

Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral Cucurbitaceae karyotype

Clement Pichot, Anis Djari, Joseph Tran, Marion Verdenaud, William Marande, Cécile C. Huneau, David Latrasse, Vivien Sommard, V. Gautier, Judit Szecsi, et al.

▶ **To cite this version:**

Clement Pichot, Anis Djari, Joseph Tran, Marion Verdenaud, William Marande, et al.. Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral Cucurbitaceae karyotype. iScience, 2022, 25 (1), pp.103696. 10.1016/j.isci.2021.103696 . hal-03507170

HAL Id: hal-03507170
https://hal.inrae.fr/hal-03507170

Submitted on 8 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



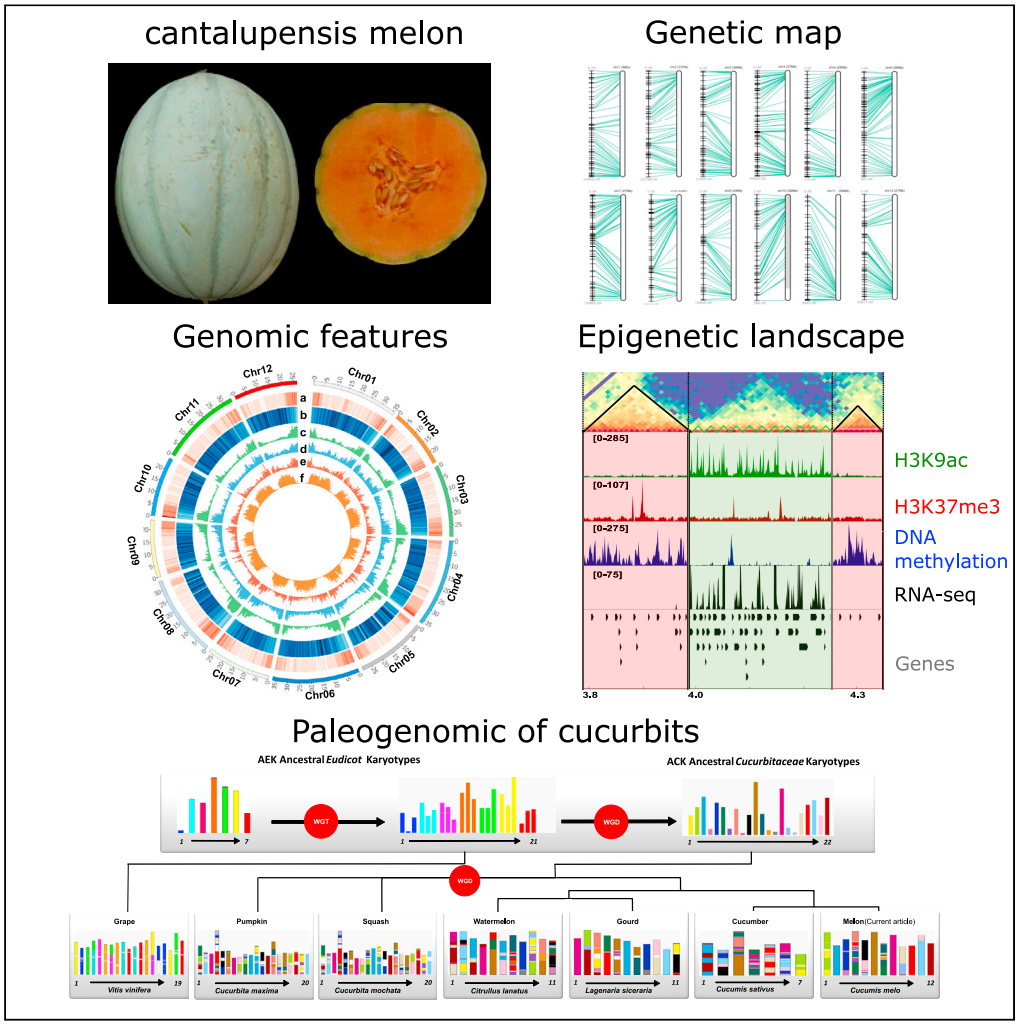
Distributed under a Creative Commons Attribution 4.0 International License

iScience

CellPress
OPEN ACCESS

Article

Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral cucurbitaceae karyotype



Clement Pichot, Anis Djari, Joseph Tran, ..., Mohamed Zouine, Adnane Boualem, Abdelhafid Bendahmane

abdelhafid.bendahmane@inrae.fr

Highlights

We provide a chromosome scale *C. melo* var. *cantalupensis* Charentais mono genome

Epigenomic analysis revealed a widespread compartmentalization of the melon genome

We reconstructed the ancestral *Cucurbitaceae* karyotype

Melon has retained the ancestral *Cucurbitaceae* genome structure

Pichot et al., iScience *25*, 103696
January 21, 2022 © 2022 The Authors.
https://doi.org/10.1016/j.isci.2021.103696

iScience

CellPress
OPEN ACCESS

Article

Cantaloupe melon genome reveals 3D chromatin features and structural relationship with the ancestral cucurbitaceae karyotype

Clement Pichot,[1,2] Anis Djari,[3] Joseph Tran,[1,2] Marion Verdenaud,[1,2] William Marande,[4] Cecile Huneau,[5] Veronique Gautier,[5] David Latrasse,[1,2] Sandrine Arribat,[4] Vivien Sommard,[1,2] Christelle Troadec,[1,2] Charles Poncet,[5] Mohammed Bendahmane,[6] Judit Szecsi,[6] Catherine Dogimont,[7] Jerome Salse,[5] Moussa Benhamed,[1,2] Mohamed Zouine,[3] Adnane Boualem,[1,2] and Abdelhafid Bendahmane[1,2,8,*]

SUMMARY

***Cucumis melo* displays a large diversity of horticultural groups with cantaloupe melon the most cultivated type. Using a combination of single-molecule sequencing, 10X Genomics link-reads, high-density optical and genetic maps, and chromosome conformation capture (Hi-C), we assembled a chromosome scale *C. melo* var. *cantalupensis* Charentais mono genome. Integration of RNA-seq, MeDip-seq, ChIP-seq, and Hi-C data revealed a widespread compartmentalization of the melon genome, segregating constitutive heterochromatin and euchromatin. Genome-wide comparative and evolutionary analysis between melon botanical groups identified Charentais mono genome increasingly more divergent from Harukei-3 (*reticulatus*), Payzawat (*inodorus*), and HS (ssp. *agrestis*) genomes. To assess the paleohistory of the *Cucurbitaceae*, we reconstructed the ancestral *Cucurbitaceae* karyotype and compared it to sequenced cucurbit genomes. In contrast to other species that experienced massive chromosome shuffling, melon has retained the ancestral genome structure. We provide comprehensive genomic resources and new insights in the diversity of melon horticultural groups and evolution of cucurbits.**

INTRODUCTION

Melon (*Cucumis melo L.*) is a diploid species (2n = 2x = 24) belonging to the Cucurbitaceae family that consists of about 965 species (Christenhusz and Byng, 2016). The Cucurbitaceae family is positioned in the front list of plant families for number of species used as human food and for cultural, medicinal, and botanical significance. Among cucurbits, melon is one of the most ancient fruit crops in the word; its cultivation was traced to about 3700 and 3500 BC in ancient Egypt (El Hadidi and Hosni, 1996; Janick et al., 2007; Paris, 2016; van Zeist and de Roller, 1993). Phylogenetically, *C. melo* clade is related to *C. trigonus* and *C. picrocarpus*, two species that grow in India and Australia, respectively (Endl et al., 2018). Melon domestication history is still under discussion with a consensus around two domestication events that took place in parallel in Africa and in Asia (Endl et al., 2018). Most domesticated melons can be attributed to two wild lineages, *C. melo* ssp. *melo* and *C. melo* ssp. *meloides*, that diverged by 1–3 Ma. The subspecies *meloides* is widespread throughout Africa and gave rise to the 'Tibish' and 'Fadasi' melons grown in the Sudanese region. The subspecies *melo* is restricted to Asia and gave rise to all modern cultivars grown worldwide (Chomicki et al., 2020). Asian melon clade also comprises cultivars belonging to the *C. melo* subspecies *agrestis*, also known as wild melon (Lian et al., 2021).

Because of the large diversity of horticultural types, melon cultivars are also classified into botanical groups including *cantalupensis*, *reticulatus* and *inodorus* in the ssp. *melo*, and *momordica*, *makuwa*, *conomon* and *chinensis* in the ssp. *agrestis* (Pitrat, 2013). A large-scale metabolomic profiling of melon accessions belonging to different botanical groups revealed a hierarchical grouping similar but not identical to the phenotypic or the sequence marker-based classification (Moing et al., 2020), highlighting the need of genome-wide comparative analysis.

[1]Université Paris-Saclay, CNRS, INRAE, Univ Evry, Institute of Plant Sciences Paris-Saclay (IPS2), 91190 Gif sur Yvette, France

[2]Université de Paris, Institute of Plant Sciences Paris-Saclay (IPS2), 91190 Gif sur Yvette, France

[3]Laboratoire de Recherche en Sciences Végétales, Université de Toulouse, CNRS, UPS, Toulouse INP, 31320 Auzeville-Tolosane, France

[4]INRAE, Centre National de Ressources Génomiques Végétales, 31326 Castanet-Tolosan, France

[5]INRAE-UCA UMR 1095 GDEC, 5 chemin de Beaulieu, 63000 Clermont-Ferrand, France

[6]Laboratoire Reproduction et Développement des Plantes, INRAE, CNRS, Université Lyon 1-ENSL, Ecole Normale Supérieure de Lyon, 69364 Lyon Cedex 07, France

[7]INRAE GAFL, Génétique et Amélioration des Fruits et Légumes, 84143 Montfavet, France

[8]Lead contact

*Correspondence: abdelhafid.bendahmane@inrae.fr

https://doi.org/10.1016/j.isci.2021.103696

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

For decades, the melon has been a central model species for investigating vascular fluxes, sex determination, and fruit ripening processes. It entered into the genomic era in 2012 by the sequencing of a hybrid line DHL92, which derives from a cross between an *inodorus* type of melon and an accession corresponding to *C. melo* ssp. *agrestis* var. chinensis (Garcia-Mas et al., 2012). Despite, the first published genome has allowed important progress in the characterization of the melon genome, the small reads sequencing technologies used, led to highly fragmented genome of thousands of sequence gaps (Ruggieri et al., 2018). Another consequence of the use of short reads sequencing in genome assembly is the low representation of genome duplication and repeated elements, ultimately leading to bias in our understanding of genome organization and function. More recently, two long-read technologies, known as Pacific Biosciences' (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore Technologies' (ONT) have permitted to assemble with high accuracy repetitive sequences (Choi et al., 2020; Hosmani et al., 2019; Sakai et al., 2015; Wang et al., 2017). Combined with high throughput chromosome interactions (Hi-C) mapping and Bionano optical mapping, long-read technologies sequencing technologies are delivering high quality assembled genomes. For instance, two high-quality melon genomes have been recently assembled from the *inodorus* and *reticulatus* horticultural groups of *C. melo* ssp. *melo* (Yano et al., 2020; Zhang et al., 2019). Two high quality melon genomes were also assembled for *C. melo* ssp. *agrestis* (Yang et al., 2020) and from DHL92, a hybrid line between *inodorus* and *agrestis* melon (Castanera et al., 2020). However, to bring new insight to the genome organization of melon in relation to the large diversity of melon botanical groups, it is required to sequence and assemble high quality genomes from more botanical groups.

In this current study, we report the genome sequencing of a climacteric Charentais mono, hereafter Charmono, breeding line belonging to *cantalupensis* horticultural group. According to FAO, cultivars belonging to *cantalupensis* horticultural group contribute to two-third of the world melon production (Figure S1). To sequence the Charmono genome, we used a combination of SMRT sequencing technology, Bionano optical mapping, Hi-C and 10× Genomics. Chromosome scale genome was assembled and genes and transposable elements annotated. As 3D chromatin features are central in the regulation of gene expression, we analyzed the chromatin architecture by integrating RNA-seq, Hi-C, and histone and DNA methylation marks (Dong et al., 2017; Feng et al., 2014). To bring new insight in melon genome evolution, we also investigated sequence divergence of sequenced melon botanical groups and the paleohistory of melon genome relative to constructed ancestral *Cucurbitaceae* karyotype. Our investigations are expected to further help understanding the phenotypic diversity and melon botanical groups and evolutionary history of cucurbits.

RESULTS

de novo Charmono genome assembly

Charmono is a European cantaloupe melon that is climacteric and sexually monoecious (Figures 1A–1D). Genetically, Charmono is classified *C.melo* ssp. *melo* and belongs to *cantalupensis* horticultural group (Naudin, 1859) which differs from the sequenced melon genomes (Figure 1E). To sequence Charmono, we generated 45 Gb of PacBio reads equivalent to about 100-fold melon genome (Table S1). PacBio reads were trimmed, corrected, and assembled with CANU 1.6 (Koren et al., 2017). After consensus generation, we obtained 236 contigs with N50 of 6.8Mb and a total length of 367 Mb. The longest contig was 21.9 Mb in length and 50% of the genome was represented in 18 contigs (Table S2). To organize the contigs in scaffolds, we generated optical maps of Charmono genome (Table S1). In total 95.5 Gb of data corresponding to molecules of more than 150 kb were generated and used for optical map *de novo* assembly. In total, we produced 329 optical maps with an N50 of 1.9 Mbp for a genome map length of 383.8 Mbp (Table S2). By mapping the whole genome assembly on the 329 optical genome maps, the 236 contigs were scaffolded into 50 sequences of about 366.2 Mb in length. The longest scaffold was 24.6 Mb in length and 50% of the genome was represented in 11 scaffolds (Table S2). To assign the scaffolds to chromosomes, we exploited the previously generated melon genetic map of 4,888 genotyping based sequencing markers (Pereira et al., 2018). Based on this analysis, 42 scaffolds, for which the order of the markers was found consistent with the hybrid scaffolding consensus, were ordered into 12 pseudomolecules corresponding to the 12 melon chromosomes (Figure 1F). The remaining 6 contigs that do not match any genetic marker were concatenated into a virtual chromosome called chromosome zero (CMiso1.1chr00; Table S3). To assess further the quality of the assembly we mapped Hi-C sequencing reads to the consensus sequence of the 12 melon chromosomes. Like the complete collinearity obtained between the genetic map and the consensus sequence, and with very few remaining gaps, perfect sequence contiguity was also obtained with the Hi-C data (Figures 1G and S1). To correct sequencing errors the genome was first polished by

