

Citation: Bandaranayake PCG, Naranpanawa N, Chandrasekara CHWMRB, Samarakoon H, Lokuge S, Jayasundara S, et al. (2023) Chloroplast genome, nuclear ITS regions, mitogenome regions, and Skmer analysis resolved the genetic relationship among *Cinnamomum* species in Sri Lanka. PLoS ONE 18(9): e0291763. https://doi.org/ 10.1371/journal.pone.0291763

Editor: Elvira Hörandl, Georg-August-Universitat Gottingen, GERMANY

Received: January 23, 2023

Accepted: September 5, 2023

Published: September 20, 2023

Copyright: © 2023 Bandaranayake et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are included in Tables, Figures and Supporting Information files.

Funding: PCG is the PI of the grant NSF SP/CIN/ 2016/01 funded by the Ministry of Primary Industries and Social Empowerment through the National Science Foundation of Sri Lanka (<u>http://</u> <u>www.nsf.ac.lk/</u>) under the special Cinnamon project –Molecular And Biochemical RESEARCH ARTICLE

Chloroplast genome, nuclear ITS regions, mitogenome regions, and Skmer analysis resolved the genetic relationship among *Cinnamomum* species in Sri Lanka

Pradeepa C. G. Bandaranayake¹¹*, Nathasha Naranpanawa^{1,2}², C. H. W. M. R. Bhagya Chandrasekara¹, Hiruna Samarakoon¹, S. Lokuge¹, S. Jayasundara¹, Asitha U. Bandaranayake³, D. K. N. G. Pushpakumara⁴, D. Siril A. Wijesundara⁵

Faculty of Agriculture, Agricultural Biotechnology Centre, University of Peradeniya, Peradeniya, Sri Lanka,
Postgraduate Institute of Science, University of Peradeniya, Peradeniya, Sri Lanka,
Faculty of Peradeniya, Sri Lanka,

Engineering, Department of Computer Engineering, University of Peradeniya, Peradeniya, Sri Lanka,

4 Faculty of Agriculture, Department of Crop Science, University of Peradeniya, Peradeniya, Sri Lanka,

5 National Institute of Fundamental Studies, Kandy, Sri Lanka

So These authors contributed equally to this work.

¤ Current address: School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, Queensland, Australia

* pradeepag@agri.pdn.ac.lk, pgunathilake@ucdavis.edu

Abstract

Cinnamomum species have gained worldwide attention because of their economic benefits. Among them, C. verum (synonymous with C. zeylanicum Blume), commonly known as Cey-Ion Cinnamon or True Cinnamon is mainly produced in Sri Lanka. In addition, Sri Lanka is home to seven endemic wild cinnamon species, C. capparu-coronde, C. citriodorum, C. dubium, C. litseifolium, C. ovalifolium, C. rivulorum and C. sinharajaense. Proper identification and genetic characterization are fundamental for the conservation and commercialization of these species. While some species can be identified based on distinct morphological or chemical traits, others cannot be identified easily morphologically or chemically. The DNA barcoding using rbcL, matK, and trnH-psbA regions could not also resolve the identification of Cinnamomum species in Sri Lanka. Therefore, we generated Illumina Hiseq data of about 20x coverage for each identified species and a C. verum sample (India) and assembled the chloroplast genome, nuclear ITS regions, and several mitochondrial genes, and conducted Skmer analysis. Chloroplast genomes of all eight species were assembled using a seedbased method. According to the Bayesian phylogenomic tree constructed with the complete chloroplast genomes, the C. verum (Sri Lanka) is sister to previously sequenced C. verum (NC_035236.1, KY635878.1), C. dubium and C. rivulorum. The C. verum sample from India is sister to C. litseifolium and C. ovalifolium. According to the ITS regions studied, C. verum (Sri Lanka) is sister to C. verum (NC_035236.1), C. dubium and C. rivulorum. Cinnamomum verum (India) shares an identical ITS region with C. ovalifolium, C. litseifolium, C. citriodorum, and C. capparu-coronde. According to the Skmer analysis C. verum (Sri Lanka) is sister to C. dubium and C. rivulorum, whereas C. verum (India) is sister to C. ovalifolium, and C. litseifolium. The chloroplast gene ycf1 was identified as a chloroplast barcode for the

Characterizations Of Sri Lankan Cinnamon And Their Wild Relatives And Expression Analysis Of Major Biochemical Genes Under Different Environmental Conditions And Plant Parts To Enhance Utilization Value Of Cinnamon In Sri Lanka. Grant No: NSF SP/CIN/2016/01. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

identification of *Cinnamomum* species. We identified an 18 bp indel region in the *ycf1* gene, that could differentiate *C. verum* (India) and *C. verum* (Sri Lanka) samples tested.

Introduction

The genus *Cinnamomum* Schaeff. of Lauraceae consists of about 247 species (https://powo. science.kew.org/taxon/urn:lsid:ipni.org:names:328262-2#children), found in Asiatic mainland to Formosa, the Malaysian region, northeastern Australia and some Pacific Islands [1,2]. It is a pantropical genus comprised of evergreen trees and shrubs. Biogeographical analysis revealed *Cinnnamomum* originated in the widespread boreotropical paleoflora of Laurasia during the early Eocene (ca. 55 Ma) [3]. *Cinnamomum* species are macromorphologically characterized by buds perulate or not, leaves alternate or opposite, pinnately veined or tripliveined, domatia presence in axil of lateral veins and floral characters such as inflorescences paniculate with cymes bearing strictly opposite lateral flowers. Recent characterization efforts include leaf epidermal micromorphology, having a reticulate periclinal wall or non-reticulate periclinal wall and pollen morphology [4,5].

Cinnamomum verum J. Presl (Synonymous with C. zeylanicum Blume), C. aromaticum Nees and C. camphora (L.) J. Presl (= Camphora officinarum Nees) are commercially traded worldwide, while several other species such as Cinnamomum burmannii (Nees & T. Nees) Blume (Indonesian cinnamon) and Cinnamomum loureiroi Nees (Vietnamese cinnamon) [6], and C. tamala T. Nees and Eberm (India and Nepal) [7] also have some economic benefits. Cinnamomum verum, recognized as Ceylon cinnamon or true cinnamon in the world market, has recently gained special attention due to scientific evidence of its medicinal benefits [8,9]. Historical evidence suggests that C. verum has first been identified in the natural rainforests in the upcountry region of Sri Lanka and brought to cultivation after 1500 AC [10]. Apart from that, Sri Lankan rainforests are home to seven endemic wild species of cinnamon. They are C. capparu-coronde Blume, C. citriodorum Thwaites, C. dubium Nees, C. litseifolium Thwaites, C. ovalifolium Wight, C. rivulorum Kosterm, and C. sinharajaense Kosterm [11,12]. Among them, species such as C. sinharajaense, C. rivulorum and C. dubium are restricted to specific environments, while others such as C. litseifolium and C. ovalifolium are naturally grown in several agro-ecological zones. The morphology of some species drastically changes when they are grown or cultivated under other agroecological conditions [13,14]. In addition, the natural cross-pollination behavior [5,15] of Cinnamomum has created considerable intraspecies diversity [16], reflected by both morphological and biochemical traits [17]. Therefore, morphology based identification of *Cinnamomum* species is challenging [3,18–20].

Molecular biological tools have been used for studying the inter-species diversity of several *Cinnamomum* species, including a few in Sri Lanka [21–33]. Among them, few early studies depended on PCR amplification and sequence analysis, Randomly Amplified Polymorphic DNA (RAPD), and sequence-related amplified polymorphism (SRAP) [23,33]. While some authors could fully resolve the phylogeny of *Cinnamomum* species [23], others could not reach the expected results [33]. DNA barcoding has also been successfully used for molecular identification of some *Cinnamomum* species [34–37]. However, recent work showed that the universal barcoding regions *rbcL*, *matK*, and *trnH-psbA* do not have sufficient polymorphism for the clear identification of *Cinnamomum* species found in Sri Lanka [25]. Similarly, several other authors have also showed that chloroplast genes, *trnL-trnF*, *trnT-trnL*, *psbA-trnH*, *rpl16*, *matK*, and nuclear DNA were not powerful to resolve the phylogenomic relationships of

Lauraceae family members [<u>38–41</u>]. Nevertheless, correct species identification is critical for conservation and sustainable utilization and industrial applications.

Chloroplast genome is used as an ultra-barcode in recent phylogenomic studies because of the advancement of next-generation sequencing technology, the improvement of sequence assembly software, and the reduction of sequencing costs [34–36]. Scientists have assessed both coding and non-coding regions of the plastome [42]. In addition, the nuclear ribosomal Internal Transcribed Spacer regions (ITS) are also utilized widely [43]. Skmer is a sample separation tool that uses genome skimming data without assembling or aligning sequences [44]. Illumina data are utilized in recent studies for assembling both chloroplast and mitochondrial regions [45]. Therefore, we utilized 20x coverage Illumina Hiseq data to assemble chloroplast genomes, a few mitochondrial regions, nuclear ITS regions, and Skmer analysis to resolve the genetic relationship among endemic *Cinnamomum* species in Sri Lanka. We also included several *C. verum* samples from India and publicly available sequencing data in the analysis. We carefully assessed the data to identify a suitable region for PCR-based identification of closely related *Cinnamomum* species.

Materials and methods

Our analysis includes the chloroplast genome, ITS regions, several mitogenome regions, and the Skmer procedure including randomly selected 500,000 reads of all *Cinnamomum* species present in Sri Lanka. The wet lab data were generated to confirm some results of bioinformatics analysis. We present the method under subtopics including the relationship among considered species based on our data and the publicly available sequencing data.

Sample collection, DNA extraction, sequencing

Sample collection. The Research Committee of the Department of Wildlife Conservation, Sri Lanka, and the Department of Forestry, Sri Lanka, granted permission to collect wild *Cinnamomum* species from the rainforests in Sri Lanka and the permission was granted from the Department of Export Agriculture to collect cultivated *C. verum* (Sri Lanka) samples from the respective research stations. Furthermore, all the samples were collected according to relevant institutional, national, and international guidelines and legislation.

Nine cultivated *C. verum* (Sri Lanka) accessions previously identified as different from each other based on morphological and biochemical traits [17] were collected from a vegetatively propagated plantation at Nillambe, Sub Research Station, and germplasm collection at the National Cinnamon Research and Training Center, Thihagoda, Palolpitiya Department of Export Agriculture (DEA).

