

RESEARCH ARTICLE OPEN ACCESS

A Super-Pangenome for Cultivated Citrus Reveals Evolutive Features During the Allopatric Phase of Their Reticulate Evolution

Gaetan Droc^{1,2}  | Delphine Giraud³  | Caroline Belser⁴  | Karine Labadie⁵ | Simone Duprat⁴ | Corinne Cruaud⁵  | Benjamin Istace⁴ | Fredson Dos Santos Menezes⁶ | Edson Mario de Andrade Silva⁷ | Franck Curk⁸  | Gilles Costantino³ | Alexandre Soriano^{1,2} | Pierre Mournet^{1,2}  | Alexis Dereeper⁹  | Maëva Miranda^{1,2} | Elodie Marchi³ | Sylvain Santoni⁸ | Raner José Santana Silva⁶  | Stéphanie Sidibe-Bocs^{1,2}  | François Luro³ | Nathalie Choisne¹⁰ | Florian Maumus¹⁰  | Barbara Hufnagel^{1,11}  | Fabienne Micheli^{1,2}  | Patrick Wincker⁴ | Jean-Marc Aury⁴  | Arnaud Lemainque⁵ | Patrick Ollitrault^{1,2} 

¹UMR AGAP, Institut Agro Montpellier, CIRAD, INRAE, University of Montpellier, Montpellier, France | ²UMR AGAP, CIRAD, Montpellier, France | ³UMR AGAP, INRAE, Institut Agro Montpellier, CIRAD, University of Montpellier, San Giuliano, France | ⁴Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France | ⁵Genoscope, Institut François Jacob, CEA, CNRS, Univ Evry, Université Paris-Saclay, Evry, France | ⁶Departamento de Ciências Biológicas (DCB), Centro de Biotecnologia e Genética (CBG), Universidade Estadual de Santa Cruz (UESC), Ilhéus, BA, Brazil | ⁷Horticultural Sciences Department, University of Florida, Gainesville, Florida, USA | ⁸UMR AGAP, INRAE, Institut Agro Montpellier, CIRAD, University of Montpellier, Montpellier, France | ⁹UMR PHIM, IRD, Montpellier, France | ¹⁰INRAE, URGI, Université Paris-Saclay, Versailles, France | ¹¹UMR AGAP, CIRAD, Petit-Bourg, France

Correspondence: Gaetan Droc (gaetan.droc@cirad.fr) | Patrick Ollitrault (patrick_ollitrault@hotmail.com)

Received: 10 April 2025 | **Revised:** 25 October 2025 | **Accepted:** 5 January 2026

Keywords: *Citrus* | evolution | genome assembly | genome hub | pangenome | phylogenomics | resequencing

ABSTRACT

The main genetic diversity observed in cultivated citrus results from a reticulate evolution involving four ancestral taxa whose radiation occurred in allopatry. In such context, GWAS analysis, genome diversity and transcriptomic studies will be significantly enhanced through pangenome approaches. We report the implementation of a super-pangenome for cultivated citrus, established with de novo assemblies of *C. medica*, *C. reticulata* and *C. micrantha*, released for the first time alongside a published chromosome-scale assembly of *C. maxima*. Repetitive element annotation revealed that half of each genome consisted of transposable elements or DNA-satellites. The new genome assemblies display strong synteny and collinearity, while discrepancies are observed with the *C. maxima* assembly. Resequencing information from 55 accessions helped to explore the intra- and inter-specific diversity of the ancestral taxa and their relationships with horticultural groups. Diagnostic SNPs of the ancestral taxa revealed interspecific introgressions in several representative accessions of *C. reticulata*, *C. maxima* and *C. medica* as well as insights into the origin and phylogenomic structures of horticultural groups. PAV analysis revealed a gene whose absence or presence was specific to one of the ancestral taxa. Diagnostic PAV analysis uncovered a large chloroplastic introgression in *C. medica* chromosome 4. The analysis of the functional enrichment and species-specific adaptations in the citrus super-pangenome revealed distinct functional specialisations. This highlights the evolutionary paths that have shaped species, contributing to the diversity in the citrus super-pangenome while maintaining a shared foundation of essential biological processes. We established a Genome Hub, offering a platform for continuous genomic research.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2026 The Author(s). *Plant Biotechnology Journal* published by Society for Experimental Biology and The Association of Applied Biologists and John Wiley & Sons Ltd.

1 | Introduction

Most modern citrus varieties have arisen from reticulate evolution involving four ancestral taxa that evolved in allopatry (Swingle and Reece 1967; Wu et al. 2018). The radiation period of Asian *Citrus* species was estimated to be in the late Miocene (6–8 million years ago), corresponding to a significant climate transition from wet to drier conditions (Wu et al. 2018). *Citrus medica* L. originally evolved in north-eastern India and neighbouring regions of Burma and China. *Citrus reticulata* Blanco originated in China and diversified between China and Japan. The Malay Archipelago and Indonesia have been home to the evolution of *Citrus maxima* (Burm.) Merr., whereas *Citrus micrantha* Wester is native to the Philippines. Hybridisations between these ancestral taxa followed this initial allopatric differentiation, giving rise to the major modern horticultural groups (Figure 1; Nicolosi et al. 2000; Curk et al. 2016; Wu et al. 2018). Among them, mandarins, pummelos and citrons are derived mainly from *C. reticulata*, *C. maxima* and *C. medica*, respectively. However, WGS and GBS studies revealed evidence of *C. maxima* and *C. reticulata* introgressions in cultivated mandarins and some pummelos, respectively (Oueslati et al. 2017; Wu et al. 2018).

Facultative apomixis, through polyembryony of nucellar origin, is present in several *Citrus* species (Wang et al. 2017, 2018; Wang, Song, et al. 2022) and has constrained the number of interspecific recombination cycles. This resulted (Figure 1) in the fixation of genomic structures either in complete interspecific heterozygosity, such as for the sour orange (*C. × aurantium* L. = *C. maxima* × *C. reticulata*), ‘Mexican’ lime and ‘Alemow’ (*C. × aurantiifolia* [Christm.] Swing. = *C. micrantha* × *C. medica*), Rangpur lime and Volkamer lemon (*C. × limonia* Osb. = *C. reticulata* × *C. medica*), or in interspecific mosaics combining large genome fragments from different species (Wu et al. 2014; Wang et al. 2018). These mosaics may involve two species, such as sweet orange or pummelo (respectively *C. × sinensis* L. Osb. and *C. × paradisi* Macf. resulting from *C. reticulata*/*C. maxima* admixture), or even three ancestors, such as Mediterranean lemon (*C. × lemon* [L.] Burm.) resulting from admixture of *C. medica*, *C. reticulata* and *C. maxima*. There can even be four ancestors, such as for the ‘Tahiti’ lime tree (*C. × latifolia* Tan. = *C. reticulata*/*C. micrantha*/*C. medica*/*C. maxima* admixtures; Ahmed et al. 2019).

Phenome analysis has shown that a large part of the actual phenotypic diversity in cultivated citrus results from the differentiation between these four ancestral species during allopatric evolution. This is true for the content of some important secondary metabolites such as coumarins, furanocoumarins (Dugrand-Judek et al. 2015), and carotenoids (Fanciullino et al. 2006), as well as pomological and adaptation traits (Santini et al. 2012; Wu et al. 2018). Therefore, the ancestral species and admixed horticultural groups (i.e., oranges, grapefruits, lemons, and limes) constitute fundamental genetic resources to breed new citrus varieties. Thus, it is essential to decipher the determinants of the traits supporting the cultivated ideotypes of modern *Citrus* and the variability between these four ancestral taxa.

In this context of reticulate evolution, genome diversity studies, GWAS analysis, and transcriptomic studies will

be significantly improved through pangenome approaches. Pangenomes limit the biases of analyses based on a single species reference genome (Stevenson et al. 2013; Markello et al. 2022), and pangenome graphs should be particularly useful for allele-specific expression studies in interspecific hybrids and admixed genotypes (Salavati et al. 2019; Sibbesen et al. 2023). Recent studies in plants were conducted to capture a wide range of genome diversity in species or genera (Liu et al. 2020; Li et al. 2023; Guo 2023) and to decipher the genomic control of some traits, such as fruit quality (Gao et al. 2019; Sun et al. 2020; Yang et al. 2023; Yocca et al. 2023; Ding et al. 2023). Few pangenomic studies on citrus have been published (Zia et al. 2022; Liu et al. 2022; Gao et al. 2023; Huang et al. 2023), and none encompass the whole diversity of cultivated citrus because they miss *C. micrantha* and the integration of intraspecific variability.

Here, we report the implementation of a super-pangenome for cultivated citrus, established for the first time with de novo assemblies in pseudochromosomes of *C. medica*, *C. reticulata* and *C. micrantha*, complemented by a published *C. maxima* chromosome scale assembly (Wang et al. 2017). We annotated transposable elements (TEs) and protein-coding genes. Special attention was devoted to pattern recognition receptor (PRR) genes, coding for immune system receptors. The definition of core and dispensable genomes integrates (i) the PAV information from resequencing data from 43 accessions representative of the four ancestral species and (ii) a comparative genomics approach based on homology clusters predicted from the four ancestral proteomes. Functional analysis of genes whose presence or absence is specific to a single ancestral taxon reveals key evolutionary features during the allopatric phase of reticulate evolution in cultivated citrus.

2 | Results

2.1 | De Novo Reference Genomes for Three Ancestral Species of Cultivated Citrus

2.1.1 | Assembly in Pseudochromosomes

The estimation of genome size and heterozygosity with Genomescope2.0 from Illumina reads provided the following values (Figure S1): 312Mb (0.9% heterozygosity) for CITMI (*C. micrantha*), 415Mb (0.242% heterozygosity) for CITME (*C. medica* cv ‘Corsican citron’) and 343Mb (0.712% heterozygosity) for CITRE (*C. reticulata* cv ‘Cleopatra mandarin’). We produced 42.74, 32.34 and 52.33 Gb of long-read sequences (Nanopore) for CITMI, CITME and CITRE, respectively (Table S2), with N50 values of 18.0, 26.3 and 19.7 kb, respectively. The metrics for optical mapping are available in Table S3.

The combination of Nanopore long-range sequencing, Illumina PCR free sequencing and optical mapping resulted in 141, 314 and 310 scaffolds encompassing 346.4, 332.9 and 332.1 Mb for the CITME, CITMI and CITRE assemblies, respectively. These assemblies showed high completeness, with BUSCO scores exceeding 98.4% for complete genes, and robust quality metrics as indicated by Merqury QV scores around 29 (Table S4, Figure S2).

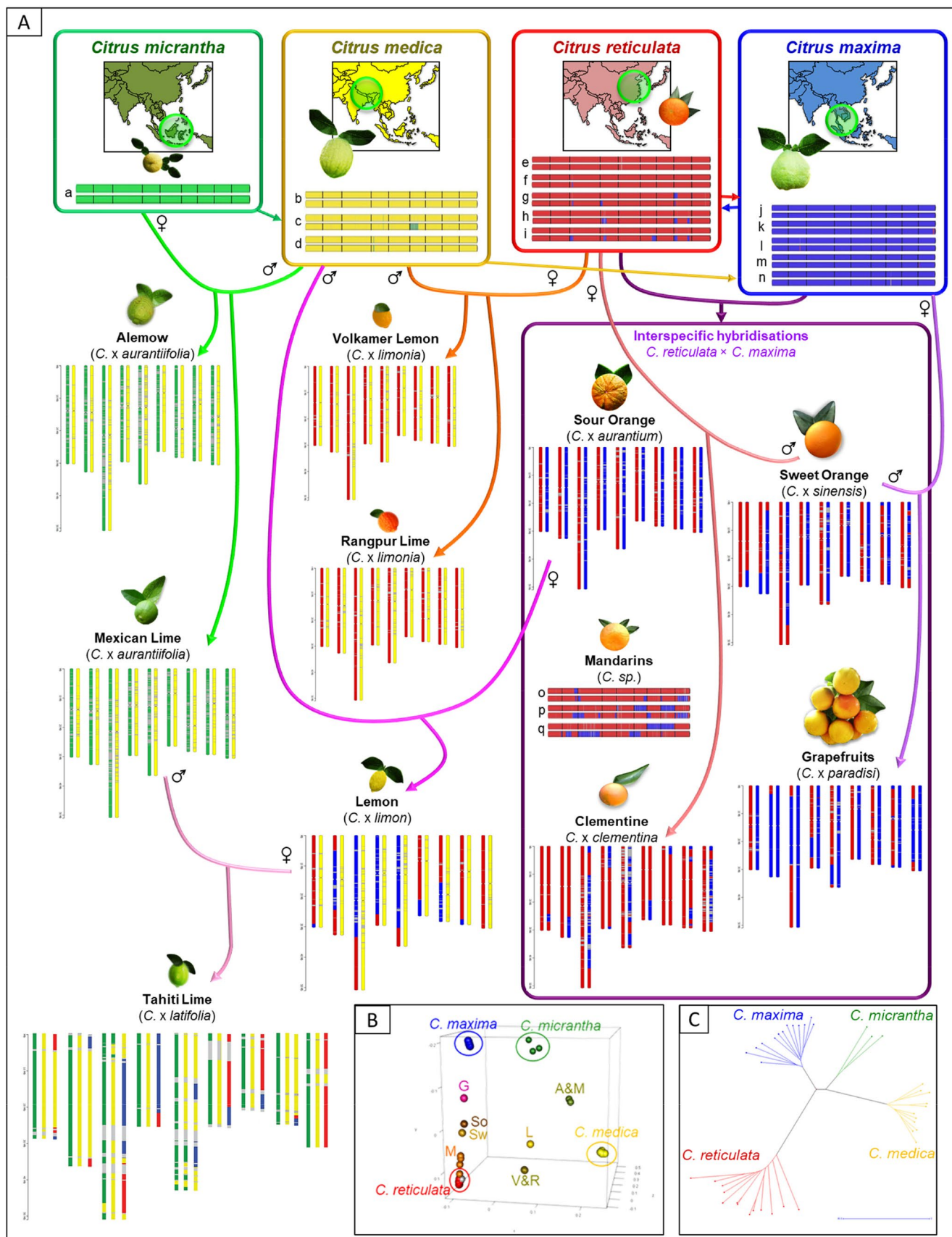


FIGURE 1 | Legend on next page.

FIGURE 1 | The reticulate evolution of the main horticultural groups of cultivated citrus and their interspecific mosaic structures. (A) Origin and phylogenomic karyotypes of *Citrus* species derived from the four considered ancestral taxa (a: Pap-macrD, pap-melaD, pap-mic1D; b: Co-ed, cit-cor1D, cit-diamD, cit-humpD, cit-macnD, cit-macvD, cit-wc1D, cit-wc2D, cit-wc3D; c: Cit-chinD; d: Cit-e861D; e: Man-daodD, man-HezWD, man-JiaWD, man-ManWD, man-scsH, man-sunkD, man-tachD; f: Man-clo1D, man-shekD; g: Man-nfmcD; h: Man-DaoW1D, man-szinD; i: Man-avanD; j: Pum-ChonD, pum-florD, pum-hf22D, pum-HuazD, pum-indeD, pum-reinD, pum-tah1D; k: Pum-HuanD; l: Pum-cha2D, pum-chaH, pum-kaopD; m: Pum-GuilD; n: Pum-timoD; o: Man-ponkD; p: Man-kingD; q: Man-owarD [correspondence between codes and variety names in Table S1]). (B) Factorial analysis displaying the intermediate position of the horticultural groups resulting from reticulate evolution (G: Grapefruits, So: Sour oranges; Sw: Sweet oranges; M: Cultivated mandarins; A&M: 'Alemow' and 'Mexican' limes; L: Mediterranean lemon); (C) Diversity of the representative accessions of the four ancestral species (neighbour-joining tree based on Manhattan distances). Figure (A–C) analyses were conducted using CITRE as the reference genome.

