UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Scienze Naturali

ELABORATO DI LAUREA

Chromosome-scale genome assembly of the Goethe’s palm (*Chamaerops humilis* L.)

Tutor: Professor Francesco Dal Grande
Dipartimento di Biologia

Co-tutor: Núria Beltrán-Sanz
Centro di Ateneo Orto Botanico Università di Padova

Laureanda: Alice Bordignon
Matricola 2000442

ANNO ACCADEMICO 2022/2023

ABSTRACT

Chamaerops humilis L., also known as dwarf palm, is a species of palm belonging to the *Arecaeae*. A specimen of this species - the so-called Goethe's palm - is kept at the Botanical Garden of Padua, famous for having inspired the German poet Johann Wolfgang von Goethe in drafting his treatise "Essay on the metamorphoses of plants", published in 1890. *Chamaerops humilis* is the only species of *Arecaeae* to have survived the glaciations that affected Europe 12,000 years ago, as well as the only one able to effectively colonize the northernmost latitudes. In this project we have obtained an almost chromosomal-level assembly of *C. humilis* genome. This information will allow to investigate not only the evolutionary origins of this plant and its phylogenetic relationships, but also the genetics underlying the mechanisms that allowed it to expand towards higher latitudes. The Hi-C contact map will provide fundamental information about genome architecture, three-dimensional chromatin organization and gene interactions. Moreover, being a species that is also of considerable interest in the biomedical field, the information stored in its genome can be investigated to explore the potential of this plant in the treatment of cardiovascular and neurodegenerative diseases.

TABLE OF CONTENTS

1. Introduction

1.1 Genome references and biodiversity preservation

Biodiversity is the variety of plant and animal life in the world or in a particular habitat: a high level of biodiversity is important and desirable as it promotes the stability of the ecosystem of reference (*Biodiversity Study | Dartmoor*, no date). It can be effectively characterized with genomics tools, but their full implementation in conservation practices is still restricted. During the last decades numerous projects aimed at the sequencing of reference genomes, i.e. a digital database of nucleotide sequences representative of the set of genes that characterize the ideal organisms of a species, or a specific representation of the organization of the genome of a species or a population, have been initiated, such as the Human genome project, an international project that aims at sequencing all the nitrogenous bases present in the human genome and at mapping the genes present, identifying their physical position and functional role (*Human Genome Project Fact Sheet*, no date). This project began in October 1990 and concluded in April 2003 with a genome sequence equal to 90% of the total genome (*Human Genome Project Fact Sheet*, no date).

Reference genomes are essential for characterizing the genetic information of a species in a specific and unequivocal way. Therefore, it is essential for them to be complete, correct and with a high degree of chromosomal sequences. Such complete, high-quality genomes provide an overview of the three-dimensional structure of the genome, including duplicate regions, centromeres, and telomeres (Formenti *et al.*, 2022). Hence, reference genomes play a fundamental role in the reconstruction of the Tree of Life, of the phylogenetic relationships between organisms and in the conception of global biodiversity conservation initiatives, since they facilitate research and conservation precisely through the tree of life, therefore they pose as an essential best practice for conservation genomics (Theissinger *et al.*, 2023).

The history of Earth's biodiversity has been characterized by five major mass extinctions, all of which coincided with catastrophic natural events (Cowie, Bouchet and Fontaine, 2022). Many scientists agree that today we are in the midst of the sixth great mass extinction, entirely due to the human impact on natural ecosystems (Cowie, Bouchet and Fontaine, 2022). In 2020 both the European Environment Agency (EEA) and the United Nations Biodiversity Summit drew attention to the alarming speed of biodiversity loss (Formenti *et al.*, 2022) and it is estimated that over 42,100 species are currently at risk of extinction (*The IUCN Red List of Threatened Species*, no date). The human population, with its exponential growth in the last century and the overexploitation of resources, is jeopardizing the biological variety of living beings without taking steps to reverse this trend (Cowie, Bouchet and Fontaine, 2022). The loss of biodiversity is accelerated not only by anthropic pressure, but also by the devastating consequences of climate change, such as the increase of carbon dioxide in the atmosphere (with consequent warming of the surface and acidification of the oceans) and the increase in mortality of many species due to changes in environmental parameters.

The conservation of biodiversity mainly depends on the recognition of the problem, the identification of taxa at risk and the constant monitoring of the most critical situations in order to be able to develop ecosystem restoration and recovery projects. Knowledge of the genome allows us to provide a targeted approach to protect biodiversity, improving the efforts made to protect ecosystems and endangered species. Reference genomes and population genetic data are generally obtained with DNA barcoding and metabarcoding techniques, sequencing of DNA and RNA fragments and representation of entire genomes, but the success of these approaches depends on the quality of the starting biological material, on the laboratory and the availability of funds (Theissinger *et al.*, 2023).

1.11 DNA barcoding

Nowadays there are drones, remote sensing, and many other tech-driven solutions that allow scientists to assess the health of ecosystems and implement protection, conservation, and restoration projects. DNA barcoding is an efficient and effective approach that allows to uniquely identify a species and monitor the degree of biodiversity: the barcodes are short sequences of nucleotides obtained from a standardized portion of the genome and are collected and catalogued in special digital libraries (Kress *et al.*, 2015). While the barcoding for the animal world is based on the mitochondrial cytochrome c oxidase subunit I (COI), a valid equivalent for the plant world has not yet been identified. Metabarcoding has proved to be an effective approach for sequencing even fragments of degraded or damaged DNA, but the limited length of the sequences often requires further investigations to accurately identify taxa at the species rank.

Numerous taxonomists have suggested that in order to accurately distinguish plant species, it may be necessary to analyze the plastid genome, also known as plastome, which contains a high percentage of highly conserved genes characterized by relatively limited mutations during evolution (Li *et al.*, 2015; Schwarz *et al.*, 2015). Despite this, it has not yet been possible to identify a universally valid method for all plant species, although the Consortium for the Barcode of Life (Group, 2009) has suggested the study of seven main candidates that can be used individually as single-locus barcoding or in several combinations via multi-locus barcoding (atpF-atpH spacer*, matK, rbcL, rpoB, rpoC1,* psbK-psbl spacer*,* and trnH-psbA spacer) (Hebert *et al.*, 2003; Group, 2009; Li *et al.*, 2015). In 2009, CBOL proposed the combination *rbcL* + *matK*, which combines the retrieval ability of *rbcL* with the discriminatory power of *matK*, but this has not proved sufficient to create a universally valid barcode for all plant species (Group, 2009; Li *et al.*, 2015).

1.12 Whole genome sequencing (WGS)

The data obtained from whole genome sequencing provide high-resolution information on the demographic history, gene recombination and natural selection that has acted on the species under examination, allowing to identify structural variations and investigate the three-dimensional architecture of the genome. Since samples of material from endangered species are rare, sequencing must be able to maximize the information gleaned from each sample (Theissinger *et al.*, 2023).

As previously highlighted, genomic references play a fundamental role in obtaining genetic information and therefore they need to be complete, correct and rich in chromosomal sequences. This not only allows to reconstruct the Tree of

Life and to establish the phylogenetic relationships between organisms, but also allows for the development and improvement of biodiversity conservation and restoration techniques. An example of a technique for whole genome sequencing is Hi-C, a system that provides a complete mapping of chromosomes (Lieberman-Aiden *et al.*, 2009).

1.2 Hi-C: a new genomic and epigenetic analysis technique

Although chromatin folding plays an essential role in genome compartmentalization, ordinary microscopy is unable to distinguish and detect the loci at acceptable resolution. Currently, techniques such as 3C (Chromosome Conformation Capture) and NGS (Next Generation Sequencing) are used but they require the selection of target loci and do not allow the study of the entire genome.

Hi-C, also known as "Traditional Hi-C" or "Hi-C standard", is a high-throughput genomic and epigenetic analysis technique. It was first described by Lieberman-Aiden *et al*. in 2009 to capture and analyze the chromatin conformation thanks to maps that highlight bases spatial proximity. This allowed the three-dimensional architecture of the genome to be reconstructed through parallel sequencing, producing spatial proximity maps with a resolution of 1Mb. The obtained results demonstrate the enormous resolving power of the Hi-C technique to map the conformation of entire genomes by generating high resolving power maps (Lieberman-Aiden *et al.*, 2009).

This technique detects chromatin interactions within the nuclear genome thanks to the combination of 3C and NGS approaches: the frequency with which two DNA fragments associate in three-dimensional space is measured linking the structure of the chromosome to the structure of the genome. Hi-C is compatible with NGS systems, enabling detection of chromatin interactions on a whole new level. The enormous resolving power of this sequencing technique allows us to explore the biophysical behavior of chromatin and the role of its structures within the biological functions of the cell nucleus (Belton *et al.*, 2012).

The traditional Hi-C sequencing protocol involves first cross-linking chromatin using formaldehyde, then solubilizing, fragmenting, and binding the loci to build a genomic library of chimeric DNA molecules, or “ligation products”. In this phase, chromatin is digested by a restriction enzyme that acts at the level of the 5’ end, which is treated with biotinylated bases. Next, the relative abundance of chimeras is analyzed, which is related to the probability that the chromatin fragments interact in three-dimensional space across the cell population. Hi-C “labels” the chromatin fragments with a nucleotide labeled with biotin, which is then eliminated using magnetic beads that purify the junctions. Finally, data on chromatin interactions can be obtained directly from the sequencing of the obtained library (Lieberman-Aiden *et al.*, 2009; Belton *et al.*, 2012).

Hi-C sequencing provides several advantages over previous techniques, including the ability to map the overall structure of the mammalian genome and chromosomes and to gain insight into the biophysical properties of chromatin; furthermore, they reveal the location of physical contact between distant genomic elements (Belton *et al.*, 2012; Eagen, 2018; Kim *et al.*, 2022). Hi-C is potentially capable of capturing infinite interactions between chimeric DNA fragments; therefore, the sample under analysis must be large enough to highlight one-of-a-kind interactions present only in a small circle compared to the general population (Belton *et al.*, 2012).

Currently, we still know very little about the molecular basis that allows the folding of chromatin inside the interphase nuclei. Hi-C combines chemical crosslinking of chromatin with DNA fragmentation, labeling and sequencing, using high-throughput techniques to detect closely spaced loci within the genome. In this way it is possible to study the degree of chromatin disaggregation in the folded, active, and inactive sequences of the DNA. These analyses produce contact maps highlighting the overlapping and folding state of chromatin, as demonstrated by Eagen (2018).

More recent studies, such as those conducted by Kyukwang Kim *et al.* (2022), also demonstrate that the development of Hi-C analysis techniques has played a major role in understanding the three-dimensional structure of chromatin and the genome even at higher levels. The power of Hi-C in detecting large-scale structural variations of the mammalian genome and its rearrangement in the identification of cancer-related pathologies has been widely demonstrated (Kim *et al.*, 2022).

