Comparative genetic mapping and a consensus interspecific genetic map reveal strong synteny and collinearity within the Citrus genus

Patrick Ollitrault, Bárbara Hufnagel, Franck Curk, Aude Perdereau, Pierre Mournet, Maëva Miranda, Gilles Costantino, Yann Froelicher, Mônica Alves, Maria Angeles Forner Giner, et al.

frontiers

Comparative genetic mapping and a consensus interspecific genetic map reveals strong synteny and collinearity within the Citrus genus

Patrick Ollitrault[1*], Barbara HUFNAGEL[2], Franck Curk[3], Aude Perdereau[4], Pierre Mournet[1], Maëva Miranda[1], Gilles Costantino[5], Yann Froelicher[6], Mônica Alves[7], Maria Angeles Forner-Giner[8], Malcom W. Smith[9], Pablo Aleza[8], Francois L. Luro[5], Nelson A. Wulff[7], Leandro Peña[10], Raphael Morillon[1], Arnaud Lemainque[4]

[1]UMR AGAP-Institut, CIRAD, France, [2]UMR AGAP-Institut, CIRAD, Guadeloupe, [3]UMR AGAP-Institut, INRAE, France, [4]Genoscope, Institut de Biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, France, [5]UMR AGAP-Institut, INRAE, France, [6]UMR AGAP-Institut, CIRAD, France, [7]Fundo de Defesa da Citricultura, Brazil, [8]Valencian Institute for Agricultural Research (IVIA), Spain, [9]Department of Agriculture and Fisheries, Bundaberg Research Station, Australia, [10]Instituto de Biologia Molecular y Celular de Plantas, Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia, Spain

Submitted to Journal:
Frontiers in Plant Science

Specialty Section:
Plant Breeding

Article type:
Original Research Article

Manuscript ID:
1475965

Received on:
04 Aug 2024

Journal website link:
www.frontiersin.org

Scope Statement

Comparative genetic mapping provide evolutive information at structural level essential for breeding projects based on wide genetic diversity. Moreover consensus genetic maps established from multiple individual maps allow to enhances the explored diversity and improves genome coverage. Such studies are particularly important for citrus, where the useful diversity for breeding covers a wide range of interspecific diversity. In this paper, we performed a GBS analysis of 1,216 diploid hybrids from 10 progenies to establish 9 individual genetic maps at interspecific level. The predominance of the cutting site of Apek1 in gene sequences permitted the development of a consensus map based on common gene-based markers between the various individual maps. With 10,756 loci our consensus genetic map constitutes the most saturated genetic framework published in citrus. The anchoring of the consensus map on 15 published citrus chromosome-scale genome assemblies revealed a high synteny and collinearity for the most recent assemblies, whereas discrepancies were observed for some older ones. The evidenced structural conservation at interspecific level is a favorable situation for further interspecific sexual breeding as well as for translational genomics. The consensus map may be easily extended to other species and horticultural groups, taking advantage of the gene-based marker approach.

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Credit Author Statement

Pablo Aleza: Funding acquisition, Investigation, Project administration, Resources, Writing – review & editing. Aude Perdereau: Investigation, Methodology, Validation, Writing – review & editing. Arnaud Lemainque: Funding acquisition, Investigation, Project administration, Supervision, Writing – review & editing. Barbara HUFNAGEL: Formal Analysis, Investigation, Writing – review & editing. Maria Angeles Forner-Giner: Investigation, Resources, Writing – review & editing. Franck Curk: Formal Analysis, Investigation, Resources, Visualization, Writing – review & editing. Francois Leon Luro: Investigation, Resources, Writing – review & editing. Gilles Costantino: Data curation, Investigation, Validation, Writing – review & editing. Leandro Peña: Funding acquisition, Investigation, Project administration, Writing – review & editing. Maëva Miranda: Investigation, Methodology, Writing – review & editing. Malcom W Smith: Funding acquisition, Investigation, Resources, Writing – review & editing. Mônica Alves: Formal Analysis, Investigation, Resources, Writing – review & editing. Nelson Arno Wulff: Investigation, Project administration, Resources, Writing – review & editing. Patrick Ollitrault: Conceptualization, Formal Analysis, Funding acquisition, Investigation, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing. Pierre Mournet: Formal Analysis, Investigation, Methodology, Validation, Writing – review & editing. Raphael Morillon: Funding acquisition, Project administration, Writing – review & editing. Yann Froelicher: Investigation, Resources, Writing – review & editing.

Keywords

Genotyping by sequencing (GBS), single nucelotide polymorphisms, Citrus, Genetic map comparison, Structural genome variations, Consensus genetic linkage map, recombination landscape, skewed segregation

Abstract

Word count: 330

Useful germplasm for citrus breeding includes all sexually compatible species of the former genera Citrus, Clymenia, Eremocitrus, Fortunella, Microcitrus, Oxanthera, and Poncirus, now merged in the single Citrus genus. An improved knowledge on the synteny/collinearity between the genome of these different species, and on their recombination landscapes, is essential to optimize interspecific breeding schemes. We have performed a large comparative genetic mapping study including several main clades of the Citrus genus. It concerns five species (C. maxima, C. medica, C. reticulata, C. trifoliata and C. glauca), two horticultural groups resulting from interspecific admixture (clementine and lemon) and two recent interspecific hybrids (C. australis x C. australasica and C. maxima x C. reticulata). The nine individual genetic maps were established from GBS data of 1,216 hybrids. The number of SNPs mapped for each parent varies from 760 for C. medica to 4,436 for the C. maxima x C. reticulata hybrid, with an average of 2,162.3 markers by map. Their comparison with C. clementina v1.0 assembly and inter-map comparisons revealed a high synteny and collinearity between the nine genetic maps. Non-Mendelian segregation was frequent and specific for each parental combination. The recombination landscape was similar for the nine mapped parents, and large genomic regions with very low recombination were identified. A consensus genetic map was successfully established. It encompasses 10,756 loci, including 7,915 gene-based markers and 2,841 SNPs out-gene sequences. The anchoring of the consensus map on 15 published

citrus chromosome-scale genomeassemblies revealed a high synteny and collinearity for the most recent assemblies, whereas discrepancies were observed for some older ones. Large structural variations do not seem to have played a major role in the differentiation of the main species of the Citrus genus. The consensus genetic map is a useful tool to check the accuracy of genome assemblies, identify large structural variation and focus on analyzing potential relationships with phenotypic variations. It should also be a reference framework to integrate the positions of QTLs and useful genes identified in different analyses.

Funding information

This research was funded by France Génomique (ANR-10-INBS-09–08; “Dynamo” project), European Feder-Guadeloupe region “Cavalbio” project, H2020 Innovation Action Program (grant #817526, preHLB project), French National Research Agency (grant ANR-23-CE20-0038), Ministry of Science and Innovation Spain (project PID2021-1281150R-I00), MCIN/AEI/10.13039/501100011033 /FEDER, UE., and Hort. Innovation, Australia.

Funding statement

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article.

Ethics statements

Studies involving animal subjects
Generated Statement: No animal studies are presented in this manuscript.

Studies involving human subjects
Generated Statement: No human studies are presented in the manuscript.

Inclusion of identifiable human data
Generated Statement: No potentially identifiable images or data are presented in this study.

Data availability statement

Generated Statement: The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

Comparative genetic mapping and a consensus interspecific genetic map reveals strong synteny and collinearity within the *Citrus* genus

Patrick Ollitrault[1,2] *, Barbara Hufnagel[1,3], Franck Curk[4], Aude Perdereau[5], Pierre Mournet[1,2], Maëva Miranda[1,2], Gilles Costantino[6], Yann Froelicher[1,7], Mônica Alves[8], Maria Angeles Forner Giner[9], Malcom W. Smith[10], Pablo Aleza[9], François Luro[6], Nelson A. Wulff[8], Leandro Peña [11], Raphaël Morillon[1,2] and Arnaud Lemainque[5]

1 UMR AGAP, Institut Agro, CIRAD, INRAE, University of Montpellier, F-34060 Montpellier, France;
2 UMR AGAP, CIRAD, F-34398 Montpellier, France
3 UMR AGAP, CIRAD, F-97170 Petit-Bourg, France
4 UMR AGAP, INRAE, Institut Agro, CIRAD, University of Montpellier, F-34060 Montpellier, France
5 Genoscope, Institut de Biologie François-Jacob, Commissariat à l'Energie Atomique (CEA), Université Paris-Saclay, F-91000 Evry, France
6 UMR AGAP, INRAE, Institut Agro, CIRAD, University of Montpellier, F-20230 San Giuliano, France
7 UMR AGAP, CIRAD, F-20230 San Giuliano, France
8 Fundo de Defesa da Citricultura, Araraquara, Brazil
9 Departamento de Citricultura y Producción Vegetal, Instituto Valenciano de Investigaciones Agrarias (IVIA), Moncada, Valencia, Spain
10 Department of Agriculture and Fisheries, Bundaberg Research Station, Bundaberg, QLD, 4670, Australia
11 Instituto de Biologia Molecular y Celular de Plantas – Consejo Superior de Investigaciones Científicas, Universidad Politécnica de Valencia, Valencia, Spain
* Correspondence: patrick.ollitrault@cirad.fr

Abstract

Useful germplasm for citrus breeding includes all sexually compatible species of the former genera *Citrus, Clymenia, Eremocitrus, Fortunella, Microcitrus, Oxanthera*, and *Poncirus*, now merged in the single *Citrus* genus. An improved knowledge on the synteny/collinearity between the genome of these different species, and on their recombination landscapes, is essential to optimize interspecific breeding schemes. We have performed a large comparative genetic mapping study including several main clades of the *Citrus* genus. It concerns five species (*C. maxima, C. medica, C. reticulata, C. trifoliata* and *C. glauca*), two horticultural groups resulting from interspecific admixture (clementine and lemon) and two recent interspecific hybrids (*C. australis x C. australasica* and C. *maxima x C. reticulata*). The nine individual genetic maps were established from GBS data of 1,216 hybrids. The number of SNPs mapped for each parent varies from 760 for *C. medica* to 4,436 for the *C. maxima x C. reticulata* hybrid, with an average of 2,162.3 markers by map. Their comparison with *C. clementina* v1.0 assembly and inter-map comparisons revealed a high synteny and collinearity between the nine genetic maps. Non-Mendelian segregation was frequent and specific for each parental combination. The recombination landscape was similar for the nine mapped parents, and large genomic regions with very low recombination were identified. A consensus genetic map was successfully established. It encompasses 10,756 loci, including 7,915 gene-based markers and 2,841 SNPs out-gene sequences. The anchoring of the consensus map on 15 published citrus chromosome-scale genome assemblies revealed a high synteny and collinearity for the most recent assemblies, whereas discrepancies were observed for some older ones. Large structural variations do not seem to have played a major role in the differentiation of the main species of the *Citrus* genus. The consensus genetic map is a useful tool to check the accuracy of genome assemblies, identify large structural variation and focus on analyzing potential relationships with phenotypic variations. It should also be a reference framework to integrate the positions of QTLs and useful genes identified in different analyses.

1. Introduction

Cultivated citrus and related sexually compatible species constitute a highly polymorphic group whose taxonomic treatment is still controversial. This group is part of the Citrinae subtribe in the Citreae tribe of the aurantioideae subfamily. Swingle and Reece (Swingle and Reece, 1967) proposed to subdivide the Citrinae into three groups. According to their taxonomic system that is still widely used, one group, the "true citrus," includes the *Citrus* genus with most cultivated species and five other genera: *Poncirus*, *Fortunella*, *Eremocitrus*, *Microcitrus* and *Clymenia*. Within *Citrus*, five ancestral species of the actual horticultural groups are clearly identified by phylogenetics (Garcia-Lor et al., 2011; Curk et al., 2016) and phylogenomic studies (Wu et al., 2018; Ahmed et al., 2019): *C. maxima* (pummelos), *C. medica* (citrons), *C. reticulata* (mandarins) and two wild species classified in the subgenus papeda by Swingle and Reece: *C. micrantha* and *C. ichangensis*. The other horticultural group with high economic importance such as sweet-and-sour oranges, grapefruits, lemons and limes results from admixture between these ancestral taxa. However, there is biological evidence that is inconsistent with the circumscription of the genus *Citrus*, as proposed by Swingle and Reece (Swingle and Reece, 1967). The different species of the other true citrus genera display sexual compatibility with the *Citrus* species (Iwamasa, M. and Nito, N., 1988). Moreover, chloroplast and nuclear phylogenetic studies (Bayer et al., 2009; Carbonell-Caballero et al., 2015; Wu et al., 2018) reveal that all "true citrus" species plus *Oxanthera* species constitute a monophyletic clade whose internal organization does not fit with the Swingle and Reece [1] classification. Wu et al. (Wu et al., 2018) proposed a first radiation of the true citrus in the late Miocene (6–8 Ma) and a more recent radiation between Australian species occurring during the early Pliocene epoch, around 4 Ma (Figure 1). The sexual compatibility and the phylogenetic studies support the proposal of Mabberley (Mabberley, 1998) and Zhang and Mabberley (Zhang Dianxiang and Mabberley David J., 2008) to integrate *Poncirus*, *Fortunella*, *Microcitrus*, *Eremocitrus* and *Clymenia* into the genus *Citrus*. According to chloroplastic phylogeny, the *Oxanthera* species should also be integrated into the *Citrus* genus to respect its monophyletic status (Ollitrault et al., 2020; Mabberley, 2022). In this paper, we adopted the Zhang and Mabberley concept of the *Citrus* genus for the former species of the Swingle and Reece true citrus group (Swingle and Reece, 1967). For horticultural groups resulting from interspecific admixture, we retained the trinomial taxonomic system proposed by Ollitrault et al. (Ollitrault et al., 2020) that provides an unambiguous conceptual framework for *Citrus* classification based on the phylogenomic information. All considerate *Citrus* species are diploid containing 2n=18 chromosomes.

Sources for adaptation to abiotic stresses and tolerance/resistance to pest and diseases are dispersed in the different subclades of the *Citrus* genus. According to the review of Krueger and Navarro (Krueger and Navarro, 2007), tolerance to salinity is present in Rangpur lime (*C.* x *limon* var. *limonia* Osb.), Cleopatra mandarin (*C. reticulata* var. *reshni*) and Alemow (*C.* x *aurantiifolia* var. *macrophylla* Wester). *C. glauca* and, at lower level, the Rangpur lime displays tolerance to water deficit. Adaptation to iron chlorosis is found in Rough lemon (*C.* x *limonia* var. *jambhiri* Lush), Volkamer lemon (*C.* x *limonia* var. volkameriana Ined) and Nasnaran (*C.* x *amblycarpa* (Hassk.) Ochse and Alemow. The Satsuma mandarins (*C.* × *aurantium* var. *unshiu* ined.), the kumquats (*C. japonica* (Thunb.) Swing.) and trifoliate orange (*C. trifoliata* (L.) Raf.) are cold tolerant. Sour orange (*C. x aurantium* var. *aurantium*), Volkamer lemon, Nasnaran, trifoliate orange and certain pummelos and mandarins are tolerant to *Phytophthora* sp. Tolerance to the African citrus cercosporiosis (fungal disease due to *Phaemularia angolensis* De Carvalho and O. Mendes P.M. Kirk) is found in pummelos, lemons (*C. x limon* var. *limon*) and Satsuma and Beauty (*C.* × *aurantium* var. *tangerina* ined.) mandarins. The susceptibility to another fungal disease, *Alternaria alternata* of tangerine, is only found in a limited set of mandarins related to the Dancy variety (*C.* × *aurantium* var. *tangerina* ined.). Behavior against citrus tristeza virus (CTV) is highly variable. Trifoliate orange was considered for a long time as immune but some resistance-breaking or "RB" strains of CTV have been described by Harper et al. (2010) to cause systemic infection. Partial resistance to CTV is found in some pummelos and kumquats and tolerance in different germplasm used as rootstocks such as Nasnaran, Cleopatra mandarin, Rangpur lime, Rough lemon and Volkamer lemon. Different levels of tolerance to citrus canker (bacteriosis due to *Xanthomonas citri* subsp. *citri*) are found in kumquats, Satsuma and Dancy mandarins. Trifoliate orange is a source of resistance to the nematode *Tylenchulus semipenetrans* Cobb and has been described in some studies as tolerant to Huanglongbing (the most devastating citrus disease due to the bacteria *Candidatus Liberibacter* sps.; (Folimonova et al., 2009; Albrecht and Bowman,

2011; Ramadugu et al., 2016)) and its vector (George and Lapointe, 2019). However, the determinism and consistence of the resistance/tolerance of trifoliate orange to HLB are still debated today (Alves et al., 2021). Recently, complete resistances to Huanglongbing were described in *C. australis, C. glauca, C. warburgiana, C. wintersii,* and some of their hybrids (Alves et al., 2021). Therefore, under its new definition, the entire *Citrus* genus can be considered the fundamental germplasm to improve citrus by sexual hybridization, and all species need to be integrated into international efforts to map and exploit the citrus genome.

Meiotic recombination is a major component of the evolution of sexually reproducing species, and the resulting allele shuffling between the parental homolog chromosomes is fundamental in sexual breeding strategies. The crossover distribution is heterogeneous within and among chromosomes (Boulton et al., 1997; Myers et al., 2010), but also between species (Dumont and Payseur, 2011) and genotypes (Kong et al., 2010; Kawakami et al., 2017), and defines the recombination landscape. The centromeric and pericentromeric genome regions generally present a very low recombination level, where allele shuffling is strongly hampered in comparison with other genomic regions. High inter-homolog sequence divergence or structural variation in interspecific hybrids can also strongly affect the recombination rates in non-centromeric regions. From an evolutionary point of view, sexual recombination creates new haplotypes, which in turn can have an impact on natural selection (Hill and Robertson, 2007; Maynard and Haigh, 2007) and, consequently, on the distribution of diversity on a genome-wide scale (Ellegren and Galtier, 2016)[4]. For breeding, linkage drag due to close linkage between favorable genes/alleles and genes with unfavorable alleles can hamper the transfer of the former genes to elite material. The recombination limitation due to sequence divergence and structural variation can also be an important bottleneck for introgression breeding strategies aiming to transfer resistance genes from wild relatives in crop's genome. An improved knowledge of the synteny and collinearity between the different species, as well as of the recombination landscapes, is essential to optimize breeding schemes involving the different species of the *Citrus* genus.