For the final assemblies in pseudochromosomes, the CITRE and CITME assemblies were anchored on the available genetic maps of Cleopatra mandarin and Corsican citron (Ollitrault et al. 2024). For Cleopatra, the genetic map was saturated and allowed for the establishment of the final CITRE assembly of 9 pseudochromosomes by anchoring 14 scaffolds that represent 89.2% of the whole CITRE assembly (Table S5). The synteny between the CITRE assembly and Cleopatra genetic map was high (98.5%) as was collinearity (Spearman coefficient = 0.956 ± 0.018 on average for the 9 chromosomes; Table S5, Figure S3a). The anchoring of the Cleopatra genetic map to the CITRE genome revealed a recombination landscape closely associated with gene sequence density along the genome (Figure S3c). For Corsican citron, the genetic map presented large gaps due to homozygosity related to the inbred origin of Corsican citron (Luro et al. 2012). However, sequence anchoring allowed for the integration of 13 scaffolds in 9 pseudochromosomes, encompassing 96.5% of the whole CITME assembly. Synteny (99.2%) and collinearity (Spearman coefficient = 0.935) between the assembly and the partial genetic map were high (Table S6, Figure S3b). For CITMI, no genetic map was available, and the synteny and collinearity with the *C. clementina* V1.0 assembly (Wu et al. 2014) drove the final assembly in pseudochromosomes (Figure S4). Twenty-one scaffolds were included in the 9-pseudochromosome assembly, representing 90.1% of the whole CITME assembly (Table S7). High synteny and collinearity were observed (Figures S5 and S6) between our 3 genome assemblies and the 2 sweet orange haplotypes published by Wu et al. (2022), *C. trifoliata* (Peng et al. 2020), *C. hindsii* and *C. australis* (Nakandala et al. 2023) genome assemblies. The 2 major discrepancies systematically observed between *C. clementina* V1.0 assembly Chr 3 and Chr 5 and our 3 new assemblies (Figure S5a–c) were previously described in genetic mapping studies and attributed to local mistakes in *C. clementina* assembly (Ollitrault et al. 2024).

2.1.2 | Chloroplast Genome Assemblies

The chloroplast genome assemblies of CITRE, CITME and CITMI encompass 160601, 160016 and 159936 bp, respectively. They contain, respectively, 97, 91 and 95 genes, and each assembly displays 10 rRNAs and 37 tRNAs (Table S8). These genomes exhibit the classical chloroplast structure, featuring two copies of an inverted-repeated region that separate large and small single-copy regions (Figure S7). The phylogenetic position of the 3 new assembled genomes is consistent with previously published chloroplast genomes for *C. reticulata*, *C. medica* and

C. micrantha (Figure S8). The sequences and annotations of the 3 chloroplast genomes are available on our website, <https://citrus-genome-hub.southgreen.fr>.

2.1.3 | Annotation of the 3 Nuclear Genome Assemblies and Comparison With *C. maxima* Genome

To accurately compare representative genome assemblies of the four ancestral species, we performed with the same pipeline the annotation of the 3 new genome assemblies (CITME, CITMI and CITRE) and of the chromosome-scale *C. maxima* assembly published by Wang et al. (2017) and called CITMA in the present paper. Two main citrus chromosome numbering methods are currently used at the international level. They result from the first 2 pseudochromosome assemblies for citrus: the one of *C. sinensis* v1.0 (Xu et al. 2013) used for the *C. maxima* assembly (Wang et al. 2017) and the one of *C. clementina* v1.0 (Wu et al. 2014) that we used for CITME, CITMI and CITRE. In addition, the chromosome orientations appeared to be more diverse (Ollitrault et al. 2024). This inconsistency in chromosome numbering and orientation may result in confusion when one tries to integrate information based on different reference genomes. For synteny and collinearity analyses, we considered a numbering and orientation for *C. maxima* identical to our 3 new genome assemblies.

2.1.3.1 | Genes. RNA-seq was performed with young and mature leaves, entire flowers, and roots of the 3 accessions used to establish the de novo assembly, with the Chandler pummelo as a representative of *C. maxima*. A total of 307 532 Mb of raw data were produced, corresponding to an average coverage of 561× for each combination accession/organ (Table S9).

A total of 30 101, 28 090, 29 258 and 29 477 high-confidence protein-coding genes were retained respectively for CITMA (345 Mb), CITME (346 Mb), CITMI (333 Mb) and CITRE (332 Mb) genome assemblies, respectively, by using a combination of homologous proteins, transcriptomic data and ab initio prediction. Gene completeness was evaluated with BUSCO v5.0.0 against the eudicots_odb10 lineage, yielding high BUSCO scores of 94.8%, 94.7%, 93.3% and 93.1% for CITME, CITMI, CITRE, and CITMA, respectively. The number of genes per chromosome along with gene completeness is detailed in Table S10. Functional annotations were performed based on InterProScan domain searches and similarity searches using the UniProt/SwissProt and UniProt/TrEMBL

databases (BlastP). Gene ontology (GO) terms were derived from InterProScan results (Jones et al. 2014), and enzyme classification (EC) numbers were predicted using KofamScan (Table S11). Gene density varies along the chromosomes with similar patterns for the homologous chromosomes of the 4 assemblies (Figure 2).

2.1.3.2 | Transposable Elements and DNA-Satellite Sequences. Transposable elements (TEs) constitute a significant portion of the analysed genomes, with 42.8% in CITMI, 43.6% in CITRE, 47.3% in CITME and 42.5% in CITMA (Figure S9, Table S12). Most of them (74%–77%) are LTR-retrotransposons, including *Copia* *Tork*, *Sire* and *Retrofit* elements (comprising 12.6%–15.2% of genomes across species) as well as *Gypsy*, *Athila* and *Tat* elements (comprising

12.0%–15.0% of genomes) (Figure S10). The higher TE content in *C. medica* compared to the other 3 ancestors (164Mb vs. 142Mb–147Mb) was explained by an accumulation of *Gypsy*, *Athila* and *Copia Sire* elements in the former. Analyses of intraspecific variability also confirmed the significantly larger amount of TEs in *C. medica* (*t*-test with *p*-value < 0.001) (Figure S11). Additionally, we estimated that approximately 10% of genomes consisted of Class II elements, predominantly TIR elements such as *Mutator*, *hAT* and *CACTA*. Non-autonomous MITE elements constituted from 1.70% to 1.92% of genomes in *C. medica* and *C. reticulata*, respectively. Analysis of chromosome distribution indicated that TEs were dispersed throughout the chromosomes but were more concentrated in specific regions (Figure 2). Such concentrated regions are generally central, but in some chromosomes,

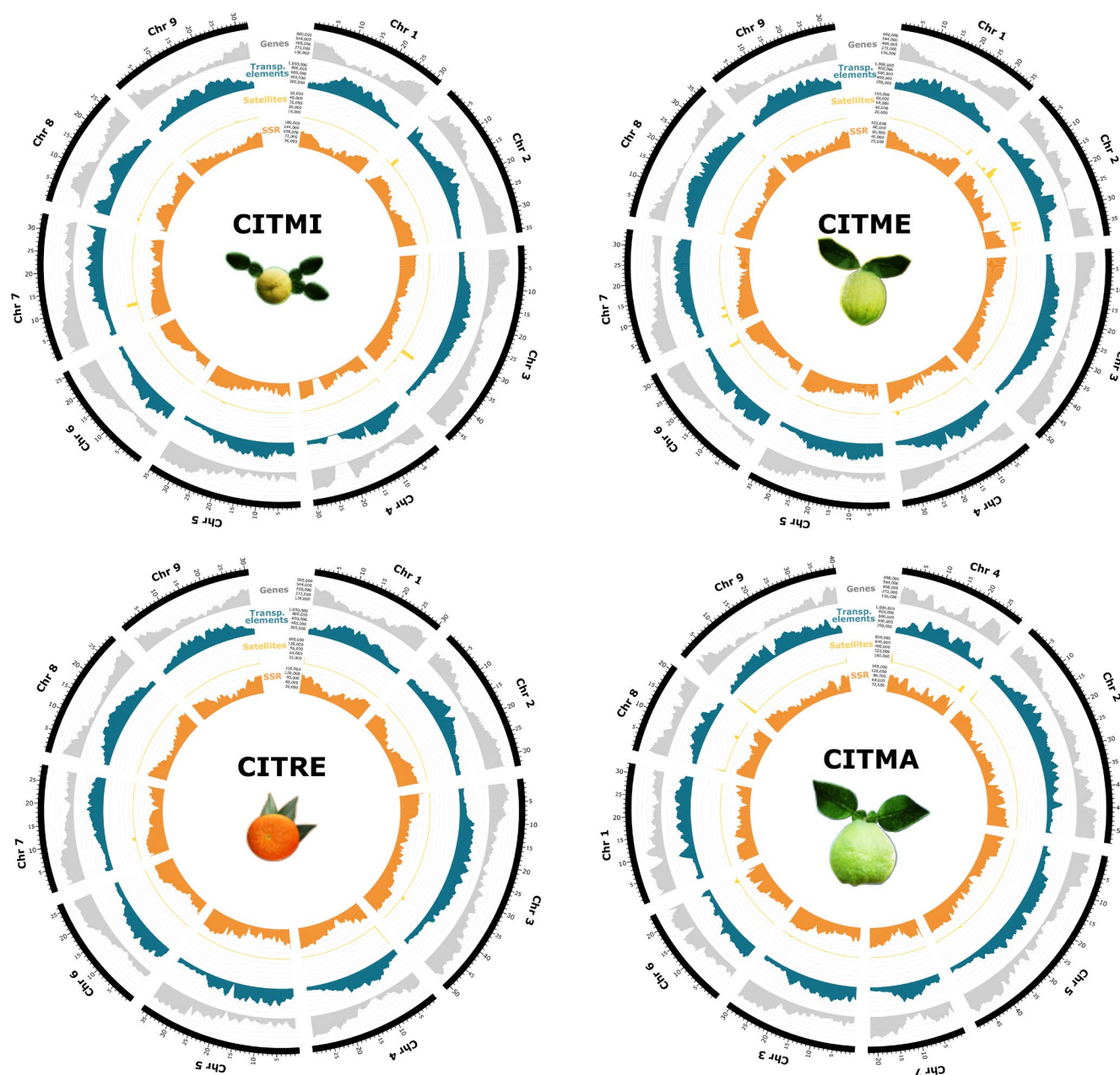


FIGURE 2 | Density of genes (grey histogram), transposable elements (blue histogram), DNA satellites (yellow histogram) and SSRs (orange histogram) along the 9 assembled pseudo-chromosomes of the 4 *Citrus* ancestors. Densities were calculated with a 1-Mb sliding window. CITMA chromosomes numbered per Wang et al. (2017), but ordering in the figure follows that of the 3 other genome assemblies.

such as Chr 6 for all 4 ancestors, TEs are notably accumulated on one chromosome side. In addition, we identified 2 distinct DNA-satellite sequences in the four species under study, accounting for 0.53% of the *C. micrantha* genome to 1.95% of the *C. reticulata* genome. Both DNA satellites, 181 and 141 bp in size, are tandemly repeated in the chromosome ends, but also in the inner regions of the chromosomes. The 181-bp satellite varies in abundance across species, with 1.7Mb in *C. micrantha*, 2.8Mb in *C. maxima*, 5.3Mb in *C. medica* and 6.4Mb in *C. reticulata*.

2.2 | Nuclear Genome Diversity in Cultivated Citrus According to Individual Reference Genomes

2.2.1 | Structural Variation Between Chromosome Scale Genome Assemblies of the 4 Ancestral Species

A strong synteny and collinearity are observed for the 3 new de novo assemblies, except for an inversion in the central region of the CITME Chr 2 when compared to CITRE and CITMI (Figure 3). In addition to different numbering and orientations of the CITMA pseudo-chromosomes compared with the 3 other assemblies, the riparian plot reveals many structural differences between CITMA and the 3 other assemblies. It concerns non-syntenic blocs, inverted regions and blocs in different positions in one chromosome.

2.2.2 | PRR Genes Diversity and Organisation Between the 4 Ancestral Genomes

The total number of PRRs and R proteins (named as PRRs in the subsequent paragraphs) varied between 1627 for CITME and 2014 for CITRE (Table S13). Across the 4 genome assemblies, a substantial portion (average of 64.3%) of the identified PRRs fell into the 'unknown' class as previously defined (Santana Silva and Micheli 2020). Among the identified classes, CNL, NL and RLK had the greatest relative contributions across all 4 genome assemblies (Figure S12).

Their combined proportions accounted for 51.9%, 57.4%, 66.4% and 67.5% of the identified PRRs for CITME, CITMI, CITRE and CITMA, respectively. The relative contribution of T, TN and TNL was lower in CITMA (6.2% in total) than in the 3 other assembled genomes (20.3%, 17.3% and 11.5% for the CITME, CITMI and CITRE genomes, respectively). The distribution of PRRs from the 12 identified classes is uneven among the chromosomes (Table S14). For the 4 genome assemblies, their number is low in Chr 4, 6, 8 and 9 (from 1.9% in CITMI-Chr4 and CITME-Chr6 to 6.1% in CITMA-Chr9). Chr 3 concentrates the largest proportion of PRRs (over 25%) for CITME, CITMI and CITRE, whereas the largest proportion for CITMA is found in Chr 2 (28.7%). For the 4 genome assemblies, a large proportion of PRRs (from 19.2% for CITME to 24.7% in CITMI) was also observed in Chr 7 and to a lesser extent in Chr 5 (ranging from 13.0% in CITMI to 21.7% in CITRE).

For these 2 chromosomes, the predominant class is CNL in the 4 genomes (Figure 4). RLK is the predominant class in Chr 2 of the 4 genomes, whereas TNL is concentrated in Chr 3 for CITME,

CITMI, and CITRE. Cluster analysis, including the PRR genes of the unknown class, revealed that a very large proportion (ranging from 85.5% in CITRE to 87.2% in CITMA) of the PRR genes were organised in clusters (Table S15). Depending on the genomes and chromosomes, the clusters range from 70% to 95.3% (CITMA-Chr 3 and CITMI-Chr 7, respectively). The number of clusters is similar for all 4 genomes, varying from 298 to 332 for CITRE and CITMA, respectively, with an average number of 4.61 ± 0.17 genes per cluster. The rates of duplication events appear to vary among the 4 genomes (Figure 4; Table S16). It is more frequent in CITMA, with 46 duplications, mainly involving the CNL (25), RLK (5), and NL (5) classes. The lowest rate is observed in CITRE, with only 14 duplications, primarily for RLKGNK2 (4) and RLK (3). Sixteen duplication events, primarily involving RLK (6) and RLKGNK2 (3), were observed in CITME, whereas CITMI displayed 21 duplication events, with the largest amounts present in CNL (6) and RLK (5). Most duplications appear to be interchromosomal for the 4 genome assemblies (Figure 4).

2.2.3 | Intra- and Interspecific Diversities Based on Resequencing Data of 55 Accessions of Cultivated Citrus

2.2.3.1 | SNP and Small Indel Variation. The intraspecific diversity was analysed for each genome assembly using the corresponding representative accessions (Table S17, Figure S13). It reveals a lower polymorphism for *C. medica* than for the other species. The total number of polymorphisms (SNPs and indels) was 3924735 for *C. medica*, whereas it varies between 6905527 (*C. maxima*) and 8471249 (*C. reticulata*) for the other species. The average heterozygosity of the *C. medica* accessions was also very low (0.95 heterozygous sites per kilobases) when compared with the other species (5.10, 4.17 and 3.70 for *C. micrantha*, *C. reticulata* and *C. maxima*, respectively). The proportion of indels over the whole variants is remarkably similar ($10.78\% \pm 0.32\%$) for all 4 species. Analysis of all representatives of the 4 species on each genome assembly yielded similar results regardless of the reference genome, with 19633997 ± 501376 SNPs and 3006432 ± 122037 and a proportion of indels ($13.28\% \pm 0.22\%$) slightly increasing compared to intraspecific data. There was a limited increase in total diversity (24169356 ± 760674) when admixed accessions were included, with no significant change in the indel rate relative to whole polymorphism (13.46 ± 0.22). The diversity of the representative accessions of the 4 ancestral species represents, respectively, 92.75%, 93.68%, 94.04% and 94.27% of the total diversity for the CITRE, CITME, CITMA and CITMI genome assemblies, with interspecific diversity constituting the main component (77.94%, 65.21%, 62.20% and 58.32% for the variant calling using CITME, CITMA, CITMI and CITRE, respectively; Figure S13). Admixed accessions displayed a logical, significant increase in heterozygosity rate (on average 3.43 times higher) compared to accessions representative of the ancestral species. Phylogenetic trees of the representative accessions of the 4 ancestors and the factorial analysis of the whole set of accessions (Figure 1B,C, respectively, for example, for the CITRE reference and Figures S14 and S15) confirm this high level of interspecific differentiation and the strong structuration between the ancestral species. Admixed accessions are positioned in the factorial analysis according to the contributions of their ancestral taxa.

