Article

Unlocking the complete chloroplast genome of a native tree species from the Amazon basin, capirona (*Calycophyllum spruceanum* Benth., Rubiaceae), and its comparative analysis with other Ixoroideae species

Carla L. Saldaña [1], Pedro Rodriguez-Grados [1,2], Julio C. Chávez-Galarza [1], Shefferson Feijoo [3], Juan Carlos Guerrero-Abad [4], Héctor V. Vásquez [1], Jorge L. Maicelo [1], Jorge H. Jhoncon [5,6], Carlos I. Arbizu [1,*]

[1] Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Perú; carla18317@gmail.com (C.L.S); pmrg1711@gmail.com (P.R-G.); jcchavez-galarza@gmail.com (J.C.C-G.); hvasquez@inia.gob.pe (H.V.V.); jmaicelo@inia.gob.pe (J.L.M)

[2] Universidad Nacional José Faustino Sánchez Carrión, Av. Mercedes Indacochea Nro. 609 Huacho, Lima; pmrg1711@gmail.com (P.R-G)

[3] Estación Experimental Agraria San Bernardo. Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Carretera Cusco, Puerto Maldonado, Tambopata, Madre de Dios. Perú; sfeijoo@inia.gob.pe (S.F)

[4] Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Lima, Perú; jguerreroa@inia.gob.pe (J.C.G-A)

[5] Centro Internacional de Investigación para la Sustentabilidad (CIIS), Universidad Nacional de Cañete, Jr. San Agustín N° 124 San Vicente de Cañete, 15701 Cañete, Peru; j_jhoncon@undc.edu.pe (J.H.J.)

[6] Centro de Investigación de Plantas Andinas y Nativas. Facultad de Ciencias. Universidad Nacional de Educación Enrique Guzmán y Valle. Av. Enrique Guzmán y Valle s/n. Lima 15472. Lima. Peru. jjhoncon@une.edu.pe (J.H.J.)

*Correspondence: carbizu@inia.gob.pe

Abstract: Capirona (*Calycophyllum spruceanum* Benth.) belongs to subfamily Ixoroideae, one of de major lineages in the Rubiaceae family, and is an important timber tree, with origin in the Amazon Basin and has widespread distribution in Bolivia, Peru, Colombia, and Brazil. In this study, we obtained the first complete chloroplast (cp) genome of capirona from department of Madre de Dios located in the Peruvian Amazon. High-quality genomic DNA was used to construct libraries. Pair-end clean reads were obtained by PE 150 library and the Illumina HiSeq 2500 platform. The complete cp genome of *C. spruceanum* has a 154,480 bp in length with typical quadripartite structure, containing a large single copy (LSC) region (84,813 bp) and a small single-copy (SSC) region (18,101 bp), separated by two inverted repeat (IR) regions (25,783 bp). The annotation of *C. spruceanum* cp genome predicted 87 protein-coding genes (CDS), 8 ribosomal RNA (rRNA) genes, 37 transfer RNA (tRNA) genes and 01 pseudogene. A total of 41 simple sequence repeats (SSR) of this cp genome were divided into mononucleotides (29), dinucleotides (5), trinucleotides (3), and tetranucleotide (4). Most of these repeats were distributed in the noncoding regions. Whole chloroplast genome comparison with the other six Ixoroideae species revealed that the small single copy and large single copy regions showed more divergence than invert regions. Finally, phylogenetic analysis resolved that *C. spruceanum* is a sister species to *Emmenopterys henryi*, and confirms its position within the subfamily Ixoroideae. This study reports for the first time the genome organization, gene content, and structural features of the chloroplast genome of *C. spruceanum*, providing valuable information for genetic and evolutionary studies in the genus *Calycophyllum* and beyond.