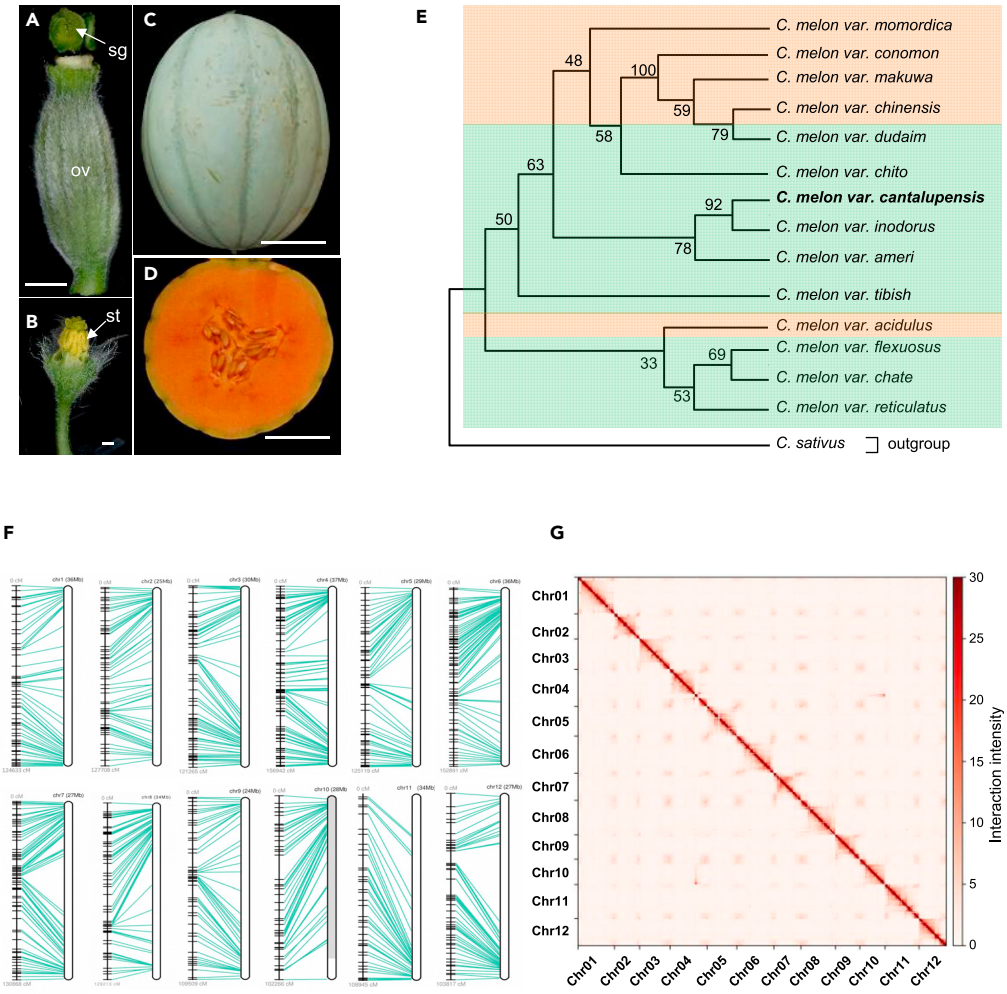


Figure 1. Charmono chromosome-level-assembly

(A–D) Flowers and fruit of the monoecious *C.melo var. cantalupensis* Charmono cultivar. Female (A) and male flowers (B) and the whole (C) or cut fruit (D). Ov, ovary; sg, stigmate; st, stamen. Scale bars: (A), 5mm; (B), 2mm; (C and D), 5cm.

(E) Neighbor-joining phylogenetics trees of 14 different *C. melo* accessions/varieties with 1,000 bootstrap replicates. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test are shown next to the branches. *C. melo ssp. melo* and *C. melo ssp. agrestis* are highlighted in green and orange, respectively. Cucumber, *C. sativus*, was used as outgroup.

(F) Representations of chromosome connections between the physical positions on the reconstructed chromosome and genetic-map positions in Charmono genome.

(G) Hi-C map of the 12 chromosome of Charmono genome at 50 Kb of resolution. Heatmap intensity scale is indicated by the interaction intensity color bar.

Quiver 49 (Chin et al., 2013). A second round of polishing was performed using genomic 10X Illumina paired-end reads and RNAseq data. In total 26,984 sequence errors were identified and edited in the consensus genome sequence. Four gaps were also filled using contigs generated based on the assembly of 200 million 10X Genomics reads (Table S2). In altogether, we obtained a final genome assembly length of 365,081 742 bp covering the 12 melon chromosomes (Table S3).

Charmono genome annotation

To predict protein-coding genes and long non-coding RNAs (lncRNAs) in Charmono genome, we used EuGene software (Sallet et al., 2014; Figure S2) that integrates several weighted features such as genome masking, output of transcript and protein alignments and probabilistic sequence models training. RNAseq mapping was carried out using STAR (Dobin et al., 2013) and transcript prediction was carried out by

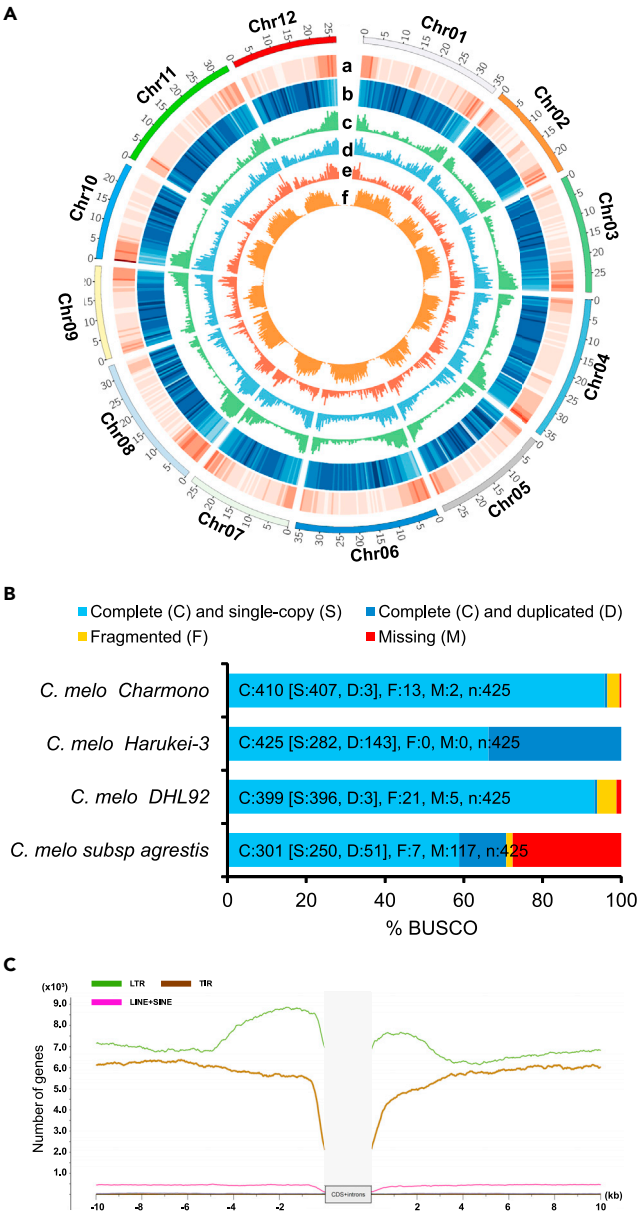


Figure 2. Charmono genome annotation and TE distribution

(Aand B) Circos plot showing features of *C. melo var. cantalupensis*Charmono genome in 500 kb bins. Gene (A) and transposable element (B) density heatmaps.

(C–F) Distribution of the epigenetic features. (C) H3K9ac, (D) Accessible chromatin, (E) H3K27me3 and (F) Global DNA methylation.(B) BUSCO completeness based on *Viridiplantae* dataset comparison between available melon annotations. Colors in the bar represent the different classes of the BUSCO assessment results. Red, yellow, purple, and blue indicate Missing, Fragmented, Duplicated, and single-copy category, respectively.(C) TE landscape surrounding Charmono genes. For all genes, 10 kb upstream the transcription start site (TSS) and 10 kb downstream transcription termination site (TTS) were analyzed.

StringTie (Kovaka et al., 2019). For more accuracy, the gene models were predicted separately on the positive and the negative DNA strands and merged. After several rounds of optimization of the parameters and manual assessment of gene model prediction, we predicted 31,348 protein-coding genes and 1961 lncRNAs (Figure 2A, Tables 1 and S4). To assess the integrity of the Charmono genome, we used Plantae Benchmarking Universal Single-Copy Orthologs (BUSCO) as a proxy of genome completeness. BUSCO

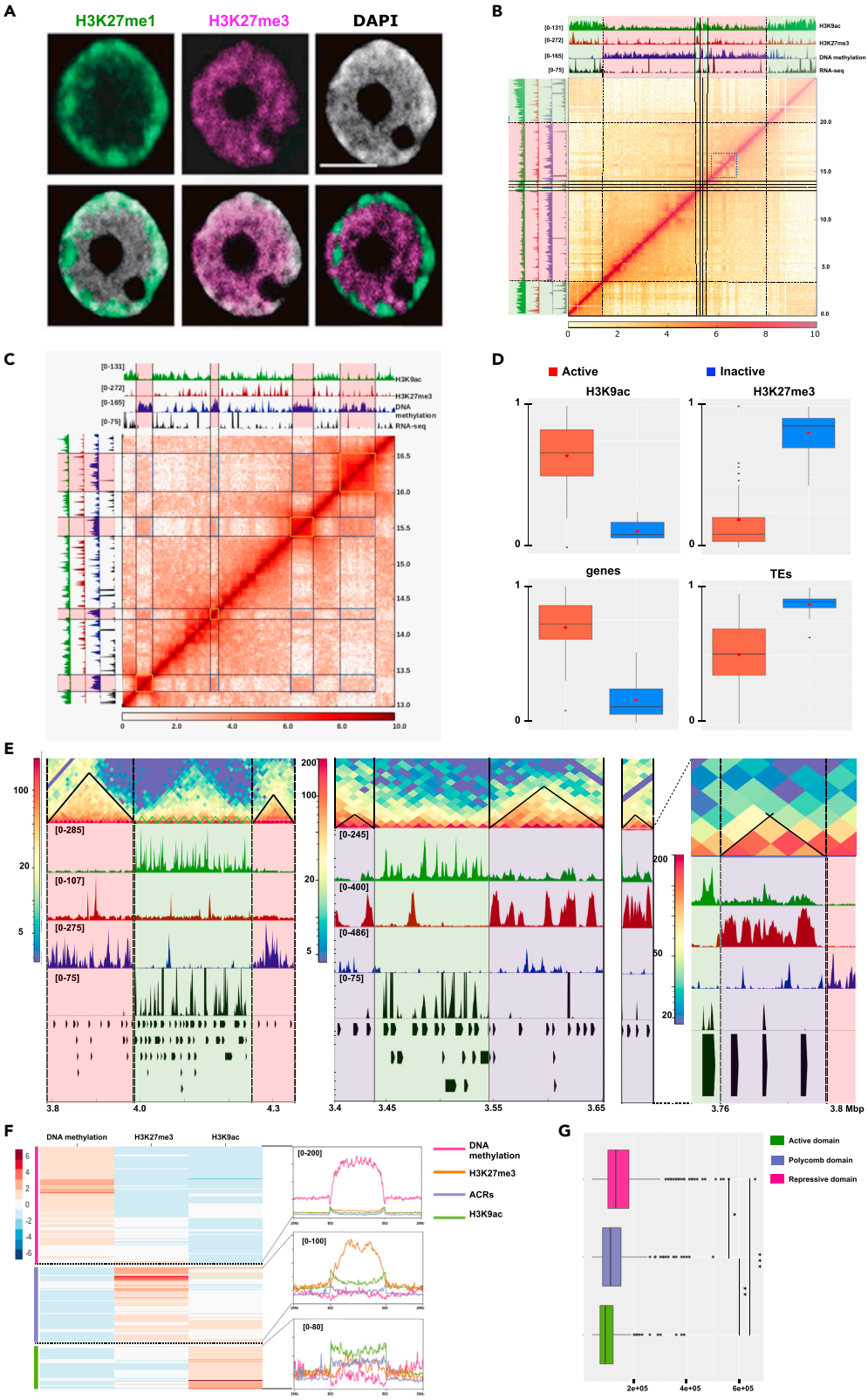

A
H3K27me1
H3K27me3
DAPI
B
H3K9ac
H3K27me3
DNA methylation
RNA-seq
C
D
Active
Inactive
H3K9ac
H3K27me3
genes
TEs
E
F
DNA methylation
H3K27me3
H3K9ac
ACRs
G
Active domain
Polycomb domain
Repressive domain

Figure 3. Characterization of the leaf Charmonoepigenome

(A) Immunofluorescence detection of H3K27me1 (green) and H3K27me3 (purple) and DAPI (white) in isolated nucleus. Scale bar represents 10 μm.

(B) Integration of Hi-C, ChIP-seq, MeDIP-seq and RNA-seq data at the chromosome scale. The heatmap of intra-chromosomal interaction frequency of chromosome 2 is presented. The color bar shows the interaction frequency scale.

(C) Integration of Hi-C, ChIP-seq, MeDIP-seq and RNA-seq data at the megabase scale. A 2D heatmap with the interaction frequency in region chr2-13000000-17,000,000 is presented.

(D) Genomic and chromatin features of the melon A/B compartments. Boxplots showing comparisons of H3K9ac level, H3K27me3 level, gene density and TE density between A and B compartment.

(E) Triangle heatmap of Hi-C interaction frequency. The H3K9ac, H3K27me3 and DNA methylation signals are presented in green, red and blue, respectively. The RNA-seq signal is represented in black.

(F) Interacting domains are associated to specific chromatin marks. Unsupervised clustering of chromatin domains associated to DNA methylation, H3K27me3 and H3K9ac marks. The median read count plot of each classified domain related to the histone marks H3K9ac (green), H3K27me3 (orange), global DNA methylation (pink) and accessible chromatin regions signal (purple) is presented.

(G) Repressive domains are larger than transcriptionally active domains. Boxplots representing the median size of the transcriptionally active domains (green), polycomb domains (purple) and repressive domains (pink). *** indicates pvalue $=6.04 \times 10^{-16}$, ** pvalue = 0.00164, *pvalue $= 9.62 \times 10^{-8}$ with Wilcoxon rank-sum test.

analysis revealed complete match with 96.1% of the investigated genes, 3% were found fragmented and less than 1% were missing (Figure 2B, Table S5). These data place the *Charmono* genome among the most complete published plant genomes. Protein-coding genes were then subjected to functional annotation using standard annotation resources. This include reciprocal best hits of Charmono proteins with UniProtKB Viridiplantae proteins, the description of the previously annotated enzymes, transcription factors and kinases identified by iTAK (Zheng et al., 2016), the transcription factors identified by PlantTFCat (Dai et al., 2013), and the InterPro analysis matching 31,853 proteins (Finn et al., 2017). To finish, the functional descriptions were merged and edited to follow GenBank functional annotation. Altogether, we attributed functional annotation to 61% of the predicted genes (Table S6).

