First complete mitogenome assembly of Castanea sativa: structure, comparative genomics, and phylogeny

First complete mitogenome assembly of *Castanea sativa*: structure, comparative genomics, and phylogeny

Sara Villa[1], Alexis Marchesini[2,3], Sara Torre[1], Luca Bianco[4], Paolo Fontana[4], Concetta De Quattro[1], Mirko Moser[4], Stefano Piazza[4], Sara Alessandri[5], Vera Pavese[6], Paola Pollegioni[2,3], Cristiano Vernesi[4], Mickael Malnoy[4], Daniela Torello Marinoni[6], Sergio Murolo[7], Luca Dondini[5], Claudia Mattioni[2,3], Roberto Botta[6], Diego Micheletti[4], Luisa Palmieri[4], Federico Sebastiani[1]*

* Corresponding author: federico.sebastiani@cnr.it

Affiliations and addresses:

[1] Institute for Sustainable Plant Protection, National Research Council, via Madonna del Piano 10, 50019, Sesto Fiorentino, Italy

[2] Research Institute on Terrestrial Ecosystems (IRET), National Research Council of Italy (CNR), via Marconi 2, Porano, Terni 05010, Italy

[3] NBFC, National Biodiversity Future Center, Piazza Marina 61, Palermo 90133, Italy

[4] Research and Innovation Centre, Edmund Mach Foundation, via Mach 1, San Michele all'Adige, TN, 38098, Italy

[5] Department of Agricultural and Food Sciences, University of Bologna, Via Zamboni 33, Bologna, BO, 40126, Italy

[6] Department of Agricultural, Forest and Food Science, University of Turin, Largo P. Braccini 2, Grugliasco, TO, 10095, Italy

[7] Department of Agricultural, Food and Environmental Science, Marche Polytechnic University, via Brecce Bianche, Ancona, AN, 60131, Italy

Sara Villa ORCID ID: 0000-0002-9238-9839

Sara Torre ORCID ID: 0000-0002-3496-4057

Federico Sebastiani ORCID ID: 0000-0003-4676-7381

Alexis Marchesini ORCID ID: 0000-0001-9487-5737

Paola Pollegioni ORCID ID 0000-0001-6388-1931

Abstract

The sweet chestnut (*Castanea sativa* Mill.) is one of the most widespread cultivated temperate trees in Europe, valued both for its edible nuts and high-quality timber. Due to its ecological, economic, and cultural importance, it has been the focus of extensive genetic studies. However, its mitochondrial genome has remained largely unexplored. Here, we present the first complete mitochondrial genome of *C. sativa* (cultivar 'Marrone di Chiusa Pesio'), assembled using high-throughput sequencing and characterised through comparative analyses with closely related species. The final assembly consists of six contigs with a total length of 402,729 bp, comprising 35 protein-coding genes, 33 tRNA genes, and 3 rRNA genes. Compared to the congeneric *C. mollissima*, *C. sativa* shows a relatively low repeat content in terms of both number and length. Codon usage patterns were found to be highly similar among *C. sativa*, *C. mollissima*, and *C. henryi*. Additionally, homologous fragments between the plastid and mitochondrial genomes were identified, totaling 4,671 bp (1.16% of the mitogenome), and including several tRNA genes. A phylogenetic analysis based on mitochondrial coding sequences from *C. sativa* and 11 other Fagaceae species confirmed its close relationship with *C. mollissima*, *C. henryi*, and *Castanopsis carlesii*. Discrepancies observed among mitochondrial, plastid, and nuclear gene trees likely reflect either inherent genomic characteristics or extensive hybridisation, particularly within the genus Quercus.

Keywords: *Castanea sativa* Mill., gene transfer, mitochondrial genome, organellar DNA, phylogenetic analysis, sweet chestnut.

Key Message: Complete assembly of *Castanea sativa* mitogenome is presented, along with comparative and phylogenetic analyses.

Introduction

The sweet chestnut *Castanea sativa* Mill. (Fagaceae) is the only species of the genus native to Europe, where it grows in temperate deciduous forests. The species has a noticeable economic value as an important source of food and high-quality timber across centuries (Conedera *et al*. 2004). Along with other congeneric species, *C. sativa* is one of the most widely cultivated temperate tree species in the northern hemisphere, as also suggested by the specific epithet (*sativa* means "cultivated by humans"). It is native of Southern Europe, Asia Minor and Transcaucasia (Krebs *et al.* 2019), with an East-West range extending from the Northern Iberian Peninsula to the Caspian Sea; however, its current range

has been strongly shaped by direct and indirect anthropic pressure (Conedera *et al.* 2016). Based on palynological data, the spread of sweet chestnut trees is hypothesized to have benefited from large-scale forest clearings for the cultivation of other plant species, as early as 8,600 years before present (B.P.). Direct, human-mediated diffusion of the species, that can be interpreted as evidence of cultivation, dates back to around 3,700 B.P., in several regions of the Anatolian peninsula, North-eastern Greece and South-eastern Bulgaria (Conedera *et al.* 2004). It was, however, during the Middle Ages that the species range started to significantly increase in Europe, as witnessed by evidences of systematic tree planting (Conedera *et al.* 2004; 2016). This close, millenary relationship between humans and sweet chestnut has led to the selection of several cultivars, on which the economy of vast regions in Europe and Turkey is based (https://www.fao.org/faostat/en/#data/QCL). Because of its economic, environmental and cultural value, it is also a target for genetic and genomic studies, along with other Asian and American chestnut species. Recently, the *de novo* assembly at chromosome-level of the nuclear genome of *C. sativa* cv. 'Marrone di Chiusa Pesio' has been published, as the first reported genome assembly of sweet chestnut (Bianco *et al.* 2024), while in 2021, the complete plastid genome of *C. sativa* was released (Xu *et al.* 2021). Nevertheless, a complete assembly of the sweet chestnut mitochondrial genome has never been published to date.

Plant mitogenomes have become a valuable tool for studies in phylogenetics and population genetics (Hiesel *et al.* 1994; Wu *et al.* 2022; Tang *et al.* 2023, Lin *et al*. 2025). The release of plant mitogenomes assemblies has increased exponentially in recent years (with almost 30 papers published in 2024), also thanks to the enhanced availability of dedicated tools (Al-Nakeeb *et al.* 2017; Jin *et al.* 2020; Fischer *et al.* 2022; He *et al.* 2023; Bi *et al.* 2024; Štorchová and Krüger 2024). Among the growing number of bioinformatics tools designed for assembling mitochondrial genomes from whole-genome sequencing (WGS) data, some leverage long-read sequencing. Notably, GSAT (He *et al.* 2022) combines long and short reads, while two recent tools, PMAT (Bi *et al.* 2024) and the more recent Oatk (Zhou *et al.* 2025), are specifically designed for long-read data. However, the complexity of the mitochondrial genome architecture makes *de novo* assembly rather challenging in plants. In contrast to metazoa, plant mitogenomes, especially in angiosperms, can reach considerable sizes, exceeding 2 Mb in some families such as Cucurbitaceae (Alverson *et al.* 2010). Despite their larger size, mitochondrial genomes in plants do not contain significantly more genes than in animals. Mitochondrial genes in land plants encode for tRNAs, rRNAs and proteins involved in the respiratory chain and in the cytochrome c biogenesis (Unseld *et al.* 1997; Burger *et al.* 2003; Morley and Nielsen, 2017). Plant mitogenome size is increased by intron gains, the extension of non-coding intergenic regions and copies of plastid or nuclear sequences, also gained through horizontal gene transfer (Bergthorsson *et al.* 2003; Richardson and Palmer 2007). Mitochondrial non-coding regions undergo

