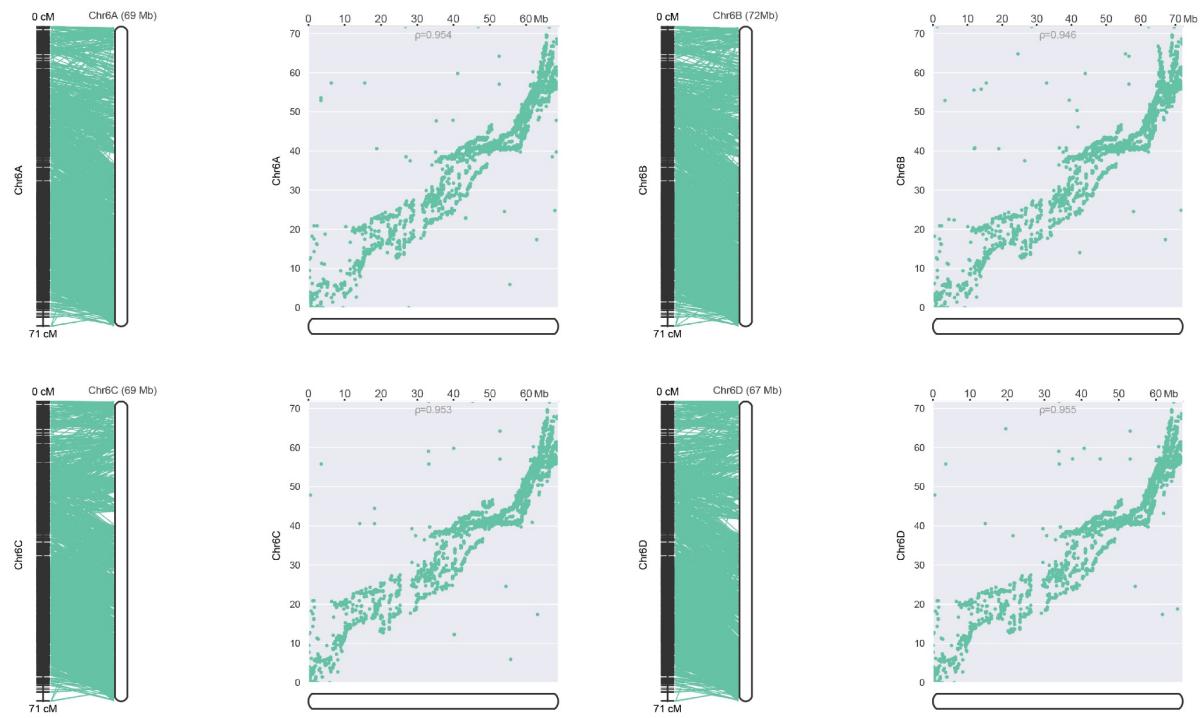
Supplementary Fig. 5 | Continued.