Transposable elements (TEs) are major components of the structure and the function of genomes. To annotate TEs in the Charmono genome, we used the REPET package that exploits two pipelines, TE*denovo* and TE*annot* (Flutre et al., 2011). The TE*denovo* pipeline permitted us to determine TE consensus sequences in the melon genome. The output of TE*denovo* and melon genome were then analyzed by the TE*annot* pipeline (Quesneville et al., 2005) to determine genome-wide annotation of TE copies (Figure S3). We found TEs spanning 46.4% of the assembly, with 24.5% being LTR retrotransposons and 18.7% being TIR transposons (Figure S4A, Table S7). We analyzed the distribution of the insertion sites of the different class of TEs. As expected, most insertions mapped outside of coding sequences (Figures 2C, S4A, and S4B). However, we found LTR transposons more frequent within 3 kb intervals flanking coding regions of genes, while TIR transposons are more frequent further away (Figure 2C). This contrasted distribution of TIR DNA transposons and LTR retrotransposons may indicate selective pressures tolerating only LTR elements, to be close to genes.

The melon genome presents a complex chromatin topology, characterized by different degrees of organization

To gain insight into melon chromatin organization, we performed immuno-detection of H3K27me3 and H3K27me1, two histone post-translational modifications associated with constitutive and facultative heterochromatin, respectively (Liu et al., 2018). Interestingly, we observed a heterogeneous distribution of these covalent histone modifications, suggesting that distinct constitutive and facultative heterochromatin compartments exist in melon (Figures 3A and S5). We then examined active chromatin by immunostaining with an antibody directed against H3K9ac, an euchromatic mark linked to active transcription (Liu et al., 2018). We observed that H3K9ac was localized in foci scattered across the center of the nucleus, suggesting a physical proximity of clusters from transcriptionally active regions (Figure S5).

To further analyze Charmono chromatin architecture, we used Hi-C, a genome-wide chromatin conformation capture assay that detects DNA–DNA physical interactions (Liu, 2017; van Berkum et al., 2010) using leaf tissue (Tables S8 and S9). Analysis of whole-genome interaction matrix revealed as expected strong intra-chromosomic interactions, which is consistent with the chromosome territory model (Figures 1G and S6) (Cavalli and Misteli, 2013; Lanctôt et al., 2007; Woodcock and Ghosh, 2010).

To investigate the relationship between chromatin 3D folding and the observed compartmentalization, we integrated H3K27me3 ChIP-seq, H3K9ac ChIP-seq, MeDIP-seq, RNA-seq data and Hi-C data at chromosome-wide level (Figure 3). The visualization of the interaction matrix revealed a widespread compartmentalization allowing a physical interaction either between (1) gene rich genomic regions or (2) between large DNA methylation rich genomic regions, segregating constitutive heterochromatin and euchromatin (Figures 3B and 3C).

In mammals and plant, it was established at the megabase scale that the genome could be divided into two compartments, called A/B compartments by computing PCA eigenvector analysis of its genome-wide interaction matrix (Dixon et al., 2012; Dong et al., 2017; Fortin and Hansen, 2015; Hou et al., 2012). By inspecting the Charmono interaction map with principal component analysis, we found that chromatin regions could be divided into two groups in which the A compartments are associated with higher gene density and active epigenetic marks, while the B compartment has higher TE density and repressive epigenetic marks (Figures 3D and S7), supporting the fact that melon chromatin organization displays a strong compartmentalization.

Topologically associating domains (TADs) are defined as genomic regions containing sequences that interact more frequently with others in the same TAD than with those in different TADs (Dixon et al., 2012; Feng et al., 2014; Hou et al., 2012; Nora et al., 2012; Phillips-Cremins et al., 2013; Sexton et al., 2012). To detect the presence of folding domains in melon, we analyzed Hi-C interaction matrixes with the "TAD-separation score" technique after ICE normalization (Ramírez et al., 2018). We found a thousand folding domains with an average size of 100 kb (Figure 3E). To classify those folding domains regarding their chromatin status, we used a hierarchical clustering approach. We observed 3 types of folding domains: (1) Associated with the euchromatin mark H3K9ac, (2) associated with the facultative heterochromatin mark H3K27me3, and (3) associated with DNA methylation, a constitutive heterochromatin in which repressive domains are larger than transcriptionally active domains (Figures 3F, 3G, S8, S9, and S10). Altogether our results suggest that melon has a complex 3D chromatin architecture in which the local compartments are important structural units.

Genome-wide comparison of Charmono genome to other melon botanical group genomes

To investigate genomic variation between sequenced melon botanical groups, we compared the Charmono genome to the genomes of DHL92, Harukei-3, Payzawat, and HS lines. As expected, we found complete collinearity between the four genomes (Figure 4A). To estimate genome divergence, we computed the total number of SNPs in the genomes. We found Charmono, belonging to the *cantalupensis* botanical group increasingly more divergent from Harukei-3 (*reticulatus*), Payzawat (*inodorus*), and HS (ssp. *agrestis*) genomes, respectively (Figures 4B and 4C; Tables S10 and S11). The hybrid line DHL92 showed a global sequence divergence, estimated based on the total number of SNPs, intermediate between Harukei-3 and Payzawat. SNPs and Indels were shown to contribute differently to species divergence (Cheng et al., 2019). We calculated the total number of insertion and deletions between the genomes. Interestingly, we found twice more Indels in the comparison of Charmono to Payzawat than with Harukei-3 and HS (Table S11 and Figures 4B and 4C).

Indels are expected to be deleterious when they occur in coding regions where they can induce frameshifts (Figure S11). To test this prediction, we calculated the number of Indels as well as the SNPs in coding and noncoding regions. Two results stand out. First, we found as expected the majority of the SNPs and InDels, with an average of ~95%, in non-coding regions (Figure 4D, Table S12). The total number of SNPs in coding and non-coding regions showed similar propensity, with increasingly more SNPs with Harukei-3, Payzawat and HS genomes, respectively (Figure 4D). Non-triple Indels are predicted to be deleterious. We found about 86% of Indels in Harukei-3 and HS to be non-triple. This number reached more than 92% in the comparison with Payzawat (Figures 4E and 4F, Table S12).

Sequence divergence is a key in plant breeding. To test whether alleles associated with traits related to leaf, root, flower, and fruit development are shared between the sequenced melon botanical groups, we compared SNPs and Indels in genes specifically expressed (GSE) in leaf, fruits, and roots. We refer to these genes as Organ Specifically Expressed Genes (OSEG). As for the genome wide sequence divergence analysis, we found globally the degree of divergence of SNPs and InDels in OSEG reflecting the degree of divergence of Harukei-3, Payzawat, and HS genomes, relative to Charmono (Figure S12).

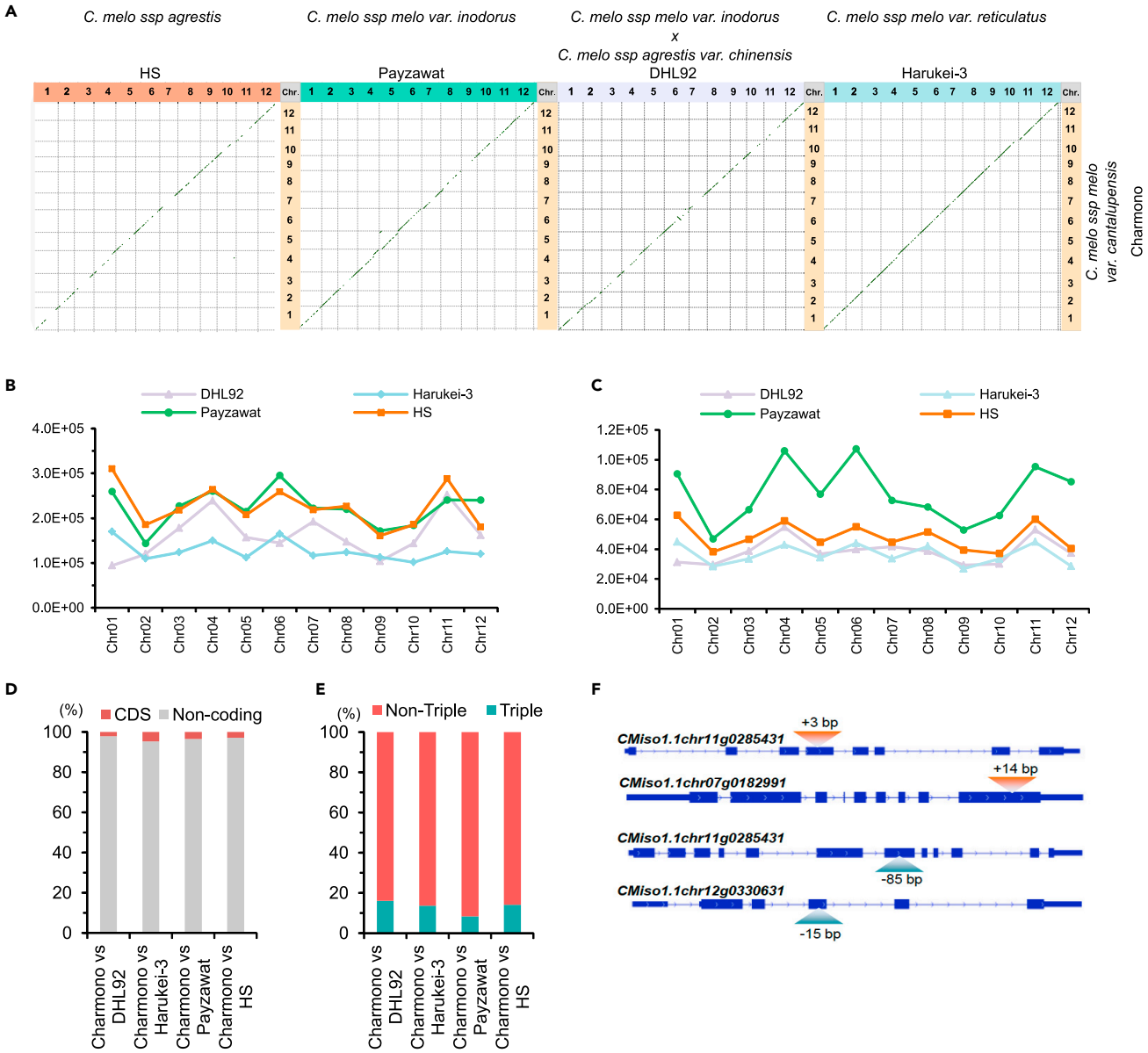


Figure 4. Genome-wide comparison of melon genomes

(A) Dot plot alignments comparing the chromosome sequence identity between Charmono and melon sequenced genomes (DHL92, Payzawat, Harukei and HS). Percentage of identity filter >75%.

(B and C) Number of substitutions (B) and InDels (C) per chromosome. Orange, green, purple and cyan colors indicate Charmonovs HS, CharmonovsPayzawat, Charmonovs DHL92 and CharmonovsHarukei-3 comparisons, respectively.

(D) Percentage of substitutions affecting a coding sequence region (CDS) or a non-coding region for each genome comparison.

(E) Percentage of non-triple or triple InDels for each genome comparison.

(F) Examples of CharmonovsHarukei-3 InDels affecting genes. Red and blue triangles indicate the position of the insertion or deletion, respectively.

Melon has retained the ancestral cucurbitaceae genome structure

In order to assess the paleohistory of the *Cucurbitaceae* within the eudicot family, we performed a comparative genomic investigation of squash (*Cucurbita moschata*), watermelon (*Citrullus lanatus*), bottle gourd (*Lagenaria siceraria*), melon (*Cucumis melo var.*Charmono), cucumber (*Cucumis sativus*), and pumpkin (*Cucurbita maxima*) genomes, together with the known ancestral eudicot karyotype (AEK structured with 7 protochromosomes and 9,022 protogenes, Murat et al., 2017), using the genome alignment parameters and ancestral genome reconstruction methods described in Pont et al. (2019) (Figure 5).

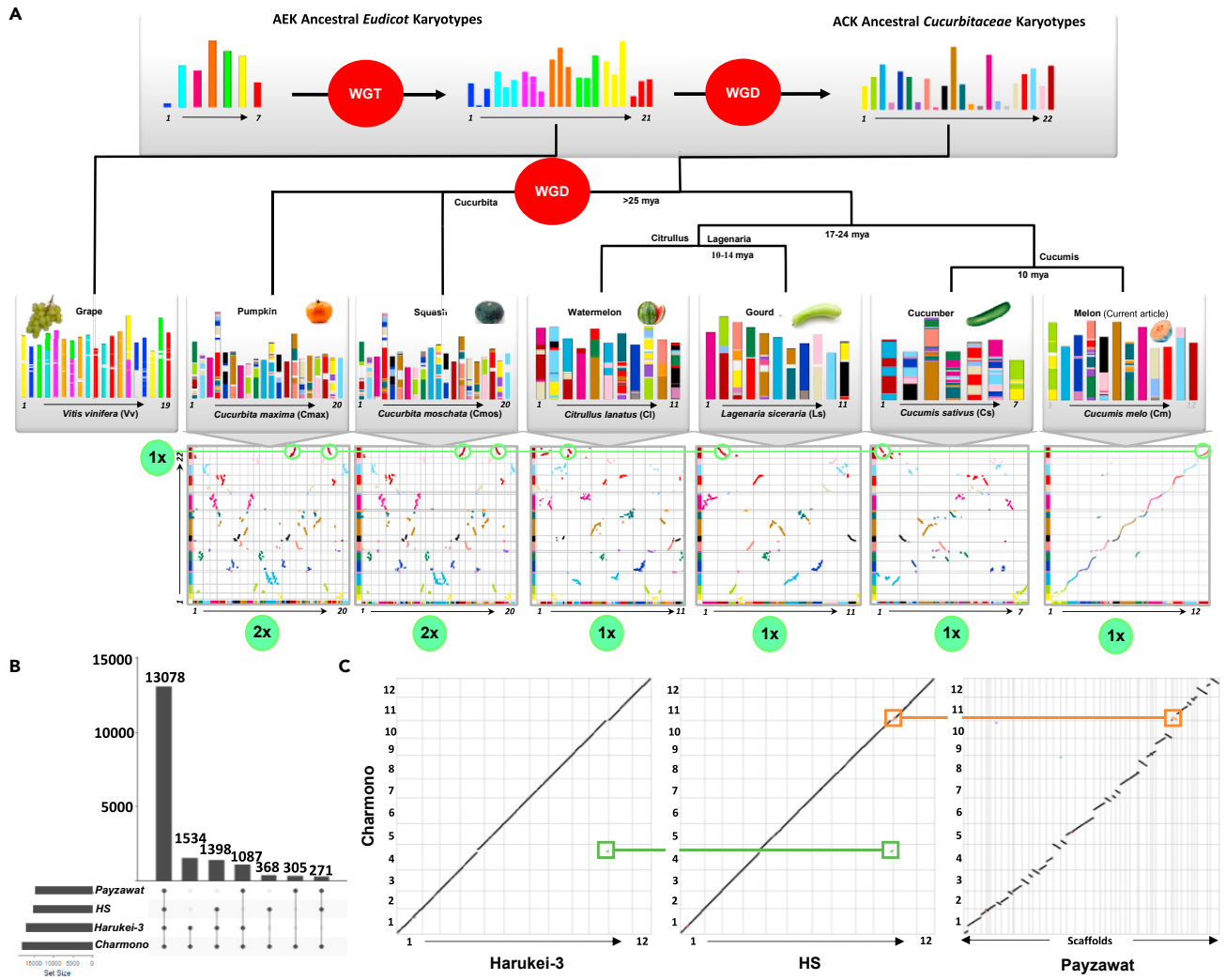


Figure 5. *Cucurbitaceae* paleoevolutionary history

(A) TOP-Evolutionary scenario of the modern *Cucurbitaceae* (squash, watermelon, gourd, melon, pumpkin and cucumber) genomes from the ancestral *Cucurbitaceae* karyotype (ACK) and the ancestral eudicot karyotype (AEK). The modern genomes are illustrated at the bottom with different colors reflecting the origin from the 22 ancestral chromosomes from ACK. Duplication (WGD) and triplication (WGT) events are shown with red dots on the tree branches, along with the shuffling events (fusions and fissions). BOTTOM- Complete dot-plot based deconvolution of the observed synteny and paralogy (dot-plot diagonals) between ACK (yaxis) and the investigated species (dot-plot xaxis). The synteny (paralogous and orthologous genes) relationships delivered between the modern *Cucurbitaceae* genomes and ACK are illustrated in green circles, as case example for ACK protochromosome 22.

