

The first nuclear genome assembly of quina (*Cinchona officinalis*), Peru's national tree and historical source of quinine

Cristian Mendoza

Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM)

Carla L. Saldaña

Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM)

Amelia Menacho

Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM)

Wilian Salazar

Universidad Nacional Agraria La Molina

Carlos I. Arbizu

`carlos.arbizu@untrm.edu.pe`

Universidad Nacional Toribio Rodríguez de Mendoza de Amazonas (UNTRM)

Research Article

Keywords: Rubiaceae, draft genome assembly, Andean forests, genomics

Posted Date: June 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9762034/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Cinchona officinalis L. is an endemic Andean tree of major historical, pharmacological, and cultural significance, recognized as the national tree of Peru and the primary natural source of quinine, the principal antimalarial alkaloid of the pre-synthetic era. Despite its importance, genomic resources for the species remain extremely limited, constraining research on alkaloid biosynthesis, comparative genomics, and conservation genetics. Here we report the first draft nuclear genome assembly and genome-wide simple sequence repeat (SSR) survey for *C. officinalis*, generated from a specimen collected in the Cajamarca region of Peru. Paired-end sequencing on the Illumina HiSeq 2500 platform produced 347.6 million reads representing 52.1 Gbp of data. K-mer frequency analysis estimated a haploid genome size of ~1,081 Mbp, with a genome-wide heterozygosity of ~1.10% and a repeat content of ~13.8%. De novo assembly yielded 293,692 nuclear contigs spanning 791.2 Mbp, with a contig N50 of 10,657 bp and 75.2% BUSCO completeness. Reference-guided scaffolding against the chromosome-level genome of *C. calisaya* increased scaffold N50 to 39.9 Mbp and improved BUSCO completeness to 98.6%, comparable to other published Rubiaceae assemblies. Repeat annotation revealed that 62.63% of the genome is composed of repetitive elements, dominated by LTR retrotransposons, consistent with patterns reported across the family. A total of 457,680 SSR loci were identified, revealing interspecific differences in abundance and motif composition relative to *C. calisaya* and *C. pubescens*. This assembly establishes the first nuclear genomic resource for Peru's national tree and provides a foundation for future investigations into quinine biosynthesis pathways, population genetics, and conservation-oriented management of threatened *Cinchona* populations.

Introduction

The Tropical Andes constitute one of the most biologically diverse regions on Earth, harboring at least 20,000 endemic plant species (Myers et al., 2000). Within this hotspot, Peru represents the principal center of diversity for the genus *Cinchona*, containing 20 of the 24 species currently recognized worldwide (Álvarez, 2013). *Cinchona* L. (Rubiaceae) comprises a lineage of Andean tree species distributed from southern Venezuela to Bolivia, with the highest concentration of diversity and endemism occurring in southern Ecuador and northern Peru (González-Orozco et al., 2023; Gonzalez-Orozco, 2021). Most species inhabit montane cloud forests between 600 and 3,200 m a.s.l., where their distributions are often geographically restricted and only partially overlapping despite documented cases of interspecific hybridization (Andersson, 1998). Beyond its biogeographic importance, the genus occupies a singular position within medicinal plant history. Economically, *Cinchona* ranks second only to *Coffea* within Rubiaceae (Canales et al., 2022), largely due to its ability to synthesize at least 35 quinoline alkaloids, including quinine, quinidine, cinchonidine, and cinchonine (Kacprzak, 2013; Lombe et al., 2026). These compounds constitute a chemically distinctive class of metabolites whose biosynthetic pathways have only recently begun to be comprehensively resolved. Among them, quinine remains particularly notable because it served for more than four centuries as the principal treatment against malaria and later provided the structural basis for the development of synthetic antimalarial drugs (Kaufman & Rúveda, 2005; Sullivan, 2011).

Among the species within the genus, *Cinchona officinalis* L. occupies a unique historical and scientific position. It was the first species formally described within *Cinchona* (Andersson, 1998), named by Linnaeus in 1753 from material collected near Loja, Ecuador, by La Condamine in 1737 (Garmendia, 2017). Commonly referred to as the “quina tree,” it is recognized as the national tree of Peru, although its current distribution within the country is restricted to fragmented Andean forests in the northern departments of Cajamarca and Piura (Huamán et al., 2019). The medicinal relevance of the species predates European contact. During the colonial period, Spanish missionaries acquired knowledge of the febrifuge properties of *Cinchona* bark from Indigenous communities (Jaramillo-Arango, 1949; Espejo, 1792/1993), after which the bark entered European medicine under names such as *Cortex peruanus* and “Jesuit’s powder” (Gachelin et al., 2017). Following its introduction into European pharmacology around 1640, *Cinchona* bark remained the dominant antimalarial therapy until the isolation of pure quinine in 1820 displaced the use of crude botanical preparations (Kacprzak, 2013). Within the genus, *C. officinalis* is especially notable because its bark can contain alkaloid concentrations reaching up to 14% of dry weight, among the highest reported for commercially relevant *Cinchona* species (Kacprzak, 2013; Aslam et al., 2023).

Despite its historical, pharmacological, and cultural importance, *C. officinalis* currently faces severe conservation pressures throughout its native range. Deforestation, agricultural expansion, livestock grazing, and road construction have collectively reduced and fragmented Andean forest habitats across northern Peru and Ecuador (González-Orozco et al., 2023; Huamán et al., 2019). Population genetic analyses conducted in Ecuador indicate that the remaining populations retain only moderate levels of genetic diversity and are frequently restricted to degraded pasture margins (Cueva-Agila et al., 2019). At broader spatial scales, species distribution modeling suggests that climatically suitable habitat for *Cinchona* currently occupies only a limited fraction of the Peruvian territory, while future climate projections indicate restricted recovery potential and uneven conservation suitability among regions (Vergara et al., 2023; Tafur et al., 2025; Coronel-Castro et al., 2024; García et al., 2022). These findings collectively underscore the need for genomic resources capable of supporting evidence-based conservation, population management, and long-term monitoring strategies for the species.

The development of such resources remains surprisingly limited considering the global relevance of the genus. Available genomic information for *Cinchona* is currently restricted to the chloroplast genome of *C. officinalis* published in 2021 (Arbizu et al., 2021), a draft nuclear assembly of *C. pubescens* generated using hybrid long-read and short-read sequencing (Canales et al., 2022), and a near chromosome-level assembly of *C. calisaya* anchored into 17 pseudochromosomes (Du et al., 2025). Consequently, *C. officinalis*, despite being the type species of one of the most pharmacologically significant plant genera in history, still lacks a nuclear genome assembly. This absence limits investigation into the molecular basis of quinine biosynthesis, constrains comparative genomic analyses across the genus, and restricts the development of population-level conservation tools. Although advances in sequencing technologies have accelerated medicinal plant genomics substantially, achieving high assembly continuity and accurate annotation in complex plant genomes remains challenging, particularly in highly repetitive regions (Sahu et al., 2022) and biosynthetic gene clusters (Nützmann et al., 2016).

