

Repetita iuvant. The extraordinary evolutionary history of the endemic palm *Chamaerops humilis* L. in the Mediterranean

Mónica Labella-Ortega

`mlabellaortega@gmail.com`

University of Sannio <https://orcid.org/0000-0001-8485-3446>

Carmine Guarino

University of Sannio

Guido Cipriani

University of Udine

Rosario Schicchi

University of Palermo

Giuseppe Bazan

Università degli Studi di Palermo

Hengchi Chen

Ghent University

Maria Antonietta Ranauda

University of Sannio

Maria Maisto

University of Sannio

Daniela Zuzolo

University of Sannio <https://orcid.org/0000-0003-4376-9509>

Maria Tartaglia

University of Sannio

Antonello Prigioniero

Department of Science and Technology, University of Sannio <https://orcid.org/0000-0002-7801-7430>

Article

Keywords:

Posted Date: May 6th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-6595308/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Abstract

Chamaerops humilis L. is the only Arecaceae species endemic to the western Mediterranean. Here, de novo genome sequencing allowed discovering reasons behind its exclusive persistence within the natural distribution area, although the troubled environmental changes occurred. The genome, estimated at 3.44 Gbp, and its annotation enabled the identification of 60,555 genes and 2.87 Gbp of repetitive regions. Evolutionary analyses revealed a whole-genome duplication event 48.02 mya, shared with other palm species, and a later divergence from *Phoenix dactylifera* (17 mya). The characterization of gene duplication types, and the distribution of K_S values, uncovered multiple recent duplication waves, and the presence of *C. humilis*-specific duplicated genes involved in responses to stress. Furthermore, expansions of transposable elements were detected and appear to be associated with major paleoclimatic events in the Mediterranean. These genomic features likely contributed to the resilience and long-term survival of *C. humilis* within the Mediterranean.

1. Introduction

Analysis of the current global biodiversity landscape confronts us with the evidence that the sixth mass extinction is underway ¹. This conscious or unconscious awareness, coupled with technological advancement, is strongly directing the studies and project activities of research groups worldwide towards the acquisition of deep genomic data, through whole genome sequencing (WGS), in response to the need to crystallise the molecular basis of the evolution and adaptation of living species to environmental change ². The tools deemed most effective in achieving this aim are reference genomes and their systematisation in increasingly comprehensive databases useful for collecting and sharing information for biodiversity assessment, conservation strategies and restoration efforts ³. In the Tree of Life definition, a gap in sequenced plant genomes, which to date remains largely unexplored, is evident ². This shortcoming is probably rooted in the challenges posed by the high complexity of plant genomes sequencing and functional analysis, leading to gaps in research ^{2,4}. Due to their greater plasticity and speed of adaptation and evolution, plant genomes tend to be more diverse and larger than the rest of the eukaryotes ⁴. Polyploidy, whole genome duplications (WGD), or partial genome duplications, which occurred over the last 200 million years and are still occurring today, add further complexity to the panorama ^{5,6}. Increasing the amount of reference plant genomes is of fundamental importance for understanding the biological and evolutionary meanings - diverse and always new - defined by the WGD and polyploidy ⁶. Genome duplication has definitely led to the flowering of new evolutionarily beneficial traits to plant species or communities, influencing gene expression and function, and improving fitness in harsh environments ⁶. Ancestral events of genome duplication have led to changes in the functions of organisms, resulting in new arrangements that selective environmental pressure may have favoured in situations of extreme dynamic conditions ⁷. Genome duplication events can increase the vigour of individuals, giving rise to a more probable transgressive segregation, reducing the risk of extinction and increasing the rate of evolution and adaptation ⁵. Today, the Mediterranean basin is considered a major

biodiversity hotspot, considering the wonderful intra- and interspecific variety of populations that have adapted there ⁸. What we can observe and study today is the result of a series of upheavals in this relatively small portion of the globe over geological eras, shaping biodiversity through speciation, adaptation and extinction events caused by the sometimes dramatic fluctuations of the paleoclimatic features ⁹. Of all the events that occurred in the Mediterranean, the one that probably influenced the shaping of the biosphere, and the echoes of which still resonate in today's biodiversity, is the Miocene ¹⁰. The global changes caused by tectonic movement in the Cenozoic led to a rearrangement of ocean currents and a series of climate fluctuations culminating in the Miocene, during the course of which the Mediterranean basin underwent profound structural changes ¹¹. The Mediterranean Basin was to be considered part of the tropical biodiversity hotspot of the Western Tethys before its separation from it, and the Mid-Miocene Climate Optimum (MMCO) certainly represented a peak moment for tropical fauna and flora ^{10,12}. In the Late Miocene, the Mediterranean began the evolutionary process that would lead it to a temporary isolation from the Atlantic Ocean that would culminate in the salinity crisis during the Messinian ¹³. This sequence of events will lead the Mediterranean to a rapid desiccation that will only be resolved in the Zanclean when, as a result of the Zanclean Flood and the reopening of the Strait of Gibraltar, the sea level will rise again at significant rate, supposedly up to 10 meters per day ¹². The impacts of these geological changes, the relatively sudden change from a tropical to a cold climate, from a humid to a highly arid and saline environment, on the taxa that inhabited the waters and coastlines of the Mediterranean are still the subject of discussion and speculation, with evidence primarily drawn from Miocene fossil records ¹².

The vegetation of the Mediterranean was fully affected by the enormity of these changes and in the transition between the Miocene and Quaternary periods underwent a series of structural and compositional remodeling influenced by complex abiotic factors ¹⁴. Analysing the pollen remains and other biomarkers, it was possible to ascertain the transition from thermophilous and hygrophilous vegetation to sclerophyllous formations adapted to aridity. The transition defined a new vegetation landscape, moving from tropical and subtropical forest systems to shrublands and grasslands rich in sclerophyllous species that have marked and defined Mediterranean vegetation to date ¹⁵. The transition from tropical to current Mediterranean vegetation has left traces that can be seen in the almost total disappearance of some taxonomic groups from the Mediterranean basin, whose ancient presence has been confirmed by the discovery of pollen and macroremains ¹⁶. One of these ecological gaps is to be attributed to the Arecaceae family, dominant in the tropical and subtropical biome, present in the pre-Miocene Mediterranean and today represented only by one narrow endemic species: *Chamaerops humilis* L., the Mediterranean dwarf palm ¹⁷.

Chamaerops humilis is the only species in the *Chamaerops* L. genus, and is narrow endemic to the western Mediterranean area. In addition, *C. humilis* is a key indicator of thermo-Mediterranean vegetation, characterizing its early successional stages due to its adaptability. It is a species intimately connected with the history of the civilizations developed along the Mediterranean coasts, having

represented and still represents a crucial source of food, fibers, ethnobotanical products and being used for ornamental purposes ¹⁸.