(B) Gene conservation between Charmono*(cantalupensis)*,Harukei-3 *(reticulatus)*, Payzawat (*inodorus*) and HS (ssp. *agrestis*).

(C) Dotplot-based deconvolution of the synteny between Charmono*(cantalupensis)*, *vs*Harukei-3 *(reticulatus)*, vs HS (ssp *agrestis*) and *vs*Payzawat (*inodorus*), illustrating shared inversions between Charmono*vs*Payzawat and *vs* HS (illuminated in brown) on chromosome 11 and shared translocation on chromosome 4 between Charmono versus Harukei-3 and versus HS (illuminated in green).

AEK experienced a known whole genome triplication (WGT γ) to reach a 21-chromosomes intermediate, with the grape genome that resembles the most to the post-γ ancestor (Murat et al., 2017). AEK then experienced a whole genome duplication (WGD) specific to the *Cucurbitaceae* delivering an ancestral *Cucurbitaceae* karyotype (ACK) consisting of 22 protochromosomes (or Conserved Ancestral Regions, CARs) with 17,969 protogenes (i.e., conserved gene adjacencies between the investigated *Cucurbitaceae* genomes, Figure 5A top and Table S13). The complete dot-plot based deconvolution into the reconstructed CARs of the observed synteny and paralogy between ACK and the investigated species validates the 22 proposed protochromosomes as the origin of *Cucurbitaceae* (Figure 5A, bottom).

Our evolutionary scenario, reconciling the modern genome structures to the founder ACK, clearly established that melon has retained the ACK genome structure in contrast to the other investigated species that experienced massive shuffling events with, for example, melon chromosomes 2-6-8-10-12 that correspond to quasi-entire non-rearranged protochromosomes (Figure 5A, bottom). The 11 watermelon chromosomes derive from ACK followed by 27 chromosomal fissions and 28 fusions. The modern gourd genome (11 chromosomes) has been shaped from ACK through 19 chromosomal fissions and 20 fusions, and cucumber (7 chromosomes) through 6 fissions and 11 fusions. Finally, the squash and pumpkin genomes experienced a specific whole-genome duplication (delivering a 24-chromosomes ancestor intermediate) and massive chromosomal rearrangements to reach their modern structure into 20 chromosomes (Figure 5A, bottom).

We then investigated the relatedness between melon botanical groups, Charmono (*cantaloupensis*), Harukei-3 (*reticulatus*), Payzawat (*inodorus*), and HS (ssp. a*grestis*) at both the gene and karyotype conservation levels. Overall, the proposed evolutionary scenario complements and refines previous proposed *Cucurbitaceae* paleohistory analyses from 12 ancestral chromosomes (Wu et al., 2017) or 15 ancestral chromosomes (Xie et al., 2019) at the basis of modern *Cucurbitaceae* genomes. Pattern of gene conservation observed between the four cultivars illustrates higher gene retention between Charmono*(cantaloupensis*) and Harukei-3 (*reticulatus*) compared with the two others *(*Payzawat and HS) (Figure 5B). Dot plot-based deconvolution of the synteny between the four cultivars allows identifying shared chromosomal rearrangements illustrating common ancestry between Payzawat-HS sharing an inversion on chromosome 11 and between Harukei-3-HS sharing a chromosomal translocation on chromosome 4. Overall, cumulative evidences of conserved genes and conserved chromosomal structures established the following genomics relatedness between the four investigated cultivars: Charmono>Harukei-3 >Payzawat> HS.

We deliver the *Cucurbitaceae* paleohistory at both the inter-specific (comparing 6 genomes) and intra-specific (comparing 4 genomes) levels from the first reconstructed ancestral *Cucurbitaceae* karyotype (ACK) and update the publicly available catalog of paralogous and orthologous gene relationships between extant Cucurbit genomes (https://urgi.versailles.inra.fr/synteny/cucurbit). The current study delivers comparative genomics data among major *Cucurbitaceae* species to conduct translational research of conserved genes driving key agronomical traits in *Cucurbitaceae*.

DISCUSSION

Melon displays a large genetic diversity divided into subspecies and botanical groups. Using several phenotypic descriptors, Pitrat reported ten and five groups for subspecies *melo* and *agrestis*, respectively (Pitrat, 2008); and the International Union for the Protection of New Varieties of Plants (UPOV) used more exhaustive list of characters to describe melon accessions. More recently, resequencing of 1,175 accessions, representing the global diversity of the species, identified three independent domestication events, two in India and one in Africa, likely leading to all melon botanical groups (Zhao et al., 2019). Sixteen agronomic traits and loci significantly associated with fruit mass, quality, and morphological characters likely selected during domestications were identified. Still, the best way to estimate sequence diversity is genome sequencing of accessions representing different botanical groups. This will not only identify sequence divergences within a set of genes or traits, but also estimate genome-wide introgressions that cannot be detected based on phenotype divergence or sequence comparisons to a reference genome.

Cantaloupe melon varieties represent one of the most cultivated melon botanical groups. We assembled *C. melo* var. *cantalupensis* genome at the chromosome level. The genome was generated using the PacBio long-read sequencing technology and scaffolding using genetic and Bionano optical mapping and Hi-C contact maps (Figure 1). The N50 of the contigs and scaffolds were over 8.5 Mb and 12.3 Mb, respectively (Table S2). The scaffolds also almost corresponded to the arms of the chromosomes. The *C. melo* var. *cantalupensis* Charmono reference genome sequence constitutes a significant contribution for research in cucurbits. It will ease QTL cloning, functional characterization and comparative genomic analysis with other melon botanical groups.

Annotation of Charmono genome identified 31,348 coding genes and 2,588 non-protein coding gene, approaching what was described in other cucurbitaceae (Table 1). Still significant variation in the gene number is to be reported. For instance, in Payzawat genome only 22,924 genes were identified (Zhang et al., 2019). In contrast, in Harukei-3 and DHL92 genomes, 33,829 and 29,980 genes were identified respectively

Table 1. Coding and non-coding genes predicted using Eugene

Gene set	Number
Protein coding genes	
Coding genes	31,348
Mean gene length (bp)	3706
GC per cent in exons	40.85
GC per cent in introns	33.42
Identified classes of non-coding RNA	
lncRNA	1961
mirRNA	113
SnoRNA	110
rRNA	72
Other rfam annotation	194
tRNA	627

(Yano et al., 2020; Castanera et al., 2020). These differences are likely due to the methods used to automatically annotate the sequenced genomes and genome divergence between the botanical groups. To improve our use of the sequenced genomes, high quality gene annotation is a prerequisite. This will help to define the core genes that are shared between the botanical groups and a set of genes that are partially shared. Such melon pangenome could help genome-wide comparison of sequenced accessions, genome wide association studies, and gene functional analysis. To do this, a consortium of specialists need to be hired with the objective of establishing a common strategy for automatic annotation, evaluation of the quality of the annotation, and defining the unique descriptor for gene naming and functional annotation.

TEs are the main constituents of plant genomes (Zhang and Wessler, 2004). Nevertheless, genome-wide annotation of transposons is still considered an option in investigation of genome organization and gene functional analysis. Previous analyses on the melon genome revealed a massive amplification of TEs in centromeric and pericentromeric regions, likely as a consequence of the counter-selection of deleterious insertions in genic regions (Castanera et al., 2020). Similarly, we found a massive amplification of TEs in centromeric and pericentromeric regions in the Charmono genome. Interestingly, we found insertion of LTR transposons more tolerated at the vicinity of coding sequences than TIR elements (Figures 2,3 and S4). This is in contrast with what was reported in the barley genome (Wicker et al., 2017).

Hi-C analysis revealed three levels of chromatin 3D organization: (1) compartments, (2) domains, and (3) loops. Even though plant and metazoan nuclei share common characteristics, they also show significant variations. For instance, in line with our work on the melon, analysis of several plant species, including Arabidopsis, rice, barley, maize, tomato, sorghum, foxtail millet, and cotton showed that plant genomes are partitioned into global A/B compartments (Concia et al., 2020; Dong et al., 2017; Feng et al., 2014; Grob et al., 2014; Mascher et al., 2017; Rodriguez-Granados et al., 2016). In plants, global A/B compartment partitions are stable across tissues while rare domain border switches associated with changes in gene expression can be observed (Dong et al., 2020). However, chromatin organization could diverge between horticultural groups and it will be interesting to compare chromatin topology between melon horticultural groups, for instance, between *melo* and *agrestis* subspecies. For genetic validation, such analysis could be carried out both in parental lines and in homozygous recombinants. Carrying such investigations will likely identify phenotypes that are associated with chromatin 3D organization. In metazoans, chromatin is also organized in TADs, structures known to facilitate the formation of enhancer-promoter contacts and to mediate gene regulation (Göndör and Ohlsson, 2018; Gonzalez-Sandoval and Gasser, 2016). We discovered a widespread presence of TAD-like domains in melon, which is consistent with previous findings (Figure S6) (Concia et al., 2020; Dong et al., 2017). Our chromatin 3D organization will likely open new avenues to unravel the complex connection between chromatin architecture and control of gene expression in crops.

Through annotation and genome-wide comparison of Charmono genome to Harukei-3 (*reticulatus*), Payzawat (*inodorus*), and HS (ssp. *agrestis*) genomes, we calculated genome divergence using the density

of SNPs or Indels as proxy (Figure 4, Table S12). Our data show that Charmono *cantalupensis* melon is more related to a *reticulatus* type of melon (Harukei-3) than to an *inodorus* (Payzawat) or an agrestis (HS). The line DHL92, a hybrid between *inodorus* and *agrestis* melon, showed a global sequence divergence intermediate between *reticulatus* (Harukei-3) and *inodorus* (Payzawat) horticulture groups. However, we found some DHL92 chromosomes more related to *agrestis* genome (Figure 4B, HS, chromosome 12) others more related to *inodorus* genome (Figure 4B, Payzawat, chromosome 11).

Rates of SNPs and Indels are known to vary between coding and non-coding regions of the genome. The majority of the SNPs, estimated to 93%–99% of the total, were in non-coding regions and thus they are likely neutral. Non-triple indels are supposed to alter the function of the coding genes leading to changes in the reading frame. We found the majority of the indels, estimated to 84%–92%, to be non-triple (Figure 4).

Assuming that OSEG are indicators of root, leaf, flower, or fruit development, we used the sequence divergence in OSEG as a proxy of organ specific sequence divergence between melon botanical groups. We found sequence divergence in OSEG reflecting the degree of divergence at the genome level, suggesting no enrichment of polymorphism in sequences associated with root, leaf, flower, or fruit development between the four melon botanical groups. To increase sequence diversity, the production of populations such as Multi-parent advanced generation intercrosses, known as MAGIC populations, from different melon botanical groups will likely provide important sources of genetic diversity for melon improvement (Huang et al., 2015).

The recent *de novo* genome sequencing technologies are providing an unprecedented opportunity to compare modern genomes with each other and to infer their evolutionary history from the reconstructed genomes of their most recent common ancestors. Using Charmono high-quality genome assembly and its annotation and publicly available cucurbits genomes, we reconstructed the paleohistory of the *Cucurbitaceae* within the eudicot family. We found the ancestral *Cucurbitaceae* karyotype consisting of 22 protochromosomes and 17,969 protogenes (Figure 5). Genome-wide comparison of Charmono genome to the *Cucurbitaceae* ancestral genome revealed that the melon has retained the ancestral genome structure whereas other cucurbit species experienced massive shuffling events and whole-genome duplication (Figure 5). This reinforces the choice of melon genome as a reference genome for structural and functional analysis of paralogous and orthologous gene relationships among major *Cucurbitaceae* species to conduct translational research of conserved genes driving key agronomical traits. Because paleogenomic studies rely on gene synteny analysis, this case study also illustrates the importance of high quality genome assembly and annotation for functional analysis and comparative genomics.

Limitations of the study

We sequenced the genome of *Cucumis melo var cantalupensis* Charmono cultivar and revealed that this genome presents a complex 3D chromatin architecture in which local compartments are important structural units. However, more functional studies are needed to assess the importance of chromatin topology in melon development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Plant material and growth conditions
- METHOD DETAILS
  - Genome sequencing and assembly
  - Charmono genome annotation
  - Melon epigenome
  - Methylated-DNA Immunoprecipitation, MeDIP-seq sequencing
  - ATAC-seq

- Gene expression analysis
- Genome wide comparison of melon genomes
- Sequence polymorphisms in Organ Specifically Expressed Genes
- Evolutionary analysis of *Cucurbitaceae* genomes

- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103696.