frequent rearrangements, and are supposed to play a role in genomic functions, although this still remains a subject of debate (Han *et al.* 2024; Wang *et al.* 2024). Interestingly, mitochondrial genome size can vary considerably in angiosperms, ranging from 66 kb in *Viscum scurruloideum* Barlow to over 11.3 Mb in *Silene conica* L. (Sloan *et al.* 2012; Skippington *et al.* 2015). This heterogeneity can also be found between strongly related species: within the same genus *Silene*, differences of more than four orders of magnitude can be observed, such as between the same *Silene conica* and *Silene latifolia* Poir. (253 kb; Sloan *et al.* 2010). In flowering plants, mitochondrial genomes can assume a variety of alternative configurations, *e.g*. linear, circular, sigma-like, fragmented, branched or multi-chromosomal, and are frequently characterised by a multipartite organisation and high rates of recombination among repeated sequences, which are common and varied (Kitazaki and Kubo 2010; Mower *et al.* 2012; Chen *et al.* 2017; Gualberto and Newton 2017). For these reasons, a 'master-circle' representation in a single circular molecule is likely a stretch and a simplification of the real complexity of the plant mitogenome (Bendich 1993).

Among Fagaceae, complete mitogenomes were assembled and published for *Fagus sylvatica* (Mader *et al.* 2020), some oak species (Bi *et al.* 2019, Liu *et al.* 2022), *Castanopsis carlesii* (Tu *et al.* 2024), the Chinese chestnut *Castanea mollissima* (Guo *et al.* 2023), and Henry's chestnut *Castanea henryi* (Tu *et al.* 2024). Among Fagales, the use of mitochondrial genes for phylogenetic reconstructions is challenging, due to their high conservation level, the plasticity of the genomic structure and the poor inter-specific synteny (Feng *et al.* 2021). Moreover, mitochondrial genes are characterized by high levels of RNA-edited sites, which may contain homoplasious signals leading to artefactual phylogenetic inferences (Bowe and dePamphilis, 1996; Small *et al.* 2020). Finally, mitochondrial genomes in Fagales show evidence of mosaicism due to receptivity to foreign genes from other seed plants (Feng *et al.* 2021). The target genomic region could strongly affect the result of the phylogenetic survey. Phylogenomic conflicts are a widespread and well-known phenomenon, whether considering nuclear rather than organellar datasets, or plastid rather than mitochondrial genomes (Som 2015; Vargas *et al.* 2017; Li *et al.* 2019; Tyszka *et al.* 2023; Xue *et al.* 2024). The aim of the present study was to provide the first complete mitochondrial genome assembly and characterisation for *C. sativa*. Different methods of assembly, based on Illumina short reads and Oxford Nanopore Technologies (ONT) long reads were tested and discussed. Comparative analyses were performed to evaluate key genomic features in *C. sativa* and related species. In addition, we reconstructed the phylogenetic relationships within Fagaceae using protein-coding genes from both the mitochondrial and plastid genomes.

Materials and Methods

The *C. sativa* mitogenome was assembled from the reads of Cv. Marrone Di Chiusa Pesio, publicly available on GenBank (ONT: SRA Accession ID SRX24152490; Bianco *et al.* 2024). ONT reads were corrected and trimmed using Canu v2.2 (Koren *et al.* 2017).

Several assembly strategies were tested (see Supplementary Information for details), but the final mitochondrial genome assembly was obtained using Oatk (Zhou *et al.* 2025). The coverage threshold (-c) was set to 80, corresponding to five times the nuclear sequence coverage, while all other parameters were left at their default values. Based on Bandage visualization (v0.9, Wick *et al.* 2015), only six nodes were retained (u7, u8, u9, u10, u11, and u12).

Annotation of the assembled contigs was performed by using the GeSeq online tool (Tillich *et al.* 2017, available at: https://chlorobox.mpimp-golm.mpg.de/geseq.html), selecting *Fagus sylvatica* mitogenome as NCBI RefSeq. To compile the complete gene composition of *C. sativa* mitogenome, the annotation automatically performed by Oatk was also considered. In particular, the tRNA genes were detected more efficiently by the Oatk annotation than the GeSeq annotation. Thus, tRNA genes annotated directly by Oatk but not detected by GeSeq were added to the gene list. Since the Oatk annotation does not distinguish complete gene sequences from fragments, only the sequences with length in accordance with the reference sequence were retained. Assembled mitogenome was visualised both with the OGDRAW online tool (Greiner *et al.* 2019, available at: https://chlorobox.mpimp-golm.mpg.de/OGDraw.html) and Bandage (Wick *et al.* 2015).

Relative Synonym Codon Usage (RSCU) in mitochondrial protein-coding sequences was calculated for *C. sativa*, *C. mollissima* and *C. henryi*. The assembled mitogenome of the three chestnut species was imported in Phylosuite (Zhang *et al.* 2020) and protein-coding sequences were extracted, aligned, manually trimmed and concatenated. Finally, RSCU for each species was calculated by using MEGA11 (Tamura *et al.* 2021) and visualised with *ggplot2* (Wickham 2016). To compare the synonymous codon usage in *C. sativa* and other congeneric species, differences in RSCU values for all codons in *C. sativa*, *C. mollissima* and *C. henryi* were calculated. Finally, differences between RSCU values found in *C. mollissima* and results from Guo *et al.* (2023) were calculated to compare results from the same assembled mitogenome.

Simple sequence repeats (SSRs) were identified by using the MISA web-service with default settings (Thiel *et al.* 2003; Beier *et al.* 2017). Tandem repeats longer than 6 nucleotides were identified by using Tandem Repeats Finder with default settings (Benson 1999). Finally, dispersed repeats (forward, reverse, palindromic and complement) were detected by using REPuter, setting minimal repeat size = 20, Hamming distance = 2 and e-value $\geq 10^{-3}$ (Kurtz *et al.* 2001). The number of repeats of each class was finally visualised with *ggplot2* (Wickham 2016).

Homologous DNA fragments between *C. sativa* plastid and mitochondrial genome were detected with the BLASTN online tool (Zhang *et al.* 2000; last accessed 20 February 2025), setting the assembled mitochondrial contigs as query sequences and the complete plastid genome (GenBank ID: NC_054204) as search set. Homologous regions were visualised by using Circos (Krzywinski *et al.* 2009) implemented in the online platform Galaxy (The Galaxy Community 2024).