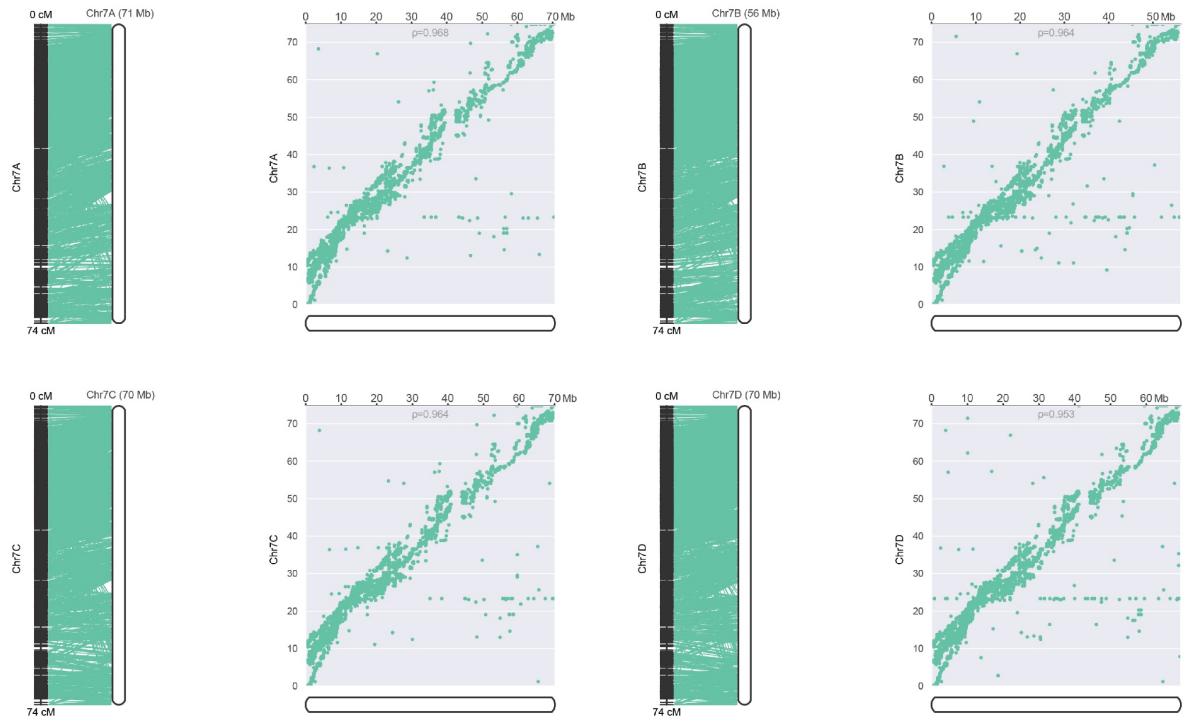
Supplementary Fig. 5 | Continued.

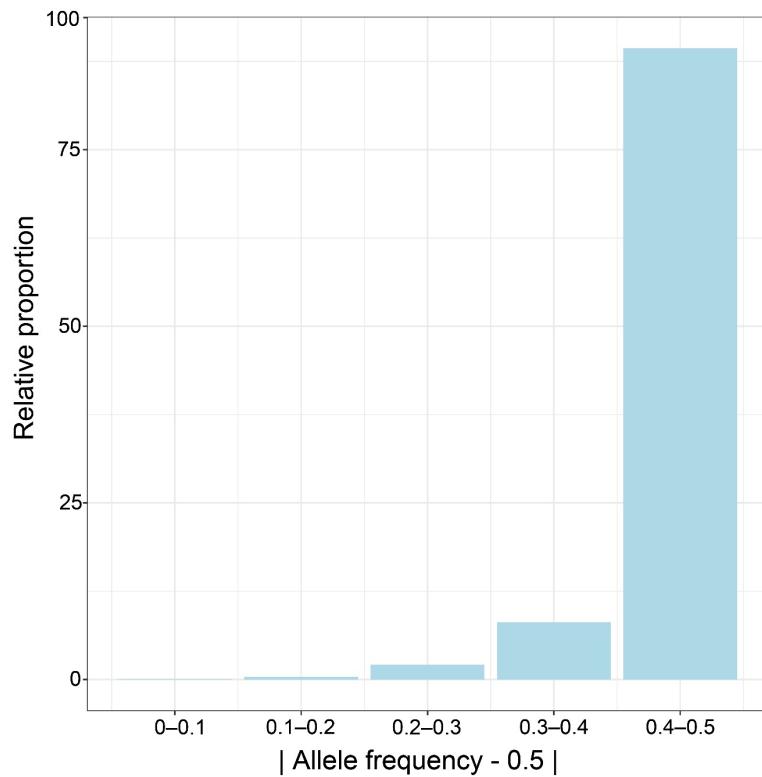


Supplementary Fig. 5 | Continued.