ACKNOWLEDGMENTS

The authors thank Pascal Audigier, Florie Vion, and Holger Ornstrup for taking care of the plant and the research facilities provided by the Institute of Plant-Science Paris-Saclay (IPS2, France). We thank the Genotoul bioinformatics platform Toulouse Midi-Pyrenées and gratefully acknowledge support from the Pôle Scientifique de Modélisation Numérique of the ENS de Lyon for providing computing resources. JSa thanks Raphael Flores (URGIINRA University Paris Saclay Versailles France) for making the Cucurbit synteny data publicly available at https://urgi.versailles.inra.fr/synteny/in the frame of the Institut Carnot Plant2Pro Project 'SyntenyViewer'. Financial support was provided by the European Research Council (ERC-SEXY-PARTH, 341076), the ANR EPISEX Project (ANR-17-CE20-0019), the ANR NECTAR Project (ANR-19-CE20-0023), LabEx Saclay Plant Sciences (SPS) (ANR-10-LABX-40-SPS), and the Plant Biology and Breeding Department of INRAE.

AUTHOR CONTRIBUTIONS

ABe, ABo, and CD initiated the project and chose the material for whole genome sequencing; ABe and ABo designed the research; CD obtained and provided the breeding line Charentais Mono; CPi, VG, DL, SA, CT, CPo, JSz and ABo performed experiments; CPi, AD, JT, MV, WM, VS, MB, MBenh, MZ, ABo and ABe analyzed data and conducted bioinformatic analysis; CH and JSa conducted the comparative genomics and ancestral genome analyses. CPi, MBenh, ABo and ABe wrote the paper. All authors discussed results, commented on the paper and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 5, 2021
Revised: December 1, 2021
Accepted: December 20, 2021
Published: January 21, 2022

REFERENCES

Akdemir, K.C., and Chin, L. (2015). HiCPlotter integrates genomic data with interaction matrices. Genome Biol. *16*, 198. https://doi.org/10.1186/s13059-015-0767-1.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics *43*, 1–33. https://doi.org/10.1002/0471250953.bi1110s43.

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., and Lander, E.S. (2010). Hi-C: amethod to study the three-dimensional architecture of genomes. J.Vis. Exp. *1869*. https://doi.org/10.3791/1869.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: amethod for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. *109*, 21–21 29 29. https://doi.org/10.1002/0471142727.mb2129s109.

Castanera, R., Ruggieri, V., Pujol, M., Garcia-Mas, J., and Casacuberta, J.M. (2020). An improved melon reference genome with single-molecule sequencing uncovers arecent burst of transposable elements with potential impact on genes. Front. Plant Sci. *10*, 1815. https://doi.org/10.3389/fpls.2019.01815.

Cavalli, G., and Misteli, T. (2013). Functional implications of genome topology. Nat. Struct. Mol. Biol. *20*, 290–299. https://doi.org/10.1038/nsmb.2474.

Cheng, H., Liu, J., Wen, J., Nie, X., Xu, L., Chen, N., Li, Z., Wang, Q., Zheng, Z., Li, M., et al. (2019). Frequent intra- and inter-species introgression shapes the landscape of genetic variation in bread wheat. Genome Biol. *20*, 136. https://doi.org/10.1186/s13059-019-1744-x.

Chin, C.-S., Alexander, D.H., Marks, P., Klammer, A.A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E.E., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat. Methods *10*, 563–569. https://doi.org/10.1038/nmeth.2474.

Choi, J.Y., Lye, Z.N., Groen, S.C., Dai, X., Rughani, P., Zaaijer, S., Harrington, E.D., Juul, S., and Purugganan, M.D. (2020). Nanopore sequencing-based genome assembly and evolutionary genomics of circum-basmati rice. Genome Biol. *21*, 21. https://doi.org/10.1186/s13059-020-1938-2.

Chomicki, G., Schaefer, H., and Renner, S.S. (2020). Origin and domestication of *Cucurbitaceae* crops: insights from phylogenies, genomics and archaeology. New Phytol. *226*, 1240–1255. https://doi.org/10.1111/nph.16015.

Christenhusz, M.J.M., and Byng, J.W. (2016). The number of known plants species in the world and its annual increase. Phytotaxa *261*, 201–217. https://doi.org/10.11646/phytotaxa.261.3.1.

Concia, L., Veluchamy, A., Ramirez-Prado, J.S., Martin-Ramirez, A., Huang, Y., Perez, M., Domenichini, S., Rodriguez Granados, N.Y., Kim, S., Blein, T., et al. (2020). Wheat chromatin architecture is organized in genome territories and transcription factories. Genome Biol. *21*, 104. https://doi.org/10.1186/s13059-020-01998-1.

Dai, X., Sinharoy, S., Udvardi, M., and Zhao, P.X. (2013). PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. BMC Bioinformatics *14*, 321. https://doi.org/10.1186/1471-2105-14-321.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dong, P., Tu, X., Chu, P.-Y., Lü, P., Zhu, N., Grierson, D., Du, B., Li, P., and Zhong, S. (2017). 3D chromatin architecture of large plant genomes determined by local A/B compartments. Mol. Plant *10*, 1497–1509. https://doi.org/10.1016/j.molp.2017.11.005.

Dong, P., Tu, X., Li, H., Zhang, J., Grierson, D., Li, P., and Zhong, S. (2020). Tissue-specific Hi-C analyses of rice, foxtail millet and maize suggest non-canonical function of plant chromatin domains. J.Integr. Plant Biol. *62*, 201–217. https://doi.org/10.1111/jipb.12809.

El Hadidi, M.N., and Hosni, H.A. (1996). Biodiversity in the flora of Egypt. In The Biodiversity of African Plants, L.J.G. van der Maesen, X.M. van der Burgt, and J.M. van Medenbach de Rooy, eds. (Springer), pp. 785–787.

Ellinghaus, D., Kurtz, S., and Willhoeft, U. (2008). LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics *9*, 18. https://doi.org/10.1186/1471-2105-9-18.

Endl, J., Achigan-Dako, E.G., Pandey, A.K., Monforte, A.J., Pico, B., and Schaefer, H. (2018). Repeated domestication of melon (Cucumis melo) in Africa and Asia and a new close relative from India. Am. J.Bot. *105*, 1662–1671. https://doi.org/10.1002/ajb2.1172.

Feng, S., Cokus, S.J., Schubert, V., Zhai, J., Pellegrini, M., and Jacobsen, S.E. (2014). Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in Arabidopsis. Mol.Cell *55*, 694–707. https://doi.org/10.1016/j.molcel.2014.07.008.

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: the protein families database. Nucleic Acids Res. *42*, D222–D230. https://doi.org/10.1093/nar/gkt1223.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztanyi, Z., El-Gebali, S., Fraser, M., et al. (2017). InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. *45*, D190–D199. https://doi.org/10.1093/nar/gkw1107.

Flutre, T., Duprat, E., Feuillet, C., and Quesneville, H. (2011). Considering transposable element diversification in de novo annotation approaches. PLoS One *6*, e16526. https://doi.org/10.1371/journal.pone.0016526.

Fortin, J.-P., and Hansen, K.D. (2015). Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. Genome Biol. *16*, 180. https://doi.org/10.1186/s13059-015-0741-y.

Garcia-Mas, J., Benjak, A., Sanseverino, W., Bourgeois, M., Mir, G., González, V.M., Hénaff, E., Câmara, F., Cozzuto, L., Lowy, E., et al. (2012). The genome of melon (Cucumis melo L.). Proc. Natl. Acad. Sci. U S A *109*, 11872–11877. https://doi.org/10.1073/pnas.1205415109.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. arXiv, 1207.3907 [q-bio]. https://arxiv.org/abs/1207.3907.

Girgis, H.Z. (2015). Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinform *16*, 227. https://doi.org/10.1186/s12859-015-0654-5.

Göndör, A., and Ohlsson, R. (2018). Enhancer functions in three dimensions: beyond the flat world perspective. F1000Res. *7*, F1000. Faculty Rev-1681. https://doi.org/10.12688/f1000research.13842.1.

Gonzalez-Sandoval, A., and Gasser, S.M. (2016). On TADs and LADs: spatial control over geneexpression. Trends Genet. *32*, 485–495. https://doi.org/10.1016/j.tig.2016.05.004.

Grob, S., Schmid, M.W., and Grossniklaus, U. (2014). Hi-C analysis in Arabidopsis identifies the KNOT, astructure with similarities to the flamenco locus of Drosophila. Mol.Cell *55*, 678–693. https://doi.org/10.1016/j.molcel.2014.07.009.

Guo, S., Zhang, J., Sun, H., Salse, J., Lucas, W.J., Zhang, H., Zheng, Y., Mao, L., Ren, Y., Wang, Z., et al. (2013). The draft genome of watermelon (Citrullus lanatus) and resequencing of 20 diverse accessions. Nat.Genet. *45*, 51–58. https://doi.org/10.1038/ng.2470.

Guo, S., Zhao, S., Sun, H., Wang, X., Wu, S., Lin, T., Ren, Y., Gao, L., Deng, Y., Zhang, J., et al. (2019). Resequencing of 414 cultivated and wild watermelon accessions identifies selection for fruit quality traits. Nat.Genet. *51*, 1616–1623. https://doi.org/10.1038/s41588-019-0518-4.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hoede, C., Arnoux, S., Moisset, M., Chaumier, T., Inizan, O., Jamilloux, V., and Quesneville, H. (2014). PASTEC: an automatic transposable element classification tool. PLoS One *9*, e91929. https://doi.org/10.1371/journal.pone.0091929.

Hosmani, P.S., Flores-Gonzalez, M., Geest, H.v.d., Maumus, F., Bakker, L.V., Schijlen, E., Haarst, J.v., Cordewener, J., Sanchez-Perez, G., Peters, S., et al. (2019). An improved de novo assembly and annotation of the tomato reference genome using single-molecule sequencing, Hi-C proximity ligation and optical maps. bioRxiv, 767764. https://doi.org/10.1101/767764.

Hou, C., Li, L., Qin, Z.S., and Corces, V.G. (2012). Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. Mol.Cell *48*, 471–484. https://doi.org/10.1016/j.molcel.2012.08.031.

Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., et al. (2009). The genome of the cucumber, Cucumis sativus L. Nat.Genet. *41*, 1275–1281. https://doi.org/10.1038/ng.475.

Huang, B.E., Verbyla, K.L., Verbyla, A.P., Raghavan, C., Singh, V.K., Gaur, P., Leung, H., Varshney, R.K., and Cavanagh, C.R. (2015). MAGIC populations in crops: current status and future prospects. TAG.Theoretical and applied genetics. Theor.Angew.Genet. *128*, 999–1017. https://doi.org/10.1007/s00122-015-2506-0.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S., Cook, H., Mende, D., Letunic, I., Rattei, T., Jensen, L., et al. (2018). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic AcidsRes. *47*. https://doi.org/10.1093/nar/gky1085.

Janick, J., Paris, H.S., and Parrish, D.C. (2007). The cucurbits of mediterranean antiquity: identification of taxa from ancient images and descriptions. Ann. Bot. London *100*, 1441–1457. https://doi.org/10.1093/aob/mcm242.

Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: genome-scale protein function classification. Bioinformatics *30*, 1236–1240. https://doi.org/10.1093/bioinformatics/btu031.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet.Genome Res. *110*, 462–467. https://doi.org/10.1159/000084979.

Khelik, K., Lagesen, K., Sandve, G.K., Rognes, T., and Nederbragt, A.J. (2017). NucDiff: in-depth characterization and annotation of differences between two sets of DNA sequences. BMC Bioinformatics *18*, 338. https://doi.org/10.1186/s12859-017-1748-z.

Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., and Frith, M.C. (2011). Adaptive seeds tame genomic sequence comparison. Genome Res. https://doi.org/10.1101/gr.113985.110.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. *27*, 722–736. https://doi.org/10.1101/gr.215087.116.

Koster, J., and Rahmann, S. (2018). Snakemake-a scalable bioinformatics workflow engine. Bioinformatics *34*, 3600. https://doi.org/10.1093/bioinformatics/bty350.

Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278. https://doi.org/10.1186/s13059-019-1910-1.

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., and Ussery, D.W. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. *35*, 3100–3108. https://doi.org/10.1093/nar/gkm160.

Lanctôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. Nat. Rev. Genet. *8*, 104–115. https://doi.org/10.1038/nrg2041.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.

Latrasse, D., Rodriguez-Granados, N.Y., Veluchamy, A., Mariappan, K.G., Bevilacqua, C., Crapart, N., Camps, C., Sommard, V., Raynaud, C., Dogimont, C., et al. (2017). The quest for epigenetic regulation underlying unisexual flower development in Cucumis melo. Epigenet.Chromatin. *10*, 22. https://doi.org/10.1186/s13072-017-0132-6.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lian, Q., Fu, Q., Xu, Y., Hu, Z., Zheng, J., Zhang, A., He, Y., Wang, C., Xu, C., Chen, B., et al. (2021). QTLs and candidate genes analyses for fruit size under domestication and differentiation in melon (*Cucumis melo* L.) based on high resolution maps. BMC Plant Biol. *21*, 126. https://doi.org/10.1186/s12870-021-02904-y.

Liu, C. (2017). In situ Hi-C library preparation for plants to study their three-dimensional chromatin interactions on a genome-wide scale. Methods Mol.Biol. *1629*, 155–166. https://doi.org/10.1007/978-1-4939-7125-1_11.

Liu, C., Cheng, Y.-J., Wang, J.-W., and Weigel, D. (2017). Prominent topologically associated domains differentiate global chromatin packing in rice from Arabidopsis. Nat. Plants *3*, 742–748. https://doi.org/10.1038/s41477-017-0005-9.

Liu, Y., Tian, T., Zhang, K., You, Q., Yan, H., Zhao, N., Yi, X., Xu, W., and Su, Z. (2018). PCSD: a plant chromatin state database. Nucleic Acids Res. *46*, D1157–D1167. https://doi.org/10.1093/nar/gkx919.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. Embnet J. *17*, 10–12. https://doi.org/10.14806/ej.17.1.200.

Lowe, T.M., and Chan, P.P. (2016). tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. *8*, 54–57. https://doi.org/10.1093/nar/gkw413.

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: a fast and versatile genome alignment system. PLoS Comput.Biol. *14*, e1005944. https://doi.org/10.1371/journal.pcbi.1005944.

Mascher, M., Gundlach, H., Himmelbach, A., Beier, S., Twardziok, S.O., Wicker, T., Radchuk, V., Dockter, C., Hedley, P.E., Russell, J., et al. (2017). A chromosome conformation capture ordered sequence of the barley genome. Nature *544*, 427–433. https://doi.org/10.1038/nature22043.

Mayjonade, B., Gouzy, J., Donnadieu, C., Pouilly, N., Marande, W., Callot, C., Langlade, N., and Muños, S. (2016). Extraction of high-molecular-weight genomic DNA for long-read sequencing of single molecules. BioTechniques *61*, 203–205. https://doi.org/10.2144/000114460.