Due to the high heterozygosity of citrus germplasm, most citrus genetic maps were developed from first-generation crosses, and segregation analyses allowed genetic maps to be developed for each of the parents and, in some cases, consensus genetic maps. The first high-density genetic map of citrus, based on codominant markers, was established in the framework of the project of the International Citrus Genome Consortium (ICGC) aiming to establish the citrus genome reference sequence from a haploid clementine (*C.* × *aurantium* var. *clementina* ined.) (Wu et al., 2014). The clementine genetic map based on 961 SNP, SSR and Indel markers (Ollitrault et al., 2012) was used for the final assembly of the genome in pseudomolecules. High-density maps of sweet orange (*C. x aurantium var. sinensis*) with 943 markers (Xu et al., 2013) and mandarin with 706 markers (Shimada et al., 2014) were also published during the same period. More recently, the new generation sequencing method (NGS) combined with the reduction of genome complexity were used to produce medium- to high-density genetic maps. Guo et al. (Guo et al., 2015) produced two pummelos maps from an F1 cross using RADSeq. DARTSeq was successfully applied to produce a synthetic map of Murcott [(*C.* x *aurantium* var. *sinensis*) x *C. reticulata*] and sweet orange (Curtolo et al., 2017a) as well as Sunki mandarin (*C. reticulata* var. *austera*) and Rubidoux trifoliate orange maps (Curtolo et al., 2017b). Genotyping by sequencing (GBS) allowed researchers to establish saturated genetic maps of trifoliate orange and sweet orange (Huang et al., 2018) as well as Ellendale tangor (*C.* × *aurantium* var. *tangerina ined.* X *C.* x *aurantium* var. *sinensis*) and Fortune mandarin (*C.* x *aurantium* var. *clementina* X (*C.* × *aurantium* var. *tangerina ined.* X *C.* x *aurantium* var. *paradisi*) (Ollitrault et al., 2021). The higher density integrated linkage map (4,163 markers) was published by Xu et al. (Xu et al., 2021) from a clementine x trifoliate orange family using specific locus amplified fragment sequencing (SLAF-seq) technology. A few comparative genetic mapping studies analyzed the synteny and collinearity in cultivated citrus and trifoliate orange. The conservation of synteny was complete between sweet orange, pummelo and clementine, and the linear order of markers also appeared to be highly conserved between these species (Ollitrault et al., 2012). From partial genetic mapping, Bernet et al. (Bernet et al., 2010) reported high synteny and collinearity between Fortune mandarin, Chandler pummelo (*C. maxima*), sour orange and trifoliate orange. More recently, the availability of pseudo chromosome assemblies of citrus genomes allowed the comparison of genetic and physical maps and globally confirmed the good conservation of marker order between different cultivated citrus species (*C. reticulata*, *C.* x *aurantium* var. *sinensis* and *C.* x *aurantium* var. *clementina*) and *C. trifoliata* (Curtolo et al., 2017b; Huang et al., 2018; Xu et al., 2021)

The aim of the present work was to perform a large comparative mapping study that includes most of the main clades of the *Citrus* genus (Figure 1) and, particularly, the Oceanian clade considering its importance in breeding projects for HLB resistance. It concerns five species (*C. maxima, C. medica, C. reticulata, C. trifoliata* and *C. glauca*), two horticultural groups resulting from interspecific admixture (clementine, and lemon) and two recent interspecific hybrids: *C. australis* x *C. inodora* and *C. maxima* x *C. reticulata*. The recombination landscape and non-Mendelian segregation along the genome were analyzed. As a consequence of the high synteny and collinearity observed, we were able to establish a consensus map. It includes 10,756 loci (7,915 gene-based markers and 2,841 SNPs located-out gene sequences) and encompasses 1,005.3 cM. This consensus map was anchored on most of the published citrus genomes assembled in pseudomolecules. The synteny and the collinearity between our consensus genetic map and the different genome assemblies were analyzed.

2. Material and method

2.1 Progenies

Ten progenies were used to establish nine genetic maps (Supplementary Table 1) with a total of 1,216 hybrids analyzed by GBS. Individual maps were generally established from a single parental combination. However, the *C. reticulata* map was established using three progenies involving the same *C. reticulata* Cleopatra mandarin variety (Cleopatra mandarin x trifoliate orange, Cleopatra mandarin x Troyer citrange and Troyer citrange x Cleopatra mandarin). *C.* x *aurantium* var. *clementina* was implemented with two progenies (Chandler pummelo x Clementine and Clementine x Finger limes [*C. australasica*]). All progenies were diploid except for the Mediterranean lemon x "Giant Key" lime progenies that were triploid hybrids. Indeed, the Giant Key lime is tetraploid, resulting from chromosome doubling of the Mexican lime (*C.* x *aurantiifolia* var. *aurantiifolia*) (Ahmed et al., 2019). For all families, two replicates of each parent were included in the analysis.

2.2 GBS analysis

Library preparation: Genomic DNA was isolated using the Plant DNAeasy kit (Qiagen) according to the manufacturer's instructions. The genomic DNA concentration of each sample was adjusted to 20 ng/µL, and ApeKI GBS libraries were prepared following the protocol described by Eslhire et al. (Elshire et al., 2011) with 96 DNA samples multiplexed per the GBS library. 10 µL of each DNA sample (200 ng) were digested with the ApeKI enzyme (New England Biolabs, Hitchin, UK). Digestion took place at 75 °C for 2 h. The ligation reaction was completed in the same plate as the digestion using the T4 DNA ligase enzyme (New England Biolabs, Hitchin, UK) at 22 °C for 1 h. Then, the ligase was inactivated prior to pooling the samples by holding it at 65 °C for 20 min. For each library, ligated samples were pooled and PCR amplified in a single tube. Genome complexity was reduced using PCR primers with one selective base (A) as described by Sonah et al. (Sonah et al., 2013).

Sequencing: For progenies 1 to 7, single-end (150 pb) sequencing was performed on two lanes of an Illumina HiSeq4000 platform at the Genoscope facilities (Paris, France). For progenies, 8 to 10 pair-end sequencing was performed on one lane per library of an Illumina HiSeq4000 platform at Genewiz facilities.

Variant calling: RAW sequencing data were cleaned with cutadapt (Martin, 2011) and demultiplexed with GBSX (Herten et al., 2015). SNP genotype calling was then performed with the VCF-Hunter 2.1.0 pipeline (https://github.com/SouthGreenPlatform/VcfHunter) as described in Baurens et al. (Baurens et al., 2019), using the Clementine v1.0 genome assembly https://phytozome-next.jgi.doe.gov/info/Cclementina_v1_0) as the nuclear reference genome. Positions with less than 10 reads were considered as missing data. Polymorphic positions were filtered for diallelic SNPs and minor allele frequency greater than 0.05.

For progenies involving species widely different from the reference genome, we filtered to keep only those SNPs within annotated gene sequences. This reduced error rates and allowed for more accurate comparative

mapping of shared marked genes. This strategy was applied for the *C. trifoliata*, *C. glauca* and *C. australis* x *C. australasica* maps, as well as for the *C.* x *aurantium* var. *clementina* map, for which an interspecific progeny with *C. australasica* was used in part. Only one SNP per gene was retained, selecting the one with the least amount of missing data. For the other progenies, a filter was made on a minimal distance of 5 kb between successive markers.

2.3 Genetic mapping

The two-way pseudo-testcross mapping strategy implemented for genetic mapping from progenies resulting from crosses between two heterozygous parents (Ritter et al., 1990) and used in previous high-density mapping studies in citrus (Ollitrault et al., 2012; Guo et al., 2015; Curtolo et al., 2017a; Huang et al., 2018) was applied to establish parental genetic maps. For each map, SNP markers were selected according to their respective heterozygosity for the mapped parent and homozygosity for the other one. Each set of data was filtered to retain markers and hybrids with less than 15% of missing data. The triploid progenies lemon x Giant Key lime was treated as a diploid one with heterozygous genotype X/Y attributed for both XXY and XYY allele doses. For the selected markers, heterozygous XY for lemon and homozygous YYYY for Giant Key, it allowed for an unambiguous inference to be made about the haploid lemon gamete from the genotyping data.

Linkage analysis and genetic mapping were then performed from the inferred gamete genotypes using JoinMap5 (https://www.kyazma.nl/index.php/JoinMap/). Linkage mapping was performed in the Hap option. Markers were grouped using the independence LOD score. Phases (coupling and repulsion) of the linked marker loci were automatically detected by the software. Map distances were estimated in centimorgan (cM) using the regression mapping algorithm and the Kosambi distance. After a first mapping round, singletons were identified. On the high-density maps, the probability of having two successive crossovers within a small genomic area is very low, whereas genotyping errors strongly affect the estimation of genetic distances that erroneously expand the genetic linkage groups. Therefore, as recommended by Van Os et al. (van Os et al., 2005), we replaced singletons with missing data using a homemade excel page routine and performed a second mapping round. At the same time, a few individuals displaying an aberrant number of recombination (according to the distribution of the number of recombination events across all hybrids) were removed, because of possible poor quality genotype-calling.

2.4 Analysis of segregation distortion

The matrix of phased data resulting from each genetic map analysis was used to study the skewed segregation all along the genome. The p-values for the chi-square test according to a 0.5 theoretical frequency for each allele were computed with Excel, and we used the approach proposed by Benjamini and Hochberg (Benjamini and Hochberg, 1995) to limit the false discovery rate (FDR) in multiple testing; the approach was performed according to the method of Storey (Storey, 2002) with a q-value threshold of 0.05. The results were visualized in a Circos plot.

2.5 Analysis of recombination landscapes

The recombination landscape was estimated from the genetic position of the different genetic maps and the physical one in the Clementine v1.0 assembly, removing the markers displaying discrepancies for synteny and the ones of the misplaced and inverted area of the chr3 of the clementine assembly. Local recombination rates were estimated with MareyMap Online (Siberchicot et al., 2017) using the Loess function adjustment (wherein a two-degree polynomial is fitted in each sliding window) with a span parameter of 0.10. The results were visualized in a Circos plot.

2.6 Anchoring of genetic maps on published assemblies

The anchorage of the genetic maps on different published genome assemblies in pseudochromosomes was performed using the locOnRef tool of the Scaffhunter toolbox (Martin et al., 2016).

We analyzed the synteny and collinearity of our genetic maps with related published genome assemblies. The clementine genetic map was compared with the *C. clementina* V1.0 genome (Wu et al., 2014). For pummelo, the Chandler map was anchored in the *Citrus maxima* (*C. grandis*) genome v1.0 (Wang et al., 2017) and *Citrus maxima* Cupi Majiayou v1.0 genome (Lu et al., 2022). For the sweet orange, the maps of its two ancestral species *C. reticulata* (Cleopatra mandarin) and *C. maxima* (Chandler pummelo) and the *C. maxima* x *C. reticulata* hybrid (Pink x Tardia) were anchored on the following: (i) the first citrus genome assembly *Citrus sinensis* Valencia genome v1.0 (Xu et al., 2013), (ii) the *Citrus sinensis* Di-Haploid Sweet Orange (DHSO) v3.0 (Wang et al., 2021) and (iii) the recent haplotypes genome assembly *Citrus sinensis* cv. Valencia DVS_A genome v1.0 and *Citrus sinensis* cv. Valencia DVS_B genome v1.0 published by Wu et al. (Wu et al., 2022). The *C. trifoliata* map was anchored on *C. trifoliata* v1.3.1 (Peng et al., 2020) and ASM1835013v1 (Huang et al., 2021). The C. lemon map was anchored on (i) the two haplotype assemblies of Di Guardo et al. (Guardo et al., 2021): *Citrus limon* L. Burm f. genome v1.0 – Primary, *Citrus limon* L. Burm f. genome v1.0 – Alternative and (ii) the two haplotypes chromosome-scale assembly of *Citrus limon* cv. Eureka genome v1.0 (Bao et al., 2023). The *C. australis* x *C. inodora* map was anchored on the *C. australis* genome v1.0 (Nakandala et al., 2023).

The consensus genetic map was anchored on all genome assemblies mentioned above plus the *C. japonica* var. *hindsii* S3y-45 genome v2.0 (Wang et al., 2022).

2.7 Synteny and collinearity

The synteny and collinearity of genetic maps with genome assemblies and between genetic maps were visualized using Circos (Krzywinski et al., 2009) in Galaxy (Rasche and Hiltemann, 2020) and drawing Marey maps using Excel. Collinearity was estimated with the Spearman's rank correlation coefficient.

2.8 Consensus genetic maps

A composite map was constructed using LPmerge v1.7 (Endelman and Plomion, 2014) for each chromosome, choosing the linkage group with the least root mean-squared error (RMSE) over the "max.interval" parameter range (1–10) evaluated. Four parameters were tested: weighting for population size (N = 68-187), weighting for the number of markers (N = 987-4436), ratio between the number of markers and population size and an unweighting model. The map weighting for the number of markers presented the least RMSE among the max. interval tested, and it was chosen.

3. Results

3.1 Individual genetic maps and comparison with C. clementina v1.0 reference genome

The number of mapped SNPs for each parent varies from 760 for *C. medica* to 4,436 for the *C. maxima* x *C. reticulata* hybrid, with an average of 2,162.3 markers/map (Table 1; Supplementary Table 2). Nine linkage groups (LGs) corresponding to the nine chromosomes of the citrus haploid genome are found for most parents, with the exception of *C. medica* with 11 LGs (2 LGs for chr3 and chr6) and *C. trifoliata* (2 LGs for chr8). For *C. medica*, large regions of the genome appear to be uncovered by heterozygous markers; therefore, it is not possible to map them genetically. A similar situation, with a large genomic region in complete homozygosity in the center of chr8 of trifoliate orange, results in its division into two LGs. The nine genetic maps span from 599.7 cM for *C. medica* to 1,150.4 cM for *C. trifoliata*, with an average of 938.2. The number of unique positions by map varies between 372 for *C. medica* and 1,000 for the *C. maxima* x *C. reticulata* interspecific hybrid, with an average of 706.1 positions. The average gap size over individual maps varies between 0.96 cM for *C. maxima* and 2.17 cM for *C. trifoliata* (average over all maps: 1.5 cM), whereas the biggest gap size varies between 5.35 cM for *C. maxima* and 25.34 cM for *C. medica*.

The number of genes from the *Clementine* v1.0 assembly anchored by a mapped SNP marker varies respectively between 501 and 3,432 for *C. medica* and *C. maxima* x *C. reticulata* (Supplementary Table 2).

Circos representations of the links between genetic positions and physical ones in the *C. clementina* v1.0 genome assembly (Figure 2) testify to a good coverage of the whole genome for most parents. However, as previously mentioned, the *C. medica* map displays very large gaps, and important gaps are also identified for *C. clementina*, particularly in chr2, chr5 and chr8, as well as in the middle region of chr4. The Circos representations (Figure 2) and Marey maps (Supplementary Figure 1a) for the other parents with high marker density reveal discrepancies between genetic maps and *C. clementina* assembly for several genomic regions. Most of these discrepancies are shared by all genetic maps. They are particularly clearly displayed for the *C. maxima* x *C. reticulata* parent that has the higher marker density (Supplementary Figure 1b). These systematic discrepancies concern the following: (i) genomic regions of chr5 (12.8-19.4 Mb) and chr4 (around 5 Mb) genetically mapped at the end of LG7; (ii) genomic regions of chr3 (around 35 Mb) and 9 (11.1-14.5 Mb) genetically mapped on the middle of LG8; (iii) a genomic region at the beginning of chr2 (0.7-4.2 Mb) mapping at the beginning of LG4; and a small genomic region of chr8 (around 16.5 Mb) mapping at the beginning of LG6. In addition, a misplaced and inverted region of 5.2 Mb is revealed by Marey representation (Supplementary Figure 1) in chr3 by the different genetic maps, including those of *C.* x *aurantium* var. *clementina* and *C. medica*.

Despite these small discrepancies, the global synteny of all our genetic maps with the *C. clementina* v1.0 assembly is high, with insignificant variations between the nine maps (from 0.961+/−0.019 to 0.997+/-0.026, respectively, for *C. maxima* x *C. reticulata* and clementine; Supplementary Table 3). The lowest mean values across the nine maps by linkage group for LG7, LG8 and LG4 were found to be consistent with our observations from the Marey maps and Circos representations. For the syntenic markers, the collinearity between the different genetic maps and the *C. clementina* v1.0 assembly is high, with the Spearman's rank correlation varying between 0.935+/−0.054 and 0.997+/−0.001, respectively, for *C.* x *limon* var. *limon* and for *C. trifoliata* (Supplementary Table 4).

Taking advantage of the good collinearity between the different maps and the *C. clementina* v1.0 genome, this genome assembly was used as a template to analyze and compare the distribution of skewed segregation and the recombination landscape of the different parents. It is based on syntenic markers of each map with the genome assembly. The probably misplaced region of chr3 of the clementine assembly was removed for these analyses. The distributions along the genome of mapped markers, skewed markers and the recombination landscape are synthetized in Circos plots (Figure 3, Supplementary Figure 2 and Table 2).