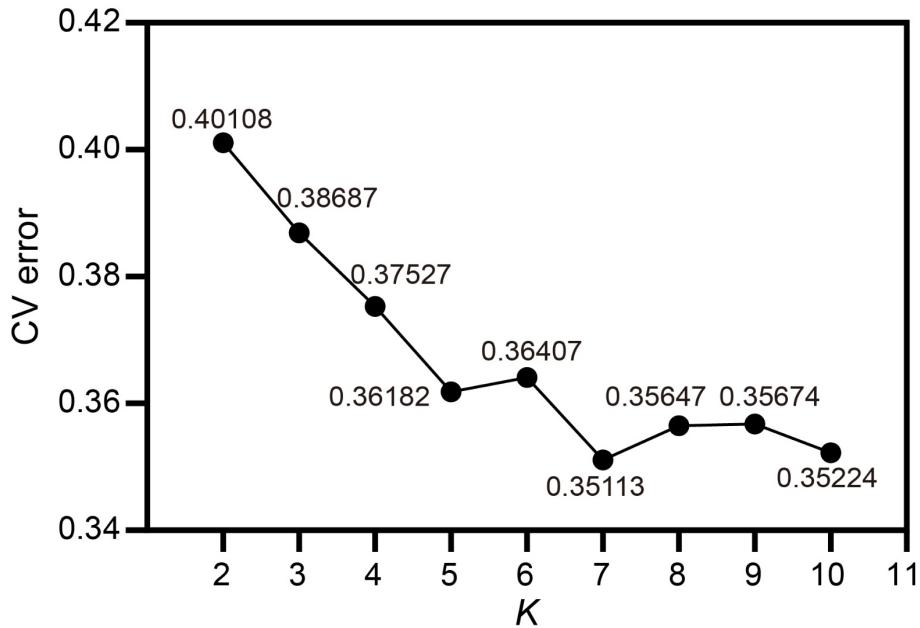


Supplementary Fig. 5 | Continued.