Whole-genome sequencing additionally enables the identification of molecular markers directly applicable to conservation genetics. Simple sequence repeats (SSRs) represent highly polymorphic and codominantly inherited markers widely used in studies of genetic diversity, population structure, and conservation management (Kalia et al., 2011; Annapurna et al., 2021). For a species such as *C. officinalis*, whose populations are fragmented, geographically restricted, and poorly represented within protected areas, the development of genome-wide SSR resources is therefore of direct practical relevance.

Here, we present the first draft nuclear genome assembly of *C. officinalis* generated from a Peruvian specimen, together with a genome-wide characterization of SSR loci. This study establishes a foundational genomic resource for Peru's national tree and provides a framework for future research into quinine biosynthesis, comparative genomics, conservation genetics, and the long-term management of one of the most emblematic medicinal tree species of the Andes.

Materials and methods

Sample collection and DNA extraction

Young leaves of *Cinchona officinalis* L. were collected from the Cajamarca region of northern Peru (6°00'58.0"S, 79°30'27.8"W) from an individual maintained within the Forestry Program of the National Institute of Agricultural Innovation (INIA). A voucher specimen corresponding to the sampled individual was deposited at the Herbarium of the Universidad Nacional de Cajamarca under accession number CPUN 24126. Total genomic DNA was extracted following the CTAB protocol of Doyle and Doyle (1987), and DNA integrity was verified by electrophoretic visualization of 900 ng on a 1% agarose gel. Paired-end sequencing libraries were generated and sequenced on the Illumina HiSeq 2500 platform using 150 bp reads.

DNA sequencing and estimation of genome size

Raw sequencing quality was initially evaluated with FastQC v0.12.1 (Andrews 2014). Adapter removal and quality trimming were conducted using three strategies: Trimmomatic v0.39 (Bolger et al. 2014), TrimGalore v0.6.11 (Krueger 2012) with default parameters, and TrimGalore v0.6.11 with a minimum read length threshold of 50 bp (--length 50). In all TrimGalore runs, adapter sequences were identified using the default automatic detection mode; the Illumina TruSeq adapter (5'-AGATCGGAAGAGC-3') was detected and removed with a minimum overlap of 5 bp, a maximum error rate of 0.1, and a Phred quality score cutoff of 20, consistent with TrimGalore defaults. Post-filtering quality metrics were subsequently summarized with MultiQC v1.33 (Ewels et al. 2016).

Genome size, heterozygosity, and repetitive content were estimated through k-mer frequency analysis. Canonical k-mers were counted using Jellyfish v2.3.1 (Marçais and Kingsford 2011) at three k-mer sizes ($k = 17$, $k = 21$, $k = 31$), and the resulting distributions were analyzed in GenomeScope v2.0.1 (Ranallo-

Benavidez et al. 2020) under a diploid model (ploidy = 2), consistent with the known ploidy of *Cinchona officinalis*, with a maximum k-mer coverage cutoff of 120.

Genome assembly and quality assessment

De novo genome assembly was performed using MaSuRCA v4.1.4 (Zimin et al. 2013) with an insert size of 300 bp and standard deviation of 45 bp, with GRAPH_KMER_SIZE=auto, LHE_COVERAGE = 25, and CA_PARAMETERS=cgwErrorRate = 0.10. All remaining parameters were set to default values. Assembly continuity statistics were calculated with QUAST v5.3.0 (Gurevich et al. 2013). Gene space completeness was evaluated using BUSCO v6.0.0 (Manni et al. 2021) against the embryophyta_odb10 (n = 1,614) dataset. A k-mer database was constructed from trimmed paired-end reads using Meryl v1.4.1 with k = 21, and consensus quality value (QV) and k-mer completeness were assessed using Merqury v1.3 (Rhie et al. 2020). Read mapping performance was assessed by aligning trimmed paired-end reads against the draft assembly using Bowtie2 v2.4.5 (Langmead and Salzberg 2012) with default parameters and 48 threads (-p 48). The resulting SAM file was converted to BAM format, coordinate-sorted, and indexed using SAMtools v1.21 (Li et al. 2009). Per-contig mean sequencing coverage was calculated using samtools coverage v1.21 applied to the coordinate-sorted BAM file, which reports the mean depth of coverage per reference sequence. The resulting per-contig coverage values were used to estimate the mean nuclear genome coverage and to support contamination filtering decisions.

Potential vector contamination was screened using BLASTn v2.16.0 (Altschul et al. 1990) against the UniVec database with thresholds of $\geq 90\%$ identity and $e\text{-value} \leq 0.01$. Taxonomic assignment of assembled contigs was performed with DIAMOND v2.1.23 (Buchfink et al. 2021) in blastx mode against the NCBI non-redundant protein database (downloaded May 2026) using an e-value cutoff of $\leq 1 \times 10^{-25}$. Contigs assigned to non-Embryophyta_odb10 taxa and presenting mean sequencing coverage below 5 $\times$ were removed as probable contaminants following the criteria proposed by Laetsch and Blaxter (2017).

Organellar-derived sequences were identified by aligning the draft assembly against the published *C. officinalis* chloroplast genome (MZ151891) and the *C. calisaya* mitochondrial genome (PQ963858) using minimap2 v2.28 with the assembly-to-assembly preset (-ax asm5). Contigs producing at least one mapped alignment were classified as organellar and excluded from the nuclear assembly. The filtered nuclear assembly was then scaffolded with RagTag v2.1.0 (Alonge et al. 2022) using the chromosome-level genome of *C. calisaya* (GCA_046055965) as reference, with default alignment parameters and no gap-filling. Following scaffolding, assembly continuity and completeness were re-evaluated with QUAST v5.3.0 and BUSCO v6.0.0.

Repeat annotation and SSR data mining

Repetitive elements were characterized de novo with RepeatModeler v2.0.7 (Flynn et al. 2020) using the LTRStruct pipeline to improve structural detection of LTR retrotransposons. Analyses were conducted on the scaffolded assembly to maximize the recovery of repetitive element boundaries and consensus sequences, generating a species-specific repeat library for *C. officinalis*. No external curated databases

such as Dfam or Repbase were incorporated into the library construction or annotation steps; the repeat library used for masking consisted exclusively of families identified de novo from the *C. officinalis* genome. This library was subsequently used by RepeatMasker v4.1.5 (Smit et al. 2013) to annotate and soft-mask repetitive regions across the scaffolded genome, using the following parameters: -lib (species-specific de novo library), -pa 60 (parallel threads), -xsmall (soft-masking mode), and -gff (GFF output format).