Seven endemics wild *Cinnamomum* species were collected from the rainforests, and the germplasm collections at the National Cinnamon Research and Training Center, Thihagoda, Palolpitiya, and Mid Country Research Station, Dalpitiya, DEA. The identity of the collected wild samples was verified using typical morphological characters by Siril A. Wijesundara. Sample collection details are given (S1 Table). The collected specimens were further verified with the voucher specimens at the National Herbarium, (PDA), the Royal Botanic Gardens, Peradeniya. The standard herbarium specimens were prepared by mounting on herbarium sheets and deposited at the National Herbarium (PDA) the Royal Botanic Gardens, Peradeniya as previously described [25]. Some morphological, biochemical, and molecular traits of considered samples are included in recent publications [16,19,20,25].

In addition, we included two authentic *C. verum* (India) bark samples and a market sample from India in the analysis. Altogether, a total of twenty *Cinnamomum* samples, ten *C. verum*

(Sri Lanka), seven wild species and three *C. verum* (India) samples were included in the analysis.

DNA extraction. Total genomic DNA from all *C. verum* (Sri Lanka) samples and *C. verum* (India) was extracted using Promega Wizard® Genomic DNA Purification kit (Cat. No: A1120) following the manufacturers' guidelines. Total genomic DNA from all the wild samples was extracted using the Cetyltrimethylammonium bromide (CTAB) method [46] with modifications as previously described [25,47]. The extracted DNA samples were re-suspended in 50 μ L nuclease-free water, quantity and quality assessed with a NanoDrop spectrophotometer (NanoDrop 2000, Thermo Scientific) and running on 0.8% agarose gels before storing at 4°C.

Chloroplast DNA sequencing

A total of 1 µg DNA from each sample was sent to Admera Health, USA, for chloroplast genome sequencing using Illumina Hiseq with 20-30x coverage per sample. The DNA quality and quantity were tested with a Qubit fluorometer (ThermoFisher) and TapeStation system (Agilent). The DNA Libraries were prepared using KAPA Hyper Prep kit (Roche, Switzerland) and sequenced on an Illumina Hiseq platform with a read length of 2 x 150 bp paired-end while giving 80–90 M paired-end reads per sample.

Chloroplast genome assembly and analysis

Chloroplast genome assembly. Using NOVOPlasty assembler (version 2.7.2) [40] we first assembled each dataset into contigs. The assembler used *Zea mays* chloroplast gene of the large subunit of RUBP (V00171.1) as a seed to jump-start the assembly. This method extends the seed iteratively in both directions by adding overlapping reads from the dataset.

We configured assembly parameters such as the insert size and read length of the dataset according to sequencing specifications, and set the kmer size to 25. As per the assembler developer's instructions, we only trimmed the adapters from the raw data and did not filter any low-quality reads before assembly. For each dataset, NOVOPlasty filtered the chloroplast reads from the total DNA and generated a set of contigs. We imported each set of contigs into Geneious (version 11.0.5) [48] and assembled them into a consensus sequence using the '*De novo* assembly' option and built-in Geneious assembler.

Annotation. We used the GeSeq web [49] annotation tool to annotate the chloroplast genomes. Selected annotation options for a given fasta file included 'Annotate plastid IRs' and ARAGORN software [50] for *de novo* tRNA annotation. As references for BLAST search, we selected the three taxa from the genus *Cinnamomum* from the NCBI nucleotide database: *C. camphora* (NC_035882.1), *C. micranthum* (NC_035802.1), and *C. verum* (NC_035236). The web tool generated several output files including the GenBank format of the annotation and the GFF3 file. When provided with the corrected GenBank file as input, the OGDRAW web tool (https://bio.tools/ogdraw) produced a graphical annotation map of each species.

Considering that the *C. verum* (India) sample was a bark while all other samples were leaves, it was necessary to determine whether there was sufficient chloroplast DNA within the total DNA of all samples. For that, we mapped the total genomic raw reads of each sample to its seed-based chloroplast assembly with the Bowtie2 [51] program embedded in Geneious. The number of mapped reads out of total reads, maximum coverage, and mean coverage of reads were the statistics recorded.

Chloroplast sequence alignments and phylogenomic analysis. A total of 82 complete chloroplast genomes, which included nine *Cinnamomum* samples sequenced and assembled in this study and 73 complete chloroplast data deposited in NCBI, representing 26 species of

Cinnamomum, were aligned using MAFFT v7.450 (algorithm FFT-NS-2) built-in Geneious (version 11.0.6) [52].

The Geneious plugin program MrBayes v3.2.6 [53] was used to perform a phylogenomic analysis. MrBayes is a program for Bayesian inference that uses Markov chain Monte Carlo (MCMC) methods to estimate the posterior distribution of model parameters. Under the default settings, the MCMC process was run for 1,100,000 and the first 100,000 generations were discarded as burn-in. The remaining raw trees were sub-sampled at a frequency of 200 to build a consensus tree to represent the summary of the samples. Then the statistical support for each branch of the consensus tree was calculated using the frequencies of the sampled trees in which a particular branch appeared. Bayesian Posterior Probability (BPP) was calculated as the proportion of sampled trees that contain a specific branch, out of the total number of sampled trees. Finally, the majority-rule consensus tree was generated by grouping the strongly supported branches together. *Ocotea porosa* was selected as the out group because *Ocotea* is one of the largest genera in the Lauraceae (400 spp.) and it has been known to be paraphyletic with respect to most other genera of the New Word Lauraceae for almost 20 years [54,55]. Finally, the majority-rule consensus tree was generated using the raw trees, sub-sampled at a sample frequency of 200th iterations.

Nucleotide diversity analysis. The alignment was examined for nucleotide diversity among the *Cinnamomum* species using the DnaSP v6.12.03 [56] DNA polymorphism analyzer. Nucleotide variability (Pi) was calculated using a sliding window method (window length 600 bp and step size 200 bp).

Simple Sequence Repeat (SSR) analysis. To find perfect Simple Sequence Repeats (SSR), Krait v1.3.3 [57] was used with the minimum number of repeats set to 8, 4, 3, 3 and 3 for mono-, di-, tri-, tetra-, and pentanucleotide SSRs, respectively.

ITS regions assembly

ITS regions extraction. First, we extracted nuclear reads by mapping the total DNA sequence reads of each species to the assembled chloroplast genome using Bowtie2 [51]. The non-mapped reads were considered the nuclear reads assuming the number of reads from the mitochondrial genome was negligible.

The time complexity of SPAdes is proportional to the size of the input data and can be estimated as O (N log N), where N is the total size of the input reads [58]. However, the actual time complexity of SPAdes may be much higher in practice, especially for large and complex genomes or datasets with high levels of sequencing coverage. Therefore, considering the computational time and the RAM capacity required to assemble the reads using SPAdes [58], we randomly selected 5GB reads for the contig assembly. Then, the ITSxpress [59] was executed, taking the assembled contigs as the input. The output sequences were taken as the candidate ITS regions for each *Cinnamomum* species. Each assembled ITS sequence contains an 18S ribosomal RNA gene (rRNA) (Partial sequence), ITS 1, 5.8S rRNA, ITS 2, and 26S rRNA (partial sequence). ITS region for a *C. verum* sample was extracted from the raw data deposited in the NCBI (SRX2990994) [60].

ITS regions validation. We validated the obtained ITS regions using NCBI Blastn with default parameters. The BLAST results for each species ITS regions indicated 100% query coverage and more than 99% per identity for *Cinnamomum verum*. Further, we computationally validated the obtained ITS regions using the ITS2 annotator [61]. This tool uses HMMer to annotate ITS2 regions of eukaryotes with Hidden Markov Models (HMMs). It returns the sequence between the conserved 5.8S and 28S (or 26S) rRNA according to the ITS2 definition.

We annotated ITS2 sequences of each ITS region, and the results confirmed that the assembled ITS regions are adequate for further analyses.

ITS Sequence alignment and phylogenomic analysis. We carried out a multiple sequence alignment for extracted ITS regions (18S ribosomal RNA gene (rRNA) (Partial sequence), ITS 1, 5.8S rRNA, ITS 2, and 28S rRNA (partial sequence) for 10 *Cinnamomum* samples, including nine *Cinnamomum* samples from this study and one from NCBI. Furthermore, we compared the ITS regions of nine *Cinnamomum* samples with additional sequences retrieved from NCBI. Considering the sequence data availability and query coverage, we downloaded sequence data of ITS 1, 5.8S rRNA, and partial sequence of ITS 2 region for additional 38 samples representing 17 *Cinnamomum* species. A total of 50 samples were subjected to multiple sequence alignment, and Bayesian phylogenomic analysis was carried out as mentioned under the above chloroplast sequence alignments and phylogenomic analysis using Geneious (version 11.0.6). *Ocotea porosa* (MF110078, MK507282) was selected as the out group.

Mitochondrial genes-assembly and analysis. The mitochondrial genes *atp1*, *atp6*, and cox1 are highly polymorphic in Silene vulgaris, [62]. Additionally, matR and atpA genes are also used in plant phylogeny work [63]. Therefore, we included these mitochondrial genes in our analysis. Initially, we downloaded the C. camphora atp1 gene (AF197681), Laurus nobilis atp6 gene (AY831985), and C. verum cox1 gene (AY009440), C. camphora matR gene (AF197797) and C. verum atpA gene (AY009415) from NCBI. We mapped the raw reads of nine Cinnamomum samples to the reference gene sequences using a custom script (https:// github.com/AgBC-UoP/mapNclean-nf). This script uses Bowtie2 v2.3.3.1 to align reads and samclip v0.4.0 (https://github.com/tseemann/samclip) to remove clipped reads. Consensus sequences for the read alignments were generated using the Geneious prime software followed by manual adjustment to remove ambiguity codes. During the alignment, any ambiguous bases in our consensus sequences were identified and replaced with the respective base observed in the majority of the samples. Specifically, if a particular nucleotide base was present in at least 80% of the other samples or if every other sample had the same nucleotide base, we replaced the ambiguous base in our consensus sequence with that nucleotide base. The same regions for a C. verum sample were extracted from the raw data deposited in the NCBI (SRX2990994) (60). Then we generated a Neighbor-Joining Consensus Tree for a combined data set of atp6 (631 bp) and cox1 (1415 bp) for 10 Cinnamomum species with the Tamura-Nei genetic distance model and 100 bootstrap replicates for node supports. The *atp6* (631 bp) and cox1 (1415 bp) genes were selected for the Neighbor-Joining Consensus tree construction as they have variable sites among the studied mitochondrial genes. Since there was no publicly available convincing data for selected genes for other Cinnamomum species except for C. verum (SRX2990994), only the previously generated C. verum data (SRX29990994) was included in this analysis.

Skmer analyses. Skmer analysis is a computational technique used for analyzing genomic or metagenomic sequence data [44]. It involves breaking up the sequence data into overlapping k-mers, and then counting the frequency of occurrence of each k-mer in the dataset. The resulting k-mer frequencies can be used to generate an unrooted tree. Here we used the entire skimming dataset of a species without separating reads into nuclear, chloroplast or mitochondrial DNA.