Phylogenetic analyses

The phylogenetic trees were built with sequences of *C. sativa* and other species of Fagaceae available on GenBank, also including *Betula pendula* as outgroup (see Table 1 for the list of species and GenBank IDs). The plastid genome of *Q. ilex* was assembled from raw reads available on GenBank (accession ID: ERX13410334) by using GetOrganelle (Jin *et al.* 2020).

Annotated mitochondrial and plastid genome sequences were imported in Phylosuite (Zhang *et al.* 2020; Xiang *et al.* 2023) directly from GenBank when available; otherwise, were previously annotated by using GeSeq, setting *Fagus sylvatica* as reference for mitochondrial genome, and *C. henryi*, *C. mollissima*, *F. sylvatica*, *Q. acutissima*, *Q. robur*, *Q. variabilis* and *B. pendula* var. 'carelica' as reference for plastid genome. Protein coding sequences for both the organellar genomes were extracted, setting the standard code and the Bacterial, Archaeal and Plant plastid code as code table for mitochondrion and chloroplast, respectively. rRNAs, tRNAs and non-coding regions were excluded. Coding sequences common to all the 13 species were used to construct the phylogenetic trees. The sequences of each gene were aligned by using MEGA11 (ClustalW option; Tamura *et al.* 2021) and manually edited. In particular, for the plastid gene NAD1 two partial sequences were retained, discarding the central part with alignment mismatches. Sequences of each organelle were concatenated, imported in Phylosuite, and trimmed with trimAL (*no-gaps* trimming option; Capella-Gutiérrez *et al.* 2009). Maximum likelihood phylogenetic trees were calculated with IQ-Tree v2.2.0 selecting *B. pendula* as outgroup and leaving the other settings as default ('*auto'* option and 5,000 ultrafast bootstrap; Minh *et al.* 2013; Nguyen *et al.* 2015). Finally, the output treefile was visualised with iTOL v.7.0 (last access February 2025; Ciccarelli *et al.* 2006; Letunic and Bork 2024)

Results

Although different assembly strategies were tested (see Supplementary Information), all produced assemblies with the same predicted coding capacity. The assembly obtained with Oatk was

the most accurate, as it had the lowest number of contigs - indicating larger and more continuous sequences - and a total length (402,729 bp) very close to that of the *Castanea mollissima* mitogenome.

The mitogenome of *C. sativa* was assembled into six contigs and deposited in the European Nucleotide Archive (ENA, accession number: ERZ26184703). The six contigs have a total length of 402,729 bp (average coverage = 180×), with single molecules spanning for 183,905 bp (u7), 37,466 bp (u8), 33,834 bp (u9), 86,453 bp (u10), 46,386 bp (u11) and 14,685 bp (u12). Gene annotation found 56 genes, including 35 protein-coding genes, 33 tRNA genes (of which 22 present in multiple copies) and 3 rRNA genes (Fig. 1, Table 2).

Analysis of codon usage in protein-coding genes revealed differences in the relative abundance of synonym codons, but consistent results in the three chestnut species under investigation (Fig. 2). The stop codon TAA exhibited the highest RSCU value (2.1), followed by GCT (alanine, RSCU = 1.79). The less represented codon was TAG (stop, RSCU = 0). Moreover, the preference for NNA and NNT codons with respect to NNG and NNC is evident in almost all the aminoacids. Similar RSCU values for all codons were found in the three analysed *Castanea* species, with differences in RSCU values ranging from 0 to 0.08. A difference of 0.08 in the RSCU values was due to the higher and lower usage of the stop codons TAA and TGA, respectively, in *C. mollissima* compared to the other species. As for *C. mollissima*, the comparison between the RSCU values found in the present work and in Guo *et al.* (2023) revealed generally a slight difference, comprised between 0 and 0.26, except for the two stop codons TAA and TAG, with the first one being found over-represented in our analysis (RSCU = 2.18) and the second one under-represented (RSCU = 0).

Analysis of SSRs revealed 5 monomeric and 9 dimeric repeats (Fig. 3). Analysis of tandem repeats revealed the presence of 11 repeats, with length ranging from 15 bp to 24 bp. Finally, the most abundant class of repeats was represented by dispersed repeats, distributed as follows: 40 palindromic, 25 forward and 2 reverse. No complementary repeats were found. Length of dispersed repeats ranged from 22 to 918 bp.

The analysis of homologous regions revealed 15 fragments homologous between the mitochondrial and plastid genome, for a total length of 4,671 bp (1.16% of the entire mitogenome; Fig. 4; Supplementary Information Table S2). Fragments length spanned from 29 to 1,149 bp, and comprised complete sequences of 7 tRNA genes (*trnD-GUC*, *trnH-GUG*, *trnI-CAU*, *trnM-CAU*, *trnN-GUU*, *trnP-UGG*, *trnW-CCA*), in addition to the protein-coding genes: *petL*, *petG*, *ndhC*, and the incomplete sequences of *psbC*, *psbD*, *psbE*, *ndhF*, *ndhK*, *rps11*, *rrn16*, *ycf1*.

Phylogenetic analysis

The phylogenetic tree based on mitochondrial protein-coding sequences was constructed considering 12 genes shared by the 13 target species as extracted by Phylosuite (*atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cytb*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*). Best substitution model according to BIC was TIM+F+I (BIC = 36,700.2; AICc = 36,478.3). The resulting tree showed a sharp distinction among *B. pendula*, *F. sylvatica* and species of the subfamily Quercoideae, comprising the genera *Quercus*, *Lithocarpus*, *Castanea* and *Castanopsis* (Fig. 5A). Within this branch, two main clusters were observed: the first comprising *Q. petraea*, *Q. robur*, *Q. ilex* and *L. litseifolius*, and the second comprising oak species of the subgenus *Cerris* (*Q. cerris*, *Q. variabilis* and *Q. acutissima*), *Castanopsis carlesii* and species of the genus *Castanea*. However, this partition was poorly supported, as indicated by the ultrafast bootstrap support percentages of the two branches (44 and 42%, respectively). Within the first group, a further partition separated *Q. petraea* and *Q. robur* (98% bootstrap support) from *Q. ilex* and *L. litseifolius*. Again, this latter branch was poorly supported (40% bootstrap support). Within the second group, the branch comprising oak species of the subgenus *Cerris* is well supported (97% bootstrap support), while the branch with chestnut species and *C. carlesii* showed only the 54% of bootstrap support.

