BMC Plant Biology

RESEARCH

Open Access

Uncovering the genome evolutionary dynamics of the Mediterranean endemic palm *Chamaerops humilis*

Mónica Labella-Ortega[1], Maria Tartaglia[1], Daniela Zuzolo[1], Antonello Prigioniero[1*], Maria Maisto[1], Maria Antonietta Ranauda[1], Emanuele Fosso[1], Hengchi Chen[2,3], Rosario Schicchi[4], Giuseppe Bazan[5], Guido Cipriani[6], Douglas E. Soltis[7,8,9,10], Pamela S. Soltis[7,9,10] and Carmine Guarino[1]

Abstract

Background *Chamaerops humilis* L. is the only endemic palm species of the western Mediterranean. Here, we present a *de novo* genome assembly with structural and functional annotation that reveals major structural changes shaping the evolutionary history of this species.

Results The genome, estimated at 3.44 Gbp, comprises 41,738 genes and 2.87 Gbp of repetitive elements. Evolutionary analyses identified a whole-genome duplication event ~ 48.0 Mya, shared with other palms, followed by divergence from its sister genus *Phoenix* ~ 16.6 Mya. Functional enrichment of rapidly evolving genes highlighted associations with genome plasticity and stress response pathways. Analyses of gene duplication types and K_S distributions uncovered recent lineage-specific duplication waves and *C. humilis*-exclusive duplicated genes, providing novel insights into palm genome evolution and the adaptive potential of this Mediterranean endemic.

Conclusions This study provides a high-quality reference genome for *C. humilis* and insights into the possible genomic basis for Mediterranean adaptation. Future work combining comparative genomics, pangenomics, and functional studies will be key to fully understanding its evolution.

Keywords Chamaerops humilis, Genome duplication, Evolution, Transposon duplication, Genome sequencing

*Correspondence:
Antonello Prigioniero
prigioniero@unisannio.it
[1]Department of Science and Technology, University of Sannio, Via de Sanctis 6, Benevento 82100, Italy
[2]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent 9052, Belgium
[3]VIB Center for Plant Systems Biology, VIB, Ghent 9052, Belgium
[4]Agricultural, Food and Forest Department, University of Palermo, Viale delle Scienze, Ed. 4, Palermo 90128, Italy
[5]Department of Biological, Chemical and Pharmaceutical Sciences and Technologies (STEBICEF), University of Palermo, Viale delle Scienze, ed. 17, Palermo 90128, Italy
[6]Department of Agriculture, Food, Environmental and Animal Sciences, University of Udine, Via delle Scienze 206, Udine 33100, Italy
[7]Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[8]Department of Biology, University of Florida, Gainesville, FL 32611, USA
[9]Biodiversity Institute, University of Florida, Gainesville, FL 32611, USA
[10]Genetics Institute, University of Florida, Gainesville, FL 32608, USA

Background

The current global biodiversity scenario clearly indicates that the sixth mass extinction is underway [1]. This awareness, together with rapid technological advances, is strongly driving research efforts worldwide. Whole-genome sequencing (WGS) is increasingly used to gain insights into the molecular basis of evolution and adaptation of living species to environmental change [2]. Reference genomes and their integration into comprehensive databases are regarded as the most effective tools for achieving this goal. They enable biodiversity assessment, the development of conservation strategies, and ecosystem restoration [3].

Across the Tree of Life, plant genomes remain underrepresented, largely due to their higher complexity in terms of sequencing and functional analysis than animal or fungal genomes [2, 4]. Owing to their structural diversity, rapid adaptation, and large size, plant genomes are often more diverse than those of other eukaryotes [5]. Modern plant genomes are shaped by polyploidy, transposable element activity, gene duplication and loss, and chromosomal rearrangements [6]. Indeed, all angiosperms retain signatures of ancient and more recent whole-genome duplication (WGD) events [7–9]. The retention of paralogs following both WGD and small-scale duplication is considered a key driver of plant diversification. However, the underlying selective mechanisms remain poorly understood [10]. Genomic variation and plasticity are thus central to enabling populations to evolve traits suited to their environments, thereby ensuring survival and reproductive success under changing conditions [11].

The Mediterranean Basin is recognized as one of the world's major biodiversity hotspots, characterized by exceptional intra- and interspecific diversity adapted to heterogeneous environments [12]. Its geological history, particularly the Miocene, profoundly shaped present-day biota [13]. The Messinian Salinity Crisis and subsequent Zanclean Flood [14, 15] triggered dramatic environmental fluctuations, which reshaped Mediterranean vegetation from tropical-subtropical forests to sclerophyllous shrublands and grasslands [16, 17]. During this transition, several lineages disappeared, as confirmed by pollen and macroremain records [18]. Palms exemplify this ecological gap: once widespread across the pre-Miocene Mediterranean, today they are reduced to a single endemic representative, *Chamaerops humilis* L [19].

Chamaerops humilis, the only species of its genus, is a cold-hardy, multi-stemmed, dioecious palm narrowly endemic to the western Mediterranean. It is a key indicator of thermo-Mediterranean vegetation, often dominating early successional stages due to its adaptability. Furthermore, it has long been intertwined with Mediterranean civilizations, providing food, fibers, and ethnobotanical products, in addition to its ornamental use [20].

Here, we report the genome of *C. humilis*, providing a high-quality reference assembly that sheds light on its evolutionary trajectory within palms. Through comparative and phylogenomic analyses, including whole-genome duplications and gene duplication dynamics, we identify the major genomic changes that shaped its history. Beyond reconstructing evolutionary history, this genomic resource provides a foundation for future studies on adaptation and resilience in *C. humilis* and related lineages. It also has important implications for biodiversity conservation in the Mediterranean hotspot.

Results

Genome assembly

We generated a *de novo* assembly of the *C. humilis* genome by integrating three sequencing and assembly technologies, including PacBio long reads (186.37 Gbp in Male with a mean read length of 27,939 bp, and 85.73 Gbp in Female with a mean read length of 13,634 bp) and Illumina short reads (273.81 Gbp in Male and 274.20 Gbp in Female), resulting in a single haploid-representative consensus genome assembly. *K-mer* analysis initially estimated a haploid genome size of ~2.2 Gbp, with low heterozygosity (2.3%) and a high repeat content (>70%). Using PacBio long reads, 10,638 contigs were assembled, yielding a contig N50 of 1.02 Mb, a total genome size of 3.49 Gbp, and 58.1× read depth. BUSCO analysis of the primary contig assembly indicated 84% completeness, with 9.2% duplicated, 3.1% fragmented, and 12.9% missing single-copy genes (Additional file 2: Table S1).

To improve accuracy, the assembly was polished with Illumina short reads (38.9× read depth), correcting ~20.5 Mbp (≈1% of the genome). Unmapped short reads were then assembled *de novo*, producing 7,553 contigs with a total length of 30.3 Mbp (N50 = 4.4 kb; mean length = 4.0 kb). These contigs were incorporated into the assembly and used as additional sequence evidence during the scaffolding process. BUSCO completeness increased to 91%, and a reduction in missing genes was observed (2.4%) (Additional file 2: Table S1). Redundancy assessment, based on mapping PacBio reads back to the raw assembly, revealed a single main peak, indicating a low fraction of redundant or secondary contigs (Additional file 1: Fig. S1).

For chromosome-scale scaffolding, we used Hi-C data (75.48Gbp in male, with 18.6× read depth), which merged 13,313 contigs into 250 scaffolds. A total of 77.12% of the contigs incorporated into chromosome-scale scaffolds originated from the long-read assembly, while a smaller fraction (22.87%) corresponded to short-read SPAdes contigs that could be confidently integrated based on scaffolding evidence. The remaining 4,514 unscaffolded

contigs (0.5% of the genome) were combined into an artificial scaffold. In total, 251 scaffolds were produced and aligned to the reference genome available in NCBI (GCA_042465325.1). Of these, 240 scaffolds were anchored to 18 chromosomes (Fig. 1), while 11 residual scaffolds were placed in an artificial chromosome ("chromosome 0") with 100-bp gaps inserted between consecutive contigs (Fig. 1, Additional file 1: Fig. S2).

The final assembly spanned 3.44 Gbp, with chromosome 2 being the largest (345.9 Mbp) and chromosome 18 the smallest (65.5 Mbp) (Additional file 2: Table S1).

De novo transcriptome assembly

De novo transcriptome assembly with Trinity produced 99,890 transcripts. BUSCO analysis indicated a completeness of 96.0%, with 3.8% fragmented and 0.2% missing