Only 500,000 reads were randomly extracted from each forward and reverse fastq file for each*Cinnamomum* skimming dataset. Then the forward and reverse reads were concatenated to form subsamples of 1,000,000 reads. Then, Skmer v3.0.2 [44] was used to perform an assembly-free and alignment-free c analysis of *Cinnamomum* samples. Only the datasets generated

in the current study were included in the analysis since similar *Cinnamomum* data sets were not available on the public domain.

Primer designing, PCR and Sanger sequencing to amplify highly variable regions in chloroplast

PCR amplification of chloroplast *ycf1* gene regions. Two primer pairs (Table 1) were designed to amplify two regions of the *ycf1* gene identified as highly variable regions in the chloroplast genome based on the nucleotide diversity analysis described above. The PCR amplification was carried out for 10 *C. verum* samples, eight from Sri Lanka and two from India. A total of 30 μ L contained lx PCR buffer, 1.5 mM MgCl₂, 200 mM dNTP (Promega, USA), 0.2 μ M of each primer (Integrated DNA Technologies, Singapore), 100 ng of DNA, 0.8 μ M spermidine, and 1 Unit Go Taq Flexi DNA polymerase (Promega, USA). The PCR cycle consisted of initial denaturation at 94°C for 2 minutes, followed by 35 cycles of 94°C for 1 minute, annealing at 48°C for 30 seconds, elongation at 72°C for 30 seconds, and a final extension at 72°C for 3 minutes.

Products were separated by electrophoresis (5 Vcm⁻¹) on 1.5% agarose gels and stained with safe green (Applied Biological Materials Inc. Canada). The PCR products were shipped to Macrogen Inc (Seoul, South Korea– <u>http://dna.macrogen.com</u>) for Sanger sequencing using the same primers as used for PCR.

ycf1 gene regions sequence alignments and analysis

Chromatograms of the PCR amplified products of highly variable regions were visually inspected using Geneious for sequencing errors, and the 5' and 3' noisy sequences of about 30 bp were removed. The same regions of the seven wild *Cinnamomum* species, *C. verum* (India) and *C. verum* (Sri Lanka) were extracted from the assembled chloroplast genomes.

Sequences of the PCR amplified two *ycf1* gene regions were extracted for *C. aromaticum* (NC_046019) and *C. verum* (NC_035236) samples using the chloroplast sequences deposited in the GenBank. Each region was aligned using Geneious alignment of Geneious Prime Software. The ends were trimmed and joined 846 bp and 794 bp of each region. Sequence divergence of *C. verum* samples collected from Sri Lanka and India were calculated using the Tamura-Nei model of Molecular Evolutionary Genetics Analysis (MEGA-X) software [64,65].

Results

We present the results under subtopics parallel to the methodology including the relationship among considered species based on our data and the publicly available sequencing data.

Table 1. Primers used to amplify highly varying regions of the chloroplast genome.

Primer	Sequence 5' to 3'	Product size bp
Primer set 1		952
803 ycf1 f2 1755 ycf1 r2	ATGTCCGATTCCGTGTAATC TTATCTCAGGCTAGGTGGTC	
Primer set 2	TCTGCTATTGCTATCCGTTC	992
3203 ycf1 f5 4195 ycf1 r5	ATGCTTGCCTAGAGTGTATG	

https://doi.org/10.1371/journal.pone.0291763.t001

Chloroplast genome assembly and analysis

Chloroplast genome assembly. For the chloroplast seed-based assembly using NOVO-Plasty software, data of all species were assembled with consistency (Table 2). Assembly length for all nine assemblies ranged between 152695 bp and 152797 bp without drastic deviations. *Cinnamomum capparu-coronde* was the smallest chloroplast genome (152695 bp), while *C. sinharajaense* was the largest one (152797 bp). The size of the assembled *C. verum* chloroplast genome of the samples from Sri Lanka (152765 bp) is very close to the chloroplast genome of *C. verum* published in NCBI (NC_035236), which is 152766 bp long.

Genome annotation. With GeSEQ and OGDRAW web tools, we were able to annotate all nine (9) chloroplast genomes and visualize the chloroplast genome annotations (Fig 1).

The typical angiosperm chloroplast genome consists of 4 rRNAs, approximately 30 tRNAs, and 80 protein coding genes in its gene content [71]. All *Cinnamomum* assemblies included 36 tRNA genes in each assembly (S2 Table).

Chloroplast sequence alignments and phylogenomic analysis. The distance matrix obtained for the multiple sequence alignment of all 82 *Cinnamomum* samples and *O. porosa* is given in S3 Table. There are only five (5) variable sites between *C. verum* (Sri Lanka) (ON685912) and the already published chloroplast genomes of *C. verum* (NC_035236.1, KY63578.1). However, the sample of *C. verum* from India (ON685911) differs in 387 positions from the sample from Sri Lanka, and in 384 positions from previously published *C. verum* chloroplast genomes in NCBI (NC_035236.1, KY63578.1). There are forty-one (41) polymorphic sites between chloroplast genomes of *C. ovalifolium* (ON685908) and *C. litseifolium* (ON685907). Interestingly, the variable sites between *C. verum* (India) (ON685911) and *C. ovalifolium* (ON685908) are 21. Furthermore, the variable sites between *C. verum* (India) (ON685911) and *C. sinharajaense* (ON685910) are 60, 159, 177 and 254 respectively. The variable sites between *C. rivulorum* (ON685909) and *C. dubium* (ON685906) are 86.

The Bayesian phylogeny tree constructed with the complete chloroplast genomes of 83 samples is given (Fig 2). Most of the branches received the highest posterior probability (pp) value of 1, while the others ranged from 0.64 to 0.99. *Cinnamomum verum* (Sri Lanka) (ON685912) formed a monophyletic group with previously sequenced *C. verum* (NC_035236.1, KY635878.1), *C. pingbienense* (OL943977.1, NC065106.1), *C. kotoense* (NC050346.1,

Species (GenBank acc. no.)	Total reads	Assembled reads	% of assembled cp reads	Number of contigs	N50	Assembly length (bp)	GC%
C. capparu-coronde (ON685904)	112428684	737058	0.66	3	133833	152695	39.1
C. citriodorum (ON685905)	93305002	811122	0.87	3	133892	152714	39.1
C. dubium (ON685906)	167974904	1895208	1.13	3	133881	152785	39.1
C. litseifolium (ON685907)	108723790	685568	0.63	3	133869	152745	39.1
C. ovalifolium (ON685908)	119032250	495094	0.41	3	133868	152708	39.1
C. rivulorum (ON685909)	95762972	781514	0.82	3	133873	152787	39.1
C. sinharajaense (ON685910)	102394356	2540352	2.48	3	133920	152797	39.1
<i>C. verum</i> (India) (ON685911)	107469622	581954	0.54	6	68777	152727	39.1
<i>C. verum</i> (Sri Lanka) (ON685912)	129226378	1649924	1.28	3	133852	152765	39.1

Table 2. Seed-based assembly statistics for nine Cinnamomum samples.

Further, N50 values for all assemblies were comparable with previous data [66–70], while *C. verum* (India) has a smaller N50 value among the nine (9) chloroplast assemblies than the rest. Out of total DNA, more than 0.4% of chloroplast reads were assembled for all datasets of all nine (9) species.

https://doi.org/10.1371/journal.pone.0291763.t002



MN698964.1), *C. chartophyllum* (OL943972.1, NC_065102.1), *C. verum* (India) (ON685911) and all the wild *Cinnamomum* species in Sri Lanka (pp value 1). Within that monophyletic group there were two major sub-clades. The first sub clade includes *C. verum* (Sri Lanka) (ON685912), *C. verum* NCBI (NC_035236.1, KY635878.1), *C. pingbienense* (OL943977.1, NC065106.1), *C. kotoense* (NC050346.1, MN698964.1), *C. chartophyllum* (OL943972.1, NC_065102.1) and wild species of *C. rivulorum* (ON685909) and *C. dubium* (ON685906) (pp value 1) while *C. verum* (India) (ON685911), *C. ovalifolium* (ON685908), *C. litseifolium* (ON685907), *C. citriodorum* (ON685905), *C. capparu-coronde* (ON685904) and *C. sinhara-jaense* (ON685910) belongs to the second sub clade (pp value 1). In the first sub clade *C. verum*



Fig 2. The Bayesian phylogenomic tree constructed for *Cinnamomum* species based on complete chloroplast sequences. Posterior probability values are given next to the nodes and *Ocotea porosa* was set as the outgroup.

(Sri Lanka) (ON685912) was sister to *C. verum* NCBI (NC_035236.1, KY635878.1) with a pp value of 1, while *C. pingbienense* (OL943977.1, NC065106.1) was sistering to *C. kotoense* (NC050346.1, MN698964.1) with a pp value of 1. On the other hand, *C. rivulorum* (ON685909) and *C. dubium* (ON685906) were sister to *C. chartophyllum* (OL943972.1, NC_065102.1) with a pp value of 1. Within the second sub-clade *C. verum* (India) ON685911) was sister to *C. ovalifolium* (ON685908) and *C. litseifolium* (ON685907) with a pp value of 1.

Nucleotide diversity analysis

Nucleotide diversity analysis revealed two highly variable regions in the chloroplast genomes of the *Cinnamomum* species. Both were intergenic spacer regions, *trn*H-*psb*A, and *pet*A-*psb*J, with more than 0.01 nucleotide variability (Pi). The gene *ycf1* had the highest Pi value among genes. Universal barcoding genes, *mat*K and *rbc*L, had lower variability (Fig 3).

Simple Sequence Repeat (SSR) analysis

Table 3 presents the distribution of SSR regions. Almost all the species have a similar SSR distribution. Mononucleotide microsatellites were the most abundant form of SSR, and A/T



Fig 3. The nucleotide diversity (Pi) calculated for the chloroplast genomes of *Cinnamomum* species.

motifs were the most common among them. Trinucleotide microsatellites were the second most common SSR type, and the third were Dinucleotide repeats. The number of SSRs located outside the gene coding regions was double the number of SSRs located within the coding regions.

Diversity of Internal Transcribed Spacer (ITS) regions

Each assembled ITS sequence contains an 18S ribosomal RNA gene (rRNA) (Partial sequence), ITS 1, 5.8S rRNA, ITS 2, and 26S rRNA (partial sequence). ITS regions of *C. cap-paru-coronde*, *C. citriodorum*, *C. litseifolium*, *C. ovalifolium*, *C. sinharajaense*, and *C. verum* (India) are 599 bp in length, while it is 595 bp in *C. dubium*, *C. rivulorum*, and *C. verum* (Sri Lanka). The GC contents vary between 70%-71%.