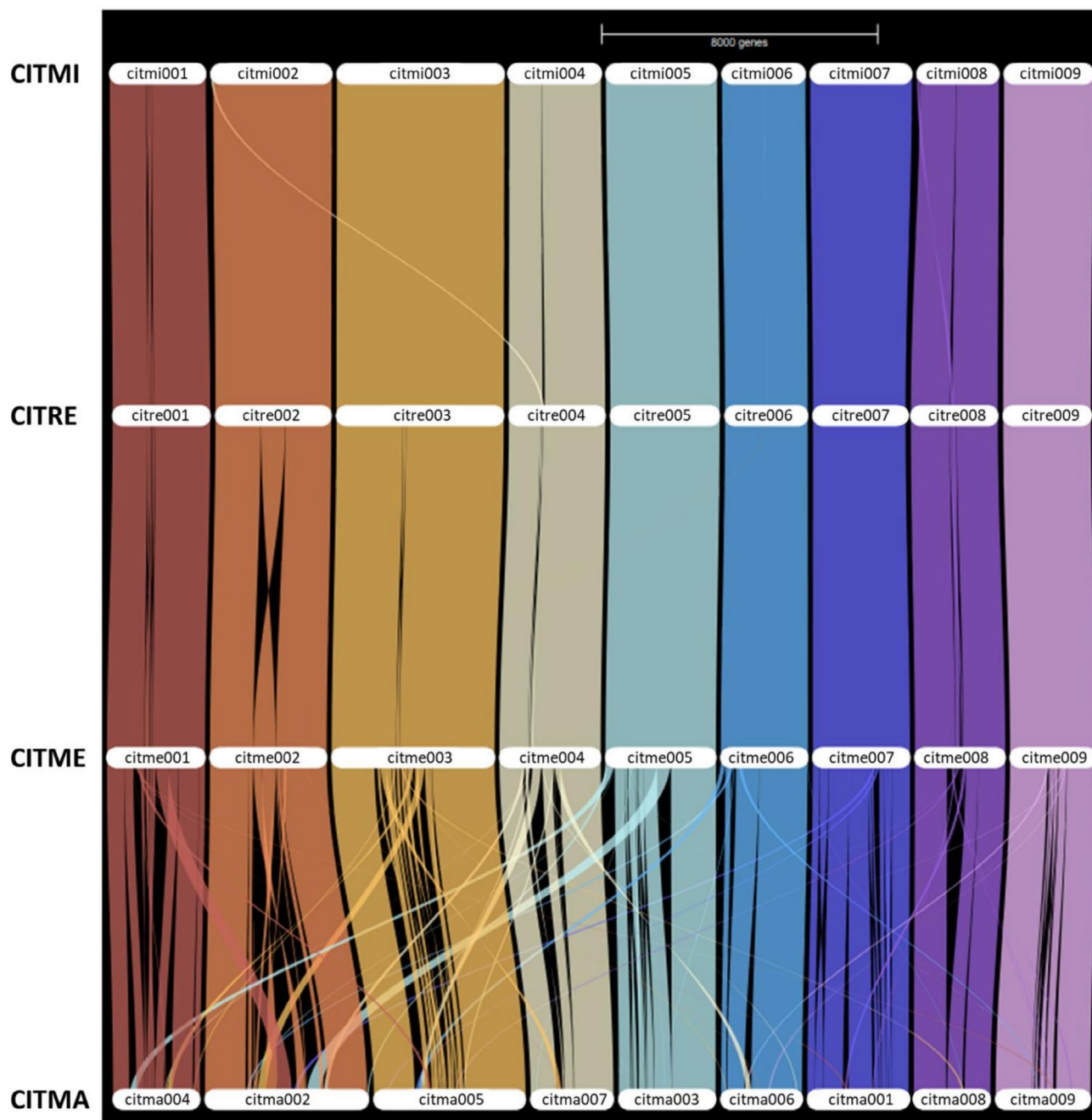


FIGURE 3 | Riparian plots representing the synteny map of orthologous genes along the 9 citrus chromosomes.

2.2.3.2 | Introgressions and Admixture in Cultivated Citrus. The first step was to identify SNPs that fully differentiated one species from the 3 other. These diagnostic SNPs (DSNPs) of the 4 ancestors were mined in the 4 genome assemblies. A total of 2 227 454, 2 298 439, 2 170 653 and 2 271 482 DSNPs were found for the CITMA, CITME, CITMI and CITRE genome assemblies, respectively (Table S18). The DSNP lists for the 4 genome assemblies are downloadable at <https://citrus-genome-hub.southgreen.fr>. DSNPs represent, respectively, 11.38%, 11.39%, 11.44% and 11.47% of the total number of SNPs found for the set of representative accessions of the 4 ancestral species, and they are dispersed all along the genome (Figure S18). Across all 4 assemblies, the relative proportion of DSNPs for each ancestral species was similar,

with a majority of *C. medica* DSNPs ($47.43\% \pm 0.88\%$) and a minority for *C. maxima* ($13.90\% \pm 0.81\%$) and *C. micrantha* ones ($12.75\% \pm 1.52\%$), with *C. reticulata* displaying intermediate values ($25.90\% \pm 1.01\%$). These results are consistent with the relative differentiation between the 4 corresponding horticultural groups that the phylogenomic analysis revealed (Figure 1C). The admixture patterns established using the CITME, CITMI and CITRE assemblies are very similar for all re-sequenced accessions (Figures S19–S22). However, the results from CITMA (Figure S21) differ slightly from those of the other three assemblies due to structural differences between CITMA and the other assemblies. Nonetheless, the same conclusions are obtained from the analysis on the 4 reference genomes. Figure 1A shows the interspecific

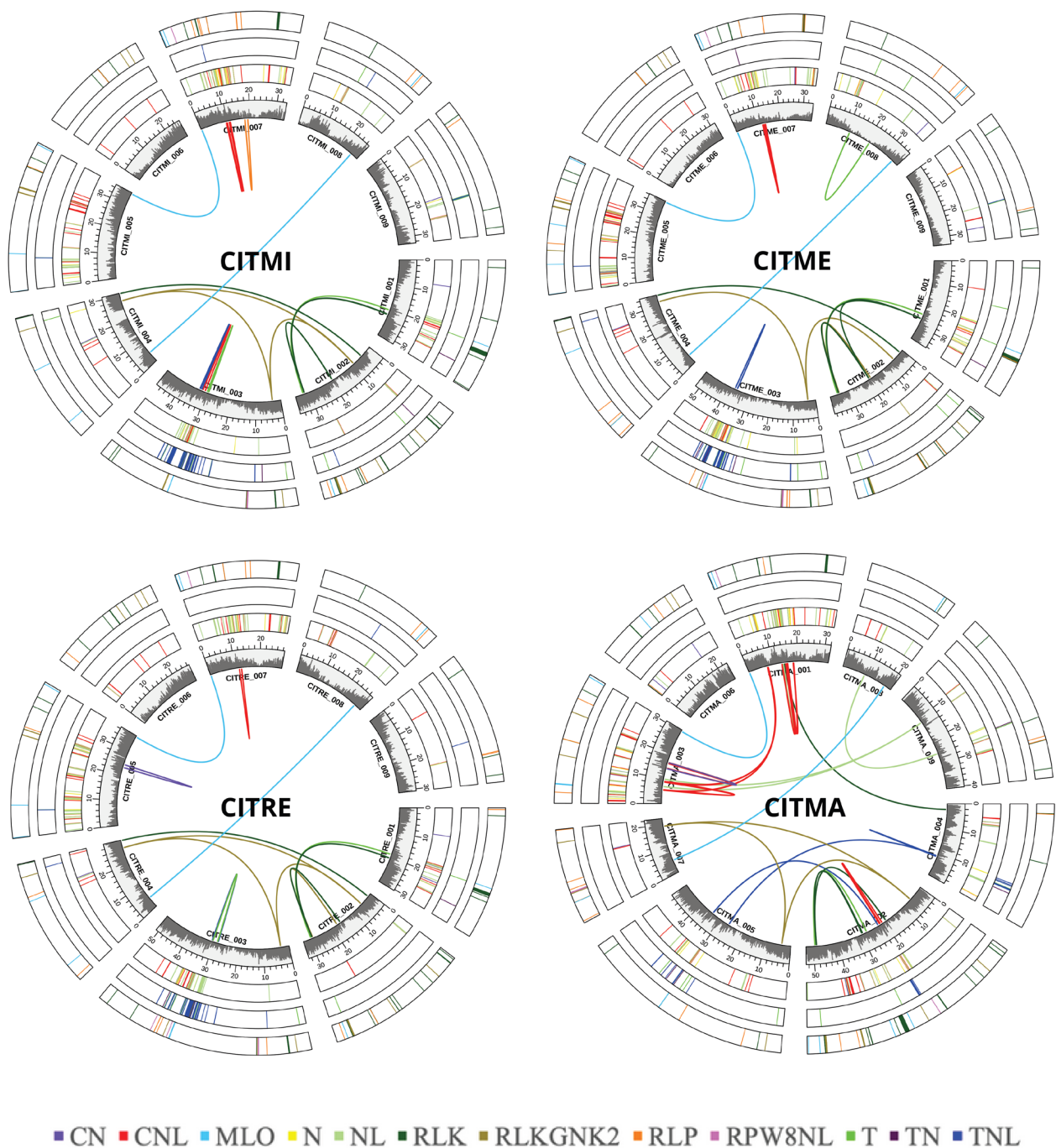


FIGURE 4 | Distribution of PRRs and/or R genes in the 12 classes previously defined (Santana Silva and Micheli 2020) along the 9 chromosomes of the CITMI, CITME, CITRE and CITMA genome assemblies. Grey histograms: Gene density. CITMA chromosomes are numbered per Wang et al. (2017), but ordering in the figure follows that of the 3 other genome assemblies.

mosaic structures revealed using CITRE as a reference. The mandarin horticultural group displays a predominance of the *C. reticulata* genome but also shows frequent introgressions from *C. maxima*. Seven of the analysed pummelos appear to be pure representatives of *C. maxima* whereas 6 display limited introgressions. For 3 varieties ('Chandler', 'Kao pan' and 'Chandler haploid'), it consists of 1.2-Mb *C. reticulata* introgressions in Chr 2. *Citrus reticulata* introgressions are also observed in Chr 9 spanning 0.5 and 1.5 Mb for 'Guil' and 'Huan' pummelos, respectively. The 'Timor' pummelo displays a 0.5-Mb *C. medica* introgression. Of the 11 citrons,

9 are pure representatives whereas the remaining 2 varieties present *C. micrantha* introgressions; large (19 Mb) and small (0.4 Mb) ones in Chr 6 and 4, respectively, for 'Chinese' citron; and a limited one (1.3 Mb) in Chr 4 for 'Etrog' citron s861. Of the horticultural groups resulting from admixture, sour orange displays a full *C. maxima*/*C. reticulata* heterozygosity all along the genome. Sweet orange and grapefruit present more complex admixture structures between the same ancestors, with heterozygous and homozygous regions. These 3 groups have a *C. maxima* chlorotype (Figure S23). 'Alemow' and 'Mexican' limes display similar phylogenomic structures

with complete *C. micrantha*/*C. medica* heterozygosity and *C. micrantha* chlorotype. The Volkamer lemon had a full *C. reticulata*/*C. medica* heterozygosity and a *C. reticulata* chlorotype. The Mediterranean lemon displays the more complex phylogenomic structure with a mosaic of *C. reticulata*/*C. medica* and *C. maxima*/*C. medica* heterozygous regions and a *C. maxima* chlorotype.

2.2.4 | Gene PAV According to Each Reference Genome

To analyse the intra- and interspecific diversity of PAVs, we adopted a methodology similar to the one used in the search for DSNPs. Regions of the representative accessions of each species found to be introgressed in previous admixture studies were excluded from the analysis. *Citrus medica* appeared to be the least polymorphic species, with only 1953 polymorphic genes, whereas *C. reticulata*, *C. maxima* and *C. micrantha* have 4577, 4691, and 3568 polymorphic genes, respectively (Table S19).

At the interspecific level, including the admixed horticultural groups, the percentages of core genes found in the CITRE, CITMA and CITMI genomes are very similar (67.6%, 67.3% and 67.4%, respectively) and slightly higher for CITME (71.7%). For each genome assembly, we identified genes that were specific to the corresponding species (i.e., found in all accessions of the species and absent from all accessions of the three other species) and genes specifically absent from one of the other ancestral species (absent in all accessions of one species and present in all accessions of the three other species). The number of these diagnostic genes (dPAV) varied from 1029 for CITMI to 1438 for CITME, with similar values for CITRE and CITMA (1158 and 1148, respectively). Several genes annotated in one of the four genome assemblies were missing in all accessions of another ancestor. *Citrus medica* is the species displaying the largest number of such missing genes in the three other species' genome assemblies (respectively, 527, 522 and 435 in CITMA, CITMI and CITRE; Table S19). The average number of genes missing for *C. reticulata*, *C. maxima* and *C. micrantha* in the genome assemblies of the three other species is 184 (average for CITMA, CITME and CITMI annotated genes), 109 (CITME, CITMI, and CITRE) and 139 (CITMA, CITME, and CITRE), respectively. CITME also has the largest number of specific genes of *C. medica* (990) whereas 497, 295 and 210 are found in CITRE, CITMA, and CITMI, to be specific, of *C. reticulata*, *C. maxima* and *C. micrantha*, respectively (Table S19). For the four reference genomes, the neighbour-joining (NJ) trees based on PAV display a strong differentiation between species (Figure S24) with increased branch lengths of the corresponding species, according to the contribution of specific genes to the interspecific differentiation.

Although the distribution of the number of diagnostic PAVs along the CITRE, CITMI and CITMA genomes is relatively homogenous, the CITME genome displays a significant peak of diagnostic PAV in the 17–18-Mb interval of Chr 4 (Figure S25). This region includes 97 dPAVs specific to *C. medica* out of 185 genes. Among the admixture horticultural groups, these genes are found in the direct hybrids of *C. medica* as the male parent

(such as 'Eureka' and 'Volkamer' lemons as well as 'Rangpur', 'Mexican' and 'Alemow' limes), but they are absent in the other admixed horticultural groups (Figure S26a). These dPAV genes are annotated as chloroplastic genes. A dot plot analysis of the chloroplastic genomes with CITME revealed homology with the entire chloroplast genome in this CITME Chr4 region, with several duplications (Figure S27a,e), whereas no homology was found between nuclear and chloroplast genomes for the three other genome assemblies (Figure S27b–d). Phylogenetic analysis of various genes with chloroplastic homology places the nuclear genes near the chloroplastic *C. medica* ones (Figure S28), suggesting an intraspecific introgression of the chloroplastic genome into Chr4 of *C. medica*. However, the nuclear orthologues are generally smaller than the original chloroplastic ones (with the exception of the NDHJ gene), displaying deletions (e.g., MatK gene in Figure S29).

To confirm the considered genes' nuclear location, we performed a new genotype calling with an additional natural or recent admixture of *C. medica* (first- and second-generation hybrids in which *C. medica* or its hybrids were the male parent). The dPAVs of the 17–18-Mb region of Chr 4 are found in all first-generation hybrids. For the second-generation hybrids, the dPAVs are found only in those that inherited the corresponding nuclear genomic region from *C. medica*, as evidenced by the interspecific mosaic analysis (Figure S30). Thus, the considered dPAV genes are absent in bergamot, clementine × 'Nestour' lime and clementine × 'Rangpur' lime, which have no *C. medica* nuclear contribution in this region of Chr 4. Additionally, all first- and second-generation hybrids inherited the chloroplast from their female parent (*C. reticulata* for 'Rough' and 'Volkamer' lemons and 'Rangpur' lime, clementine × 'Alemow', clementine × 'Nestour' lime, clementine × 'Volkamer' lemon and clementine × 'Rangpur' lime; *C. micrantha* for 'Alemow', 'Mexican' and 'Nestour' limes; *C. maxima* for 'Chandler' pummelo × 'Corsican' citron, 'Chandler' pummelo × 'Eureka' lemon and bergamot; Figure S26b). These results validated the nuclear location of the considered dPAV genes. In the same variant calling, we also expanded the diversity to other *Citrus* ancestral species, such as *C. ichangensis*, *C. indica*, *C. japonica*, *C. trifoliata*, *C. glauca*, *C. australis*, *C. australasica* and *Atalantia buxifolia*. The dPAVs of the Chr4 17–18 Mb were absent in all these species except *C. indica* (Figure S26a), which shared a phylogenetic branch with *C. medica* (Figure S26b). Twenty-six of the 97 dPAVs were present in *C. indica*.

2.3 | A Super-Pangenome for Cultivated Citrus

The cultivated citrus super-pangenome was established by combining the identification of orthologous genes among the four genome assemblies and the PAV information for 55 accessions on the same four genome assemblies (Figure 5).