Today, there are numerous versions of Hi-C analysis techniques, such as in situ Hi-C, low Hi-C, SAFE Hi-C and Micro-C. Each of them is designed to gain insight into different aspects of chromatin characteristics and functions.

In situ Hi-C studies the relative frequency of DNA recombination events to reconstruct the organization of the genome in three-dimensional space (Johanson and Allan, 2022). Low input capture Hi-C can map and compare promoter-enhancer interactions providing high resolution data and allowing to expand the study of chromatin organization levels (Tomás-Daza *et al.*, 2023). SAFE Hi-C can generate enought non-amplified ligation products to preserve chromatin interactions. This allows to overcome the gene distance biases due to the amplification of the sequences and to reduce the degree of background noise (Niu *et al.*, 2019). Micro Hi-C improves data resolution and reduces background noise from reagents used during sample analysis. This is possible through nucleases and enzymes that digest specific segments of the sequence (Burgess, 2020).

Over the last few years, Hi-C technology has been applied in numerous fields of biological study, such as the study of cell division growth mechanisms, transcription regulation processes and genome evolution (Eagen, 2018; Kim *et al.*, 2022). The integration between the contact maps obtained by Hi-C and other genome sequencing datasets will delineate the modifications and role of chromatin, the level of gene expression and the mechanisms of regulation and stability of the genome (Belton *et al.*, 2012).

1.21 Hi-C applications in genome's modern studies

Hi-C is used as an efficient and effective genome scaffolding tool, but it can also be effectively applied to highlight chromatin conformation during mitosis and meiosis. During interphase, chromatin is loose, and this allows other regulatory activities to take place, but as they enter cell division chromatin is compactly folded into chromosomes. The development of single-cell Hi-C over the last five years has enabled the representation of the entire three-dimensional structural landscape of chromatins/chromosomes during the cell cycle and many studies have found that these genomic domains remain unchanged at interphase and are deleted by silencing mechanisms when the cell enters mitosis (Nagano *et al.*, 2017; Bintu *et al.*, 2018).

Moreover, traditional Hi-C can be used to investigate global changes in chromosome structures during embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) differentiation (Kong and Zhang, 2019). Nowadays, Hi-C is considered one of the standard methods to study transcriptional activities and regulation and has confirmed that three-dimensional chromosome architecture is linked with cell fate (Dixon *et al.*, 2015; Yu and Ren, 2017). In this way, Hi-C allowed scientists and genomics to explore genome architecture during cell growth and development in mammalians: chromatin structural features gradually establish from weaker frequencies to cleaner data points after fertilization of the sperm and oocyte, followed by the zygote stage, 2-, 4-, 8-cell stage, blastocyst stage, and the embryonic stage (Niakan *et al.*, 2012; Du *et al.*, 2017; Ke *et al.*, 2017).

Furthermore, Hi-C analysis techniques permit exploration of the characteristics and evolutionary changes undergone by the genome over the course of geological eras, especially thanks to the improvement of data on the three-dimensional structures of the genome. The development of Hi-C technologies provides a new perspective on the evolution of the eukaryotic tree of life (Kong and Zhang, 2019).

1.3 How to obtain the Hi-C contact map

Hi-C scaffolding produces chimeric DNA that represents pairwise chromatin interactions or physical 3D contacts within the nucleus (Lieberman-Aiden *et al.*, 2009; Belton *et al.*, 2012; Lin *et al.*, 2018; Kong and Zhang, 2019) and the gathered data are used to generate a Hi-C contact map. There are several methods to analyze these maps to identify and explore chromosomal structural patterns and many of these data analysis can also be used with 3C sequencing.

1.31 Proximo Hi-C genome scaffolding

To implement the Hi-C proximo scaffolding it is necessary to crosslink the sample with formaldehyde to fix the DNA sequences close to each other. Chromatin assumes a complex three-dimensional architecture that brings even very distant portions closer together and formaldehyde allows DNA-protein interactions to be fixed, detected, and quantified (Hoffman *et al.*, 2015). Subsequently, endonucleases and restriction enzymes cleave the cross-linked DNA, creating fragments of various lengths that retain the covalent bonds fixed by formaldehyde. Biotin fills the gaps at the 5' ends creating chimeric junctions and uncross linked proteins are removed by dilution: the process is based on the premise that the more often two sequences are linked to each other, the closer they are in genomic space. Subsequently, the chimeric DNA fragments are cleaned of biotin and assembled in an orderly manner to reconstruct the chromatin sequences, creating an assembly draft. In this way it is possible to produce libraries of scaffolds reordered in decreasing length by mapping the three-dimensional conformation of the chromatin within the genome (van Berkum *et al.*, 2010; Belton *et al.*, 2012). Finally, PCR duplicates and amplifies the ligated fragments, preparing the interacting chromatin regions for sequencing (van Berkum *et al.*, 2010; Belton *et al.*, 2012).

1.32 Bioinformatics

Through sequencing, data is produced to be collected in a file in FASTQ format: the chimeric sequences are aligned to identify interaction pairs and the shortest sequence capable of uniquely identifying an interaction pair (Lajoie, Dekker and Kaplan, 2015; Pal, Forcato and Ferrari, 2018). Fragments are filtered to remove technical noise and isolate chromatin interactions, followed by balancing to correct errors, binning to group genome reads, and bin filtering to purify the final data (Lajoie, Dekker and Kaplan, 2015; Forcato and Bicciato, 2021; Gong *et al.*, 2021). Finally, balancing and normalization act implicitly or explicitly to correct for bias in Hi-C data (Lajoie, Dekker and Kaplan, 2015; Pal, Forcato and Ferrari, 2018).

1.33 Analyzing the Hi-C contact map

After mapping, filtering and bias-correction of the data, the final product is a square matrix representing the interactions within the genome and the interaction frequency between loci. This interaction matrix allows to identify and interpret the sequenced genome from a biological point of view, reporting the genes and their interaction frequency on the X and Y axes. It is important to note that the interaction frequency cannot be translated into an actual cell number and is therefore an unscaled representation. Each map reflects the interaction frequency captured by formaldehyde crosslinking, which is not necessarily related to the distance between two loci but rather to the three-dimensional architecture of the genome (Lajoie, Dekker and Kaplan, 2015). Map interpretation is complicated by a few factors:

- it is not possible to distinguish scenarios in which multiple interactions occur simultaneously or are mutually exclusive within a cell.
- specific models cannot be identified since each map reflects the result of a method.
- the presence of complex structures in the underlying genomes cannot be excluded a priori.
- it is possible that several models can coexist and overlap.
- being a representation of the frequency of interaction, it does not allow to formulate hypotheses regarding the physical position of the loci in terms of distance or proximity.
- it is not possible to make hypotheses about the ergodicity of the loci, since a single observation regarding one locus or pair of loci does not necessarily hold true for every other locus or pair of loci (Lajoie, Dekker and Kaplan, 2015).

Each map has specific interaction patterns that vary as a function of scale, genome interactions, and point interactions between loci, making it easier to separate pattern identification from pattern interpretation. Despite this, it is possible to outline five main patterns typically present in the genome: cis/trans interaction ratio, distance-dependent interaction frequency, genomic compartments, topological domains, and punctual interaction.

1.4 Aims of the thesis

A first genome draft of Goethe's palm (*Chamaerops humilis* L. – **Figure 1**) was obtained using Pacbio SMRT long-read sequencing. The aim of the project was to obtain an almost chromosome-level assembly applying the Hi-C ARIMA technology on the oldest plant at the Padua Botanical Garden. Thousands of previously obtained scaffolds were joined, producing a draft characterized by repetitive regions and particularly long monomers: the Hi-C contact map obtained with this information presents 18 main scaffolds, for a total of 3.3 Gbp. A ~95% complete draft with high completeness would provide essential data on the information contained in the genome, the presence of repetitive and intergenic regions, the degree of proximity between genes and the presence of polygenic blocks. The linear reconstruction of the genome and its high contiguity assembly will provide information on its regulatory mechanisms and its three-dimensional architecture, on the selective pressure that acted on the different regions and on the processes that directed its evolution.

1.41 Padua Botanical Garden and Goethe's Palm

Founded in 1545, the Padua Botanical Garden is the oldest botanical garden in the world still located in its original location. It covers an area of about 2.2 hectares and was named a UNESCO World Heritage Site in 1997. The ancient garden preserves over 6,000 cultivated plants representing 3,500 different species, among which the "historic trees" are particularly important because of their incredible longevity and cultural value. Today, the Botanical Garden welcomes many rare and unique species, among which the Goethe Palm, belonging to the *Arecaceae* family, stands out. It is estimated that the *Arecaceae* family include about 2,600 species enclosed in 181 genera (Christenhusz and Byng, 2016), mainly distributed in tropical and subtropical climates (Balslev, Bernal and Fay, 2016). The first documented fossil remains date back to the late Cretaceous (about 80 million years ago) and, despite being monocotyledonous angiosperms, they have an arboreal habitat thanks to the modifications of the parenchymal tissue.

Although most of the species are prevalent at low latitudes, *Chamaerops humilis*, also known as “Palma di San Pietro”, is the only species capable of living in the northeast latitude. It is a typical species of the Mediterranean maquis and is widespread in the western Mediterranean between Portugal, Malta, Morocco, and Libya. In Italy it is mainly present along the western coasts from Sicily to Tuscany, in Sardinia and in some islands of the Tyrrhenian Sea. This palm is a thermophilic species that prefers the coastal and sunny environments: it grows at average temperatures above 10 °C, but the optimal growth temperature is between 22-30 °C. It grows well on rocky or sandy soils and, although it can withstand temperatures down to -12°C, it normally fears intense cold.

In ancient times called “palm thrown to the ground” (from the Greek language χαμαί chamái, “on the ground” and ῥώψ rhṓps, “bush”), this evergreen bush has a stem of variable diameter (10-15cm) with a fibrous appearance whose basal portion is covered by dead leaves scaly residues. The leaves are supported by spiny petioles that branch out from the top of the stem and are green on the upper side and white on the underside. In nature this plant can reach 2m in height, but specimens grown in greenhouses can reach 5-6m. Today, it is the only native palm species to have survived the glaciations that affected Europe 12,000 years ago (*La gigante nana dell’Orto botanico*, 2019).