SSRs were identified from the nuclear contig-level assembly using MISA v2.1 (Beier et al. 2017). The contig assembly was intentionally selected for SSR mining to minimize false or truncated loci introduced by scaffold gap junctions. Minimum repeat unit thresholds were established as follows: mononucleotide (≥ 12 repeat units), dinucleotide (≥ 6), trinucleotide (≥ 4), tetranucleotide (≥ 3), pentanucleotide (≥ 3), and hexanucleotide (≥ 3), with a maximum interruption distance of 100 bp between two SSRs to define compound microsatellites (definition: 1–12 2–6 3–4 4–3 5–3 6–3; interruptions: 100). For comparative analyses across the genus, identical parameters were applied to the reference genomes of *C. calisaya* (GCA_046055965.1) and *C. pubescens* (GCA_025175665.1).

Results

The raw sequencing data generated in this study have been deposited in NCBI under BioProject [pending], BioSample [pending], and SRA accession [pending]. Sequencing of *C. officinalis* produced 347,555,802 paired-end reads with a mean read length of 150 bp, corresponding to a total yield of 52.1 Gbp and an overall GC content of 34%. Quality filtering with Trimmomatic retained 81.4% of raw reads, whereas TrimGalore retained 99.9% under default parameters and 99.8% when a minimum read length threshold of 50 bp was applied. Because the latter strategy maintained Phred quality scores above 35 across all nucleotide positions while reducing residual adapter contamination to below 1%, it was selected for all downstream analyses, retaining a total of 346,761,743 high-quality reads.

Genome survey and de novo assembly

K-mer frequency analysis at three k-mer sizes revealed consistent variation in genome size estimates: k=17 produced a biologically implausible estimate of ~274 Mbp with no detectable repeat content, while k=31 overestimated the haploid genome size at ~1,244 Mbp. K=21 yielded the most consistent estimate (~1,081 Mbp) with a model fit of 98.5–98.8%, a repeat content of ~13.8%, and a read error rate of 0.37%, and was therefore selected for downstream analyses. Based on this k-mer size, GenomeScope v2.0.1 revealed a diploid genomic profile characterized by a heterozygous coverage peak at approximately 23.5× and a homozygous peak near 43×, with a genome-wide heterozygosity of ~1.10% (Table 1; Figure 1).

Table 1 Genome survey statistics for *C. officinalis*

Parameter	Min	Max
Homozygous (aa)	98.8827%	98.9194%
Heterozygous (ab)	1.0806%	1.1173%
Genome Haploid Length (bp)	1,079,519,822	1,083,640,794
Genome Repeat Length (bp)	148,940,445	149,509,013
Genome Unique Length (bp)	930,579,377	934,131,782
Model Fit	98.5605%	98.8445%
Read Error Rate	0.3735%	0.3735%

De novo assembly with MaSuRCA produced 293,757 contigs spanning a total assembled length of 791,276,176 bp, corresponding to approximately 73.2% of the estimated haploid genome size, with a GC content of 32.82%, a contig N50 of 10,657 bp, and no ambiguous bases (Table 2). Assembly integrity was supported by multiple independent metrics: 97.43% of trimmed paired-end reads successfully remapped to the assembled contigs, of which 92.16% aligned as properly paired reads, yielding a mean nuclear coverage of $\sim 85.7\times$ as estimated from samtools coverage applied to the coordinate-sorted BAM file. In parallel, Merqury v1.3 estimated a consensus quality value (QV) of 43.45 and a k-mer completeness of 87.30%, collectively indicating high base-level accuracy of the assembly.

Table 2 Assembly statistics and BUSCO completeness of *C. officinalis* genome across pipeline stages

Statistic	Raw assembly	Nuclear assembly	Scaffolded assembly
N50 (bp)	10,657	10,656	39,877,460
N90 (bp)	4,398	4,398	25,279,696
L50	18,304	18,304	8
L90	54,016	54,015	17
Total length (bp)	791,276,176	791,220,560	802,912,260
Total length ($\geq 1,000$ bp)	708,878,855	708,848,058	735,015,081
Largest contig (bp)	80,268	80,268	50,557,737
GC (%)	32.82	32.82	32.81
Total sequences (≥ 0 bp)	293,757	293,692	176,775
# contigs ($\geq 1,000$ bp)	122,795	122,792	29,894
# contigs ($\geq 10,000$ bp)	20,416	20,414	1,506
# contigs ($\geq 50,000$ bp)	40	40	17
Ns per 100 kbp	0	0	1,666.46
Complete BUSCOs	75.20%	75.20%	98.60%
Single-copy BUSCOs	69.90%	69.90%	91.00%
Duplicated BUSCOs	5.30%	5.30%	7.60%
Fragmented BUSCOs	20.50%	20.50%	0.70%
Missing BUSCOs	4.30%	4.30%	0.70%

Contamination screening and nuclear genome definition

Screening against the UniVec database identified a single vector-derived contig, which was subsequently removed from the assembly. Additional contamination filtering was performed through taxonomic assignment with DIAMOND v2.1.23 against the NCBI nr protein database combined with per-contig coverage evaluation. This analysis identified 39 contigs predominantly assigned to non-*Viridiplantae* taxa that also exhibited mean sequencing coverage below 5 $\times$, consistent with low-abundance external contamination. These contigs were therefore excluded from the nuclear assembly. Organellar-derived sequences were identified independently through alignment against the published chloroplast genome of *C. officinalis* (MZ151891) and the mitochondrial genome of *C. calisaya* (PQ963858) using minimap2 v2.28, resulting in the detection and removal of 25 additional contigs of plastidial or mitochondrial origin.

Following decontamination and organellar filtering, the final nuclear assembly comprised 293,692 contigs with a total assembled length of 791,220,560 bp. BUSCO analysis recovered 75.2% complete

conserved orthologs, indicating that the filtering procedure removed contaminant and organellar sequences without producing substantial loss of nuclear genic content (Table 2).

Reference-guided scaffolding and assembly validation

Reference-guided scaffolding using the published chromosome-scale assembly of *C. calisaya* as scaffolding reference. produced a substantial improvement in assembly continuity relative to the pre-scaffolded nuclear assembly. Contiguity increased 3,742-fold following scaffolding, accompanied by a marked increase in BUSCO completeness from 75.2% to 98.6% (Table 2), suggesting improved reconstruction of fragmented conserved orthologs. A total of 116,934 contigs were successfully anchored into pseudochromosomes, whereas 176,758 contigs remained unplaced after scaffolding (Table 3).