The recombination landscape was analyzed for all maps except for that of Corsican citron (*C. medica*), which was too incomplete. Similar patterns were observed for the other eight maps, with regions with low levels of recombination corresponding to low gene density regions and peaks of recombination in regions of high gene density (Supplementary Figure 2). Looking at the *C. maxima* x *C. reticulata* figure as an example (Figure 3), we observe regions of low recombination in chr1, chr2, chr3, 4 and chr5 corresponding to the lowest gene density regions of these different chromosomes. Large regions with no or very few recombinations are observed in chr6 (first 10 MB of the chromosome), chr8 (5 to 15 MB) and chr9 (5 to 20 Mb). Chr7 is the only one to display recombinations events all along the genome as well as no region with low gene density.

The segregation of *C. maxima* cv. Chandler (Supplementary Figure 2a) appears to be totally Mendelian, whereas limited distortions are observed for clementine (4.26%; Supplementary Figure 2d), *C. reticulata* cv Cleopatra (7.08%; Supplementary Figure 2b) and *C. glauca* (7.53%; Supplementary Figure 2g). For *C. reticulata* cv Cleopatra, most of the skewed markers are grouped in chr7, where an imbalance between alleles reached 0.7/0.3. A similar imbalance is observed for chr9 of *C. glauca* that contains 60.4% of the skewed markers for this parent (Table 2 and Supplementary Figure 2g). For C. *trifoliata*, the global rate of skewed markers is 12.9% (Supplementary Figure 2i). No or very few (<2%) distortions were found in chr2, chr4, chr5, chr7 and chr9, whereas half of the markers have skewed segregation in chr8, reaching an imbalance of 0.94/0.06 for the two alleles. 20.7% of the markers display skewed segregation for *C. medica* (Supplementary Figure 2c). They are

located mostly in chr1 and chr7. In chr7, the distortion reaches the almost complete elimination of one haplotype. For *C. x limon*, 96.1%, 65.9%, 31.9% and 15.4% of the markers were skewed respectively for chr5, chr8, chr1 and chr4, with a global rate of distortion of 21.7% (Supplementary Figure 2e). The higher intensity of distortion is observed for chr5, where it reaches an imbalance of 0.8/0.2 between the two alleles. The rate of skewed markers for *C. australis* x *C. inodora* reaches 31.1% with a distribution across all chromosomes (Supplementary Figure 2h), and a maximum imbalance between alleles of around 0.7/0.3. *C. maxima x C. reticulata* segregation displays the higher level of skewed segregation with 62.7% of the markers. Almost all markers of chr1, chr4 and chr9; around half of chr2, chr5 and chr7; and 25.6% of chr6 display skewed segregations (Table 2; Figure 3). The level of distortion reaches imbalances of 0.85/0.15 in chr1 and chr9; 0.75/0.25 in chr3, chr4 and chr7; 0.7/0.3 in chr2; and 0.6/0.4 in chr6. We did not observe evidence for complete elimination of one allele. If we consider that the distortions result from unfavorable genes located close to the position with higher –log(Qvalue), we can observe that the evolution of the –log(Qvalue) pattern along the chromosome is closely linked with the segregation landscape (Figure 3). For chr1, the implied gene should be close to 25 Mb in a region with a high level of recombination. It results in a decreased distortion on both sides, leading to Mendelian segregation at the right end of the chromosome. On the left side, the decrease of the distortion is more limited in the 8-15 Mb region, which displays limited recombination and gene density. Then, from 8 Mb to the start of the chromosome (corresponding to high gene density and recombination rate), there is a strong decrease in distortion until reaching Mendelian segregation. For chr9, we can hypothesize that the unfavorable gene is located in the large region with very few recombination events and that its adverse effect influences segregation throughout this region. Its effect then decreases in the outer region, in line with high recombination rates and gene density. For chr7, the unfavorable gene may be located around 2.5 Mb in a region with a high recombination rate, and the maintenance of recombination all across the chromosome allows for recovery of Mendelian segregation on both sides. Similar interpretation can be made for the profile of distortions in chr2, chr3 and chr4.

3.2 Relation of the genetic maps with corresponding published genome assembly in pseudo-chromosomes

The Circos analyses of the link between our genetic maps and the different genome assemblies (Supplementary Figure 3) reveal two forms of chromosome numbering. The genomes published by the International citrus genome consortium and USA research groups (*C. clementina* v1.0, TrO-USA, Swo-USA-HapA; Swo-USA-HapB), as well as the trifoliate orange published by Huazhong Agricultural University (TrO-China; (Huang et al., 2021)), follow the numbering of *C. clementina* v1.0 that we used as a reference for our genetic mapping. The genomes published by the Chinese groups (Swo-China-V3, Pum-ChinaV1, Pum-ChinaV2, Lemon-China-HapA, and Lemon China-HapB), Italian groups (Lemon-It-Prim and Lemon-It-Alt) and Australian groups (*Australis*) adopt the numbering of the Swo-China-V1 [28]. The correspondences between the numbering of our genetic maps and considered genome assembly are provided in Supplementary Table 5. The analyses of synteny and collinearity account for these correspondences.

For the *C. clementina* v1.0 genome, the synteny with the clementine genetic map is high (0.997) despite the supposed misplaced regions of the *Clementina* v1.0. It can be explained because the concerned regions are not covered by the genetic map due to full homozygosity in these areas. The collinearity is very high for chr1, chr2, chr3, chr4, chr8 and chr9, but the Spearman's coefficient falls to 0.779 in chr6. It results in an average collinearity of 0.935+/−0.054.

For pummelo genomes, the synteny with the *C. maxima* cv. Chandler map is relatively low (0.887) with Pum-China-V1 due to numerous clusters of markers of the different chromosomes assigned in different linked groups. The synteny is high with the Pum-China-V2 assembly (0.987). For syntenic markers, collinearity with the genetic map is high for all chromosomes of both genome assemblies, except for chr9 of the Pum-China-V2, which has a Spearman's coefficient value of 0.503. This is due to an inversion affecting half of the assembled pseudo-chromosome.

For lemons, we observed low synteny of the *C. x limon* genetic maps with Lemon-It-Prim and lemon-It-Alt (0.889 and 0.895, respectively). However, it is very high with the two haplotypes of Eureka lemon (0.996 and 0.994, respectively for Lemon-China-HapA and Lemon-China-HapB). The collinearity of the *C. x limon* genetic map is low with the two haplotypes of Lemon-It (0.839+/−0.124 and 0.848+/−0.129, respectively), with a particularly low value in chr4 (Supplementary Table 6), whereas it is high with the two Eureka lemon haplotypes (0.997+/−0.001 and 0.996+/−0.001).

The SwO-China-V3 displays very clear increases of synteny and collinearity with *C. reticulata*, *C. maxima* and *C. maxima* x *C. reticulata* genetic maps when compared with SwO-China-V1 (Table 3). Very high synteny values are observed for all chromosomes with SwO-USA-HapB. Similar results are observed with SwO-HapA, except for chr1 and chr9, displaying a potential reciprocal translocation (Supplementary Figure 3). The collinearity between the three considered genetic maps and the two sweet orange haplotypes assemblies is also very high for all SwO-USA-HapB chromosomes and chr2, chr3, chr4, chr5, chr6, chr7 and chr8 of SwO-USA-HapA. The TrO-USA assembly displays very high synteny and collinearity with the *C. trifoliata* genetic map (0.991 and 0.969+/−0.014, respectively). For TrO-China, the synteny with the genetic map is a bit lower (0.976), mainly due to a cluster of markers of LG7 positioned in chr5 in the TrO-China assembly (Supplementary Figure 3). The collinearity of syntenic markers is high for both trifoliate orange genome assemblies but a little higher for TrO-China (0.990+/−0.003 versus 0.969+/−0.014 for TrO-USA).
The *C. australis* genome assembly displays very high synteny (0.992) and collinearity (0.998+/−0.001) with the *C. australis* x *C. inodora* genetic map.

3.3 Consensus genetic map and its comparison with individual genetic maps

To estimate the synteny and collinearity between the different genetic maps, we analyzed the links between the position on the *C. maxima x C. reticulata* map (that displays the higher marker density) and the eight other parental maps. The genes of *C. clementina* v1.0 were considered as markers to establish the link between genetic maps. Both synteny and collinearity estimated using the Spearman's coefficient are higher than 0.992 for all genetic maps (Table 4). The Circos representations (Supplementary Figure 4) show that un-syntenic markers are dispersed over the map, and we did not observe any cluster of un-syntenic markers.

Given the high degree of synteny and collinearity observed, it was appropriate to draw up a consensus genetic map. Considering the very incomplete coverage of the whole genome by the *C. medica* genetic map, we established the consensus genetic map with the eight other available individual maps. We also excluded the *C. trifoliata* map for the implementation of the LG8 consensus map because the *C. trifoliata* chr8 corresponded with two unlinked LGs. The markers of the eight maps positioned in the same gene of the *C.* x *aurantium* var. *clementina* genome assembly were considered to be a single "gene marker" to establish the consensus genetic map. The consensus genetic map spans 1,005.27 cM and includes 10756 markers (7915 C. *clementina* "gene markers" and 2841 "no-gene markers"; Table 5; Supplementary Table 7). The size of the nine linkage groups varies from 90.10 cM for LG9 to 166.94 CM for LG3. The total number of unique positions is 2,808 with an average gap size of 0.36 similar for all chromosomes. Respectively, 95.35% and 99.92% of the gaps between unique positions are less than 1 cM and 4 cM (Supplementary Figure 4). Only two gaps are over 4 cM. Both are located in LG6: one at the start (5.65 cM) and one at the end (8.3 cM).

The 10,756 gene markers and no-gene markers correspond to 17,745 SNP markers of the initial individual maps. The number of markers shared by the consensus map and the individual ones varies from 985 for *C.* x *aurantium* var. *clementina* to 4,435 for *C. maxima* x *C. reticulata* (Supplementary Table 8a), and the collinearity between the consensus map and all individual maps is very high (between 99,79% for *C. limon* and 99,94% for *C. reticulata*; Supplementary Table 8b).

3.4 Anchoring of the consensus genetic map on different genome assemblies in pseudochromosomes

Among the 17,745 SNPs of the individual map anchored on the consensus one, 17,718 are located on the nine chromosomes of the *C. clementina* v1.0 assembly. Probes of 100 bases on each side of these 17,718 SNPs were defined from the *C. clementina* v1.0 sequence (Supplementary Tables 9 and 10) and anchored in 14 genome assemblies in pseudochromosomes in order to perform synteny and collinearity analysis. The rate of successful anchorage (Table 6) varies between 91.57% for TrO-China (Huang et al., 2021) and 99.18% for SwO-China-V3(Wang et al., 2021). It is less than 94% for the two haplotypes of the lemon assemblies published in 2021 (Guardo et al., 2021). Interestingly, very good anchorage (>98%) is observed for the Australian species genome assemblies. For each genome assembly, one single marker by gene of the considered genome assembly was selected for synteny and collinearity analysis. The number of gene tags varies from 5,884 to 8,086, respectively, for Lemon-Italy-Alt and SwO-USA-B (Table 6). The detailed links

between the consensus genetic map and the anchored genes for each genome are given in Supplementary Table 11. For the anchorage rates and the number of tagged genes, important differences are observed between the two haplotypes of lemons published in 2021 (Guardo et al., 2021) and the ones published in 2023 from PACBIO HIFi data (Bao et al., 2023), with a strong increase for the last ones. Similar improvement is observed between the first *C. sinensis* genome assembly and the more recent ones (Wang et al., 2021; Wu et al., 2022). The trifoliate orange assembly published in 2021 (Huang et al., 2021) displays lower tagging values than that released in 2020 (Peng et al., 2020).

The synteny between the consensus map and the different genome assemblies varies between 0.890 and 0.998 (Table 6). It is over 0.996 for eight of the assemblies, including SwO-China-V2 (Figure 4.2), SwO-USA-HAPB (Figure 4.4), Pum-China-V2 (Figure 5.2), *C. hindsii* (Figure 5.4), Lemon-China HapA and HapB (Figures 6.3 and 6.4), TrO-USA (Figure 7.1) and *C. australis* (Figure 7.3). The Circos plots reveal different kinds of pictures for the assemblies with lower values of synteny. The assemblies of SWO-China v1.0 (Figure 4.1), Pum-China-V1 (Figure 5.1), Lemon-It-prim (Figure 6.1), Lemon-It-alt (Figure 6.2) display numerous small genomic regions and isolated markers assigned to different linkage groups of the consensus genetic map. Most of these discrepancies with the consensus map are common in the comparison with the two lemon haplotypes and the Pum-China v1.0 assemblies. For the *C. clementina* v1.0 assembly (Figure 5.3) five small genomic regions have different locations on the consensus map. The lower global synteny of TrO-China (Figure 7.2) is principally caused by a 5 Mb genomic region of chr5 located in the LG5 of the consensus map. Finally, the low syntenic value (0.902) with the haplotype A of the Valencia sweet orange assembly (Wu et al., 2022) (Figure 4.3) results from a reciprocal exchange of genomic regions between chr1 and chr9, whereas synteny is very high for the others chromosomes.

The collinearity with the consensus map estimated by the Spearman coefficient is very high (> 0.995) with 11 genome assemblies (Table 6; Supplementary Table 12). It is a little lower for SwO-China-V1, but still high (0.987). The relatively low global value for Pumello-ChinaV2 (0.939) is due to a displaced/inverted region in chr9, while collinearity is very high with all other chromosomes (0.996). The lower values (respectively 0.870 and 0.874) are observed for Lemon-It-Prim and Lemon-It-Alt with particularly low values (<0.7) for chr7 and chr8.

The synteny and collinearity of the consensus map with the *C. australis* assembly are very high (Nakandala et al., 2023). However, it is the only analyzed genome assembly for which genomic regions of several Mb are not anchored by the markers of the consensus genetic map. It is the case for 8 Mb at the beginning of chr4 (LG1), 7 Mb at the end of chr7 (LG4) and around 5 Mb at the beginning and end of chr8 (LG8).

4. Discussion

GBS using Apek1 as restriction enzyme is powerful for comparative genetic mapping and integration of genetic maps at the interspecific level

Numerous genetic maps were developed in citrus to explore the genetic control of several useful agronomic traits. However, initially, their comparison was hampered because most of them were established with dominant markers such as RAPD and AFLP. Multiallelic SSR markers allowed the first analyses of comparative genetic mapping to be conducted (Chen et al., 2008; Bernet et al., 2010; Ollitrault et al., 2012) but were limited by the relative low density of markers shared by the different maps. Shimada et al. (Shimada et al., 2014) proposed a gene-based marker approach to develop a framework genetic map, including 706 loci. This framework genetic map was useful to integrate other genetic maps based on common mapped genes. The same authors, also proposed that it should help to understand the regulation of gene expression by combining information on genetic loci and transcription profiles. However, the number of tagged genes and the resolution of their genetic map remained limited. Taking advantage of NGS, during the last 10 years, several medium- to high-density genetic maps of various *Citrus* species were produced using GBS (Huang et al., 2018; Ollitrault et al., 2021), RADSeq(Guo et al., 2015), DART-Seq (Curtolo et al., 2017a, 2017b) and SLAF-

seq (Xu et al., 2021). However, it remained difficult to integrate the maps obtained with different species because they shared very few common SNP markers. In the present work, we successfully combined the preferential distribution of the ApeKI enzyme restriction site on coding regions, demonstrated in maize and barley (Elshire et al., 2011), soybean (Sonah et al., 2013), *Populus* (Schilling et al., 2014) and citrus (Oueslati et al., 2017; Ahmed et al., 2019), and the concept of gene-based markers to develop a high-density consensus genetic map. It was established from an individual map of three Asian ancestral species (*C. maxima*, *C. reticulata* and *C. trifoliata*), one Australian species (*C. glauca*), two Asian secondary species (*C.* x *limon* var. *limon* and *C.* x *aurantium* var. *clementina*), one F1 Asian interspecific hybrid (*C. maxima* x *C. reticulata*) and one Australian interspecific F1 hybrid (*C. australis* x *C. inodora*). Most of the individual genetic maps displayed a regular coverage of the citrus genome. However, the *C. medica* map displays very large gaps corresponding to the full absence of heterozygous markers in large genomic regions. This result can be explained by the origin of the Corsican citron (used as the *C. medica* representative), from a self-fecundation of the "Poncire commun" variety (Luro et al., 2012). Similarly, several genomic regions in chr2, chr5 and chr8 were found uncovered by the *C.* x *aurantium* var. *clementina* genetic map. This was also observed for the first reference clementine genetic map (Ollitrault et al., 2012) and was attributed to fully homozygous regions resulting from inbreeding in the origin of Clementine (Wu et al., 2014). Due to its incompleteness, the *C. medica* map was not used to establish the consensus genetic map. The consensus map span 1,005.27 cM and encompasses 10,756 loci, including 7,915 gene-based markers and 2,841 SNPs located-out gene sequences. It presents 2,808 unique positions with an average gap size of 0.36 cM.

Comparative genetic mapping reveals a high synteny and collinearity between true citrus species and a similar recombination landscape

Previous comparative genetic mapping studies based on SSR markers suggested important synteny and collinearity between several cultivated citrus species (mandarin, clementine, pummelo, sweet and sour orange (Bernet et al., 2010; Ollitrault et al., 2012; Yu et al., 2016)) and even between cultivated citrus species (pummelo, sweet and sour orange) and *C. trifoliata* (Chen et al., 2008; Bernet et al., 2010). However, these conclusions were based on partial maps and low numbers of markers with a maximum of 418 markers shared between sweet orange and clementine (Ollitrault et al., 2012). The anchoring on the same reference genome of different high-density genetic maps, established with SNP markers (Yu et al., 2016; Huang et al., 2018; Ollitrault et al., 2021), highlighted the high synteny between genetic maps, even though they also evidenced some discrepancies between the considered genome assemblies and genetic maps. In the present work, the anchoring of the nine genetic maps on the *C. clementina* v1.0 genome assembly and the comparison between the *C. maxima* x *C. reticulata* map and the other eight individual genetic maps based on common gene-based markers revealed a very high synteny and collinearity for all genetically mapped parents. The conclusions for a highly conserved structure of the nuclear genomes therefore concern all the major clade of the true citrus group as defined by Swingle and Reece (Swingle and Reece, 1967), now joined in the new definition of the *Citrus* genus (Mabberley, 1998, 2022; Ollitrault et al., 2020). Our evidences from comparative genetic mapping are fully in agreement with the complete absence of large inter-chromosomal rearrangement between six species of the true citrus (*C. maxima, C. reticulata, C. medica, C. mangshanensis, C. trifoliata* and *C. australasica*) revealed by chromosome-specific painting (He et al., 2020). Therefore, the genome structure of the true citrus species appears to be globally highly conserved.