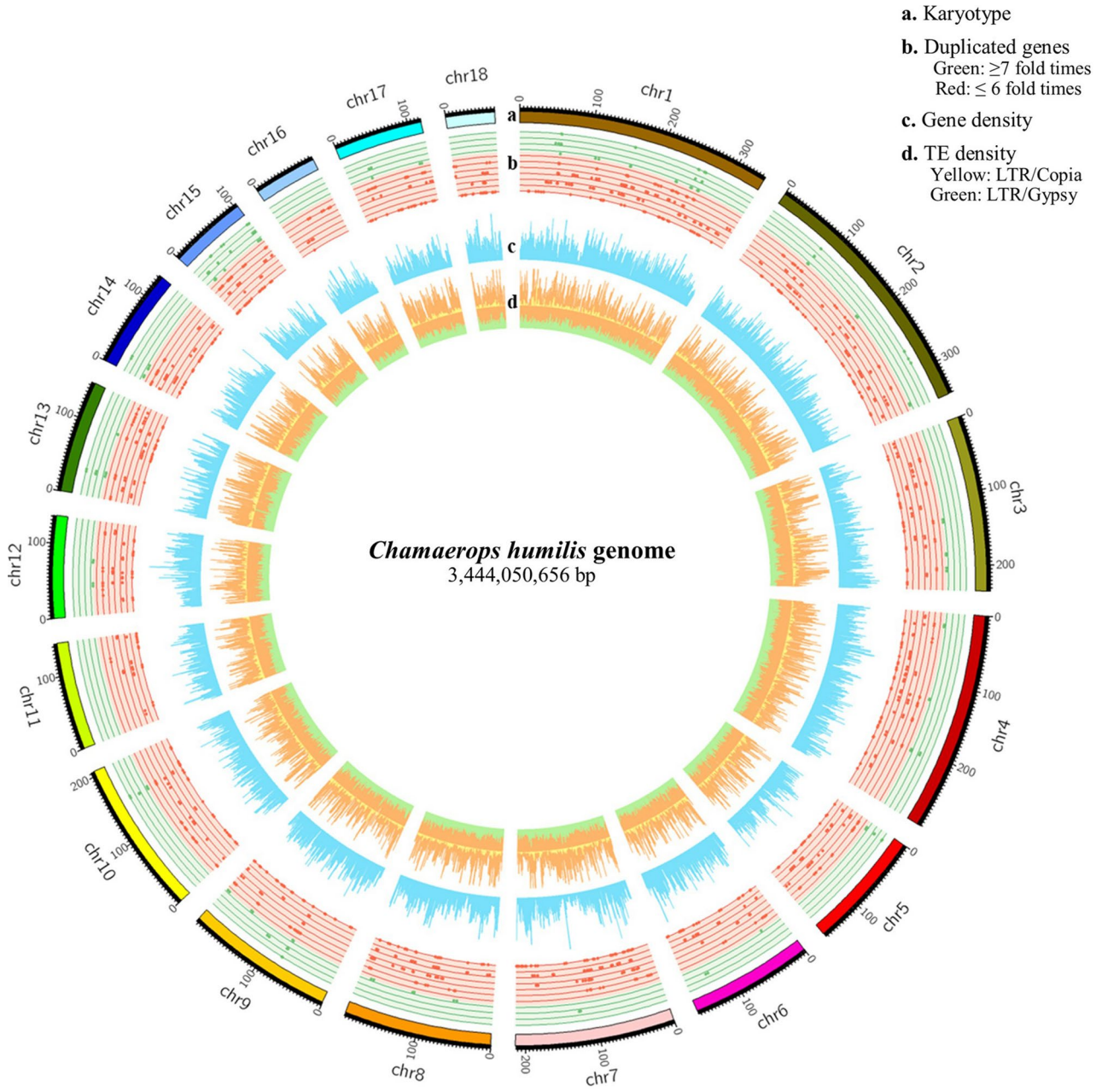

Fig. 1 Chromosomal features, functional annotation, and gene duplication landscape of *Chamaerops humilis*. Key chromosomal characteristics and duplication patterns in *C. humilis* in the 18 assembled chromosomes. From the outer to the inner rings, the tracks represent: (**a**) chromosome size in megabases (Mb); (**b**) positions of duplicated genes, in green genes ≥ 7 fold times duplicated and in red genes ≤ 6 fold times duplicated indicates genes duplicated; (**c**) gene density (annotated genes per 10-Mb); and (**d**) in orange transposable element density per 10-Mb, in green LTR/Gypsy and yellow LTR/Copia density

genes (Additional file 2: Table S1). Coding regions were predicted with TransDecoder, identifying 39,812 complete open reading frames, in addition to partial coding sequences. Functional annotation showed that 68.75% of predicted proteins had homologs in public databases or contained conserved domains. This transcriptome was subsequently used to support genome annotation.

Genome annotation

The final gene set contains 41,738 protein-coding genes, representing 9.85% of the *C. humilis* genome (Fig. 1). Of these genes, 91.47% were located on chromosomes, with chromosome 2 harbouring the highest number of genes, whereas chromosome 18 contained the fewest (Additional file 2: Table S2).

Artificial chromosome 0 contained 8.52% of the annotated genes and comprised 76.31% transposable element bases. This chromosome mainly represents genomic regions for which gene models were supported by transcriptomic evidence and annotation, but which could not be confidently anchored or assigned to specific chromosomes in the assembled genome.

Gene Ontology (GO) analysis revealed that the most represented biological processes were DNA integration, protein phosphorylation, and proteolysis. Among molecular functions, nucleic acid binding, ATP binding, and zinc ion binding were predominant, while the most enriched cellular components were the membrane, nucleus, and cytoplasm (Additional file 1: Fig. S3).

Transposable elements (TEs) accounted for the majority of the genome and were classified into two major classes [21]. Retrotransposons (Class I) comprised ~ 78% of the genome, whereas DNA transposons (Class II) represented 5.3% (Additional file 2: Table S3). The most abundant superfamilies were LTR/Copia and LTR/Gypsy, occupying 36.9% and 29.8% of the genome, respectively. The LTR Assembly Index (LAI) was 19.86, demonstrating high assembly quality across repetitive regions and surpassing values reported for other palm genomes (Additional file 2: Table S4). TE content varied among palms, with *C. humilis* (2.20 Gbp) and *Cocos nucifera* (2.19 Gbp) exhibiting the highest levels, followed by *Elaeis guineensis* (1.80 Gbp), whereas *Phoenix dactylifera* showed the lowest TE content (0.40 Gbp) (Additional file 2: Table S4).

Insertion time analysis of full-length LTR retrotransposons revealed a burst of LTR/Copia insertions around 5.5 Mya, followed by a gradual decline and a minor increase at ~ 2 Mya. For LTR/Gypsy, insertion rates steadily increased between 9 and 4.5 Mya before subsequently declining (Additional file 1: Fig. S4). Classification of full-length LTR retrotransposons into lineages showed that within LTR/Copia, the most abundant groups were Tork (21.8%), Angela (13.8%), and SIRE (11.7%), whereas within LTR/Gypsy, CRM (16.6%), Tat (13.9%), and Athila (6.9%) were the most represented (Additional file 2: Table S5).

Chloroplast genome analysis

The final estimated chloroplast genome size of *C. humilis* is 158.6 Kbp, with a GC content of 37.2%. We annotated 135 genes, of which 90 were protein-coding genes, 38 were tRNA, and 8 were rRNA (Additional file 1: Fig. S5). A phylogenetic tree across all the plastid genomes available was constructed (Additional file 1: Fig. S6, Additional file 2: Table S6). In the phylogenetic reconstruction, the subfamilies of Arecaceae were well resolved, with *C. humilis* placed in subfamily Coryphoideae and forming a clade that includes diverse species, the majority from the genus *Trachycarpus*, with the genus *Brahea* as the sister clade. Regarding chloroplast genome size, *C. humilis* has a genome size similar to *Adonidia merrillii* (158.692 Kbp [22]) and larger than that of *P. dactylifera* (158.462 Kbp [23]).

Gene family evolution

Single-copy orthologs (SOGs) from the *embryophyta_odb10* BUSCO dataset were identified across the species analyzed. The final species tree was inferred from 241 high-confidence single-copy orthologous genes. The inferred phylogenetic tree recovered two major clades—monocotyledons and eudicotyledons—that diverged ~ 130 Mya during the Early Cretaceous (Fig. 2a). Within monocots, the Arecaceae were estimated to have diverged ~ 110 Mya, and *C. humilis* branched off from its closest relatives in the genus *Phoenix* ~ 16.6 Mya (Fig. 2a).

Collinearity analysis among the Arecaceae species examined revealed a 1:1 gene ratio, suggesting that they have undergone the same number of WGD events (Additional file 1: Fig. S7). Synteny analysis between *C. humilis* and *P. dactylifera* and *E. guineensis* identified only small conserved syntenic blocks (Fig. 2b).

A targeted phylogenetic analysis of the Arecaceae, with Zingiberales as the outgroup, incorporated genome size dynamics together with gene family expansion and contraction (Additional file 1: Fig. S6). The ancestral genome size of *C. humilis* and *Phoenix* ancestor was estimated to have had a genome size of 1.81 Gbp. Our analysis indicates that the *C. humilis* lineage subsequently underwent genome expansion, whereas the *Phoenix* clade experienced a reduction over evolutionary time (Additional file 1: Fig. S8).

Across nine Arecaceae species and three Zingiberales, OrthoFinder identified 32,615 orthogroups, of which 9,085 were shared across all species. In *C. humilis*, 1,139 gene families were shared with other palms, while 1,057 families of genes were unique (Fig. 2c).

Focusing on rapidly evolving gene families, we detected 1,521 families that expanded or contracted in *C. humilis*