The ITS region is the most variable region (Table 4) compared to the adjacent rDNA regions. Compared to the other species examined, *C. dubium*, *C. rivulorum*, *C. verum* NCBI, (SRX2990994), and *C. verum* (Sri Lanka) (OQ867307) have a 4 bp deletion in alignment positions 603–606. In the ITS 2 region, *C. dubium*, *C. rivulorum*, *C. verum* NCBI (SRX2990994), and *C. verum* (Sri Lanka) (OQ867307) share identical sequences, while another identical pattern is found in *C. capparu-coronde*, *C. citriodorum*, *C. litseifolium*, *C. ovalifolium*, *C. verum* (India) (OQ867448) and *C. sinharajaense*. Interestingly, *C. verum* Sri Lanka (OQ867307) and

Table 3.	SSR type,	location	in the	genome, a	nd the cou	nt among	Cinnamomum s	species.

Species			S	Location					
	Mono	Di	Tri	Tetra	Penta	Hexa	Non-coding	Coding	Total
C. capparu-coronde (ON685904)	130	36	68	8	1	1	160	84	244
C. citriodorum (ON685905)	130	36	68	8	1	1	160	84	244
C. dubium (ON685906)	132	38	69	8	1	1	164	85	249
C. litseifolium (ON685907)	131	36	68	8	1	1	160	85	245
C. ovalifolium (ON685908)	131	36	68	8	1	1	160	85	245
C. rivulorum (ON685909)	132	38	69	8	1	1	164	85	249
C. sinharajaense (ON685910)	130	36	69	8	1	1	160	85	245
<i>C. verum</i> (India) (ON685911)	131	36	68	8	1	1	156	89	245
<i>C. verum</i> (NC_035236)	130	38	68	8	1	1	161	85	246
<i>C. verum</i> (Sri Lanka) (NC685912)	130	38	68	8	1	1	157	89	246

https://doi.org/10.1371/journal.pone.0291763.t003

Species									Positio	on							
	185	18S ITS1							ITS2								
	1	78	112	160	161	171	218	241	494	546	561	566	593	603	604	605	606
C. dubium (OQ874796)	C	Т	C	С	A	G	A	C	A	G	Т	A	G	-	-	-	-
C. rivulorum (OQ888700)	С	Т	С	С	A	G	A	С	A	G	Т	A	G	-	-	-	-
C. verum NCBI (SRX2990994)	С	Т	С	С	G	G	С	G	A	G	Т	A	G	-	-	-	-
<i>C. verum</i> (Sri Lanka) (OQ867307)	С	Т	С	С	G	G	С	Т	A	G	Т	A	G	-	-	-	-
C. capparu-coronde (OQ888687)	Т	C	Т	A	A	A	C	Т	Т	A	С	G	A	C	C	A	Т
C. citriodorum (OQ874734)	Т	С	Т	Α	Α	А	С	Т	Т	A	С	G	A	С	С	А	Т
C. litseifolium (OQ874733)	Т	С	Т	A	A	A	С	Т	Т	A	С	G	A	С	С	A	Т
C. ovalifolium (OQ888686)	Т	С	Т	A	A	A	С	Т	Т	A	С	G	A	С	С	A	Т
<i>C. verum</i> (India) (OQ867448)	Т	С	Т	A	A	A	С	Т	Т	A	С	G	A	С	С	A	Т
C. sinharajaense (OQ867450)	Т	C	Т	A	A	A	С	С	Т	A	С	G	A	С	С	A	Т

Table 4. Nucleotide diversity in ITS regions among 10 Cinnamomum species.

https://doi.org/10.1371/journal.pone.0291763.t004

C. verum NCBI (SRX2990994) have only one (01) bp difference in the ITS region. *Cinnamomum verum* (India) (OQ867448) shares an identical ITS region with *C. ovalifolium*, *C. litseifolium*, *C. citriodorum*, and *C. capparu-coronde*. *Cinnamomum sinharajaense* has a single base difference at the 241st alignment position compared to the group including *C. verum* (India) (OQ867448).

The distance matrix obtained for the multiple sequence alignments of ITS regions of nine *Cinnamomum* samples sequenced in this study and the NCBI samples is given in S4 Table. The number of variable sites ranged from 0 to 325. There was no variation between C. verum (Sri Lanka) (OQ867307) and three sequences of C. verum from NCBI (KU139902, KU139903, KX766399), while there was a single base pair difference between C. verum (Sri Lanka) (OQ867307) and another five sequences from NCBI (SRX2990994, MF110059, MF110060, MF110061, and KX509827). In contrast, there are twenty (20) variable sites between C. verum (Sri Lanka) (OQ867307) and C. verum (India) (OQ867448). Variable sites between C. verum (Sri Lanka) (OQ867307) and wild Cinnamomum species from Sri Lanka ranged from 3 to 21, where the difference between C. verum (Sri Lanka) (OQ867307) compared to C. dubium (OQ874796) and C. rivulorum (OQ888700) was three (3), and it was twenty (20) compared to C. capparu-coronde (OQ888687), C. citriodorum (OQ874734), C. litseifolium (OQ874733) and C. ovalifolium (OQ888686), and twenty-one (21) compared to C. sinharajaense (OQ867450). Interestingly, the number of variable sites between C. verum (India) (OQ867448) and wild Cinnamomum species from Sri Lanka ranged from 0 to 21 as well. There was no difference between C. verum (India) (OQ867448) and C. capparu-coronde (OQ888687), C. citriodorum (OQ874734), C. litseifolium (OQ874733), and C. ovalifolium (OQ888686), while there was one nucleotide difference compared to C. sinharajaense (OQ867450). Cinnamomum verum India (OQ867448) differs from the wild species C. dubium (OQ874796) and C. rivulorum (OQ888700) in twenty-one (21) base pairs.





The Bayesian phylogenomic tree constructed for ITS sequences of 50 Cinnamomum samples show taxa relationship (Fig 4). Cinnamomum verum (Sri Lanka) (OQ867307) formed a monophyletic group with C. capparu-coronde (OQ888687), C. citriodorum (OQ874734), C. litseifolium (OQ874733), C. ovalifolium (OQ888686), C. verum India (OQ867448), C. sinharajaense (OQ867450), C. dubium (OQ874796), C. rivulorum (OQ888700), C. verum NCBI (SRX2990994, MF110059, MF110060, MF110061, KX509827, KU139902, KU139903, KX766399) with pp value of 0.99. There were two major sub clades within that monophyletic group. Cinnamomum verum (Sri Lanka) belongs to the first sub-clade with eight C. verum NCBI samples (SRX29990994, MF110059, MF110060, MF110061, KX509827, KU139902, KU139903, KX766399) and two wild species C. dubium (OQ874796) and C. rivulorum (OQ888700) (pp value of 1). Among the *C. verum* samples, SRX299094, MF110059, MF110060, MF110061 and KX509827 were closer (pp value 0.77), compared to C. verum samples of KU139902, KU139903, KX766399 and C. verum Sri Lanka (OQ867307). Cinnamomum verum (India) (OQ867448) and the remaining wild species of C. capparu-coronde (OQ888687), C. citriodorum (OQ874734), C. litseifolium (OQ874733), C. ovalifolium (OQ888686) and C. sinharajaense (OQ867450) belong to the second subclade (pp value 1).

Mitochondrial genes-assembly and analysis

Assembly of the complete plant mitochondrial genome is challenging [72]. Therefore, we looked at several regions *atp6* (631 bp), *cox1*(1415 bp), *atp1* (1262 bp), *matR* (1777 bp), and *atpA* (1239 bp) of the mitochondrial genome. Among them, *atp1* (1530 bp), *matR* (1777 bp), and *atpA* (1239 bp) regions were identical in the species examined, while *atp6* (631bp) had nine, and cox1 (1415 bp) had one variable site. While SNPs in the *atp6* gene cause 4 synonymous and 5 nonsynonymous changes, the SNP in the *cox1* gene is synonymous (S5 Table). Interestingly, *C. verum* (Sri Lanka), *C. verum* NCBI (SRX2990994), and *C. dubium* share an indel region between 766 and 771 bp and share identical *atp6* genes and proteins.

According to the Neighbor-Joining Consensus tree constructed for combined datasets of *atp6* (631 bp) and cox1 (1415 bp) genes *C. verum* (India) (OQ863233,OQ863248), *C. ovalifo-lium* (OQ863235,OQ876858) *C. dubium* (OQ863237,OQ863245) *C. capparu-coronde* (OQ863240 OQ863247), *C. sinharajaense* (OQ863234,OQ863242), *C. verum* (Sri Lanka) (OQ863232, OQ863241) *C. verum* NCBI (SRX2990994) and *C. verum* (India) (OQ863233, OQ863248) clustered together (bootstrap 96) while *C. verum* (Sri Lanka) (OQ863224) and *C. verum* NCBI (SRX2990994) further clustering together (bootstrap value 71) (Fig 5).



Fig 5. The unrooted Neighbor Joining Consensus tree constructed for *Cinnamomum* species based on mitochondrial *atp6* and cox1 genes. Bootstrap values are given next to the nodes. GenBank accession numbers of *atp6* (631 bp) and cox1 (1415 bp) genes are given respectively next to the species names.

https://doi.org/10.1371/journal.pone.0291763.g005

Skmer analysis

Interestingly, Skmer V 3.02 grouped Sri Lankan *Cinnamomum* species into three clades (Fig 6). *Cinnamomum verum* (Sri Lanka) is sister to C. *dubium* and C. *rivulorum* with a boot-strap value of 0.0042, whereas C. *verum* (India) is sister to C. *ovalifolium*, C. *litseifolium* and C. *citriodorum* (8.2x10⁻⁴). Furthermore, C. *sinharajaense* and C. *capparu-coronde* formed a clade with bootstrap value of 0.0014.

Primer design, PCR, Sanger sequencing and data analysis

Since *C. verum* (India) and *C. verum* (Sri Lanka) had considerable differences in all the analyses, we analyzed more samples from these two groups. Additional eight *C. verum* (Sri Lanka) accessions showing considerably different morphological and chemical traits [16], and two *C. verum* (India) were assessed with the most variable regions (*ycf1* gene) in the chloroplast genome. The two regions of the *ycf1* gene were PCR amplified, and all the samples resulted in good-quality sequencing data. The alignment included the same regions as extracted from Illumina data and already deposited in GenBank (S6 Table). The first region of the *ycf1* gene has thirteen variable sites among the three *C. verum* samples from India and the examined samples from Sri Lanka including wild species. The second *ycf1* region includes seventeen variable sites and an indel region of 18 bp. Altogether, 30 variable sites and an indel region were found among the examined species. While the variable sites vary among samples, all the *C. verum* (Sri Lanka) samples studied show an 18 bp insertion in the second *ycf1* region, which is not present in three samples collected from India. The same insertion is present in *C. verum* NCBI (NC_035236) and *C. sinharajaense*, while it is absent in all the other Sri Lankan wild species and *C. aromaticum* (NC046019).