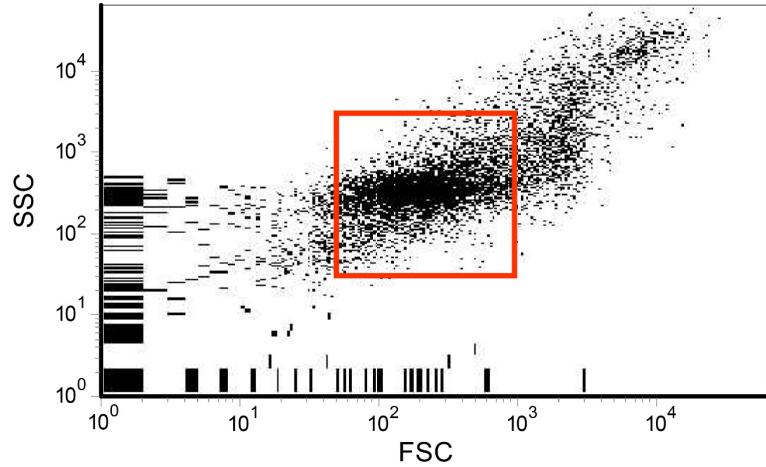




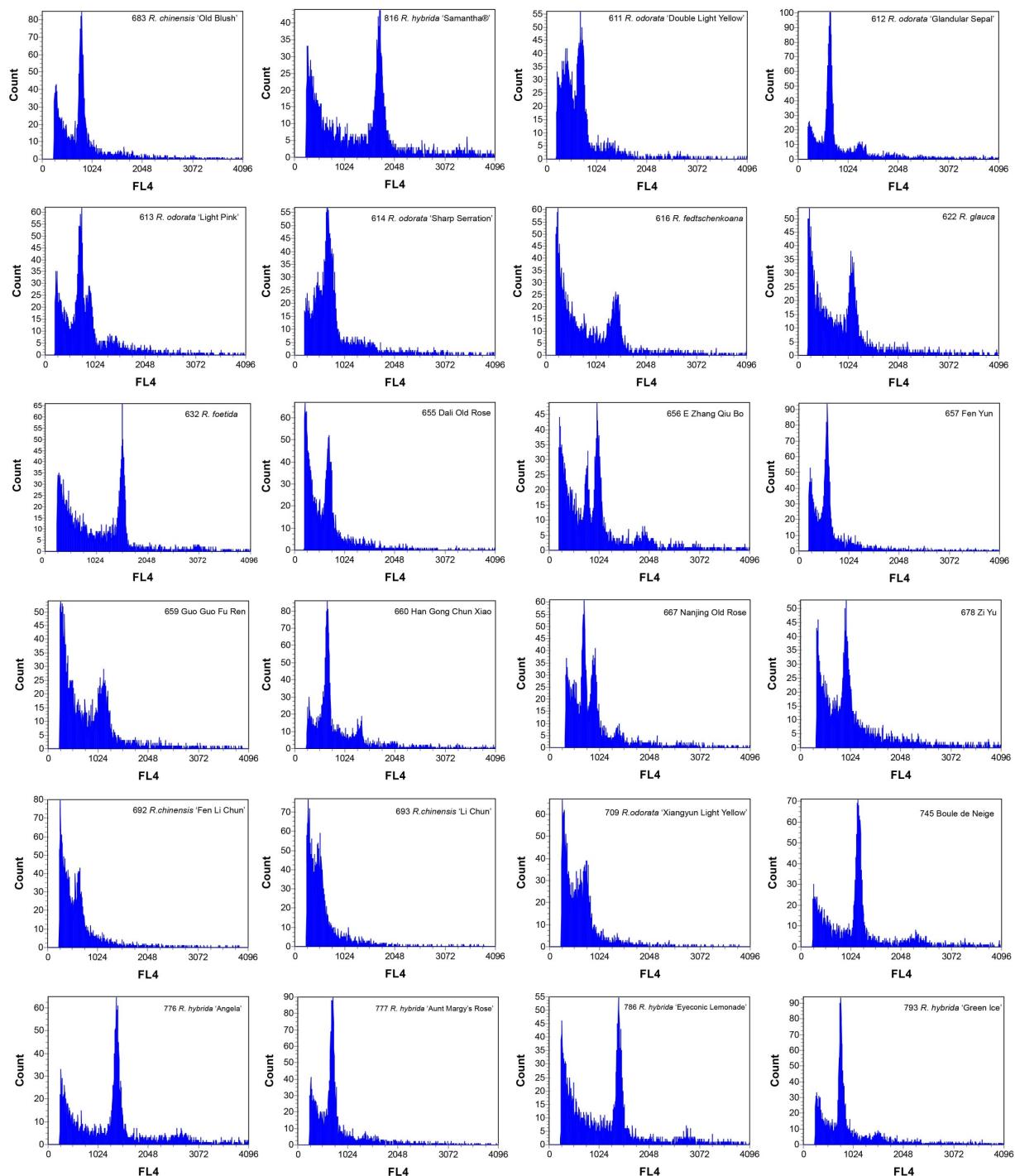
Supplementary Fig. 6 | Validating the accuracy of genome phasing. Allele frequency was used to verify the accuracy of the phasing of the ‘Samantha®’ genome assembly using nPhase (<https://github.com/OmarOakheart/nPhase>). An allele frequency close to 0.5 or the absolute value of the allele frequency minus 0.5 ($|\text{Allele frequency} - 0.5|$) close to zero is considered a potential incorrectly phased region. A very small portion (0.07%) of the assembled ‘Samantha®’ genome had $|\text{Allele frequency} - 0.5|$ close to zero, indicating a very low potential phase switching error rate in the ‘Samantha®’ genome assembly.