Taking advantage of this high synteny and collinearity, we analyzed the distribution of non-Mendelian segregation and the recombination landscape along the genome using the anchorage of the different genetic maps in the same reference (i.e., *C. clementina* v1.0 used for the genotype calling from GBS data). The genetic map size of the eight maps with good global coverage of the *C. clementina* v1.0 genome assembly (excluding *C. medica*) varies respectively between 852.3 and 1,099.7 cM for *C. glauca* and *C. australis* x *C. inodora*. In tomato, it has been proposed that sequence divergence at the interspecific level has an inhibitory effect on sexual recombination (Liharska et al., 1996; Chetelat et al., 2000). The map size variations observed between our different parents, which include pure species, F1 interspecific hybrids and complex admixture genomes, do not reveal such an effect in citrus. The recombination landscape along the genome was analyzed for the eight maps with good global coverage of the *C. clementina* v1.0 genome assembly. This landscape was similar

and highlighted low/no recombination regions in genomic areas with very low gene density of chr1, chr2, chr3, chr4 and chr5, where centromeres were located by half tetrad analysis (Aleza et al., 2015). Very large regions of chr6, chr8 and chr9 with very low recombination levels were also revealed. These last regions include the centromere location (Aleza et al., 2015) but also encompass numerous gene sequences. Chr7 is the only one displaying recombination all throughout. For sexual breeding schemes, the revealed pattern of recombination will be very useful to optimize the management of multi-loci selection of genes located on the same chromosome.

Deviations from Mendelian segregation have been frequently described in citrus, particularly for the male parent markers (Carlos de Oliveira et al., 2007; Bernet et al., 2010; Ollitrault et al., 2012, 2021; Yu et al., 2016), probably due to pollen competition (Ollitrault et al., 2012). We also noticed the higher rates of skewed segregations for the male parents *C. australis* x *C. inodora* and *C. maxima* x *C. reticulata*. However, the female *C.* x *lemon* parent displays the third highest rate of skewed markers. The lower rates of non-Mendelian segregation were observed for several pure species (*C. maxima* cv Chandler, *C. reticulata* cv Cleopatra and *C. trifoliata*) used as male or female parents. Therefore, it is probable that various factors (e.g., sex, phylogenomic structure, parental combination) may affect Mendelian segregation. Recessive unfavorable mutations can result in strong gametophytic selection and therefore important deviation from Mendelian segregations. For *C. maxima* x *C. reticulata* (Pink pummelo x Tardia mandarin), the direct relation between the extent of skewed regions and recombination landscape is a good illustration of the impact of the variation of recombination rates along the genome and the extent of linkage drag. In the case of Corsican citron (*C. medica*) used as a pollinator, we observed a quasi-complete elimination of one haplotype at the beginning of chr7, where one S locus is located for gametophytic incompatibility in citrus, based on a S-RNase system (Liang et al., 2020; Ollitrault et al., 2021). This skewed segregation is similar to those occurring in reciprocal crosses between two self-incompatible varieties (Fortune mandarin and Ellendale tangor) sharing one incompatible allele at this S locus (Ollitrault et al., 2021). This suggests that Corsican citron may share with Chandler pummelo (used as the female parent of the progenies used for Corsican citron mapping) one self-incompatibility allele at the pollen-specific S-locus F-box (SLF) gene.

The role of structural variants in reproductive isolation between species is an ongoing discussion (Zhang et al., 2021; Berdan et al., 2024), but several biological models highlight the strong negative impact of large structural heterozygosity in male and female fertility (Noor et al., 2001; Baurens et al., 2019). Therefore, the large sexual compatibility between species of the *Citrus* genus whose reticulation occured 6-8 Ma ago (Wu et al., 2018), as well as the fertility of their interspecific hybrids, can be explained by a highly conserved genomic structure. Large structural variations affecting male and female fertility have, however, been reported in some specific varieties (Ollitrault et al., 2008). It is, for example, the case in the reciprocal translocation in sweet orange cv Valencia, as evidenced by cytogenetic (Iwamasa, M., n.d.) and genome sequencing (Wu et al., 2022). This structural heterozygosity does not result from a structural differentiation between constitutive ancestral species (*C. maxima* and *C. reticulata*) but is due to a mitotic mutational event on a standard sweet orange.

The anchoring of individual and consensus genetic maps on chromosome-scale genome assemblies reveals some discrepancies

The evolution of NGS and particularly of long reads sequencing greatly improved the quality of *de novo* whole genome assembly, and recent publications propose citrus haplotype chromosome-scale assemblies from Nanopore or PacBio HiFi sequencing sometimes coupled with Hi-C sequencing data (Guardo et al., 2021; Wu et al., 2022; Bao et al., 2023; Nakandala et al., 2023, 2024). Our consensus genetic map which is highly syntenic with all individual genetic maps was anchored on 15 chromosome-scale assemblies to identify discrepancies between the consensus map and physical assemblies and therefore between assemblies.

The chromosome numbering and orientation of our consensus genetic map are the same as the first published high density citrus genetic map (Ollitrault et al., 2012) that was adopted for the *C. clementina* v1.0 chromosome assembly, published by the international citrus genome consortium (Wu et al., 2014). The anchoring of the different pseudochromosome assemblies with our individual and consensus maps clarified the relative numbering and orientation of the different genome assemblies (Supplementary Table 4)

highlighting two mains ways of numbering, deriving from the first two pseudochromosome assemblies (Supplementary Table 4): the *C. sinensis* v1.0 (Xu et al., 2013) and the *C. clementina* v1.0 one (Wu et al., 2014). The chromosome orientations appear to be more diverse. This inconsistency in chromosome numbering and orientation may result in confusion regarding the location of useful genes and the integration of QTL information based on different reference genomes or genetic maps.

We analyzed the synteny and collinearity after numbering and orienting all genomes in the same way as in the consensus genetic map. For the same horticultural group, the most recent assembly was the most congruent with the consensus genetic map. This is the case when, for example, comparing the first sweet orange chromosome-level assembly (Xu et al., 2013), the sweet orange V3 assembly (Wang et al., 2021) and the more recent haplotype-resolved assembly of Valencia sweet orange (Wu et al., 2022). Numerous discrepancies were found for synteny with the first chromosome-scale genome assembly of *C. maxima* (Wang et al., 2017), whereas synteny and collinearity were high for the second *C. maxima* assembly (Lu et al., 2022), except for one inverted region of chr9 (consensus map numbering). Similarly, a strong increase in synteny and collinearity was observed between the first haplotype-resolved lemon genomes (Guardo et al., 2021) and the most recent one (Bao et al., 2023). The first lemon haplotype assemblies displayed numerous discrepancies shared with the first *C. maxima* one. It may be explained by the use of this *C. maxima* assembly to finish the chromosome-scale assembly of lemon haplotypes (Guardo et al., 2021).

Therefore, the most recent assemblies for *C. australis*, *C. hindsii, C. maxima, C. trifoliata,* sweet orange and lemon appear to be globally highly syntenic and collinear. The discrepancies observed by the anchoring of some genome assemblies with the genetic consensus map may reveal real, large structural variations compared with most citrus species, as well as misplaced regions during the assembly. For the Valencia sweet orange haplotype-resolved assembly (Wu et al., 2022), our results for the preferentially-mandarin haplotype (SwO-USA-A) are fully concordant with the reciprocal translocation revealed by early cytogenetic studies (Iwamasa, M. and Nito, N., 1988). Moreover, its localization between chr1 (LG1) and chr9 (LG9) is in agreement with the findings of molecular cytogenetic analysis, which localized the translocation between chr4 (LG1 of consensus genetic map) and chr9 (LG9) of the sweet orange V1 assembly (Song et al., 2023). Structural differences are observed between the two published assemblies for *C. trifoliata*. One is highly syntenic and collinear with the consensus map (Peng et al., 2020), whereas the other suggests a translocation of 5 Mb from chr7 (LG7) to chr5 (LG5). It may be interesting to validate the existence of this structural variability in *C. trifoliata* and analyze its potential phenotypical implication. The *C. australis* collapsed assembly is the only one with four telomeric regions (start of chr4 -LG1-, end of chr7 -LG4-, and start and end of chr8 -LG8) not anchored by the consensus genetic map. These genomic regions of the collapsed assembly do not fit well with the individual haplotype assemblies, as shown by the published dot plots between collapsed and haplotype assemblies (Nakandala et al., 2024). The new methodologies combining PacBio HiFi and Hi-C reads have opened the way for telomere-to-telomere gapless assemblies (Sun et al., 2024) and may explain why this very recent genome assembly displays additional telomeric regions when compared with previous citrus assemblies. It may be interesting to validate these identified regions of the collapsed *C. australis* assembly and to analyze if they contain genes not present in the other citrus species and their possible involvement in the resistance/tolerance to HLB of this species (Ramadugu et al., 2016; Alves et al., 2021). For *C. clementina* v1.0 (Wu et al., 2014), previous genetic mapping studies consistently suggested that some genomic regions were misplaced. This is confirmed by the present work with accurate localization (Figure 4.3). The concerned genomic regions are (i) from chr5 (12.78-19.40 Mb) and chr4 (4.81-5.84 MB) genetically mapped at the end of LG7; (ii) genomic regions of chr3 (34.30-35.77 Mb) and 9 (11.14-14.45 Mb) genetically mapped in the middle of LG8; (iii) a genomic region of chr2 (0.74-4.23 Mb) mapping at the beginning of LG4; and (iv) a small region of chr8 (16.30-16.71 Mb) mapping at the beginning of LG6. In addition, a misplaced and inverted region is revealed in chr3 (29.03-34.23 Mb). It may be noteworthy that most of these discrepancies (except for the misplaced inverted centromeric region of chr3) concern genomic areas that were not covered by the reference clementine genetic map (Ollitrault et al., 2012) used for the final assembly in pseudochromosomes. The *C. clementina* v1.0 assembly is still used as a reference for several genetic and genomic studies, and our information about the misplaced region may be important for a better interpretation of results.

Perspective for further genetic analysis and interspecific breeding projects

Consensus genetic maps allow researchers to overcome some of the limitations of individual genetic maps. The integration of multiple populations enhances the explored diversity and improves genome coverage thanks to the complementarity of the different families, as a region that is monomorphic within a population may be polymorphic in another [81-83]. In citrus, genomic studies revealed identity by descent for numerous genomic regions between mandarins and sweet oranges (Wu et al., 2018) resulting in full homozygosity of large genomic regions in modern mandarins and tangors (mandarin x sweet orange hybrids), as well as large gaps in the genetic maps of these horticultural groups. Our consensus map partially overcomes this problem with an average gap size of 0.36 cM. However, its anchorage in the genome assembly of *C. australis* reveals some uncovered telomeric regions. The consensus may be improved by the integration of additional genetic maps, by taking advantage of the gene-based marker approach.

The combination of data from multiple families also allows researchers to capture more recombination events and therefore increase the mapping resolution (Pootakham et al., 2015; Linge et al., 2018). It therefore improves the precision of QTL analysis and the search for candidate genes and regulatory elements. Our consensus genetic map with 10,756 loci, including 7,915 gene-based markers, will constitute a very useful framework to integrate the locations of QTLs identified from different segregant progenies or genome-wide association studies in various germplasms of the *Citrus* genus. Indeed, we have shown that the rate of successful anchoring on 14 different genomes covering a wide range of the *Citrus* genus diversity, of the 17,718 SNP probes (Supplementary Table 7) associated with the 10,756 markers of the consensus map, varies between 91.57% and 99.18%. We can therefore expect to easily infer the location in the consensus map of any QTL or candidate gene identified in the different chromosome-scale genome assemblies.

Consensus genetic maps have also been used to drive the final steps of chromosome-scale assembly (Muñoz-Amatriaín et al., 2011; Mohd Sanusi et al., 2023) and identify large structural genome variations (Khan et al., 2012). The genome structural conservation evidenced by comparative genetic mapping in the *Citrus* genus and anchorage of our consensus genetic map with 15 genomes covering a large diversity of the *Citrus* genus is a favorable situation for further interspecific sexual breeding as well as for translational genomics. It also justifies the use of a unique reference genome to provide a global view of the phylogenomic structure along the genome of modern citrus, as proposed from WGS resequencing [4,26] or GBS data [5,62]. However, it appears essential to select a high-quality genome assembly for an accurate determination of interspecific breaking points in the modern citrus genome; our consensus genetic map constitutes a good template to select such high-quality assemblies.

Individual maps remain essential to analyze the distribution of skewed segregation along the genome. It is particularly important in citrus because apomixis by polyembryony (Wang et al., 2017) is present in several species and horticultural groups. Indeed, apomixis has contributed to the accumulation of recessive hidden mutations in heterozygosity, as demonstrated in sweet orange (Wang et al., 2023). The unfavorable mutations can result in gametophytic selection and non-Mendelian segregation. As we observed for our *C. maxima* x *C. reticulata* parent, the linkage drag can extend the skewed segregation in large genomic areas when counter-selected genes are located in regions with low recombination rates. This can be puzzling when interpreting the genetic determinism of useful traits controlled by a gene located in such a region, if only phenotypic segregation data are considered.

Comparative analysis of the nine individual maps and their mapping to the *C. clementina* v1.0 genome assembly revealed a similar recombination landscape for the nine populations and highlighted several large genomic regions with very low recombination rates. This information will be essential in optimizing sexual breeding strategies because these large genomic regions dramatically increase the risk of linkage drag by reducing the probability of recombination between useful and unfavorable genes.

Conclusion

The GBS data of 1,216 hybrids from 10 bi-parental families using the Apek1 restriction enzyme were powerful in developing nine uni-parental high-density genetic maps encompassing five ancestral species of the *Citrus* genus, two horticultural groups resulting from interspecific admixture and two F1 interspecific hybrids. The predominance of the cutting site of Apek1 in gene sequences permitted the development of a consensus map based on common gene-based markers between the various individual maps. It spans 1,005.27 cM and includes 10,756 loci: 7,915 being gene-based markers and 2,841 SNPs located-out gene sequences. The synteny is complete between the consensus map and the individual maps, and their collinearity is very high. The comparative genetic mapping and anchoring of the consensus map on 15 published chromosome-scale genome assemblies highlighted the differences in numbering and orientation of chromosomes within different genome assemblies. Once these parameters were homogenized, the consensus genetic map appeared to be highly syntenic and collinear with the most recent genome assemblies, whereas discrepancies were observed for some older ones. These high synteny and collinearity concern the recent genome assemblies of *C. australis, C. hindsii, C. maxima, C. trifoliata, C.* x. *aurantium* var. *sinensis* and *C.* x. *lemon* var. *lemon*. The role of large structural variations during speciation within the *Citrus* genus seems to have been limited. It may explain the high level of sexual compatibility between most *Citrus* species and the frequent good fertility of the F1 interspecific hybrids. The recombination landscape also appears to be largely conserved between ancestral species and F1 interspecific hybrids. Large genomic regions with very low recombination rates have been identified. This information will be very useful to estimate the linkage drag extent and to optimize conventional breeding schemes. Non-Mendelian segregations are frequent in citrus and were observed in specific regions for each parental combination. They may complicate the interpretation of a useful traits' genetic determinants. Our consensus genetic map constitutes the most saturated genetic framework published in citrus, and its congruence with the most recent citrus genome assemblies validates its quality. It may be easily extended to other species and horticultural groups, taking advantage of the gene-based marker approach. The consensus genetic map is a useful tool to check the accuracy of genome assemblies, identify large structural variation that may have occurred within the horticultural group (such as that of Valencia sweet orange) and provide a focus to study potential relationships with phenotypic variations. It may also be a reference framework to integrate the positions of QTLs and useful genes identified from different segregant progenies or genome-wide association studies in various germplasms of the *Citrus* genus.