Furthermore, sequence divergence calculated for combined sequences of each group revealed that there is no within-group sequence divergence for *C. verum* (Sri Lanka), while it is 0.006 for *C. verum* (India). The intergroup sequence divergence for *C. verum* (Sri Lanka) and *C. verum* (India) is 0.007. As such, when considering *C. verum* (Sri Lanka) and *C. verum*





https://doi.org/10.1371/journal.pone.0291763.g006

(India) as two groups, the sequence divergence between them is higher than the within-group sequence divergence of each species.

Discussion

In this study we assembled the chloroplast genomes, ITS regions, *atp6*, *cox1*, *atp1*, *matR*, and *atpA* of the mitochondrial genome and Skmer analysis using the 20x coverage Illumina Hiseq data from *Cinnamomum* species found in Sri Lanka. We did similar analyses using all the datasets and gene regions for easy comparisons. Interestingly, all the analyses supported a similar pattern of evolutionary relationship among *Cinnamomum* species in Sri Lanka. The picture is clearer than what we observed with the universal barcoding regions [25].

There are about 100 chloroplasts in typical mesophyll cells of plants such as *Arabidopsis*, wheat, and rice [73–75]. Regular DNA extraction protocols such as CTAB [76] and SDS [77] result in total cellular DNA, including genomic, chloroplast, and mitochondrial DNA. However, there are protocols available for additional enrichment of chloroplast DNA, including plastid isolation, enrichment via methylation-sensitive capture, hybrid bait capture, and PCR [78,79]. Nevertheless, these isolation or enrichment procedures are time-consuming and expensive [80]. Therefore, "skim sequencing" has become common now, in which the total DNA is sequenced and the chloroplast DNA is separated bioinformatically [80,81]. This approach is more cost-effective, as a total genome sequenced at a lower coverage usually results in sufficient coverage of chloroplast DNA for assembling the chloroplast genome. It is suggested that a sequencing coverage of $\sim 0.1 - 10x$ for the nuclear genome is sufficient for the genome skimming approach [82,83].

When assembling complete chloroplast genomes from total DNA, it is vital to identify the most effective assembly method and bioinformatics tools to obtain the highest accuracy in results. Even degraded herbarium material has been successfully assembled with genome skimming, but it is necessary to give special attention to the assembly process [84]. When sequencing, a suitable platform that provides read lengths larger than repeat lengths in the plastome should be chosen. Currently, most NGS platforms fulfill this requirement, and the reads generated are sufficient for *de novo* assemblies [85]. A coverage of 30x and more than 500 Mb of sequencing data is considered sufficient to generate a good quality assembly [82]. Hence, Illumina HiSeq provides a cost-effective solution and high throughput for larger genome skimming.

When performing genome skimming to assemble a chloroplast genome, it is generally a good idea to first separate the chloroplast reads from nuclear DNA before assembly. It reduces the complexity of the data, aiding the *de novo* assembly process [82]. Among suitable *de novo* assemblers for filtered chloroplast reads are Geneious [48], MIRA [86], ABySS [87], SOAPdenovo [88], SPAdes [58], and Velvet [89]. Some assemblers such as MITObim [90], Fast-plast [91], and NOVOPlasty [45] merge filtering of chloroplast reads and assembly processes. These assemblers use a known plastid sequence as a seed or a 'bait' to identify chloroplast reads within total DNABased on the assembly statistics. We considered the seed-based assemblies to be more reliable to continue with the annotation and analysis. The amount of data were sufficient to assemble the ITS regions, 18S ribosomal RNA gene (rRNA) (Partial sequence), ITS 1, 5.8S rRNA, ITS 2, and 26S rRNA (partial sequence). While the data generated were not sufficient to assemble the mitogenomes, several mitochondrial regions were assembled and included in the analysis.

We encountered several bioinformatics-related challenges during the optimization process. For example, whether sufficient coverage of data could be generated to assemble chloroplast sequences from dry bark. Based on the chloroplast analysis, it was clear that the data is comparable to the data generated from green leaves. For example, *C. verum* (India) has the smallest N50 value among the nine (9) chloroplast assemblies. This behavior in *C. verum* (India, ON685911) data could also be due to the sample being a bark. As barks do not contain much chloroplast DNA, the amount of nuclear DNA could be high within the sequenced data of *C. verum* (India, ON685911). This could potentially be a challenge for the *de novo* assembler and would provide a large number of contigs with more gaps than leaf samples. Similarly, a higher nuclear DNA amount could have affected the N50 value of *C. verum* (India, ON685911) being lower than that of other assemblies. Further, if a high repeat content is present, it could also affect the N50 value as the assembler would struggle to produce longer contigs.

However, since the *C. verum* (India, ON685911) assembly length, amount of chloroplast reads assembled, and GC content is in range with all other assemblies, it can be considered of good quality. In addition, the percentage of chloroplast reads assembled in *C. ovalifolium* is even smaller than the assembled chloroplast reads of *C. verum* (India, ON685911). Nevertheless, the *C. ovalifolium* assembly is of very good quality considering the N50 value. This indicates that the available chloroplast reads in the total DNA were sufficient to assemble a good quality *C. verum* (India, ON685911) chloroplast genome. The assembled chloroplast genomes were submitted to NCBI GenBank.

According to the complete chloroplast analysis, *C. chartophyllum* (158 kb) had a larger genome size compared with published chloroplast genomes and the other nine newly sequenced chloroplast genomes in this study. The authors, Ge *et al.* 2022, predicted that the larger size of *C. chartophyllum* is due to the IR expansion, resulting in duplication of complete *trnICAU*, *rpl32*, *rpl2*, *and ycf2* in the IR regions, which was the first case in the genus *Cinnamo-mum* [88]. Nevertheless, it might also be due to artefacts of the *de novo* plastome assembly process. Surprisingly, *C. verum* (Sri Lanka) and *C. verum* (NC_035236.1, KY635878.1) are more closely related to *C. pingbienense* (OL943977.1, NC65106.1) and *C. kotoense* (NC050346.1, MN698 964.1) than to wild *Cinnamomum* samples in Sri Lanka and *C. verum* (India). *Cinnamomum kotoense* is an endangered species in Lanyu island, Taiwan, and is an ornamental plant. It is reported that *C. kotoense* is closely related to *C. verum* and *C. aromaticum* [92]. *Cinnamomum pingbienense* is native to South-Central and Southeast China [93]. The branch support values are high (pp value1), confirming the accuracy of clustering based on chloroplast data deposited on the NCBI. However, further analysis and studying type specimens will be needed to confirm such relationships.

In our previous analyses, nucleotide diversity in *rbcL*, *mat*K and *trn*H-*psb*A regions was not sufficient to differentiate *C. verum* (Sri Lanka) from *C. sinharajaense*, or *C. litseifolium* from *C. ovalifolium* and *C. citriodorum* [25]. The current analysis suggests comparatively higher nucleotide diversity in the *trn*H-*psb*A and *pet*A-*psbJ* regions than in the common universal barcoding regions. However, a previous study has also reported inadequacy of the nucleotide diversity in the *trn*H-*psb*A region for molecular level identification of the *Cinnamomum* species in Sri Lanka [5]. The current work proposes new barcode regions of *ycf1*, ITS and mito-chondrial genes of *atp6* and cox1 for the identification of *Cinnamomum* species.

Complete chloroplast and ITS analyses suggest that *C. verum* (Sri Lanka) is more closely related to *C. verum* samples in NCBI, *C. dubium* and *C. rivulorum* than to other species. Further, *C. ovalifolium* and *C. litseifolium* always group together suggesting DNA level similarity between them. Interestingly, *C. verum* (India, ON685911) groups with *C. ovalifolium* and *C. litseifolium* and *C. litseifolium* and *C. litseifolium* always group together suggesting DNA level similarity between them. Interestingly, *C. verum* (India, ON685911) groups with *C. ovalifolium* and *C. litseifolium* but not with *C. verum* (Sri Lanka) (or at least not immediately). Further *C. litseifolium* is only found in restricted habitats above 1800 m elevation [8]. However, the sequences retrieved from NCBI of specimens identified as *C. verum* share an identical *atp6* region (SRX299099), an identical *Cox1* region (SRX299099) and a single base pair mismatch in the

ITS region with *C. verum* (Sri Lanka, OQ867307). While most of the chloroplast regions are identical there is a 5 bp mismatch in the complete chloroplast genomes between *C. verum* (Sri Lanka, OQ685912) and *C. verum* (NC_035236). We observed the same with our previous work on barcoding regions [25]. Therefore, the available chloroplast genome (ON685912) and the rest of the sequences of *C. verum* in the NCBI (NC_035236.1, KY635878.1) could have originated from Sri Lanka because some countries specially China and India grow Ceylon cinnamon.

The NCBI consisted of ninety complete chloroplast genome assemblies and more than 250 ITS regions that we could include in the analysis. However, there was no sufficient data for mitochondrial gene regions, except *C. <u>chekiangense</u>*. Similarly, we limited Skemer analysis to our dataset since there was limited skim sequencing data for the genus *Cinnamomum*. Therefore, the number of taxa in each analysis and the topology of phylogenetic tree were not comparable. However, relationship among samples included in this study were consistent.

The wild populations of C. verum (Sri Lanka) are still found in the upcountry and midcountry rain forests in Sri Lanka [13,20,94]. Such findings support the historical evidence of the origin of cultivated Cinnamon, where Portuguese and Dutch invaders started commercial cultivation in the southern part of the country. The historical evidence suggests the introduction of *C. verum* to India by taking a bunch of seeds from Sri Lanka in the 1920s [10]. Therefore, a closer relationship is expected between C. verum (Sri Lanka) and C. verum (India). However, all the above analyses suggested they are further apart when considering DNA diversity. The morphology [13], molecular [25], and biochemical [17] data suggest intraspecies diversity in both cultivated and wild Cinnamomum species in Sri Lanka. We selected a group of C. verum samples identified as most diverse in morphological, biochemical, and yieldrelated traits from our previous work to assess their molecular diversity and to compare them with the samples collected from India. PCR primers were designed for the *ycf1* gene, the most variable region of the chloroplast genome, that we identified in the current study. The ycf1 gene, a hypervariable region, is the most variable locus and achieved better phylogenomic resolutions than standard DNA barcodes in land plants for phylogenomic studies [95,96]. They also suggest it as a cost-effective method, considering that complete chloroplast genome sequencing requires high-quality DNA, a higher cost for sequencing, and bioinformatics facilities.