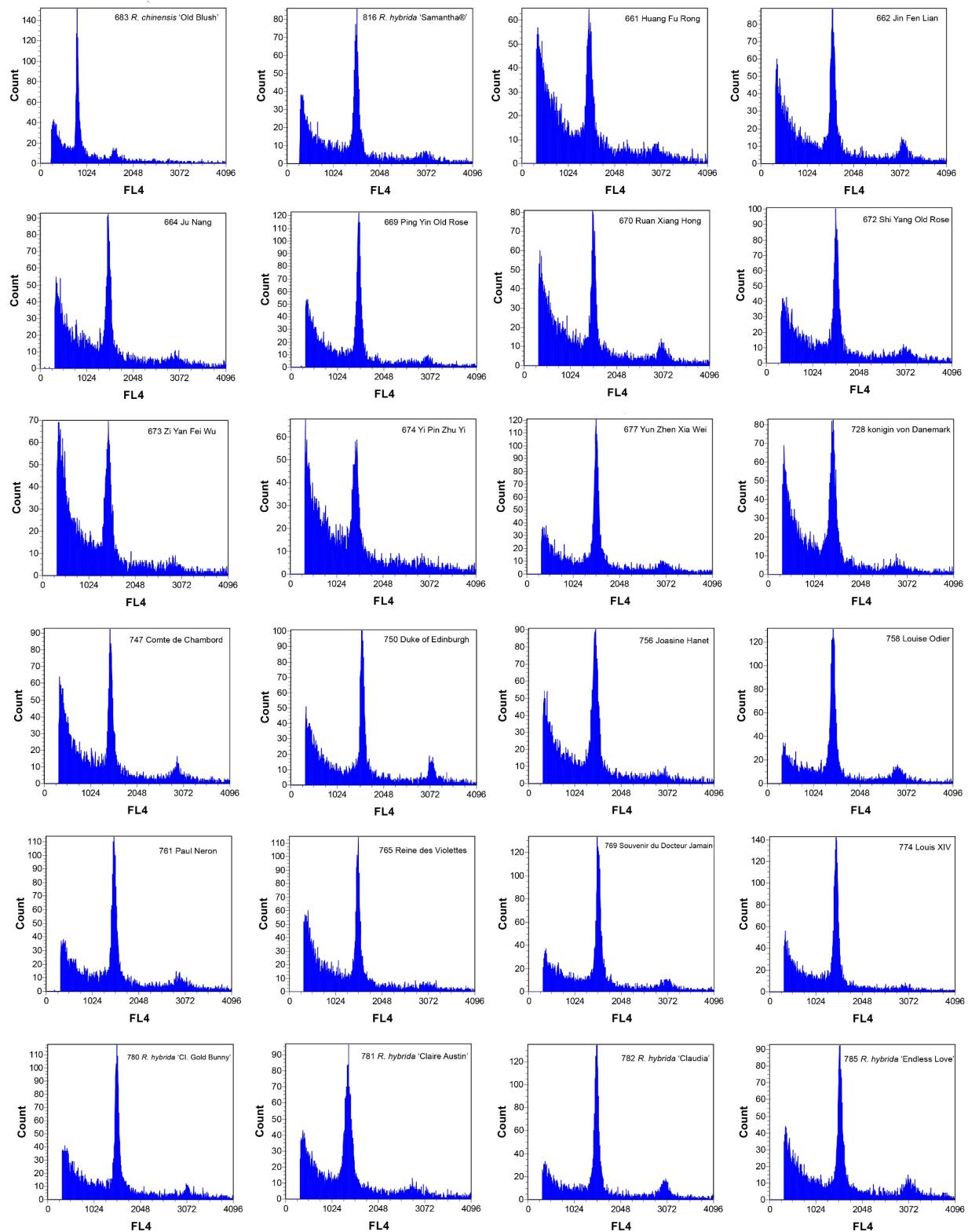
Supplementary Fig. 7 | Cross-validation (CV) error with K ranging from 2 to 10 in the population structure analysis using ADMIXTURE.