Tables

Table 1: Statistics of the nine individual genetic maps

	n	LG	1	2	3	4	5	6	7	8	9	all
C. maxima		Mk.	320	388	423	226	356	294	273	297	175	2752
		M.S.	99.61	97.31	124.31	79.98	81.55	69.68	81.57	102.17	102.27	838.46
	194	A.G.S.	0.93	0.89	0.94	0.98	0.77	0.82	0.95	1.01	1.38	0.96
		B.G.S.	4.13	4.67	3.25	4.02	2.58	2.58	4.65	3.13	5.35	5.35
		U. P.	108	110	133	83	105	84	87	102	75	887
C. reticulata		Mk.	214	233	272	118	169	169	169	147	161	1652
		M.S.	110.68	113.00	159.41	101.89	109.02	76.64	92.54	94.18	90.10	947.45
	127	A.G.S.	1.88	1.4	1.63	2.04	1.87	1.42	1.93	1.74	1.67	1.73
		B.G.S.	8.30	6.24	5.96	7.15	6.33	3.95	5.95	7.64	5.61	8.30
		U. P.	60	82	99	51	59	55	49	55	55	565
C. medica		Mk.	87	35	110+7	144	89	35+30	52	110	61	760
		M.S.	19.25	75.61	67.2 + 2.1	85.89	47.00	33.2+17.4	71.47	111.90	68.59	599.71
	203	A.G.S.	0.687	3.601	1.308	1.101	1.270	1.632	2.465	1.963	2.540	1.841
		B.G.S.	4.45	25.94	5.26	3.34	4.28	6.44	9.77	15.84	11.79	25.94
		U. P.	29	22	55	79	38	33	30	58	28	372
C. maxima x C. reticulata		Mk.	539	622	784	481	501	215	486	336	472	4436
		M.S.	106.68	130.36	179.81	77.72	96.82	54.98	96.14	110.94	73.18	926.63
	152	A.G.S.	0.79	0.81	1	0.93	0.85	1.45	0.9	1.32	0.81	0.98
		B.G.S.	2.63	2.63	10.64	3.30	5.21	9.96	4.62	8.09	3.29	10.64
		U. P.	136	161	180	85	115	39	108	85	91	1000
C. x aurantium var clementina		Mk.	92	114	198	125	131	106	55	84	82	987
		M.S.	115.99	132.43	193.98	115.40	155.03	91.39	86.31	131.34	75.63	1097.50
	187	A.G.S.	2.47	2.28	1.83	1.46	1.78	1.79	2.78	2.68	1.61	2.08
		B.G.S.	16.54	10.65	11.94	6.92	22.64	19.20	18.30	16.25	8.70	22.64
		U. P.	48	59	107	80	88	52	32	50	48	564
C. x limon var limon		Mk.	426	518	670	430	389	361	408	372	344	3918
		M.S.	88.08	85.31	171.51	109.76	117.77	71.66	109.41	111.43	66.43	931.36
	108	A.G.S.	1.01	1.09	1.08	1.14	1.09	1.17	1.14	1.43	1.13	1.14
		B.G.S.	4.65	3.71	3.71	5.23	3.71	4.65	5.17	5.63	2.81	5.63
		U. P.	88	79	160	97	109	62	97	79	60	831
C. trifoliata		Mk.	166	193	282	149	182	104	128	137	158	1499
		M.S.	141.82	138.53	191.68	116.62	129.80	93.52	118.07	47.44+54.72	118.22	1 150.44
	68	A.G.S.	2.68	1.92	1.92	2.16	2.09	2.75	2.41	1.33	2.23	2.17
		B.G.S.	10.45	5.91	5.95	5.91	7.41	11.53	8.92	4.43	6.78	11.53
		U. P.	54	73	101	55	63	35	50	42	54	527
C. glauca		Mk.	175	248	318	230	195	139	221	160	160	1846
		M.S.	96.55	95.67	156.47	78.74	99.44	82.86	73.41	88.06	81.51	852.71
	173	A.G.S.	1.18	0.99	1.09	0.89	1.18	1.34	0.86	1.38	1.24	1.13
		B.G.S.	4.35	4.06	6.21	3.25	13.67	5.61	5.50	5.17	3.47	13.67
		U. P.	83	98	144	89	85	63	86	65	67	780
C. australis x C. inodora		Mk.	211	244	367	250	202	190	181	188	186	2019
		M.S.	108.12	142.73	164.93	137.85	95.32	95.15	114.00	112.35	129.26	1099.71
	171	A.G.S.	1.3	1.41	1.17	1.21	1.27	1.15	1.52	1.56	1.7	1.37
		B.G.S.	3.84	3.96	3.63	3.67	7.17	4.15	5.86	4.46	6.59	7.17
		U. P.	84	102	142	115	76	84	76	73	77	829

n: number of hybrids, LG: linkage groups, Mk.: number of markers, M.S.: Map size, A.G.S.: Average gap size, B.G.S.: Biggest gap size, U.P.: Unique positions

787 **Table 2:** Skewed segregations
788

	Chr.	1	2	3	4	5	6	7	8	9	Total
C. reticulata	Sk. nb	1	1	14	0	1	5	89	1	5	117
	%	0.5%	0.4%	5.1%	0.0%	0.6%	3.0%	52.7%	0.7%	3.1%	7.1%
C. maxima	Sk. nb	0	0	0	0	0	0	0	0	0	0
	%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
C. medica	Sk. nb	87	6	4	1	1	13	34	10	1	157
	%	100.0%	17.1%	3.4%	0.7%	1.1%	20.0%	65.4%	9.1%	1.6%	20.7%
Clementine	Sk. nb	12	2	11	1	5	1	2	2	6	42
	%	13.0%	1.8%	5.6%	0.8%	3.8%	0.9%	3.6%	2.4%	7.3%	4.3%
C. x limon	Sk. nb	136	7	8	66	374	8	5	245	0	849
	%	31.9%	1.4%	1.2%	15.3%	96.1%	2.2%	1.2%	65.9%	0.0%	21.7%
C. maxima x C. reticulata	Sk. nb	508	267	453	481	276	55	260	7	469	2776
	%	94.6%	43.0%	57.9%	100.0%	55.1%	25.6%	53.5%	2.1%	99.4%	62.6%
C. trifoliata	Sk. nb	22	0	57	0	3	39	2	70	1	194
	%	13.3%	0.0%	20.2%	0.0%	1.6%	37.5%	1.6%	51.1%	0.6%	12.9%
C. australis x C. inodora	Sk. nb	102	95	141	37	83	19	64	64	23	628
	%	48.3%	38.9%	38.4%	14.8%	41.1%	10.0%	35.4%	34.0%	12.4%	31.1%
C. glauca	Sk. nb	0	13	30	4	0	1	7	0	84	139
	%	0.0%	5.2%	9.4%	1.7%	0.0%	0.7%	3.2%	0.0%	52.5%	7.5%

789 Sk.nb: number of significantly skewed markers; %: % of significantly skewed markers
790
791
792

Table 3: Synteny and collinearity between some genome assemblies and the related genetic maps

Genome	Genetic map	nb markers	Synteny	Global Col.
C. clementina V1.0	*C. x aurantium var clementina*	987	0.997	0.935+/-0.054
Pum-China-V1	*C. maxima*	2647	0.887	0.998+/-0.000
Pum-China-V2	*C. maxima*	2571	0.987	0.932+/-0.105
	C. maxima	2324	0.909	0.983+/-0.022
Swo-China-V1	*C. reticulata*	1384	0.909	0.989+/-0.009
	C. maxima x C. reticulata	3696	0.921	0.969+/-0.035
	C. maxima	2705	0.984	0.998+/-0.000
Swo-China-V3	*C. reticulata*	1636	0.983	0.959+/-0.016
	C. maxima x C. reticulata	4395	0.995	0.975+/-0.020
	C. maxima	2697	0.898	0.998+/-0.001
SWO-USA-A	*C. reticulata*	1620	0.885	0.998+/-0.000
	C. maxima x C. reticulata	4363	0.870	0.996+/-0.002
	C. maxima	2709	0.985	0.999+/-0.000
SWO-USA-B	*C. reticulata*	1631	0.986	0.998+/-0.001
	C. maxima x C. reticulata	4393	0.995	0.997+/-0.002
Lemon-It-prim	*C. x limon*	3540	0.889	0.839+/-0.124
Lemon-It-alt	*C. x limon*	3448	0.895	0.848+/-0.129
Lemon-China-HapB	*C. x limon*	3880	0.996	0.997+/-0.001
Lemon-China-HapA	*C. x limon*	3876	0.994	0.996+/-0.001
TrO-USA	*C. trifoliata*	1484	0.991	0.969+/-0.014
TrO-China	*C. trifoliata*	1399	0.976	0.990+/-0.003
C. australis	*C. australis x C.inodora*	1987	0.993	0.998+/-0.001

Nb markers: number of markers linking the considered genome and genetic map; Global Cl.: average collinearity over the nine chromosomes.

807 **Table 4 :** Synteny and average collinearity over the nine chromosomes between Pink x Tardia genetic map and
808 the height other genetic maps
809

	Nb Mark.	Synteny	Collinearity
C. maxima	655	0.995	0.995+/-0.003
C. reticulata	478	0.998	0.996+/-0.002
C. medica	164	1.000	0.993+/-0.002
C. x aurantium var. clementina	636	0.998	0.994+/-0.005
C. x limon var. limon	1328	0.999	0.992+/-0.009
C. trifoliata	559	0.995	0.993+/-0.004
C. glauca	703	0.994	0.992+/-0.004
C. australis x C. inodora	795	0.992	0.996+/-0.003

810 Nb Mark.: number of markers linking the Pink x Tardia and the other genetic maps.
811

812 **Table 5:** statistics of the consensus map
813

LG	GM	NGM	TM	size (cM)	AGS	BGS	UP
1	873	292	1165	113.033	0.3	3.334	381
2	1062	330	1392	115.214	0.31	1.547	367
3	1404	457	1861	166.936	0.39	2.586	424
4	883	269	1152	104.76	0.4	3.714	263
5	851	394	1245	110.079	0.34	2.233	321
6	688	235	923	93.521	0.39	8.295	242
7	759	327	1086	98.811	0.34	3.138	290
8	670	280	950	112.813	0.38	2.592	297
9	725	257	982	90.098	0.41	1.63	223
all	7915	2841	10756	1005.265	0.36	8.295	2808

814 LG: linkage group; GM: gene markers; NGM: no gene markers; TM: total markers; AGS: average gap size; BGS:
815 biggest gap size; UP: unique positions
816
817
818
819

Table 6: statistics for the anchorage of the consensus genetic map on 15 genome assemblies

Genome	NAP	NGM	Synt.	Col.
Clementine-ICGC	17718	7840	0.976	0.997+/-0.002
SwO-China-V1	17117	6589	0.928	0.987+/-0.015
SwO-China-V3	17573	7803	0.998	0.998+/-0.001
SwO-USA-A	17472	7712	0.902	0.998+/-0.001 *
SwO-USA-B	17540	8086	0.998	0.998+/-0.001
Lemon-Italy-prim	16586	5938	0.890	0.870+/-0.100
Lemon-Italy-alt	16423	5884	0.895	0.874+/-0.100
Lemon-China-A	17556	7496	0.998	0.998+/-0.001
Lemon-China-B	17485	7438	0.998	0.995+/-0.003
TrO-USA	17321	7766	0.998	0.998+/-0.001
TrO-China	16224	5996	0.985	0.996+/-0.003
Kumquat-China	17197	7555	0.998	0.998+/-0.001
C.australis- Australia	17372	7166	0.996	0.998+/-0.001
Pumello-ChinaV1	17455	7401	0.906	0.997+/-0.001
Pumello-ChinaV2	16979	6855	0.998	0.939+/-0.111

NAP: Number of anchored probes; NGM: number of gene markers; Synt: Synteny; Col.: average collinearity over the nine chromosomes; * estimated for chromosome 2 to 8

Figure captions

Figure 1: phylogenetic organization of the Citrus genus and location of the accessions considered in this paper for genetic mapping (M) and genome assemblies (G). The phylogenetic tree was constructed by Minimum Evolution (ME) approach (Desper and Gascuel, 2002) and is based on 181207 diallelic SNPs identified from variant calling of WGS data on the *C. clementina* V1.0 reference genome after filtration for a minimum distance of 1kb between successive SNPs (our unpublished data). The Asian radiation was estimated by Wu et al. (2018) to have occured in the late Miocene (6–8 Ma) and Oceanian radiation during the early Pliocene epoch (around 4 Ma).

Figure 2: Link between the nine genetic maps and the *C. clementina* v1.0 genome assembly
Chr: *C. clementina* chromosomes; LG: linkage groups of the genetic maps
1 (up /left): *C. maxima*; 2 (up/center): *C. reticulata*; 3 (up/right): *C. medica*; 4 (center/left): *C. x aurantium var clementina*; 5 (center/center): *C. limon var limon*; 6 (center/right): *C. maxima x C. reticulata*; 7 (down/left): *C. trifoliata*; 8 (down/center): *C: glauca*; 9 (down/right): *C. australis x C. inodora*

Figure 3: Gene and marker densities, recombination landscape and Non-Mendelian segregations of the *C. maxima x C. reticulata* parent.
External Outer ring: Pink histogram: gene density (scale 0-50%), blue line local recombination (scale 0-20 cM/Mb; purple dot: number of markers (scale: 0-5/100kb); inner ring: red line Threshold for Qvalue (0.05) significance; brown dot: Qvalue for Mendelian segregation (scale: 0-15)

Figure 4: Link between the Consensus genetic map and genome assemblies.
1 (up/left): SwO-China-V1; 2 (up/right): SwO-China-V3; 1 (down/left): SwO-USA-HapA; 4 (down/right): SwO-USA-HapB
LG: linkage groups of the consensus genetic map; Chr: chromosome of the genome assembly

Figure 5: Link between the Consensus genetic map and genome assemblies.

1 (up/left): Pum-China-V1; 2 (up/right): Pum-China-V2; 1 (down/left): *C. clementina* V1.0; 4 (down/right): *C. hindsii*
LG: linkage groups of the consensus genetic map; Chr: chromosome of the genome assembly

Figure 6: Link between the Consensus genetic map and genome assemblies.
1 (up/left): Lemon-It-prim; 2 (up/right): Lemon-It-alt; 1 (down/left): Lemon-China-HapA; 4 (down/right): Lemon-China-HapB
LG: linkage groups of the consensus genetic map; Chr: chromosome of the genome assembly

Figure 7: Link between the Consensus genetic map and genome assemblies.
1 (up/left): TrO-USA; 2 (up/right): TrO-China; 3 (down): *C. australis*
LG: linkage groups of the consensus genetic map; Chr: chromosome of the genome assembly

Supplementary information

Supplementary figure 1: Marey map of the genetic maps relative to C. clementina V1.0 genome assembly
Supplementary Figure 1a: Marey Map of the nine genetic maps relative to C. clementina V1.0 genome assembly
X axis physical position on C. clementina V1.0 reference genome ; Y axis: genetic position (cM)
1: *C. maxima*, 2: *C. reticulata*, 3: *C. medica*, 4: *C. x aurantium var clementina*, 5: *C. x limon*, 6: C. maxima x C. reticulata, 7: *C. trifoliata*, 8: *C. glauca*, 9: *C. australis x C. inodora*
Supplementary figure 1b: Marey Map of the *C. maxima x C. reticulata* genetic map, relative to C. clementina V1.0 genome assembly
X axis physical position on *C. clementina* V1.0 reference genome ; Y axis: genetic position (cM)
Red ellipse discreapencies with the *C. clementina* V1.0 assembly shared with others genetic maps

Supplementary Figure 2: Gene density, marker density, recombination landscape, and squewed segregation;
External Outer ring: Pink histogram: gene density (scale 0-50%), blue line local recombination (scale 0-30 cM/Mb; purple dot: number of markers (scale: 0-5/100kb); inner ring: red line Threshold for Qvalue (0.05) significance; brown dot: Qvalue for Mendelian segregation (scale: 0-15)

Supplementary Figure 3: Link between genetic maps and related genome assemblies.
Supplementary Figure 3a
1: *C. maxima* map / Pum-China-V1; 2: C. *maxima* map / Pum-China-V2
3*: C. trifoliata* map / TOr-USA; 4: *C. trifoliata* map / Tor-China
Supplementary Figure 3b: Link between genetic maps and related genome assemblies.
1: *C. x limon map* / Lemon-It-Prim; 2: : *C. x limon map* / Lemon-It-Alt
3: C. x limon map / Lemon-China-HapA; 4: *C. x limon* map / Lemon-China-HapB
Supplementary Figure 3c: Link between genetic maps and related genome assemblies.
1: *C. maxima* map / SwO-China-V1; 2: : *C. maxima* map / SwO-China-V3
3: *C. maxima* map / SwO-USA-A; 4: *C. maxima* map / SwO-USA-B
Supplementary Figure 3d: Link between genetic maps and related genome assemblies.
1: *C. reticulata* map / SwO-China-V1; 2: : *C. reticulata* map / SwO-China-V3
3: *C. reticulata* map/ SwO-USA-A; 4: *C. reticulata* map/ SwO-USA-B
Supplementary Figure 3e: Link between genetic maps and related genome assemblies.
1: *C. maxima x C. reticulata* / SwO-China-V1; 2: *C. maxima x C. reticulata* / SwO-China-V3
3: *C. maxima x C. reticulata* / SwO-USA-A; 4: *C. maxima x C. reticulata* / SwO-USA-B
Supplementary Figure 3f: Link between genetic maps and related genome assemblies.
C. australis x C. inodora genetic map / *C. australis* assembly

Supplementary Figure 4: Link between Pink x Tardia genetic maps and the height other maps

1: *C. maxima*; 2: *C. reticulata*; 3: *C. medica*; 4: *C. x aurantium var clementina,* 5: *C. x limon*; 6: *C. trifoliata*; 7: *C. australis x C. inodora*; 8: *C. glauca*

Supplementary Table 1: Populations and sequencing methods used to establish nine parental genetic maps

Supplementary Table 2: Genetic positions, physical positions on the *C. clementina* V1.0 genome assembly and corresponding genes of the markers of the nine genetic maps

Supplementary Table 3: Synteny of the nine genetic map with the *C. clementina* V1.0 assembly

Supplementary Table 4: Collinearity of the nine genetic map with the *C. clementina* V1.0 assembly (Spearman coefficien)

Supplementary Table 5: Correspondences of chromosome numbering and orientation ("+"= identical;" -"= inverted) between genetic maps (individual and consensus) and 15 chromosome scale genome assemblies

Supplementary table 6: Synteny and collinearity of individual genetic maps with related genome assemblies

Supplementary Table 7: Detailed information on the consensus genetic map; genetic and physical positions on the *C. clementina* V1.0 assembly

Supplementary Table 8: Number of shared markers and collinearity between the consensus map and the individual genetic maps

Supplementary Table 9: Detailed information for probe sequences used to anchor the consensus genetic map on the different genome assemblies

Supplementary Table 10: Fasta file of the probes to anchor the consensus genetic map on different genome assemblies

Supplementary Table 11: Anchorage of the consensus genetic map on 15 genome assemblies

Supplementary Table 12: Collinearity estimated by the Spearman coefficient between the consensus genetic map and 15 chromosome-scale genome assemblies

Conflict of Interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

P.O. conceived and coordinated the experiment, analysed the data and write the paper, F.C., Y.F., M.A., M.F., M.S., P.A. and F.L. produced the mapping progenies, A.P., P.M. and M.M. produced the GBS data, B.H., F.C., G.C. contributed to the data analysis, B.H., F.C., M.S. and N.W. contributed to the writing/review of the

manuscript, P.O., P.A., N.W., L.P., R.M. and A.L. coordinated the different projects that funded the activity. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by France Génomique (ANR-10-INBS-09–08; “Dynamo” project), European Feder-Guadeloupe region “Cavalbio” project, H2020 Innovation Action Program (grant #817526, preHLB project), French National Research Agency (grant ANR-23-CE20-0038), Ministry of Science and Innovation Spain (project PID2021-1281150R-I00), MCIN/AEI/10.13039/501100011033/FEDER, UE., and Hort. Innovation, Australia.