The present work aims to reconstruct the evolutionary and adaptive history of *C. humilis* through sequencing of its genome, highlighting WGD events that occurred over time and studying genomic underpinnings that led this iconic species to diverge within the Arecaceae family and become the only endemic palm of the western Mediterranean. In a broader sense, we believe that the results of this work can form the basis for the identification and future definition of strategies for the conservation of endemic and endangered species, drawing on the information carved into the genome of the species themselves.

2. Results

2.1 Genome annotation

A total of 147,287 model genes were predicted, and 60,555 genes were annotated, corresponding to 92,564 proteins, representing 13.01% of the whole genome of *C. humilis* (Fig. 1). From all the genes identified, 92.46% were in chromosomes and 7.54% in chromosome 0. The chromosomes with the highest and lowest number of genes were chromosome 2 and chromosome 18, respectively (**Supplementary Table 1**). Genome completeness analysis assessed with BUSCO showed that 90.8% of the core reference genes were found in the genome and just 2.1% were missing.

From the GO analysis, we observed that the biological processes with the highest representation were DNA Integration, Protein phosphorylation and regulation of DNA-templated transcription. Molecular function was primarily characterized by zinc ion binding, RNA/DNA hybrid ribonuclease and nucleic acid binding; and in cellular processes, membrane, nucleus and cytoplasm were the most prevalent (**Supplementary Fig. 1**).

Transposable elements (TEs) were classified into two major classes according to the mechanism of transposition ¹⁹, Class I Retrotransposons which represent almost 78% of the genome, and Class II DNA transposons, which represent 5.3% of the genome (**Supplementary Table 2**). Of all, the most abundant superfamily was long terminal repeats (LTR) Copia and Gypsy, with 36.94% and 29.81% of the total genome length. The LTR Assembly Index (LAI) ²⁰, a metric to evaluate the quality and completeness of genome assembly based on LTR retrotransposons, was 19.86, indicating a high assembly quality through repetitive regions, being higher than in the other palms studied, indicating the high quality of the genome assembly. The TE content in palms varies, being *C. humilis* and *Cocos nucifera* the palms with the highest amount of TE (2.20 and 2.19 Gbp respectively) followed by *E. guineensis* (1.80 Gbp) and being *P. dactylifera* the palm with less TE (0.40 Gbp) (**Supplementary Table 3**). Insertion time estimation of LTR retrotransposon revealed an increase in the insertion rate of LTR/Copia around 5.5 million years ago (mya), followed by a gradual decrease and a slight increase of approximately 2 mya. In the case of

LTR/Gypsy, a steady increase in the LTR insertion was observed from 9 to 4.5 mya, which was then succeeded by a gradual decline (**Supplementary Fig. 2**).

Full-length LTR were classified into lineages, being in the case con Copia retrotransposon, the most abundant Tork, Angela and SIRE lineage (21.78%, 13.80% and 11.72% respectively) and in the case of Gypsy retrotransposon, the most abundant were CRM, Tat and Athila (16.63%, 13.91% and 6.91%, respectively) (**Supplementary Table 4**). Previously it has been seen this lineage also abundant in *P. dactylifera*, *C. nucifera*, *Calamus simplicifolius* and *Elaeis olifera*²¹.

2.2 Gene family evolution

All-against-all comparisons were performed to identify single-copy genes of *C. humilis* with orthologues in 6 monocotyledons (*Ananas comosus*, *C. nucifera*, *E. guineensis*, *Oryza sativa*, *P. dactylifera* and *Zea mays*) and 4 dicotyledons (*Arabidopsis thaliana*, *Coffea arabica*, *Theobroma cacao* and *Vitis vinifera*). The resulting phylogenetic tree (Fig. 2a) shows two main clades, monocotyledons and dicotyledons, that split ~ 150 mya (the late Jurassic), in agreement with prior findings²². Following monocot-dicot divergence, Poaceae (*Z. mays* and *O. sativa*) diverged from the rest of the clade ~ 120 mya²³. Subsequently, ~ 87 mya Bromeliaceae (*A. comosus*) separated from the Arecaceae followed by the separation of *C. humilis* and *P. dactylifera* in one clade ~ 23 mya and the final divergence of *C. humilis* species ~ 17 mya.

Given the close evolutionary relationship between *C. humilis* with *P. dactylifera* and *E. guineensis*, the evaluation of their synteny provided a framework to assess the extent of genomic conservation and the structural rearrangements that have occurred over time, revealing only small syntenic blocks among *C. humilis* and these species (Fig. 2b). The genome dynamic nature of *C. humilis*, reflected in the low number of conserved syntenic blocks, underscores the importance of further exploring the expansion and contraction of gene families as a potential driving force behind these changes, ultimately contributing to the monotypic nature of the genus. Collinearity study between Arecaceae was performed and the proportion of genes observed was 1:1, suggesting that they have experienced the same number of whole genome duplications.

Using all annotated protein-coding genes from 7 monocotyledons and 4 dicotyledons, orthofinder identified 32,135 total orthogroups and 9,243 orthogroups with all the species present. To evaluate the overall genomic changes throughout evolution in *C. humilis*, all gene families were analyzed. Among the findings, 1,008 gene families were shared between *C. humilis* and the other palms, while 3,135 gene families were unique (Fig. 2c). Of these unique families, 2663 appear to be expanded and 472 gene families contracted.

Subsequently, only rapidly evolving gene families were analysed to focus on genes that have undergone significant changes, potentially influenced by selective pressure. 942 gene families in *C. humilis* that have rapidly evolved in gene copy number since the last common ancestor were identified, reflecting a gain of 12,428 genes and a loss of 290 genes. Based on Gene Ontology enrichment analysis, of the most

enriched biological process terms are related to DNA plasticity processes like transposition, cellular cycle control, and interstrand cross-link repair, as well as terms related to stress response like cellular response to water deprivation and response to salt (**Supplementary Table 5**).

2.3 Whole genome duplication

An analysis of the size and frequency of paralogous gene families accumulated in the *C. humilis* genome provided a record of the whole genome duplication events that occurred in this species. Based on results obtained in other palm species, we expected to find a common whole genome duplication event that occurs in all the Aceraceae (39.87–55.2 mya) as previously reported ²⁴.

WGD events were analyzed by calculating the whole anchor K_S age distribution of *C. humilis*. The presence of some retained anchor pairs between 0 and 1 suggests a recent WGD event (**Supplementary Fig. 3**), with the 95% confidence interval of the log-normal distribution indicating a peak between 0.16 and 0.60 (**Supplementary Fig. 4A**). Further phylogenetic dating analysis estimated that this WGD event occurred 48.02 mya (**Supplementary Fig. 4B**).