Chamaerops humilis is a dioecious species that produces flowers grouped in yellow and panicle inflorescences. The male flowers have 6 to 9 stamens borne on a fleshy calyx; the female flowers have up to 3 fleshy apocarpic carpels. Pollination is possible thanks to the mutualistic symbiosis with *Derelomus chamaeropsis*, a weevil that lays its eggs directly on the staminate inflorescences. The larvae complete their development and penetrate the rachis and subsequently the adults travel among the inflorescences carrying the pollen and allowing pollination, creating a symbiotic interaction: the palm requires the beetle for pollination and the insect feeds exclusively on the reproductive tissues of the palm during each of its vital stages (Anstett, 1999). The fruit is a globose or oblong drupe whose length varies between 12-14mm. It has a fibrous and green pulp, but it evolves towards a reddish-yellow color as it matures and becomes brown when ripe. Two varieties of palm are known today, distinguished by the leaves color:

- *Chamaerops humilis* var. *humilis*. Widespread in Europe (Portugal, Spain, south of France, Italy) and with green leaves.
- *Chamaerops humilis* var. *argentea* André (syn. *C. humilis* var. *cerifera* Becc.). Widespread in Africa (Morocco, Algeria, and Tunisia) and with greyish leaves.

Chamaerops humilis is currently the oldest plant in the Botanical Garden: planted in 1585 in the medicinal plants sector, it is universally known as "Goethe's Palm" since the German poet Johann Wolfgang von Goethe (28 August 1749, Frankfurt - 22 March 1832, Weimar) was fascinated by it during the "Grand Tour in Italy". On 26 September 1786, during his stop at the Botanical Garden, the writer was able to admire the palm and formulated his evolutionary intuition subsequently expressed in the "Essay on the metamorphoses of plants" published in 1890. In this essay the author acknowledges the homology of the components of different plants and of the successive phases of the life cycle of the same plant, constituting the founding principle of the so-called "Goethe's science". His observations focus on the transformations of the plant and leaves during the life cycle and led him to theorize the foundations of modern plant physiology (*La gigante nana dell'Orto botanico*, 2019).

Despite being a "dwarf plant", today Goethe's Palm reaches 12m in height and the poet himself described it by paying particular attention to "[the] leaves that rise from the ground, first simple and lanceolate, then dividing like the fingers of a hand bent". The palm has been able to reach such a height because the Paduan specimen belongs to the *arborescens* variety and is protected by a greenhouse that limits the impact of the winter cold that characterizes the Po Valley.

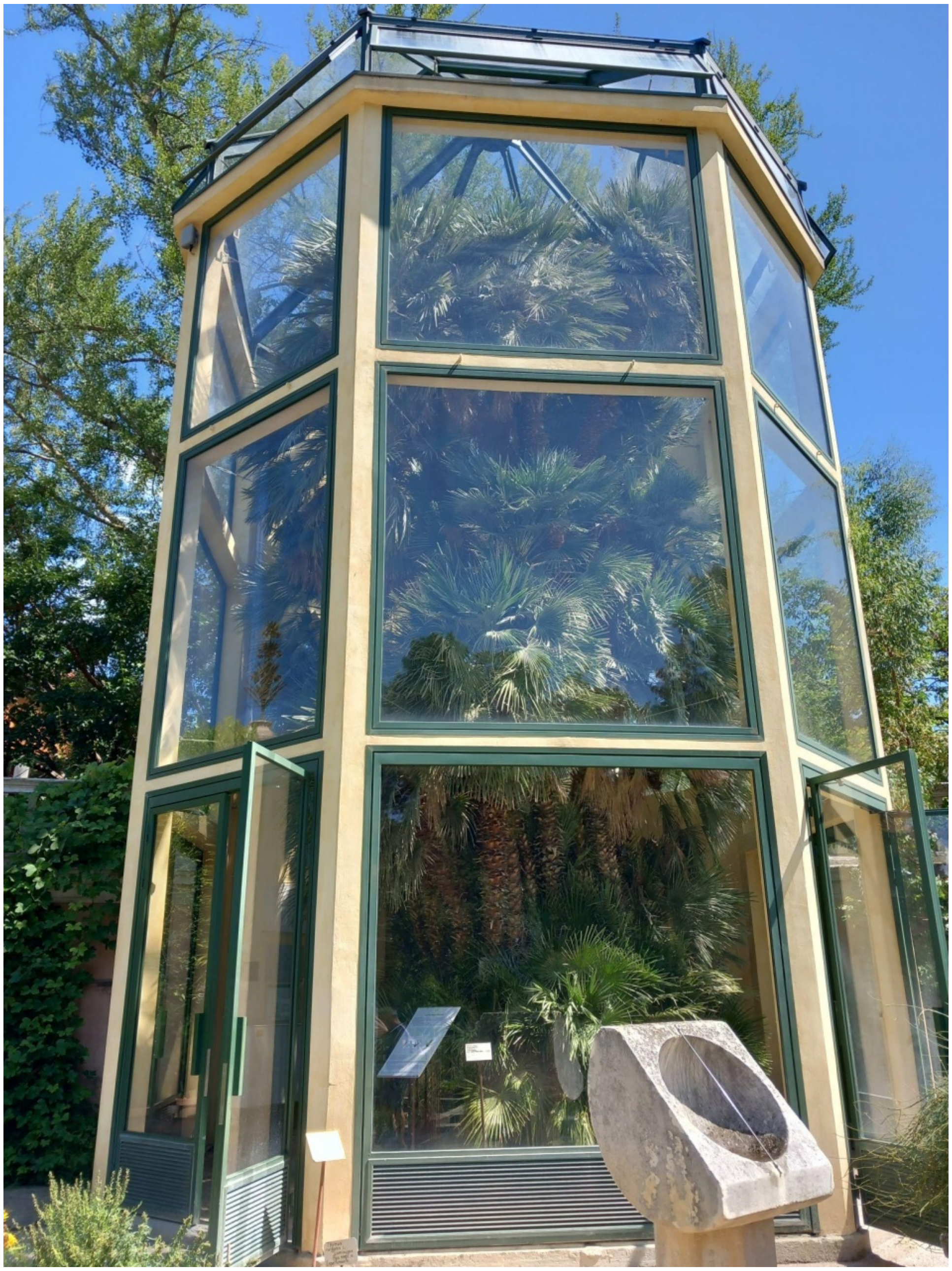

Figure 1 - *Chamaerops humilis,* Padua Botanical Garden. Currently Goethe's palm is protected by a greenhouse that creates a controlled environment, but it is possible to walk inside to observe the arrangement of the trunks and leaves. This palm inspired the "Essay on the metamorphoses of plants" written by Johann Wolfgang von Goethe and published in 1890.

2. Materials and Methods

2.1 Hi-C library preparation and sequencing

Chromatin interaction (Hi-C) libraries were generated using Arima Genomics libraries on DNA isolated using a modified CTAB method and sequenced on Illumina instruments. Arima Hi-C preparation was performed by the LOEWE Translational Biodiversity Genomics center lab using the Arima Hi-C kit that uses two enzymes (P/N: A510008). The resulting Arima Hi-C proximally ligated DNA was then sheared, size-selected around 200-600 bp using SPRI beads, and enriched for biotin-labeled proximity-ligated DNA using streptavidin beads. KAPA Hyper Prep kit (P/N: KK8504) was used to generate an Illumina-compatible library from the fragments. The resulting library was PCR amplified and purified with SPRI beads. The quality of the final library was checked with qPCR and Bioanalyzer, then sequenced on Illumina, flanked by histograms for each taxonomic group and analyzed HiSeq X at ~30× coverage following the manufacturer's protocols.

2.2 Bioinformatics

For filtering and mapping the data to the contig-level genome assembly we followed the Arima Hi-C mapping pipeline (https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima_mapping_pipeline.sh). This process was necessary in order to prepare our data before reaching the chromosome-level assembly. We first mapped the paired-end reads in FASTQ format using BWA-MEM v.0.7.17 (Li, 2013) in order to discard "chimeric reads". These "chimeric reads" were later filtered with Samtools v.1.17 (Danecek *et al.*, 2021). Subsequently, duplicated reads were removed with Picard v.3.1.0 (*Broad Institute*, 2018). The mapped and filtered reads were then analyzed with YaHS v.1.1a.2 (c-zhou, 2023; Zhou, McCarthy and Durbin, 2023) for proximity-ligation-based scaffolding, and with JuicerTools v.2.20.00 (Durand *et al.*, 2016) for generating the Hi-C contact map. The Hi-C assembly was visualized with JuicerBox v.2.3.0 website interface (Robinson *et al.*, 2018; https://aidenlab.org/juicebox/).

With Assemblathon (Bradnam *et al.*, 2013) we obtained the consensus assembly for our data. After obtaining the Hi-C assembly at chromosome-level, we used TGS-Gap Closer v.1.2.1 (Xu *et al.*, 2020) to close the N-gaps within the sequences and complete genome assembly. Racon was used as an error correction module to enhance the base-level accuracy of merged sequences.

In order to evaluate the quality and completeness of the assembly, we ran the assembly statistics with Quast v. 5.0.2 (Gurevich *et al.*, 2013) and the gene set completeness with BUSCO v. 5.2.2 (Manni *et al.*, 2021) using the Viridiplantae odb10 database. We visualized the genome with BlobTools v.1.1 (Laetsch and Blaxter, 2017) in order to detect if all scaffolds belong to our target. BlobTools v.1.1 is a program written in Python and designed for visualization, quality control and partitioning of the genome dataset. We improved the assemblies and the final screening by analyzing the guanine and cytosine content, the coverage of reading in sequencing libraries and the taxonomy of similarity matches between genome sequences.

3. Results

3.1 Illumina raw reads

Through Arima_HiC_GP5D we obtained 649,277,018 raw reads and a total of 97,391,552,700 bp raw data with an estimated error of 0.03%. The process was 100.00% effective and, as far as the quality score is concerned, 94.99% of the data had Q20; therefore, it is possible that there is an error every 100 bp (99.00% accuracy call), while 88.26% had Q30, or the possible presence of one error every 1000bp (99.90% accuracy call). This indicates a very low probability of error within the data. The GC coverage was calculated at 44.82%.

3.2 Conting-level assembly results

The contig-level genome assembly presents a total genome size of 4.4 Gb. The total length was estimated at 4,407,665,712 bp. Among the 2,897 total contings, the largest one was 28.6 Mbp long. We calculated a GC coverage of 43.80%. The contigs presented a N50 of 4.4 Mbp and a N90 of 0.8 Mbp (**Table**

1). Through Assemblathon we obtained 3,464 total contings (of which 1,225 are present in scaffolds and 2,239 are excluded) and 2,317 total scaffolds (**Tables 2-3**). Scaffolded contings present a 84.90% of assembly and unscaffolded contings present a 15.10% of assembly. The average number of contings per scaffold is 1.5 and the average length of break (> 25 Ns) between contings in scaffold is 200.

Table 1. Conting-level assembly results.