Data Availability statement

The demultiplexed raw data of Genotyping By Sequencing (GBS) have been deposited at NCBI under BioProject numbers: PRJNA1119914, PRJNA1120630, PRJNA1120237, PRJNA1120896, PRJNA1124254, PRJNA1125292, PRJNA1125912, PRJNA1128137, PRJNA1128404. Detailed data about individual and consensus genetic maps and their anchoring on 15 published genome assemblies (including probe sequence of the consensus genetic map) are provided as supplementary material.

References

Abed, A., Badea, A., Beattie, A., Khanal, R., Tucker, J., and Belzile, F. (2022). A high-resolution consensus linkage map for barley based on GBS-derived genotypes. *Genome* 65, 83–94. doi: 10.1139/gen-2021-0055

Ahmed, D., Comte, A., Curk, F., Costantino, G., Luro, F., Dereeper, A., et al. (2019). Genotyping by sequencing can reveal the complex mosaic genomes in gene pools resulting from reticulate evolution: a case study in diploid and polyploid citrus. *Ann. Bot.* 123, 1231–1251. doi: 10.1093/aob/mcz029

Albrecht, U., and Bowman, K. D. (2011). Tolerance of the Trifoliate Citrus Hybrid US-897 (Citrus reticulata Blanco × Poncirus trifoliata L. Raf.) to Huanglongbing. *HortScience* 46, 16–22. doi: 10.21273/HORTSCI.46.1.16

Aleza, P., Cuenca, J., Hernández, M., Juárez, J., Navarro, L., and Ollitrault, P. (2015). Genetic mapping of centromeres in the nine Citrus clementina chromosomes using half-tetrad analysis and recombination patterns in unreduced and haploid gametes. *BMC Plant Biol.* 15, 80. doi: 10.1186/s12870-015-0464-y

Alves, M. N., Lopes, S. A., Raiol-Junior, L. L., Wulff, N. A., Girardi, E. A., Ollitrault, P., et al. (2021). Resistance to 'Candidatus Liberibacter asiaticus,' the Huanglongbing Associated Bacterium, in Sexually and/or Graft-Compatible Citrus Relatives. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.617664

Bao, Y., Zeng, Z., Yao, W., Chen, X., Jiang, M., Sehrish, A., et al. (2023). A gap-free and haplotype-resolved lemon genome provides insights into flavor synthesis and huanglongbing (HLB) tolerance. *Hortic Res* 10, uhad020. doi: 10.1093/hr/uhad020

Baurens, F.-C., Martin, G., Hervouet, C., Salmon, F., Yohomé, D., Ricci, S., et al. (2019). Recombination and Large Structural Variations Shape Interspecific Edible Bananas Genomes. *Molecular Biology and Evolution* 36, 97–111. doi: 10.1093/molbev/msy199

Bayer, R. J., Mabberley, D. J., Morton, C., Miller, C. H., Sharma, I. K., Pfeil, B. E., et al. (2009). A molecular phylogeny of the orange subfamily(Rutaceae: Aurantioideae) using nine cpDNA sequences. *American Journal of Botany* 96, 668–685. doi: 10.3732/ajb.0800341

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57, 289–300.

Berdan, E. L., Aubier, T. G., Cozzolino, S., Faria, R., Feder, J. L., Giménez, M. D., et al. (2024). Structural Variants and Speciation: Multiple Processes at Play. *Cold Spring Harb Perspect Biol* 16, a041446. doi: 10.1101/cshperspect.a041446

Bernet, G., Fernandez-Ribacoba, J., Carbonell, E., and Asins, M. (2010). Comparative genome-wide segregation analysis and map construction using a reciprocal cross design to facilitate citrus germplasm utilization. *Mol.Breed.*, 659–673. doi: 10.1007/s11032-009-9363-y

Boulton, A., Myers, R. S., and Redfield, R. J. (1997). The hotspot conversion paradox and the evolution of meiotic recombination. *Proceedings of the National Academy of Sciences* 94, 8058–8063. doi: 10.1073/pnas.94.15.8058

Carbonell-Caballero, J., Alonso, R., Ibañez, V., Terol, J., Talon, M., and Dopazo, J. (2015). A Phylogenetic Analysis of 34 Chloroplast Genomes Elucidates the Relationships between Wild and Domestic Species within the Genus Citrus. *Molecular Biology and Evolution* 32, 2015–2035. doi: 10.1093/molbev/msv082

Carlos de Oliveira, A., Bastianel, M., Cristofani-Yaly, M., Morais do Amaral, A., and Machado, M. A. (2007). Development of genetic maps of the citrus varieties "Murcott" tangor and "Pera" sweet orange by using fluorescent AFLP markers. *J.Appl.Genet.* 48, 219–231.

Chen, C. X., Bowman, K. D., Choi, Y. A., Dang, P. M., Rao, M. N., Huang, S., et al. (2008). EST-SSR genetic maps for Citrus sinensis and Poncirus trifoliata. *Tree Genetics and Genomes* 4, 1–10.

Chetelat, R. T., Meglic, V., and Cisneros, P. (2000). A genetic map of tomato based on BC(1) Lycopersicon esculentum x Solanum lycopersicoides reveals overall synteny but suppressed recombination between these homeologous genomes. *Genetics* 154, 857–867.
Curk, F., Ollitrault, F., Garcia-Lor, A., Luro, F., Navarro, L., and Ollitrault, P. (2016). Phylogenetic origin of limes and lemons revealed by cytoplasmic and nuclear markers. *Annals of botany* 117, 565–583. doi: 10.1093/aob/mcw005 [doi]
Curtolo, M., Cristofani-Yaly, M., Gazaffi, R., Takita, M. A., Figueira, A., and Machado, M. A. (2017a). QTL mapping for fruit quality in Citrus using DArTseq markers. *BMC genomics* 18, 289-017-3629–2. doi: 10.1186/s12864-017-3629-2 [doi]
Curtolo, M., Soratto, T. A. T., Gazaffi, R., Takita, M. A., Machado, M. A., and Cristofani-Yaly, M. (2017b). High-density linkage maps for Citrus sunki and Poncirus trifoliata using DArTseq markers. *Tree Genetics & Genomes* 14, 5. doi: 10.1007/s11295-017-1218-9
Desper, R., and Gascuel, O. (2002). Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *J Comput Biol* 9, 687–705. doi: 10.1089/106652702761034136
Dumont, B. L., and Payseur, B. A. (2011). Genetic Analysis of Genome-Scale Recombination Rate Evolution in House Mice. *PLOS Genetics* 7, e1002116. doi: 10.1371/journal.pgen.1002116
Ellegren, H., and Galtier, N. (2016). Determinants of genetic diversity. *Nat Rev Genet* 17, 422–433. doi: 10.1038/nrg.2016.58
Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., et al. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE* 6, e19379. doi:10.1371/journal.pone.0019379-doi:10.1371/journal.pone.0019379.
Endelman, J. B., and Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics* 30, 1623–1624. doi: 10.1093/bioinformatics/btu091
Fallah, M., Jean, M., Boucher St-Amour, V.-T., O'Donoughue, L., and Belzile, F. (2022). The construction of a high-density consensus genetic map for soybean based on SNP markers derived from genotyping-by-sequencing. *Genome* 65, 413–425. doi: 10.1139/gen-2021-0054
Folimonova, S. Y., Robertson, C. J., Garnsey, S. M., Gowda, S., and Dawson, W. O. (2009). Examination of the Responses of Different Genotypes of Citrus to Huanglongbing (Citrus Greening) Under Different Conditions. *Phytopathology*® 99, 1346–1354. doi: 10.1094/PHYTO-99-12-1346
Garcia-Lor, A., Curk, F., Luro, F., Navarro, L., and Ollitrault, P. (2011). Nuclear and maternal phylogeny within Citrus and related genera based on nuclear gene sequence SNPs and mitochondrial indels., (. Amsterdam the Netherlands).
George, J., and Lapointe, S. L. (2019). Host-plant resistance associated with Poncirus trifoliata influence oviposition, development and adult emergence of Diaphorina citri (Hemiptera: Liviidae). *Pest Management Science* 75, 279–285. doi: 10.1002/ps.5113
Guardo, M. D., Moretto, M., Moser, M., Catalano, C., Troggio, M., Deng, Z., et al. (2021). The haplotype-resolved reference genome of lemon (Citrus limon L. Burm f.). *Tree Genetics & Genomes* 17, 46. doi: 10.1007/s11295-021-01528-5
Guo, F., Yu, H., Tang, Z., Jiang, X., Wang, L., Wang, X., et al. (2015). Construction of a SNP-based high-density genetic map for pummelo using RAD sequencing. *Tree Genetics & Genomes* 11, 2. doi: 10.1007/s11295-014-0831-0
Harper, S. J., Dawson, T. E., and Pearson, M. N. (2010). Isolates of Citrus tristeza virus that overcome Poncirus trifoliata resistance comprise a novel strain. *Arch Virol* 155, 471–480. doi: 10.1007/s00705-010-0604-5
He, L., Zhao, H., He, J., Yang, Z., Guan, B., Chen, K., et al. (2020). Extraordinarily conserved chromosomal synteny of Citrus species revealed by chromosome-specific painting. *The Plant Journal* 103, 2225–2235. doi: 10.1111/tpj.14894
Herten, K., Hestand, M. S., Vermeesch, J. R., and Van Houdt, J. K. (2015). GBSX: a toolkit for experimental design and demultiplexing genotyping by sequencing experiments. *BMC Bioinformatics* 16, 73. doi: 10.1186/s12859-015-0514-3

Hill, W. G., and Robertson, A. (2007). The effect of linkage on limits to artificial selection. *Genetics Research* 89, 311–336. doi: 10.1017/S001667230800949X
Huang, M., Roose, M. L., Yu, Q., Du, D., Yu, Y., Zhang, Y., et al. (2018). Construction of High-Density Genetic Maps and Detection of QTLs Associated With Huanglongbing Tolerance in Citrus. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01694
Huang, Y., Xu, Y., Jiang, X., Yu, H., Jia, H., Tan, C., et al. (2021). Genome of a citrus rootstock and global DNA demethylation caused by heterografting. *Hortic Res* 8, 69. doi: 10.1038/s41438-021-00505-2
Iwamasa, M. (1970). Chromosome aberrations in citrus in relation to sterility and seedlessness. 1st International Citrus Symposium, Riverside. 175-181.
Iwamasa, M. and Nito, N. (1988). Cytogenetics and the evolution of modern cultivated Citrus., in *Proceedings of the sixth International Citrus Congress*, (Tel Aviv, Israel: Margraf Scientific Books, Weikersheim), 265–275.
Kawakami, T., Mugal, C. F., Suh, A., Nater, A., Burri, R., Smeds, L., et al. (2017). Whole-genome patterns of linkage disequilibrium across flycatcher populations clarify the causes and consequences of fine-scale recombination rate variation in birds. *Molecular Ecology* 26, 4158–4172. doi: 10.1111/mec.14197
Khan, M. A., Han, Y., Zhao, Y. F., Troggio, M., and Korban, S. S. (2012). A Multi-Population Consensus Genetic Map Reveals Inconsistent Marker Order among Maps Likely Attributed to Structural Variations in the Apple Genome. *PLOS ONE* 7, e47864. doi: 10.1371/journal.pone.0047864
Kong, A., Thorleifsson, G., Gudbjartsson, D. F., Masson, G., Sigurdsson, A., Jonasdottir, A., et al. (2010). Fine-scale recombination rate differences between sexes, populations and individuals. *Nature* 467, 1099–1103. doi: 10.1038/nature09525
Krueger, R. R., and Navarro, L. (2007). “Citrus germplasm resources,” (Wallingford, UK: CAB International), 45–140.
Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., et al. (2009). Circos: an information aesthetic for comparative genomics. *Genome Res* 19, 1639–1645. doi: 10.1101/gr.092759.109
Liang, M., Cao, Z., Zhu, A., Liu, Y., Tao, M., Yang, H., et al. (2020). Evolution of self-compatibility by a mutant S m -RNase in citrus. *Nature Plants* 6, 131–142. doi: 10.1038/s41477-020-0597-3
Liharska, T., Wordragen, M., Kammen, A., Zabel, P., and Koornneef, M. (1996). Tomato chromosome 6: effect of alien chromosomal segments on recombinant frequencies. *Genome* 39, 485–491.
Linge, C. da S., Antanaviciute, L., Abdelghafar, A., Arús, P., Bassi, D., Rossini, L., et al. (2018). High-density multi-population consensus genetic linkage map for peach. *PLOS ONE* 13, e0207724. doi: 10.1371/journal.pone.0207724
Lu, Z., Huang, Y., Mao, S., Wu, F., Liu, Y., Mao, X., et al. (2022). The high-quality genome of pummelo provides insights into the tissue-specific regulation of citric acid and anthocyanin during domestication. *Hortic Res* 9, uhac175. doi: 10.1093/hr/uhac175
Luro, F., Venturini, N., Costantino, G., Paolini, J., Ollitrault, P., and Costa, J. (2012). Genetic and chemical diversity of citron (Citrus medica L.) based on nuclear and cytoplasmic markers and leaf essential oil composition. *Phytochemistry* 77, 186–196. doi: 10.1016/j.phytochem.2011.12.013
Mabberley, D. (1998). Australian Citreae with notes on other Aurantioideae (Rutaceae). doi: 10.7751/TELOPEA19982004
Mabberley, D. J. (2022). A classification for edible citrus: an update, with a note on Murraya (Rutaceae). *Telopea* 25, 271–284. doi: 10.7751/telopea15954
Martin, G., Baurens, F.-C., Droc, G., Rouard, M., Cenci, A., Kilian, A., et al. (2016). Improvement of the banana “Musa acuminata” reference sequence using NGS data and semi-automated bioinformatics methods. *BMC Genomics* 17, 243. doi: 10.1186/s12864-016-2579-4
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10–12. doi: 10.14806/ej.17.1.200

Maynard, J., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. *Genetics Research* 89, 391–403. doi: 10.1017/S0016672308009579
Mohd Sanusi, N. S. N., Rosli, R., Chan, K.-L., Halim, M. A. A., Ting, N.-C., Singh, R., et al. (2023). Integrated consensus genetic map and genomic scaffold re-ordering of oil palm (*Elaeis guineensis*) genome. *Computational Biology and Chemistry* 102, 107801. doi: 10.1016/j.compbiolchem.2022.107801
Muñoz-Amatriaín, M., Moscou, M. J., Bhat, P. R., Svensson, J. T., Bartoš, J., Suchánková, P., et al. (2011). An Improved Consensus Linkage Map of Barley Based on Flow-Sorted Chromosomes and Single Nucleotide Polymorphism Markers. *The Plant Genome* 4. doi: 10.3835/plantgenome2011.08.0023
Myers, S., Bowden, R., Tumian, A., Bontrop, R. E., Freeman, C., MacFie, T. S., et al. (2010). Drive Against Hotspot Motifs in Primates Implicates the PRDM9 Gene in Meiotic Recombination. *Science* 327, 876–879. doi: 10.1126/science.1182363
Nakandala, U., Furtado, A., Masouleh, A. K., Smith, M. W., Williams, D. C., and Henry, R. J. (2024). The genome of Citrus australasica reveals disease resistance and other species specific genes. *BMC Plant Biology* 24, 260. doi: 10.1186/s12870-024-04988-8
Nakandala, U., Masouleh, A. K., Smith, M. W., Furtado, A., Mason, P., Constantin, L., et al. (2023). Haplotype resolved chromosome level genome assembly of Citrus australis reveals disease resistance and other citrus specific genes. *Hortic Res* 10, uhad058. doi: 10.1093/hr/uhad058
Noor, M. A., Grams, K. L., Bertucci, L. A., and Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci U S A* 98, 12084–12088. doi: 10.1073/pnas.221274498
Ollitrault, P., Ahmed, D., Costantino, G., Evrard, J.-C., Cardi, C., Mournet, P., et al. (2021). Segregation Distortion for Male Parents in High Density Genetic Maps from Reciprocal Crosses between Two Self-Incompatible Cultivars Confirms a Gametophytic System for Self-Incompatibility in Citrus. *Agriculture* 11, 379. doi: 10.3390/agriculture11050379
Ollitrault, P., Curk, F., and Krueger, R. (2020). "Chapter 4 - Citrus taxonomy," in *The Genus Citrus*, eds. M. Talon, M. Caruso, and F. G. Gmitter (Woodhead Publishing), 57–81. doi: 10.1016/B978-0-12-812163-4.00004-8
Ollitrault, P., Dambier, D., Luro, F., and Froelicher, Y. (2008). Ploidy manipulation for breeding seedless triploid citrus. *Plant Breed.Rev.* 30, 323–352.
Ollitrault, P., Terol, J., Chen, C., Federici, C. T., Lotfy, S., Hippolyte, I., et al. (2012). A reference genetic map of C. clementina hort. ex Tan.; citrus evolution inferences from comparative mapping. *BMC Genomics* 13, 593. doi: 10.1186/1471-2164-13-593
Oueslati, A., Salhi-Hannachi, A., Luro, F., Vignes, H., Mournet, P., and Ollitrault, P. (2017). Genotyping by sequencing reveals the interspecific C. maxima / C. reticulata admixture along the genomes of modern citrus varieties of mandarins, tangors, tangelos, orangelos and grapefruits. *PloS one* 12, e0185618. doi: 10.1371/journal.pone.0185618 [doi]
Peng, Z., Bredeson, J. V., Wu, G. A., Shu, S., Rawat, N., Du, D., et al. (2020). A chromosome-scale reference genome of trifoliate orange (Poncirus trifoliata) provides insights into disease resistance, cold tolerance and genome evolution in Citrus. *The Plant Journal* 104, 1215–1232. doi: 10.1111/tpj.14993
Pootakham, W., Ruang-Areerate, P., Jomchai, N., Sonthirod, C., Sangsrakru, D., Yoocha, T., et al. (2015). Construction of a high-density integrated genetic linkage map of rubber tree (Hevea brasiliensis) using genotyping-by-sequencing (GBS). *Front. Plant Sci.* 6. doi: 10.3389/fpls.2015.00367
Qu, P., Wang, J., Wen, W., Gao, F., Liu, J., Xia, X., et al. (2021). Construction of Consensus Genetic Map With Applications in Gene Mapping of Wheat (Triticum aestivum L.) Using 90K SNP Array. *Front. Plant Sci.* 12. doi: 10.3389/fpls.2021.727077
Ramadugu, C., Keremane, M. L., Halbert, S. E., Duan, Y. P., Roose, M. L., Stover, E., et al. (2016). Long-Term Field Evaluation Reveals Huanglongbing Resistance in Citrus Relatives. *Plant Disease* 100, 1858–1869. doi: 10.1094/PDIS-03-16-0271-RE