The plastid gene tree was built considering 65 genes (*atpA*, *atpF*, *atpH*, *atpI*, *ccsA*, *cemA*, *clpP1*, *matK*, *ndhA*, *ndhB*, *ndhD*, *ndhE*, *ndhG*, *ndhH*, *ndhI*, *pafI*, *pafII*, *pbf1*, *petA*, *petB*, *petD*, *petG*, *petL*, *petN*, *psaA*, *psaB*, *psaC*, *psaJ*, *psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbT*, *psbZ*, *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl23*, *rpl32*, *rpl33*, *rpl36*, *rpoA*, *rpoB*, *rpoC1*, *rpoC2*, *rps2*, *rps3*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps19*, *ycf1*, *ycf2*). Best substitution model according to BIC was TVM+F+G4 (BIC = 184,678.5; AICc = 184,403.2). The resulting tree showed again a sharp separation between *B. pendula*, *F. sylvatica* and the Quercoideae species (Fig. 5B). Among Quercoideae, a highly supported branch (100% bootstrap support) grouped *Q. ilex*, *Q. robur* and *Q. petraea* together, while a less supported branch (71% bootstrap support) comprised the other species. This latter branch exhibited a further partition with oaks of the subgenus *Cerris* again grouped together (100% bootstrap support) and separated from *Castanea* spp., *Castanopis carlesii* and *L. litseifolius* (84% bootstrap support). Finally, *L. litseifolius* was separated from *Castanea* spp. and *Castanopis carlesii*, further grouped in the following couples: *C. mollissima* with *C. henryi* and *C. sativa* with *C. carlesii* (100% bootstrap support, in both cases).

Discussion

Although often depicted as a circular molecule ('master circle model'), plant mitochondrial DNA is highly complex and dynamic, existing in multiple forms even within a single individual (Sloan 2013). The challenges associated with assembling plant mitochondrial genomes arise not only from the presence of repeats, which complicate bioinformatic analyses, but also from genuine genomic rearrangements that are best represented as graph-like structures (Sloan 2013; Kozik *et al.* 2019). For these reasons, we considered the output of the *C. sativa* mitochondrial genome assembly in the form of assembly graphs and focused on genes presence without considering orientations and copy number. Among the tested assembly tools, Oatk demonstrated superior performance in *C. sativa* compared to other software.

Mitochondrial genome size was found to be slightly greater in the sweet chestnut, *C. sativa* (402,729 bp) than in the Chinese chestnut, *C. mollissima* (388,038 bp; Guo *et al.* 2023), but still smaller than in *Fagus sylvatica* (504,715 bp; Mishra *et al.* 2021). Gene annotation revealed high similarity between the mitochondrial genomes of *C. sativa* and *C. mollissima*. In both species, the rps10 gene sequence is interrupted by multiple internal stop codons, raising questions about its functional status. Although annotated as present in *C. mollissima*, the similar sequence pattern in *C. sativa* suggests that rps10 may be truncated or non-functional in both genomes.

Relative synonymous codon usage (RSCU) analysis showed that synonymous codons did not occur with equal frequencies, as evident for example in valine or leucine, with codons GTT and CTT represented almost 4 and 6 times more than the synonymous GTG and CTG, respectively. Codon usage bias, the non-random use of synonymous codons, arises from the interplay of natural selection, mutation pressure, and genetic drift, and its pattern is often species-specific (Sharp *et al*. 1986; Novoa and Ribas de Pouplana 2012). However, as expected in closely related species, the three Castanea species examined here (*C. sativa*, *C. mollissima*, and *C. henryi*) exhibited very similar RSCU values, suggesting a conserved codon usage trend (Tang *et al*. 2021). The preference for -A and -T ending codons is a common feature in plant mitochondrial genomes and, being more pronounced in algae and some bryophytes and less in vascular plants, is supposed to have an evolutionary meaning (Xu *et al.* 2015). This bias also affects stop codons, with TAA (and TGA) usually favoured over TAG across plant taxa (Xu *et al.* 2015). The 'NNA' and 'NNT' pattern was found to improve the efficiency of transcription and translation during gene expression, at least in plastid and nuclear genome (Klump *et al.* 1993; Tang *et al.* 2021). Finally, the recorded, slight differences between the RSCU values found in *C. mollissima* by Guo *et al.* (2023), and those found in *C. sativa* in the present work, were probably consequences of the manual editing phase, apparently lacking in Guo *et al.* (2023).

Compared to *C. mollissima* (Guo *et al.* 2023), *C. sativa* mitogenome showed less and shorter SSRs). Among the dispersed repeats, palindromic repetitive elements were the most abundant type both in Chinese and sweet chestnut, although in the first were three times more abundant than in the latter (n = 129 *versus* 40). Forward repeats were also widespread in *C. mollissima* (n = 119), while in *C. sativa* only 25 were found. In both species, no complementary repeats were found. Contrary to the Chinese chestnut, dispersed repeats found in sweet chestnut were rather short (20-25 bp) and thus unlikely to be subjected to recombination (Wynn and Christensen 2019). Duplications and repeats are partly responsible for the large size of mitogenome in angiosperms, together with frequent DNA sequence capture from the chloroplast and, less commonly, nuclear genome through horizontal transfer events (Marienfeld *et al.* 1999; Palmer *et al.* 2000).

Gene transfer from plastid and nuclear genome to mitochondrial genome is facilitated by the readiness of these organelles to capture DNA from external sources, primarily due to their highly dynamic nature and recombination activity (Koulintchenko *et al.* 2003). Transfer of tRNA genes from chloroplast to mitochondrial genome is a common phenomenon in angiosperms (Michaud *et al.* 2011; Yong *et al.* 2024). The overall length of fragments migrated from plastome to mitogenome, as well as the percentage represented in the entire mitogenome was found to be slightly lower in the sweet chestnut (1.16%) than in the Chinese chestnut (1.49%; Guo *et al.* 2023). All the tRNA genes found in *C. sativa* mitogenome (*trnD-GUC*, *trnH-GUG*, *trnI-CAU*, *trnM-CAU*, *trnN-GUU*; Guo *et al.* 2023) were also found in *C. mollissima*, in addition to *trnP-UGG* and *trnW-CCA*. We also found *petL* and *petG* genes, encoding for subunits of cytochrome b6-f complex and *ndhC*, as recorded in the mitogenome of some monocot species (Lin *et al.* 2015). Finally, incomplete mitochondrial sequences of both tRNA and protein-coding genes were found in *C. sativa*. In particular, *rps10* gene was present with a truncated sequence, characterised by the presence of several stop codons. This is probably the result of the transfer of this gene from the mitochondrion to the nucleus, as observed in other land plants (Wischmann and Schuster 1995; Kan *et al.* 2021).

Phylogenetic trees based on organellar genes confirmed a closer relationship between the genera *Castanopsis* and *Castanea*, in accordance with the phylogeny inferred by Zhou *et al.* (2022) based on whole genome sequencing, although the latter analysis did not include *C. sativa* (Fig. 5). However, in our plastome-based phylogeny *Castanea sativa* showed a stronger affinity with *Castanopsis carlesii*, than with the two congeneric species (Fig. 5B). Organellar gene trees also agreed in the topology of *Quercus* subgenus *Cerris* (*Q. cerris, Q. acutissima, Q. variabilis*), more related to chestnuts and *Castanopsis carlesii* than to other *Quercus* species. Conversely, a clear distinction between the *Quercus* clade and the *Castanea-Castanopsis* clade is evident from Zhou *et al.* (2022).