Reference	**Values**
Total contings	2,897
Largest (bp)	28,621,313
Total length (bp)	4,407,665,712
N50 conting length	4,424,387
N90 conting length	843,495
L50 conting count	280
L90 conting count	1,127
GC (%)	43.80
Total N's per 100 kbp	0.00

Table 2. Assemblathon consensus assembly results – contings.

Reference	**Values**	
Total number	3,464	
Contings in scaffolds	1,225	
Contings not in scaffolds	2,239	
Total size (bp)	4,407,665,712	
Largest (bp)	28,621,313	
Shortest (bp)	1,000	
Mean size (bp)	1,272,421	
Median size (bp)	353,000	
Number of contings > 1 Knt	3442	99.40%
Number of contings > 10 Knt	3431	99.00%
Number of contings > 100 Knt	2536	73.20%
Number of contings > 1 Mnt	1096	31.60%
Number of contings > 10 Mnt	39	1.10%
N50 conting length	3,866,000	
L50 conting count	320	
Conting % A	28.10	
Conting % C	21.90	

Conting % G	21.90
Conting % T	28.10
Conting % N	0.00
Conting % non-ACTGN	0.00
Total contings non-ACTGN nt	0

Table 3. Assemblathon consensus assembly results – scaffolds.

Reference	Values	
Total number	2,317	
Total size (bp)	4,407,895,112	
Largest (bp)	338,985,879	
Shortest (bp)	1,000	
Mean size (bp)	1,902,415	
Median size (bp)	141,908	
Number of scaffolds > 1 Knt	2295	99.10%
Number of scaffolds > 10 Knt	2284	98.60%
Number of scaffolds > 100 Knt	1389	59.90%
Number of scaffolds > 1 Mnt	215	9.30%
Number of scaffolds > 10 Mnt	24	1.00%
N50 scaffold lenght	195,019,846	
L50 scaffold count	9	
Scaffold % A	28.10	
Scaffold % C	21.90	
Scaffold % G	21.90	
Scaffold % T	28.10	
Scaffold % N	0.01	
Scaffold % non-ACTGN	0.00	
Total scaffolds non-ACTGN nt	0	

3.3 Results from gap-filling

The following table shows assembly statistics before and after running TGS-Gap Closer. At the end of the process, the largest scaffold among the 2,317 total ones was 339 Mbp and we calculated a total length of 4.4 Gbp. We estimated a N50 of 195 Mbp and N90 of 1.1 Mbp. GC% was 43.80%. We obtained 4.71 N's per 100 kb for a total of 207,800 N's (**Table 4**).

Table 4. Statistics and results of the Hi-C assembly before and after running TGS-Gap Closer.

Statistics (no reference)	Hi-C assembly before TGS-Gap Closer	Hi-C assembly after TGS-Gap Closer
Total scaffolds	2,317	2,317
Largest scaffold (bp)	338,985,879	339,009,363
Total length (bp)	4,407,895,112	4,408,031,063
N50 scaffold length	195,019,846	195,022,640
N90 scaffold lenght	1,112,000	1,112,000
L50 scaffold count	9	9
L90 scaffold count	197	197
GC (%)	43.8	43.8

Mismatches	Hi-C assembly before TGS-Gap Closer	Hi-C assembly after TGS-Gap Closer
Total N's per 100 kbp	5.2	4.71
Total N's	229,400	207,800

3.4 Hi-C contact map

The hi-C contact map presents 2,317 total scaffolds, among which it is possible to identify 18 chromosomes. The genome size of these 18 chromosomes is 3.3 Gbp (**Figure 2**).

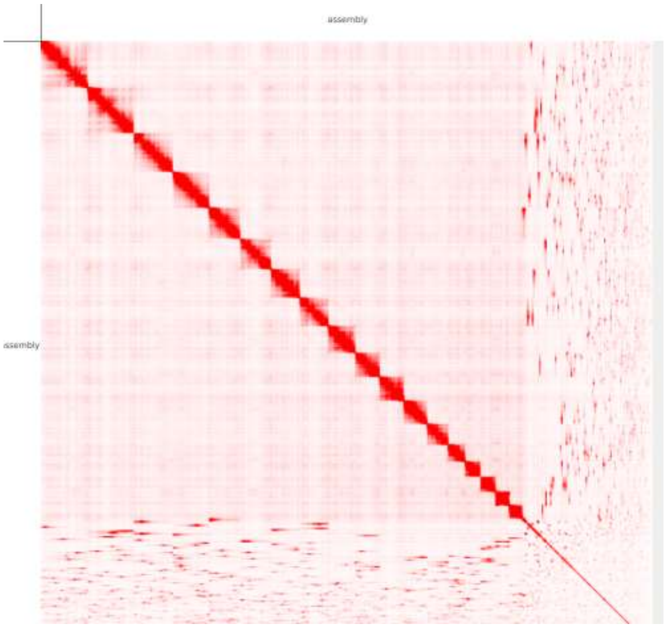

Figure 2 - Contact map for the consensus Hi-C re-assembly showing the 18 main scaffolds along the diagonal. An intense red color indicates the presence of many interactions, pink a moderate interaction and white the absence of interactions. Interactions increase within chromosomes and between neighboring chromosomes, which is why a very dark red-orange diagonal is created between the ends of the matrix.

3.5 BUSCO data analysis

The 2,317 scaffolds were searched for 425 BUSCOs genes and their analysis showed 422 complete BUSCOs, equivalent to a 99.30% completeness (**Figure 3**). Those 422 BUSCOs include 373 single-copy BUSCOs (S: 87.80%) and 49 duplicated BUSCOs (D: 11.50%). We also found 2 fragmented BUSCOs (F: 0.50%) and a missing one (M: 0.20%).

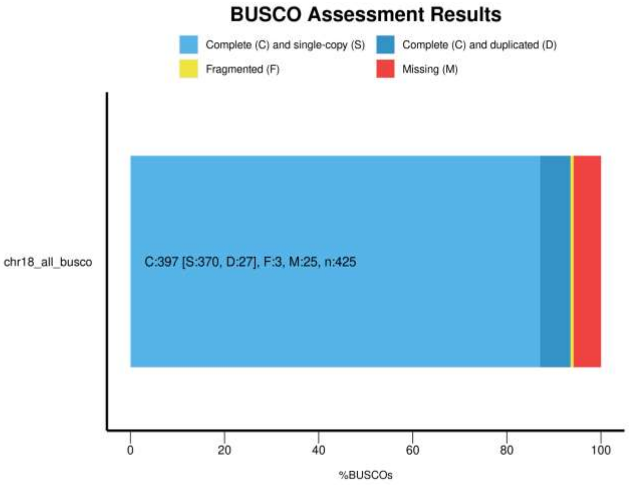

Figure 3 - BUSCO output for the Hi-C assembly. The graph shows an overall completeness of 99.30%, including single-copy (S: 87.80%), duplicated (D: 11.50%), fragmented (F: 0.50%) and missing (M: 0.20%) BUSCO genes.

From the BUSCO analysis of the 18 chromosome blocks, we identified 397 complete BUSCOs for a 93.50% completeness (**Figure 4**). These include 370 single-copy BUSCOs (S: 87.10%) and 27 duplicated BUSCOs (D: 6.40%). There are 3 fragmented BUSCOs (F: 0.70%) and 25 missing ones (M: 5.80%).

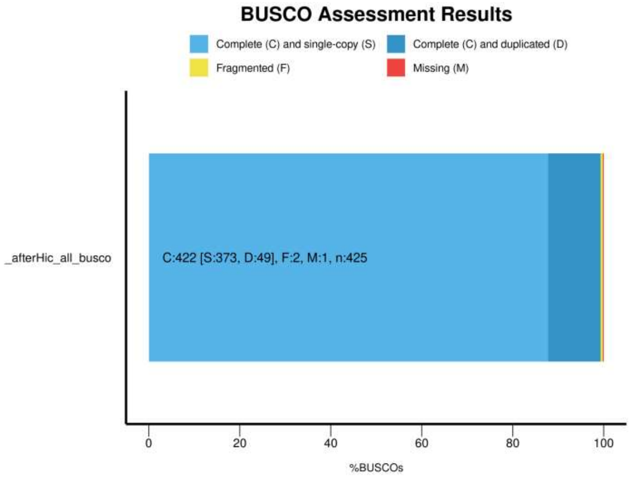

Figure 4 - BUSCO output for the 18 chromosomes. The graph shows an overall completeness of 93.50%, including single-copy (S: 87.10%), duplicated (D: 6.40%), fragmented (F: 0.70%) and missing (M: 5.80%) BUSCO genes.

3.6 BlobTools

Through BlobTools we obtained a Blobplot, i.e., a two-dimensional scatter plot in which the genomic sequences are represented by circles (**Figure 5**). For each sequence it is important to consider that the position along the Y axis is given by the base coverage of the sequence in the coverage library, a proxy for molarity of input DNA; the position along the X axis is determined by cytosine and guanine content, the proportion of which can differ substantially between genomes (Laetsch and Blaxter, 2017).

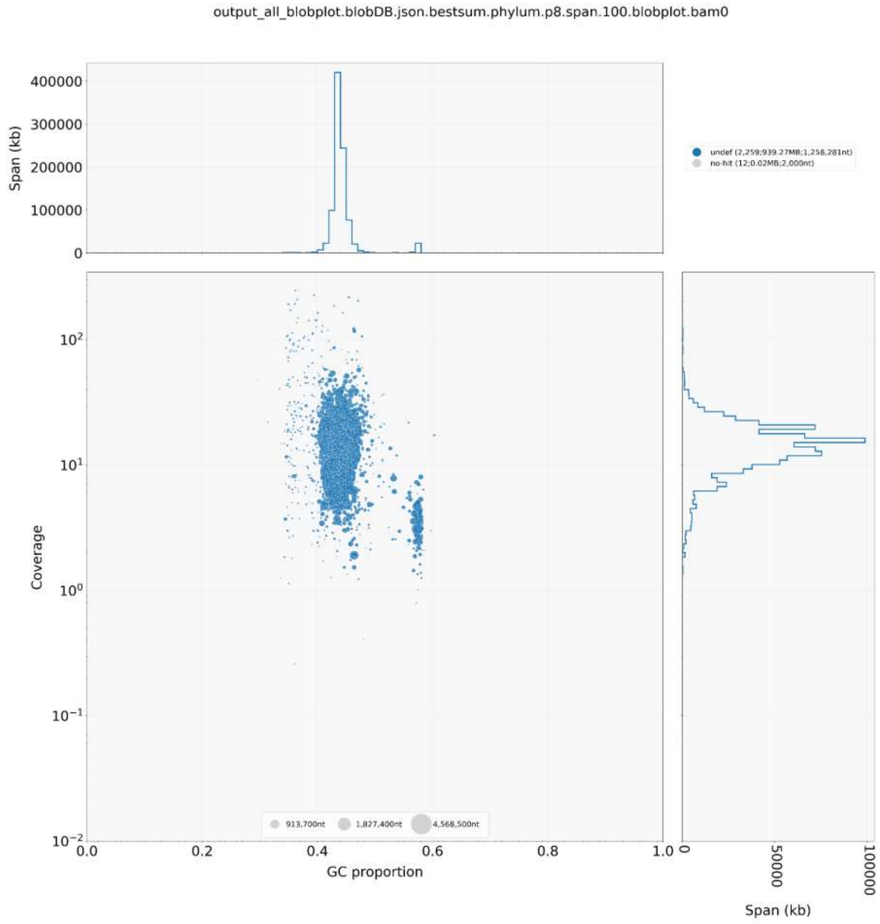

Figure 5 - Two-dimensional scatter Blobplot of the Hi-C assembly. The main blob constitutes the *C. humilis* genome and small blobs in the bottom right corner represent mainly organellar contings. The sequences are represented by circles whose diameter is proportional to the length of the sequence.