Rasche, H., and Hiltemann, S. (2020). Galactic Circos: User-friendly Circos plots within the Galaxy platform. *Gigascience* 9. doi: 10.1093/gigascience/giaa065
Ritter, E., Gebhardt, C., and Salamini, F. (1990). Estimation of recombination frequencies and construction of RFLP linkage maps in plants from crosses between heterozygous parents. *Genetics* 125, 645–654.
Schilling, M. P., Wolf, P. G., Duffy, A. M., Rai, H. S., Rowe, C. A., Richardson, B. A., et al. (2014). Genotyping-by-Sequencing for Populus Population Genomics: An Assessment of Genome Sampling Patterns and Filtering Approaches. *PLOS ONE* 9, e95292. doi: 10.1371/journal.pone.0095292
Shimada, T., Fujii, H., Endo, T., Ueda, T., Sugiyama, A., Nakano, M., et al. (2014). Construction of a citrus framework genetic map anchored by 708 gene-based markers. *Tree Genetics & Genomes* 10, 1001–1013. doi: 10.1007/s11295-014-0738-9
Siberchicot, A., Bessy, A., Guéguen, L., and Marais, G. A. (2017). MareyMap Online: A User-Friendly Web Application and Database Service for Estimating Recombination Rates Using Physical and Genetic Maps. *Genome Biology and Evolution* 9, 2506–2509. doi: 10.1093/gbe/evx178
Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., et al. (2013). An Improved Genotyping by Sequencing (GBS) Approach Offering Increased Versatility and Efficiency of SNP Discovery and Genotyping. *PLOS ONE* 8, e54603. doi: 10.1371/journal.pone.0054603
Song, S., Liu, H., Miao, L., He, L., Xie, W., Lan, H., et al. (2023). Molecular cytogenetic map visualizes the heterozygotic genome and identifies translocation chromosomes in *Citrus sinensis*. *Journal of Genetics and Genomics* 50, 410–421. doi: 10.1016/j.jgg.2022.12.003
Storey, J. D. (2002). A Direct Approach to False Discovery Rates. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 64, 479–498.
Sun, Z., Li, S., Liu, Y., Li, W., Liu, K., Cao, X., et al. (2024). Telomere-to-telomere gapless genome assembly of the Chinese sea bass (Lateolabrax maculatus). *Sci Data* 11, 175. doi: 10.1038/s41597-024-02988-9
Swingle, W. T., and Reece, P. C. (1967). “The botany of Citrus and its wild relatives,” in *The Citrus industry*, eds. W. Reuther, H. J. Webber, and D. L. Batchelor (Berkeley, CA, USA), 190–430.
van Os, H., Stam, P., Visser, R. G., and van Eck, H. J. (2005). SMOOTH: a statistical method for successful removal of genotyping errors from high-density genetic linkage data. *Theor.Appl.Genet.* 112, 187–194. doi: 10.1007/s00122-005-0124-y
Wang, L., Huang, Y., Liu, Z., He, J., Jiang, X., He, F., et al. (2021). Somatic variations led to the selection of acidic and acidless orange cultivars. *Nat. Plants* 7, 954–965. doi: 10.1038/s41477-021-00941-x
Wang, N., Chen, P., Xu, Y., Guo, L., Li, X., Yi, H., et al. (2023). Phased genomics reveals hidden somatic mutations and provides insight into fruit development in sweet orange. *Hortic Res* 11, uhad268. doi: 10.1093/hr/uhad268
Wang, N., Song, X., Ye, J., Zhang, S., Cao, Z., Zhu, C., et al. (2022). Structural variation and parallel evolution of apomixis in citrus during domestication and diversification. *National Science Review* 9, nwac114. doi: 10.1093/nsr/nwac114
Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. *Nature genetics* 49, 765–772. doi: 10.1038/ng.3839 [doi]
Wu, B., Yu, Q., Deng, Z., Duan, Y., Luo, F., and Gmitter Jr, F. (2022). A chromosome-level phased genome enabling allele-level studies in sweet orange: a case study on citrus Huanglongbing tolerance. *Hortic Res* 10, uhac247. doi: 10.1093/hr/uhac247
Wu, G. A., Prochnik, S., Jenkins, J., Salse, J., Hellsten, U., Murat, F., et al. (2014). Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nat.Biotechnol.* 32, 656–662. doi: 10.1038/nbt.2906
Wu, G. A., Terol, J., Ibanez, V., LÃ³pez-GarcÃ-a, A., PÃ©rez-RomÃ¡n, E., BorredÃ¡, C., et al. (2018). Genomics of the origin and evolution of Citrus. *Nature* 554, 311.

Xu, Q., Chen, L.-L., Ruan, X., Chen, D., Zhu, A., Chen, C., et al. (2013). The draft genome of sweet orange (Citrus sinensis). *Nature Genetics* 45, 59–66. doi: 10.1038/ng.2472

Xu, Y.-Y., Liu, S.-R., Gan, Z.-M., Zeng, R.-F., Zhang, J.-Z., and Hu, C.-G. (2021). High-Density Genetic Map Construction and Identification of QTLs Controlling Leaf Abscission Trait in Poncirus trifoliata. *International Journal of Molecular Sciences* 22, 5723. doi: 10.3390/ijms22115723

Yu, Y., Chen, C., and Gmitter, F. G. (2016). QTL mapping of mandarin (Citrus reticulata) fruit characters using high-throughput SNP markers. *Tree Genetics & Genomes* 12, 77. doi: 10.1007/s11295-016-1034-7

Zhang Dianxiang and Mabberley David J. (2008). Citrus Linnaeus, Sp. *Fl. China* 11, 90–96.

Zhang, L., Reifová, R., Halenková, Z., and Gompert, Z. (2021). How Important Are Structural Variants for Speciation? *Genes (Basel)* 12, 1084. doi: 10.3390/genes12071084

Figure 1.TIF

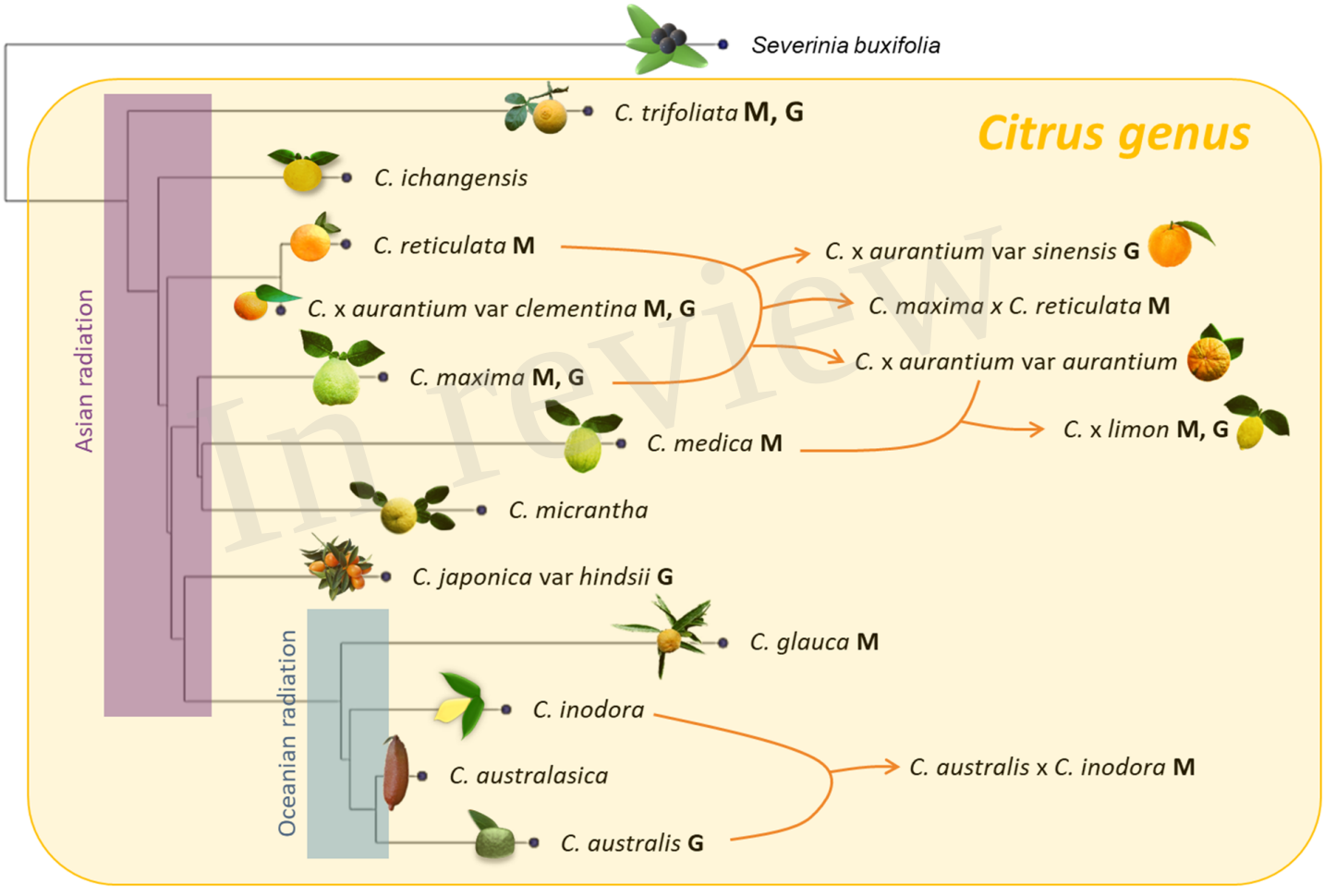

Severinia buxifolia
Citrus genus
C. trifoliata M, G
C. ichangensis
C. reticulata M
C. x aurantium var clementina M, G
C. maxima M, G
C. medica M
C. micrantha
C. japonica var hindsii G
C. glauca M
C. inodora
C. australasica
C. australis G
C. x aurantium var sinensis G
C. maxima x C. reticulata M
C. x aurantium var aurantium
C. x limon M, G
C. australis x C. inodora M
Asian radiation
Oceanian radiation
In review

Figure 2.TIF

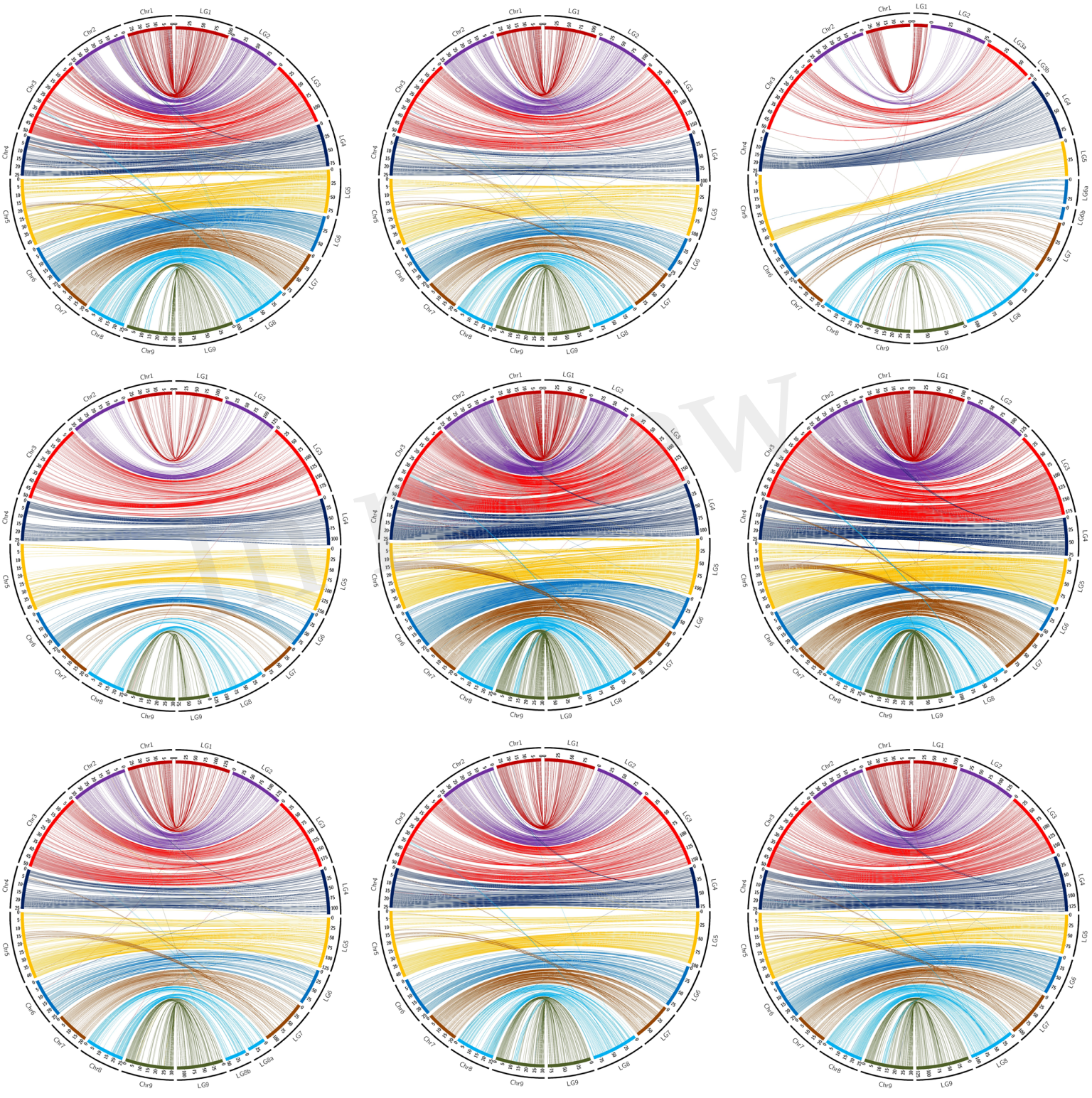

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
LG1
LG2
LG3
LG4
LG5
LG6
LG7
LG8
LG9
LG3a
LG3b
LG6a
LG6b
LG8a
LG8b