In all palms, the same WGD event was observed ²⁴. However, since this WGD is common among other palm species, the increase in the genome size in *C. humilis* might be more influenced by other types of duplications. To further explore the dynamics of genome expansion, the gene duplication modes in two palm species were analyzed and compared with *C. humilis*, offering valuable insights into the genomic strategies that shape the diversity and adaptability of these plants.

Considering the complete gene set of *C. humilis*, only 13.48% (7,490 genes) were WGD derived genes, while in *P. dactylifera* and *E. guineensis* this percentage reached 52.64% and 52.03%, respectively (**Supplementary Fig. 5A**). Although the proportion of tandem (TD) (7.10%) and transposon-derived duplications (TRD) (28.05%) in *C. humilis* was similar to that observed in other palm species, the total number of genes in these duplication types is twice as high compared to the other species (**Supplementary Fig. 5A**). Additionally, the percentage of genes associated with proximal duplication (PD) (9.92%) is slightly higher than in the other palms. Notably, dispersed duplications (DD) represent the 41.44% of the total amount of duplicated genes in *C. humilis*, while this percentage is lower in *P. dactylifera* and *E. guineensis* (9.20% and 8.00%, respectively) (**Supplementary Fig. 5A**).

The K_S distribution was analyzed for each type of gene duplication. In the case of WGD genes, a shared peak at 0.27 was detected across the three palm species (**Supplementary Fig. 5B**), consistent with a previously reported duplication event dated to 39.87–55.2 mya. This peak is also observed in TRD, suggesting that this ancient event may have played a role in shaping the genomic architecture of these species. Although WGD genes were more abundant in the other palms, *C. humilis* showed the highest peak in TRD, highlighting the significant role of transposable elements in reshaping its genome, suggesting that transposition-driven duplication has been a major force in its genome evolution.

The major difference is also found in TD, PD, TRD and DD duplications. In *C. humilis*, all these duplication types exhibit a unique recent peak at 0.05, a pattern not observed in the other palm species (**Supplementary Fig. 5B**). This exclusive signal may be linked to species-specific genome expansion and could partially explain the larger genome size observed in *C. humilis*.

Gene homology matrix dot plot shows the collinearity between chromosomes, where a remnant of gene duplications experienced in *C. humilis* can be seen in all the chromosomes (Fig. 3a). Most of the dots show a K_S value lower than 2, suggesting that many of the duplications are relatively recent, providing evidence of the lasting impact of these duplication events that continue to influence in the genomic landscape of this species.

An intraspecific collinearity plot was generated to illustrate the distribution of varying levels of collinearity in *C. humilis* (Fig. 3b), revealing a significant number of multiplicons (homeologous regions within a genome that are present in multiple copies), with some regions exhibiting up to nine copies. This underscores the extensive genomic duplication in *C. humilis*, especially when compared to other palms like *E. guineensis* and *P. dactylifera*, which shows no more than four-fold duplications (**Supplementary Fig. 6**). A comparison between multiplicons observed in the three palms was performed, and 729 multiplicons were found exclusively duplicated in *C. humilis*, being 17 of them duplicated at least seven times within *C. humilis* genome.

Multiplicons duplicated at least seven-fold times include a total of 184 genes, which are distributed across all chromosomes except for chromosomes 11, 16, and 18 (**Supplementary Fig. 7**). Notably, the multiplicons exhibiting up to nine-fold duplication are localized in chromosomes 1, 2, 5, 9, 10, 13, 14, and 15, with the highest concentration observed in chromosome 1 (9 genes) (**Supplementary Fig. 7**). In all the cases dispersed duplications were the most abundant (58%), followed by transposon derived duplication (24.6%) and whole genome duplication derived (11.2%).

From all the genes identified, 74.5% were coding genes and 25.5% were repetitive elements (Fig. 3c). Several proteins were identified that play essential roles in biotic and abiotic stress responses, particularly to cold and salinity. Among these, aspartic peptidases, such as DDI1-type, are linked to low temperatures²⁵, while other members of the aspartic peptidase family are involved in pathogen defense and regulation of responses to both biotic and abiotic stresses (e.g., drought, salinity, and cold)²⁶. Additionally, GTP-binding nuclear proteins and RING-type E3 ligases, which are integral to cellular signalling and regulatory processes during stress responses, also play a crucial role in responses to stresses like drought, salinity, and low temperatures²⁷. Another significant group includes NAC-transcription factors, such as NST1, which are involved in secondary cell wall formation and responses to abiotic stresses like cold and salinity²⁸. Other proteins, such as SUT1, regulate immune responses²⁹, and plant LRR proteins are involved in pathogen defence³⁰, contributing to the overall stress response mechanism. These duplications suggest that the plant has evolved enhanced adaptability through a more redundant and efficient stress response system.

Moreover, retrotransposons, including Gypsy and Copia³¹, as well as Ty3-gypsy retrotransposon³², are known to be sensitive to various abiotic stresses such as cold, heat, salinity, and infections. Their duplication could contribute to genomic plasticity, providing the plant with the flexibility needed to adjust to fluctuating environmental conditions.

3. Discussion

Across palm family, there had been found 58-fold change range of variation in genome size, from 0.53 Gbp in the diploid *Licuala orbicularis* and *L. sarawakensis*, to the 38-ploid *Voanioala gerardii* with a genome size of 30.63 Gbp³³. The sequenced genome of *C. humilis* in this study, falls within the average size range of known palms genomes³³. In this study we compare *C. humilis* with *P. dactylifera* and *E. guineensis*^{34,35}, as they are the phylogenetic closest relatives. Additionally, *P. dactylifera* ancestors share a historical presence in Southern Europe environments, making it a particularly relevant species for understanding the evolutionary adaptations of *C. humilis* in this region³⁶.

The genome size variation among *C. humilis* with *P. dactylifera* and *E. guineensis* palms can provide insight into how genomic architecture evolves in response to environmental pressures, leading to genome expansion. In plants, this increase in genome size is primarily driven by polyploidization events and the accumulation of substantial numbers of repetitive elements³⁴. The large genome size in *C. humilis* could represent an evolutionary advantage acquired after its divergence from *P. dactylifera*.

Intense environmental fluctuations, including significant climate changes such as glaciation events, rapid shifts in sea surface temperatures and level, and a sharp drop in global mean surface temperature occurred between 25 and 40 mya³⁷, are linked to a wave of paleopolyploidizations²⁴. The WGD event observed in *C. humilis* (48.02 mya) corresponds to the one detected in other palms, indicating that it predates their phylogenetic divergence²⁴.