4. Discussion and future perspectives

In this project we obtained a high quality, i.e., almost chromosome-level, assembly of the genome of *Chamaerops humilis*. The newly obtained *C. humilis* genome represents the highest quality genome among the eight genome references available for species of the *Arecaceae* family. The eight genome reference species are the following: *Areca catechu* L., *Calamus simplicifolius, Cocos nucifera* L., *Elaeis guineensis* Jacq., *Elaeis oleifera* (Kunth) Cortés*, Metroxylon sagu* Rottbøll, *Phoenix dactylifera* L. and *Phoenix roebelenii* O'Brien. All of their genomes have been fully sequenced and the data is available on GenBank (https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=4710).

Observing the available genomes (**Table 5**) it is possible to note that the total number of contigs of *Chamaerops humilis* is much lower than the others, and that our assembly has a considerably higher N50. An N50 greater than 1Mb is normally considered good; in our case this value is ~100 times higher.

Within the list, the closest relatives of *Chamaerops humilis* are *Phoenix dactylifera* and *Phoenix roebelenii*: all three species belong to the same subfamily, i.e., *Coryphoideae*, but different tribes. *P. dactylifera* has 18 chromosomes as found in *C. humilis*. The *Phoeniceae* can reach considerable heights in nature (even 20-30m) but, although the tribe is also widespread in the Mediterranean, they are not able to colonize higher latitudes with the same efficiency as *Chamaerops humilis*.

Table 5. *Arecaeae* genomes available on GenBank (*Areca catechu* L., *Calamus simplicifolius, Cocos nucifera* L., *Elaeis guineensis* Jacq., *Elaeis oleifera* (Kunth) Cortés*, Metroxylon sagu* Rottbøll, *Phoenix dactylifera* L. and *Phoenix roebelenii* O'Brien) compared to *Chamaerops humilis*. While our GC% coverage is only slightly higher than the rest, *Chamaerops humilis* shows a N50 ~100 times higher than the other species.

Species	**Genome size (Mb)**	**Level**	**Total contings**	**Contings N50 (Mb)**	**Total scaffolds**	**Scaffolds N50 (Mb)**	**GC%**	**Genome coverage**
Areca catechu	2,823	Chromosome	7,480	21.2	73	186.5	41.0	100.0x
Calamus simplicifolius	1,961	Scaffold	22,420	0.1369	5,116	0.8030	41.0	270.0x
Cocos nucifera	2,102	Scaffold	20,060	0.1636	7,998	0.5705	37.5	50.0x
Elaeis guineensis	1,535	Chromosome	202,467	0.0094	40,348	1.0	37.0	16.0x
Elaeis oleifera	1,403	Scaffold	219,249	0.0084	26,756	0.3313	37.5	16.0x
Metroxylon sagu	472.4	Scaffold	879	1.1	667	1.5	36.0	50.0x
Phoenix dactylifera	772.3	Chromosome	2,706	0.8972	2,389	4.7	40.0	100.0x
Phoenix roebelenii	471.7	Scaffold	84,938	0.0194	57,227	10.5	39.5	95.0x
Chamaerops humilis	4,4	Chromosome	3,464	3.866	2,317	195.022	43.8	20.0x

4.1 Previous studies conducted on *Chamaerops humilis*

Historically, traditional medicine has often made use of the medicinal properties of palm trees and over the years there have been numerous studies on the metabolic compounds produced by *Chamaerops humilis* and their possible application in the medical field, such as the research conducted by José P. Coelho *et al.* (2017) about its compounds and their possible application in the fight against cardiovascular diseases. Sandra Gonçalves *et al.* (2018) have successfully extracted compounds with antioxidant properties (ergo they are involved in contrasting the oxidative stress exerted by oxygen radicals) and capable of inhibiting the main enzymes related to neurodegenerative diseases, such as acetylcholinesterase (AChE), butyrylcholinesterase (BChE) and tyrosinase (TYR) (Gonçalves *et al.*, 2018). The antioxidant and inhibitory properties of the identified compounds vary in relation to the tissue of origin and show peaks in correspondence with the seeds, but the results achieved suggest that *Chamaerops humilis* could constitute a new frontier to be explored in the phytochemical and pharmacological fields (Gonçalves *et al.*, 2018).

Now that a high-quality reference genome of *Chamaerops humilis* is available, it will be possible to fully investigate the molecular mechanisms underlying its antioxidant and inhibitory properties, identifying the gene sequences involved and providing new research ideas in the fight against neurodegenerative and cardiovascular diseases.

4.2 *Phoenix* genus genome analysis

The *Phoenix* genus has been extensively studied for many years due to its economic importance and morphological characteristics and nowadays we can count on genetic studies. *Phoenix dactylifera* is one of the main date palms and is widespread along the coasts of North Africa, South Europe, and Western Asia. It is a monophyletic group of 14 species among *Coryphoideae,* but their ability to hybridize and produce fertile hybrids and their morphological similarities make it difficult to draw clear boundaries between one species and another.

4.21 *Phoenix* sp. systematics and evolution

Although their belonging to the *Coryphoideae* group is supported by molecular studies (Asmussen *et al.*, 2006), their relationship with the other species still remains a mystery: some recent genomic studies place them among the *Trachycarpeae*, with variable support based on the DNA portion analyzed (Asmussen *et al.*, 2006; Barrett *et al.*, 2016; *Complete Generic-Level Phylogenetic Analyses of Palms (Arecaceae) with Comparisons of Supertree and Supermatrix Approaches | Systematic Biology | Oxford Academic*, no date). Although the times of divergence from the common ancestor have not yet been estimated with certainty, the radiation from the *Trachycarpeae* is believed to have occurred around 49 million years ago (Baker and Couvreur, 2013). The reconstruction of its evolutionary history is currently compromised by the absence of in-depth information regarding the rates of molecular evolution and by the presence of recent hybridization events (Gros-Balthazard *et al.*, 2021), but the presence of complete genomes of various species will allow us to deepen its origins and the mechanisms that have favored its expansion in the Mediterranean basin. In addition, the comparison with other palm genomes, such as that of *Chamaerops humilis*, will allow us to further investigate the phylogenetic relationships with the other members of the *Coryphoideae* group.

4.22 *Phoenix dactylifera* ability to endure salinity stress

Although *Phoenix dactylifera* is a relatively salt-tolerant species, due to the high salinity of irrigation water it is subjected to considerable osmotic stress. The molecular study conducted by Mahmoud W. Yaish *et al.* (2017) on *Phoenix dactylifera* genome revealed that its leaves and roots can express respectively 2,630 and 4,687 genes when the plant is subjected to saline stress, of which 194 are found in both types of tissue. Further studies have shown that these genes are involved in the efficiency of metabolic pathways present in leaves, such as carbohydrate metabolism and oxidative phosphorylation, but also in membrane transport mechanisms and protein metabolism in root tissues (Yaish *et al.*, 2017). Among these genes, GH3, an auxin-responsive gene and probable potassium transporter, and the proton pump characteristic of the vacuolar tonoplast stand out (Yaish *et al.*, 2017).

Thanks to the relative phylogenetic proximity between *Chamaerops* and *Phoenix*, the characterization of the genes involved in tolerance to soil and irrigation water salinity can constitute an important starting point for investigating the mechanisms by which *Chamaerops humilis* has adapted to live also along the coasts of the Mediterranean (and of Sicily) to then colonize the dune environment behind it.

4.3 Molecular studies conducted on *Elaeis guineensis*

Even thought knowledge on the genome of palm trees is still limited, many steps forward have been made to date in terms of their sequencing. It is important to mention the sequencing project conducted by Le Wang *et al.* (2022) on the Dura variety of the African oil palm (*Elaeis guineensis*), one of the major exponents of the Arecaceae family. This research made it possible to assemble a genome of 1.7 Gb, equal to 94.50% of the total estimated size (1.8 Gb), furthermore it was also possible to recompose the sequences at the chromosomal level by analyzing the conserved sequences shared with the date palm and the coconut palm (Wang *et al.*, 2022). Although the study highlights the presence of possible inaccuracies and incompleteness, such as the fragmentation of the scaffolds, and therefore the need to conduct further research, it is important to underline that the acquired resources can be used to accelerate the understanding of evolution of palm trees (Wang *et al.*, 2022).

4.31 VIRESCENS role in fruit color

This study suggests that environmental stress has heavily impacted the selection of genes related to the response to changes in external conditions and adaptation to new habitats (Wang *et al.*, 2022). Furthermore, the more detailed analysis of the mutations present in the base sequence of the VIRESCENS gene suggests its probable involvement in the color change of palm fruits (Wang *et al.*, 2022). VIRESCENS is a transcription factor located on chromosome 1 and it regulates the accumulation of anthocyanins in the exocarp, resulting in a color that changes from purple to black (Singh *et al.*, 2013): such a dark color makes the fruit less visible than the others (Wang *et al.*, 2022) and it is observed that the seed dispersers prefer fruits with bright red and yellow colors, thus selecting the fruits of that color to the detriment of the darker ones (Wang *et al.*, 2022). This is in contrast with the hypothesis according to which the color evolution of neotropical palm fruits is linked to interactions with frugivores (Nascimento *et al.*, 2020) and their ability to distinguish colors (particularly red and green), providing an evolutionary advantage to all parties involved (Onstein *et al.*, 2020). Despite this, it is underlined that the genomic and molecular basis of fruit color evolution need further investigation (Wang *et al.*, 2022). The study found that oil and date palms with reddish-yellow fruits have a low concentration of anthocyanins and a mutation in the VIRESCENS genes, whose sequence is interrupted by highly repetitive regions (Wang *et al.*, 2022). It follows that frugivores negatively select the VIRESCENS gene preferring red and yellow fruits over purple or black fruits (Wang *et al.*, 2022).