Table 3 Scaffolding statistics of *C. officinalis* nuclear assembly using *C. calisaya* as reference

Metric	Value
Placed sequences	116,934.00
Placed bases (bp)	626,719,955.00
Placed bases (%)	79.20
Unplaced sequences	176,758.00
Unplaced bases (bp)	164,500,605.00
Unplaced bases (%)	20.80
Gap sequences	116,917.00
Gap bases (bp)	11,691,700.00

Repeat landscape and SSR identification

Repeat annotation identified 502,870,160 bp of repetitive and low-complexity sequence within the scaffolded *C. officinalis* genome, corresponding to 62.63% of the assembled nuclear genome (Table 4). Interspersed repeats represented the largest component, accounting for 492,215,903 bp (61.30%). Among these, retroelements were the predominant category, occupying 209,819,881 bp (26.13% of the genome). LTR retrotransposons constituted the major fraction of retroelements, spanning 182,771,230 bp (22.76%), with Gypsy/DIRS1 elements (13.80%) considerably more abundant than Ty1/Copia elements (7.94%). LINE elements represented 3.37% of the genome, largely dominated by L1/CIN4 families (2.75%), whereas no SINE elements were detected. DNA transposons occupied 13,498,471 bp, corresponding to 1.68% of the assembly, with MULE-MuDR (0.63%) and hobo-Activator (0.52%) as the most abundant superfamilies. Rolling-circle elements accounted for an additional 0.34% of the genome. A substantial proportion of repetitive sequence remained unclassified, totaling 268,897,551 bp or 33.49%

of the assembly. Simple repeats and low-complexity regions accounted for 0.74% and 0.17% of the genome, respectively (Table 4).

Table 4 Repetitive DNA content of the *C. officinalis* scaffolded genome

Repetitive element	Number of elements	Length (bp)	% of genome
Retroelements	260,687	209,819,881	26.13
SINEs	0	0	0.00
LINEs	39,412	27,048,651	3.37
RTE/Bov-B	7,410	5,006,198	0.62
L1/CIN4	32,002	22,042,453	2.75
LTR elements	221,275	182,771,230	22.76
Ty1/Copia	86,267	63,726,763	7.94
Gypsy/DIRS1	120,808	110,836,406	13.80
DNA transposons	27,735	13,498,471	1.68
hobo-Activator	10,082	4,202,257	0.52
MULE-MuDR	10,147	5,076,198	0.63
Tourist/Harbinger	958	586,471	0.07
Rolling-circles	3,176	2,698,400	0.34
Unclassified	840,888	268,897,551	33.49
Total interspersed repeats		492,215,903	61.30
Small RNA	3,922	603,342	0.08
Simple repeats	168,094	5,960,552	0.74
Low complexity	29,060	1,391,963	0.17

A total of 457,680 SSR loci were identified across the *C. officinalis* nuclear genome. Mononucleotide repeats represented the most abundant class (166,732 loci), followed by tetranucleotide (93,749), dinucleotide (84,581), trinucleotide (61,804), pentanucleotide (33,359), and hexanucleotide repeats (17,455) (Figure 2; Table 5). Among mononucleotide repeats, A/T motifs were predominant, with A accounting for 84,072 loci and T for 73,654 loci, while C and G motifs were comparatively rare (4,758 and 4,248 loci, respectively). Among dinucleotide SSRs, AT/TA repeats were the most abundant (59,642 loci combined), followed by AC/CA (7,818) and TG/GT (7,270). Within trinucleotide SSRs, AAG (12,564), TAA (12,157), and ATA (12,065) were the most frequent motifs. A consistent AT-rich pattern was observed across all repeat classes, with AT-rich motifs dominating tetranucleotide (AAAT: 10,480; TTTA: 6,724;

AATT: 5,930), pentanucleotide (AAAAT: 2,756; TTAAT: 2,049; ATTAA: 1,992), and hexanucleotide (AAAAAT: 866; AAAAAG: 570; TTTTTA: 408) SSRs (Table 5). Comparative analyses with the genomes of *C. calisaya* and *C. pubescens* revealed clear interspecific differences in both total SSR abundance and repeat motif composition, with *C. officinalis* exhibiting a particularly elevated proportion of mononucleotide repeats relative to both reference species (Figure 2).

Table 5 Distribution of the most abundant SSR motifs identified in the *C. officinalis* nuclear genome

SSR type	Total SSRs	First most abundant motif		Second most abundant motif		Third most abundant motif	
Mononucleotide	166,732	A	(84,072)	T	(73,654)	C	(4,758)
Dinucleotide	84,581	AT/TA	(59,642)	TG/GT	(7,27)	AC/CA	(7,818)
Trinucleotide	61,804	AAG	(12,564)	TAA	(12,157)	ATA	(12,065)
Tetranucleotide	93,749	AAAT	(10,480)	TTTA	(6,724)	AATT	(5,930)
Pentanucleotide	33,359	AAAAT	(2,756)	TTAATT	(2,049)	ATTAA	(1,992)
Hexanucleotide	17,455	AAAAAT	(866)	AAAAAG	(570)	TTTTTA	(408)

Discussion

Among the trimming strategies evaluated, TrimGalore with a minimum length threshold of 50 bp outperformed Trimmomatic in read retention (99.8% vs. 81.4%), consistent with previous observations showing that excessively stringent quality filtering can remove a substantial proportion of informative reads without proportional improvements in assembly quality (Del Fabbro et al., 2013). K-mer size selection also required careful evaluation, as k=17 produced a biologically implausible genome size estimate of ~274 Mbp without detectable repeat content, whereas k=31 substantially overestimated the haploid genome at ~1,244 Mbp. By contrast, k=21 generated the most coherent profile, yielding an estimated genome size of ~1,081 Mbp, a model fit of 98.5–98.8%, and a repeat content of ~13.8% consistent with the repetitive landscapes reported for related *Cinchona* species, supporting its selection for downstream analyses and aligning with the strategy previously applied for the genome characterization of *C. pubescens* (Canales et al., 2022). The estimated haploid genome size of ~1,081 Mbp places *C. officinalis* within the upper range reported for the genus, exceeding the ~822 Mbp described for *C. calisaya* (Du et al., 2025) and approaching the ~1,100 Mbp estimated for *C. pubescens* (Canales et al., 2022), consistent with the broad genome size variation documented across Rubiaceae. Similarly, the genome-wide heterozygosity of ~1.10% falls between the 0.69% reported for *N. cadamba* and the 1.32% observed in *M. officinalis* (Zhao et al., 2022; Wang et al., 2021), suggesting moderate standing genetic diversity despite the species’ fragmented distribution and prolonged history of selective exploitation.