Finally, *L. litseifoilus* was found to be more related to *Q. ilex* (although the branch is poorly supported) and separated from chestnuts when considering mitochondrial genes, but it resulted to be closely related to chestnuts and separated from oaks when considering plastid genes. Conflicts between mitochondrial and plastid phylogenies, and the deviation of these from the nuclear tree (*i.e.*, cytonuclear discordance) are common in Fagaceae (Zhou *et al.* 2022), particularly for the genus *Quercus*, and may be the result of high hybridisation and introgression rates leading to a reticulate evolution (Hipp *et al.* 2020), as observed also in other angiosperms families, such as Asteraceae (Vargas *et al.* 2017). These interspecific evolutionary dynamics add complexity to the already expected discordance between phylogenies obtained from genomic regions characterised by different inheritance and molecular evolution patterns. Discrepancies between the mitochondrial and the plastid tree may indeed be driven by the inherent features of the two genomes: the different substitution rates, the heterogeneous substitution rate within plastome and RNA editing in mitogenome (Bowe and dePamphilis 1996; Liu *et al.* 2014; Petersen *et al.* 2016; Wicke *et al.* 2016). Mitochondrial genes exhibit lower substitution rates - and thus lower homoplasy -, and are likely more effective in resolving deep and ancient relationships (Qiu *et al.* 2010; Richardson *et al.* 2013), while plastid genes are more suitable in contexts of relatively recent speciation events. As a consequence, discrepancies between nuclear and organellar gene trees are not surprising as they had been found in a variety of plant taxa, both monocots and dicots (Vargas *et al.* 2017; Li *et al.* 2019; Wang *et al.* 2023; Xue *et al.* 2024).

Conclusions

Mitogenome of the sweet chestnut was assembled and characterised for the first time. Mitochondrial DNA was assembled in six contigs covering a total length of 402,729 bp, containing 35 protein-coding genes, 32 tRNA genes and 3 rRNA genes. Comparative analyses of codon usage and repeated regions showed similarities between the mitogenome of *C. sativa* and other congeneric species. As expected and already observed in other flowering plants, homologous regions between plastome and mitogenome included mainly tRNA genes. Phylogenetic analyses based on both mitochondrial and plastid genes were consistent with the nuclear gene tree as regards the grouping of *Castanea/Castanopsis* taxa and the *Quercus* subgenus *Cerris* species, but some discrepancies were found in oaks topology. Conflicts in phylogenomic trees highlight the complex relationships within the family of Fagaceae, and especially among oaks, more prone to hybridisation and introgression events.

Funding: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript. A.M., C.M. and P.P. were funded by the National Recovery and

Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union - Next Generation EU; Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP B83C22002930006, Project title “National Biodiversity Future Center - NBFC".

Competing Interest: The authors have not conflict of interest to declare.

Author contributions: FS conceived and designed research. FS and SV conducted data analyses. SV wrote the manuscript. All authors read, edited and approved the manuscript.

Data Availability: The *Castanea sativa* mitogenome assembly is available at ENA (Accession number: ERZ26184703).

Supplementary Information

Table S1. Comparison of the results obtained from the three assembly methods

Table S2. Homologous regions in *C. sativa* mitochondrial and plastid genome

References:

Al-Nakeeb K, Petersen T, Sicheritz-Pontén T (2017) Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics* 18:510. https://doi.org/10.1186/s12859-017-1927-y

Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* 27(6):1436-1448. https://doi.org/10.1093/molbev/msq029

Arnold J. Bendich AJ (1993) Reaching for the ring: the study of mitochondrial genome structure. *Current Genetics* 24:279-290. https://api.semanticscholar.org/CorpusID:35106898

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. *Bioinformatics* 33(16):2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research* 27(2): 573-580. https://doi.org/10.1093/nar/27.2.573

Bergthorsson U, Adams K, Thomason B *et al* (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature* 424:197-201. https://doi.org/10.1038/nature01743

Bi Q, Li D, Zhao Y *et al* (2019) Complete mitochondrial genome of *Quercus variabilis* (Fagales, Fagaceae). *Mitochondrial DNA Part B* 4(2):3927-3928. https://doi.org/10.1080/23802359.2019.1687027

Bi C, Shen F, Han F *et al* (2024) PMAT: an efficient plant mitogenome assembly toolkit using low-coverage HiFi sequencing data. *Horticulture Research* 11(3):uhae023. https://doi.org/10.1093/hr/uhae023

Bianco L, Fontana P, Marchesini A *et al* (2024) The de novo, chromosome-level genome assembly of the sweet chestnut (*Castanea sativa* Mill.) Cv. Marrone Di Chiusa Pesio. *BMC Genomic Data* 25:64. https://doi.org/10.1186/s12863-024-01245-7

Bowe LM, dePamphilis CW (1996) Effects of RNA editing and gene processing on phylogenetic reconstruction. *Molecular Biology and Evolution* 13:1159-1166. https://doi.org/10.1093/oxfordjournals.molbev.a025680

Burger G, Gray MW, Lang BF. 2003. Mitochondrial genomes: anything goes. *Trends in Genetics* 19(12):709-716. https://doi.org/10.1016/j.tig.2003.10.012

Capella-Gutiérrez, Silla-Martínez JM, Gabaldón T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25(15):1972-1973. https://doi.org/10.1093/bioinformatics/btp348

Chen Z, Zhao N, Li S *et al* (2017) Plant mitochondrial genome evolution and cytoplasmic male sterility. *Critical Reviews in Plant Sciences* 36:1-15. https://doi.org/10.1080/07352689.2017.1327762.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* 311(5765):1283-1287. https://www.science.org/doi/10.1126/science.1123061

Conedera M, Krebs P, Tinner W, Pradella M, Torriani D (2004) The cultivation of *Castanea sativa* (Mill.) in Europe, from its origin to its diffusion on a continental scale. *Vegetation History and Archaeobotany* 13(3):161-179. https://doi.org/10.1007/s00334-004-0038-7.

Conedera M, Tinner W, Krebs P, de Rigo D, Caudullo G (2016) *Castanea sativa* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G,

Houston Durrant T, Mauri A (eds.), *European Atlas of Forest Tree Species*. Publ. Off. EU, Luxembourg, pp. e0125e0+

Feng Y, Xiang X, Akhter D, Pan R, Fu Z, Jin X (2021) Mitochondrial phylogenomics of Fagales provides insights into plant mitogenome mosaic evolution. *Frontiers in Plant Science Sec. Plant Cell Biology* 12. https://doi.org/10.3389/fpls.2021.762195

Fischer A, Dotzek J, Walther D, Greiner S (2022) Graph-based models of the *Oenothera* mitochondrial genome capture the enormous complexity of higher plant mitochondrial DNA organization. *NAR Genomics and Bioinformatics* 4(2):lqac027. https://doi.org/10.1093/nargab/lqac027

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research* 47:W59-W64. https://doi.org/10.1093/nar/gkz238

Gualberto JM, Newton KJ (2017) Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annual Review of Plant Biology* 68:225-252. https://doi.org/10.1146/annurev-arplant-043015-112232

Guo H, Liu Q, Chen Y *et al* (2023) Comprehensive assembly and comparative examination of the full mitochondrial genome in *Castanea mollissima* Blume. *Genomics* 115(6):110740. https://doi.org/10.1016/j.ygeno.2023.110740