2.3.1 | Search for Orthologous Genes Among the Four Genome Assemblies

The number of orthologous groups across the four assemblies is 33 520 (Table S20). The number of orthologous groups per

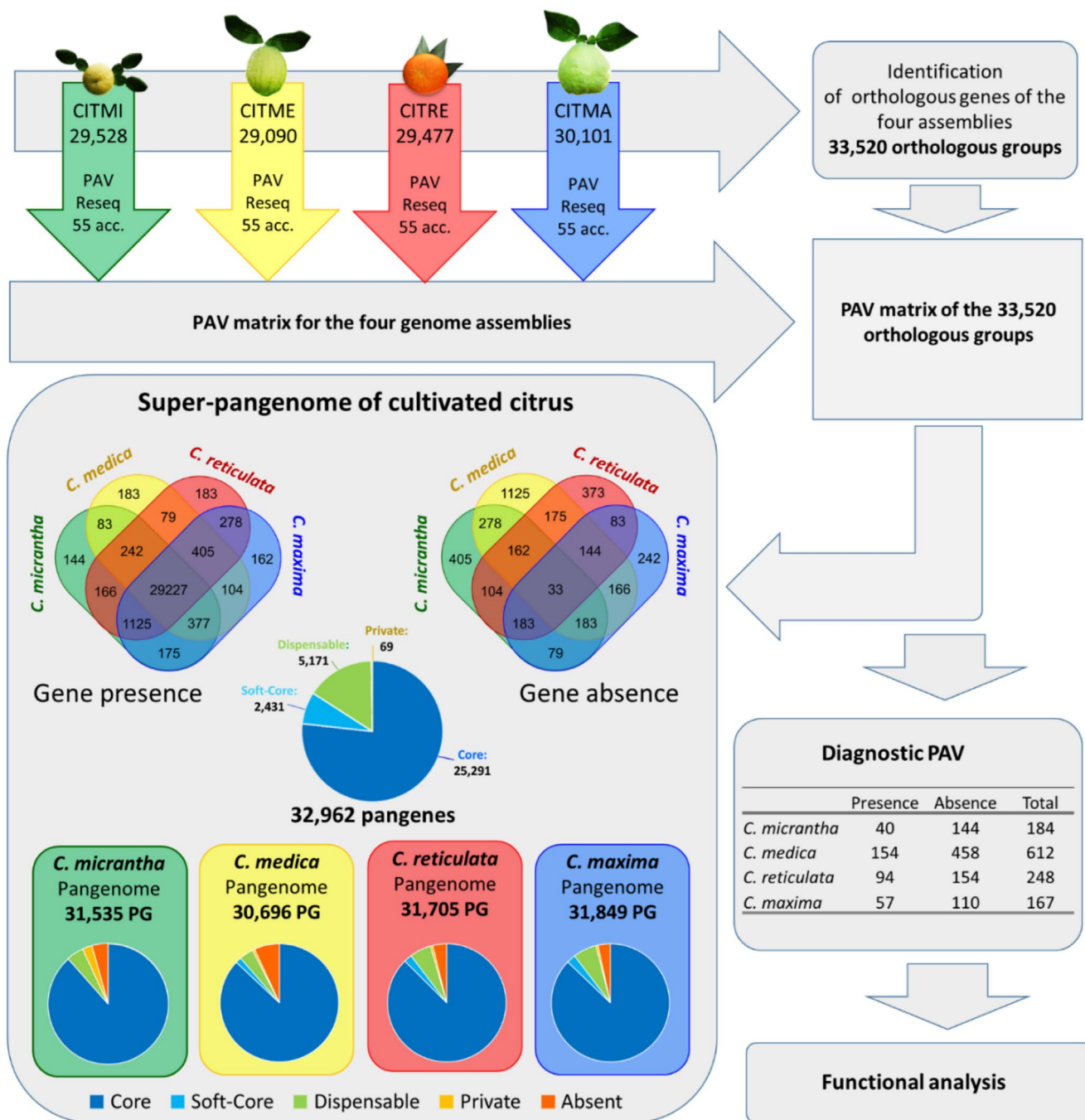


FIGURE 5 | Implementation strategy and structure of the cultivated citrus super-pangenome; acc.: Accession.

assembly ranges from 24,716 for CITMA to 25,322 for CITMI. The proportion of single copy genes by orthologous group varies from 88.98% for CITRE to 92.10% for CITME (Figure S31).

2.3.2 | PAV Analysis and Core/Dispensable Genome Identification

The PAV for the pangenes was established by combining PAV information from individual assemblies and the established orthology links (Figure 5). A pangenome was considered present in a specific accession if at least one of the corresponding orthologous genes was present in one of the four ancestral genome assemblies.

From the PAV analysis, 558 orthologous groups are found to be absent in all accessions. Therefore, the super-pangenome encompasses 32,962 pangenes (considering each orthologous group a pangenome). Of the pangenes, 33 are found only in some accessions of the admixed horticultural groups but not in the representative accessions of the four ancestral taxa. NJ tree analysis of the pangenome PAV data indicates a strong structuration among the four ancestral species (Figure 6a). Factorial analysis, based on the 55 accessions (Figure 6b), reveals the positions of admixed horticultural groups, which are fully consistent with their phylogenomic constitution.

The number of pangenes varies between 30,696 for *C. medica* and 31,849 for *C. maxima* (Table S21). The intraspecific

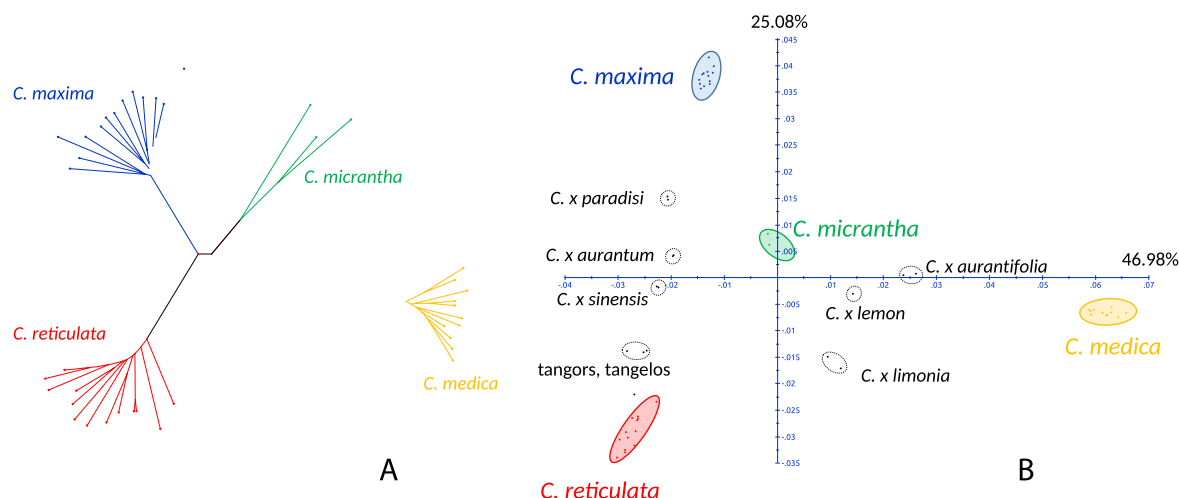


FIGURE 6 | Organisation of the PAV for 32962 validated pangenes. (A) Neighbour-joining tree of 43 accessions representative of the four ancestral species; (B) Factorial analysis of the 55 re-sequenced accessions.

core-gene frequency ranges from 90.2% for *C. maxima* to 93.5% for *C. medica* (Figure 5). At the interspecific level, the validated pangenes encompass 25291 core pangenes, 2431 soft-core genes, 5171 dispensable pangenes and 69 private genes (Tables S21 and S22). A total of 29227 pangenes are present in at least one accession of each of the four ancestral species (Figure 5).

A total of 1211 pangenes were found to be diagnostic for one of the four ancestral species (Figure 5; Table S23). Among them, the diagnostic pangenes for absences (866) are more frequent than those for presences (345). *Citrus medica* has the largest number of dPAVs, with 612 diagnostic pangenes.

2.4 | Interactomics and Functional Enrichment and Species-Specific Adaptations in the Citrus Pangenome

To elucidate the functional landscape of the core genome as well as of the four species-specific genomes in the citrus super-pangenome, protein-protein interaction (PPI) networks and functional enrichment were performed. A global network comprising all the proteins (core plus species-specific proteomes) was obtained and is available at https://string-db.org/cgi/download?sessionId=bOuaPkVaf3f9&species_text=STRG0A37MEC. Sub-PPI networks comprising species-specific proteins were obtained and analysed regarding (i) their interaction with the core genomes; (ii) their functionality (GO); and (iii) their centrality (Figure 7; Tables S24–S28). Globally, the four species-specific PPI networks contained 6 to 7 functional clusters with similar organisation. Some clusters did not show any functional enrichment (cluster 5 in all species, cluster 6 in CITMI) or no exclusive protein (cluster 6 in CITME) (Figure 7). Only CITRE contained 7 clusters. In the four species, cluster 1 mainly contained proteins related to transport and localisation processes, hormone signalisation, and membrane organisation, while cluster 2 contained proteins related to gene expression, macromolecules synthesis, metabolism and organisation, and flower and fruit development (Figure 7). Cluster 3 contained proteins involved in general metabolism of macro- and small molecules, photosynthesis

and energy production, as well as pigments and carotenoid biosynthetic processes. Cluster 4 contained proteins associated to response to oxidative stress and reactive oxygen species (ROS), and to flavonoid and phenylpropanoid metabolism (Figure 7). Cluster 6 (CITRE and CITMA) contained proteins related to cell wall metabolism and organisation, as well as dehiscence process (Figure 7c,d); interestingly, in CITRE the cell wall associated-proteins were present in both clusters 1 and 6 (Figure 7c). In CITMI, the cell wall associated-proteins were present also in the cluster 1 (Figure 7a). Proteins related to response to stress were present in different clusters according to the species: except in CITRE, they were present in the cluster 4 with some variations regarding the stress type (metals and environmental stress in CITME, toxins in CITMA), but also in cluster 2 in CITMI and in cluster 6 in CITRE (Figure 7). Regarding growth and development, cluster 1 contained proteins associated with cell or vegetative organ growth, except in CITMA; moreover, CITRE cluster 1 also contained proteins related to pollen formation and development—a process related to those present in cluster 2 in all the species (reproductive structure, flower, and fruit development). Cluster 4 of CITMI was the only one to contain proteins specifically related to jasmonic acid and oxylipin metabolism (Figure 7a). Finally, CITRE was the only species showing a seventh cluster containing proteins related to glucan and cellulose metabolism (Figure 7c). The analysis of protein centrality revealed that most of the proteins involved in the species-specific PPI networks were common proteins (Table S24). However, some hub (central), bottleneck (convergent) and hub-bottleneck (central and convergent) proteins could be highlighted. In CITMI, which presented stress response processes in two clusters (Figure 7a), two bottleneck proteins associated with disease resistance (NB-ARC, LRR, TIR) were found (Table S24). CITMI also showed a hub protein, RPB1, with DNA-dependent RNA polymerase (Table S24). CITMA showed specific biological processes associated with endo- and exocytosis that could be related to autophagy (Figure 7d, Table S28). Moreover, CITMA showed three hub proteins related to energy metabolism that could be highlighted (cytochrome *c* oxidase, cytochrome *c* asm, ATPase), as well as a wall-associated receptor kinase which could be related to the cluster 6 that in this species is strongly related to cell wall organisation, biogenesis and metabolism (Figure 7c,

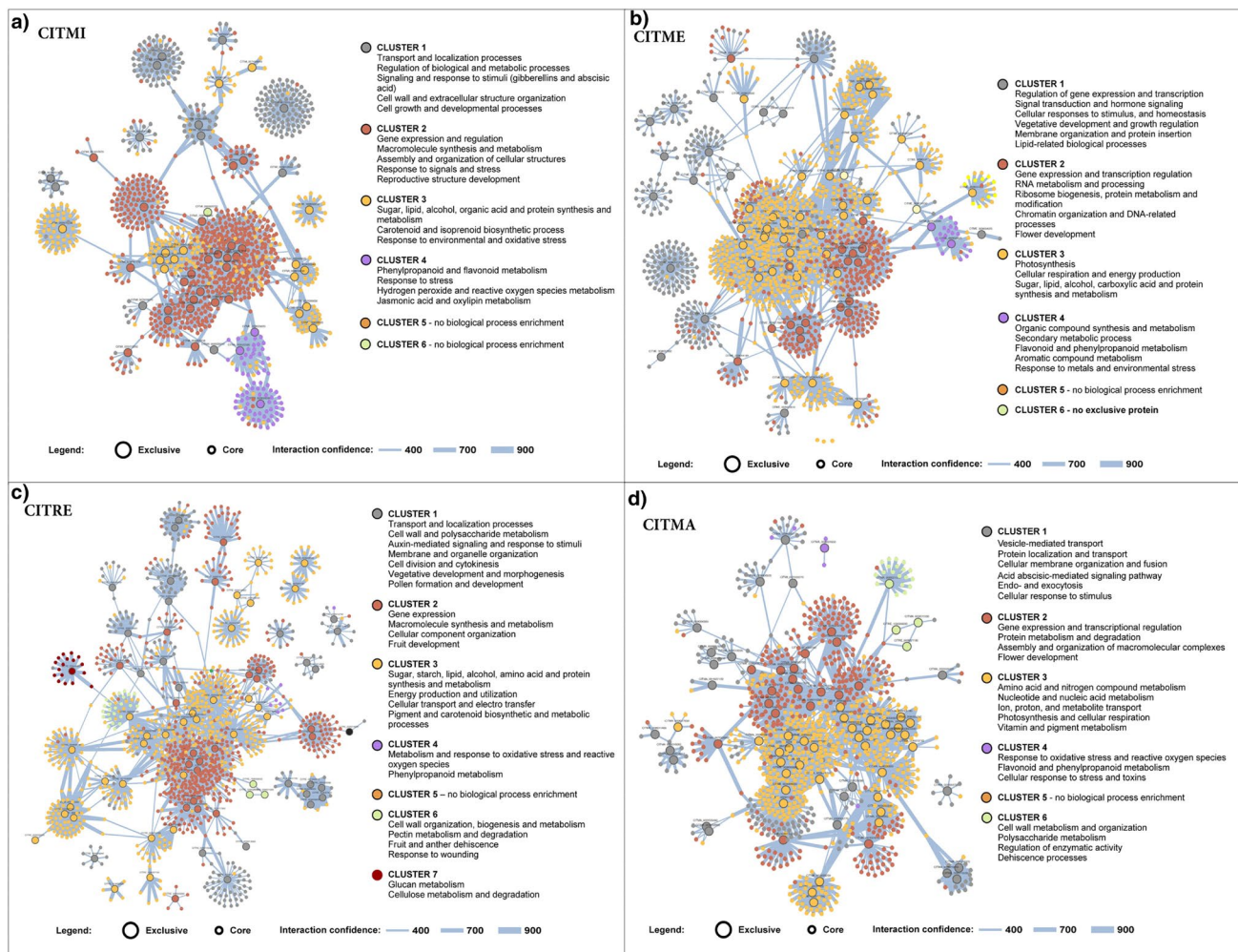


FIGURE 7 | Protein-protein interaction sub-network from (a) *C. micrantha*, (b) *C. medica*, (c) *C. reticulata* and (d) *C. maxima*. Specific clusters and their respective biological processes are indicated by different node colours (legend on the right of the figure). The size of the node indicates the *Citrus* species exclusive nodes (large nodes) vs. belonging to the core proteome (small nodes). The thickness of the node edges represents the interaction confidence level (400, 700 or 900).

Table S24). Similarly, CITRE showed a hub-bottleneck protein with exopolysaccharidase-like function that could be correlated to cell wall organisation, biogenesis and metabolism and/or glucan metabolism (clusters 6 and 7, Figure 7c). CITRE also showed three hub proteins related to pre-rRNA-processing and ribosome constitution (Table S24). CITME showed hub, hub-bottleneck and bottleneck proteins related to protein synthesis, tryptophan metabolism, membrane signalling and trafficking (PI3/P14 kinases, PLD) and energy metabolism and electron/proton transport, and homeostasis (Table S24), that could be correlated to a large exclusive protein amount from cluster 3 (Figure 7b).