There is considerable intraspecies diversity among C. verum (India, NC_035236.1,) as well as in *C. verum* (Sri Lanka, ON685912) for *ycf1*. However, the indel region between position 736 and 753 is conserved in all nine samples of C. verum (Sri Lanka), where the 18 bp motif (GTCCCTATAGAATCTTCT) is duplicated. The same sequence is present in a C. verum chloroplast genome (NC_035236.1) deposited in NCBI and in C. sinharajaense (ON685910). Only one copy of the 18 bp region is present in all the other Cinnamomum species in Sri Lanka as well as in the three C. verum samples from India. Therefore, sequence data in NCBI might be linked to Sri Lanka, though the authors have purchased it from an ornamental plant grower, Top Tropicals and the authenticity information has not been mentioned [60]. Cinnamomum sinharajaense is only found in restricted locations of the Sinharaja rainforest and is considered threatened. Therefore, there is a minimal possibility of appearing C. sinharajaense in either local or foreign markets. Therefore, this indel region is useful as a marker to differentiate C. verum (India) from C. verum (Sri Lanka). Further, there was no nucleotide diversity in the ycf1 region among the C. verum (Sri Lanka) collections included in the analysis while it was 0.006 among the C. verum (India) collection. However, the diversity between C. verum samples collected from India and Sri Lanka is higher than the intra-sample diversity values.

Conclusions

All the analysis suggests that C. verum chloroplast sequencing data (NC_035236.1) deposited in the NCBI database could be from a sample of Sri Lankan origin. Chandrasekara et al. 2021 suggested the same based on an analysis of universal barcoding regions [25]. Interestingly, C. verum (Sri Lanka) and C. sinharajaense have similar chemical profiles except for differences in the relative abundance of some compounds [94]. While C. sinharajaense can be identified morphologically, there is no evidence for morphological differences between C. verum from Sri Lanka and C. verum from India. While the name C. zeylanicum [97] has widely been used in the literature for Sri Lankan Cinnamon, it is considered a synonym for C. verum [98]. However, the Scanning Electron Microscopic analysis (SEM) of C. verum (Sri Lanka) pollen samples [5] suggests considerable differences compared to recently published SEM data of C. verum (India) [99]. The pollen size and the spine length are different between them. Such differences are to be studied comprehensively. Nevertheless, the molecular data presented here would provide complementary evidence. Although this analysis included only a single accession from each species, it consisted of the complete chloroplast genome, a considerable region in the mitochondrial genome, and many coding and non-coding regions of the nuclear genome. Individual regions and combined analyses were conducted and resulted in similar evolutionary relationships and associations. Therefore, the robustness of the data and the analysis are confirmed, similar to our previous work on transcriptomics [100]. We propose this as a cost-effective analysis method for studying phylogenomic relationships among closely related species.

Supporting information

S1 Table. Details of the wild *Cinnamomum* **samples collected.** (DOCX)

S2 Table. Features of the chloroplast genome. (XLSX)

S3 Table. Distance matrix obtained for the chloroplast genome alignment of *Cinnamo-mum* samples.

(XLSX)

S4 Table. Distance matrix obtained for the ITS alignment of *Cinnamomum* samples. (XLSX)

S5 Table. Nucleotide diversity of the mitochondrial genes *atp6* and *cox1* among 10 *Cinna-momum* species. Accession numbers in the first column represent *atp6* and *cox1* respectively. (XLSX)

S6 Table. Sequence variation in two parts of the *ycf1* gene among *Cinnamomum* species. (XLSX)

Acknowledgments

The authors thank Dr. Ardeshir B. Damania from the Department of Plant Sciences, the University of California Davis and Prof. Hashendra S. Kathriarchchi from the Department of Plant Sciences, University of Colombo for insightful comments, and for editing the manuscript. The authors appreciate the support of Prof. Ranjith Senaratne, Chairman, National Science Foundation, Sri Lanka for the insightful comments and immense support as the Chairperson of the Steering Committee Special Project on Cinnamon and the members of the

Steering Committee and the contact persons of the National Science Foundation, especially Mr. Janaka Karunasena. The authors acknowledge Dr. Ranil Rajapaksha, Department of Crop Sciences, Faculty of Agriculture, University of Peradeniya for his guidance throughout the wild cinnamon sample collection and Mr. Supun Bandusekara, University of Colombo, Institute for Agro-Technology and Rural Sciences for collecting wild cinnamon samples. The authors thank the Former Directors, Mr. K.G.G. Wijesinghe and Dr. G.G. Jayasinghe, and the staff of National Cinnamon Research and Training Center, Department of Export Agriculture, Thihagoda, Palolpitiya and Mr. R. A. A. K. Ranawaka, Assistant Director (Research) and the staff of Mid Country Research Station, Department of Export Agriculture, Dalpitiya, Atabage, for providing cinnamon samples and the assistance provide throughout the project. Special thanks to High commission of the Democratic Socialist Republic of Sri Lanka in India for providing C. verum samples from India and State Ministry of Development of Minor crops, Sri Lanka for the coordination of the process. The authors thank the Department of Wildlife Conservation, Sri Lanka, and the Forest Department for providing permission and required approvals to conduct this research work. The authors would like to thank the Director and the staff of the National Herbarium, (PDA), the Royal Botanic Gardens, Peradeniya for authentication of specimens. The authors would like to thank the other personal involved with the Special Cinnamon Project for their support, especially, Prof. K.M.S. Wimalasiri, Prof. Wasantha Kumara, Ms. N.M.N. Liyanage, Ms. H.AB.M. Hathurusinghe, Ms. W.M.K.K. Rajapaksha, Ms. W.D.N. Wickamasinghe, Ms. I.S.A. Isurumali Jayasiri and Ms. Y.M.A.D.K Yapa. The authors sincerely thank Ms. D.M.D.K Halangoda, Mr. R.A.J. Rathnayake, Ms. R. M. Mallika Wijerathne, and the staff of the Agricultural Biotechnology Centre, Faculty of Agriculture, University of Peradeniya for the continuous support for project work.

Author Contributions

Conceptualization: Pradeepa C. G. Bandaranayake.

- **Data curation:** Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Formal analysis: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Funding acquisition: Pradeepa C. G. Bandaranayake, D. K. N. G. Pushpakumara.
- Investigation: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Asitha U. Bandaranayake, D. K. N. G. Pushpakumara, D. Siril A. Wijesundara.
- Methodology: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Project administration: Pradeepa C. G. Bandaranayake.
- **Resources:** Pradeepa C. G. Bandaranayake, Asitha U. Bandaranayake, D. K. N. G. Pushpakumara, D. Siril A. Wijesundara.
- Software: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara, Asitha U. Bandaranayake.
- **Supervision:** Pradeepa C. G. Bandaranayake, Asitha U. Bandaranayake, D. K. N. G. Pushpakumara, D. Siril A. Wijesundara.

- Validation: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Visualization: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Writing original draft: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara.
- Writing review & editing: Pradeepa C. G. Bandaranayake, Nathasha Naranpanawa, C. H. W. M. R. Bhagya Chandrasekara, Hiruna Samarakoon, S. Lokuge, S. Jayasundara, Asitha U. Bandaranayake.

References

- 1. Yang Z, Liu B, Yang Y, Ferguson DK. Phylogeny and taxonomy of Cinnamomum (Lauraceae). Vol. 12, Ecology and Evolution. John Wiley & Sons, Ltd; 2022. p. e9378.
- Rohde R, Rudolph B, Ruthe K, Lorea-Hernández FG, de Moraes PLR, Li J, et al. Neither phoebe nor cinnamomum–the tetrasporangiate species of aiouea (Lauraceae). Taxon. 2017 Oct 1; 66(5):1085– 111.
- Huang JF, Li L, van der Werff H, Li HW, Rohwer JG, Crayn DM, et al. Origins and evolution of cinnamon and camphor: A phylogenetic and historical biogeographical analysis of the Cinnamomum group (Lauraceae). Mol Phylogenet Evol. 2016 Mar 1; 96:33–44. https://doi.org/10.1016/j.ympev.2015.12. 007 PMID: 26718058
- Gang Z, Liu B, Rohwer JG, Ferguson DK, Yang Y. Leaf epidermal micromorphology defining the clades in Cinnamomum (Lauraceae). PhytoKeys. 2021; 182:125–48. <u>https://doi.org/10.3897/</u> phytokeys.182.67289 PMID: 34720625
- 5. Hathurusinghe BM, Pushpakumara DKNG, Bandaranayake PCG. Macroscopic and microscopic study on floral biology and pollination of Cinnamomum verum Blume (Sri Lankan). bioRxiv. 2022 Jul 13;2022.07.12.499711.
- Chen P, Sun J, Ford P. Differentiation of the four major species of cinnamons (C. burmannii, C. verum, C. cassia, and C. loureiroi) using a flow injection mass spectrometric (FIMS) fingerprinting method. J Agric Food Chem. 2014; 62(12):2516–21. https://doi.org/10.1021/jf405580c PMID: 24628250
- Barceloux DG. Cinnamon (Cinnamomum species). Dis Mon. 2009 Jun; 55(6):327–35. <u>https://doi.org/10.1016/j.disamonth.2009.03.003 PMID: 19446676</u>
- Ranasinghe P, Pigera S, Premakumara GS, Galappaththy P, Constantine GR, Katulanda P. Medicinal properties of "true" cinnamon (Cinnamomum zeylanicum): A systematic review. BMC Complement Altern Med. 2013 Oct 22; 13(1):1–10. https://doi.org/10.1186/1472-6882-13-275 PMID: 24148965
- 9. Suriyagoda L, Mohotti AJ, Vidanarachchi JK, Kodithuwakku SP, Chathurika M, Bandaranayake PCG, et al. "Ceylon cinnamon": Much more than just a spice. Plants People Planet. 2021; 3(4):319–36.
- 10. Ravindran PN, Nirmal Babu K, Shylaja M. Cinnamon and cassia: the genus Cinnamomum. 2004.
- 11. Dassanayake MD, Trimen H. A Revised handbook to the flora of Ceylon. A.A. Balkema; 1980.
- 12. Sritharan R. The study of genus Cinnamomum. University of Peradeniya, Sri Lanka; 1984.
- Bandusekara BS, Pushpakumara DKNG, BandaranayakeWijesinghe KGG, Jayasinghe GG. Field Level Identification of Cinnamomum Species in Sri Lanka Using a Morphological Index. Trop Agric Res. 2020; 31(4):43.
- Bandusekara BS, Pushpakumara DKNG, Bandaranayake PCG, Wijesinghe KGG. Development of Morphological Index for Field Level Identification of Cinnamon Varieties. In: SLCARP International Agriculture Research Symposium. 2018. p. 27.
- Hathurusinghe B, Pushpakumara DKNG, Bandaranayake PCG. Unveiling the possible floral visitors and invisible pollination networks from Deep RNA-seq Profile. Ecol Genet Genomics. 2023 Sep; 28:100178.
- Liyanage NMN, Bandusekara BS, Kanchanamala RWMK, Hathurusinghe HABM, Rathnayaka AMRWSD, Pushpakumara DKNG, et al. Identification of superior Cinnamomum zeylanicum Blume germplasm for future true cinnamon breeding in the world. J Food Compos Anal. 2021 Mar 1; 96:103747.