The recovery of ~73.2% of the estimated genome in the contig-level assembly is consistent with the known limitations of short-read sequencing for resolving highly repetitive genomic regions (Treangen and Salzberg, 2012), particularly relevant considering that repetitive elements comprise 66.12% of the *C. calisaya* genome (Du et al., 2025). Such a repeat-rich genomic architecture likely contributed directly to the fragmented assembly structure observed (N50=10,657 bp), making reference-guided scaffolding against the chromosome-level genome of *C. calisaya* necessary to recover fragmented gene models and improve structural continuity. RagTag v2.1.0 was applied using default alignment parameters, which have previously demonstrated robust anchoring performance for plant genome scaffolding without requiring extensive manual optimization (Alonge et al., 2022). This strategy produced a 3,742-fold increase in contiguity and improved BUSCO recovery from 75.2% to 98.6%, values comparable to the 98.3% reported for *C. calisaya*, the 96.8% described for *N. cadamba*, and the 97.02% observed in *M. officinalis* (Du et al., 2025; Zhao et al., 2022; Wang et al., 2021). The 20.8% of nuclear sequence that remained unplaced likely reflects authentic genomic divergence relative to *C. calisaya* and may include highly divergent or species-specific genomic regions potentially associated with secondary metabolism, particularly considering that phylogenetic structure within *Cinchona* has been associated with substantial variation in alkaloid composition among populations (Maldonado et al., 2017). Additionally, highly repetitive or short contigs refractory to alignment-based anchoring may also contribute to the unplaced fraction (Treangen and Salzberg, 2012; Alonge et al., 2022). The modest increase in duplicated BUSCOs from 5.3% to 7.6% is expected in heterozygous genomes where allelic haplotypes assemble independently as separate contigs, artificially inflating duplication estimates without compromising biological interpretability (Waterhouse et al., 2018).

Repeat annotation revealed that 62.63% of the *C. officinalis* genome is composed of repetitive elements, highly consistent with the 66.12% reported for *C. calisaya* (Du et al., 2025), suggesting that elevated repetitive content represents a conserved genomic characteristic within *Cinchona*. Comparable proportions have also been described across Rubiaceae, including approximately 60.60% in *Coffea arabica*, 58.04% in *M. officinalis*, and 52.94% in *N. cadamba*, indicating that repeat-rich genomes are widespread across the family. LTR retrotransposons dominated the repetitive fraction (22.76%), with Gypsy/DIRS1 elements (13.80%) substantially more abundant than Ty1/Copia elements (7.94%), a pattern consistently reported across angiosperms and widely recognized as a major driver of genome size expansion (Du et al., 2025; Zhao et al., 2022). Notably, no SINE elements were detected in *C. officinalis*, consistent with their apparent absence in *C. calisaya*, potentially reflecting lineage-specific reduction within *Cinchona* or limitations associated with de novo repeat library construction. The large proportion of unclassified repeats (33.49%) is expected given that the repeat library was generated exclusively through de novo prediction without integration of curated external databases, a limitation similarly reported in *M. speciosa*, and highlights the need for improved curation of Rubiaceae-specific transposable element resources. This elevated repetitive content also provides a plausible explanation for the discrepancy between the estimated (~1,081 Mbp) and assembled (~791 Mbp) genome sizes, as highly repetitive regions cannot be reliably resolved during short-read graph assembly (Treangen and Salzberg, 2012). The scaffolded assembly was selected for repeat annotation because longer scaffolds

provide greater sequence context for accurate identification of transposable element boundaries (Flynn et al., 2020), whereas SSR detection was performed on the contig-level assembly to avoid artificial truncation of microsatellite loci at scaffold gap junctions (Beier et al., 2017).

The SSR catalog of 446,794 loci identified in *C. officinalis* is intermediate between the 479,988 reported for *C. calisaya* and the 297,570 described for *C. pubescens*, suggesting that SSR abundance within *Cinchona* is not strictly determined by genome size alone. The predominance of AT-rich motifs across SSR classes is consistent with the AT-rich repetitive landscapes characteristic of plant genomes and has been widely documented across angiosperms (Srivastava et al., 2019). The disproportionately high abundance of mononucleotide repeats observed in *C. officinalis* most likely reflects residual assembly fragmentation rather than genuine biological expansion, as homopolymeric regions are particularly susceptible to misassembly in short-read datasets (Treangen and Salzberg, 2012). In contrast, the relative consistency observed across di- through hexanucleotide SSR classes supports their reliability as molecular markers for population genetics, phylogeographic inference, and conservation-oriented studies focused on *Cinchona* germplasm in Peru.

Conclusion

This study presents the first draft nuclear genome assembly for *C. officinalis*, the national tree of Peru and one of the most historically important medicinal plant species worldwide. Despite the limitations inherent to short-read sequencing in highly repetitive plant genomes, the combined strategy of de novo assembly, contamination filtering, and reference-guided scaffolding produced a highly complete genome resource with BUSCO values comparable to other chromosome-level Rubiaceae assemblies. The genome of *C. officinalis* was characterized by substantial repetitive content, moderate heterozygosity, and extensive SSR diversity, collectively expanding current knowledge of genome architecture within *Cinchona* and Rubiaceae more broadly.

The genome-wide SSR catalog generated here provides an important molecular resource for future studies of population structure, phylogeography, and conservation genetics in fragmented Andean populations. In parallel, the assembly establishes a foundational framework for investigating the genetic basis of quinoline alkaloid biosynthesis, including pathways associated with quinine production that underpin the historical pharmacological relevance of the genus.

Although unresolved repetitive regions and unplaced contigs remain, the present assembly substantially reduces the genomic gap for *Cinchona*. Future integration of long-read sequencing technologies and chromatin conformation scaffolding approaches will further improve assembly continuity and facilitate the characterization of complex repetitive regions, structural variation, and biosynthetic gene clusters associated with pharmacologically relevant traits.

Declarations

Competing interests

The authors declare no competing interests.

Funding

This research was funded by the project “Creación del servicio de agricultura de precisión en los Departamentos de Lambayeque, Huancavelica, Ucayali y San Martín 4 Departamentos” of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government with grant number CUI 2449640. C.I.A. was supported by PP0068 “Reducción de la vulnerabilidad y atención de emergencias por desastres” and Vicerectorado de Investigación of UNTRM. C.M. is supported by project “Desarrollo y validación de tecnologías para la valorización del rocoto: selección genómica y evaluación agronómica con inteligencia artificial aplicada a su escalamiento productivo en Amazonas” with grant number PE501097536-2025-PROCIENCIA.

Author Contribution

Conceptualization: C.I.A. and W.S.; Methodology: C.M., C.L.S. and A.M.; Software: C.M. and C.L.S.; Validation: C.M., C.L.S. and A.M.; Formal analysis: C.M.; Investigation: C.L.S, W.S. and C.I.A.; Data curation: C.M. and C.L.S.; Writing—original draft preparation: C.M.; Review and editing: C.L.S., W.S. and C.I.A.; Supervision: W.S. and C.I.A. All authors have read and agreed to the published version of the manuscript.