Figure 3.TIF

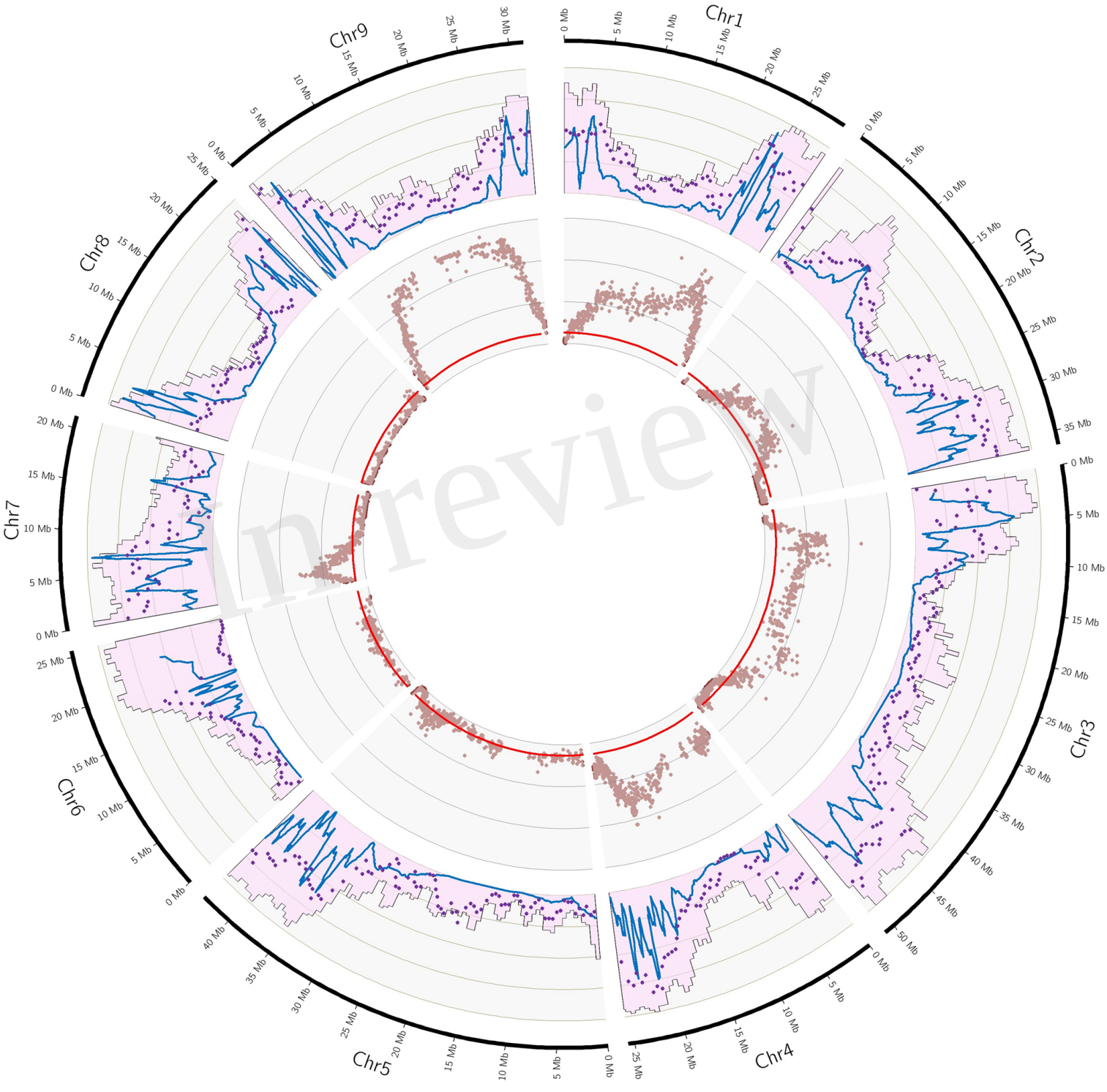

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
0 Mb
5 Mb
10 Mb
15 Mb
20 Mb
25 Mb
30 Mb
35 Mb
40 Mb
45 Mb
50 Mb
Inreview

Figure 4.TIF

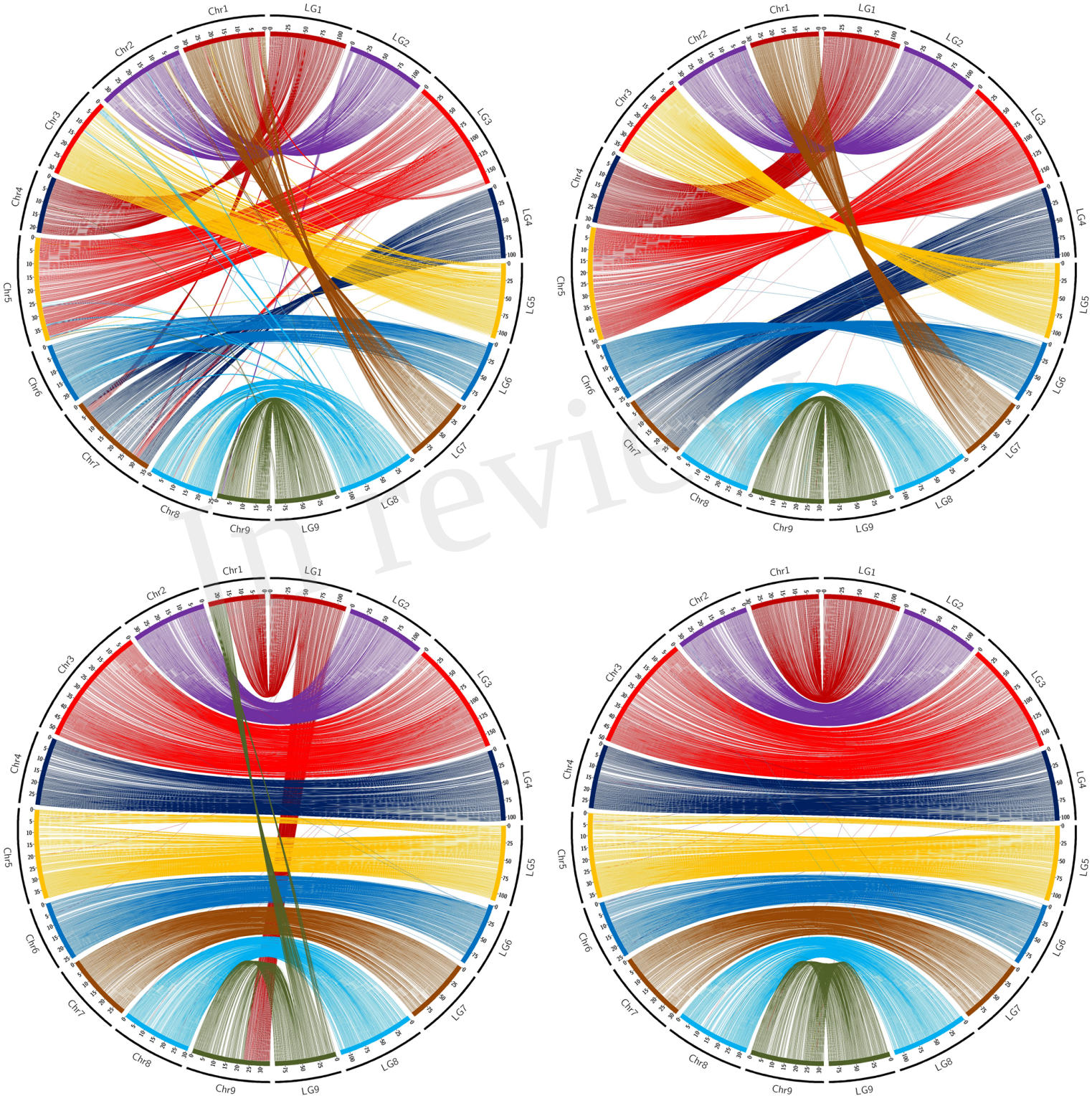

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
LG1
LG2
LG3
LG4
LG5
LG6
LG7
LG8
LG9

Figure 5.TIF

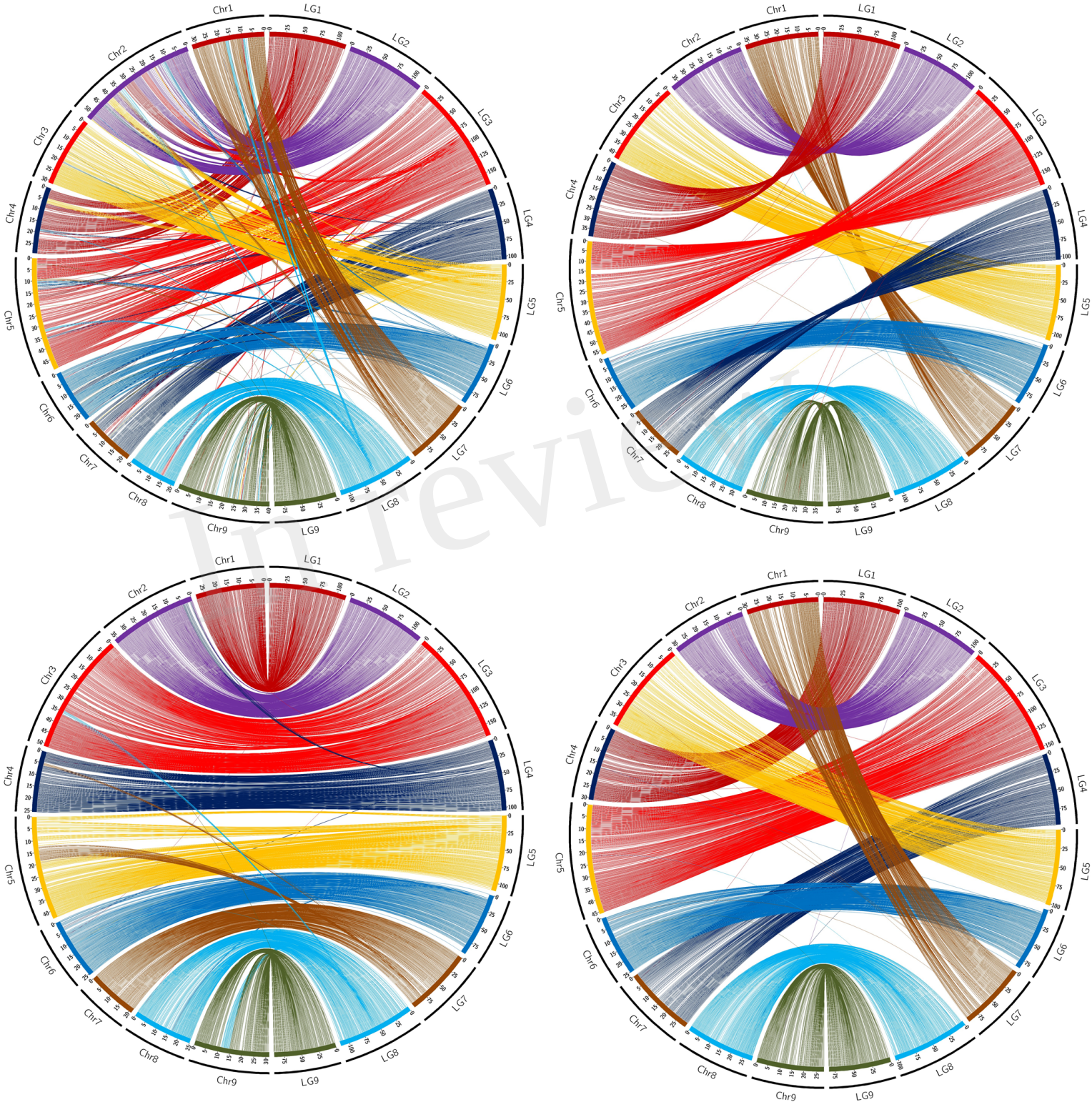

Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
LG1
LG2
LG3
LG4
LG5
LG6
LG7
LG8
LG9

Figure 6.TIF

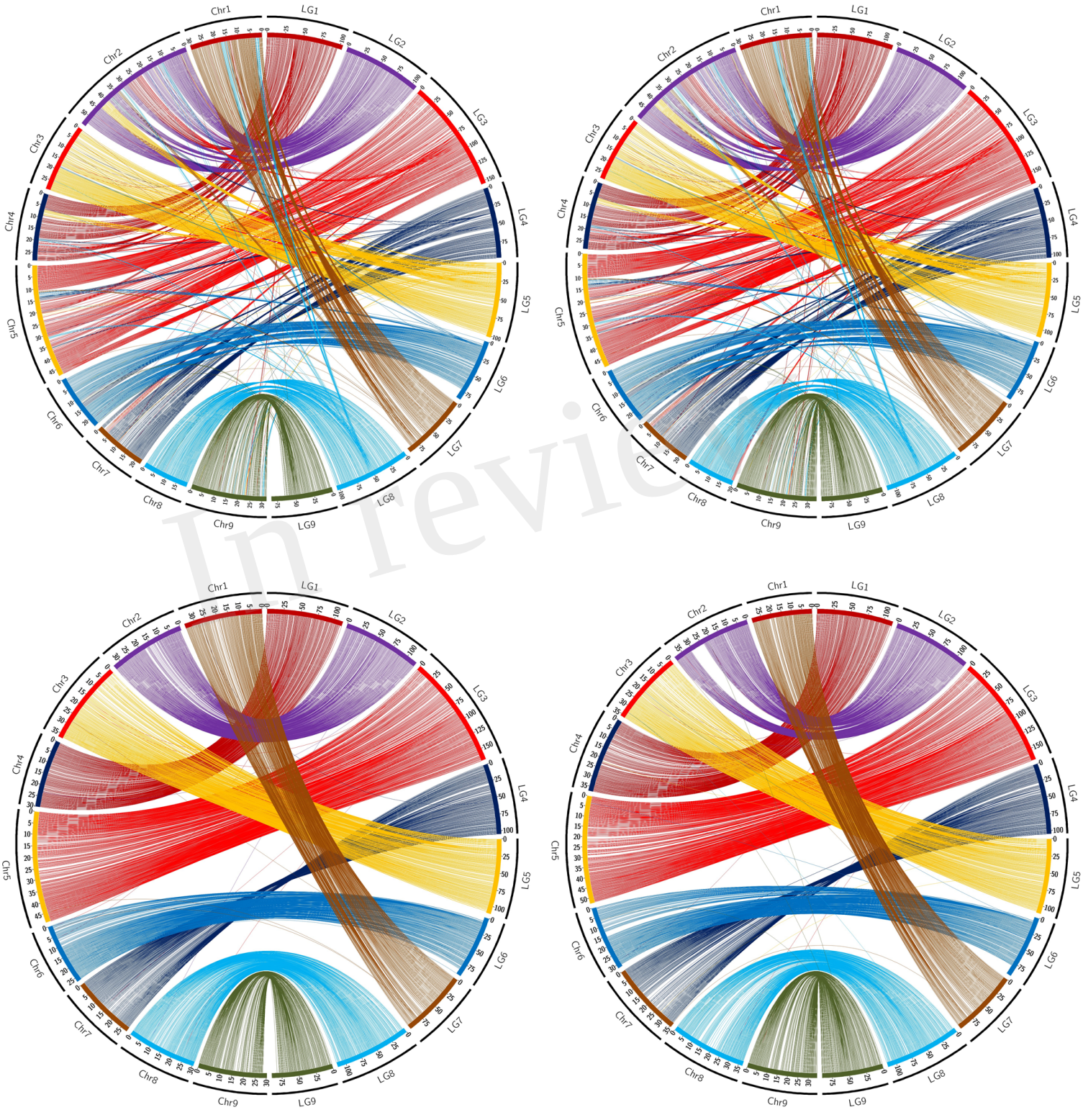
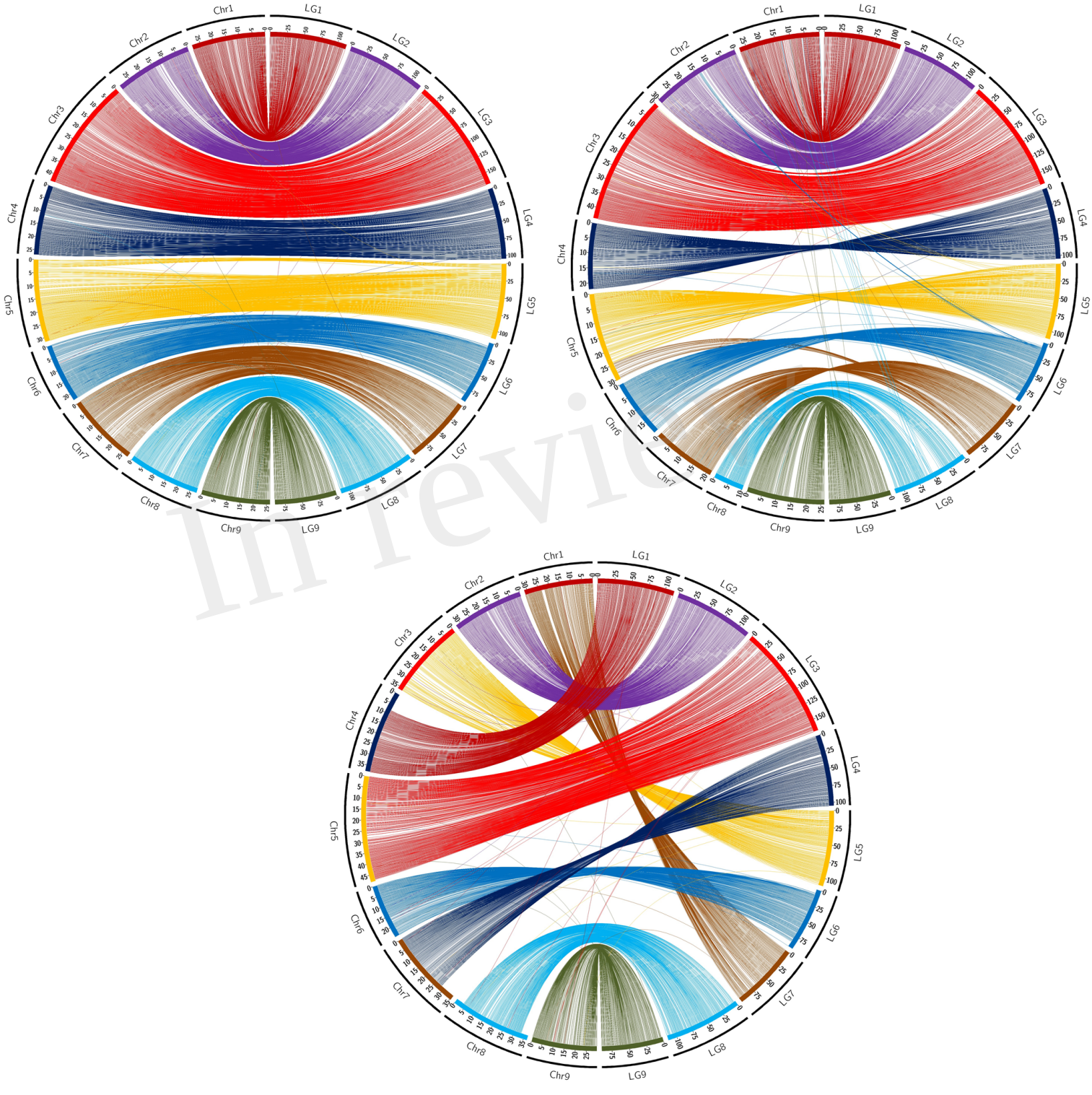


Figure 7.TIF


Chr1
LG1
Chr2
LG2
Chr3
LG3
Chr4
LG4
Chr5
LG5
Chr6
LG6
Chr7
LG7
Chr8
LG8
Chr9
LG9