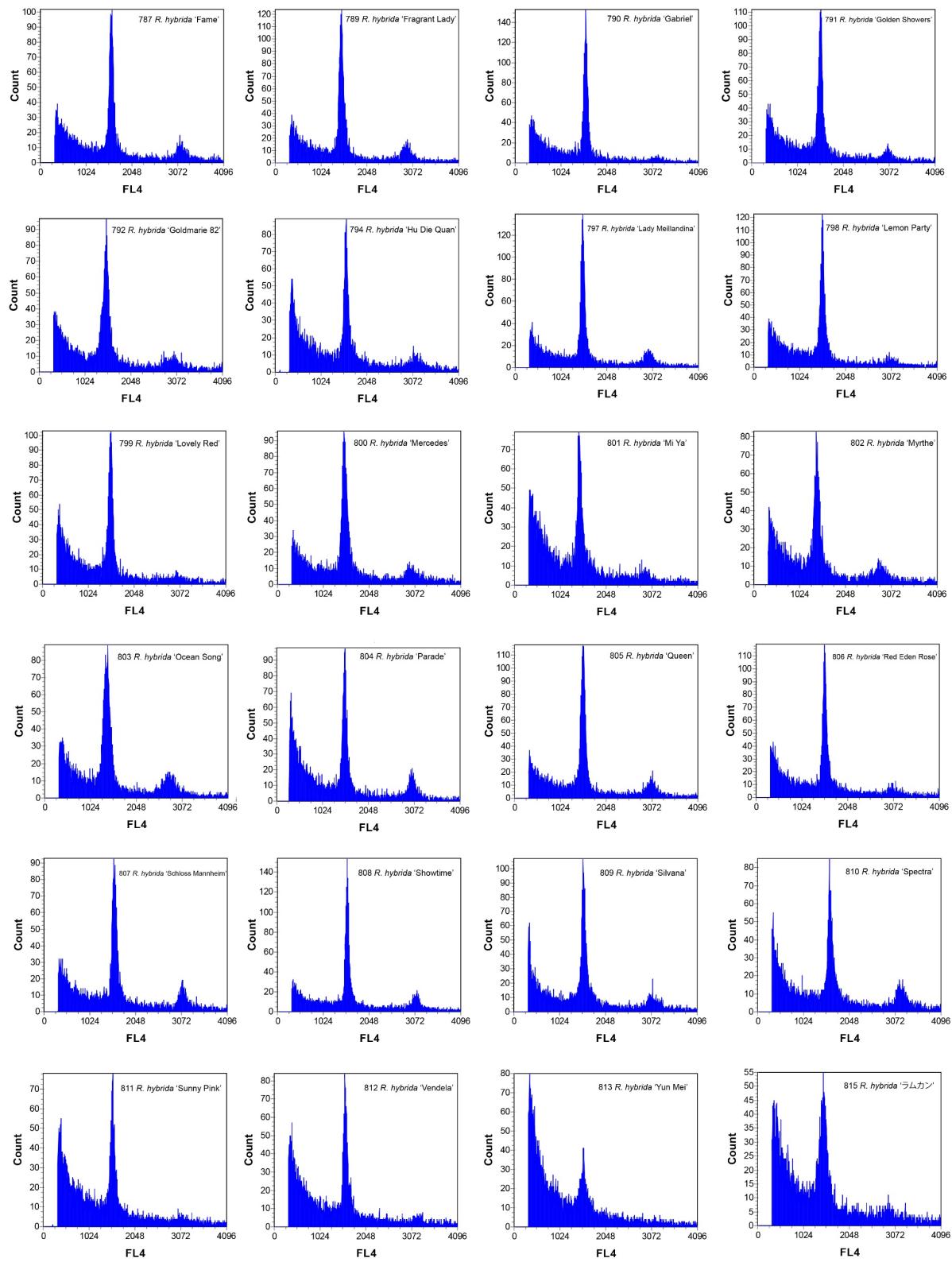


Supplementary Fig. 8 | Illustration of the gating strategy for flow cytometry analysis. The red box highlights the selected region. FSC and SSC represent forward scatter and side scatter, respectively.



Supplementary Fig. 9 | Flow cytometry analysis of ploidy levels in different rose accessions. The x-axis represents the DNA content, and the y-axis represents cell count. *R. chinensis* 'Old Blush' and *R. hybrida* 'Samantha®' are used to represent diploid and tetraploid, respectively.





Supplementary Fig. 9 | Continued.