Chamaerops humilis has green fruits in the immature stage, but they turn bright yellow, orange, and red once they are ripe, and then brown at the end of the life cycle phase. It would be interesting to investigate the role of the VIRESCENS gene within the coloration of the fruit and how it impacts the mutualistic relationship with the pollinator *Derelomus chamaeropsis*, currently known as the only pollinator of this species.

4.32 Transposons increase, genome enhancement and palm specimen

Genome expansion and increased diversity among species are mainly associated with polyploidy and transposon increase: these two events increase the amount of available genetic material, creating a favorable substrate for genetic variations (Gregory *et al.*, 2007; Marburger *et al.*, 2018). Today we know that the palm genome is between 800Mb and 3Gb in size, but this is due to the two main Whole Genome Duplication (WGD) events that occurred respectively ~150 Mya for all monocots and ~75 Mya for the last ancestor of palm trees (Barrett *et al.*, 2019). Despite the notable difference in genome size, date palms (800 Mb) show no evidence of loss of DNA portions compared to oil palms (1.8 Gb): this suggests that oil palms have undergone a proliferation of transposons which favored the increase in DNA size and the diversification of the function of different regions, leading to greater speciation (Wang *et al.*, 2022). This phenomenon, associated with the movement of transposons, may have favored a reorganization of the chromosomes and gene recombination, determining speciation (Tenaillon, Hollister and Gaut, 2010).

The study of the transposable elements present in *Chamaerops* could provide new information on its evolutionary origins and on the genome's duplication events that favored its speciation and made it particularly suitable for surviving at higher latitudes. Transposons have been shown to play a fundamental role in the response and adaptation to environmental stress: they are able to influence genes, create regulatory networks and provide additional genetic material to acquire new capabilities (Casacuberta and González, 2013). Given our high quality, completeness and contiguity genome, a more in-depth study of transposons could provide fundamental information on this species and its evolutionary adaptations.

4.33 How PRs counteracts pathogen infection

Pathogen-related proteins (PRs) are proteins produced in plants following the attack of a pathogen, because of the acquired resistance system. Although many of the mechanisms involving them remain to be investigated, the study reveals the presence of 505 PR in oil palm, 319 in date palm, 382 in coconut palm and 427 in banana palm and the presence of duplications of their genes (Wang *et al.*, 2022). It was observed that in oil palm 97% of duplications were in tandem and the remaining 3% were the result of older translocations or duplications (Wang *et al.*, 2022). This suggests that PRs exhibit more intense activity at birth and death of the organism and can rearrange their position within the genome. Furthermore, the analysis of the consequences on the genome following the infection by *Ganoderma boninense* (Bahari *et al.*, 2018) suggests the involvement of large portions of duplicated PRs in the response to the pathogen (Wang *et al.*, 2022).

Although the Goethe palm is currently kept in a controlled environment such as that of the botanical garden and protected by the greenhouse, this does not guarantee complete protection from the attack of pathogens and fungal infections. The deepening of the role of genes linked to PR proteins could favor the development of new research lines on the response of palms to pathogens to protect also the wild species present in the Mediterranean and currently threatened by infectious agents imported from warmer regions following overheating global and foreign trade. A first example is what happened in 2006 in Sicily, when the type of rot caused by the infection by *Thielaviopsis paradoxa* was found on the trunk of 10 date palms belonging to *Phoenix dactylifera* (Polizzi *et al.*, 2006). It is a typical pathogen of date palms, capable of infecting any part of them and is currently endemic in northern Italy: deepening our understanding of the function of PR in *Chamaerops humilis* could open the way to new applications to boost defense mechanisms against this type of pathogen.

4.34 Consequences of selective pressure and environmental stress

Through the scanning of the genome, 317 genes closely related to the response to stress and pressure exerted by external factors, such as excess ultraviolet radiation, regulation of autophagy processes and defense mechanisms against oxidative stress have been identified (Wang *et al.*, 2022). This suggests that the genomic regions subjected to abiotic stress are also those most affected by natural selection and play a fundamental role in the adaptive evolution of oil palm. The analyzed genes are closely related to the response to ionic and water stress, often related to drought, but some are also partially involved in the response to pathogens (Wang *et al.*, 2022).

Finally, since *Chamaerops humilis* is the only palm capable of surviving at higher latitudes, it would be interesting to investigate which are the portions of DNA most involved and the physiological mechanisms that have allowed the colonization of the Mediterranean maquis.

5. Conclusions

The sequencing of the genome of *Chamaerops humilis* stands as a new point of reference for the *Arecaceae*: the information obtained from its chromosome-level assembly will provide a new perspective on its origins, genome evolution, chromatin three-dimensional organization and adaptation strategies. It will be possible to delve into its evolutionary history and the phylogenetic relationships with the current genomic references since it is the only palm species capable of reaching the northernmost latitudes. A comparison between its genome and that of its closest relatives will allow us to understand which factors have favored its expansion and survival in the Mediterranean Sea and among non-tropical environments at medium and high latitudes, where temperatures decrease considerably compared to the more tropical areas.

This array may also have interesting results in the medical field, thanks to the in-depth study of the antioxidant and inhibitory properties of the compounds produced by *Chamaerops humilis*, opening new scenarios in the field of neurodegenerative and cardiovascular diseases. To date, several studies have followed on the antioxidant and inhibitory properties of the compounds produced by palms and the presence of new genomic references will allow us to further investigate their origins and potential not only in *Chamaerops* but also in the *Arecaeae* family.

It may be feasible to investigate the role of genes involved in fruit color and how this is possibly related to the mutualistic symbiosis with pollinators, transposons and their role in genome duplication, proteins involved in the stress response and the role of selective pressure in the adaptation to new habitats, thus investigating the micro-evolutionary processes that have affected it and that have influenced its distribution.

Therefore, this project will allow to delve the origins and properties of Goethe’s palm, inaugurating new research scenarios in the biogenomic, phylogenetic and biomedical fields.

ACKNOWLEDGEMENTS

Anstett, M.C., 1999. An experimental study of the interaction between the dwarf palm (Chamaerops humilis) and its floral visitor Derelomus chamaeropsis throughout the life cycle of the weevil. Acta Oecologica 20, 551–558. https://doi.org/10.1016/S1146-609X(00)86622-9

Asmussen, C.B., Dransfield, J., Deickmann, V., Barfod, A.S., Pintaud, J.-C., Baker, W.J., 2006. A new subfamily classification of the palm family (Arecaceae): evidence from plastid DNA phylogeny. Botanical Journal of the Linnean Society 151, 15–38. https://doi.org/10.1111/j.1095-8339.2006.00521.x

Bahari, M.N.A., Sakeh, N.M., Abdullah, S.N.A., Ramli, R.R., Kadkhodaei, S., 2018. Transciptome profiling at early infection of Elaeis guineensis by Ganoderma boninense provides novel insights on fungal transition from biotrophic to necrotrophic phase. BMC Plant Biology 18, 377. https://doi.org/10.1186/s12870-018-1594-9

Baker, W.J., Couvreur, T.L.P., 2013. Global biogeography and diversification of palms sheds light on the evolution of tropical lineages. I. Historical biogeography. Journal of Biogeography 40, 274–285. https://doi.org/10.1111/j.1365-2699.2012.02795.x

Balslev, H., Bernal, R., Fay, M.F., 2016. Palms – emblems of tropical forests. Botanical Journal of the Linnean Society 182, 195–200. https://doi.org/10.1111/boj.12465

Barrett, C.F., Baker, W.J., Comer, J.R., Conran, J.G., Lahmeyer, S.C., Leebens-Mack, J.H., Li, J., Lim, G.S., Mayfield-Jones, D.R., Perez, L., Medina, J., Pires, J.C., Santos, C., Wm. Stevenson, D., Zomlefer, W.B., Davis, J.I., 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palms and other commelinid monocots. New Phytologist 209, 855–870. https://doi.org/10.1111/nph.13617

Barrett, C.F., McKain, M.R., Sinn, B.T., Ge, X.-J., Zhang, Y., Antonelli, A., Bacon, C.D., 2019. Ancient Polyploidy and Genome Evolution in Palms. Genome Biology and Evolution 11, 1501–1511. https://doi.org/10.1093/gbe/evz092

Belton, J.-M., McCord, R.P., Gibcus, J., Naumova, N., Zhan, Y., Dekker, J., 2012. Hi-C: A comprehensive technique to capture the conformation of genomes. Methods 58, 10.1016/j.ymeth.2012.05.001. https://doi.org/10.1016/j.ymeth.2012.05.001

Bintu, B., Mateo, L.J., Su, J.-H., Sinnott-Armstrong, N.A., Parker, M., Kinrot, S., Yamaya, K., Boettiger, A.N., Zhuang, X., 2018. Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. Science 362, eaau1783. https://doi.org/10.1126/science.aau1783

Biodiversity Study | Dartmoor [WWW Document], n.d. URL https://www.dartmoor.gov.uk/learning/teachers-educators/biodiversity-study (accessed 7.28.23).

Bradnam, K.R., Fass, J.N., Alexandrov, A., Baranay, P., Bechner, M., Birol, I., Boisvert, S., Chapman, J.A., Chapuis, G., Chikhi, R., Chitsaz, H., Chou, W.-C., Corbeil, J., Del Fabbro, C., Docking, T.R., Durbin, R., Earl, D., Emrich, S., Fedotov, P., Fonseca, N.A., Ganapathy, G., Gibbs, R.A., Gnerre, S., Godzaridis, É., Goldstein, S., Haimel, M., Hall, G., Haussler, D., Hiatt, J.B., Ho, I.Y., Howard, J., Hunt, M., Jackman, S.D., Jaffe, D.B., Jarvis, E.D., Jiang, H., Kazakov, S., Kersey, P.J., Kitzman, J.O., Knight, J.R., Koren, S., Lam, T.-W., Lavenier, D., Laviolette, F., Li, Y., Li, Z., Liu, B., Liu, Y., Luo, R., MacCallum, I., MacManes, M.D., Maillet, N., Melnikov, S., Naquin, D., Ning, Z., Otto, T.D., Paten, B., Paulo, O.S., Phillippy, A.M., Pina-Martins, F., Place, M., Przybylski, D., Qin, X., Qu, C., Ribeiro, F.J., Richards, S., Rokhsar, D.S., Ruby, J.G., Scalabrin, S., Schatz, M.C., Schwartz, D.C., Sergushichev, A., Sharpe, T., Shaw, T.I., Shendure, J., Shi, Y., Simpson, J.T., Song, H., Tsarev, F., Vezzi, F., Vicedomini, R., Vieira, B.M., Wang, J., Worley, K.C., Yin, S., Yiu, S.-M., Yuan, J., Zhang, G., Zhang, H., Zhou, S., Korf, I.F., 2013. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. GigaScience 2, 2047-217X-2–10. https://doi.org/10.1186/2047-217X-2-10

Burgess, D.J., 2020. Chromosome structure at micro-scale. Nat Rev Genet 21, 337–337. https://doi.org/10.1038/s41576-020-0243-y

Casacuberta, E., González, J., 2013. The impact of transposable elements in environmental adaptation. Molecular Ecology 22, 1503–1517. https://doi.org/10.1111/mec.12170

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

Coelho, J.P., Veiga, J.G., Elvas-Leitão, R., Brigas, A.F., Dias, A.M., Oliveira, M.C., 2017. Composition and in vitro antioxidants activity of Chamaerops humilis L., in: 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG). Presented at the 2017 IEEE 5th Portuguese Meeting on Bioengineering (ENBENG), pp. 1–4 https://doi.org/10.1109/ENBENG.2017.7889422

Complete Generic-Level Phylogenetic Analyses of Palms (Arecaceae) with Comparisons of Supertree and Supermatrix Approaches | Systematic Biology | Oxford Academic [WWW Document], n.d. URL https://academic.oup.com/sysbio/article/58/2/240/1672113 (accessed 8.19.23).