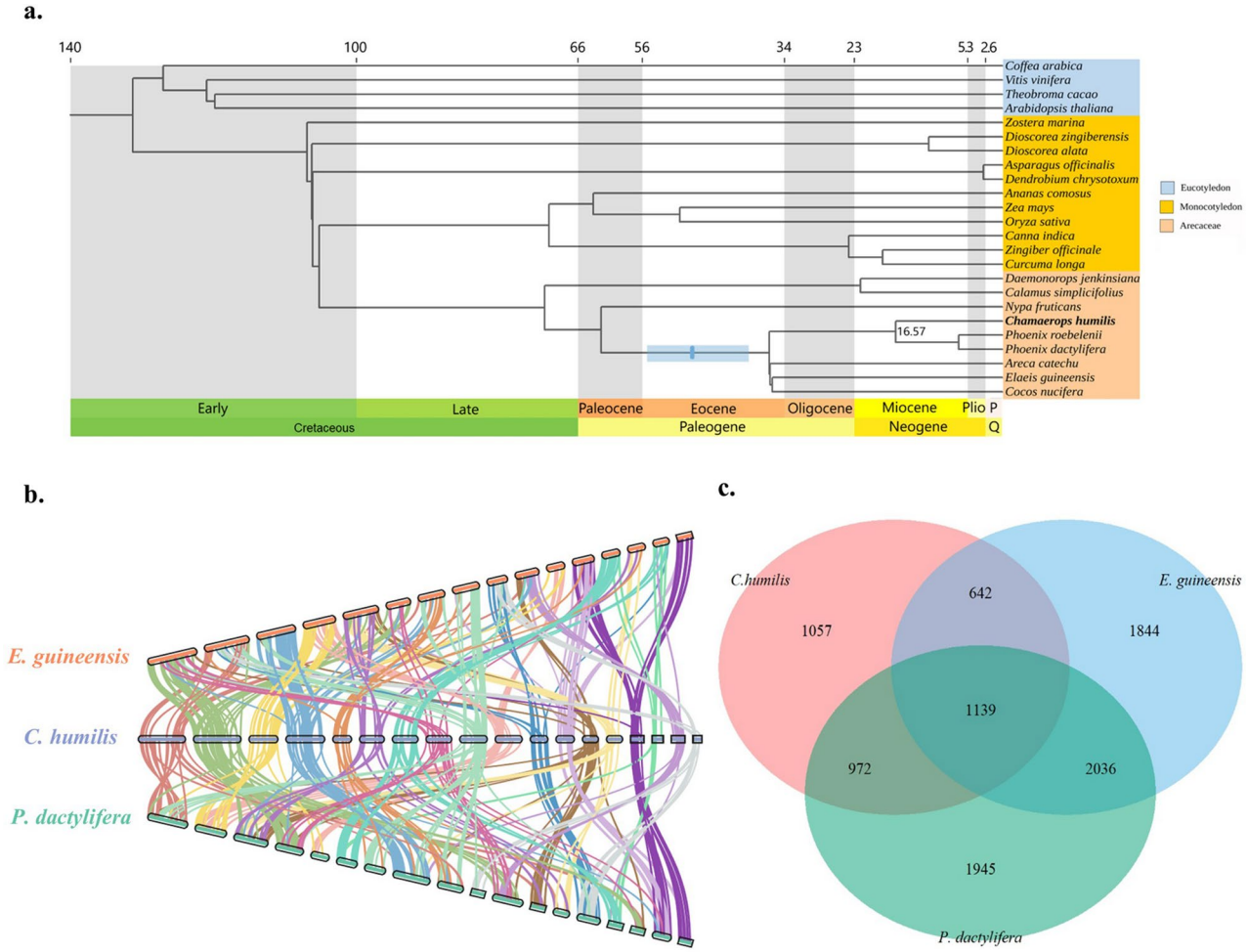

Fig. 2 Genome evolution of *C. humilis*. (**a**) Phylogenetic tree of angiosperm species showing their divergence times in million years. The blue box indicated the WGD event observed across Arecaceae species (39.87–55.2 mya), highlighting the event detected in *C. humilis* (48.02 mya) (Additional file 1: Fig. S7). Abbreviations in the figure correspond to geological epochs: Plio (Pliocene), P (Pleistocene), and Q (Quaternary). (**b**) Synteny blocks identified among the chromosomes of *C. humilis*, *E. guineensis*, and *P. dactylifera*. (**c**) Venn diagram illustrating the shared gene families changed among three palm species: *C. humilis*, *E. guineensis*, and *P. dactylifera*

since its divergence from Phoenix genus, comprising 3,123 gene gains and 2,325 losses (Additional file 1: Fig. S8). Gene Ontology enrichment analysis revealed that these families were predominantly associated with DNA stability processes (including mismatch repair, interstrand cross-link repair, DNA repair and regulation of cell cycle process), followed by stress response categories such as cellular response to water deprivation, response to abiotic stimulus and immune response (Additional file 2: Table S7).

Whole-genome duplication

Analysis of paralogous gene families in the *C. humilis* genome provided evidence for historical WGD events. WGD was assessed by calculating the genome-wide distribution of synonymous substitution rates (K_S) and non-synonymous substitution rates (K_A) for anchor gene pairs (Additional file 1: Fig. S9). Retained anchor pairs were detected within the 0–1 K_S range, with the 95% confidence interval of the log-normal distribution indicating a peak between 0.16 and 0.60 (Additional file 1: Fig. S10A). Phylogenetic dating analysis placed this WGD event at ~48.0 Mya (Additional file 1: Fig. S10B).

To further investigate genome expansion, gene duplication modes in *C. humilis* were compared with those in *P. dactylifera* and *E. guineensis*. In *C. humilis*, only 20.62% (7,520 genes) of the gene set were WGD-derived, compared with 50.6% in *P. dactylifera* and 52.0% in *E. guineensis* (Additional file 1: Fig. S11A). The proportions of tandem duplications (TD, 3.61%) were lower than in the other palms, while transposed derived duplications (TRD, 32.94%) and proximal duplications (PD, 4.56%) were slightly higher than those of other palms; *C. humilis* contained approximately twice as many genes in the case of TRD genes. Notably, DD accounted for 38.24% of all duplicated genes in *C. humilis*, compared with only 8.91%

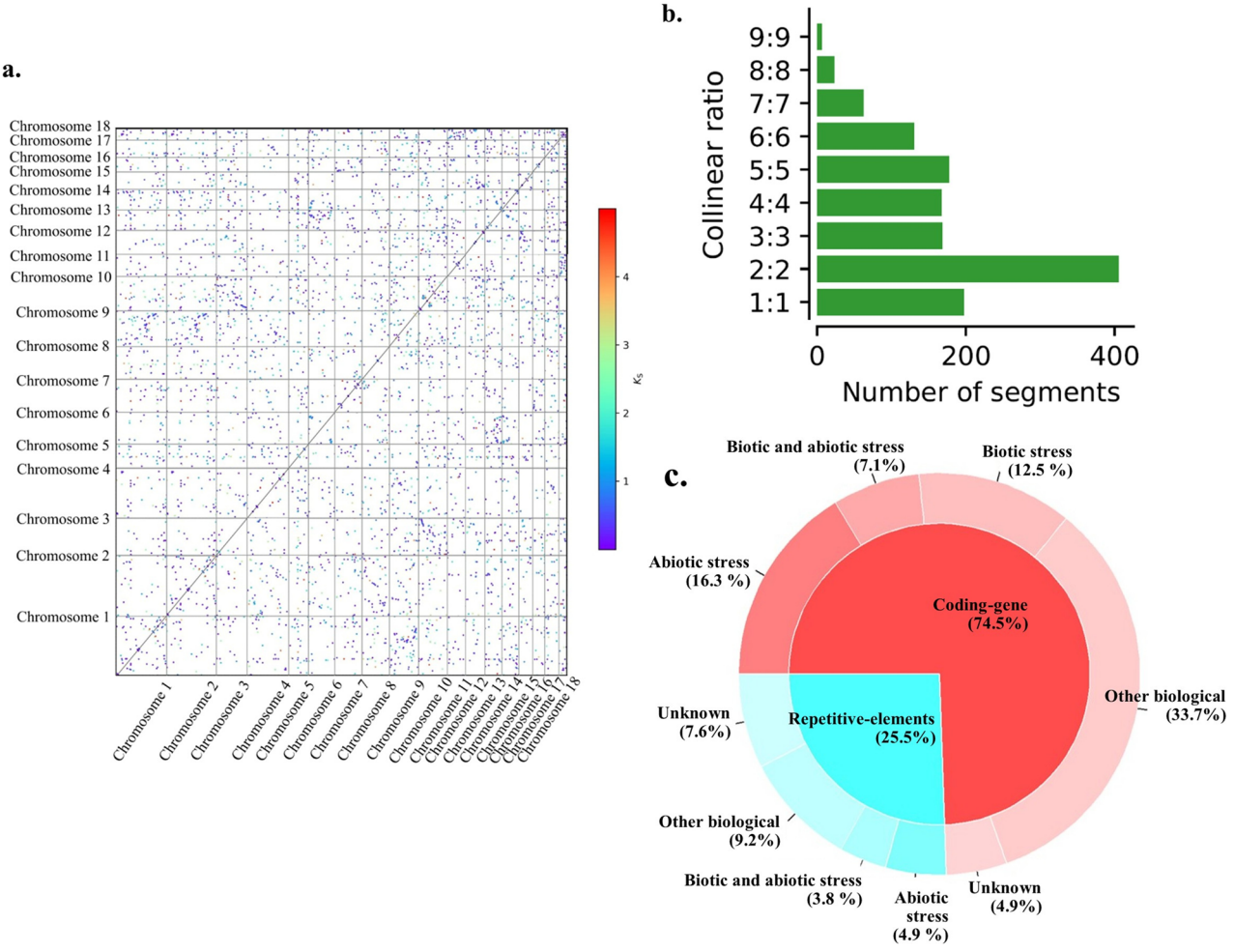

Fig. 3 Genomic duplication profile and functional insights in *C. humilis*. (**a**) Intraspecific genome homology dot plots for *C. humilis*. Anchor pairs (homologous genes residing in homeologous segments) are represented as dots colored by their associated K_S values. Axes show genes on corresponding chromosomes. (**b**) Intraspecific homology collinearity levels for *C. humilis*. The minimum length for segments is set to 100 Kbp. (**c**) Donut plot of the main functions of genes and repetitive elements duplicated exclusively in *C. humilis* more than seven times

in *P. dactylifera* and 8.0% in *E. guineensis* (Additional file 1: Fig. S11A).

K_S distributions were then analyzed by duplication type. For WGD-derived genes, a shared peak at $K_S = 0.27$ was observed across all three palm species (Additional file 1: Fig. S11B). This peak was also detected in TRD genes. The most striking difference lies in the distribution of dispersed duplications, which shows a nearly constant frequency across the entire K_S range, a pattern observed exclusively in *C. humilis* (Additional file 1: Fig. S11B).

Homology dot plot analysis revealed extensive collinearity among *C. humilis* chromosomes, with remnants of gene duplications distributed across all chromosomes (Fig. 3a). Most of the highlighted duplicated gene pairs showed K_S values < 2, consistent with relatively recent duplication events.

Intraspecific collinearity analysis of *C. humilis* (Fig. 3b) revealed an extensive number of multiplicons (homeologous regions present in multiple copies within the genome), with some regions duplicated up to nine times. Such extensive duplication was not observed in other palms: *E. guineensis* and *P. dactylifera* displayed no more than four-fold duplications (Additional file 1: Fig. S12). Comparative analysis identified 729 multiplicons uniquely duplicated in *C. humilis*, 17 of which were duplicated at least seven times.