After WGD events, rediploidization restores diploid-like inheritance through a combination of genomic reorganization and selective gene retention³⁸. In *C. humilis*, after the WGD event, phylogenetic differentiation allowed the identification of a set of expanded gene families. From all of them, the presence of terms related to DNA regulation, such as transposition, as well as terms related to environmental stress, suggests a connection between genomic plasticity and adaptation to environmental changes³⁹. The activation of TE in response to environmental stress generates a high genetic variability⁴⁰ that must be controlled, justifying the presence of terms associated with DNA repair. Terms related to response to environmental stress indicate that the availability of water, salt and hormones like brassinosteroids, which have been linked to response to various stresses⁴¹, play a significant role. These observations reinforce the idea that genomic plasticity is a dynamic and regulated response that allows adaptation to adverse conditions⁴².

When the *C. humilis* genome is compared to other palms phylogenetically closed, low synteny is observed. This statement has been seen previously in other monotypic genera, such as *Ginkgo biloba*⁴³ and *Amborella trichopoda*⁴⁴, which has low synteny and significant genomic rearrangements, that could be linked to long term isolation and adaptation to specific ecological niches.

Additionally, certain gene duplications appear to be unique to *C. humilis*. These duplications may have played a crucial role in the species' adaptation to environments with both biotic and abiotic stresses, aligning with its status as one of the most cold-tolerant palms in the Northern Hemisphere⁴⁵. Similar patterns have been observed in other species, where duplicated genes, particularly tandem duplications, contribute to stress response, similarly to *Arabidopsis* and *Oryza*, where they enhance biotic stress resilience⁴. Dispersed duplications were the most abundant type of gene duplication in *C. humilis*. These duplications may result from larger-scale genomic rearrangements, such as duplications, inversions, and transpositions, which can drive gene family's expansion, functional diversification, and increased genomic flexibility, potentially linked to evolutionary adaptations⁴⁶. This suggests that the duplications in *C. humilis* may have similarly contributed to its enhanced resilience to environmental stresses.

The accumulation of repetitive elements and TE expansions can drive rapid genome size increases, either independently or in concert with WGD⁴⁷. A striking example is maize, where post-WGD TE expansion has led to a genome composed of 85% transposable elements⁴⁸. Similarly, *C. humilis* exhibits a genome with 82% TE content, suggesting a major role for TEs in its genomic architecture. Previous studies in palms have shown that repeat expansions can be influenced by environmental factors such as aridity³³. The expansion and regulation of TEs are tightly controlled through mechanisms that suppress their activity and mitigate their effects when necessary. However, under environmental stress or genomic instability, TEs may escape these controls, leading to bursts of amplification⁴⁹. Events such as polyploidy and hybridization can further weaken TE suppression, facilitating their proliferation⁵⁰. While TE insertions are often considered deleterious, they can also contribute to adaptation by generating genetic diversity and regulatory innovations.

The observed fluctuation in Copia and Gypsy transposon activity align with major geological and climatic events in the Mediterranean region, particularly the Messinian Salinity Crisis and the subsequent Zanclean flood. Between 5.33 and 5.97 million years ago, the Mediterranean Sea experienced a dramatic isolation from the Atlantic ocean during the Messinian Salinity Crisis, an extraordinary geological event that triggered a precipitous drop in sea levels and led to the extensive deposition of vast salt layers across the basin⁵¹. Such stressors may disrupt *C. humilis* genome, destabilizing epigenetic controls and enabling transposon amplification and insertion⁵². The progressive decline of LTR insertions fit with the end of the Messinian salinity crisis, marked by the cataclysmic Zanclean megaflood⁵³. This event imposed strong selective pressures in the Mediterranean, favouring genome expansion mechanisms that may have conferred adaptive advantages, such as the amplification of stress-response genetic elements.

More recently, during the Quaternary period, alternating glacial and interglacial cycles led to repeated phases of species contraction and expansion, favouring the extinction of those less adapted to extreme environmental changes⁵⁴. Before the onset of the Quaternary, the Mediterranean flora still contained a considerable number of tropical and subtropical tree taxa⁵⁵. In the case of *C. humilis*, its present-day distribution could be linked to its ability to persist in climatic refugia during these periods⁵⁶.

Chamaerops humilis exemplifies a species that, despite these fluctuations, has persisted and adapted. Fossil evidence from Santorini (Greece) suggests that it was present in the region during the Late Pleistocene but later disappeared from that area, reflecting a contraction in its range⁵⁷. However, rather than facing extinction, *C. humilis* has survived across fragmented Mediterranean habitats, demonstrating its resilience. Despite the tumultuous geoclimatic evolution of the Mediterranean, which has fluctuated between warmer, subtropical conditions, arid and cooler climate, the species has managed to persist. Its genome is a testament to the species' ability to withstand both environmental stressors and historical climate changes.

4. Materials and methods

4.1. Phylogenetic and Genomic Comparative Analysis

To investigate the evolutionary history of *C. humilis*, gene family construction was performed using OrthoFinder (v2.5.5)⁵⁸. The analysis included additional palms, including palm species *P. dactylifera*⁵⁹, *E. guineensis*⁶⁰, and *C. nucifera*⁶¹, as well as other monocotyledons including *Ananas comosus*⁶², *O. sativa*⁶³, and *Z. mays*⁶⁴, and some dicotyledons plants as outgroups including *A. thaliana*, *C. arabica*⁶⁵, *V. vinifera*⁶⁶ and *T. cacao*⁶⁷. The protein sequences of 70 single-copy orthologs were concatenated and aligned using MAFFT (Katoh et al., 2002) with the L-INS-i iterative refinement method. The most suitable model (JTT + G + F) was identified with ModelGenerator v0.85 (<http://bioinf.nuim.ie/software/modelgenerator/>). Divergence times between species were estimated with BEAST v2.6.3⁶⁸, utilizing the JTT substitution model, four gamma categories, the Yule model, and an uncorrelated relaxed clock with lognormal distribution. The MCMC chain was run for 10,000,000 generations, recording parameters every 1000 generations. Calibration points for BEAST were obtained from the TimeTree database (<http://www.timetree.org/>), *C. nucifera* versus *V. vinifera* (145 mya), *O. sativa* versus *Z. mays* (60 mya), *P. dactylifera* versus *C. nucifera* (70 mya) and *A. thaliana* versus *C. arabica* (90 mya). The contraction and expansion of the gene families were observed by comparing the differences in cluster size between *C. humilis* and each species using CAFE (v.5.1.0)⁶⁹. Contracted and expanded gene families were analyzed, and GO enrichment was conducted using topGO⁷⁰. iTOLs⁷¹ was used to visualize the species tree and annotations. Synteny analysis was performed between the two reference Arecaceae genomes: *P. dactylifera* and *E. guineensis* using JCVI (v 1.4.25)⁷².