Cowie, R.H., Bouchet, P., Fontaine, B., 2022. The Sixth Mass Extinction: fact, fiction or speculation? Biological Reviews 97, 640–663. https://doi.org/10.1111/brv.12816

c-zhou, 2023. YaHS: yet another Hi-C scaffolding tool.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., Diao, Y., Liang, J., Zhao, H., Lobanenkov, V.V., Ecker, J.R., Thomson, J., Ren, B., 2015. Chromatin Architecture Reorganization during Stem Cell Differentiation. Nature 518, 331–336. https://doi.org/10.1038/nature14222

Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, Xianglin, He, J., Xiang, Y., Wang, Q., Li, Y., Ma, J., Zhang, Xu, Zhang, K., Wang, Y., Zhang, M.Q., Gao, J., Dixon, J.R., Wang, X., Zeng, J., Xie, W., 2017. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature 547, 232–235. https://doi.org/10.1038/nature23263

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., Aiden, E.L., 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Systems 3, 95–98. https://doi.org/10.1016/j.cels.2016.07.002

Eagen, K.P., 2018. Principles of Chromosome Architecture Revealed by Hi-C. Trends Biochem Sci 43, 469–478. https://doi.org/10.1016/j.tibs.2018.03.006

Forcato, M., Bicciato, S., 2021. Computational Analysis of Hi-C Data, in: Bodega, B., Lanzuolo, C. (Eds.), Capturing Chromosome Conformation: Methods and Protocols, Methods in Molecular Biology. Springer US, New York, NY, pp. 103–125. https://doi.org/10.1007/978-1-0716-0664-3_7

Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Ciofi, C., Crottini, A., Godoy, J.A., Höglund, J., Malukiewicz, J., Mouton, A., Oomen, R.A., Paez, S., Palsbøll, P.J., Pampoulie, C., Ruiz-López, María J., Svardal, H., Theofanopoulou, C., Vries, J. de, Waldvogel, A.-M., Zhang, Guojie, Mazzoni, C.J., Jarvis, E.D., Bálint, M., Formenti, G., Theissinger, K., Fernandes, C., Bista, I., Bombarely, A., Bleidorn, C., Čiampor, F., Ciofi, C., Crottini, A., Godoy, J.A., Hoglund, J., Malukiewicz, J., Mouton, A., Oomen, R.A., Paez, S., Palsbøll, P., Pampoulie, C., Ruiz-López, María José, Svardal, H., Theofanopoulou, C., Vries, J. de, Waldvogel, A.-M., Zhang, Goujie, Mazzoni, C.J., Jarvis, E., Bálint, M., Aghayan, S.A., Alioto, T.S., Almudi, I., Alvarez, N., Alves, P.C., Amorim, I.R., Antunes, A., Arribas, P., Baldrian, P., Berg, P.R., Bertorelle, G., Böhne, A., Bonisoli-Alquati, A., Boštjančić, L.L., Boussau, B., Breton, C.M., Buzan, E., Campos, P.F., Carreras, C., Castro, L.Fi., Chueca, L.J., Conti, E., Cook-Deegan, R., Croll, D., Cunha, M.V., Delsuc, F., Dennis, A.B., Dimitrov, D., Faria, R., Favre, A., Fedrigo, O.D., Fernández, R., Ficetola, G.F., Flot, J.-F., Gabaldón, T., Agius, D.R.G., Gallo, G.R., Giani, A.M., Gilbert, M.T.P., Grebenc, T., Guschanski, K., Guyot, R., Hausdorf, B., Hawlitschek, O., Heintzman, P.D., Heinze, B., Hiller, M., Husemann, M., Iannucci, A., Irisarri, I., Jakobsen, K.S., Jentoft, S., Klinga, P., Kloch, A., Kratochwil, C.F., Kusche, H., Layton, K.K.S., Leonard, J.A., Lerat, E., Liti, G., Manousaki, T., Marques-Bonet, T., Matos-Maraví, P., Matschiner, M., Maumus, F., Cartney, A.M.M., Meiri, S., Melo-Ferreira, J., Mengual, X., Monaghan, M.T., Montagna, M., Mysłajek, R.W., Neiber, M.T., Nicolas, V., Novo, M., Ozretić, P., Palero, F., Pârvulescu, L., Pascual, M., Paulo, O.S., Pavlek, M., Pegueroles, C., Pellissier, L., Pesole, G., Primmer, C.R., Riesgo, A., Rüber, L., Rubolini, D., Salvi, D., Seehausen, O., Seidel, M., Secomandi, S., Studer, B., Theodoridis, S., Thines, M., Urban, L., Vasemägi, A., Vella, A., Vella, N., Vernes, S.C., Vernesi, C., Vieites, D.R., Waterhouse, R.M., Wheat, C.W., Wörheide, G., Wurm, Y., Zammit, G., 2022. The era of reference genomes in conservation genomics. Trends in Ecology & Evolution 37, 197–202. https://doi.org/10.1016/j.tree.2021.11.008

Genome [WWW Document], n.d. . NCBI. URL https://www.ncbi.nlm.nih.gov/datasets/genome/?taxon=4710 (accessed 8.27.23).

Gonçalves, S., Medronho, J., Moreira, E., Grosso, C., Andrade, P.B., Valentão, P., Romano, A., 2018. Bioactive properties of Chamaerops humilis L.: antioxidant and enzyme inhibiting activities of extracts from leaves, seeds, pulp and peel. 3 Biotech 8, 88. https://doi.org/10.1007/s13205-018-1110-9

Gong, H., Yang, Y., Zhang, S., Li, M., Zhang, X., 2021. Application of Hi-C and other omics data analysis in human cancer and cell differentiation research. Comput Struct Biotechnol J 19, 2070–2083. https://doi.org/10.1016/j.csbj.2021.04.016

Gregory, T.R., Nicol, J.A., Tamm, H., Kullman, B., Kullman, K., Leitch, I.J., Murray, B.G., Kapraun, D.F., Greilhuber, J., Bennett, M.D., 2007. Eukaryotic genome size databases. Nucleic Acids Research 35, D332–D338. https://doi.org/10.1093/nar/gkl828

Gros-Balthazard, M., Baker, W.J., Leitch, I.J., Pellicer, J., Powell, R.F., Bellot, S., 2021. Systematics and Evolution of the Genus Phoenix: Towards Understanding Date Palm Origins, in: Al-Khayri, J.M., Jain, S.M., Johnson, D.V. (Eds.), The Date Palm Genome, Vol. 1: Phylogeny, Biodiversity and Mapping, Compendium of Plant Genomes. Springer International Publishing, Cham, pp. 29–54. https://doi.org/10.1007/978-3-030-73746-7_2

Group, C.P.W., 2009. A DNA Barcode for Land Plants. Proceedings of the National Academy of Sciences of the United States of America 106, 12794–12797.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hebert, P.D.N., Cywinska, A., Ball, S.L., deWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc Biol Sci 270, 313–321. https://doi.org/10.1098/rspb.2002.2218

Hoffman, E.A., Frey, B.L., Smith, L.M., Auble, D.T., 2015. Formaldehyde Crosslinking: A Tool for the Study of Chromatin Complexes. J Biol Chem 290, 26404–26411. https://doi.org/10.1074/jbc.R115.651679

Human Genome Project Fact Sheet [WWW Document], n.d. . Genome.gov. URL https://www.genome.gov/about-genomics/educational-resources/fact-sheets/human-genome-project (accessed 7.28.23).

Johanson, T.M., Allan, R.S., 2022. In Situ HiC. Methods Mol Biol 2458, 333–343. https://doi.org/10.1007/978-1-0716-2140-0_18

Juicebox [WWW Document], n.d. URL https://aidenlab.org/juicebox/ (accessed 8.20.23).

Ke, Y., Xu, Y., Chen, X., Feng, S., Liu, Z., Sun, Y., Yao, X., Li, F., Zhu, W., Gao, L., Chen, H., Du, Z., Xie, W., Xu, X., Huang, X., Liu, J., 2017. 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. Cell 170, 367-381.e20. https://doi.org/10.1016/j.cell.2017.06.029

Kim, K., Kim, M., Kim, Y., Lee, D., Jung, I., 2022. Hi-C as a molecular rangefinder to examine genomic rearrangements. Seminars in Cell & Developmental Biology, Special issue: Novel concepts of molecular mechanisms in spermatogenesis by Yan Cheng / Special issue: 3D genome organization, genetic variation and disease by Justin O'Sullivan and Tayaza Fadason 121, 161–170. https://doi.org/10.1016/j.semcdb.2021.04.024

Kong, S., Zhang, Y., 2019. Deciphering Hi-C: from 3D genome to function. Cell Biol Toxicol 35, 15–32. https://doi.org/10.1007/s10565-018-09456-2

Kress, W.J., García-Robledo, C., Uriarte, M., Erickson, D.L., 2015. DNA barcodes for ecology, evolution, and conservation. Trends in Ecology & Evolution 30, 25–35. https://doi.org/10.1016/j.tree.2014.10.008

La gigante nana dell'Orto botanico [WWW Document], 2019. . Il Bo Live UniPD. URL http://ilbolive.unipd.it/it/news/orto-botanico-palma-goethe (accessed 7.10.23).