Multiplicons duplicated ≥ 7-fold comprised 184 genes distributed across all chromosomes except chromosomes 11, 16, and 18 (Fig. 1, Additional file 1: Fig. S13). The regions exhibiting up to nine-fold duplication were located on chromosomes 1, 2, 5, 9, 10, 13, 14, and 15, with chromosome 1 showing the highest concentration (nine genes) (Additional file 1: Fig. S13). Overall, 74.5% of the duplicated elements corresponded to coding genes, whereas 25.5% were repetitive sequences (Fig. 3c). Among these duplications, DD were the most abundant

(58%), followed by TRD (24.6%), and WGD-derived duplications (11.2%).

Examination of K_S values of genes duplicated ≥ 7-fold, revealed that all duplications exhibit values of $\log_{10}(K_S)$ below 0. Genes derived from WGD and TRD show an increased number of copies with similar substitution rates ($\log_{10}(K_S)$ ~ -0.4) (Additional file 1: Fig. S14A), reflecting the gene expansions associated with the WGD event. In contrast, DD-derived genes display substitution rates closer to 0.13, indicating that these duplications are relatively more recent. An additional peak of genes with substitution rates near 0.01 may correspond to the burst of LTR insertions observed approximately 4 million years ago (Additional file 1: Fig S4, Fig. S14A). Analysis of K_A/K_S values of DD and TRD genes clustering around 0.4 and 0.3, respectively, while WGD genes are around 0.1, suggesting stronger functional constraint on the WGD-derived duplicates (Additional file 1: Fig. S14B).

Discussion

Chamaerops humilis is the only palm native to the western Mediterranean at day, being one of the most northerly palm species in Europe and is also one of the most cold-tolerant [24]. In this context, the distinct genomic features observed in *C. humilis*, likely underpin its adaptation to the Mediterranean's marginal conditions, differentiating it from tropical relatives. This study did not include analyses capable of demonstrating that genomic structural dynamics conferred resilience to the environmental fluctuations that occurred in the Mediterranean over geological timescales. Therefore, this putative adaptation remains a hypothesis requiring further testing.

Within the palm family, genome size shows remarkable variation, ranging from 0.53 Gbp in diploid *Licuala orbicularis* and *L. sarawakensis* to 30.63 Gbp in the 38-ploid *Voanioala gerardii* [25]. The *C. humilis* genome sequenced here spans 3.44 Gbp, assembled into 18 chromosomes [26], with an LAI value of 19.86, indicating high assembly quality [27]. The genome size estimated from *k-mer* analysis was smaller than the assembled genome size. This discrepancy likely reflects the high abundance of repetitive sequences, which are often underrepresented in *k-mer* spectra, but more comprehensively captured by assembly-based approaches [28].

Compared with the recently published *C. humilis* assembly [29], our genome shows an approximately 276 Mb difference in total chromosome sequence length. This can be explained we had sequenced both male and female individuals, whereas the previous assembly was derived from a single sex. Sex-associated differences in genome size have also been reported in other palms, like in *P. dactylifera* [30]. Differences are even greater in the unplaced scaffold fraction, most likely due to the high repeat content of these regions, which difficult the assembly and chromosomal anchoring. Comparisons with *P. dactylifera* and *E. guineensis*, the closest phylogenetic relatives with high-quality genomes [31, 32], highlight that *C. humilis* possesses a genome 3.6 and 2.0 times larger, respectively [33, 34]. Its genome is also ~ 1.9 times larger than that of the inferred common ancestor with *Phoenix* (Additional file 1: Fig. S8). These results suggest that *C. humilis* may have underwent genome expansion relative to its close relatives, likely influenced by repetitive element dynamics.

Increases in genome size in plants can result from polyploidization and repetitive element accumulation [32]. In palms, genome size variation appears to be primarily driven by repetitive DNA rather than polyploidy [25, 35, 36]. In *C. humilis*, TEs account for ~ 85% of the genome, a proportion similar to that of *C. nucifera*, but much higher than in *P. roebelenii* (~ 35%) (Additional file 2: Table S4). TE proliferation is known to drive genome expansion, sometimes in association with WGD [37]. Comparable examples include maize, where TE activity after WGD has resulted in a genome composed of ~ 85% TEs [38]. As in other plant lineages, the preferential expansion of specific TE lineages appears to contribute to genome size diversity in palms [25, 39–41]. In *C. humilis*, LTR/Copia and LTR/Gypsy were the dominant retrotransposons, with lineage composition largely consistent with other palms [42].

LTR/Copia families, such as SIRE, are reported to expand preferentially under environmental stress [25, 43]. In *C. humilis*, we detected a burst of LTR/Copia insertions around 4 Mya. Although this study does not demonstrate a direct causal link, the temporal concordance suggests that environmental stress associated with the Messinian Salinity Crisis and the consequent aridification of the Mediterranean Basin [14, 44] may have contributed to TE proliferation in this lineage.

The WGD event identified in *C. humilis* (~ 48 Mya) corresponds to the ancestral WGD reported for Arecaceae [45]. This event is inferred to have occurred after the origin of the family in the Late Cretaceous (~ 110 Mya) [46, 47], but before the diversification of the major Arecoid clade. Rediploidization following WGD is thought to restore diploid-like inheritance through genome restructuring and selective retention of paralogs [48]. In *C. humilis*, functional enrichment of rapidly evolving gene families highlighted categories related to genome regulation (e.g., cell cycle control, interstrand cross-link repair, and transposition) as well as environmental stress responses. These included brassinosteroid-related terms, associated in other plants with drought and salt responses [49], and auxin activated signalling pathway, which has been reported in *E. guineensis* and *C. nucifera* to be involved in abiotic stresses, including cold, drought, and osmotic stress [50, 51]. These findings suggest that

hormonal regulation may also have contributed to stress resilience.

Interestingly, only 20.62% of duplicated genes in *C. humilis* derive from WGD, compared with >50% in *P. dactylifera* and *E. guineensis*. This suggests that other duplication mechanisms, particularly TRD and DD, have played a more prominent role, like it had been seen in other species [52–54]. DD result in non-colinear gene copies, a mechanism observed across plant genomes but still not fully understood [55–57]. Notably, a subset of DD may represent remnants of ancient tandem or proximal duplications. Their original syntenic context may have been obscured by extensive genome rearrangements, gene loss, and TE insertions. These processes disrupt collinearity and can cause formerly syntenic paralogs to be misclassified as dispersed duplicates [58, 59].

Relatively constant K_S distributions for TD and PD have been reported in other plant species [60]. Accordingly, the apparent continuous occurrence of DD across a broad K_S range likely reflects their heterogeneous evolutionary origins rather than a single, temporally restricted duplication burst.

Uniquely duplicated genes in *C. humilis*, in some cases expanded up to 7–9 copies, included proteins implicated in stress responses. These comprised aspartic peptidases (linked to cold tolerance and pathogen defense [61, 62]), GTP-binding proteins and RING-type E3 ligases (stress signaling [63]), NAC transcription factors (abiotic stress responses [64]), SUT1 transporters (immune regulation [65]), and LRR proteins (pathogen defense [66]). Retrotransposons such as LTR/Gypsy and LTR/Copia, which are stress-responsive in other species [67, 68], may also have contributed to genomic variations. While these expansions are consistent with potential adaptive advantages, functional validation will be required to confirm their roles in stress tolerance.

WGD-derived duplicated genes exhibited comparatively lower K_A/K_S ratios than those originating from other duplication mechanisms, which may indicate that these genes have been subject to stronger long-term purifying selection [69]. In contrast, relatively higher K_A/K_S ratios were observed in DD and TRD genes, which have relatively low K_S values. This pattern may align with previous studies proposing that positive selection may contribute to the early retention and functional divergence of duplicated genes [70–72]. However, additional analyses will be necessary to distinguish selective pressures from neutral processes for understanding duplicate gene evolution.

The Messinian Salinity Crisis and subsequent Zanclean flood represented dramatic environmental shifts in the Mediterranean Basin, leading to widespread extinctions [15]. The persistence of *C. humilis* despite these events is consistent with the hypothesis that genome plasticity and stress-related gene family expansions may have facilitated survival under fluctuating conditions. However, direct evidence for this link is currently lacking and requires further investigation.

Conclusion

Overall, our study provides a high-quality reference genome for *C. humilis* and new insights into the genomic mechanisms potentially underpinning its persistence in the Mediterranean Basin. Future approaches integrating comparative and population genomics, pangenome analysis, transcriptomics, proteomics, epigenomics, and functional validation will be essential to test these hypotheses and to fully understand the adaptive evolution of this emblematic palm.

Materials and methods

Plant materials and nucleic acid extraction

The plant material from which nucleic acid was extracted consisted of young leaves taken from the yearlings of individuals of *C. humilis* each carrying remnants of male and female inflorescences separately. Fresh leaves were sampled from individuals living in the Circeo National Park (Latina, Italy).

DNA extraction was performed on leaves from individuals separately following Russo et al. [73], with slight modifications as outlined below. Young leaves were collected, flash-frozen in liquid nitrogen, and stored at − 80 °C until DNA extraction. A 100-mg aliquot of frozen leaf tissue was ground to a fine powder in liquid nitrogen using a mortar and pestle. The powder was then transferred to a sterile 2-mL centrifuge tube containing 600 µL of SDS lysis buffer (1% polyvinylpyrrolidone 40 (PVP40), 1% sodium metabisulphite, 0.5 M sodium chloride, 100 mM Tris–HCl (pH 8), 50 mM EDTA (pH 8), 2% β-mercaptoethanol (β-ME), and 1.5% sodium dodecyl sulfate (SDS)). The sample was vortexed for 3–5 s and incubated in a thermomixer with gentle agitation (400 rpm for 20 min at 55 °C), adding 4 µL of 100 mg/mL DNase-free RNase A (Qiagen, Germantown, MD, USA). After the incubation, 200 µL of 5 M potassium acetate was added, and the mixture was inverted for mixing. Next, 800 µL of phenol: chloroform: isoamyl alcohol (25:24:1, v/v, pH 8) was added, and the sample was incubated for 10 min at room temperature (RT). The sample was then centrifuged for 10 min at 10,000 × g at RT. The supernatant was carefully transferred to a new 2-mL tube, and the extraction was repeated one more time. The resulting sample was purified using magnetic beads (Sera-Mag SpeedBeads™ Carboxyl Magnetic Beads, GE Healthcare 65152105050250, Fisher Scientific). The sample underwent three washes with 1 mL of 70% ethanol (EtOH) before being eluted in 50 µL of 10 mM Tris–HCl (pH 8.5). Quality control was performed on the extracted

DNA before proceeding with library preparation and sequencing.