4.2 Whole genome duplication (WGD) analysis

WGD analysis was conducted to determine the presence of lineage-specific and/or shared WD events among Aracaceae, using wgd (v2.0.38)⁷³.

The whole paranome (complete set of paralogous genes within a genome) was constructed using 'wgd dmd' with default parameters. To infer genomic collinearity, 'wgd syn' was applied using i-ADHoRe (v3.0.01)⁷⁴. 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)⁷³. 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, facilitating the effective identification of potential WGD events. With the K_S distribution of the anchor pairs, only those within the 95% of confidence level were retained using 'wgd peak'. Finally, using a phylogenetic tree and fossil calibration of 17 related species previously reported by Chen et al⁷³, 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)⁷⁵.

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

Duplicated gene pairs were identified and classified into WGD-derived and SSD-derived gene pairs using the R package doubletrouble⁷⁶. Paralog genes that are used as anchor pairs in syntenic regions are classified as SD, 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 termed TD; those separated by only a few genes (default is 10, adjustable) are classified as PD; and all others fall under DD. Additionally, duplicates that arise from transposon-derived duplications are known as TRD. The location of duplicated genes and classification were plotted using chromomap in R⁷⁷.

Declarations

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.

Author contributions

G. C., Z. D., T. M., P. A conceived, write and revised the article, L-O. M. performed bioinformatic analysis, write and revised the article, T.M., R. M-A and M.M extract nucleic acid material, C. G., R. revised the

article, S. R. and B. G provided biological material, C. H contributed with bioinformatic analysis and revised the manuscript.

Data availability

The genome sequence and annotation data for *C. humilis*, the sequencing data and the transcriptome data are accessible under NCBI BioProject no. PRJNA1247256.

References

1. Theissinger, K. *et al.* How genomics can help biodiversity conservation. *Trends Genet.* **39**, 545–559 (2023).
2. Twyford, A. D. The road to 10,000 plant genomes. *Nat. Plants* **4**, 312–313 (2018).
3. Mc Cartney, A. M. *et al.* The European Reference Genome Atlas: piloting a decentralised approach to equitable biodiversity genomics. *Npj Biodivers.* **3**, 1–17 (2024).
4. Panchy, N., Lehti-Shiu, M. & Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiol.* **171**, 2294–2316 (2016).
5. Van de Peer, Y., Maere, S. & Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **10**, 725–732 (2009).
6. Soltis, P. S., Marchant, D. B., Van de Peer, Y. & Soltis, D. E. Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* **35**, 119–125 (2015).
7. Bowers, J. E., Chapman, B. A., Rong, J. & Paterson, A. H. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* **422**, 433–438 (2003).
8. Aurelle, D. *et al.* Biodiversity, climate change, and adaptation in the Mediterranean. *Ecosphere* **13**, e3915 (2022).
9. Svenning, J.-C., Eiserhardt, W. L., Normand, S., Ordonez, A. & Sandel, B. The Influence of Paleoclimate on Present-Day Patterns in Biodiversity and Ecosystems. *Annu. Rev. Ecol. Evol. Syst.* **46**, 551–572 (2015).
10. Cornacchia, I., Brandano, M. & Agostini, S. Miocene paleoceanographic evolution of the Mediterranean area and carbonate production changes: A review. *Earth-Sci. Rev.* **221**, 103785 (2021).
11. Zachos, J., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, Rhythms, and Aberrations in Global Climate 65 Ma to Present. *Science* **292**, 686–693 (2001).
12. Agiadi, K. *et al.* Late Miocene transformation of Mediterranean Sea biodiversity. *Sci. Adv.* **10**, eadp1134 (2024).
13. Krijgsman, W., Hilgen, F. J., Raffi, I., Sierro, F. J. & Wilson, D. S. Chronology, causes and progression of the Messinian salinity crisis. *Nature* **400**, 652–655 (1999).

14. 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.* **162**, 403–415 (2010).
15. Salocchi, A. C. *et al.* Biomarker constraints on Mediterranean climate and ecosystem transitions during the Early-Middle Miocene. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **562**, 110092 (2021).
16. Jiménez-Moreno, G. & Suc, J.-P. Middle Miocene latitudinal climatic gradient in Western Europe: Evidence from pollen records. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **253**, 208–225 (2007).
17. Yao, G. *et al.* A plastid phylogenomic framework for the palm family (Arecaceae). *BMC Biol.* **21**, 50 (2023).
18. Guzmán, B., Fedriani, J. M., Delibes, M. & Vargas, P. The colonization history of the Mediterranean dwarf palm (*Chamaerops humilis* L., Palmae). *Tree Genet. Genomes* **13**, 24 (2017).
19. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982 (2007).
20. Ou, S., Chen, J. & Jiang, N. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126–e126 (2018).
21. Ibrahim, M. A. *et al.* Genome-wide comparative analysis of transposable elements in Palmae genomes. *Front. Biosci.-Landmark* **26**, 1119–1131 (2021).
22. 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.* **58**, 424–441 (2004).
23. Gallaher, T. J. *et al.* Grasses through space and time: An overview of the biogeographical and macroevolutionary history of Poaceae. *J. Syst. Evol.* **60**, 522–569 (2022).
24. Chen, H., Almeida-Silva, F., Logghe, G., Bonte, D. & Peer, Y. V. de. The Rise of Polyploids During Environmental Catastrophes. 2024.11.22.624806 Preprint at <https://doi.org/10.1101/2024.11.22.624806> (2024).
25. Wang, B., Wang, G. & Zhu, S. DNA Damage Inducible Protein 1 is Involved in Cold Adaption of Harvested Cucumber Fruit. *Front. Plant Sci.* **10**, (2020).
26. Figueiredo, L., Santos, R. B. & Figueiredo, A. Defense and Offense Strategies: The Role of Aspartic Proteases in Plant–Pathogen Interactions. *Biology* **10**, 75 (2021).
27. Shu, K. & Yang, W. E3 Ubiquitin Ligases: Ubiquitous Actors in Plant Development and Abiotic Stress Responses. *Plant Cell Physiol.* **58**, 1461–1476 (2017).
28. 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* **17**, 2993–3006 (2005).
29. Yan, 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.* **100**, 357–373 (2019).