Laetsch, D.R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies. https://doi.org/10.12688/f1000research.12232.1

Lajoie, B.R., Dekker, J., Kaplan, N., 2015. The Hitchhiker's Guide to Hi-C Analysis:

Practical guidelines. Methods 72, 65–75. https://doi.org/10.1016/j.ymeth.2014.10.031

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997

Li, X., Yang, Y., Henry, R.J., Rossetto, M., Wang, Y., Chen, S., 2015. Plant DNA barcoding: from gene to genome. Biological Reviews 90, 157–166. https://doi.org/10.1111/brv.12104

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., Sandstrom, R., Bernstein, B., Bender, M.A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., Mirny, L.A., Lander, E.S., Dekker, J., 2009. Comprehensive mapping of long range interactions reveals folding principles of the human genome. Science 326, 289–293. https://doi.org/10.1126/science.1181369

Manni, M., Berkeley, M.R., Seppey, M., Simão, F.A., Zdobnov, E.M., 2021. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution 38, 4647–4654. https://doi.org/10.1093/molbev/msab199

Marburger, S., Alexandrou, M.A., Taggart, J.B., Creer, S., Carvalho, G., Oliveira, C., Taylor, M.I., 2018. Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes. Proceedings of the Royal Society B: Biological Sciences 285, 20172732. https://doi.org/10.1098/rspb.2017.2732

Nagano, T., Lubling, Y., Várnai, C., Dudley, C., Leung, W., Baran, Y., Mendelson-Cohen, N., Wingett, S., Fraser, P., Tanay, A., 2017. Cell-cycle dynamics of chromosomal organisation at single-cell resolution. Nature 547, 61–67. https://doi.org/10.1038/nature23001

Nascimento, L.F. do, Guimarães, P.R., Onstein, R.E., Kissling, W.D., Pires, M.M., 2020. Associated evolution of fruit size, fruit colour and spines in Neotropical palms. Journal of Evolutionary Biology 33, 858–868. https://doi.org/10.1111/jeb.13619

Niakan, K.K., Han, J., Pedersen, R.A., Simon, C., Pera, R.A.R., 2012. Human pre-implantation embryo development. Development 139, 829–841. https://doi.org/10.1242/dev.060426

Niu, L., Shen, W., Huang, Y., He, N., Zhang, Y., Sun, J., Wan, J., Jiang, D., Yang, M., Tse, Y.C., Li, L., Hou, C., 2019. Amplification-free library preparation with SAFE Hi-C uses ligation products for deep sequencing to improve traditional Hi-C analysis. Commun Biol 2, 1–8. https://doi.org/10.1038/s42003-019-0519-y

Onstein, R.E., Vink, D.N., Veen, J., Barratt, C.D., Flantua, S.G.A., Wich, S.A., Kissling, W.D., 2020. Palm fruit colours are linked to the broad-scale distribution and diversification of primate colour vision systems. Proceedings of the Royal Society B: Biological Sciences 287, 20192731. https://doi.org/10.1098/rspb.2019.2731

Pal, K., Forcato, M., Ferrari, F., 2018. Hi-C analysis: from data generation to integration. Biophys Rev 11, 67–78. https://doi.org/10.1007/s12551-018-0489-1

Picard Tools - By Broad Institute [WWW Document], n.d. URL https://broadinstitute.github.io/picard/ (accessed 8.20.23).

Polizzi, G., Castello, I., Vitale, A., Catara, V., Tamburino, V., 2006. First Report of Thielaviopsis Trunk Rot of Date Palm in Italy. Plant Disease 90, 972–972. https://doi.org/10.1094/PD-90-0972C

Robinson, J.T., Turner, D., Durand, N.C., Thorvaldsdóttir, H., Mesirov, J.P., Aiden, E.L., 2018. Juicebox.js Provides a Cloud-Based Visualization System for Hi-C Data. cels 6, 256-258.e1. https://doi.org/10.1016/j.cels.2018.01.001

Schwarz, E.N., Ruhlman, T.A., Sabir, J.S.M., Hajrah, N.H., Alharbi, N.S., Al-Malki, A.L., Bailey, C.D., Jansen, R.K., 2015. Plastid genome sequences of legumes reveal parallel inversions and multiple losses of rps16 in papilionoids. Journal of Systematics and Evolution 53, 458–468. https://doi.org/10.1111/jse.12179

Singh, R., Ong-Abdullah, M., Low, E.-T.L., Manaf, M.A.A., Rosli, R., Nookiah, R., Ooi, L.C.-L., Ooi, S.-E., Chan, K.-L., Halim, M.A., Azizi, N., Nagappan, J., Bacher, B., Lakey, N., Smith, S.W., He, D., Hogan, M., Budiman, M.A., Lee, E.K., DeSalle, R., Kudrna, D., Goicoechea, J.L., Wing, R.A., Wilson, R.K., Fulton, R.S., Ordway, J.M., Martienssen, R.A., Sambanthamurthi, R., 2013. Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds. Nature 500, 335–339. https://doi.org/10.1038/nature12309

Tenaillon, M.I., Hollister, J.D., Gaut, B.S., 2010. A triptych of the evolution of plant transposable elements. Trends in Plant Science 15, 471–478. https://doi.org/10.1016/j.tplants.2010.05.003

The IUCN Red List of Threatened Species [WWW Document], n.d. . IUCN Red List of Threatened Species. URL https://www.iucnredlist.org/en (accessed 7.28.23).

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P.R., Bleidorn, C., Bombarely, A., Crottini, A., Gallo, G.R., Godoy, J.A., Jentoft, S., Malukiewicz, J., Mouton, A., Oomen, R.A., Paez, S., Palsbøll, P.J., Pampoulie, C., Ruiz-López, M.J., Secomandi, S., Svardal, H., Theofanopoulou, C., Vries, J. de, Waldvogel, A.-M., Zhang, G., Jarvis, E.D., Bálint, M., Ciofi, C., Waterhouse, R.M., Mazzoni, C.J., Höglund, J., Aghayan, S.A., Alioto, T.S., Almudi, I., Alvarez, N., Alves, P.C., Rosario, I.R.A. do, Antunes, A., Arribas, P., Baldrian, P., Bertorelle, G., Böhne, A., Bonisoli-Alquati, A., Boštjančić, L.L., Boussau, B., Breton, C.M., Buzan, E., Campos, P.F., Carreras, C., Castro, L.Fi.C., Chueca, L.J., Čiampor, F., Conti, E., Cook-Deegan, R., Croll, D., Cunha, M.V., Delsuc, F., Dennis, A.B., Dimitrov, D., Faria, R., Favre, A., Fedrigo, O.D., Fernández, R., Ficetola, G.F., Flot, J.-F., Gabaldón, T., Agius, D.R., Giani, A.M., Gilbert, M.T.P., Grebenc, T., Guschanski, K., Guyot, R., Hausdorf, B., Hawlitschek, O., Heintzman, P.D., Heinze, B., Hiller, M., Husemann, M., Iannucci, A., Irisarri, I., Jakobsen, K.S., Klinga, P., Kloch, A., Kratochwil, C.F., Kusche, H., Layton, K.K.S., Leonard, J.A., Lerat, E., Liti, G., Manousaki, T., Marques-Bonet, T., Matos-Maraví, P., Matschiner, M., Maumus, F., Cartney, A.M.M., Meiri, S., Melo-Ferreira, J., Mengual, X., Monaghan, M.T., Montagna, M., Mysłajek, R.W., Neiber, M.T., Nicolas, V., Novo, M., Ozretić, P., Palero, F., Pârvulescu, L., Pascual, M., Paulo, O.S., Pavlek, M., Pegueroles, C., Pellissier, L., Pesole, G., Primmer, C.R., Riesgo, A., Rüber, L., Rubolini, D., Salvi, D., Seehausen, O., Seidel, M., Studer, B., Theodoridis, S., Thines, M., Urban, L., Vasemägi, A., Vella, A., Vella, N., Vernes, S.C., Vernesi, C., Vieites, D.R., Wheat, C.W., Wörheide, G., Wurm, Y., Zammit, G., 2023. How genomics can help biodiversity conservation. Trends in

Genetics 39, 545–559. https://doi.org/10.1016/j.tig.2023.01.005

Tomás-Daza, L., Rovirosa, L., López-Martí, P., Nieto-Aliseda, A., Serra, F., Planas-Riverola, A., Molina, O., McDonald, R., Ghevaert, C., Cuatrecasas, E., Costa, D., Camós, M., Bueno, C., Menéndez, P., Valencia, A., Javierre, B.M., 2023. Low input capture Hi-C (liCHi-C) identifies promoter-enhancer interactions at high-resolution. Nat Commun 14, 268. https://doi.org/10.1038/s41467-023-35911-8

Twelve years of SAMtools and BCFtools | GigaScience | Oxford Academic [WWW Document], n.d. URL https://academic.oup.com/gigascience/article/10/2/giab008/6137722 (accessed 8.20.23).

van Berkum, N.L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L.A., Dekker, J., Lander, E.S., 2010. Hi-C: A Method to Study the Three-dimensional Architecture of Genomes. J Vis Exp 1869. https://doi.org/10.3791/1869

vgp-assembly/pipeline/salsa/arima_mapping_pipeline.sh at master · VGP/vgp-assembly [WWW Document], n.d. . GitHub. URL https://github.com/VGP/vgp-assembly/blob/master/pipeline/salsa/arima_mapping_pipeline.sh (accessed 8.10.23).

Wang, L., Lee, M., Yi Wan, Z., Bai, B., Ye, B., Alfiko, Y., Rahmadsyah, R., Purwantomo, S., Song, Z., Suwanto, A., Hua Yue, G., 2022. Chromosome-level Reference Genome Provides Insights into Divergence and Stress Adaptation of the African Oil Palm. Genomics, Proteomics & Bioinformatics. https://doi.org/10.1016/j.gpb.2022.11.002

Xu, M., Guo, L., Gu, S., Wang, O., Zhang, R., Peters, B.A., Fan, G., Liu, X., Xu, X., Deng, L., Zhang, Y., 2020. TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience 9, giaa094. https://doi.org/10.1093/gigascience/giaa094

Yaish, M.W., Patankar, H.V., Assaha, D.V.M., Zheng, Y., Al-Yahyai, R., Sunkar, R., 2017. Genome-wide expression profiling in leaves and roots of date palm (Phoenix dactylifera L.) exposed to salinity. BMC Genomics 18, 246. https://doi.org/10.1186/s12864-017-3633-6

Yu, M., Ren, B., 2017. The Three-Dimensional Organization of Mammalian Genomes. Annu Rev Cell Dev Biol 33, 265–289. https://doi.org/10.1146/annurev-cellbio-100616-060531

Zhou, C., McCarthy, S.A., Durbin, R., 2023. YaHS: yet another Hi-C scaffolding tool. Bioinformatics 39, btac808. https://doi.org/10.1093/bioinformatics/btac808