Crosslinking of the solid tissue samples was done for HiC [74]. Briefly, 2 g of young male leaf tissue was ground into a fine powder in a liquid nitrogen-chilled mortar; once the tissue was powdered, it was resuspended in 10 times its volume in 1% formaldehyde (37%). After resuspension, the sample was incubated at room temperature for 20 min, with periodic mixing or vortexing to ensure thorough mixing. Following the incubation, glycine was added to the sample to achieve a final concentration of 125 mM, and the sample was incubated at room temperature for an additional 15 min, with periodic mixing. After the second incubation, the sample was centrifuged (1000 × g for 1 min) to pellet the crosslinked tissue powder. The pellet was then stored at − 80 °C until further analysis.

To aid in gene annotation and phylogenomic analyses, fresh leaves from male and female individuals of *C. humilis* were collected for RNA sequencing (RNA-seq). RNA was extracted from leaf samples, which were collected and immediately frozen in liquid nitrogen and transported to the laboratory without interrupting the cold chain. The samples were pulverized in liquid nitrogen using a mortar and pestle, and 50 mg of each sample was taken for subsequent extractions using the RNeasy Plant Mini Kit (QIAGEN) following the manufacturer's instructions. High-quality RNA-seq libraries were prepared and sequenced with the Illumina NovaSeq 6000 platform.

Genome sequencing and assembly

Genomic DNA from male and female individuals of *C. humilis* was utilized to construct PacBio and Illumina sequencing libraries, while DNA from male individuals was employed for Hi-C library preparation. A SMRTbell library was constructed with Sequel 1.0 reagents using PacBio Sequel, and paired-end libraries for Illumina were prepared following the manufacturer's protocol. Hi-C library construction was performed with the Proximo™ Hi-C Plant kit (Phase Genomics) following the manufacturer's protocol. Raw reads were evaluated for quality and filtered using Filtlong (v0.2.0) (https://github.com/rrwick/Filtlong) and NextDenovo (v2.4.0) [75] and Seqtk (v4) (https://github.com/lh3/seqtk), removing poor-quality sequences and filtering out sequences < 2.5 Kbp and > 50 Kb, which showed skewed nucleotide frequencies. A *k-mer*-based statistical approach was used to estimate the genome size, heterozygosity, and repeat content of *C. humilis*. The analysis was conducted using short reads from whole-genome shotgun and Hi-C libraries with Jellyfish [76] and GenomeScope tools [77], but the convergence of the k-mer model was achieved only with the Hi-C reads.

PacBio reads were utilized for a *de novo* assembly using NextDenovo v2.5.0 [75]. The process involved two stages, an initial correction of the raw reads, followed by the assembly of a consensus sequence of the corrected reads using a string graph algorithm. The completeness of the genome assembly, generated by using primary contigs, was evaluated using BUSCO (v5.0) against the *viridiplantae_odb10* database [78]. To improve the assembly quality, a polishing step was performed with NextPolish [79] using both long and short reads, to correct small insertions, deletions, and single-base substitutions.

To recover the potential portion of the genome not assembled in the long-reads assembly process, all short reads were mapped against the draft genome. Approximately 95–98% of the reads were successfully mapped, and unmapped reads, accounting for about 15 Gbp of data, were then extracted and used as input for a *de novo* assembly approach using SPAdes (v3.15.5) [80].

Contigs were analysed to detect possible separation of primary contigs from secondary/alternatives using the Purge_haplotigs pipeline [81]. Based on the plot distribution, specific cutoffs were set to identify contigs with very low or high coverage (representing < 1%), and these contigs were marked and excluded from the assembly. Contigs from the sequence assemblers were linked to form the scaffold-scale assembly, based on the ALLHi-C pipeline [82]. The completeness of the genome assembly was evaluated using BUSCO (v5.0) [78].

Scaffolds were reordered and oriented to chromosome level using RagTag [83], with the chromosome-based genome assembly of *C. humilis* (NCBI Accession: GCA_042465385.1 [29]), used as reference. Rag Tag was applied exclusively for scaffold ordering and orientation, without altering scaffold sequences. Consequently, chromosome-scale structural interpretations should be made with consideration of this reference-guided step.

Transcriptome assembly

Prior to further analysis, a quality check was performed on the raw sequencing data, removing low-quality portions while preserving the longest high-quality part of reads using BBDuck, normalized with BBNorm (https://sourceforge.net/projects/bbmap/) and joined to generate a comprehensive dataset. Normalized high-quality reads were assembled with Trinity (v2.4.0) and the accuracy and completeness of the created assembly were checked using BUSCO (v5.0) against the *viridiplantae_odb10* database [78]. Open reading frames were predicted with TransDecoder (https://github.com/TransDecoder/TransDecoder), generating a reference set of putative protein-coding transcripts.

Genome annotation

Gene prediction was performed using the chromosome-level genome assembly together with transcriptomic evidence. The Trinity-derived transcripts and their corresponding predicted proteins (99,980 sequences) were incorporated as evidence. Additionally, RNA-seq reads were mapped to the assembly using STAR [84]; the total number of mapped reads was 28.8 million, 92.5% of which mapped uniquely.

MAKER2 [85] was used as a framework to organize the structural genome annotation workflow. For the genome annotation process, repeats were soft masked using the Red algorithm [86], masking 82.3% of the genome (2.8 Gbp). A protein database combining 121,927 sequences from transcriptome assemblies, Viridiplantae UniRef90, and Coryphoideae was constructed. The soft-masked genome, protein database, and RNA-seq were used to generate training models and a draft annotation using Braker3 [87], with additional annotations provided by the Helixer pipeline [88]. Proteins from *P. dactylifera* were mapped using Miniprot [89], and along with all other results, were integrated using TSEBRA [90]. TransDecoder annotated CDS sequences [91], and functional annotation was performed with Pannzer2 [92]. To reduce transcript redundancy and retain a single representative transcript per gene, we further processed the annotation set using EvidentialGene pipeline [93], yielding a non-redundant gene set.

A high-quality non-redundant TE library was generated by the EDTA pipeline [94] and the inbuilt RepeatModeler [95]. The identification of TE was performed by using RepeatMasker (v1.332) [96] utilizing the NCBI/RMBLAST search engine (v2.6.0). The long terminal repeats (LTR) Assembly Index (LAI) [27], a metric to evaluate the quality and completeness of genome assembly based on LTR retrotransposons, was determined using LTR_retriever [97]. Full-length LTR retrotransposons were annotated based on the Viridiplantae database v4.0 [98].

The genome karyotype, gene distribution, and transposon density were visualized using Circos [99]. Figures were generated in R v.4.4.2 (R Core team, 2021) using ggplot [100] and ggraphExtra [101] packages.

Chloroplast genome analysis

High-accuracy PacBio reads from both male and female individuals were used to assemble the chloroplast genome. The selected high-quality reads were processed with Oatk software (v. 1.0) [102] to reconstruct the complete plastid genome, annotation were obtained using CPGAVAS2 [103] taking as reference *P. dactylifera* chloroplast [23]. The assembly was aligned using MAFFT with the other reference plastid genomes of Arecaceae (Additional file 2:Table S8), and a maximum-likelihood phylogenetic tree was constructed with IQ-TREE3 [104, 105].

Phylogenetic and genomic comparative analysis

To investigate the evolutionary history and genomic diversification of *C. humilis*, a phylogenetic tree was conducted, using a diverse set of eudicots (4 genomes) and monocotyledons (20 genomes) (Additional file 2: Table S5), including all the Arecaceae genomes available (9 genomes) (Additional file 2: Table S4), to evaluate changes in genome size and gene family evolution.

To obtain putative orthologues from each genome, HMMER [106] was used taking as reference single-copy orthogroups (SOG) from the *embryophyta_odb10* BUSCO dataset [78], and when multiple hits were observed, the best hit was selected. The final species tree was inferred using 241 high-confidence SOGs genes. Multiple sequence alignment was inferred for each SOG using MAFFT [107], and a maximum likelihood tree for each alignment was constructed using IQ-TREE3 [104, 105]. The resulting individual trees were supplied to ASTRAL-Pro to summarize the species tree under default parameters [108]. To obtain the time-calibrated tree, SOGs observed in all the species were used using mcmctree (v4.10) [109] and parameter set was set as default (use data = 3, clock = 2, model = 0, alpha = 0) [110]. The following calibration points were used: *Coffea arabica* and *Arabidopsis thaliana* divergence (114-123.9 million years ago (mya), *Zea mays* and *Oryza sativa* divergence (41.1–51.9 mya), and *Oryza sativa* and *Elaeis guineensis* divergence (103-119.6 mya). The resulting tree was plotted using MCMCtreeR package in R [111].

The phytools package in R [112] was used to estimate the ancestral genome size of Arecaceae, and gene families were determined in primary proteins by Orthofinder v 2.5.5 [113]. Expansion and contraction of gene families were observed using CAFE (v.5.1.0) [114]. Contracted and expanded gene families were analyzed, and GO enrichment was conducted using topGO [115]. Synteny analysis was performed between the two reference Arecaceae genomes, *P. dactylifera* and *E. guineensis*, using JCVI (v 1.4.25) [116].

Whole-Genome Duplication analysis

WGD analysis was conducted to determine the presence of lineage-specific and/or shared WGD events among Arecaceae, using wgd pipeline (v2.0.38) [117].

The whole paranome (complete set of paralogous genes within a genome) was constructed using 'wgd dmd' with default parameters. Only one transcript per gene was included to avoid redundancy, gene family clustering was conducted using the Markov Cluster Algorithm (MCL) with an inflation factor of 2. To infer genomic collinearity, 'wgd syn' was applied using i-ADHoRe (v3.0.01) [118].