3 | Discussion

3.1 | Chromosome-Scale Genome Assemblies for *C. micrantha*, *C. medica* and Pure *C. reticulata* Are Provided for the First Time

We produced for the first time a chromosome-scale genome assembly for *C. medica* and *C. micrantha*. A scaffold scale assembly for *C. medica* was previously published by Wang et al. (2017)

and no genomic resources except resequencing data were available for *C. micrantha*. The first chromosome-scale assemblies for mandarins have recently been released (Isobe et al. 2023). They concern the Satsuma mandarin and its two supposed parents, the 'Kunembo' and Kishu mandarins. The Satsuma and 'Kunembo' display very large introgressions of *C. maxima* (Oueslati et al. 2017; Wu et al. 2018) and therefore cannot be considered a pure representative of the ancestral *C. reticulata*. The 'Kishu' is a Japanese mandarin variety producing small fruits classified as *C. kinokuni* by Tanaka (1961), who considered it a different species from *C. reticulata*; however, its phylogenomic position remains to be clarified. In this work, we presented a chromosome-scale genome assembly for *C. reticulata* established with the Cleopatra mandarin that displays a pure *C. reticulata* phylogenomic ancestry all along its genome, without any interspecific introgression. The three new assemblies did not display any introgression; and therefore, CITRE, CITME, and CITMI are the first chromosome-scale assemblies of pure *C. reticulata*, *C. medica* and *C. micrantha* genomes, respectively.

These three new assemblies were annotated alongside a previously published chromosome-scale assembly of

C. maxima (Wang et al. 2017). 30101, 29477, 29258 and 28090 high-confidence protein-coding genes were identified respectively for the CITMA, CITRE, CITMI, and CITME genome assemblies. We focused on resistance genes. PRRs and/or R genes are widely distributed across plant genomes, playing crucial roles in recognising pathogen-associated molecular patterns (PAMPs) or effectors. This recognition triggers varying responses: PAMP-triggered immunity (PTI) or effector-triggered immunity (ETI) with different intensities and specificities (Jones and Dangl 2006; Katagiri and Tsuda 2010). In this study, these genes were identified in four ancestral citrus species: *C. medica* (1627 genes), *C. micrantha* (1899), *C. maxima* (1851), and *C. reticulata* (2014). Similar numbers were found in other species, such as in *C. sinensis* (2345), *C. clementina* (1617), *A. thaliana* (1756), and others (Santana Silva and Micheli 2020). The main receptors activating PTI are receptor-like kinases (RLKs), characterised by LRR, transmembrane, and kinase domains. Each species showed the following number of RLK genes—*C. medica* (88), *C. micrantha* (97), *C. maxima* (95), and *C. reticulata* (102)—comparable to other citrus and plant species (Santana Silva and Micheli 2020). For ETI, NBS-LRR genes, specifically CNL and TNL classes, are crucial. The numbers of CNL genes were 139, 182, 227, and 250 for *C. medica*, *C. micrantha*, *C. maxima*, and *C. reticulata*, respectively, whereas the numbers of TNL genes were 55, 73, 19, and 56 for *C. medica*, *C. micrantha*, *C. maxima*, and *C. reticulata*, respectively, with comparisons to other species showing similar distributions (Wang et al. 2015; Santana Silva and Micheli 2020). These genes tend to cluster on specific chromosomes across citrus species, such as Chr 3, 5, and 7 in *C. medica*, *C. micrantha*, and *C. reticulata* and Chr 2, 5, and 7 in *C. maxima*. High concentrations were also observed on specific chromosomes for RLKs (*C. medica* and *C. micrantha*), CNLs (*C. medica* and *C. reticulata*), and TNLs (all species). Additionally, gene duplication events were identified that contributed to genetic diversity and adaptive responses in citrus species. These events play a crucial role in evolutionary processes, potentially enhancing resistance and adaptation in response to environmental stresses (Qian and Zhang 2014; Panchy et al. 2016).

The annotation and the quantification of repeated sequences in the 4 *Citrus* ancestors revealed that despite their small sizes (from 332 to 347 Mb by haploid genome), almost half of each genome is composed of retrotransposons and DNA satellites (47.8% *C. micrantha*, 50.3% *C. reticulata*, 52.7% *C. medica*, 48.2% *C. maxima*). The larger number of *Gypsy Athila* and *Copia Sire* elements in *C. medica* compared with that in the other three species indicates a specific dynamic of TEs during *Citrus* radiation, particularly since the divergence of the four ancestors. The observed dynamic was congruent with previous studies that have shown the timing of retrotransposon activity and accumulation in *Citrus* species (Liu et al. 2019; Borredá et al. 2019). The number of TEs estimated in each reference genome appears to be representative of all ancestor varieties (Figure S11). However, minor variations identified between varieties reflect the more recent dynamics of TEs, which could explain the diversity of observed phenotypes in each *Citrus* species. In addition, we found that the TE density varied along the chromosomes, with more TEs on one side of Chr 5, 6, and 7 in CITMI, CITRE, and CITME. Similar distribution patterns were observed for the corresponding *C. maxima* chromosomes (Chr 1, 3, and 6, respectively). This was already observed in Chr 1, 3, 6 of *C. × sinensis*

(Xu et al. 2013) and in Chr 5, 6, and 7 of *C. clementina* (Wu et al. 2014). Aleza et al. (2015) located the centromere position in the nine clementine chromosomes by half-tetrad analysis and determined the acrocentric and submetacentric configuration of Chr 5, 6, and 7. The analysis of the recombination landscape of *C. reticulata* combined with gene density along the genome yielded the same conclusions. Moreover, a comparative genetic mapping study (Ollitrault et al. 2024) revealed a strong similarity between the recombination landscape along the genomes of several *Citrus* species and admixed horticultural groups, suggesting a conservation of the centromere location across species. Therefore, TEs were highly concentrated in pericentromeric regions in all chromosomes, but the position of centromeres varies across chromosomes, explaining our observations for TE density. Regarding the DNA satellite, the 181-bp satellite has previously been identified in *C. × limon*, *C. × sinensis*, *C. × aurantium*, *C. ichangensis* and *C. trifoliata* (Fann et al. 2001; Kang et al. 2008). Furthermore, fluorescence in situ hybridisation analyses conducted by Deng et al. (2019) confirmed its localisation in telomeric and subtelomeric regions across several *Citrus* species, such as *C. × clementina*, *C. × aurantium*, *C. × paradisi*, and *C. × tachibana*. This satellite sequence appears to be conserved during the diversification of the *Citrus* genus.

3.2 | Synteny and Collinearity of the Ancestral Taxa of Cultivated Citrus

The synteny and collinearity are high among the three new genome assemblies. The main structural variation concerns an inversion in the centromeric region of Chr 2 of CITME when compared with CITMI and CITRE. In addition, dot-plot analysis reveals important synteny and collinearity of these three genomes, with five citrus genome assemblies recently published (Peng et al. 2020; Wang, Song, et al. 2022; Wu et al. 2022; Nakandala et al. 2023). Conversely, the *C. maxima* assembly (Wang et al. 2017) published displays numerous differences from the three new assemblies and other published citrus genomes. A recent comparative genetic study at the interspecific level (Ollitrault et al. 2024) revealed a large structural conservation of the genome of the various *Citrus* species, including *C. maxima*. The same authors also suggested probable local assembly mistakes in the *C. maxima* genome published by Wang et al. (2017) that displayed numerous differences from their *C. maxima* genetic map. The genome structural conservation across cultivated *Citrus* species is a favourable situation for further interspecific sexual breeding and for translational genomics.

3.3 | Introgressions and Admixtures in Cultivated Citrus

Regardless of which reference genome is used as a template, the general conclusions from the interspecific mosaic analyses were congruent. For mandarins, they confirm the previous observation of frequent introgressions of *C. maxima* genome in a *C. reticulata* background (Oueslati et al. 2017; Wu et al. 2018) and the probable role of *C. maxima* introgression in the domestication of mandarins (Wu et al. 2018). The small *C. reticulata* introgressions observed in Chr 2 of 'Chandler' and 'Kao pan' pummelos by Wu et al. (2018) are confirmed, suggesting a kinship between

these two varieties. We also identified small *C. reticulata* introgressions in Chr 9 for ‘Guil’ and ‘Huan’ pummelos, whereas a 0–5Mb *C. medica* introgression was identified in the ‘Timor’ pummelo. Although previous studies (Wu et al. 2018; Ahmed et al. 2019) did not show interspecific introgression in *C. medica* accessions, here we revealed *C. micrantha* introgression in two citron varieties (‘Chinese’ and ‘Etrog e861’). The mosaic structure of the horticultural groups resulting from admixture as well as the assignment of cytoplasmic genomes corroborate previous results obtained using *C. clementina* V1.0 as a template genome (Wu et al. 2018) for sweet and sour oranges, grapefruits, Mediterranean lemons and Mexican limes. In addition, we confirmed the origin of the ‘Volkamer’ lemon and ‘Alemow’ lime from direct interspecific hybridisations between *C. reticulata* and *C. medica* on one side and *C. micrantha* and *C. medica* on the other side, as GBS data suggested (Ahmed et al. 2019). The two lemons and the two limes inherited, respectively, the *C. reticulata* and *C. micrantha* chloroplast, as previously proposed from cytoplasmic markers analysis (Curk et al. 2016). The contribution of *C. medica* as the male parent for these four interspecific hybrids aligned with the cleistogamy of this species. Introgression could have played an important role in citrus domestication. It would be interesting to conduct in-depth studies on the impact of ancestral species-specific variants (DSNPs, Indels, dPAV) involved in the introgressions in the cultivated citrus.

3.4 | Integration of Plastid Genome Sequences in Chromosome 4 of *C. medica* Ancestor and Its Conservation in Hybrid/Admixed Progenies

Our results revealed a complete chloroplast sequence insertion with several duplicated regions in the 17–18-Mb region of the Chr 4 of CITME genome assembly, whereas no evidence of large chloroplast sequence insertion was found in the genome assemblies of the three other ancestors of cultivated citrus. Moreover, this region's chloroplastic genes were found in all *C. medica* accessions, in first-generation hybrids having *C. medica* as the male parent and in second-generation hybrids having inherited *C. medica* genome in the considered region, validating the nuclear location of these genes of chloroplast origin. Part of these genes was also found in *C. indica*, an Indian species sharing a monophyletic origin with *C. medica*. Therefore, the considered nuclear insertion may have occurred after the first radiation of the Asian *Citrus* species (6–8 million years ago; Wu et al. 2018) and before the secondary radiation between *C. indica* and *C. maxima* (Figure S26b). Nuclear plastid DNA sequences (NUPTs) are frequently found in plant genomes, and it is accepted that the transfer of organellar DNA into the nuclear genome occurs continuously and contributes to genome structure and evolution (Leister and Kleine 2011; Zhang et al. 2020). Plastid genes generally show very low or null expression in the nucleus because they often lack the regulatory element for gene expression in the nucleus or do not encode for completely open reading frames (Zhao et al. 2019; Zhang et al. 2020). Furthermore, epigenetic regulation has also been associated with the transcriptional repression of integrated chloroplast genes (Yoshida et al. 2019). In the absence of functional implication, the evolution of NUPTs may be neutral and even selected against (Marczuk-Rojas et al. 2024). This is consistent with the fact that all plastid genes inserted in

the Chr 4 of the CITME genome, except NDHJ, display large deletions compared with the original chloroplast gene sequences. Moreover, the expression of the integrated plastid genes appears insignificant based on our RNASeq data of the ‘Corsica’ citron (data not shown). The 17–18-Mb region of CITME Chr4—in which the large plastid insertion is located—was included in a genomic area with very low gene sequence frequency and a high repeat-element level that probably corresponds to the centromeric region of Chr 4. Such insertion of plastid sequence in the centromere's vicinity has been described in Arabidopsis and rice genomes (Michalovova et al. 2013). In a pioneer study of whole-chloroplast genome diversity based on WGS data (paired-end reads) of 34 citrus accessions mapped against the *C. sinensis* chloroplast genome (Bausher et al. 2006; Carbonell-Caballero et al. 2015) found frequent heteroplasmy in the *Citrus* genus after interspecific hybridisation. Their conclusions were based on the observation of SNPs with strong preeminent alleles and low frequency alleles, corresponding to those of the female and the male parents, respectively. The main gene pool sustaining their hypothesis was *C. medica* and F1 interspecific hybrids with *C. medica* as the male parent. According to our results, it is more likely that the observations of Carbonell-Caballero et al. (2015) for this gene pool resulted from the nuclear integration of the large *C. medica* plastid sequence in the *C. medica* nuclear genome. Our hypothesis is in agreement with the conclusion of Wang, Li, et al. (2022) for strict maternal inheritance of the chloroplast genome in citrus.

3.5 | A Super-Pangenome for Cultivated Citrus

The implemented super-pangenome of cultivated citrus integrates the identification of orthologous genes of the four genome de novo assemblies and PAVs from resequencing data of 55 accessions. This analysis revealed 32 962 pangenes, 25 291 classified as core, 2 431 as soft-Core, 5 171 as dispensable and 69 as private, reflecting intricate genomic complexity and evolutionary history of citrus. The strong genetic structuration observed among the ancestral species and the alignment of horticultural groups with their phylogenomic backgrounds underscored this pangenome's value in our understanding of the genetic diversity underlying modern cultivated citrus. Among 1 211 pangenes were found to be diagnostic for one of the four ancestral species, and specific absences (866) were more frequent than specific presences (345). Of the four ancestral taxa, *C. medica* displays the largest number of dPAV (612 pangenes). The comparison of the unique gene enrichments in each species with the core genome revealed distinct functional specialisations that complement the conserved processes shared across the pangenome. Whereas the core genome is enriched in broad, essential functions, such as transcription regulation, stress response and protein turnover (Tranchant-Dubreuil et al. 2019), individual species display unique adaptations that enhance their survival and reproduction in specific environments. *Citrus reticulata*'s enrichment of auxin signalling and RNA repair likely provides refined growth control and stress resilience (Aslam et al. 2020; Jing et al. 2023) whereas *C. maxima*'s emphasis on autophagy and energy metabolism suggests robust mechanisms for maintaining cellular homeostasis (Magen et al. 2022). The absence of certain biosynthetic pathways in *C. medica* contrasts with the core genome, possibly reflecting a trade-off or specialisation in

other areas. Meanwhile, *C. micrantha* enrichment in antiviral response and RNA processing indicates a specialised defence mechanism (Yang and Li 2018), aligning with the core genome's general stress response but with a more targeted focus. These differences highlight the evolutionary paths that have shaped each species, contributing to the diversity in the citrus super-pangenome while maintaining a shared foundation of essential biological processes.

3.6 | Functional Protein Enrichment Allowed the Identification of Central and Species-Specific Proteins

PPI network analysis allowed the identification of specific and significant enrichment in several critical processes present in all the species (core) such as gene expression and transcription regulation, stress response mechanisms, macro- and small molecules metabolism, vegetative but also reproductive growth and development, among others (Figure 7; Tables S24–S28). These functions underscore the core genome's essential roles in maintaining cellular homeostasis, regulating gene expression and ensuring successful development and reproduction across these citrus species. However, variations in cluster organisation and PPI between species may reveal differences in specific related processes and help identify key proteins involved in one or more functional pathways. For example, the disease resistance proteins containing the NB-ARC domain found in CITMI and identified as bottlenecks (Figure 7, Table S24) could be correlated to the large duplication events in CNL and RLK PRR proteins also observed in this species (Figure 4). In CITMI, the interplay between ROS metabolism, jasmonic acid, and oxylipins forms a complex regulatory network that enables plants to perceive, respond to, and survive stress conditions. This balance is critical for plant growth, defence, and adaptation in fluctuating environments (Figure 7). CITMA exhibited enrichment in autophagy and energy metabolism, as well as specific cytochrome c hub proteins, reflecting its robust mechanisms for cellular quality control and efficient energy management. These processes are essential for degrading faulty mRNAs, recycling cellular components and sustaining energy production, which are vital under stress. CITRE showed specific proteins related to the auxin-mediated signalling pathway, RNA processing and macromolecule transport and localisation, with potential impact on growth regulation, maintaining RNA integrity and precise protein trafficking. These functions are crucial for responding to environmental stimuli, ensuring accurate protein synthesis under stress and maintaining cellular organisation. CITME PPI network showed a very large number of exclusive proteins in cluster 3 that specifically contained the photosynthesis biological process (Figure 7). This finding could be directly related to the identification of a large chloroplast DNA insertion in the 17–18 Mb region of Chr 4 in the CITME genome, absent in other citrus ancestors (Figures S25 and S26).