Data Availability

Data were deposited into the NCBI database and are available at the following URL:
<https://www.ncbi.nlm.nih.gov/nucleotide>

References

1. Alonge, M., Lebeigle, L., Kirsche, M., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato systems genetics platform. *Genome Biology*, 23(1), 258. <https://doi.org/10.1186/s13059-022-02823-7>
2. Álvarez, A.J. (2013). El árbol de la calentura. *Boletín Instituto Nacional de Salud*, 19, 214–215.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
4. Andersson, L. (1998). A revision of the genus *Cinchona* (Rubiaceae-Cinchoneae). *Memoirs of the New York Botanical Garden*, 80, 1–75.

5. Andrews, S. (2014). FastQC: A quality control tool for high throughput sequence data. Babraham Bioinformatics. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
6. Annapurna, D., Bhatt, R. K., Bhargav, J., Bhatt, V., & Rawat, S. (2021). Simple sequence repeat (SSR) markers: an overview. *International Journal of Current Microbiology and Applied Sciences*, 10(1), 3679–3694.
7. Arbizu, C. I., Ferro-Mauricio, R. D., Chávez-Galarza, J. C., Guerrero-Abad, J. C., Vásquez, H. V., & Maicelo, J. L. (2021). The complete chloroplast genome of the national tree of Peru, quina (*Cinchona officinalis* L., Rubiaceae). *Mitochondrial DNA Part B*, 6(9), 2781–2783. <https://doi.org/10.1080/23802359.2021.1969697>
8. Aslam, S., Jabeen, T., Ahmad, M., & AL-Huqail, A. A. (2023). Cinchona. In M. Zia-Ul-Haq, A. Abdulkreem AL-Huqail, M. Riaz, & U. Farooq Gohar (Eds.), *Essentials of Medicinal and Aromatic Crops*. Springer. https://doi.org/10.1007/978-3-031-35403-8_10
9. 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>
10. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
11. Buchfink, B., Reuter, K., & Drost, H. G. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nature Methods*, 18(4), 366–368. <https://doi.org/10.1038/s41592-021-01101-x>
12. Canales, N. A., Pérez-Escobar, O. A., Powell, R. F., Töpel, M., Kidner, C., Nesbitt, M., Maldonado, C., Barnes, C. J., Rønsted, N., Przelomska, N. A. S., Leitch, I. J., & Antonelli, A. (2022). A highly contiguous, scaffold-level nuclear genome assembly for the fever tree (*Cinchona pubescens* Vahl) as a novel resource for Rubiaceae research. *GigaByte*, 2022, gigabyte71. <https://doi.org/10.46471/gigabyte.71>
13. Coronel-Castro, E., Meza-Mori, G., Torres, J. M. C., Mondragón, E. P., Cotrina-Sanchez, A., Oliva Cruz, M., López, R. S., & Campo Ramos, R. E. (2024). Potential distribution and identification of critical areas for the preservation and recovery of three species of *Cinchona* L. (Rubiaceae) in northeastern Peru. *Forests*, 15(2), 321. <https://doi.org/10.3390/f15020321>
14. Cueva-Agila, A., Vélez-Mora, D., Arias, D., et al. (2019). Genetic characterization of fragmented populations of *Cinchona officinalis* L. (Rubiaceae), a threatened tree of the northern Andean cloud forests. *Tree Genetics & Genomes*, 15, 81. <https://doi.org/10.1007/s11295-019-1393-y>
15. Del Fabbro, C., Scalabrin, S., Morgante, M., & Giorgi, F. M. (2013). An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS ONE*, 8(12), e85024. <https://doi.org/10.1371/journal.pone.0085024>
16. Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19, 11–15.
17. Du, C., Xiao, Y., Liu, J., He, Q., Li, W., & Xie, W. (2025). A near-complete genome assembly of *Cinchona calisaya*. *Scientific Data*, 12, 122. <https://doi.org/10.1038/s41597-025-04449-3>

18. Espejo, E. (1792/1993). *Voto de un ministro togado de la Audiencia de Quito: edición facsimilar*. Comisión nacional permanente de conmemoraciones cívicas.
19. Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19), 3047–3048. <https://doi.org/10.1093/bioinformatics/btw354>
20. Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., & Smit, A. F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences*, 117(17), 9451–9457. <https://doi.org/10.1073/pnas.1921046117>
21. Gachelin, G., Garner, P., Ferroni, E., Tröhler, U., & Chalmers, I. (2017). Evaluating Cinchona bark and quinine for treating and preventing malaria. *Journal of the Royal Society of Medicine*, 110(1), 31–40. <https://doi.org/10.1177/0141076816681421>
22. García, L., Veneros, J., Chavez, S. G., Oliva, M., & Rojas-Briceño, N. B. (2022). World historical mapping and potential distribution of *Cinchona* spp. in Peru as a contribution for its restoration and conservation. *Journal for Nature Conservation*, 70, 126290. <https://doi.org/10.1016/j.jnc.2022.126290>
23. Garmendia, A. (2017). *Evaluación del paisaje y recursos escénicos después de 350 años de explotación de la "cascarilla", Cinchona officinalis L. (Rubiaceae) en el sector Cajanuma-Rumishitana, Ecuador*. *Arnaldia*, 23(2), 461–474.
24. González-Orozco, C. E., Guillén, E. G., & Cuví, N. (2023). Changes of *Cinchona* distribution over the past two centuries in the northern Andes. *Royal Society Open Science*, 10(4), 230229. <https://doi.org/10.1098/rsos.230229>
25. Gonzalez-Orozco, C. (2021). Regiones biogeográficas del género *Cinchona* L. (Rubiaceae-Cinchoneae). *Revista Novedades Colombianas*, 16(1), 135–156. <https://doi.org/10.47374/novcol.2021.v16.1987>
26. Gawayed, S. M. H., & Abd El-Moneim, D. (2021). Detection of genetic divergence among some wheat (*Triticum aestivum* L.) genotypes using molecular and biochemical indicators under salinity stress. *PLOS ONE*, 16(3), e0248890. <https://doi.org/10.1371/journal.pone.0248890>
27. Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QAST: quality assessment tool for genome assemblies. *Bioinformatics*, 29(8), 1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
28. Huamán, L., Albán, J., & Chilquillo, E. (2019). Aspectos taxonómicos y avances en el conocimiento del estado actual del árbol de la quina (*Cinchona officinalis* L.) en el norte de Perú. *Ecología Aplicada*, 18(2), 145. <https://doi.org/10.21704/rea.v18i2.1333>
29. Jaramillo-Arango, J. (1949). A critical review of the basic facts in the history of Cinchona. *Journal of the Linnean Society of London, Botany*, 53(352), 272–311.
30. Kacprzak, K. M. (2013). Chemistry and biology of Cinchona alkaloids. In *Natural Products* (pp. 605–641). Springer. https://doi.org/10.1007/978-3-642-22144-6_22
31. Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica*, 177(3), 309–334. <https://doi.org/10.1007/s10681-010->