Subsequently, 'wgd ksd' was used to construct K_S distributions for both the paranome and anchor pairs (pairs of duplicated genes derived from WGD events which reside in duplicated segments) [117]. K_S estimation was performed using the non-pairwise mode (default setting), in which KS values are inferred from full multiple sequence alignments, ensuring that the evolutionary history of each alignment column is traced back to the root of the gene family, thereby accounting for all gene duplicates simultaneously and providing a more biologically conserved estimate of K_S.

To identify potential WGD components, the exponential-lognormal mixture model (ELMM) and the log-scale Gaussian mixture model (GMM) were employed within 'wgd viz' to identify potential WGD components from the whole paranome and anchor pair K_S distributions, respectively. For anchor pairs, only K_S values falling within the 95% of confidence interval were retained using 'wgd peak', excluding outliers and reducing the influence of K_S saturation. Using the K_S distribution of anchor pairs, WGD peaks were identified through segment-guided anchor pair clustering, as implemented in wgd peak. Default parameters were used, including a K_S saturation cutoff of 5, a confidence interval of 95% for log-normal distributions, and a component range of 1–4 for mixture model fitting.

Using a phylogenetic tree and fossil calibration of 17 related species previously reported by Chen et al. [117], orthogroups were constructed by combining the anchor pairs with related homologous genes. WGD events were dated with 'wgd focus' using mcmctree from PAML (v4.9) [109].

When the paralogs of *C. humilis* and orthologs with other Arecaceae species were identified, only those exclusive of *C. humilis* and duplicated at least 7 times were evaluated. A functional analysis (Gene Ontology analysis and literature review) was conducted to determine their biological significance.

Duplicated gene pairs were identified and classified into segmental duplications (WGD-derived) and small-scale duplications (SSD) derived gene pairs using the R package doubletrouble [119]. Paralog genes that are used as anchor pairs in syntenic regions are classified as WGD-derived, typically originating from whole-genome duplications. All other duplicates are classified as SSD. These can be further classified based on their proximity: those adjacent in the genome are tandem duplications (TD); those separated by only a few genes (default is 10, adjustable) are classified as proximal duplications (PD); and all others fall under dispersed duplications (DD). Additionally, some duplicates arise from transposon-derived duplications (TRD). The location of duplicated genes and classification were plotted using chromomap in R [120].

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-026-08500-2.

Additional File 1: Figure S1, GenomeScope analysis of *C. humilis* genome. Figure S2, Chromosome reconstruction of *C. humilis*. Figure S3, GO classification of *C. humilis* genes. Figure S4, Distribution of LTR transposon insertion times. Figure S5, Chloroplast genome of *C. humilis*. Figure S6, Phylogenetic tree of chloroplast genome from Arecaceae family. Figure S7, Synteny depth analysis between *C. humilis* with *P. dactylifera* (A) and *E. guineensis* (B). Figure S8, Genome size evolution and gene family dynamics in Arecaceae. Figure S9, Anchor Ks and Ka distribution of *Chamaerops humilis*. Figure S10, Peaks emerged in the anchor pair Ks and posterior date distributions of WGD dating. Figure S11, Classification of genes based on their duplication type (A) and Ks distribution (B). Figure S12, Intraspecific homology collinearity levels for *P. dactylifera* (A) and *E. guineensis* (B). Figure S13, The karyotype of *C. humilis* showing duplicated genes.

Additional File 2: Table S1, BUSCO analysis during genome assembly. Table S2, Genome annotation summary. Table S3, Whole-genome TE annotation obtained by the EDTA pipeline. Table S4, Summary statistics of genome assembly of *C. humilis* and other Arecaceae. Table S5, Full length LTR annotation. Table S6, Chloroplast genomes used for phylogenetic analysis. Table S7, Gene Ontology enrichment of rapidly evolving gene families in *Chamaerops humilis*. Table S8, Nuclear genomes used for phylogenetic analysis.

Acknowledgements
There are no contributions from non-authors.

Authors' contributions
G. C., Z. D., T. M., P. (A) conceived, wrote, and revised the article, L-O. M. performed bioinformatic analysis and wrote and revised the article, T. M., R. M-A., E. F. and M. M. extracted nucleic acid material, C. G., S. D. and S. P. revised the article, S. R. and (B) G. provided biological material, (C) H. contributed with bioinformatic analysis and revised the manuscript.

Funding
This work was supported by the European Union - Next-GenerationEU - National Recovery and Resilience Plan (NRRP) – MISSION 4 COMPONENT 2, INVESTIMENT N. 1.1, project reference: ChAMPION - ***Chamaerops humilis*** reference genome sequencing for Active conservation of an endemic Mediterranean Palm Including analyses Of biotic interactions Network, CUP F53D2300817000.

Data availability
The ***C. humilis*** genome assembly was deposited in NCBI GenBank under accession number JBNOGQ000000000.1 in the BioProject PRJNA1247256. Chromosome assembly FASTA files have been uploaded and are publicly available at the link https://www.ncbi.nlm.nih.gov/Traces/wgs/JBNOGQ01?display=contigs .The chloroplast assembly was deposited in the Zenodo repository (https://doi.org/10.5281/zenodo.18671330) [121].

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 10 December 2025 / Accepted: 1 March 2026
Published online: 04 March 2026