- Liyanage NMN, Ranawake AL, Bandaranayake PCG. Cross-pollination effects on morphological, molecular, and biochemical diversity of a selected cinnamon (Cinnamomum zeylanicum Blume) seedling population. J Crop Improv. 2021; 35(1):21–37.
- Ariyarathne HBMA, Weerasuriya SN, Senarath WTPSK. Comparison of morphological and chemical characteristics of two selected accessions and six wild species of genus Cinnamomum Schaeff. Sri Lankan J Biol. 2018; 3(1):11.
- 19. Bandaranayake PCG, Pushpakumara DKNG. Genetics and Molecular Characterization of Genus Cinnamomum. In: Cinnamon. Cham: Springer International Publishing; 2020. p. 119–46.
- 20. Abeysinghe PD, Bandaranayake PCG, Pathirana R. Botany of Endemic Cinnamomum Species of Sri Lanka. In: Cinnamon. Cham: Springer International Publishing; 2020. p. 85–118.
- Lin TP, Cheng YP, Huang SG. Allozyme variation in four geographic areas of Cinnamomum kanehirae. J Hered. 1997; 88(5):433–8.
- Kojoma M, Kurihara K, Yamada K, Sekita S, Satake M, lida O. Genetic identification of cinnamon (Cinnamomum spp.) based on the trnL-trnF chloroplast DNA. Planta Med. 2002; 68(1):94–6. https://doi.org/10.1055/s-2002-20051 PMID: 11842343
- Abeysinghe PD, Samarajeewa NGCD, Li G, Wijesinghe KGG. Preliminary investigation for the identification of Sri Lankan Cinnamomum species using randomly amplified polymorphic DNA (RAPD) and sequence related amplified polymorphic (SRAP) markers. J Natl Sci Found Sri Lanka. 2014; 42 (3):175–82.
- 24. Hathurusinghe HABM, Bandusekara BS, Pushpakumar DKNG, Ranawaka RAAK, Bandaranayake PCG. Possibility of utilizing Inter Simple Sequence Repeat regions, bark powder morphology and floral morphometry to characterize the Cinnamomum species in Sri Lanka. Trop Agric Res. 2023 Jan 1; 34 (1):65–79.
- Bhagya Chandrasekara CHWMR Naranpanawa DNU, Bandusekara BS Pushpakumara DKNG, Wijesundera DSA Bandaranayake PCG. Universal barcoding regions, rbcL, matK and trnH-psbA do not discriminate Cinnamomum species in Sri Lanka. PLoS One. 2021; 16(2 February):1–16.
- 26. Soulange JG, Ranghoo-Sanmukhiya VM, Seeburrun SD. Tissue culture and RAPD analysis of Cinnamomum camphora and Cinnamomum verum. Biotechnology. 2007; 6(2):239–44.
- Joy P, Maridass M. Inter Species Relationship of Cinnamomum Species Using RAPD Marker Analysis. Ethnobot Leafl. 2008; 12:476–80.
- Kuo DC, Lin CC, Ho KC, Cheng YP, Hwang SY, Lin TP. Two genetic divergence centers revealed by chloroplastic DNA variation in populations of Cinnamomum kanehirae Hay. Conserv Genet. 2010 Jun; 11(3):803–12.
- Lee SC, Lee CH, Lin MY, Ho KY. Genetic identification of Cinnamomum species based on partial internal transcribed spacer 2 of ribosomal DNA. J Food Drug Anal. 2010; 18(4):225–31.
- **30.** Sandigawad AM, Patil CG. Genetic diversity in cinnamomum zeylanicum blume. (lauraceae) using random amplified polymorphic dna. African J Biotechnol. 2011; 10(19):3682–8.
- Ho KY, Hung TY. Cladistic relationships within the genus cinnamomum (Lauraceae) in Taiwan based on analysis of leaf morphology and inter-simple sequence repeat (ISSR) and internal transcribed spacer (ITS) molecular markers. African J Biotechnol. 2011 Sep 12; 10(24):4802–15.
- Kameyama Y. Development of microsatellite markers for cinnamomum camphora (lauraceae). Am J Bot. 2012 Jan; 99(1). https://doi.org/10.3732/ajb.1100231 PMID: 22223691
- 33. Abeysinghe PD, Wijesinghe KGG, Tachida H, Yoshda T. Molecular Characterization of Cinnamon (Cinnamomum Verum Presl) Accessions and Evaluation of Genetic Relatedness of Cinnamon Species in Sri Lanka Based on TrnL Intron Region, Intergenic Spacers Between trnT-trnL, trnL-trnF, trnH -psbA and nuclear ITS. Agric Biol Sci. 2009; 5(6):1079–88.
- Swetha VP, Parvathy VA, Sheeja TE, Sasikumar B. DNA Barcoding for Discriminating the Economically Important Cinnamomum verum from Its Adulterants. Food Biotechnol. 2014; 28(3):183–94.
- Purushothaman N, Newmaster SG, Ragupathy S, Stalin N, Suresh D, Arunraj DR, et al. A tiered barcode authentication tool to differentiate medicinal Cassia species in India. Genet Mol Res. 2014; 13 (2):2959–68. https://doi.org/10.4238/2014.April.16.4 PMID: 24782130
- Seethapathy GS, Ganesh D, Santhosh Kumar JU, Senthilkumar U, Newmaster SG, Ragupathy S, et al. Assessing product adulteration in natural health products for laxative yielding plants, Cassia, Senna, and Chamaecrista, in Southern India using DNA barcoding. Int J Legal Med. 2015; 129 (4):693–700. https://doi.org/10.1007/s00414-014-1120-z PMID: 25425095
- Liu ZF, Ci XQ, Li L, Li HW, Conran JG, Li J. DNA barcoding evaluation and implications for phylogenetic relationships in Lauraceae from China. PLoS One. 2017; 12(4):1–20. <u>https://doi.org/10.1371/journal.pone.0175788</u> PMID: 28414813

- Chanderbali AS, Van Der Werff H, Renner SS. Phylogeny and historical biogeography of Lauraceae: Evidence from the chloroplast and nuclear genomes. Ann Missouri Bot Gard. 2001; 88(1):104–34.
- Rohwer JG. Toward a phylogenetic classification of the Lauraceae: Evidence from matK sequences. Syst Bot. 2000; 25(1):60–71.
- 40. Li J, Christophel DC, Conran JG, Li HW. Phylogenetic relationships within the "core" Laureae (Litsea complex, Lauraceae) inferred from sequences of the chloroplast gene matK and nuclear ribosomal DNA ITS regions. Plant Syst Evol. 2004; 246(1–2):19–34.
- Tian Y, Zhou J, Zhang Y, Wang S, Wang Y, Liu H, et al. Research progress in plant molecular systematics of lauraceae. Biology (Basel). 2021 May 1; 10(5):391. https://doi.org/10.3390/biology10050391 PMID: 34062846
- Daniell H, Lin CS, Yu M, Chang WJ. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. Genome Biol. 2016 Dec 23; 17(1):134. https://doi.org/10.1186/s13059-016-1004-2 PMID: 27339192
- 43. Yao H, Song J, Liu C, Luo K, Han J, Li Y, et al. Use of ITS2 region as the universal DNA barcode for plants and animals. PLoS One. 2010; 5(10):e13102. https://doi.org/10.1371/journal.pone.0013102 PMID: 20957043
- Sarmashghi S, Bohmann K, Gilbert PMT, Bafna V, Mirarab S. Skmer: Assembly-free and alignmentfree sample identification using genome skims. Genome Biol. 2019 Feb 13; 20(1):1–20.
- Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017; 45(4):18. https://doi.org/10.1093/nar/gkw955 PMID: 28204566
- Doyle JJ, Doyle JL. A rapid DNA isolation procedure from small quantities of fresh leaf material. Phytochem Bull. 1987; 19:11–5.
- 47. Chandrasekara C, Bandaranayake P, PushpaKumara D. DNA extraction from Cinnamomum zeylanicum cinnamon: a simple and efficient method. In: Proceedings of the 7th YSF SymposiumYoung Scientists Forum National Science and Technology Commission. 2018. p. 22–5.
- Assembly and Mapping | Academic & Government | Geneious Prime [Internet]. Available from: <u>https://</u> www.geneious.com/academic/features/assembly-mapping/.
- 49. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-versatile and accurate annotation of organelle genomes. Nucleic Acids Res [Internet]. 2017 Jul 3; 45(W1):W6–11. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkx391.
- Laslett D, Canback B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res [Internet]. 2004 Jan 2; 32(1):11–6. Available from: <u>https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh152</u>. PMID: 14704338
- Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012 Mar 4; 9 (4):357–9. https://doi.org/10.1038/nmeth.1923 PMID: 22388286
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013 Apr 1; 30(4):772–80. https://doi.org/10.1093/molbev/ mst010 PMID: 23329690
- 53. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001 Aug 1; 17(8):754–5. https://doi.org/10.1093/bioinformatics/17.8.754 PMID: 11524383
- Trofimov D, De Moraes PLR, Rohwer JG. Towards a phylogenetic classification of the Ocotea complex (Lauraceae): Classification principles and reinstatement of Mespilodaphne. Bot J Linn Soc. 2019 Apr 26; 190(1):25–50.
- 55. Trofimov D, Rohwer JG. Towards a phylogenetic classification of the Ocotea complex (Lauraceae): An analysis with emphasis on the Old World taxa and description of the new genus Kuloa. Bot J Linn Soc. 2020 Feb 27; 192(3):510–35.
- 56. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017 Dec 1; 34 (12):3299–302. https://doi.org/10.1093/molbev/msx248 PMID: 29029172
- Du L, Zhang C, Liu Q, Zhang X, Yue B. Krait: An ultrafast tool for genome-wide survey of microsatellites and primer design. Bioinformatics. 2018 Feb 15; 34(4):681–3. <u>https://doi.org/10.1093/</u> bioinformatics/btx665 PMID: 29048524
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012 May; 19 (5):455–77. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599
- Rivers AR, Weber KC, Gardner TG, Liu S, Armstrong SD. ITSxpress: Software to rapidly trim internally transcribed spacer sequences with quality scores for marker gene analysis. F1000Research. 2018;7.