30. Afzal, A. J., Wood, A. J. & Lightfoot, D. A. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol. Plant-Microbe Interact. MPMI* **21**, 507–517 (2008).
31. Papolu, P. K. *et al.* Retrotransposons: How the continuous evolutionary front shapes plant genomes for response to heat stress. *Front. Plant Sci.* **13**, (2022).
32. 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* **181**, 1183–1193 (2009).
33. Schley, R. J. *et al.* The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. *New Phytol.* **236**, 433–446 (2022).
34. He, B. *et al.* Evolution of Plant Genome Size and Composition. *Genomics Proteomics Bioinformatics* **22**, qzae078 (2024).
35. Pellicer, J., Hidalgo, O., Dodsworth, S. & Leitch, I. J. Genome Size Diversity and Its Impact on the Evolution of Land Plants. *Genes* **9**, 88 (2018).
36. Pereira, D. S. & Phillips, A. J. L. Exploring the Diversity and Ecological Dynamics of Palm Leaf Spotting Fungi—A Case Study on Ornamental Palms in Portugal. *J. Fungi* **11**, 43 (2025).
37. Cather, S. M., Dunbar, N. W., McDowell, F. W., McIntosh, W. C. & Scholle, P. A. Climate forcing by iron fertilization from repeated ignimbrite eruptions: The icehouse–silicic large igneous province (SLIP) hypothesis. *Geosphere* **5**, 315–324 (2009).
38. Wolfe, K. H. Yesterday's polyploids and the mystery of diploidization. *Nat. Rev. Genet.* **2**, 333–341 (2001).
39. Magadum, S., Banerjee, U., Murugan, P., Gangapur, D. & Ravikesavan, R. Gene duplication as a major force in evolution. *J. Genet.* **92**, 155–161 (2013).
40. Casacuberta, E. & González, J. The impact of transposable elements in environmental adaptation. *Mol. Ecol.* **22**, 1503–1517 (2013).
41. Miao, R. *et al.* The Role of Endogenous Brassinosteroids in the Mechanisms Regulating Plant Reactions to Various Abiotic Stresses. *Agronomy* **14**, 356 (2024).
42. Napier, J. D., Heckman, R. W. & Juenger, T. E. Gene-by-environment interactions in plants: Molecular mechanisms, environmental drivers, and adaptive plasticity. *Plant Cell* **35**, 109–124 (2023).
43. Guan, R. *et al.* Draft genome of the living fossil *Ginkgo biloba*. *GigaScience* **5**, s13742-016-0154–1 (2016).
44. Amborella Genome Project *et al.* The Amborella Genome and the Evolution of Flowering Plants. *Science* **342**, 1241089 (2013).
45. Reichgelt, T., West, C. K. & Greenwood, D. R. The relation between global palm distribution and climate. *Sci. Rep.* **8**, 4721 (2018).
46. Lallemand, T., Leduc, M., Landès, C., Rizzon, C. & Lerat, E. An Overview of Duplicated Gene Detection Methods: Why the Duplication Mechanism Has to Be Accounted for in Their Choice. *Genes* **11**, 1046 (2020).

47. Bennetzen, J. L. & Wang, H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.* **65**, 505–530 (2014).
48. Schnable, P. S. *et al.* The B73 maize genome: complexity, diversity, and dynamics. *Science* **326**, 1112–1115 (2009).
49. Pulido, M. & Casacuberta, J. M. Transposable element evolution in plant genome ecosystems. *Curr. Opin. Plant Biol.* **75**, 102418 (2023).
50. Oliver, K. R. & Greene, W. K. Transposable elements: powerful facilitators of evolution. *BioEssays* **31**, 703–714 (2009).
51. Krijgsman, W. *et al.* Causes and consequences of the Messinian salinity crisis. *Nat. Rev. Earth Environ.* **5**, 335–350 (2024).
52. Zhou, W., Liang, G., Molloy, P. L. & Jones, P. A. DNA methylation enables transposable element-driven genome expansion. *Proc. Natl. Acad. Sci.* **117**, 19359–19366 (2020).
53. Garcia-Castellanos, D. *et al.* The Zanclean megaflood of the Mediterranean – Searching for independent evidence. *Earth-Sci. Rev.* **201**, 103061 (2020).
54. Svenning, J.-C. Deterministic Plio-Pleistocene extinctions in the European cool-temperate tree flora. *Ecol. Lett.* **6**, 646–653 (2003).
55. Sadori, L. *et al.* Palynology and Mediterranean vegetation history. *Flora Mediterr.* **23**, 156 (2013).
56. Willis, K. J. & Whittaker, R. J. The Refugial Debate. *Science* **287**, 1406–1407 (2000).
57. Velitzelos, D., Bouchal, J. M. & Denk, T. Review of the Cenozoic floras and vegetation of Greece. *Rev. Palaeobot. Palynol.* **204**, 56–117 (2014).
58. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
59. Hazzouri, K. M. *et al.* Genome-wide association mapping of date palm fruit traits. *Nat. Commun.* **10**, 4680 (2019).
60. Singh, R. *et al.* Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. *Nature* **500**, 335–339 (2013).
61. Xiao, Y. *et al.* The genome draft of coconut (*Cocos nucifera*). *GigaScience* **6**, gix095 (2017).
62. Ming, R. *et al.* The pineapple genome and the evolution of CAM photosynthesis. *Nat. Genet.* **47**, 1435–1442 (2015).
63. Shang, L. *et al.* A complete assembly of the rice Nipponbare reference genome. *Mol. Plant* **16**, 1232–1236 (2023).
64. Hufford, M. B. *et al.* De novo assembly, annotation, and comparative analysis of 26 diverse maize genomes. *Science* **373**, 655–662 (2021).
65. Medrano, J. F. *et al.* De novo whole-genome assembly and annotation of *Coffea arabica* var. *Geisha*, a high-quality coffee variety from the primary origin of coffee. *G3 Bethesda Md* **15**, jkae262 (2025).
66. Shi, X. *et al.* The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Hortic. Res.* **10**, uhad061 (2023).

67. Argout, X. *et al.* The cacao Criollo genome v2.0: an improved version of the genome for genetic and functional genomic studies. *BMC Genomics* **18**, 730 (2017).
68. Bouckaert, R. *et al.* BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol.* **15**, e1006650 (2019).
69. Mendes, F. K., Vanderpool, D., Fulton, B. & Hahn, M. W. CAFE 5 models variation in evolutionary rates among gene families. *Bioinformatics* **36**, 5516–5518 (2021).
70. Adrian Alexa, J. R. topGO. Bioconductor <https://doi.org/10.18129/B9.BIOC.TOPGO> (2017).
71. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**, W293–W296 (2021).
72. Tang, H. *et al.* Synteny and Collinearity in Plant Genomes. *Science* **320**, 486–488 (2008).
73. 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* **40**, btae272 (2024).
74. Proost, S. *et al.* i-ADHoRe 3.0—fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **40**, e11 (2012).
75. Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
76. Almeida-Silva, F. & Van de Peer, Y. doubletrouble: an R/Bioconductor package for the identification, classification, and analysis of gene and genome duplications. *Bioinformatics* **41**, btaf043 (2025).
77. Anand, L. & Rodriguez Lopez, C. M. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC Bioinformatics* **23**, 33 (2022).