32. Kaufman, T. S., & Rúveda, E. A. (2005). The quest for quinine: those who won the battles and those who won the war. *Angewandte Chemie*, 44(6), 854–885. <https://doi.org/10.1002/anie.200400663>
33. Krueger, F. (2012). Trim Galore: A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Bioinformatics. http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/
34. Laetsch, D. R., & Blaxter, M. L. (2017). BlobTools: Interrogation of genome assemblies. *F1000Research*, 6, 1287. <https://doi.org/10.12688/f1000research.12232.1>
35. Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
36. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
37. Lombe, B. K., Zhou, T., Kang, G., et al. (2026). Biosynthesis of cinchona alkaloids. *Nature*. <https://doi.org/10.1038/s41586-026-10227-x>
38. Maldonado, C., Barnes, C. J., Cornett, C., Holmfred, E., Hansen, S. H., Persson, C., Antonelli, A., & Rønsted, N. (2017). Phylogeny predicts the quantity of antimalarial alkaloids within the iconic yellow cinchona bark (Rubiaceae: *Cinchona calisaya*). *Frontiers in Plant Science*, 8, 391. <https://doi.org/10.3389/fpls.2017.00391>
39. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution*, 38(10), 4647–4654. <https://doi.org/10.1093/molbev/msab199>
40. Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
41. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, 403(6772), 853–858. <https://doi.org/10.1038/35002501>
42. Nützmann, H. W., Huang, A., & Osbourn, A. (2016). Plant metabolic clusters – from genetics to genomics. *New Phytologist*, 211(3), 771–789. <https://doi.org/10.1111/nph.13981>
43. Pootakham, W., Yoocha, T., Jomchai, N., Kongkachana, W., Naktang, C., Sonthirod, C., Chowpongpan, S., Aumpuchin, P., & Tangphatsornruang, S. (2022). A chromosome-scale genome assembly of *Mitragyna speciosa* (kratom) and the assessment of its genetic diversity in Thailand. *Biology*, 11(10), 1492. <https://doi.org/10.3390/biology11101492>
44. Ranallo-Benavidez, T. R., Jaron, K. S., & Schatz, M. C. (2020). GenomeScope 2.0 and Smudgeplots characterize genome heterozygosity in polyploid organisms. *Nature Communications*, 11, 1432. <https://doi.org/10.1038/s41467-020-14998-3>

45. Rhie, A., Walenz, B. P., Koren, S., & Phillippy, A. M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology*, 21, 245. <https://doi.org/10.1186/s13059-020-02134-9>
46. Sahu, S. K., Thangaraj, M., & Kundu, S. (2022). Advances and challenges in plant genome assembly. *Frontiers in Plant Science*, 13, 1012247. <https://doi.org/10.3389/fpls.2022.1012247>
47. Salojärvi, J., Rambani, A., Yu, Z., et al. (2024). The genome and population genomics of allopolyploid *Coffea arabica* reveal the diversification history of modern coffee cultivars. *Nature Genetics*, 56, 721–731. <https://doi.org/10.1038/s41588-024-01695-w>
48. Smit, A. F., Hubley, R., & Green, P. (2013). RepeatMasker Open-4.0. <http://www.repeatmasker.org>
49. Sullivan, D. J. (2011). Cinchona alkaloids: quinine and quinidine. In H. Staines & S. Krishna (Eds.), *Treatment and Prevention of Malaria*. Springer. https://doi.org/10.1007/978-3-0346-0480-2_3
50. Tafur, E., Cuchca, S., García, L., Rojas-Briceño, N. B., & Veneros, J. (2025). Generalized linear models to estimate the probability of occurrence of *Cinchona officinalis* L., *Cinchona pubescens* Vahl, and *Cinchona calisaya* Wedd in Peru. *Journal of Sustainable Forestry*, 44(1–5), 103–125. <https://doi.org/10.1080/10549811.2025.2513230>
51. Treangen, T. J., & Salzberg, S. L. (2012). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics*, 13(1), 36–46. <https://doi.org/10.1038/nrg3117>
52. Vergara, A. J., Cieza-Tarrillo, D., Ocaña, C., Quiñonez, L., Idrogo-Vasquez, G., Muñoz-Astecker, L. D., Auquiñivin-Silva, E. A., Cruzalegui, R. J., & Arbizu, C. I. (2023). Current and future spatial distribution of the genus *Cinchona* in Peru: opportunities for conservation in the face of climate change. *Sustainability*, 15(19), 14109. <https://doi.org/10.3390/su151914109>
53. Wang, J., Xu, S., Mei, Y., Cai, S., Gu, Y., Sun, M., Liang, Z., Xiao, Y., Zhang, M., & Yang, S. (2021). A high-quality genome assembly of *Morinda officinalis*, a famous native southern herb in the Lingnan region of southern China. *Horticulture Research*, 8(1), 135. <https://doi.org/10.1038/s41438-021-00551-w>
54. Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. *Molecular Biology and Evolution*, 35(3), 543–548. <https://doi.org/10.1093/molbev/msx319>
55. Zhao, X., Hu, X., OuYang, K., Yang, J., Que, Q., Long, J., Zhang, J., Zhang, T., Wang, X., Gao, J., Hu, X., Yang, S., Zhang, L., Li, S., Gao, W., Li, B., Jiang, W., Nielsen, E., Chen, X., & Peng, C. (2022). Chromosome-level assembly of the *Neolamarckia cadamba* genome provides insights into the evolution of cadambine biosynthesis. *The Plant Journal*, 109(4), 891–908. <https://doi.org/10.1111/tpj.15600>
56. Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>