Moing, A., Allwood, J.W., Aharoni, A., Baker, J., Beale, M.H., Ben-Dor, S., Biais, B., Brigante, F., Burger, Y., Deborde, C., et al. (2020). Comparative metabolomics and molecular phylogenetics of melon (Cucumis melo, cucurbitaceae) biodiversity. Metabolites *10*, 121. https://doi.org/10.3390/Metabo10030121.

Mölder, F., Jablonski, K.P., Letcher, B., Hall, M.B., Tomkins-Tinch, C.H., Sochat, V., Forster, J., Lee, S., Twardziok, S.O., Kanitz, A., et al. (2021). Sustainable data analysis with Snakemake. F1000Res. *10*, 33. https://doi.org/10.12688/f1000research.29032.2.

Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A.C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. *35*, W182–W185. https://doi.org/10.1093/nar/gkm321.

Murat, F., Armero, A., Pont, C., Klopp, C., and Salse, J. (2017). Reconstructing the genome of the most recent common ancestor of flowering plants. Nat.Genet. *49*, 490–496. https://doi.org/10.1038/ng.3813.

Naudin, C. (1859). Essais d'une monographie des espèces et des variétés du genre *Cucumis*. Ann.des Sci.Nat. Botanique *4*, 5–87.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385. https://doi.org/10.1038/nature11049.

Paris, H.S. (2016). Overview of the origins and history of the five major cucurbit crops: issues for ancient DNA analysis of archaeological specimens. Vegetation Hist.Archaeobot. *25*, 405–414. https://doi.org/10.1007/s00334-016-0555-1.

Pereira, L., Ruggieri, V., Pérez, S., Alexiou, K.G., Fernández, M., Jahrmann, T., Pujol, M., and Garcia-Mas, J. (2018). QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. BMC Plant Biol. *18*. https://doi.org/10.1186/s12870-018-1537-5.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3-D organization of genomes during lineage commitment. Cell *153*, 1281–1295. https://doi.org/10.1016/j.cell.2013.04.053.

Pitrat, M. (2008). Melon. In Handbook of Crop Breeding, N.F. Prohens J, ed. (Springer), pp. 283–315. https://doi.org/10.1007/978-0-387-30443-4_9.

Pitrat, M. (2013). Phenotypic diversity in wild and cultivated melons (Cucumis melo). Plant Biotechnol. *30*, 273–278. https://doi.org/10.5511/plantbiotechnology.13.0813a.

Pont, C., Wagner, S., Kremer, A., Orlando, L., Plomion, C., and Salse, J. (2019). Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. Genome Biol. *20*, 29. https://doi.org/10.1186/s13059-019-1627-1.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Auwera, G.A.V.d., Kling, D.E., Gauthier, L.D., Levy-Moonshine, A., Roazen, D., et al. (2017). Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv, 201178. https://doi.org/10.1101/201178.

Quesneville, H., Bergman, C.M., Andrieu, O., Autard, D., Nouaud, D., Ashburner, M., and Anxolabehere, D. (2005). Combined evidence annotation of transposable elements in genome sequences. PLoS Comput.Biol. *1*, e22. https://doi.org/10.1371/journal.pcbi.0010022.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

Ramírez, F., Bhardwaj, V., Arrigoni, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., and Manke, T. (2018). High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nat. Commun. *9*, 189. https://doi.org/10.1038/s41467-017-02525-w.

Rodriguez-Granados, N.Y., Ramirez-Prado, J.S., Veluchamy, A., Latrasse, D., Raynaud, C., Crespi, M., Ariel, F., and Benhamed, M. (2016). Put your 3D glasses on: plant chromatin is on show. J.Exp.Bot. *67*, 3205–3221. https://doi.org/10.1093/jxb/erw168.

Ruggieri, V., Alexiou, K.G., Morata, J., Argyris, J., Pujol, M., Yano, R., Nonaka, S., Ezura, H., Latrasse, D., Boualem, A., et al. (2018). An improved assembly and annotation of the melon (Cucumis melo L.) reference genome. Sci. Rep. *8*, 8088. https://doi.org/10.1038/s41598-018-26416-2.

Sakai, H., Naito, K., Ogiso-Tanaka, E., Takahashi, Y., Iseki, K., Muto, C., Satou, K., Teruya, K., Shiroma, A., Shimoji, M., et al. (2015). The power of single molecule real-time sequencing technology in the de novo assembly of a

eukaryotic genome. Sci. Rep. *5*, 16780. https://doi.org/10.1038/srep16780.

Sallet, E., Gouzy, J., and Schiex, T. (2014). EuGene-PP: a next-generation automated annotation pipeline for prokaryotic genomes. Bioinformatics *30*, 2659–2661. https://doi.org/10.1093/bioinformatics/btu366.

Salse, J. (2016). Ancestors of modern plant crops. Curr. Opin. Plant Biol. *30*, 134–142. https://doi.org/10.1016/j.pbi.2016.02.005.

Schwarzer, W., Abdennur, N., Goloborodko, A., Pekowska, A., Fudenberg, G., Loe-Mie, Y., Fonseca, N.A., Huber, W., Haering, C.H., Mirny, L., and Spitz, F. (2017). Two independent modes of chromatin organization revealed by cohesin removal. Nature *551*, 51–56. https://doi.org/10.1038/nature24281.

Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472. https://doi.org/10.1016/j.cell.2012.01.010.

Sun, H., Wu, S., Zhang, G., Jiao, C., Guo, S., Ren, Y., Zhang, J., Zhang, H., Gong, G., Jia, Z., et al. (2017). Karyotype stability and unbiased fractionation in the paleo-allotetraploid *Cucurbita*genomes. Mol. Plant *10*, 1293–1306. https://doi.org/10.1016/j.molp.2017.09.003.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat.Biotechnol. *28*, 511–515. https://doi.org/10.1038/nbt.1621.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One *9*, e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang, X., Xu, Y., Zhang, S., Cao, L., Huang, Y., Cheng, J., Wu, G., Tian, S., Chen, C., Liu, Y., et al. (2017). Genomic analyses of primitive, wild and cultivated citrus provide insights into asexual reproduction. Nat.Genet. *49*, 765–772. https://doi.org/10.1038/ng.3839.

Weisenfeld, N.I., Kumar, V., Shah, P., Church, D.M., and Jaffe, D.B. (2017). Direct determination of diploid genome sequences. Genome Res. *27*, 757–767. https://doi.org/10.1101/gr.214874.116.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., et al. (2007). A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet. *8*, 973–982. https://doi.org/10.1038/nrg2165.

Wicker, T., Schulman, A.H., Tanskanen, J., Spannagl, M., Twardziok, S., Mascher, M., Springer, N.M., Li, Q., Waugh, R., Li, C., et al. (2017). The repetitive landscape of the 5100 Mbp barley genome. Mobile DNA *8*, 22. https://doi.org/10.1186/s13100-017-0102-3.

Wingett, S.W., and Andrews, S. (2018). FastQ Screen: a tool for multi-genome mapping and quality control. F1000Res. *7*, 1338. https://doi.org/10.12688/f1000research.15931.2.

Wolff, J., Bhardwaj, V., Nothjunge, S., Richard, G., Renschler, G., Gilsbach, R., Manke, T., Backofen, R., Ramírez, F., and Grüning, B.A. (2018). Galaxy HiCExplorer: a web server for reproducible Hi-C data analysis, quality control and visualization. Nucleic Acids Res. *46*, W11–W16. https://doi.org/10.1093/nar/gky504.

Wolff, J., Rabbani, L., Gilsbach, R., Richard, G., Manke, T., Backofen, R., and Grüning, B.A. (2020). Galaxy HiCExplorer 3: a web server for reproducible Hi-C, capture Hi-C and single-cell Hi-C data analysis, quality control and visualization. Nucleic Acids Res. *48*, W177–W184. https://doi.org/10.1093/nar/gkaa220.

Woodcock, C.L., and Ghosh, R.P. (2010). Chromatin higher-order structure and dynamics. Cold Spring Harb. Perspect. Biol. *2*. https://doi.org/10.1101/cshperspect.a000596.

Wu, S., Shamimuzzaman, M., Sun, H., Salse, J., Sui, X., Wilder, A., Wu, Z., Levi, A., Xu, Y., Ling, K.-S., and Fei, Z. (2017). The bottle gourd genome provides insights into Cucurbitaceae evolution and facilitates mapping of a Papaya ring-spot virus resistance locus. Plant J. *92*, 963–975. https://doi.org/10.1111/tpj.13722.

Wu, S., Wang, X., Reddy, U., Sun, H., Bao, K., Gao, L., Mao, L., Patel, T., Ortiz, C., et al. (2019). Genome of 'Charleston Gray', the principal American watermelon cultivar, and genetic characterization of 1,365 accessions in the US National Plant Germplasm System watermelon collection. Plant Biotechnol.J. *17*, 2246–2258. https://doi.org/10.1111/pbi.13136.

Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C.-Y., and Wei, L. (2011). Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. *39*, W316–W322. https://doi.org/10.1093/nar/gkr483.

Xie, D., Xu, Y., Wang, J., Liu, W., Zhou, Q., Luo, S., Huang, W., He, X., Li, Q., Peng, Q., et al. (2019). The wax gourd genomes offer insights into the genetic diversity and ancestral cucurbit karyotype. Nat.Commun. *10*, 5158. https://doi.org/10.1038/s41467-019-13185-3.

Yang, J., Deng, G., Lian, J., Garraway, J., Niu, Y., Hu, Z., Yu, J., and Zhang, M. (2020). The chromosome-scale genome of melon dissects genetic architecture of important agronomic traits. iScience *23*, 101422. https://doi.org/10.1016/j.isci.2020.101422.

Yano, R., Ariizumi, T., Nonaka, S., Kawazu, Y., Zhong, S., Mueller, L., Giovannoni, J.J., Rose, J.K.C., and Ezura, H. (2020). Comparative genomics of muskmelon reveals a potential role for retrotransposons in the modification of gene expression. Commun.Biol. *3*, 1–13. https://doi.org/10.1038/s42003-020-01172-0.

Zang, C., Schones, D.E., Zeng, C., Cui, K., Zhao, K., and Peng, W. (2009). A clustering approach for identification of enriched domains from histone modification ChIP-Seq data. Bioinformatics *25*, 1952–1958. https://doi.org/10.1093/bioinformatics/btp340.

van Zeist, W., and de Roller, G.J. (1993). Plant remains from Maadi, a predynastic site in lower Egypt. Veg. Hist. Archaeobot. *2*, 1–14. https://doi.org/10.2307/23416027.

Zhang, X., and Wessler, S.R. (2004). Genome-wide comparative analysis of the transposable elements in the related species Arabidopsis thaliana and Brassica oleracea. Proc. Natl. Acad. Sci. U S A *101*, 5589–5594. https://doi.org/10.1073/pnas.0401243101.

Zhang, H., Li, X., Yu, H., Zhang, Y., Li, M., Wang, H., Wang, D., Wang, H., Fu, Q., Liu, M., et al. (2019). Ahigh-quality melon genome assembly provides insights into genetic basis of fruit trait improvement. iScience *22*, 16–27. https://doi.org/10.1016/j.isci.2019.10.049.

Zhao, G., Lian, Q., Zhang, Z., Fu, Q., He, Y., Ma, S., Ruggieri, V., Monforte, A.J., Wang, P., Julca, I., et al. (2019). A comprehensive genome variation map of melon identifies multiple domestication events and loci influencing agronomic traits. Nat.Genet. *51*, 1607–1615. https://doi.org/10.1038/s41588-019-0522-8.

Zheng, Y., Jiao, C., Sun, H., Rosli, H.G., Pombo, M.A., Zhang, P., Banf, M., Dai, X., Martin, G.B., Giovannoni, J.J., et al. (2016). iTAK: aprogram for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. Mol. Plant *9*, 1667–1670. https://doi.org/10.1016/j.molp.2016.09.014.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
H3K27me1 (Rabbit)	Diagenode	Cat# C15410045-10
H3K4me3 (Rabbit)	Diagenode	Cat#C15310003
anti-5-methylcytosine antibody	Diagenode	Cat#NA8133D3
H3K27me3 (Mouse)	Diagenode	Cat# C15200181-50
H3K27me3 (Rabbit)	Diagenode	Cat# C15410069
H3K9me2 (Mouse)	Diagenode	Cat# C15200154
H3K9ac (Rabbit)	Diagenode	Cat# C15410004
Biological samples		
Cucumis melo L. ssp. *melo* var *cantalupensis*Charentais mono cultivar	This paper	Institute of Plant of Paris –Saclay (IPS2)
Chemicals, peptides, and recombinant proteins		
AMpure PB Beads	Pacific Biosciences	Cat#100-265-900
DpnII restriction enzyme	New Englands Biolabs	Cat#R0543S
SPRI magnetic beads	Beckman Coulter	Cat#B23317
Fixing solution (Formaldehyde 36.5%-38% in H2O)	Sigma-Aldrich	Cat#F8775
Lysis Buffer	Bionano Genomics	Cat#20255
RNAse A	Qiagen	Cat#19101
1X Wash Buffer	Bionano Genomics	Cat#20256
1X TE Buffer	ThermoFisher Scientific	Cat#12090015
Qubit dsDNA BR Assay	Invitrogen	Cat#12102
Nb.BssSI	New England Biolabs	Cat#R0681S
10X NEB 3.1	New England Biolabs	Cat#B7203S
50X repair Mix	Bionano Genomics	Cat#20250
NAD+	New England Biolabs	Cat#B9007S
10X Labeling Buffer	Bionano Genomics	Cat#20168
Taq polymerase	New England Biolabs	Cat#M0267S
Taq DNA Ligase	New England Biolabs	Cat#M0208S
10X ThermoPol Rxn Buffer	New England Biolabs	Cat#B9004S
DNA Stain YOYO	Bionano Genomics	Cat#80005
Polylysine slides	Sigma-Aldrich	Cat#25988-63-0
PBS	Beckman Coulter	Cat#6603369
Tween20	Sigma-Aldrich (Merck)	Cat#P9416
BSA	Vector Laboratories	Cat#H1200
DAPI	Merck	Cat#10236276001
MTSB buffer	This paper	NA
Vectashield	Vector Laboratories	Cat#H1200
Critical commercial assays		
bluePippin	Sage Science	Cat#B06010003
Femto Pulse	Agilent Technologies	Cat#M5330AA
Qubit fluorimeter	Thermofisher	Cat#Q33216