Figures

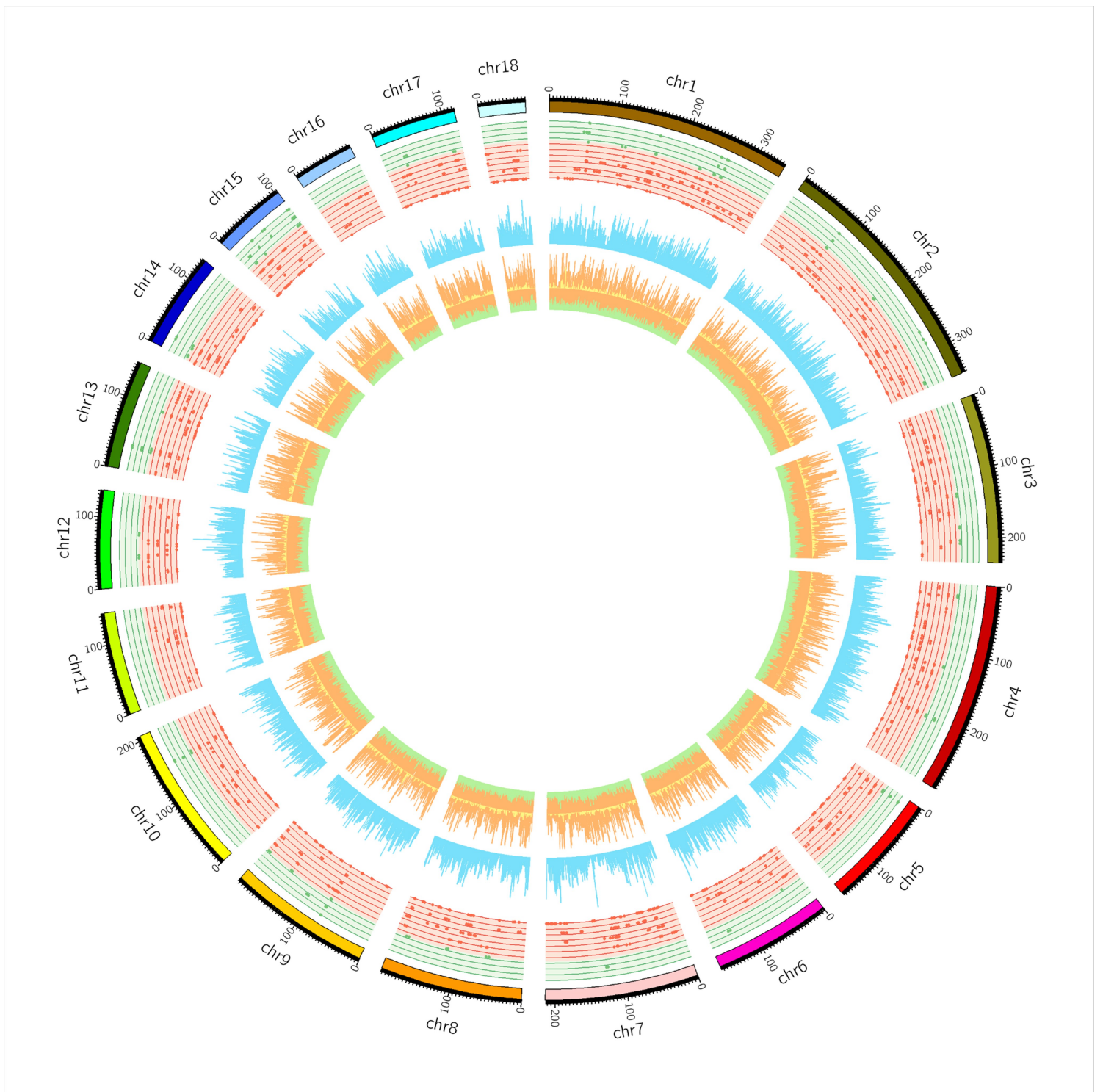


Figure 1

Chromosomal features, functional annotation, and gene duplication landscape of *Chamaerops humilis*. The figure illustrates key chromosomal characteristics and duplication patterns in *C. humilis* in the 18 principal chromosomes. From the outer to the inner rings, the tracks represent: **(a)** chromosome size in megabases (Mb); **(b)** positions of duplicated genes, with green indicating genes duplicated 7 to 9 times and red indicating genes duplicated fewer than 6 times; **(c)** gene density, shown in blue, calculated as the number of annotated genes per 10-Mb bin; and **(d)** transposable element (TE) density per 10-Mb bin,

with orange representing the overall density and further differentiation of LTR types: green for Gypsy and yellow for Copia.