Han F, Bi C, Zhao Y, Gao M, wang Y, Chen Y (2024) Unraveling the complex evolutionary features of the *Cinnamomum camphora* mitochondrial genome. *Plant Cell Reports* 43:183. https://doi.org/10.1007/s00299-024-03256-1

He W, Xiang K, Chen C, Wang J, Wu Z (2023) Master graph: an essential integrated assembly model for the plant mitogenome based on a graph-based framework. *Briefings in Bioinformatics* 24(1): bbac522. https://doi.org/10.1093/bib/bbac522

Hiesel R, von Haeseler A, Brennicke A (1994) Plant mitochondrial nucleic acid sequences as a tool for phylogenetic analysis. *Proceedings of the National Academy of Sciences of the United States of America* 91:634–638. https://doi.org/10.1073/pnas.91.2.634

Hipp AL, Manos PS, Hahn M *et al* (2020) Genomic landscape of the global oak phylogeny. *New Phytologist* 226(4):1198-1212. https://doi.org/10.1111/nph.16162

Jin JJ, Yu WB, Yang JB *et al* (2020) GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biology* 21:241. https://doi.org/10.1186/s13059-020-02154-5

Kan SL, Shen TT, Ran JH *et al* (2021) Both Conifer II and Gnetales are characterized by a high frequency of ancient mitochondrial gene transfer to the nuclear genome. *BMC Biology* 19:146. https://doi.org/10.1186/s12915-021-01096-z

Kitazaki K, Kubo T (2010) Cost of having the largest mitochondrial genome: evolutionary mechanism of plant mitochondrial genome. *Journal of Botany* 2010:620137. https://doi.org/10.1155/2010/620137

Klump HH (1993) Correlation between genome size, observed codon preference and Gibbs energy of codon–anticodon interaction. *Pure and Applied Chemistry* 65(9):1947–1950. https://doi.org/10.1351/PAC199365091947

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM (2017) Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Resources* 27(5):722-736. http://www.genome.org/cgi/doi/10.1101/gr.215087.116.

Koulintchenko M, Konstantinov Y, Dietrich A (2003) Plant mitochondria actively import DNA via the permeability transition pore complex. *The EMBO Journal* 22(6):1245-1254. https://doi.org/10.1093/emboj/cdg128

Kozik A, Rowan BA, Lavelle D *et al* (2019) The alternative reality of plant mitochondrial DNA: One ring does not rule them all. *PLOS Genetics* 15(8):e1008373. https://doi.org/10.1371/journal.pgen.1008373

Krebs P, Pezzatti GB, Beffa G, Tinner W, Conedera M (2019) Revising the sweet chestnut (*Castanea sativa* Mill.) refugia history of the last glacial period with extended pollen and macrofossil evidence. *Quaternary Science Reviews* 206:111-128. https://doi.org/10.1016/j.quascirev.2019.01.002

Krzywinski M, Schein J, İnanç Birol Connors J *et al* (2009) Circos: An information aesthetic for comparative genomics. *Genome Research* 19(9):1639-1645. https://doi.org/10.1101/gr.092759.109

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Resources* 29(22):4633-4642. https://doi.org/10.1093/nar/29.22.4633

Letunic I, Bork P (2024) Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* 52(W1):W78-W82. https://doi.org/10.1093/nar/gkae268

Li Y-X, Li Z-H, Schuiteman A *et al* (2019) Phylogenomics of Orchidaceae based on plastid and mitochondrial genomes. *Molecular Phylogenetic and Evolution* 139:106540. https://doi.org/10.1016/j.ympev.2019.106540

Lin CS, Chen J, Huang YT *et al* (2015) The location and translocation of *ndh* genes of chloroplast origin in the Orchidaceae family. *Scientific Report* 5:9040. https://doi.org/10.1038/srep09040

Lin, D., Shao, B., Gao, Z. *et al* (2025) Phylogenomics of angiosperms based on mitochondrial genes: insights into deep node relationships. *BMC Biology* 23 (45). https://doi.org/10.1186/s12915-025-02135-9

Liu Y, Cox CJ, Wang W, Goffinet B (2014) Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* 63(6):862-878. https://doi.org/10.1093/sysbio/syu049

Liu D, Guo H, Zhu J *et al* (2022) Complex physical structure of complete mitochondrial genome of *Quercus acutissima* (Fagaceae): A significant energy plant. *Genes* 13(8):1321. https://doi.org/10.3390/genes13081321

Mader M, Schroeder H, Schott T *et al* (2020) Mitochondrial genome of *Fagus sylvatica* L. as a source for taxonomic marker development in the Fagales. *Plants* 9(10):1274. https://doi.org/10.3390/plants9101274

Marienfeld J, Unseld M, Brennicke A (1999) The mitochondrial genome of *Arabidopsis* is composed of both native and immigrant information. *Trends in Plant Science* 4(12):495-502. https://doi.org/10.1016/S1360-1385(99)01502-2

Michaud M, Cognat V, Duchêne A-M, Maréchal-Drouard L (2011) A global picture of tRNA genes in plant genomes. *The Plant Journal* 66:80-93. https://doi.org/10.1111/j.1365-313X.2011.04490.x

Minh BQ, Nguyen MA, von Haeseler A (2013) Ultrafast approximation for phylogenetic bootstrap. *Molecular Biology and Evolution* 30(5):1188-1195. https://doi.org/10.1093/molbev/mst024

Mishra B, Ulaszewski B, Meger J, Ploch S, Burczyk J, Thines M (2021) A comparison of three circular mitochondrial genomes of *Fagus sylvatica* from Germany and Poland reveals low

variation and complete identity of the gene space. *Forests* 12(5):571. https://doi.org/10.3390/f12050571

Morley SA, Nielsen BL (2017) Plant mitochondrial DNA. *Frontiers in Bioscience* (Landmark Ed) 22(6):1023-1032. https://doi.org/10.2741/4531

Mower JP, Sloan D, Alverson A (2012) Plant mitochondrial genome diversity: the genomics revolution. In: Wendel J, Greilhuber J, Dolezel J, Leitch I (eds.), *Plant Genome Diversity* Volume 1. Springer, Vienna, pp 123-144. https://doi.org/10.1007/978-3-7091-1130-7_9

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* 32(1):268-274. https://doi.org/10.1093/molbev/msu300

Novoa EM, Ribas de Pouplana L (2012) Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics* 28(11):574-581. http://dx.doi.org/10.1016/j.tig.2012.07.006

Palmer JD, Adams KL, Cho Y, Song K (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* 97(13):6960-6966. https://doi.org/10.1073/pnas.97.13.6960

Petersen G, Seberg O, Davis JI *et al* (2006) Mitochondrial data in monocot phylologenetics. *Aliso* 22(1):52-62. https://scholarship.claremont.edu/aliso/vol22/iss1/5

Qiu Y-L, Li L, Wang B (2010) Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48(6):391-425. https://doi.org/10.1111/j.1759-6831.2010.00097.x

Richardson AO, Palmer JD (2007) Horizontal gene transfer in plants. *Journal of Experimental Botany* 58(1):1-9. https://doi.org/10.1093/jxb/erl148

Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD (2013) The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* 11:29. https://doi.org/10.1186/1741-7007-11-29

Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. *Journal of Molecular Evolution* 24:28-38. https://doi.org/10.1007/BF02099948 (1986).

Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all *nad* genes. *Proceedings of the National Academy of Sciences of the United States of America* 112:E3515–E3524. https://doi.org/10.1073/pnas.1504491112

Sloan DB (2013) One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. *New Phytologist* 200(4):978-985. https://doi.org/10.1111/nph.12395

Sloan DB, Alverson AJ, Chuckalovcak JP *et al* (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLOS Biology* 10(1):e1001241. https://doi.org/10.1371/journal.pbio.1001241

Sloan DB, Alverson AJ, Štorchová H *et al* (2010) Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. *BMC Evolutionary Biology* 10:274. https://doi.org/10.1186/1471-2148-10-274

Small ID, Schallenberg-Rudinger M, Takenaka M, Mireau H, Ostersetzer-Biran O (2020) Plant organellar RNA editing: what 30 years of research has revealed. *The Plant Journal* 101(5):1040-1056. https://doi.org/10.1111/tpj.14578

Som A (2015) Causes, consequences and solutions of phylogenetic incongruence. *Briefings in Bioinformatics* 16(3):536-548. https://doi.org/10.1093/bib/bbu015

Štorchová H, Krüger M (2024) Methods for assembling complex mitochondrial genomes in land plants. *Journal of Experimental Botany* 75(17):5169-5174. https://doi.org/10.1093/jxb/erae034

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38(7):3022-3027. https://doi.org/10.1093/molbev/msab120

Tang D, Wei F, Cai Z *et al* (2021) Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. *Development Genes and Evolution* 231(1-2):1-9. https://doi.org/10.1007/s00427-020-00670-9

Tang J-Y, Wei R, Zhang X-C, Xiang Q-P (2023) Mitogenome-based phylogenomics provides insights into the positions of the enigmatic *sinensis* group and the *sanguinolenta* group in Selaginellaceae (Lycophyte). *Molecular Phylogenetics and Evolution* 179:107673. https://doi.org/10.1016/j.ympev.2022.107673

560 The Galaxy Community (2024) The Galaxy platform for accessible, reproducible, and collaborative
561 data analyses: 2024 update. *Nucleic Acids Research* 2024:gkae410,
562 https://doi.org/10.1093/nar/gkae410

563 Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development
564 and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theoretical*
565 *and Applied Genetics* 106(3):411-422. https://doi.org/10.1007/s00122-002-1031-0

566 Tillich M, Lehwark P, Pellizzer T *et al* (2017) GeSeq – versatile and accurate annotation of organelle
567 genomes. *Nucleic Acids Research* 45(W1):W6-W11. https://doi.org/10.1093/nar/gkx391

568 Tu XD, Xin YX, Fu HH *et al* (2024) The complete mitochondrial genome of *Castanopsis carlesii*
569 and *Castanea henryi* reveals the rearrangement and size differences of mitochondrial DNA
570 molecules. *BMC Plant Biology* 24:988. https://doi.org/10.1186/s12870-024-05618-z

571 Tyszka AS, Bretz EC, Robertson HM, *et al* (2023) Characterizing conflict and congruence of
572 molecular evolution across organellar genome sequences for phylogenetics in land plants.
573 *Frontiers in Plant Science Sec. Bioinformatics* 14:1125107.
574 https://doi.org/10.3389/fpls.2023.1125107

575 Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis*
576 *thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics* 15(1):57-61.
577 https://doi.org/10.1038/ng0197-57

578 Vargas OM, Ortiz EM, Simpson BB (2017) Conflicting phylogenomic signals reveal a pattern of
579 reticulate evolution in a recent high-Andean diversification (Asteraceae: Astereae:
580 *Diplostephium*). *New Phytologist* 214(4):1736-1750. https://doi.org/10.1111/nph.14530.

581 Wang MT, Hou Z-Y, Li C *et al* (2023) Rapid structural evolution of *Dendrobium* mitogenomes and
582 mito-nuclear phylogeny discordances in *Dendrobium* (Orchidaceae). *Journal of Systematics*
583 *and Evolution* 61(5):790-805. https://doi.org/10.1111/jse.12912

584 Wang J, Kan S, Liao X *et al* (2024) Plant organellar genomes: much done, much more to do. *Trends*
585 *in Plant Science* 29(7):754-769. https://doi.org/10.1016/j.tplants.2023.12.014

586 Wick RR, Schultz MB, Zobel J, Holt KE (2015) Bandage: interactive visualisation of de novo genome
587 assemblies. *Bioinformatics* 31(20):3350-3352. https://doi.org/10.1093/bioinformatics/btv383

588 Wicke S, Müller KF, dePamphilis CW, Schneeweiss GM (2016) Mechanistic model of evolutionary
589 rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National*

590 *Academy of Sciences of the United States of America* 113(32):9045-9050.
591 https://orcid.org/0000-0003-2811-3317

592 Wickham H (2016) *ggplot2: elegant graphics for data analysis*. Springer-Verlag New York.
593 https://ggplot2.tidyverse.org.

594 Wischmann C, Schuster W (1995) Transfer of *rps10* from the mitochondrion to the nucleus in
595 *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration
596 site. *FEBS Letters* 374(2):152-156. https://doi.org/10.1016/0014-5793(95)01100-S

597 Wu Z-Q, Liao X-Z, Zhang X-N, Tembrock LR, Broz A (2022) Genomic architectural variation of
598 plant mitochondria—A review of multichromosomal structuring. *Journal of Systematics and*
599 *Evolution* 60(1):160-168. https://doi.org/10.1111/jse.12655

600 Wynn EL, Christensen AC (2019) Repeats of unusual size in plant mitochondrial genomes:
601 identification, incidence and evolution. *G3 Genes, Genomes, Genetics* 9(2):549-559.
602 https://doi.org/10.1534/g3.118.200948

603 Xiang C-Y, Gao F, Jakovlić I *et al* (2023) Using PhyloSuite for molecular phylogeny and tree-based
604 analyses. *iMeta* 2(1):e87. https://doi.org/10.1002/imt2.87

605 Xu W, Xing T, Zhao M *et al* (2015) Synonymous codon usage bias in plant mitochondrial genes is
606 associated with intron number and mirrors species evolution. *PLOS ONE* 10(6):e0131508.
607 https://doi.org/10.1371/journal.pone.0131508

608 Xu H, Zhang Q, Wang G, Xiang K, Liu Q, Chen X (2021) The complete chloroplast genome of
609 *Castanea sativa*, an endemic to Europe. *Mitochondrial DNA B Resources* 6(4):1300-1302.
610 https://doi.org/10.1080/23802359.2021.1907246

611 Xue T-T, Janssens SB, Liu B-B, Yu S-X (2024) Phylogenomic conflict analyses of the plastid and
612 mitochondrial genomes via deep genome skimming highlight their independent evolutionary
613 histories: A case study in the cinquefoil genus *Potentilla sensu lato* (Potentilleae, Rosaceae).
614 *Molecular Phylogenetics and Evolution* 190:107956.
615 https://doi.org/10.1016/j.ympev.2023.107956

616 Yong Y, Hu S, Zhong M *et al* (2024) Horizontal gene transfer from chloroplast to mitochondria of
617 seagrasses in the yellow–Bohai seas. *Genomics* 116(5):110940
618 https://doi.org/10.1016/j.ygeno.2024.110940.

Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Molecular Ecology Resources* 20(1):348-355. https://doi.org/10.1111/1755-0998.13096

Zhang Z, Schwartz S, Wagner L, Miller W (2000) A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology* 7(1-2):203-214. https://doi.org/10.1089/10665270050081478

Zhou C, Brown M, Blaxter M, The Darwin Tree of Life Project Consortium, McCarthy SA, Durbin R (2025) Oatk: a de novo assembly tool for complex plant organelle genomes. https://doi.org/10.1101/2024.10.23.619857

List of captions

Fig. 1 Visualisation of mitogenome assembly of *Castanea sativa*. (a) Mastercircle representation with gene annotation obtained from OGDRAW (Greiner *et al.* 2019) and (b) Bandage visualisation of the six contigs (u7, u8, u9, u10, u11, u12; Wick *et al.* 2015)

Fig. 2 Random Synonym Codon Usage (RSCU) in *Castanea sativa, C. henryi* and *C. mollissima*. RSCU values are reported on the *x*-axis, while codons and their respective amino-acids are listed on the *y*-axis. Amino-acid names followed the three-letter code

Fig. 3 Sequence repeats abundance. Abundance of sequence repeats found in *C. sativa* mitogenome, grouped as Simple Sequence Repeats (SSR), Tandem repeats or Dispersed repeats

Fig. 4 Homologous regions between plastid and mitochondrion. Homologous regions between the plastome (cp, in green) and the six mitochondrial contigs (mt_u7, mt_u8, mt_u9, mt_u10, mt_u11, mt_u12, in blue) obtained during the Oatk assembly. Red arcs link homologous regions and labels indicate genes (both complete and incomplete) identified through annotation, with position along the genome highlighted by bands. To make the diagram more readable, segments from 70 to 80 kb in the plastome and from 75 to 85 kb in mt_u7 have been zoomed in by a factor of 7. Fragments details are reported in Supplementary Information Table S2

Fig. 5 Phylogenetic trees. Mitochondrial (a) and plastid (b) ML gene tree of 12 Fagaceae species based on the coding sequences of conserved plastid protein-coding genes. Numbers on each node indicate ultrafast bootstrap support (%). *Betula pendula* was included as outgroup

Table 1. List of species included in gene trees and relative accession ID

Species	GenBank accession ID	
	chloroplast	mitochondrion
Castanea sativa Mill.	NC_054204	assembled *de-novo*
Castanea mollissima Blume	NC_014674	OP895669, OP895670
Castanea henryi Rehder & E.H.Wilson	NC_033881	PP856681, PP856682, PP856683
Castanopsis carlesii (Hemsl.) Hayata	NC_057119	PP853255
Lithocarpus litseifolius Chun	NC_063927	NC_065018
Quercus robur L.	MG678035	OW028777
Quercus cerris L.	OY770020	OY770018, OY770019
Quercus ilex L.	assembled *de-novo*	OZ205167
Quercus petraea (Matt.) Liebl.	LT996899	OZ066324
Quercus variabilis Blume	NC_031356	CP129458
Quercus acutissima Carruth.	NC_039429	MZ636519
Fagus sylvatica L.	NC_041437	OZ125618
Betula pendula Roth	NC_072281	LT855379

Table 2. List of annotated genes in *C. sativa* mitochondrial genome. Genes are grouped by function, numbers in brackets indicate the number of copies for each gene, and 'x' indicate the node where genes were found. Asterisks indicate tRNA found only by Oatk direct annotation

Group of genes	**Gene name**	**u7 (183,905 bp)**	**u8 (37466 bp)**	**u9 (33,834 bp)**	**u10 (86,453 bp)**	**u11 (46,386 bp)**	**u12 (14,685 bp)**
ATP synthase	*atp1* (3)	x		x	x		
	atp4			x			
	atp6	x					
	atp8	x					
	atp9 (2)	x				x	
NADH dehydrogenase	nad1 (2)	x		x			
	nad2 (2)	x	x				

	nad3 (2)	x				x	
	nad4 (2)	x					x
	nad4L			x			
	nad5 (2)	x				x	
	nad6 (2)			x	x		
	nad7	x					
	nad9	x					
Cytochrome c biogenesis	*cob*	x					
Ubiquinol cytochrome c reductase	*ccmB*				x		
	ccmC (2)	x					x
	ccmFc		x				
	ccmFn (2)	x			x		
Cytochrome c oxidase	*cox1*	x					
	cox2					x	
	cox3	x					
Maturases	*matR*	x					
Transport membrane protein	*mttB* (2)		x	x			
Large subunit of ribosome	*rpl5*	x					
	rpl10				x		
	rpl16		x				
Small subunit of ribosome	*rps1*				x		
	rps3		x				
	rps4				x		
	rps12	x					
	rps14	x					
	rps19 (3)	x	x				x
Succinate dehydrogenase	*sdh3*	x					
	sdh4	x					
Ribosome RNA	*rrn5*					x	
	rrnS					x	
	rrnL (2)	x				x	
Transfer RNA	*trnA-UGC* (4)*	x*		x*	x*		x*
	trnC-GCA				x		
	trnD-GUC (2)					x	x
	trnE-UUC				x		
	trnF-GAA (2)					x	
	trnG-GCC				x		x*
	*trnG-UCC**				x*	x*	
	trnH-GUG				x		
	trnI-CAU (4)*	x*	x*		x*		x*
	trnI-UAU (2)	x					x
	trnK-UUU				x		

	trnL-CAA (2)*					x*	x*
	trnL-UAA (2)*		x*		x*		
	trnM-CAU (2)		x		x		
	trnN-GUU (2)				xx		
	trnP-GGG (3)*	x*			x*	x*	
	trnP-UGG (3)	x		x		x	
	trnQ-UUG				x		
	trnR-ACG (4)*	x*			x*	x*	x*
	trnR-CCG (2)*	x*			x*		
	trnR-UCG (4)*	x*			x*	x*	x*
	trnR-UCU (4)*	x*			x*	x*	x*
	trnS-CGA (4)*	x*		x	x*		x*
	trnS-CGA (3)*	x*			x*		x*
	trnS-GCU					x	
	trnS-GGA (4)*	x*	x*		x*	x*	
	trnS-UGA (2)	x					x
	trnT-GGU (4)*	x*			x*	x*	x*
	trnT-UGU (3)	x			x*	x*	
	trnV-GAC				x*		
	trnV-UAC				x*		
	trnW-CCA	x					
	trnY-GUA (2)		x		x		

659

660