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NEBNext Ultra II DNA library preparation Kit	New Englands Biolabs	Cat#E7103
Qubit dsDNA HS reagent assay kit	Thermofisher	Cat#Q32851
Binding Kit 2.0	Pacific Biosciences	Cat#101-811-600
E.Z.N.A Plant DNA Kit	Promega	Cat#D3485-01
Ipure kit v2	Diagenode	Cat#C03010010
PacBio SMRT Cells v2.0	Pacific Biosciences	Cat#101-389-001
Deposited data		
Genome assembly	This paper	http://cmelogenome.gbfwebtools.fr
RNA-seq dataset of *C.melo*charmono plant organs	Latrasse et al. (2017)	PRJNA383830
ChIP-seq dataset of *C.melo*charmono plant organs	Latrasse et al. (2017)	PRJNA383830
MeDIP-seq dataset of *C.melo*charmono young leaf	Ruggieri et al. (2018)	https://doi.org/10.1038/s41598-018-26416-2
Hi-C raw data dataset of *C.melo*charmono young leaf	This paper	PRJNA768100
ATAC-seq raw dataset of *C.melo*charmono young leaf	This paper	PRJNA768100
Experimental models: Organisms/strains		
Cucumis melo Charmono (*Cucumis melo* L. ssp. *melo* var *cantalupensis*)	This paper	NA
Software and algorithms		
CANU 1.6	Koren et al. (2017)	https://github.com/marbl/canu
Quiver 49	Chin et al., (2013)	
Longranger 2.2.0	Reference Support-Software –Genome & Exome –Official 10x Genomics Support	https://github.com/10XGenomics/longranger
GATK 3.8-1	Poplin et al. (2017), Van der Auwera et al. (2013)	https://gatk.broadinstitute.org/hc/en-us
Pilon 1.22	Walker et al. (2014)	https://github.com/broadinstitute/pilon
Bwa-mem 0.7.17	Li and Durbin (2009)	https://github.com/lh3/bwa
FreeBayes 1.2.0	Garrison and Marth (2012)	https://github.com/freebayes/freebayes
Bcftools consensus 1.8	Bcftools by samtools s. d.	https://samtools.github.io/bcftools/bcftools.html
Supernova 2.0.0	Weisenfeld et al. (2017)	https://www.nuget.org/packages/SuperNova/2.0.0
Snakemake workflow	Mölder et al., 2021	https://snakemake.readthedocs.io/en/stable/tutorial/basics.html
Last 963	Kiełbasa et al. (2011)	https://gitlab.com/mcfrith/last
Bedtools 2.27.1	Quinlan and Hall (2010)	https://github.com/arq5x/bedtools2
LabelDensityCalculator software	Bionano Genomics	https://bionanogenomics.com/support/software-downloads/
Bionano genomics Irysview software	Bionano Genomics	https://bionanogenomics.com/support/software-downloads/
HiC-pro software	Servant et al., (2015)	https://github.com/nservant/HiC-Pro/blob/master/doc/COMPATIBILITY.md
HiCPlotter software	Akdemir and Chin (2015)	https://github.com/kcakdemir/HiCPlotter
EuGene pipeline release 4.2b	Sallet et al. (2014)	https://www.genesisenergy.com/operations/pipeline/offshore/tariffs-eugene/
tRNAscan-SE	Lowe and Chan (2016)	https://github.com/UCSC-LoweLab/tRNAscan-SE
RNAmmer	Lagesen et al. (2007)	http://www.cbs.dtu.dk/services/RNAmmer/
Red	Girgis (2015)	https://github.com/BioinformaticsToolsmith/Red

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LTRharvest	Ellinghaus et al. (2008)	https://github.com/HajkD/LTRpred
REPET package v2.5	NA	https://urgi.versailles.inra.fr/Tools/REPET/TEdenovo-tuto
TE*denovo* pipeline	Flutre et al. (2011)	https://urgi.versailles.inra.fr/Tools/REPET/TEdenovo-tuto
PASTEC	Hoede et al. (2014)	https://urgi.versailles.inra.fr/Tools/PASTEClassifier
TEs from RepBase database v20.05	Jurka et al. (2005)	https://www.girinst.org/repbase/
Pfam27.0	Finn et al. (2014)	ftp://ftp.ebi.ac.uk/pub/databases/Pfam/releases/Pfam27.0/Pfam-B.hmm.gz
InterProScan v5.31-70.0	Jones et al. (2014)	https://github.com/tabotaab/docker-interproscan
KAAS	Moriya et al. (2007)	http://www.genome.jp/kegg/kaas/
KOBAS 3.0	Xie et al., 2011	http://kobas.cbi.pku.edu.cn/download/
PlantTFcat	Dai et al., 2013	https://bio.tools/planttfcat
EggNog database of orthology relationship	Huerta-Cepas et al. (2018)	http://eggnog5.embl.de/#/app/home
Trimmomatic tool	Bolger et al. (2014)	http://www.usadellab.org/cms/?page=trimmomatic
MACS2	NA	https://github.com/macs3-project/MACS
SICER	Zang et al. (2009)	https://bio.tools/sicer
HOMER	Heinz et al., 2010	http://homer.ucsd.edu/homer/interactions/
HICexplorer pipeline	Wolff et al. (2018), 2020; Ramírez et al. (2018)	https://github.com/deeptools/HiCExplorer
STAR 2.7.3a	NA	https://github.com/alexdobin/STAR
featureCounts	NA	https://bioinformaticshome.com/tools/rna-seq/descriptions/FeatureCounts.html
Other		
Agilent 2100 bioanalyzer	Agilent Technologies	https://www.agilent.com/en/product/automated-electrophoresis/bioanalyzer-systems/bioanalyzer-instrument/2100-bioanalyzer-instrument-228250
Bionano Genomics Irys System	Bionano Genomics	Cat#IN-011-01
Confocal microscope	ZEISS	Cat#LM880
Diagenode Bioruptor 200 UCD-300	Diagenode	Cat# B01020001

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Abdelhafid Bendahmane (abdelhafid.bendahmane@inrae.fr).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Accession numbers are listed in the key resources table. This paper does not report original code.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Plant material and growth conditions

The monoecious Charmono melon cultivar (*Cucumis melo* L. ssp. *melo* var *cantalupensis*) was used for all the experiments described below. Charmono plants were grown in 10L pots in a growth chamber at 27°C/21°C day/night temperature with 16h/8h day/night supplemental light provided by sodium vapor and metal halide bulbs. For genomic DNA sequencing, young leaves were sampled and immediately frozen in liquid nitrogen. For nuclei preparation, fresh leaves were samples and processed as described below.

METHOD DETAILS

Genome sequencing and assembly

High molecular weight (HMW) DNA isolation and DNA sequencing. High molecular weight DNA was prepared from ~ 10 g of Charmono young leaves using the procedure described in Mayjonade et al. (2016).

Single-molecule real-time long reads sequencing was performed at Gentyane Sequencing Platform (INRAE Clermont-Ferrand, France) with a PacBio Sequel Sequencer (Pacific Biosciences, Menlo Park, CA, USA) according to the manufacturer's recommendations. The SMRTbell library was prepared using a SMRTbell Express Template prep kit, following the "procedure and checklist -preparing >30kb librairies using SMRTbell Express Template preparation kit" protocol. Genomic DNA (5 µg) was sheared with the 50 kb program using a Diagenode Megaruptor (Diagenode) generating DNA fragments of approximately 25 kb. Sheared genomic DNA was carried into the first enzymatic reaction to repair any damages that may be present on the DNA backbone. A blunt adapter ligation was conducted to generate the SMRT Bell template. The samples were size-selected using the BluePippin (Sage Science, Beverly, MA, USA) in order to recover all the material above 15 kb. The eluted DNA was then repaired again and purified with 0,45X AMPure PB Beads to obtain the final libraries around 20 kb. The SMRTBell libraries were quality inspected and quantified on a Femto Pulse (Agilent Technologies) and a Qubit fluorimeter with Qubit dsDNA HS reagent Assay kit (Life Technologies). A ready-to-sequence SMRTBell Polymerase Complex was created using a Binding Kit 2.0 (PacBio) and the primer V3, the diffusion loading protocol was used, according to the manufacturer's instructions. The PacBio Sequel instrument was programmed to load and sequenced the sample on PacBio SMRT cells v2.0 (Pacific Biosciences), acquiring one movie of 600 min per SMRTcell. PacBio RSII and 10X Genomics library construction and sequencing reads were generated using the service of INRAE GeT-PlaGe Genomic platform (INRAE Toulouse, France).

*In situ*Hi-C assay. Young melo leaves were used to perform the *in situ* Hi-C experiment. It was carried out in accordance with the protocol published in Liu et al. (2017) with some modifications related to the library building in which we use NEBNext Ultra II DNA library preparation Kit (New Englands Biolabs). The restriction enzyme use was DpnII (New England Biolabs). According to the recommendation, nine PCR cycles were performed during the library amplification and Hi-C libraries were purified using SPRI magnetic beads (Beckman Coulter) and eluted in a final volume of 20 µL of nuclease-free water. The quality control of the libraries was assessed using Agilent 2100 bioanalyzer (Agilent) and subjected to 2 x 75bp paired end high-throughput sequencing using the service of INRAE EPITRANS platform (Orsay, France).

Genome assembly, chromosome construction and quality estimation. The 1 733 644 PacBio raw subreads were trimmed, corrected and assembled with CANU 1.6 (Koren et al., 2017). After consensus generation, the total corrected reads were assembled in 236 contigs with a total assembled genome length of 366 812 256 bp. By mapping the contigs of this assembly on the optical, the 236 contigs were scaffolded into 43 sequences. Using the genetic map, the 43 scaffolds were ordered into the 12 pseudomolecules corresponding to the 12 melon chromosomes. The remaining 6 contigs that do not match any genetic marker were concatenated into a virtual chromosome called CMiso1.1chr00 (Tables S2 and S3). The genome was first polished by Quiver 49 (Chin et al., 2013). A second round of polishing was performed using 10X Genomics paired-end library (114x coverage) using longranger 2.2.0 (Reference Support -Software -Genome & Exome -Official 10x Genomics Support), GATK 3.8-1 (Poplin et al., 2017) and pilon 1.22 (Walker et al., 2014) softwares. To further polish the genome sequence, we identified variants using different paired-end library datasets: the same 10X Genomics dataset and RNA-seq whole flower dataset (Table S14). Paired-end reads were mapped using longranger 2.2.0, bwa-mem 0.7.17 (Li and Durbin, 2009) and STAR 2.7.0d (Dobin et al., 2013) respectively, and variants identified using FreeBayes 1.2.0 (Garrison and Marth, 2012) and GATK 3.8-1 (Poplin et al. (2017), Van der Auwera et al. (2013). A total of 26, 984 variants were used to edit the genome

sequence, skipping 34 positions as they conflicted with other variants, using bcftools consensus 1.8 (« Bcftools by samtools » s. d.). Finally, we closed four gaps filled by contigs, assembled using supernova 2.0.0 (Weisenfeld et al., 2017) on a subset of 200M 10X reads (57x coverage; N50 scaffold length 9.87 Mb; Long scaffolds (>= 10 kb) count 406; assembly size 334.71 Mb), with a snakemake workflow (Mölder et al., 2021) using last 963 (Kiełbasa et al., 2011) and bedtools 2.27.1 (Quinlan and Hall, 2010). The corrected and final genome assembly length summed up to 365 081 742 bp.

Bionano optical maps construction and hybrid assembly. HMW DNA was purified from one gram of fresh young leaves according to the Bionano Prep Plant tissue DNA Isolation Base Protocol (30068, Bionano Genomics) with the following specifications and modifications. Briefly, the leaves were fixed using fixing solution (Bionano Genomics) containing formaldehyde (Sigma-Aldrich) and then disrupt with rotor stator in homogenization buffer. Nuclei were washed several times and then embedded in agarose plugs. After overnight proteinase K digestion in the presence of Lysis Buffer (Bionano Genomics) and one hour treatment with RNAse A (Qiagen), plugs were washed four times in 1x Wash Buffer (Bionano Genomics) and five times in 1x TE Buffer (ThermoFisher Scientific). Then, plugs were melted two min at 70°C and solubilized with 2 µL of 0.5 U/µL AGARase enzyme (ThermoFisher Scientific) for 45 min at 43°C. A dialysis step was performed in 1x TE Buffer (ThermoFisher Scientific) for 45 minutes to purify DNA from any residues. The DNA samples were quantified by using the Qubit dsDNA BR Assay (Invitrogen). The presence of mega base size DNA was visualized by pulsed field gel electrophoresis (PFGE).

To assess the frequency of recognition sites for different nicking enzymes, the contig sequences of Charmono was analyzed with LabelDensityCalculator software (Bionano Genomics). The optimal labeling frequency was calculated for endonuclease Nb.BssSI (13.5 labels/100kb). Nicking, labeling, repair and staining of the HMW DNA were performed according to the Bionano Prep Labeling NLRS Protocol (30024, Bionano Genomics). Briefly, 300 ng of genomic DNA was nicked using 2 µL of 20 U/µL of Nb.BssSI (New England BioLabs) for two hours at 37°C in presence of 10X NEB 3.1 (New England BioLabs). The nicked DNA was labeled with the 10X Labeling Mix (Bionano Genomics) using 1 µL of 5 U/µL Taq polymerase (New England BioLabs) for one hour at 72°C in presence of 10X Labeling Buffer (Bionano Genomics). After labeling, the nicks were ligated with 1µL of 40 U/µL Taq DNA ligase (New England BioLabs) in the presence of 10x ThermoPol Rxn Buffer (New England BioLabs), 50X repair Mix (Bionano Genomics) and NAD+ (New England BioLabs) for 30 minutes at 37°C. The backbone of the labeled DNA was stained with the DNA Stain YOYO (Bionano Genomics). The NLRS DNA concentration was measured with the Qubit dsDNA HS Assay (Invitrogen).

Labeled and stained DNA was loaded on the Irys chip. Loading of the chip and running of the Bionano Genomics Irys System were all performed according to the Irys User Guide (30047, Bionano Genomics). Data processing was performed using the Bionano Genomics Irysview software. For the optical map *de novo* assembly, 150kb and 9 labels were used for the minimum length of molecules and the minimum number of label sites per molecule in the BNX sort step. The hybrid scaffolding was performed between the whole genome assembly and the optical genome maps with default parameters.

Pseudo-chromosomes validation using Hi-C. To evaluate the Charmono pseudo-chromosomal genome assembly, we constructed spatial proximity maps of the Charmono genome using chromosome conformation capture sequencing (Hi-C) at a resolution of 50 kb. To process the Hi-C data we used HiC-Pro software (Servant et al., 2015). For the analysis presented here, FASTQ paired end reads 75bp were aligned against the Charmono genome and filtered through the HiC-pro pipeline following these paramaters : –very-sensitive -L 30 –score-min L,-0.6,-0.2 –end-to-end –reorder. To remove duplicates reads, filter valid interactions and finally generate Hi-C interaction matrices (50kb, 100kb, 500kb and 1Mb), we set the parameters: MIN_FRAG_SIZE = 50; MAX_FRAG_SIZE = 100000; MIN_INSERT_SIZE = 100; and MAX_INSERT_SIZE = default. Moreover, all experimental artefacts, such as circularized reads and re-ligations, singletons have been filtered out using HiC-Pro. The genome was divided into equally sized bins and number of contacts observed between each pair of bins, was reported. Finally contact maps were plotted with HICPlotter software (Akdemir and Chin, 2015). The high collinearity between the genetic map based pseudomolecules anchoring and Hi-C based contact map information corroborated the overall assembly quality (Figure 1).