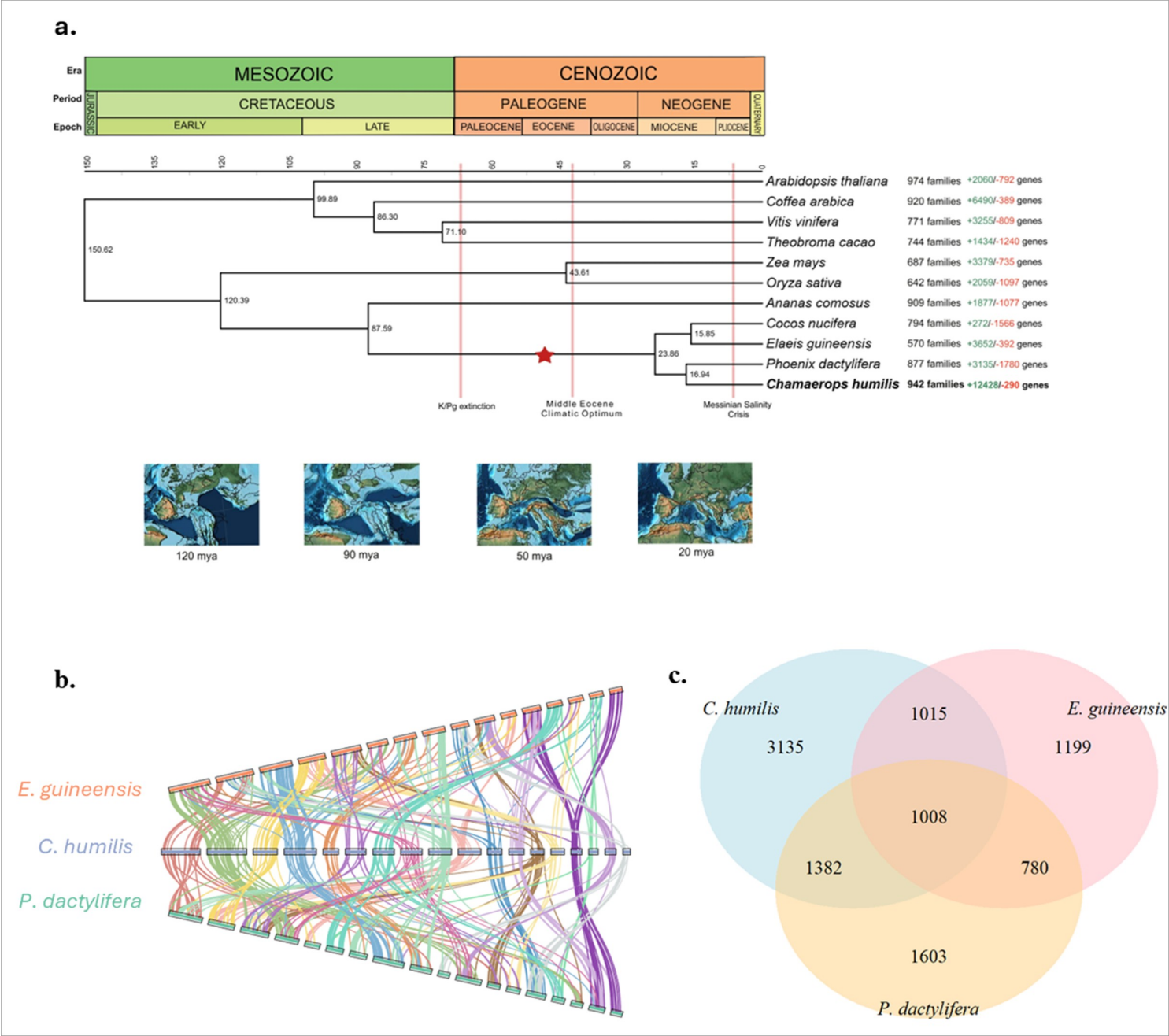


Figure 2

Genome evolution of *C. humilis*. **(a)** Phylogenetic tree of monocotyledon and dicotyledon species showing their divergence times in million years, calculated based on orthologs within single gene families. The star marks a genome duplication event in the *C. humilis* lineage. Total CAFE detected gene families with significant accelerated gene changes of each species; green and red numbers indicate the number of gene expansion or contraction, respectively. At the bottom of the phylogenetic tree, the geological evolution of the *C. humilis* current natural range of distribution, coinciding with the actual western Mediterranean, is represented. **(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*. Each number represents the number of gene families shared within the corresponding intersections, while exclusive numbers indicate families unique to a specific species.

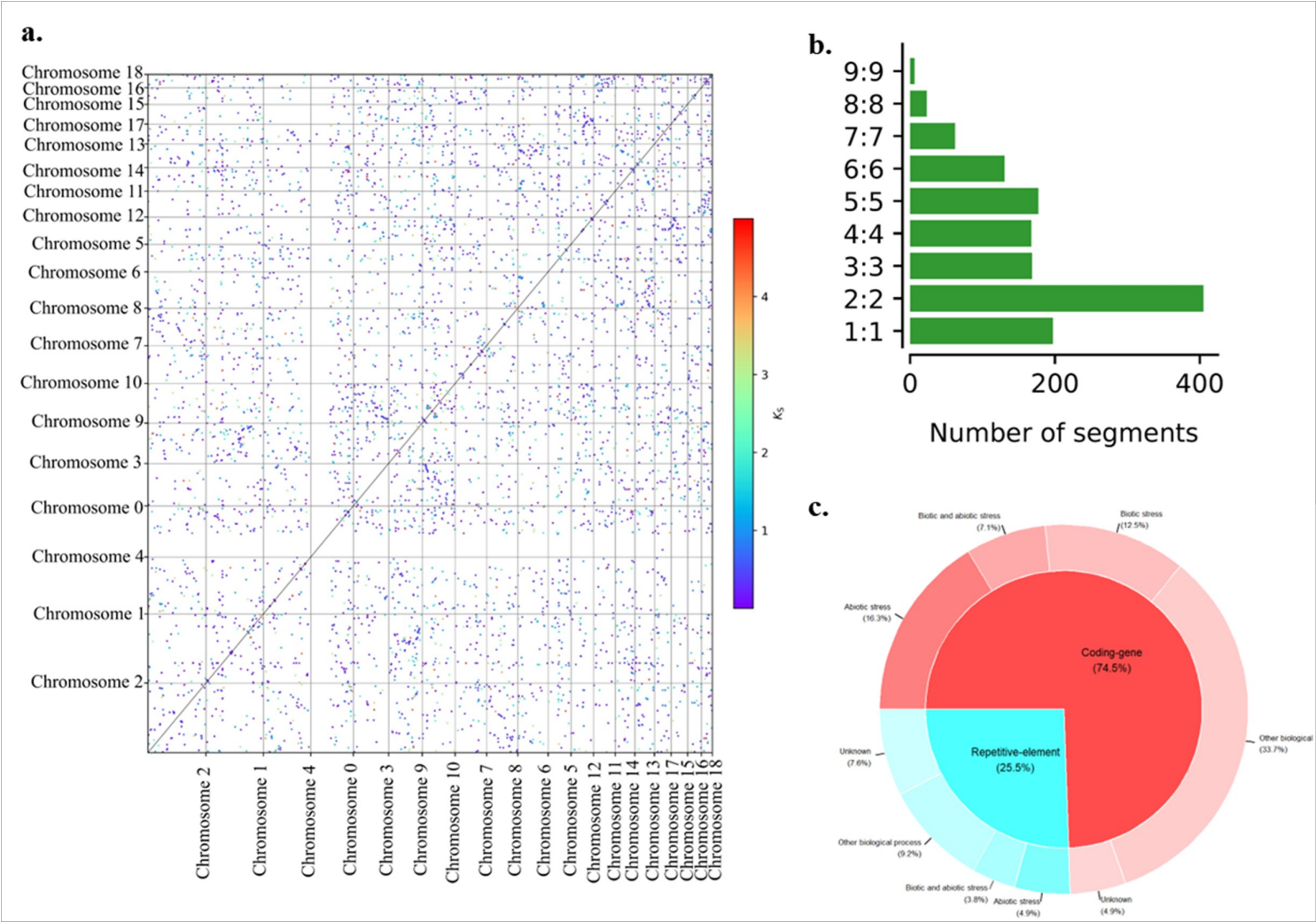


Figure 3

Genomic Duplication Profile and Functional Insights in *Chamaerops 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 100kb. **(c)** Donut plot of the main functions of genes and repetitive elements duplicated exclusively in *Chamaerops humilis* more than 7 times.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [RepetitaiuvantSupplementaryTables.xlsx](#)

- [RepetitaiuvantSupplementaryMaterial.pdf](#)