- Rabah SO, Lee C, Hajrah NH, Makki RM, Alharby HF, Alhebshi AM, et al. Plastome Sequencing of Ten Nonmodel Crop Species Uncovers a Large Insertion of Mitochondrial DNA in Cashew. Plant Genome. 2017 Nov; 10(3). https://doi.org/10.3835/plantgenome2017.03.0020 PMID: 29293812
- Keller A, Schleicher T, Schultz J, Müller T, Dandekar T, Wolf M. 5.8S-28S rRNA interaction and HMMbased ITS2 annotation. Gene. 2009 Feb 1; 430(1–2):50–7. <u>https://doi.org/10.1016/j.gene.2008.10.</u> 012 PMID: 19026726
- Sloan DB, Müller K, Mccauley DE, Taylor DR, Štorchová H. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in Silene vulgaris, an angiosperm with pervasive cytoplasmic male sterility. New Phytol. 2012; 196(4):1228–39. <u>https://doi.org/10.1111/j.1469-8137.2012</u>. 04340.x PMID: 23009072
- Barkman TJ, Chenery G, McNeal JR, Lyons-Weiler J, Ellisens WJ, Moore G, et al. Independent and combined analyses of sequences from all three genomic compartments converge on the root of flowering plant phylogeny. Proc Natl Acad Sci U S A. 2000; 97(24):13166–71. <u>https://doi.org/10.1073/pnas.</u> 220427497 PMID: 11069280
- 64. Tamura K, Nei M. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol. 1993; 10(3):512–26. <u>https://doi.org/10.1093/oxfordjournals.molbev.a040023 PMID: 8336541</u>
- 65. Tamura K, Stecher G, Kumar S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol Biol Evol. 2021 Jun 25; 38(7):3022–7. https://doi.org/10.1093/molbev/msab120 PMID: 33892491
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ. Thirteen Camellia chloroplast genome sequences determined by high-throughput sequencing: Genome structure and phylogenetic relationships. BMC Evol Biol. 2014 Jul 7; 14(1):151. https://doi.org/10.1186/1471-2148-14-151 PMID: 25001059
- Wang W, Messing J. High-Throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. Badger JH, editor. PLoS One. 2011 Sep 9; 6(9):e24670. <u>https://doi.org/10.1371/journal.pone.0024670</u> PMID: 21931804
- Molina J, Hazzouri KM, Nickrent D, Geisler M, Meyer RS, Pentony MM, et al. Possible loss of the chloroplast genome in the parasitic flowering plant Rafflesia lagascae (Rafflesiaceae). Mol Biol Evol. 2014; 31(4):793–803. https://doi.org/10.1093/molbev/msu051 PMID: 24458431
- Yang M, Zhang X, Liu G, Yin Y, Chen K, Yun Q, et al. The complete chloroplast genome sequence of date palm (phoenix dactylifera L.). Badger JH, editor. PLoS One. 2010 Sep 15; 5(9):1–14. <u>https://doi.org/10.1371/journal.pone.0012762</u> PMID: 20856810
- 70. Zhang YJ, Ma PF, Li DZ. High-throughput sequencing of six bamboo chloroplast genomes: Phylogenetic implications for temperate woody bamboos (Poaceae: Bambusoideae). Poon AFY, editor. PLoS One. 2011 May 31; 6(5):e20596. https://doi.org/10.1371/journal.pone.0020596 PMID: 21655229
- Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, et al. The complete chloroplast genome sequence of Pelargonium × hortorum: Organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol. 2006 Nov 1; 23(11):2175–90.
- 72. lorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol. 2012; 12(1):1–17. <u>https://doi.org/10.1186/1471-2229-12-61 PMID: 22548759</u>
- 73. Pyke KA, Leech RM. Chloroplast division and expansion is radically altered by nuclear mutations in Arabidopsis thaliana. Plant Physiol. 1992 Jul 1; 99(3):1005–8. <u>https://doi.org/10.1104/pp.99.3.1005</u> PMID: 16668963
- 74. Boffey SA, Ellis JR, Selldén G, Leech RM. Chloroplast Division and DNA Synthesis in Light-grown Wheat Leaves. Plant Physiol. 1979 Sep 1; 64(3):502–5. https://doi.org/10.1104/pp.64.3.502 PMID: 16660998
- 75. Hassan L, Wazuddin M. Colchicine-induced variation of cell size and chloroplast number in leaf mesophyll of rice. Plant Breed. 2000 Dec; 119(6):531–3.
- Porebski S, Bailey LG, Baum BR. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Plant Mol Biol Report. 1997 Mar; 15(1):8–15.
- 77. Wang W, Vignani R, Scali M, Cresti M. A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. Electrophoresis. 2006; 27(13):2782–6. https://doi.org/10. 1002/elps.200500722 PMID: 16732618
- 78. Vieira LDN, Faoro H, De Freitas Fraga HP, Rogalski M, De Souza EM, De Oliveira Pedrosa F, et al. An Improved Protocol for Intact Chloroplasts and cpDNA Isolation in Conifers. PLoS One. 2014 Jan 2; 9(1):e84792. https://doi.org/10.1371/journal.pone.0084792 PMID: 24392157

- 79. Liu D, Cui Y, Li S, Bai G, Li Q, Zhao Z, et al. A New Chloroplast DNA Extraction Protocol Significantly Improves the Chloroplast Genome Sequence Quality of Foxtail Millet (Setaria italica (L.) P. Beauv.). Sci Reports 2019 91. 2019 Nov 7; 9(1):1–9. https://doi.org/10.1038/s41598-019-52786-2 PMID: 31700055
- Garaycochea S, Speranza P, Alvarez-Valin F. A Strategy to Recover a High-Quality, Complete Plastid Sequence from Low-Coverage Whole-Genome Sequencing. Appl Plant Sci. 2015 Oct; 3 (10):1500022.
- Straub SCK, Parks M, Weitemier K, Fishbein M, Cronn RC, Liston A. Navigating the tip of the genomic iceberg: Next-generation sequencing for plant systematics. Am J Bot. 2012 Feb; 99(2):349–64. https://doi.org/10.3732/ajb.1100335 PMID: 22174336
- Twyford AD, Ness RW. Strategies for complete plastid genome sequencing. Mol Ecol Resour. 2017; 17(5):858–68. https://doi.org/10.1111/1755-0998.12626 PMID: 27790830
- Coissac E, Hollingsworth PM, Lavergne S, Taberlet P. From barcodes to genomes: extending the concept of DNA barcoding. Mol Ecol. 2016 Apr 1; 25(7):1423–8. <u>https://doi.org/10.1111/mec.13549</u> PMID: 26821259
- Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraaijeveld K, Stielow B, et al. Genomic Treasure Troves: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. Caramelli D, editor. PLoS One. 2013 Jul 29; 8(7):e69189. <u>https://doi.org/10.1371/journal.pone.0069189</u> PMID: 23922691
- Malé PJG, Bardon L, Besnard G, Coissac E, Delsuc F, Engel J, et al. Genome skimming by shotgun sequencing helps resolve the phylogeny of a pantropical tree family. Mol Ecol Resour. 2014 Apr 1; 14 (5):966–75. https://doi.org/10.1111/1755-0998.12246 PMID: 24606032
- Biophysik P:, Suhai S. MIRA: An Automated Genome and EST Assembler. Available from: http://archiv.ub.uni-heidelberg.de/volltextserver/7871/1/thesis_zusammenfassung.pdf.
- Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. ABySS: A parallel assembler for short read sequence data. Genome Res. 2009 Jun 1; 19(6):1117–23. <u>https://doi.org/10.1101/gr.</u> 089532.108 PMID: 19251739
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 20(2):265–72. <u>https://doi.org/10.1101/gr.097261</u>. 109 PMID: 20019144
- Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008 May; 18(5):821–9. https://doi.org/10.1101/gr.074492.107 PMID: 18349386
- Hahn C, Bachmann L, Chevreux B. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—A baiting and iterative mapping approach. Nucleic Acids Res. 2013 Jul; 41(13):e129. https://doi.org/10.1093/nar/gkt371 PMID: 23661685
- McKain, Michael R., Wilson M. Fast-Plast: Rapid de novo assembly and finishing for whole chloroplast genomes. Https://Github.Com/Mrmckain. 2019.
- 92. Fayaz A. Encyclopedia of Tropical Plants. UNSW Press; 2011. 720 p.
- Spongberg SA, Boufford DE. Acta Phytotaxonomica Sinica—a bibliographic summary of published volumes. Taxon. 1982; 31(4):705–7.
- 94. Bandusekara BS, Pushpakumara DKNG, Bandaranayake PCG. Comparison of High-Performance Liquid Chromatography (HPLC) Profiles and Antimicrobial Activity of Different Cinnamomum Species in Sri Lanka. Trop Agric Res. 2023 Apr 1; 34(2):126–35.
- 95. Dong W, Xu C, Li C, Sun J, Zuo Y, Shi S, et al. ycf1, the most promising plastid DNA barcode of land plants. Sci Rep. 2015 Feb 12; 5(1):1–5. https://doi.org/10.1038/srep08348 PMID: 25672218
- **96.** Xiao TW, Ge XJ. Plastome structure, phylogenomics, and divergence times of tribe Cinnamomeae (Lauraceae). BMC Genomics. 2022 Dec 1; 23(1):1–14.
- Blume KL. Bijdragen tot de flora van Nederlandsch Indië /uitgegeven door C.L. Blume. Vols. 1825–26 pt, Bijdragen tot de flora van Nederlandsch Indië /uitgegeven door C.L. Blume. Ter Lands Drukkerij; 2011.
- 98. Berchtold F, Presl JS. O Prirozenosti Rostlin aneb Rostlinár. Vol. 2. 1823. 37-44 p.
- RV RK, Santhoshkumar ES, Radhamany PM. Systematic significance of pollen morphology in south indian species of Cinnamomum schaeffer (Lauraceae). Int J Recent Sci Res. 2019; 10(11):35918–24.
- 100. Naranpanawa DN, Chandrasekara C, Bandaranayake PC, Bandaranayake AU. Raw transcriptomics data to gene specific SSRs: a validated free bioinformatics workflow for biologists. Sci Rep 10 (1): 18236. https://doi.org/10.1038/s41598-020-75270-8 PMID: 33106560