Charmono genome annotation

Gene model, ncRNA and TEs prediction

Transcriptome data used for the prediction of gene models. For the transcriptome-based gene prediction, RNA-seq data were retrieved from GEO data repository GSE98054 (Latrasse et al., 2017).

Gene and ncRNA models. Transcripts were predicted using a Snakemake pipeline (Koster and Rahmann, 2018). Briefly, RNAseq read quality was assessed with fastQC screen (Wingett and Andrews, 2018) and adapters and low quality bases were trimmed with Cutadapt (Martin, 2011). Filtered reads were then mapped to Charmono genome with STAR (Langmead and Salzberg, 2012) and transcripts predicted with stringtie (Pertea et al., 2015). We choose to predict transcripts on strand + and – separately and to merge prediction with stringtie merge.

Annotation of protein-coding genes and lncRNAs were predicted with a fully automated and parallelized pipeline, egn-ep, that carries out probabilistic sequence model training, genome masking, transcript and protein alignment computation and integrative gene modeling in EuGene software (release 4.2b) (Figure S2; Sallet et al., 2014). We use a 2 steps prediction, as we predicted genes model on each strand before merging annotation.

De novo transposable elements annotation. Prior to transposable elements (TEs) annotation in the Charmono assembled genome, we excluded gaps longer than 11 undefined bases (Ns) and unanchored sequences of CMiso1.1chr00. TEs were predicted using the REPET package v2.5 (https://urgi.versailles.inra.fr/Tools/REPET). TE*denovo* pipeline (Flutre et al., 2011) was used to align all-versus-all the contig sequences and identify clusters of related TE sequences. Only TE consensus characterized by at least 5 clustered sequences were retained. Then, the consensus sequences were classified into TE orders according to Wicker's classification (Wicker et al., 2007) using PASTEC (Hoede et al., 2014), based on their similarities to characterized TEs from RepBase database v20.05 (Jurka et al., 2005) and domains from Pfam27.0 (Finn et al., 2014). Consensus sequences identified as satellites (labeled SSR) and rDNA were discarded, as well as unclassified consensus sequences assembled with less than ten copies, providing an initial library of consensus sequences to annotate TE copies against the whole genome. TE copies annotation was carried out by TE*annot* pipeline (Quesneville et al., 2005) with default parameters using three iterations. At each iteration, copies of the consensus sequences were identified on the genome, starting with the initial consensus sequences library at iteration one, and kept consensus sequences showing at least one full-length fragment (i.e., fragments covering more than 95% of the consensus sequence) as input for the next iteration. A manual curation step was carried out to refine TE annotation. The curated library of consensus sequences was used to run a fourth TE*annot* iteration to obtain the final TE annotation.

Functional gene annotation. Protein-coding genes were annotated through integration of different databases (https://lipm-gitlab.toulouse.inra.fr/LIPM-BIOINFO/nextflow-functionnalannotation– commit 17f5efb9dcf7c67f74769cf6a1930cf9ca986e47). Predicted genes were functionally annotated by performing a BLASTP search against the UniProtKB Viridiplantae database and the NCBI non-redundant protein database with an e-value threshold of 1e-10. In addition, a comprehensive annotation was also achieved using InterProScan (v5.31–70.0) (Jones et al., 2014), which includes motifs/domains prediction, functional classifications, protein family identification, transmembrane topology, predicted signal peptides and GO annotations. KAAS (Moriya et al., 2007) and KOBAS 3.0 (Xie et al., 2011) were used to search the KEGG GENES database for KO (KEGG Orthology) assignments and generating a KEGG pathway membership. PlantTFcat (Dai et al., 2013) was also used to systematically analyze InterProScan domain and categorize possible chromatin regulators (CRs), transcription factors (TFs) and other transcriptional regulators (TRs) in the current assembly. In parallel, to complete the previous annotation, we perform a second run of BLASTP search against a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses, called EggNog database of orthology relationships (Huerta-Cepas et al., 2018). Finally, all GO, KO and KEGG functional custom annotation were merged together.

Melon epigenome

ChIP-seq assay. The ChIP-seq expression data were retrieved from GEO data repository GSE98054 (Latrasse et al., 2017).

Methylated-DNA Immunoprecipitation, MeDIP-seq sequencing

Genomic DNA was isolated from roots and leaves of the Charmono melon cultivar (*Cucumis melo* L. ssp. *melo* var *cantalupensis*) using E.Z.N.A Plant DNA Kit (Omega). Fragmentation was performed using Diagenode Bioruptor 200 UCD-300. MeDIP-Seq libraries construction was performed as described in Ruggieri et al. (2018). Briefly, methylated DNA was precipitated with anti-5-methylcytosine antibody (NA8133D3, Merck Millipore, Diagenode), then purified using Auto Ipure kit v2 (Diagenode). Libraries were synthetized using NebNext Ultra DNA Library Preparation Kit (NEB) according to the manufacturer's instructions and sequenced by Illumina Sequencing technology using the service of INRAE EPITRANS platform (Orsay, France).

ATAC-seq

ATACseq assay was performed on ~ 10 g of Charmono young leaves using the procedure described in the original method (Buenrostro et al., 2015). To prepare and purify ATACseq libraries, we used NEBNext Ultra II DNA library preparation Kit (New Englands Biolabs) and AMPure beads (E6220, BioLabs), respectively. The quality of the libraries was assessed with Agilent 2100 Bioanalyzer (Agilent, LabChip Caliper). ATACseq libraries were subjected to high-throughput sequencing by Illumina Sequencing technology using the service of INRAE EPITRANS platform (Orsay, France).

Computational analysis of the epigenomic data

Computational analysis of raw data from ChIPseq, meDip and ATAC-seq. The ChIPseq and meDip libraries were sequenced by illumina sequencing with reads of 75Pb in "single-end" and 50 million, 40 million reads were obtained from this sequencing. For each type of experiment, Nextera adapters were trimmed and filtered using the Trimmomatic tool (Bolger et al., 2014), following the same parameters as for the ATACseq pipeline. Alignment to the PacBio genome was then performed with a maximum incompatibility of 1 Pb and only a single read alignment was retained. 20 million, 34 million reads were obtained for the ChIPseq and meDip experiments respectively. To determine the target regions of H3K9ac and DNA methylation, the model-based analysis of ChIP-seq (MACS2, http://liulabdfci.hardvard.edu/MACS/) was used following the parameters by default. Simultaneously, the detection of the H3K27me3 peaks was carried out with the SICER tool (Zang et al., 2009) according to these parameters: - rt 1, -w 200, -fs 150, -f 0.95, -cuoff 0.1.

Computational analysis of Hi-C raw data. For the genome compartmentalization analysis, we generated Hi-C interaction matrices at different resolutions (5kb,10kb,25kb,50kb,100kb,500kb and 1Mb) using HiC-Pro pipeline as described in Servant et al. (2015).

For sub-nuclear compartment analysis, we used HOMER (http://homer.ucsd.edu/homer/interactions/) and HiCExplorer pipeline (Wolff et al., 2018, 2020). To compute global compartmentalization signal (genome-wide) we used proposed method in HiCexplorer introduced by Schwarzer et al. (2017). We compute the global strength for compartmentalization as (AA + BB) / (AB + BA) after reranging the bins obs/exp based on their corresponding principle composant 1 values. These values are reordered incrementally to then follow and then the same order of bins is reused to rearrange the bins in obs/exp matrix. In order to reassign the correct sign of the eigenvector (-; B or +; A) of the compartments, the global DNA methylation signal from meDIP-seq data was used. Any compartment positively correlated with this signal was annotated as compartment B.

Sub-domain analysis was carried out using HiCExplorer pipeline on normalized interactions matrix of 10 kb, more precisely using the hicFindTADs functions with the following parameters: –correct For Multiple Testing 'fdr' –minDepth'30000' (x3 of the bins size of the interaction matrix) –maxDepth100000 (x10 of the bins size of the interaction matrix) –step '10000' (the number of bins resolution of the interactin matrix) –thresholdComparisons'0.05' –delta '0.01'. For the further downstream analysis we select the ones associated a maximum correlation with epigenetic marks. To group the domains, we do hierarchical clustering based on their H3K9ac (H3K9ac peaks coverage), H3K27me3 (H3K27me3 peaks coverage) and global DNA methylation levels.

H3K9Ac and H3K27me3 immunofluorescence assay on Charmono nuclei

Isolation of nuclei. The plant tissues were fixed using 1% formaldehyde solution during 15 min under vacuum. Then 2.5 ml of glycine was added and tissues placed again for 5 min under vacuum to stop fixation.

Fixed tissues were finely chopped with a razor blade in 500 µL of MTSB buffer (50 mM PIPES, 5 mM EGTA, 5 mM MgSO4, pH 7.2 / 7.4 adjusted with KOH) + 0.1% Triton. Then, the chopped samples were filtered on a 50µm filter. After filtration on 50µm filters, the nuclei were centrifuged for 10 min at 1500g at 4°C and the supernatant was discarded. The nuclei pellet was then resuspended in 500 µl of MTSB + 0.1% Triton and centrifuged again for 10 min at 1500 g at 4°C. Finally, the pellet was resuspended in 100 µL of MTSB.

Immunofluorescence assay isolated melon nuclei. The immunofluorescence assay was carried out on polylysine slides (Sigma-Aldrich). First, samples was deposited on the hydrophobic slide area and dried at 37°C. Then, a first brief wash with PBS was applied to the slides, followed by two sample washes of 10 min with PBS +0.1% tween20. The slide blocking and saturation was performed for 20 min in PBS + 0.1% Tween20 + 3% BSA. Next, 70 µl of primary antibodies previously diluted at 1: 400 in PBS were applied to the sample and incubated overnight at 4°C in a humid room. The slides were then washed twice for 10 min in PBS + 0.1% Tween20. Then, 70 µL of secondary antibodies diluted at 1: 400 in PBS were added and the slides were incubated for 1 h at 37°C. After incubation, the slides were washed 10 x 5 min in PBS + 0.1% Tween20 and once 5 min with PBS only. Finally, a drop of Vectashield + DAPI (Vector Laboratories, H1200) was deposited and the slide was covered with a coverslip. The slides were finally observed under a confocal microscope (LM880, ZEISS). Each combination of primary antibodies with secondary antibodies, according to the pairs of marks of interest chosen, is shown in Table S15.

Gene expression analysis

The transcriptome RNA-seq expression data were retrieved from GEO data repository GSE98054 (Latrasse et al., 2017). RNA-seq raw reads were trimmed for adaptator and quality filtered (mapq 10) using Trimmomatic software (Bolger et al., 2014). Filtered mapped reads were mapped against our Charmono genome version using STAR 2.7.3a software following the parameter setup as below:

–outFilterTypeBySJout \ –outSAMunmapped Within \ –outSAMattributes All \ –outFilterMultimapNmax 20 \ –alignSJoverhangMin 8 \ –alignSJDBoverhangMin 1 \ –outFilterMismatchNmax 999 \ –outFilterMismatchNoverReadLmax 0.04 \ –alignIntronMin 20 \ –alignIntronMax 20000 \ –alignMatesGapMax 20000 \ –twopassMode Basic \–quantMode'TranscriptomeSAM GeneCounts'.

To read counts table was generated using featureCounts tools with the following paramaters :"-p -C -Q 10 –primary -t exon -g gene_id -B".

Genome wide comparison of melon genomes

Whole genome alignments between the genome assemblies of Charmono and DHL92, Payzawat, Harukei-3 and HS were performed with MUMMER (version 4.0) (Marçais et al., 2018), using the nucmer algorithm with the –mum parameter. Dot-plots were generated with mummerplot. For SNP comparison, we used nucmer module of NuCdiff package (Khelik et al., 2017).

Sequence polymorphisms in Organ Specifically Expressed Genes

To identify OSEG, we analyzed RNAseq data from root, leaf, flower and fruit, using CuttDiff software from CuffLinks tools suite (Trapnell et al., 2010) and the following parameters : p-value 0.05, statistical correction: Benjamini HoChberg; FDR: 0.05. OSEG were selected using log2(FC)>1 or <-1. Polymorphisms in OSEG were called using NuCdiff package and Charmono genome as a reference. SNPs and InDels were classified as triple or no-triple. Sequence diversity in OSEG was defined as the density of InDels and SNPs per Kb normalized by the size of the corresponding gene. All p-values were calculated using Wilcoxon test and adjusted using Benjamini HoChberg method.

Evolutionary analysis of *Cucurbitaceae* genomes

The ancestral *Cucurbitaceae* karyotype (ACK), a 'median' or 'intermediate' genome consisting of a clean reference gene order common to the extant species investigated, and derive evolutionary scenario were obtained following the method described in Salse (2016) based on the orthologous and paralogous relationships identified between *Cucurbita moschata* (20 chromosomes, 32205 genes; Sun et al., 2017), *Cucurbita maxima* (20 chromosomes, 32076 genes; Sun et al., 2017) *Citrullus lanatus* (11 chromosomes, 23440 genes; Guo et al., 2013, 2019; Wu et al., 2019), *Lagenaria siceraria* (11 chromosomes, 22472 genes; Wu et al., 2017), *Cucumis melo* (current article), *Cucumis sativus* (7 chromosomes, 23248 genes;

Huang et al., 2009) and the ancestral eudicot karyotype (7 protochromosomes, 9022 genes; Murat et al., 2017). Briefly, the first step consists of aligning the investigated genomes to define conserved/duplicated gene pairs on the basis of alignment parameters (CIP for Cumulative Identity Percentage and CALP Cumulative Alignment Length Percentage). The second step consists of clustering or chaining groups of conserved and duplicated genes into ancestral protochromosomes (also referred to as CARs for Contiguous Ancestral Regions) corresponding to independent sets of blocks sharing paralogous and/or orthologous relationships in modern species. In the third step, conserved gene pairs (or conserved groups of gene-to-gene adjacencies) between the investigated species are then considered as potentially ancestral in the same order and orientation. From the reconstructed ancestral karyotype (ACK) an evolutionary scenario can then be inferred taking into account the fewest number of genomic rearrangements (including inversions, deletions, fusions, fissions, translocations) which may have operated between ACK and the modern genomes.

QUANTIFICATION AND STATISTICAL ANALYSIS

Details of the statistical tests applied, including the statistical methods, number of replicates, mean and error bar details and significances are indicated in the relevant figure legends.