3.7 | Genome-Hub for the Cultivated Citrus Super-Pangenome

The Citrus Genome Hub (CGH, <https://citrus-genome-hub.southgreen.fr>) was developed to support post-genomics efforts

by centralising citrus genomic information. It offers a set of user-friendly interfaces for the analysis and visualisation of genomic data, providing a comprehensive platform for researchers to explore and interpret citrus genomes efficiently. Users can search for genes in two ways, by querying using various filters, such as name, putative function, chromosomal position and functional annotation (e.g., putative function, InterPro domain, gene ontology) and by sequence similarity using Blast. In both methods, results are displayed in a dynamic table that summarises the relevant information about the search. Each entry includes a link to the gene report and the genome browser (Jbrowse), providing easy access to detailed gene information and visualisations. Comparative analysis at the genome scale is supported by Synvisio, an interactive multiscale synteny visualisation tool. This feature allows users to explore syntenic relationships between genomes at various scales, facilitating a detailed understanding of genomic structural variation and conservation across species. A download section provides users with direct access to the data used by the various tools that compose the hub.

4 | Conclusion

To implement a super-pangenome for cultivated *Citrus*, we performed de novo chromosome scale assemblies for *C. medica* (346 Mb), *C. micrantha* (333 Mb) and *C. reticulata* (332 Mb) for the first time. These assemblies were annotated for genes and repetitive elements, alongside a previously published assembly of *C. maxima*. We identified 28 090, 29 258, 29 477 and 30 101 genes in the respective genomes, with 1627, 1899, 2014 and 1851 of those genes categorised as PRRs and/or R genes. Despite their relatively small sizes, nearly half of each genome consists of transposable elements or DNA satellites. Strong synteny and collinearity were observed among the three new genome assemblies whereas significant discrepancies were observed with the *C. maxima* one.

Intra- and interspecific diversity were explored by resequencing data from 55 citrus accessions. The phylogenomic structure along the genome of these accessions was revealed, highlighting interspecific introgressions in several representative accessions of *C. reticulata*, *C. maxima* and *C. medica* as well as the origin of horticultural groups shaped by reticulate evolution. A super-pangenome for cultivated *Citrus* was established by combining orthologous gene identification across the four assemblies and gene PAV based on resequencing data. This analysis revealed significant inter- and intraspecific variations, including 25 291 core genes, 2431 soft-core genes and 5171 dispensable genes. PAV analysis also revealed genes with presence or absence specific to particular ancestral taxa. In particular, these dPAVs revealed a large chloroplastic introgression on chromosome 4 of *C. medica*, which is inherited by horticultural groups and recent hybrids having this species as the male parent. Comparing the unique gene enrichments or those lost in each species with the core genome revealed distinct functional specialisations that complement the conserved processes shared across the pangenome. Whereas the core genome is characterised by broad, essential functions, such as transcription regulation, stress response and protein turnover, individual species exhibit unique adaptations enhancing their survival and reproduction in specific environments.

We established a dedicated Genome Hub to provide a platform for ongoing genomic research, allowing for continuous updates and the future inclusion of additional ancestral species. This super-pangenome offers a comprehensive resource that captures the vast genetic diversity in modern cultivated citrus. Together with specialised tools, it will enhance the efficiency of association genetics and transcriptomics studies, expanding our understanding of the molecular basis of phenotypic diversity in cultivated citrus. Ultimately, this resource will contribute to breeding programs focused on sustainable citriculture.

5 | Experimental Procedures

5.1 | Genomic Sequences of Citrus Accessions

5.1.1 | Germplasm Accession Used for Sequencing and Optical Mapping for De Novo Assembly of Three Accessions

The genome assemblies of *C. reticulata*, *C. medica* and *C. micrantha* were implemented respectively from the ‘Cleopatra’ mandarin (ICVN-0110274), ‘Corsican’ citron (SRA-613), and ‘Biasong’ (SRA-1115) localised on the Corsican field collection hosted by the Citrus Biological Resource Center (Citrus BCR-INRAE-CIRAD, San Giuliano, Corsica, France) (Luro et al. 2017).

5.1.2 | Resequencing of Citrus Accessions

For variant calling on the four genome assemblies, a set of 40 citrus varieties was re-sequenced (Table S1). This set included 30 representatives of the four ancestral species and 10 representatives of admixed horticultural groups. Additionally, for the analysis of chloroplastic sequences, we re-sequenced 6 accessions representing 5 additional pure species, 3 admixed varieties, and 6 controlled hybrids. All these accessions were also sourced from the Corsican field collection hosted by the Citrus BRC (INRAE-CIRAD) and Cirad-INRAE breeding program in San Giuliano, Corsica, France. Total DNA was extracted using the Plant DNAeasy kit (Qiagen), following the manufacturer's instructions. Genomic libraries were prepared according to standardised Illumina protocols for plant DNA and pair-end sequences (2×150) using an Illumina HiSeq 4000 platform at Genoscope, with an average coverage of 30× per accession. Demultiplexing, quality filtering, and adapter removal were carried out according to the standardised protocol at Genoscope.

5.2 | De Novo Nuclear Genome Assembly

The de novo genome assemblies were obtained by combining nanopore long-read sequencing, Illumina PCR-Free sequencing and bionano optical mapping.

5.2.1 | PromethION Library Preparation and Sequencing

The libraries were prepared according to the ‘1D Genomic DNA by ligation (Kit 9 chemistry)’ protocol, provided by Oxford

Nanopore. Genomic DNA was first repaired and end-prepped with the NEBNext FFPE Repair Mix (New England Biolabs, Ipswich, MA, USA) and the NEBNext Ultra II End Repair/dA-Tailing Module (NEB). DNA was then purified with AMPure XP beads (Beckmann Coulter, Brea, CA, USA), and sequencing adapters provided by Oxford Nanopore Technologies were ligated using Concentrated T4 DNA Ligase 2M U/mL (NEB). After purification with AMPure XP beads (Beckman Coulter) using dilution buffer (ONT) and wash buffer (ONT), the library was mixed with the sequencing buffer (ONT) and the Library Loading Bead (ONT) and loaded on the PromethION R9.4.1 Flow Cells. The nanopore long reads were not cleaned, and raw reads were used for genome assembly.

5.2.2 | Illumina PCR-Free Library Preparation and Sequencing

PCR free libraries were prepared using the Kapa Hyper Prep Kit (Roche, Basel, Switzerland). Briefly, genomic DNA (1–1.5 µg) was sonicated using a Covaris E220 sonicator (Covaris, Woburn, MA, USA). Fragments were end-repaired and 3'-adenylated, and Illumina adapters (Bioo Scientific, Austin, TX, USA) were ligated following the manufacturer's instructions. Ligation products were purified twice with AMPure XP beads (Beckman Coulter Genomics, Danvers, MA, USA) and quantified by qPCR (MxPro, Agilent Technologies, Santa Clara, CA, USA) using the KAPA Library Quantification Kit for Illumina Libraries (Roche). The library profile was assessed using an Agilent High Sensitivity DNA kit on the Agilent 2100 Bioanalyser. Libraries were paired-end sequenced on an Illumina HiSeq2500 instrument (Illumina, San Diego, CA, USA) using 250 base-length read chemistry. After the Illumina sequencing, an in-house quality control process was applied to the reads that passed the Illumina quality filters. The first step discards low-quality nucleotides ($Q < 20$) from both ends of the reads. Next, Illumina sequencing adapters and primer sequences were removed from the reads. Then, reads shorter than 30 nucleotides after trimming were discarded. These trimming and removal steps were completed using in-house-designed software based on the FastX package (Engelen and Aury; <https://www.genoscope.cns.fr/fastxtend>). The last step is to identify and discard read pairs that are mapped to the phage phiX genome, using SOAP aligner (Li et al. 2008), and the Enterobacteria phage PhiX174 reference sequence (GenBank: NC_001422.1). This processing, described in Alberti et al. (2020), resulting in high-quality data.

5.2.3 | Optical Mapping

The direct label and stain labelling (using the DLE-1 enzyme) and the nick label repair and stain (NLRS) labelling (using the BspQI enzyme) protocols and Bionano Genomics protocols were followed with 750 and 600 ng of DNA, respectively. The chip loadings were performed as Bionano Genomics recommended.

5.2.4 | Long-Read-Based Genome Assembly

First, we estimated each organism's genome size and heterozygosity rate using Genomescope2.0 (Ranallo-Benavidez

et al. 2020) with the Illumina reads as inputs. Using these values, we launched the Necat assembler (Chen et al. 2021) (git commit 0c5fc85) with the appropriate genome size and with all Nanopore reads as inputs. Each assembly was then polished two times with HAPLO-G (Aury and Istace 2021) and Illumina reads.

5.2.5 | Chromosome Sequence Reconstruction

Hybrid scaffolding: the BspQI and DLE-1 preparations were run on a single flow cell each. The generated molecules were assembled to produce optical maps using software provided by Bionano Genomics with the following two options: ‘add pre-assembly’ and ‘non haplotype without extend and split’ (Bionano Solve and tools Version: 3.3_10252018). The two optical maps and the ONT contigs were subjected to the 2-enzyme hybrid scaffolding pipeline to generate the hybrid scaffolds. BisCoT (Istace et al. 2020) was used to correct artifactual duplications (negative gaps) introduced during the scaffolding process.

Pseudochromosome construction: for the ‘Cleopatra’ mandarin and ‘Corsican’ citron, the final genome assembly in pseudochromosomes was guided by anchoring their scaffolds to the genetic maps of the same varieties (Ollitrault et al. 2024), using the locOnRef tool from the Scaffhunter toolbox (Martin et al. 2016). For *C. micrantha*, for which no genetic map is available, the final assembly was guided by the synteny and collinearity with the *C. clementina* V1.0 pseudochromosome assembly (https://phytozome-next.jgi.doe.gov/info/Cclementina_v1_0). Chromosome numbering and orientation of the three de novo genomes are identical to those of *C. clementina* V1.0 (https://phytozome-next.jgi.doe.gov/info/Cclementina_v1_0) and *P. trifoliata* V1.3.1 (https://phytozome-next.jgi.doe.gov/info/Ptrifoliata_v1_3_1). Unassembled scaffolds derived from cytoplasmic genomes were removed after alignment on assembled chloroplast sequences (see next section) and mitochondrial sequence (NC_037463.1) using blastn (evalue < 10e-6, match on 80% of the sequence).

5.3 | Chloroplast Genome Assembly

Raw WGS-paired Illumina HiSeq 4000 sequences of ‘Corsican’ citron, *C. micrantha* and ‘Cleopatra’ mandarins were treated with fastQC v0.11.7, and de novo assembly of CITME, CITMI and CITRE chloroplastic genomes was performed with NovoPlast v4.3.1 (Dierckxsens et al. 2017), using the default parameters and the Pineapple sweet orange chloroplastic genome NC_008334.1 as an assembly guide. The assembled circular sequence was annotated with GeSeq v2.03 (Dec 18, 2020) using default parameters. A phylogeny was realised with these three new assembled chloroplastic genomes and was completed with 23 already published assembled genomes of Rutaceae, 21 *Aurantioideae* species (Shi, Song, Liu, et al. 2023; Shi, Song, Zhao, et al. 2023) and 2 outgroup species (*Zanthoxylum madagascariense* NC_046744 and *Zanthoxylum asiaticum* NC_056094). All sequences were aligned with MAFFT v7.045 (Katoh and Standley 2013), and maximum likelihood analysis was conducted with phyML 3.0 (Guindon et al. 2010) after selection of the best evolutionary model by SMS (Lefort et al. 2017).

5.4 | Annotation of the Three De Novo Genomes and the Chinese *C. maxima* Assemblies

To compare our three de novo assemblies and the *C. maxima* (HWB.v1.0) available at <http://citrus.hzau.edu.cn/download.php>, we annotated the four genome assemblies following the procedure described below.

5.4.1 | Annotation of Genes

5.4.1.1 | RNA-Seq Experiment and Transcriptome Assembly.

Three replicates of young leaves, mature leaves and flowers from the Cleopatra mandarin (ICVN-0110274), Corsican citron (SRA-613), Chandler pummelo (SRA-608) and *C. micrantha* (SRA-1115) were harvested on adult trees of the Corsican field collection hosted by the Citrus Biological Resource Center (Citrus BCR) (INRAE-CIRAD, San Giuliano, Corsica, France). Three root replicates from the same genotypes were taken from cuttings in greenhouses. RNA was extracted using the NucleoMag kit (Macherey-Nagel GmbH & Co.KG, Düren, Germany) with the Kingfisher automated system (KingFisher Flex Purification System, ThermoFisher Scientific, Waltham, MA) according to the manufacturer's protocol. RNA integrity and quality were checked on TapeStation, screentape D5000. RNAseq libraries were prepared to obtain labelled and matched sequences using UDI indexes, following the protocol and recommendations from the Illumina TruSeq kit. The libraries' integrity and quality were checked on TapeStation, screentape D5000, and assayed by qPCR. An equimolar pool of the libraries was then built up using the TruSeq Stranded mRNA Sample Preparation kit from Illumina. Pair-end sequencing (2×150 nt) was performed on an Illumina HiSeq 4000 platform at Genewiz. Clean RNA-seq reads, which had been filtered by Trimmomatic (v0.39), were mapped onto their respective assemblies using Hisat2 (v2.1.0) with default parameters. Each of these files was then coordinate-sorted and converted to BAM format. The resulting BAM file was used for genome-guided transcriptome assembly using Trinity.

5.4.1.2 | Prediction of Protein-Coding Gene Models.

Automatic gene prediction was performed using the EuGene Eukaryotic Pipeline (EGNEP v1.5 with EuGene v4.2a), an integrative gene finder software that can combine several sources of information to predict genes (Sallet et al. 2019).

A EuGene internal prediction model can be built and trained using proteomic and transcriptomic data. Therefore, a set of protein-coding gene models of *C. maxima* and *C. medica* proteins (Wang et al. 2017) as well as the UniProt/SwissProt Viridiplantae database were used for the proteomic evidence, with the prior exclusion of proteins that were similar to those present in RepBase (v. 20.05). The CIRAD obtained an institutional licence from <https://phoenixbioinfo.org/rebase> (ref: 00Do0J6b5_5001J10Do2W:ref: 07/23/2021-07/22/2024, access extended to 08/11/2024) to use the Repbase database (<https://www.girinst.org/rebase>). Two datasets of transcribed sequences were aligned on the genome and used by EuGene as transcription evidence: (i) the respective genome-guided transcriptome assembly produced by Trinity and (ii) a collection of more than 560 000 ESTs retrieved from

the NCBI database using the query “Citrus”[Organism] AND plants[filter] AND biomol_mrna[PROP] AND is_est[filter]. Finally, ab initio was predicted using Augustus v3.3.2. To enhance prediction accuracy, a custom statistical model for splice-site detection was applied, leveraging information from both transcriptomic assemblies, using the *egn_build_wam.pl* companion script. Additionally, tRNAscan-SE was used to predict tRNA genes. Other non-coding RNAs, such as rRNA, miRNA and snRNA, were annotated using INFERNAL v1.1.2 through a search against the Rfam database v9.1.

5.4.1.3 | Functional Annotation of Protein-Coding Genes. To identify gene functions, we first used Diamond (v2.0.11) to annotate protein-coding genes’ functions based on items in the Swiss-Prot and TrEMBL databases (v2022-06-07; <https://www.uniprot.org>). Additionally, structural domains and GO annotations of predicted protein-coding genes were generated using InterProScan (v5.48-83.0). KEGG orthology (KO) terms were assigned through searches against the hidden Markov model (Kofam) database using KofamScan v1.3.036 with default parameters.

5.4.1.4 | Annotation of PRR. Domain annotation of the PRR proteins from the four genomes was performed using InterProScan 5.48–83.0 (Jones et al. 2014) with the default parameters. Subsequently, the RRGPredictor tool was employed to identify and classify the proteins into 13 classes (T, TN, TNL, RPW8NL, RLP, RLK, RLKGNK2, N, NL, CN, CNL, MLO and UNKNOWN) according to the methodology (Santana Silva and Micheli 2020) described. The UNKNOWN class was used only to estimate the global PRR quantity and was excluded from other analyses. The genes on the chromosomes were graphically visualised using the TBtools tool (Chen et al. 2020). Gene duplication analysis for each PRR class and each genome was conducted using the MCScanX tool (Wang et al. 2012) with the default parameters. The search for PRR gene clusters was carried out by subdividing this analysis into two steps. First, clusters with three or more genes were identified following (Holub 2001) methodology for genes included in a region smaller than or equal to 200 kb. Then, for the remaining genes, the clusters of two genes were identified following the methodology of Jupe et al. (2012).