Figures

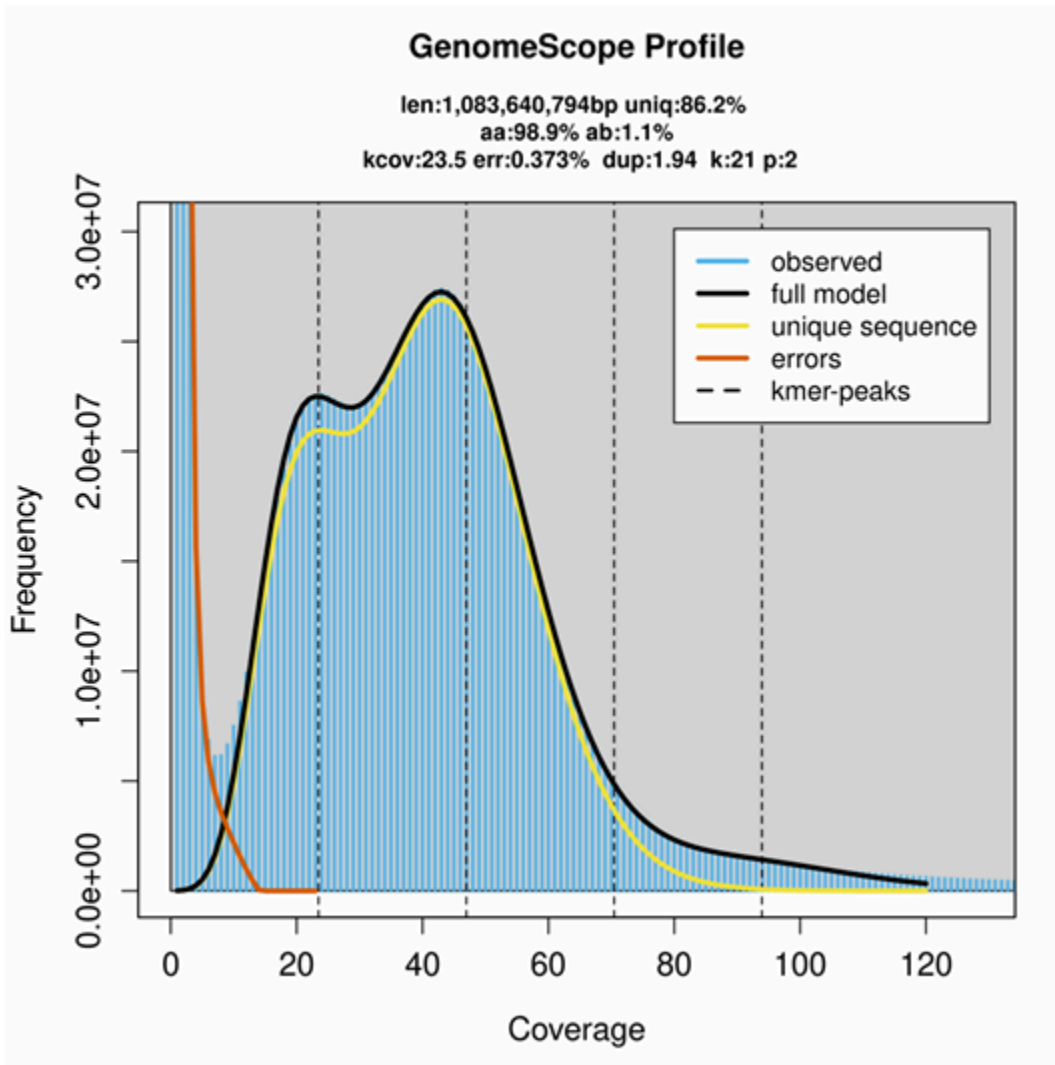


Figure 1

K-mer frequency analysis of *C. officinalis* draft genome. The heterozygous coverage peak at ~23.5x and homozygous peak at ~43x are consistent with a diploid genome

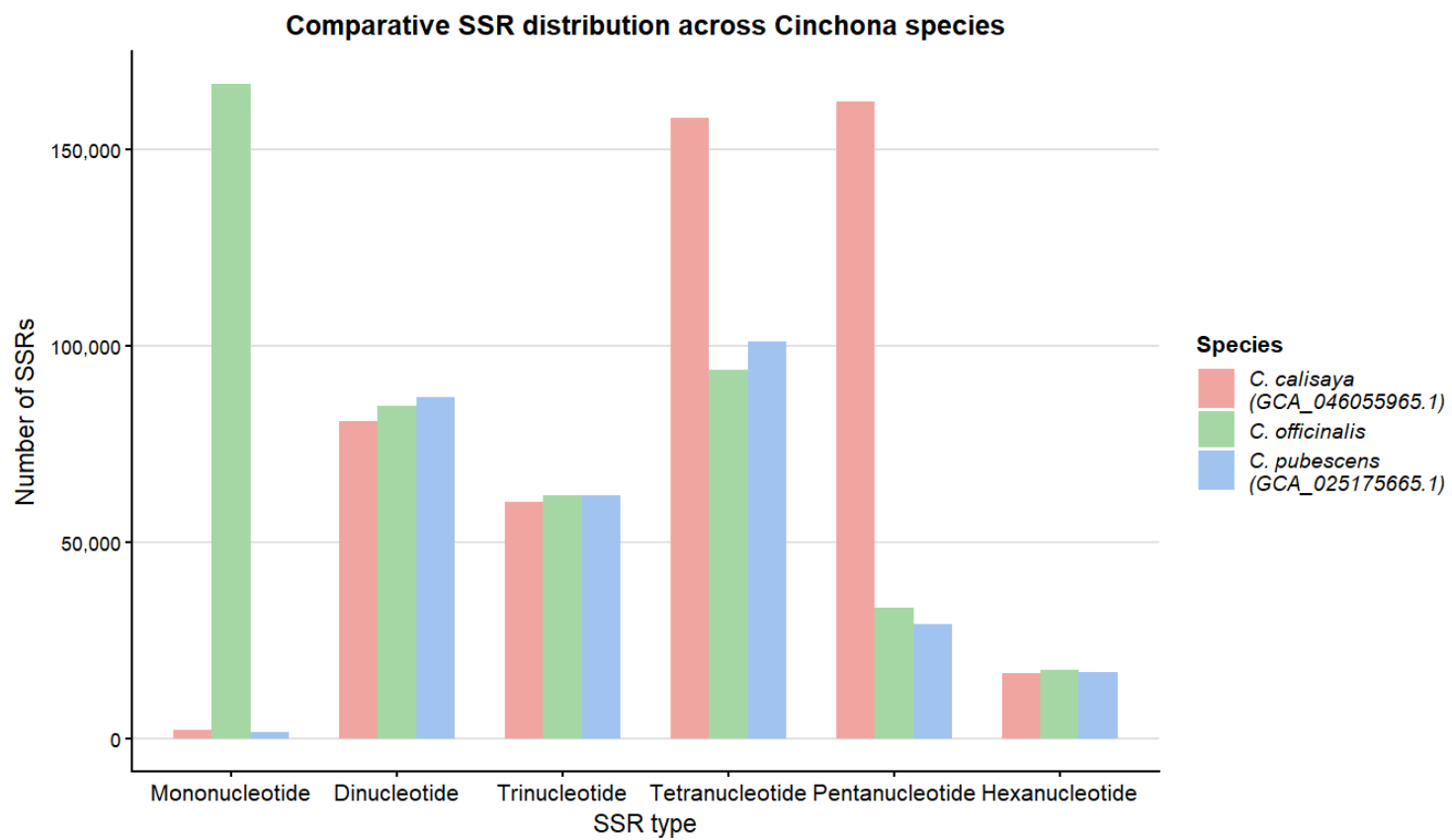


Figure 2

Comparative distribution of SSR types across *Cinchona* species