References

1. Theissinger K, Fernandes C, Formenti G, Bista I, Berg PR, Bleidorn C, et al. How genomics can help biodiversity conservation. Trends Genet. 2023;39:545–59. https://doi.org/10.1016/j.tig.2023.01.005.
2. Twyford AD. The road to 10,000 plant genomes. Nat Plants. 2018;4:312–3. https://doi.org/10.1038/s41477-018-0165-2.
3. Mc Cartney AM, Formenti G, Mouton A, De Panis D, Marins LS, Leitão HG, et al. The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. Npj Biodivers. 2024;3:1–17. https://doi.org/10.1038/s44185-024-00054-6.
4. Kress WJ, Soltis DE, Kersey PJ, Wegrzyn JL, Leebens-Mack JH, Gostel MR et al. Green plant genomes: What we know in an era of rapidly expanding opportunities. Proceedings of the National Academy of Sciences. 2022;119:e2115640118. https://doi.org/10.1073/pnas.2115640118
5. Panchy N, Lehti-Shiu M, Shiu S-H. Evolution of Gene Duplication in Plants. Plant Physiol. 2016;171:2294–316. https://doi.org/10.1104/pp.16.00523.
6. Wendel JF, Jackson SA, Meyers BC, Wing RA. Evolution of plant genome architecture. Genome Biol. 2016;17:37. https://doi.org/10.1186/s13059-016-0908-1.
7. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. Nat Rev Genet. 2009;10:725–32. https://doi.org/10.1038/nrg2600.
8. Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, et al. Ancestral polyploidy in seed plants and angiosperms. Nature. 2011;473:97–100. https://doi.org/10.1038/nature09916.
9. Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. Curr Opin Genet Dev. 2015;35:119–25. https://doi.org/10.1016/j.gde.2015.11.003.
10. Rensing SA. Gene duplication as a driver of plant morphogenetic evolution. Curr Opin Plant Biol. 2014;17:43–8. https://doi.org/10.1016/j.pbi.2013.11.002.
11. Pozzi CM, Gaiti A, Spada A. Climate change and plant genomic plasticity. Theor Appl Genet. 2025;138:231. https://doi.org/10.1007/s00122-025-05010-x.
12. Aurelle D, Thomas S, Albert C, Bally M, Bondeau A, Boudouresque C-F, et al. Biodiversity, climate change, and adaptation in the Mediterranean. Ecosphere. 2022;13:e3915. https://doi.org/10.1002/ecs2.3915.
13. Cornacchia I, Brandano M, Agostini S. Miocene paleoceanographic evolution of the Mediterranean area and carbonate production changes: A review. Earth Sci Rev. 2021;221:103785. https://doi.org/10.1016/j.earscirev.2021.103785.
14. Krijgsman W, Hilgen FJ, Raffi I, Sierro FJ, Wilson DS. Chronology, causes and progression of the Messinian salinity crisis. Nature. 1999;400:652–5. https://doi.org/10.1038/23231.
15. Agiadi K, Hohmann N, Gliozzi E, Thivaiou D, Bosellini FR, Taviani M, et al. Late Miocene transformation of Mediterranean Sea biodiversity. Sci Adv. 2024;10:eadp1134. https://doi.org/10.1126/sciadv.adp1134.
16. Jiménez-Moreno G, Fauquette S, Suc J-P. Miocene to Pliocene vegetation reconstruction and climate estimates in the Iberian Peninsula from pollen data. Rev Palaeobot Palynol. 2010;162:403–15. https://doi.org/10.1016/j.revpalbo.2009.08.001.
17. Salocchi AC, Krawielicki J, Eglinton TI, Fioroni C, Fontana D, Conti S, et al. Biomarker constraints on Mediterranean climate and ecosystem transitions during the Early-Middle Miocene. Palaeogeogr Palaeoclimatol Palaeoecol. 2021;562:110092. https://doi.org/10.1016/j.palaeo.2020.110092.
18. Jiménez-Moreno G, Suc J-P. Middle Miocene latitudinal climatic gradient in Western Europe: Evidence from pollen records. Palaeogeogr Palaeoclimatol Palaeoecol. 2007;253:208–25. https://doi.org/10.1016/j.palaeo.2007.03.040.
19. Yao G, Zhang Y-Q, Barrett C, Xue B, Bellot S, Baker WJ, et al. A plastid phylogenomic framework for the palm family (Arecaceae). BMC Biol. 2023;21:50. https://doi.org/10.1186/s12915-023-01544-y.
20. Guzmán B, Fedriani JM, Delibes M, Vargas P. The colonization history of the Mediterranean dwarf palm (Chamaerops humilis L., Palmae). Tree Genet Genomes. 2017;13:24. https://doi.org/10.1007/s11295-017-1108-1.
21. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82. https://doi.org/10.1038/nrg2165.
22. Chen D-J, Landis JB, Wang H-X, Sun Q-H, Wang Q, Wang H-F. Plastome structure, phylogenomic analyses and molecular dating of Arecaceae. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.960588.
23. Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (Phoenix dactylifera L). PLoS ONE. 2010;5:e12762. https://doi.org/10.1371/journal.pone.0012762.
24. Giovino A, Scibetta S, Saia S, Guarino C. Genetic and morphologic diversity of European fan palm (Chamaerops humilis L.) populations from different environments from Sicily. Bot J Linn Soc. 2014;176:66–81. https://doi.org/10.1111/boj.12195.
25. Schley RJ, Pellicer J, Ge X-J, Barrett C, Bellot S, Guignard MS, et al. The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. New Phytol. 2022;236:433–46. https://doi.org/10.1111/nph.18323.
26. Röser M. Variation and Evolution of Karyotype Characters in Palm Subfamily Coryphoideae s.l. Bot Acta. 1993;106:170–82. https://doi.org/10.1111/j.1438-8677.1993.tb00354.x.
27. Ou S, Chen J, Jiang N. Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. 2018;46:e126–126. https://doi.org/10.1093/nar/gky730.
28. Jenike KM, Campos-Domínguez L, Boddé M, Cerca J, Hodson CN, Schatz MC, et al. k-mer approaches for biodiversity genomics. Genome Res. 2025;35:219–30. https://doi.org/10.1101/gr.279452.124.
29. Beltran-Sanz N, Prost S, Malavasi V, Moschin S, Greve C, Schell T, et al. Chromosome-level assembly of the 400-year-old Goethe's Palm (Chamaerops humilis L). Sci Data. 2025;12:1542. https://doi.org/10.1038/s41597-025-05673-7.
30. Torres MF, Mathew LS, Ahmed I, Al-Azwani IK, Krueger R, Rivera-Nuñez D, et al. Genus-wide sequencing supports a two-locus model for sex-determination in Phoenix. Nat Commun. 2018;9:3969. https://doi.org/10.1038/s41467-018-06375-y.
31. Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. Genome Size Diversity and Its Impact on the Evolution of Land Plants. Genes. 2018;9:88. https://doi.org/10.3390/genes9020088.
32. He B, Liu W, Li J, Xiong S, Jia J, Lin Q, Genomics, et al. Proteom Bioinf. 2024;22:qzae078. https://doi.org/10.1093/gpbjnl/qzae078.
33. Wang L, Lee M, Yi Wan Z, Bai B, Ye B, Alfiko Y, et al. A Chromosome-Level Reference Genome of African Oil Palm Provides Insights into Its Divergence and Stress Adaptation. Genom Proteom Bioinform. 2023;21:440–54. https://doi.org/10.1016/j.gpb.2022.11.002.
34. Hazzouri KM, Gros-Balthazard M, Flowers JM, Copetti D, Lemansour A, Lebrun M, et al. Genome-wide association mapping of date palm fruit traits. Nat Commun. 2019;10:4680. https://doi.org/10.1038/s41467-019-12604-9.
35. Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. Genera Palmarum: the evolution and classification of palms. Richmond: Kew Publ; 2008.
36. Barrett CF, McKain MR, Sinn BT, Ge X-J, Zhang Y, Antonelli A, et al. Ancient Polyploidy and Genome Evolution in Palms. Genome Biol Evol. 2019;11:1501–11. https://doi.org/10.1093/gbe/evz092.
37. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 2014;65:505–30. https://doi.org/10.1146/annurev-arplant-050213-035811.
38. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326:1112–5. https://doi.org/10.1126/science.1178534.
39. Macas J, Novák P, Pellicer J, Čížková J, Koblížková A, Neumann P, et al. In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae. PLoS ONE. 2015;10:e0143424. https://doi.org/10.1371/journal.pone.0143424.
40. Beric A, Mabry ME, Harkess AE, Brose J, Schranz ME, Conant GC et al. Comparative phylogenetics of repetitive elements in a diverse order of flowering plants (Brassicales). G3 Genes|Genomes|Genetics 2021;11:jkab140. https://doi.org/10.1093/g3journal/jkab140
41. Pellicer J, Fernández P, Fay MF, Michálková E, Leitch IJ. Genome Size Doubling Arises From the Differential Repetitive DNA Dynamics in the Genus Heloniopsis (Melanthiaceae). Front Genet. 2021;12. https://doi.org/10.3389/fgene.2021.726211.
42. Chaw S-M, Chang C-C, Chen H-L, Li W-H. Dating the Monocot–Dicot Divergence and the Origin of Core Eudicots Using Whole Chloroplast Genomes. J Mol Evol. 2004;58:424–41. https://doi.org/10.1007/s00239-003-2564-9.
43. Galindo-González L, Mhiri C, Deyholos MK, Grandbastien M-A. LTR-retrotransposons in plants: Engines of evolution. Gene. 2017;626:14–25. https://doi.org/10.1016/j.gene.2017.04.051.
44. Krijgsman W, Rohling EJ, Palcu DV, Raad F, Amarathunga U, Flecker R, et al. Causes and consequences of the Messinian salinity crisis. Nat Rev Earth Environ. 2024;5:335–50. https://doi.org/10.1038/s43017-024-00533-1.
45. Chen H, Almeida-Silva F, Logghe G, Bonte D, de Peer YV. The Rise of Polyploids During Environmental Catastrophes 2024:2024.11.22.624806. https://doi.org/10.1101/2024.11.22.624806

46. Couvreur TL, Forest F, Baker WJ. Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. BMC Biol. 2011;9:44. https://doi.org/10.1186/1741-7007-9-44.
47. Foster CSP, Sauquet H, van der Merwe M, McPherson H, Rossetto M, Ho SYW. Evaluating the Impact of Genomic Data and Priors on Bayesian Estimates of the Angiosperm Evolutionary Timescale. Syst Biol. 2017;66:338–51. https://doi.org/10.1093/sysbio/syw086.
48. Wolfe KH. Yesterday's polyploids and the mystery of diploidization. Nat Rev Genet. 2001;2:333–41. https://doi.org/10.1038/35072009.
49. Miao R, Li C, Liu Z, Zhou X, Chen S, Zhang D, et al. The Role of Endogenous Brassinosteroids in the Mechanisms Regulating Plant Reactions to Various Abiotic Stresses. Agronomy. 2024;14:356. https://doi.org/10.3390/agronomy14020356.
50. Santhi CKV, Rajesh MK, Ramesh SV, Muralikrishna KS, Gangaraj KP, Gupta P, et al. Genome-wide exploration of auxin response factors (ARFs) and their expression dynamics in response to abiotic stresses and growth regulators in coconut (Cocos nucifera L). Plant Gene. 2021;28:100344. https://doi.org/10.1016/j.plgene.2021.100344.
51. Jin L, Yarra R, Zhou L, Cao H. The auxin response factor (ARF) gene family in Oil palm (Elaeis guineensis Jacq.): Genome-wide identification and their expression profiling under abiotic stresses. Protoplasma. 2022;259:47–60. https://doi.org/10.1007/s00709-021-01639-9.
52. Lekired A, Mudge J, Laidemitt MR, Mutuku MW, Loker ES, Zhang S-M. Whole-genome sequencing and genome-wide transcriptome profiling of the freshwater planorbid snail Bulinus ugandae (Mollusca: Gastropoda), a Nilotic bulinine refractory to Schistosoma haematobium. BMC Genomics. 2025;26:1089. https://doi.org/10.1186/s12864-025-12320-3.
53. Gaccione L, Toppino L, Bolger M, Schmidt M, Tassone MR, Sulli M, et al. Graph-based pangenomes and pan-phenome provide a cornerstone for eggplant biology and breeding. Nat Commun. 2025;16:9919. https://doi.org/10.1038/s41467-025-64866-1.
54. Karbstein K, Choudhary N, Xie T, Tomasello S, Wagner ND, Barke BH, et al. Assembling genomes of non-model plants: A case study with evolutionary insights from Ranunculus (Ranunculaceae). Plant J. 2025;123:e70390. https://doi.org/10.1111/tpj.70390.
55. Ganko EW, Meyers BC, Vision TJ. Divergence in Expression between Duplicated Genes in Arabidopsis. Mol Biol Evol. 2007;24:2298–309. https://doi.org/10.1093/molbev/msm158.
56. Wang Y, Ficklin SP, Wang X, Feltus FA, Paterson AH. Large-Scale Gene Relocations following an Ancient Genome Triplication Associated with the Diversification of Core Eudicots. PLoS ONE. 2016;11:e0155637. https://doi.org/10.1371/journal.pone.0155637.
57. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, et al. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 2019;20:38. https://doi.org/10.1186/s13059-019-1650-2.
58. Abdullah M, Sabir IA, Shah IH, Sajid M, Liu X, Jiu S, et al. The role of gene duplication in the divergence of the sweet cherry. Plant Gene. 2022;32:100379. https://doi.org/10.1016/j.plgene.2022.100379.
59. Chen H, Zhang Y, Feng S. Whole-genome and dispersed duplication, including transposed duplication, jointly advance the evolution of TLP genes in seven representative Poaceae lineages. BMC Genomics. 2023;24:290. https://doi.org/10.1186/s12864-023-09389-z.
60. Lekired A, Mudge J, Laidemitt MR, Mutuku MW, Loker ES, Zhang S-M. Whole-genome sequencing and genome-wide transcriptome profiling of the freshwater planorbid snail Bulinus ugandae (Mollusca: Gastropoda), a Nilotic bulinine refractory to Schistosoma haematobium. BMC Genomics. 2025;26:1089. https://doi.org/10.1186/s12864-025-12320-3.
61. Wang B, Wang G, Zhu S. DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit. Front Plant Sci. 2020;10. https://doi.org/10.3389/fpls.2019.01723.
62. Figueiredo L, Santos RB, Figueiredo A. Defense and Offense Strategies: The Role of Aspartic Proteases in Plant–Pathogen Interactions. Biology. 2021;10:75. https://doi.org/10.3390/biology10020075.
63. Shu K, Yang W. E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses. Plant Cell Physiol. 2017;58:1461–76. https://doi.org/10.1093/pcp/pcx071.
64. Mitsuda N, Seki M, Shinozaki K, Ohme-Takagi M. The NAC Transcription Factors NST1 and NST2 of Arabidopsis Regulate Secondary Wall Thickenings and Are Required for Anther Dehiscence. Plant Cell. 2005;17:2993–3006. https://doi.org/10.1105/tpc.105.036004.
65. Yan J, Liu Y, Huang X, Li L, Hu Z, Zhang J, et al. An unreported NB - LRR protein SUT 1 is required for the autoimmune response mediated by type one protein phosphatase 4 mutation (topp4-1) in *Arabidopsis*. Plant J. 2019;100:357–73. https://doi.org/10.1111/tpj.14447.
66. Afzal AJ, Wood AJ, Lightfoot DA. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. Mol Plant Microbe Interact. 2008;21:507–17. https://doi.org/10.1094/MPMI-21-5-0507.
67. Cheng X, Zhang D, Cheng Z, Keller B, Ling H-Q. A New Family of Ty1-copia-Like Retrotransposons Originated in the Tomato Genome by a Recent Horizontal Transfer Event. Genetics. 2009;181:1183–93. https://doi.org/10.1534/genetics.108.099150.
68. Papolu PK, Ramakrishnan M, Mullasseri S, Kalendar R, Wei Q, Zou L-H, et al. Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress. Front Plant Sci. 2022;13. https://doi.org/10.3389/fpls.2022.1064847.
69. Wang Y, Ficklin SP, Wang X, Feltus FA, Paterson AH. Large-Scale Gene Relocations following an Ancient Genome Triplication Associated with the Diversification of Core Eudicots. PLoS ONE. 2016;11:e0155637. https://doi.org/10.1371/journal.pone.0155637.
70. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, et al. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 2019;20:38. https://doi.org/10.1186/s13059-019-1650-2.
71. Ren L-L, Liu Y-J, Liu H-J, Qian T-T, Qi L-W, Wang X-R, et al. Subcellular Relocalization and Positive Selection Play Key Roles in the Retention of Duplicate Genes of Populus Class III Peroxidase Family. Plant Cell. 2014;26:2404–19. https://doi.org/10.1105/tpc.114.124750.
72. Hanada K, Kuromori T, Myouga F, Toyoda T, Li W-H, Shinozaki K. Evolutionary Persistence of Functional Compensation by Duplicate Genes in Arabidopsis. Genome Biol Evol. 2009;1:409–14. https://doi.org/10.1093/gbe/evp043.
73. Russo A, Mayjonade B, Frei D, Potente G, Kellenberger RT, Frachon L, et al. Low-Input High-Molecular-Weight DNA Extraction for Long-Read Sequencing From Plants of Diverse Families. Front Plant Sci. 2022;13:883897. https://doi.org/10.3389/fpls.2022.883897.
74. Padmarasu S, Himmelbach A, Mascher M, Stein N. In Situ Hi-C for Plants: An Improved Method to Detect Long-Range Chromatin Interactions. In: Chekanova JA, Wang H-LV, editors. Plant Long Non-Coding RNAs: Methods and Protocols. New York, NY: Springer; 2019. pp. 441–72. https://doi.org/10.1007/978-1-4939-9045-0_28.
75. Hu J, Wang Z, Sun Z, Hu B, Ayoola AO, Liang F, et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. Genome Biol. 2024;25:107. https://doi.org/10.1186/s13059-024-03252-4.
76. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27:764–70. https://doi.org/10.1093/bioinformatics/btr011.
77. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33:2202–4. https://doi.org/10.1093/bioinformatics/btx153.
78. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.
79. Hu J, Fan J, Sun Z, Liu S. NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics. 2020;36:2253–5. https://doi.org/10.1093/bioinformatics/btz891.
80. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes De Novo Assembler. Curr Protocols Bioinf. 2020;70:e102. https://doi.org/10.1002/cpbi.102.
81. Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 2018;19:460. https://doi.org/10.1186/s12859-018-2485-7.
82. Zhang X, Zhang S, Zhao Q, Ming R, Tang H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. Nat Plants. 2019;5:833–45. https://doi.org/10.1038/s41477-019-0487-8.
83. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, et al. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 2019;20:224. https://doi.org/10.1186/s13059-019-1829-6.
84. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
85. Campbell MS, Law M, Holt C, Stein JC, Moghe GD, Hufnagel DE, et al. MAKER-P: A Tool Kit for the Rapid Creation, Management, and Quality Control of Plant Genome Annotations. Plant Physiol. 2014;164:513–24. https://doi.org/10.1104/pp.113.230144.

86. Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinformatics. 2015;16:227. https://doi.org/10.1186/s12859-015-0654-5.
87. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024;34:769–77. https://doi.org/10.1101/gr.278090.123.
88. Stiehler F, Steinborn M, Scholz S, Dey D, Weber APM, Denton AK. Helixer: cross-species gene annotation of large eukaryotic genomes using deep learning. Bioinformatics. 2021;36:5291–8. https://doi.org/10.1093/bioinformatics/btaa1044.
89. Li H. Protein-to-genome alignment with miniprot. Bioinformatics. 2023;39:btad014. https://doi.org/10.1093/bioinformatics/btad014.
90. Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. TSEBRA: transcript selector for BRAKER. BMC Bioinformatics. 2021;22:566. https://doi.org/10.1186/s12859-021-04482-0.
91. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512. https://doi.org/10.1038/nprot.2013.084.
92. Törönen P, Medlar A, Holm L. PANNZER2: a rapid functional annotation web server. Nucleic Acids Res. 2018;46:W84–8. https://doi.org/10.1093/nar/gky350.
93. Gilbert DG. Longest protein, longest transcript or most expression, for accurate gene reconstruction of transcriptomes? 2019:829184. https://doi.org/10.1101/829184
94. Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 2019;20:275. https://doi.org/10.1186/s13059-019-1905-y.
95. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci. 2020;117:9451–7. https://doi.org/10.1073/pnas.1921046117.
96. Tarailo-Graovac M, Chen N. Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. Curr Protocols Bioinf. 2009;25. 4.10.1–4.10.14.
97. Ou S, Jiang N, LTR_retriever:. A Highly Accurate and Sensitive Program for Identification of Long Terminal Repeat Retrotransposons. Plant Physiol. 2018;176:1410–22. https://doi.org/10.1104/pp.17.01310.
98. Neumann P, Novák P, Hoštáková N, Macas J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mob DNA. 2019;10:1. https://doi.org/10.1186/s13100-018-0144-1.
99. Krzywinski M, Schein J, Birol İ, Connors J, Gascoyne R, Horsman D, et al. Circos: An information aesthetic for comparative genomics. Genome Res. 2009;19:1639–45. https://doi.org/10.1101/gr.092759.109.
100. Wilkinson L, Biometrics. 2011;67:678–9. https://doi.org/10.1111/j.1541-0420.2011.01616.x
101. Moon K-W, ggiraphExtra. Make Interactive ggplot2. Extension to ggplot2 and ggiraph 2016:0.3.0. https://doi.org/10.32614/CRAN.package.ggiraphExtra
102. Zhou C, Brown M, Blaxter M, McCarthy SA, Durbin R, Darwin Tree of Life Project Consortium. Oatk: a de novo assembly tool for complex plant organelle genomes. Genome Biol. 2025;26:235. https://doi.org/10.1186/s13059-025-03676-6.
103. Shi L, Chen H, Jiang M, Wang L, Wu X, Huang L, et al. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 2019;47:W65–73. https://doi.org/10.1093/nar/gkz345.
104. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol Biol Evol. 2020;37:1530–4. https://doi.org/10.1093/molbev/msaa015.
105. Wong T, Ly-Trong N, Ren H, Baños H, Roger A, Susko E et al. IQ-TREE 3: Phylogenomic Inference Software using Complex Evolutionary Models 2025. https://doi.org/10.32942/X2P62N
106. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. Nucleic Acids Res. 2018;46:W200–4. https://doi.org/10.1093/nar/gky448.
107. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436.
108. Zhang C, Mirarab S. ASTRAL-Pro 2: ultrafast species tree reconstruction from multi-copy gene family trees. Bioinformatics. 2022;38:4949–50. https://doi.org/10.1093/bioinformatics/btac620.
109. Yang Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.
110. Álvarez-Carretero S, Goswami A, Yang Z, Dos Reis M. Bayesian Estimation of Species Divergence Times Using Correlated Quantitative Characters. Syst Biol. 2019;68:967–86. https://doi.org/10.1093/sysbio/syz015.
111. Puttick MN. MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics. 2019;35:5321–2. https://doi.org/10.1093/bioinformatics/btz554.
112. Revell LJ. phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). PeerJ. 2024;12:e16505. https://doi.org/10.7717/peerj.16505.
113. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y.
114. Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021;36:5516–8. https://doi.org/10.1093/bioinformatics/btaa1022.
115. Alexa A, Rahnenführer J, Lengauer T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. Bioinformatics. 2006;22:1600–7. https://doi.org/10.1093/bioinformatics/btl140.
116. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and Collinearity in Plant Genomes. Science. 2008;320:486–8. https://doi.org/10.1126/science.1153917.
117. Chen H, Zwaenepoel A, Van de Peer Y. wgd v2: a suite of tools to uncover and date ancient polyploidy and whole-genome duplication. Bioinformatics. 2024;40:btae272. https://doi.org/10.1093/bioinformatics/btae272.
118. Proost S, Fostier J, De Witte D, Dhoedt B, Demeester P, Van de Peer Y, et al. i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. Nucleic Acids Res. 2012;40:e11. https://doi.org/10.1093/nar/gkr955.
119. Almeida-Silva F, Van de Peer Y. Whole-genome Duplications and the Long-term Evolution of Gene Regulatory Networks in Angiosperms. Mol Biol Evol. 2023;40:msad141. https://doi.org/10.1093/molbev/msad141.
120. Anand L, Rodriguez Lopez CM. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics. 2022;23:33. https://doi.org/10.1186/s12859-021-04556-z.
121. Labella Ortega M. Uncovering the genome evolutionary dynamics of the Mediterranean endemic palm *Chamaerops humilis*. 2026. https://doi.org/10.5281/ZENODO.18671330

Publisher's note