5.4.2 | Annotation of Repeated Sequences in Genomes

The number of repeated sequences was estimated using the *Citrus* repeat databank, developed by Giraud et al. (2025). This bank was built from the genome assemblies and contained all the diversity of repeats found in the four ancestral species of cultivated citrus, including diversity of TEs and DNA satellites. The TEannot pipeline from the REPET package was used for annotating the four studied genomes (Quesneville et al. 2005; Flutre et al. 2011; Hoede et al. 2014). This pipeline aligns consensus of repeated sequences against the four reference genome sequences, using such tools as NCBI-Blast, RepeatMasker, CENSOR, BLASTER, SSR annotation with TRF, Mreps and RMSSR. Matches with a size larger than 100 bp and with at least 80% homology with consensus sequences (following the Wicker rule) were considered in the final annotation. Additionally, intraspecific variability analyses were conducted to confirm that

the quantity of repeated sequences annotated in assembled genomes was representative of the diversity of repeated sequences found in all the varieties of the four *Citrus* species ancestors. For this purpose, raw Illumina sequencing libraries from 10 varieties for each ancestor were aligned against the *Citrus* repeat databank using Bowtie2 (Langmead and Salzberg 2012) (settings parameters: very sensitive local alignment, considered not-paired reads). The 10 varieties selected by ancestor (only four varieties for *C. micrantha*) were chosen because the phylogenomic analysis showed they were pure representatives of the ancestors (Table S1). The Illumina libraries used to assemble the reference genomes were included in the analyses. Finally, to enhance gene annotations, predicted genes showing sequence homology with repeated sequences were removed. Genes were considered repeated sequences when more than 50% of their sequences were aligned with one repeated sequence (comparison of gene and repeat positions) and when they contained a coding protein domain associated with repeated sequences.

5.4.3 | Syntenic Gene Alignments

We estimated orthology and synteny between protein-coding genes for the 4 taxa using the R package GENESPACE v1.1.8 (Lovell et al. 2022). This package leverages MCScanX (Wang et al. 2012) to infer syntenic gene blocks and incorporates OrthoFinder v2.5.4 (Emms and Kelly 2019) and DIAMOND v2.1.4.158 (Buchfink et al. 2021) to identify orthogroups in these syntenic blocks. This approach allowed us to analyse conserved gene relationships and structural genomic features across the taxa.

5.5 | Variant Calling From Resequencing Data

The re-sequencing data from 40 new accessions generated for this study were combined with raw data from NCBI for 15 additional accessions (Table S1). Of these, 11 were previously published by Wang et al. (2017, 2018) and 4 by Wu et al. (2014, 2018). A dedicated pipeline was developed using Snakemake to handle sequencing data (<https://gitlab.cirad.fr/agap/workflows/GATK>). This pipeline automates various steps of data handling, ensuring reproducibility and scalability for large-scale genomic analyses. The raw data were cleaned using fastp (v 0.20.1). Variant calling was performed with GATK4 (<https://github.com/gatk-workflows>), using a standardised procedure and parameters, including the following steps: BWA alignment, SortSam, Mark Duplicate, HAPloTypeCaller0, Select-Variant, Variant-Filtration, Exclude-Filtered, BaseRecalibrator, ApplyBQSR, HaplotypeCaller1, GenomicsDB and GenotypeGVCFs. Four reference genome assemblies were used for variant calling: the three new ones released in this paper and the previously published HWB.v1.0 assembly for *C. maxima* (Wang et al. 2017), called CITMA in this work.

5.6 | Admixture Analyses

The detailed method for interspecific mosaic analysis is provided in Text S1 and Figures S16 and S17.

5.6.1 | Identification of Ancestral Species' Diagnostic SNPs

We applied the two-step method (Ahmed et al. 2019) proposed to identify diallelic DSNPs for each of the four ancestral species. Among the accessions classified as *C. reticulata*, we discarded clementine, 'Satsuma' and 'King' mandarins, found to be highly introgressed by *C. maxima*. The identification of DSNPs was therefore based on 11 citrons, 13 mandarins, 13 pummelos and 3 papedas (Table S1). The first step of the DSNP search was intended to identify introgressions on the representative cultivars. It relied on the heterozygosity level and each accession's similarity to the centroid of each ancestral taxon along the genome. SNPs with a differentiation parameter (GST) value ≥ 0.5 were selected to contrast areas with and without interspecific introgressions. Subsequently, the introgressed areas of each accession were removed and a second run was performed to select DSNPs with GST = 1 for the differentiation of one of the ancestral species from the three others.

5.6.2 | Admixture Along the Four Ancestral Genomes

TraceAncestor software (Ahmed et al. 2019) was used for admixture analysis using the WGS resequencing data. For each ancestor, read counts of ancestral and alternative alleles were used to perform maximum-likelihood tests of ancestor contribution (i.e., allelic doses = 0, 0.5, 1) within genomic windows of 20 DSNPs specific to the ancestor in question. At this stage, the number and position of windows varied between ancestral species depending on the density and positions of the DSNPs for each of the four species considered. To integrate the information obtained for the four ancestral species, the allelic doses of each ancestral taxon were then transferred from the 20 DSNPs windows to corresponding 100-kb consecutive windows all along the genome. In cases in which the sum of ancestral contributions exceeded 2, the corresponding windows were considered undetermined for both haplotypes.

The identification of DSNPs and mosaic structure analyses were performed using the SniPlay web tool (<https://sniplay.southgreen.fr/cgi-bin/home.cgi>). The DSNP databases for the four genome assemblies are available at CGH (<https://citrus-genome-hub.southgreen.fr>).

5.7 | Gene Presence/Absence Variation From Resequencing Data

Gene PAV analysis was conducted by aligning DNA sequencing reads from each accession to the four reference genomes using Bowtie2 v2.4.2 in the 'very-sensitive-local' mode.

Gene PAVs were determined using SGSGeneLoss v0.1 (Golicz et al. 2015) according to the cumulative reads coverage of exons within each protein-coding gene. A gene was classified as 'present' if at least 90% of its exons were covered by a minimum of five reads (parameters: minCov = 5, lostCutoff = 0.9); otherwise, it was considered 'absent'.

5.8 | Gene-Based Pangenome Construction

5.8.1 | Identification of Orthologous Genes Among the Four Genome Assemblies and Pangene Identification

To enable a comparison of CITME, CITMI, CITRE and CITMA genome assemblies, we used PanTools v3.4.0 (Jonkheer et al. 2022) to calculate homologous relationships in a predicted panproteome of these 4 species. Initially, we determined the optimal setting for relaxation using the optimal grouping. Based on the results, we selected a value of 2 for the relaxation parameter for homology grouping. Subsequently, gene classification of the homology groups was also done with PanTools.

5.8.2 | Identification of Core and Dispensable Genes Among the Pangenes From the PAV Analysis of Resequencing Data From 55 Cultivated Citrus Accessions

The pangenome was defined by combining the PAV information from resequencing data mapped on the four ancestor assemblies and the identification of the orthologous genes of the four genome assemblies (Figure 6). A pangene was considered present in a given accession if at least one of the corresponding orthologous genes identified in the four assemblies was found present for this accession. The number of shared groups was visualised using the tool available on the following page: https://bioinformatics.psb.ugent.be/cgi-bin/liste/Venn/calculate_venn.html. This approach allowed us to identify common homologous gene groups among the four *Citrus* species, providing insights into their evolutionary relationships and shared genetic features.

5.9 | Interactomics

5.9.1 | Protein-Protein Interaction Network Construction

The proteins from private and core genomes were used to obtain the protein-protein interaction network, built using the STRING database (version 12.0) using the "Add any organism to STRING/Annotate proteome" parameter. The complete network data, including associated metadata, was downloaded (https://string-db.org/cgi/download?sessionId=bOuaPkVaf3f9&species_text=STRG0A37MEC). To ensure the quality of the interactions, an initial filtering step was performed to select only interactions with medium or higher confidence levels (score > 400). The obtained network, named global network, was used in the subsequent analysis.

5.9.2 | Cluster, Centrality Analysis and Subnetwork Analysis

The centrality and cluster analysis of the global network were conducted using scripts developed in R, available in a repository on GitHub (<https://github.com/Donandrade/interactome.github.io.git>). Central proteins were identified using metrics such as degree (hub proteins), betweenness centrality (bottleneck

proteins), and a combination of both (hub-bottleneck proteins), with the mean values of degree and betweenness serving as cut-offs. The centrality and cluster data were then imported into the global network using Cytoscape (version 3.10.3, Java 17.0.5; more information available at <https://cytoscape.org/>). Subnetworks were constructed by selecting central and private proteins and generating new subnetworks containing only their direct interactions. The edges of these subnetworks were exported as tables in Cytoscape, and Bash commands were used to filter first-degree interactions. This approach isolated a more focused set of relevant interactions, enhancing biological interpretation.

This refinement step resulted in a simpler and more informative subnetwork, highlighting the most meaningful connections between exclusive and central proteins in the functional context. This methodology was designed to ensure the reproducibility of the analysis, from the initial network construction to the creation and refinement of specific subnetworks, providing detailed insights into the protein interactions in the studied organism.

5.9.3 | Functional Enrichment Analysis of Clusters

A functional enrichment analysis was performed for each of the clusters in the created subnetworks using the Biological Network Gene Ontology (BiNGO v3.0.5) software, a Cytoscape plugin available at <http://www.cytoscape.org>. The degree of functional enrichment for a given cluster and category was quantitatively assessed (*p*-value) through a hypergeometric distribution, with a multiple test correction applied using the false discovery rate (FDR) algorithm, fully implemented in the BiNGO software. Overrepresented biological process categories were generated after FDR correction, with a significance level of 0.05.

Author Contributions

P.O. conceived and coordinated the experiments and analysis; M.M., E.M., S.S., C.C. and K.L. performed DNA/RNA extractions; C.C., K.L. and A.L. coordinated DNA sequencing; C.B., S.D., C.C., K.L., B.I. and J.-M.A. managed the sequencing data and produced the nuclear assemblies; G.C. produced the chloroplast assemblies, annotation and phylogenetic studies; P.M., M.M. and G.C. performed RNA sequencing and analysis; D.G., N.C., F.M. and S.S.-B. annotated repeated sequences; G.D. performed the gene annotation and structural variation studies; F.D.S.M., R.S. and F.M. annotated PRR genes; F.C., P.O., A.D. and F.L. conducted the analysis of intra and interspecific diversity from resequencing data; G.D., A.S. and P.O. conducted the gene PAV analysis; G.D., A.D., A.S., B.H. and P.O. implemented the super-pangenome; A.S., E.S. and B.H. conducted the analysis of functional enrichment; P.O., D.G., G.D., F.D.S.M., F.M., B.H., G.C. F.C. and A.S. wrote the manuscript and prepared the tables and figures; A.L., P.W. and J.-M.A. coordinated the project at Genoscope; and P.O. and G.D. coordinated the project at CIRAD.

Acknowledgements

This study was conducted with financial support from CIRAD, the Genoscope - Centre National de Séquençage, the *Commissariat à l'Énergie Atomique et aux Énergies Alternatives* and *France Génomique* (ANR-10-INBS-09-08). This work received support from MESO@LR-Platform at the University of Montpellier and technical support from the bioinformatics group of the UMR AGAP Institute, member of the French Institute of Bioinformatics (IFB)-South Green Bioinformatics Platform.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw data from long-read sequencing and genome assemblies of *C. micrantha*, *C. medica* and *C. reticulata* were deposited at EBI under the following accession numbers: CITMI (*C. micrantha*): Raw data: PRJEB78372—Assemblage: PRJEB78515—Umbrella: PRJEB78526. CITME (*C. medica*): Raw data: PRJEB78445—Assemblage: PRJEB78498—Umbrella: PRJEB78525. CITRE (*C. reticulata*): Raw data: PRJEB78477—Assemblage: PRJEB78537—Umbrella: PRJEB78524. Resequencing raw reads derived from 58 citrus accessions was deposited in the EBI database under AC experiment accessions ERX12760151, ERX12780957, ERX12791201 and ERX12817346 to ERX12817400 (Table S1). The raw RNA-seq data for *C. medica*, *C. reticulata*, *C. maxima*, and *C. micrantha* were deposited in the European Nucleotide Archive under accession code PRJEB77075. The pseudochromosome and chloroplast genome assemblies of *C. medica*, *C. reticulata*, and *C. micrantha* and their annotations are available on our website, <https://citrus-genome-hub.southgreen.fr>.

References

- Ahmed, D., A. Comte, F. Curk, 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." *Annals of Botany* 123: 1231–1251.
- Alberti, A., J. Poulain, S. Engelen, et al. 2020. "Sequencing and Data Quality Control." <https://doi.org/10.17504/protocols.io.qwjdxcn>.
- Aleza, P., J. Cuenca, M. Hernández, J. Juárez, L. Navarro, and P. Ollitrault. 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 Biology* 15: 80.
- Aslam, M., K. Sugita, Y. Qin, and A. Rahman. 2020. "Aux/IAA14 Regulates MicroRNA-Mediated Cold Stress Response in Arabidopsis Roots." *International Journal of Molecular Sciences* 21: 8441.
- Aury, J.-M., and B. Istace. 2021. "Hapo-G, Haplotype-Aware Polishing of Genome Assemblies With Accurate Reads." *NAR Genomics and Bioinformatics* 3: lqab034.
- Bausher, M. G., N. D. Singh, S.-B. Lee, R. K. Jansen, and H. Daniell. 2006. "The Complete Chloroplast Genome Sequence of *Citrus sinensis* (L.) Osbeck Var "Ridge Pineapple": Organization and Phylogenetic Relationships to Other Angiosperms." *BMC Plant Biology* 6: 21.
- Borredá, C., E. Pérez-Román, V. Ibanez, J. Terol, and M. Talon. 2019. "Reprogramming of Retrotransposon Activity During Speciation of the Genus Citrus." *Genome Biology and Evolution* 11: 3478–3495.
- Buchfink, B., K. Reuter, and H.-G. Drost. 2021. "Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND." *Nature Methods* 18: 366–368.
- Carbonell-Caballero, J., R. Alonso, V. Ibañez, J. Terol, M. Talon, and J. Dopazo. 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.
- Chen, C., H. Chen, Y. Zhang, et al. 2020. "TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data." *Molecular Plant* 13: 1194–1202.
- Chen, Y., F. Nie, S.-Q. Xie, et al. 2021. "Efficient Assembly of Nanopore Reads via Highly Accurate and Intact Error Correction." *Nature Communications* 12: 60.
- Curk, F., F. Ollitrault, A. Garcia-Lor, F. Luro, L. Navarro, and P. Ollitrault. 2016. "Phylogenetic Origin of Limes and Lemons Revealed by Cytoplasmic and Nuclear Markers." *Annals of Botany* 117: 565–583.

- Deng, H., S. Xiang, Q. Guo, W. Jin, Z. Cai, and G. Liang. 2019. "Molecular Cytogenetic Analysis of Genome-Specific Repetitive Elements in *Citrus clementina* Hort. Ex Tan. and Its Taxonomic Implications." *BMC Plant Biology* 19: 77.
- Dierckxsens, N., P. Mardulyn, and G. Smits. 2017. "NOVOPlasty: De Novo Assembly of Organelle Genomes From Whole Genome Data." *Nucleic Acids Research* 45: e18.
- Ding, B., H. Hu, T. Liu, et al. 2023. "Pan-Genome of Pear Provides Insights Into the Fruit Quality Traits Differentiation Between Asian and European Pears." *bioRxiv*. <https://doi.org/10.1101/2023.09.29.560244>.
- Dugrand-Judek, A., A. Olry, A. Hehn, et al. 2015. "The Distribution of Coumarins and Furanocoumarins in *Citrus* Species Closely Matches Citrus Phylogeny and Reflects the Organization of Biosynthetic Pathways." *PLoS One* 10: e0142757.
- Emms, D. M., and S. Kelly. 2019. "OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics." *Genome Biology* 20: 238.
- Fanciullino, A. L., C. Dhuique-Mayer, F. Luro, J. Casanova, R. Morillon, and P. Ollitrault. 2006. "Carotenoid Diversity in Cultivated Citrus Is Highly Influenced by Genetic Factors." *Journal of Agricultural and Food Chemistry* 54: 4397–4406.
- Fann, J.-Y., A. Kovarik, V. Hemleben, N. I. Tsirekidze, and T. G. Beridze. 2001. "Molecular and Structural Evolution of Citrus Satellite DNA." *Theoretical and Applied Genetics* 103: 1068–1073.
- Flutre, T., E. Duprat, C. Feuillet, and H. Quesneville. 2011. "Considering Transposable Element Diversification in *De Novo* Annotation Approaches." *PLoS One* 6: e16526.
- Gao, L., I. Gonda, H. Sun, et al. 2019. "The Tomato Pan-Genome Uncovers New Genes and a Rare Allele Regulating Fruit Flavor." *Nature Genetics* 51: 1044–1051.
- Gao, Y., J. Xu, Z. Li, et al. 2023. "Citrus Genomic Resources Unravel Putative Genetic Determinants of Huanglongbing Pathogenicity." *iScience* 26: 106024.
- Giraud, D., N. Choise, M. Summo, et al. 2025. "Construction of a Comprehensive Library of Repeated Sequences for the Annotation of Citrus Genomes." *BMC Genomic Data* 26: 30.
- Golicz, A. A., P. A. Martinez, M. Zander, et al. 2015. "Gene Loss in the Fungal Canola Pathogen *Leptosphaeria maculans*." *Functional & Integrative Genomics* 15: 189–196.
- Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk, and O. Gascuel. 2010. "New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0." *Systematic Biology* 59: 307–321.
- Guo, Y. 2023. "Pangenome and the Diversity of Potato Species." *Nature Food* 4: 638.
- Hoede, C., S. Arnoux, M. Moisset, et al. 2014. "PASTEC: An Automatic Transposable Element Classification Tool." *PLoS One* 9: e91929.
- Holub, E. B. 2001. "The Arms Race Is Ancient History in Arabidopsis, the Wildflower." *Nature Reviews. Genetics* 2: 516–527.
- Huang, Y., J. He, Y. Xu, et al. 2023. "Pangenome Analysis Provides Insight Into the Evolution of the Orange Subfamily and a Key Gene for Citric Acid Accumulation in Citrus Fruits." *Nature Genetics* 55: 1964–1975.
- Isobe, S., H. Fujii, K. Shirasawa, Y. Kawahara, T. Endo, and T. Shimada. 2023. "Haploid-Resolved and Chromosome-Scale Genome Assembly in *Citrus unshiu* and Its Parental Species, *C. nobilis* and *C. kinokuni*." *bioRxiv*. <https://doi.org/10.1101/2023.06.02.543356>.
- Istace, B., C. Belser, and J.-M. Aury. 2020. "BiSCoT: Improving Large Eukaryotic Genome Assemblies With Optical Maps." *PeerJ* 8: e10150.
- Jing, H., E. G. Wilkinson, K. Sageman-Furnas, and L. C. Strader. 2023. "Auxin and Abiotic Stress Responses." *Journal of Experimental Botany* 74: 7000–7014.
- Jones, J. D. G., and J. L. Dangl. 2006. "The Plant Immune System." *Nature* 444: 323–329.
- Jones, P., D. Binns, H.-Y. Chang, et al. 2014. "InterProScan 5: Genome-Scale Protein Function Classification." *Bioinformatics* 30: 1236–1240.
- Jonkheer, E. M., D.-J. M. van Workum, S. Sheikhzadeh Anari, et al. 2022. "PanTools v3: Functional Annotation, Classification and Phylogenomics." *Bioinformatics* 38: 4403–4405.
- Jupe, F., L. Pritchard, G. J. Etherington, et al. 2012. "Identification and Localisation of the NB-LRR Gene Family Within the Potato Genome." *BMC Genomics* 13: 75.
- Kang, S.-K., D.-H. Lee, H.-J. An, et al. 2008. "Extensive Chromosomal Polymorphism Revealed by Ribosomal DNA and Satellite DNA Loci in 13 Citrus Species." *Molecules and Cells* 26: 319–322.
- Katagiri, F., and K. Tsuda. 2010. "Understanding the Plant Immune System." *Molecular Plant-Microbe Interactions* 23: 1531–1536.
- Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30: 772–780.
- Langmead, B., and S. L. Salzberg. 2012. "Fast Gapped-Read Alignment With Bowtie 2." *Nature Methods* 9: 357–359.
- Lefort, V., J.-E. Longueville, and O. Gascuel. 2017. "SMS: Smart Model Selection in PhyML." *Molecular Biology and Evolution* 34: 2422–2424.
- Leister, D., and T. Kleine. 2011. "Chapter Three – Role of Intercompartmental DNA Transfer in Producing Genetic Diversity." In *International Review of Cell and Molecular Biology*, edited by K. W. Jeon, 73–114. Academic Press.
- Li, N., Q. He, J. Wang, et al. 2023. "Super-Pangenome Analyses Highlight Genomic Diversity and Structural Variation Across Wild and Cultivated Tomato Species." *Nature Genetics* 55: 852–860.
- Li, R., Y. Li, K. Kristiansen, and J. Wang. 2008. "SOAP: Short Oligonucleotide Alignment Program." *Bioinformatics* 24: 713–714.
- Liu, H., X. Wang, S. Liu, et al. 2022. "Citrus Pan-Genome to Breeding Database (CPBD): A Comprehensive Genome Database for Citrus Breeding." *Molecular Plant* 15: 1503–1505.
- Liu, Y., H. Du, P. Li, et al. 2020. "Pan-Genome of Wild and Cultivated Soybeans." *Cell* 182: 162–176.e13.
- Liu, Y., M. Tahir ul Qamar, J.-W. Feng, et al. 2019. "Comparative Analysis of Miniature Inverted-Repeat Transposable Elements (MITEs) and Long Terminal Repeat (LTR) Retrotransposons in Six Citrus Species." *BMC Plant Biology* 19: 140.
- Lovell, J. T., A. Sreedasyam, M. E. Schranz, et al. 2022. "GENESPACE Tracks Regions of Interest and Gene Copy Number Variation Across Multiple Genomes." *eLife* 11: e78526.
- Luro, F., E. Bloquel, B. Tomu, et al. 2017. "The INRA-CIRAD Citrus Germplasm Collection of San Giuliano, Corsica." In *AGRUMED: Archaeology and History of Citrus Fruit in the Mediterranean: Acclimatization, Diversifications, Uses*, edited by V. Zech-Matterne and G. Fiorentino. Publications du Centre Jean Bérard.
- Luro, F., N. Venturini, G. Costantino, J. Paolini, P. Ollitrault, and J. Costa. 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.
- Magen, S., H. Seybold, D. Laloum, and T. Avin-Wittenberg. 2022. "Metabolism and Autophagy in Plants—A Perfect Match." *FEBS Letters* 596: 2133–2151.
- Marczuk-Rojas, J. P., A. Salmerón, A. Alcayde, V. Isanbaev, and L. Carretero-Paulet. 2024. "Plastid DNA Is a Major Source of Nuclear Genome Complexity and of RNA Genes in the Orphan Crop Moringa." *BMC Plant Biology* 24: 437.

- Markello, C., C. Huang, A. Rodriguez, et al. 2022. "A Complete Pedigree-Based Graph Workflow for Rare Candidate Variant Analysis." *Genome Research* 32: 893–903.
- Martin, G., F.-C. Baurens, G. Droc, et al. 2016. "Improvement of the Banana "*Musa acuminata*" Reference Sequence Using NGS Data and Semi-Automated Bioinformatics Methods." *BMC Genomics* 17: 243.
- Michalovova, M., B. Vyskot, and E. Kejnovsky. 2013. "Analysis of Plastid and Mitochondrial DNA Insertions in the Nucleus (NUPTs and NUMTs) of Six Plant Species: Size, Relative Age and Chromosomal Localization." *Heredity* 111: 314–320.
- Nakandala, U., A. K. Masouleh, M. W. Smith, et al. 2023. "Haplotype Resolved Chromosome Level Genome Assembly of Citrus Australis Reveals Disease Resistance and Other Citrus Specific Genes." *Horticulture Research* 10: uhad058.
- Nicolosi, E., Z. N. Deng, A. Gentile, S. la Malfa, G. Continella, and E. Tribulato. 2000. "Citrus Phylogeny and Genetic Origin of Important Species as Investigated by Molecular Markers." *Theoretical and Applied Genetics* 100: 1155–1166.
- Ollitrault, P., B. Hufnagel, F. Curk, et al. 2024. "Comparative Genetic Mapping and a Consensus Interspecific Genetic Map Reveal Strong Synteny and Collinearity Within the Citrus Genus." *Frontiers in Plant Science* 15: 1475965.
- Oueslati, A., A. Salhi-Hannachi, F. Luro, H. Vignes, P. Mournet, and P. Ollitrault. 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.
- Panchy, N., M. Lehti-Shiu, and S.-H. Shiu. 2016. "Evolution of Gene Duplication in Plants." *Plant Physiology* 171: 2294–2316.
- Peng, Z., J. V. Bredeson, G. A. Wu, 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." *Plant Journal* 104: 1215–1232.
- Qian, W., and J. Zhang. 2014. "Genomic Evidence for Adaptation by Gene Duplication." *Genome Research* 24: 1356–1362.
- Quesneville, H., C. M. Bergman, O. Andrieu, et al. 2005. "Combined Evidence Annotation of Transposable Elements in Genome Sequences." *PLoS Computational Biology* 1: 166–175.
- Ranallo-Benavidez, T. R., K. S. Jaron, and M. C. Schatz. 2020. "GenomeScope 2.0 and Smudgeplot for Reference-Free Profiling of Polyploid Genomes." *Nature Communications* 11: 1432.
- Salavati, M., S. J. Bush, S. Palma-Vera, M. E. B. McCulloch, D. A. Hume, and E. L. Clark. 2019. "Elimination of Reference Mapping Bias Reveals Robust Immune Related Allele-Specific Expression in Crossbred Sheep." *Frontiers in Genetics* 10: 863.
- Sallet, E., J. Gouzy, and T. Schiex. 2019. "EuGene: An Automated Integrative Gene Finder for Eukaryotes and Prokaryotes." *Methods in Molecular Biology (Clifton, N.J.)* 1962: 97–120.
- Santana Silva, R. J., and F. Micheli. 2020. "RRGPredictor, a Set-Theory-Based Tool for Predicting Pathogen-Associated Molecular Pattern Receptors (PRRs) and Resistance (R) Proteins From Plants." *Genomics* 112: 2666–2676.
- Santini, J., J. Giannettini, S. Herbette, et al. 2012. "Physiological and Biochemical Response to Photooxidative Stress of the Fundamental Citrus Species." *Scientia Horticulturae* 147: 126–135.
- Shi, W., W. Song, J. Liu, C. Shi, and S. Wang. 2023. "Comparative Chloroplast Genome Analysis of *Citrus* (Rutaceae) Species: Insights Into Genomic Characterization, Phylogenetic Relationships, and Discrimination of Subgenera." *Scientia Horticulturae* 313: 111909.
- Shi, W., W. Song, Y. Zhao, C. Shi, and S. Wang. 2023. "Complete Chloroplast Genomes of Four *Atalantia* (Rutaceae) Species: Insights Into Comparative Analysis, Phylogenetic Relationships, and Divergence Time Estimation." *Plant Systematics and Evolution* 309: 31.
- Sibbesen, J. A., J. M. Eizenga, A. M. Novak, et al. 2023. "Haplotype-Aware Pantranscriptome Analyses Using Spliced Pangenome Graphs." *Nature Methods* 20: 239–247.
- Stevenson, K. R., J. D. Coolon, and P. J. Wittkopp. 2013. "Sources of Bias in Measures of Allele-Specific Expression Derived From RNA-Seq Data Aligned to a Single Reference Genome." *BMC Genomics* 14: 536.
- Sun, X., C. Jiao, H. Schwaninger, et al. 2020. "Phased Diploid Genome Assemblies and Pan-Genomes Provide Insights Into the Genetic History of Apple Domestication." *Nature Genetics* 52: 1423–1432.
- Swingle, W. T., and P. C. Reece. 1967. "The Botany of Citrus and Its Wild Relatives." In *The Citrus Industry*, edited by W. Reuther, H. J. Webber, and D. L. Batchelor, 190–430. CA, USA.
- Tanaka, T. 1961. *Citologia: Semi-Centennial Commemoration Papers on Citrus Studies*. Citologia Supporting Foundation.
- Tranchant-Dubreuil, C., M. Rouard, and F. Sabot. 2019. "Plant Pangenome: Impacts on Phenotypes and Evolution." In *Annual Plant Reviews Online*, 453–478. John Wiley & Sons, Ltd.
- Wang, L., F. He, Y. Huang, et al. 2018. "Genome of Wild Mandarin and Domestication History of Mandarin." *Molecular Plant* 11: 1024–1037.
- Wang, N., C. Li, L. Kuang, et al. 2022. "Pan-Mitogenomics Reveals the Genetic Basis of Cytonuclear Conflicts in Citrus Hybridization, Domestication, and Diversification." *Proceedings of the National Academy of Sciences of the United States of America* 119: e2206076119.
- Wang, N., X. Song, J. Ye, et al. 2022. "Structural Variation and Parallel Evolution of Apomixis in Citrus During Domestication and Diversification." *National Science Review* 9: nwac114.
- Wang, X., Y. Xu, S. Zhang, et al. 2017. "Genomic Analyses of Primitive, Wild and Cultivated Citrus Provide Insights Into Asexual Reproduction." *Nature Genetics* 49: 765–772.
- Wang, Y., H. Tang, J. D. Debarry, et al. 2012. "MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity." *Nucleic Acids Research* 40: e49.
- Wang, Y., L. Zhou, D. Li, et al. 2015. "Genome-Wide Comparative Analysis Reveals Similar Types of NBS Genes in Hybrid *Citrus sinensis* Genome and Original Citrus Clementine Genome and Provides New Insights Into Non-TIR NBS Genes." *PLoS One* 10: e0121893.
- Wu, B., Q. Yu, Z. Deng, Y. Duan, F. Luo, and F. Gmitter Jr. 2022. "A Chromosome-Level Phased Genome Enabling Allele-Level Studies in Sweet Orange: A Case Study on Citrus Huanglongbing Tolerance." *Horticulture Research* 10: uhac247.
- Wu, G. A., S. Prochnik, J. Jenkins, et al. 2014. "Sequencing of Diverse Mandarin, Pummelo and Orange Genomes Reveals Complex History of Admixture During Citrus Domestication." *Nature Biotechnology* 32: 656–662.
- Wu, G. A., J. Terol, V. Ibanez, et al. 2018. "Genomics of the Origin and Evolution of Citrus." *Nature* 554: 311–316.
- Xu, Q., L.-L. Chen, X. Ruan, et al. 2013. "The Draft Genome of Sweet Orange (*Citrus sinensis*)." *Nature Genetics* 45: 59–66.
- Yang, S., J. Cai, M. Wang, et al. 2023. "The Construction and Analysis of Wax Gourd Pangenome Uncover Fruit Quality-Related and Resistance Genes." *Scientia Horticulturae* 318: 112084.
- Yang, Z., and Y. Li. 2018. "Dissection of RNAi-Based Antiviral Immunity in Plants." *Current Opinion in Virology* 32: 88–99.
- Yocca, A. E., A. Platts, E. Alger, et al. 2023. "Blueberry and Cranberry Pangenomes as a Resource for Future Genetic Studies and Breeding Efforts." *bioRxiv*. <https://doi.org/10.1101/2023.07.31.551392>.
- Yoshida, T., H. Y. Furihata, T. K. To, T. Kakutani, and A. Kawabe. 2019. "Genome Defense Against Integrated Organellar DNA Fragments From

Plastids Into Plant Nuclear Genomes Through DNA Methylation.” *Scientific Reports* 9: 2060.

Zhang, G.-J., R. Dong, L.-N. Lan, S.-F. Li, W.-J. Gao, and H.-X. Niu. 2020. “Nuclear Integrants of Organellar DNA Contribute to Genome Structure and Evolution in Plants.” *International Journal of Molecular Sciences* 21: 707.

Zhao, N., C. E. Grover, Z. Chen, J. F. Wendel, and J. Hua. 2019. “Intergenomic Gene Transfer in Diploid and Allopolyploid *Gossypium*.” *BMC Plant Biology* 19: 492.

Zia, K., M. J. Rao, M. Sadaqat, et al. 2022. “Pangenome-Wide Analysis of Cyclic Nucleotide-Gated Channel (CNGC) Gene Family in *Citrus* Spp. Revealed Their Intraspecies Diversity and Potential Roles in Abiotic Stress Tolerance.” *Frontiers in Genetics* 13: 1034921.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** Supporting Information. **Table S20:** pbi70553-sup-0002-TableS20.xlsx. **Table S22:** pbi70553-sup-0003-TableS22.xlsx. **Table S23:** pbi70553-sup-0004-TableS23.xlsx. **Table S24:** pbi70553-sup-0005-TableS24.xlsx. **Table S25:** pbi70553-sup-0006-TableS25.xlsx. **Table S26:** pbi70553-sup-0007-TableS26.xlsx. **Table S27:** pbi70553-sup-0008-TableS27.xlsx. **Table S28:** pbi70553-sup-0009-TableS28.xlsx.