Keywords: chloroplast; genetic resources; genomics capirona; phylogenomics

1. Introduction

The family Rubiaceae is one of the largest and most diverse families of angiosperms, and include the economically important genus, *Coffea*, and the horticulturally important *Gardenia* and *Ixora,* all part of Ixoroideae subfamily [1][2]. This subfamily comprising about 4000 species, of pantropical and subtropical distributions and is one of three major lineages in the Rubiaceae family, and include *Coffea canephora*, *Fosbergia shweliensis*, *Scyphiphora hydrophyllacea*, *Emmenopterys henryi* and *Calycophyllum spruceanum* "capirona"[1]. Capirona is an important timber tree [3], with origin in the Amazon Basin and has widespread distribution from Bolivia, Peru, Colombia, and Brazil [4]. It is a rainforest hardwood tree and is also exported around the world for high density wood, durable lumber and building materials, but also as a medicinal plant. Moreover, it is used for the construction of economically valuable products [5]. It has excellent qualities for field planting or in agroforestry systems combinations. In addition, capirona has good natural regeneration, therefore, it is an ideal species for the management of secondary successions [3]. On the other hand, to date, *C. sprucearum* is considered an orphan forest organism since genetic and genomic resources for this species are limited today. Russell et al.,[3], Tauchen et al., [5] and Saldaña et al.,[6] determined the genetic variation of capirona using molecular markers, such as amplified fragment length polymorphisms (AFLP), internal transcribed spacer (ITS), and random-amplified polymorphic DNA (RAPD) in different populations of capirona from the Peruvian Amazon. Their results demonstrated a greater variation within provenances than among provenances. Also, on the contrary, Dávila-Lara et al [7] used AFLP and reported lower genetic diversity parameters across 13 populations of capirona in Nicaragua (Central America). Capirona is currently the object of research in the Peruvian Amazon basin, in the context of increased deforestation through unsustainable slash and burn agriculture, and also for conservation strategies [3].

Chloroplasts, as metabolic organelles responsible for photosynthesis and the synthesis of amino acids, nucleotides, fatty acids, phytohormones, vitamins, and other metabolites, play an important role in the physiology and development of land plants and algae [8,9] .They have their own genetic replication mechanisms, and they transcribe their own genome relatively independently [10]. In most terrestrial plants, chloroplast genomes resolve highly conserved and organized structures, and occur as circular DNA molecules with a size of 120-170 kb [11] and have a highly conserved quadripartite structure and normally encodes approximately 110-130 genes involved in photosynthesis, transcription and translation process. Also, chloroplast genomes contain two inverted repeat sequences (IR), as well as a large single copy region (LSC) and a small single copy region (SSC) [12,13]. Although the chloroplast genomes of angiosperms are highly conserved, mutational events occur, such as structural rearrangement, insertions, and deletions (InDels), inversions, translocations and variations in the number of copies (CNV). This polymorphism in the chloroplast genome provides invaluable information about population genetics and structure, phylogeny, species barcode analysis, and endangered species conservation and breeding improvement [14].

To date, there is no report on the application of Next Generation Sequencing (NGS) techniques to study *Calycophyllum* spp. genomes, so in the present work, we present the first complete chloroplast genome sequence of *C. spruceanum*, based on the Illumina sequencing technology. Then, a comparative analysis of *C. spruceanum* with other six closely related species that belong to Ixoroideae subfamily is reported. Our study provided useful information on genome organization, gene content, and structure variation in the *C. spruceanum* cp genome, and also provided important clues to its phylogenetic relationships, which will contribute to genetic and evolutionary studies in *C. spruceanum* and beyond.

2. Results

2.1 C. spruceanum chloroplast Genome Assembly and Its Features

The overall length of the *C. sprucearum* chloroplast genome is 154,480 bp, exhibiting the circular quadripartite structure characteristic of major angiosperm plants. After annotation and modification, the entire chloroplast (cp) genome sequence was submitted to the GenBank database with accession number: OK326865.1. The chloroplast genome of capirona consists of a pair of the inverted repeat (IR) regions (25,783 bp) separated by a large single-copy (LSC) region of 84,813 bp and a small single-copy (SSC) region of 18,101 bp. A circular representation of the complete chloroplast genome is shown in Figure 1. The GC content of the IR region (43.14 %) was much higher than that of the LSC (31.89 %) and SSC regions (35.48%) in the *C. spruceanum* cp genome (Table 1). The annotation of cp genome predicted a total of 133 genes, of which 114 are unique consisting of 80 protein-coding genes, 30 transfer RNA (tRNA) genes, four ribosomal RNA (rRNA) genes, and one pseudogene (Table 2). Of these, seven protein-coding genes, four rRNAs and seven tRNAs are duplicated in the IR regions. A total of ten protein-coding genes and eight tRNAs genes contained a single intron, whereas three genes exhibited two introns each. The *rps12* gene was predicted to be trans-spliced with its 5′ end located at the LSC region and the 3′ end with a copy located in each of the two IR regions.

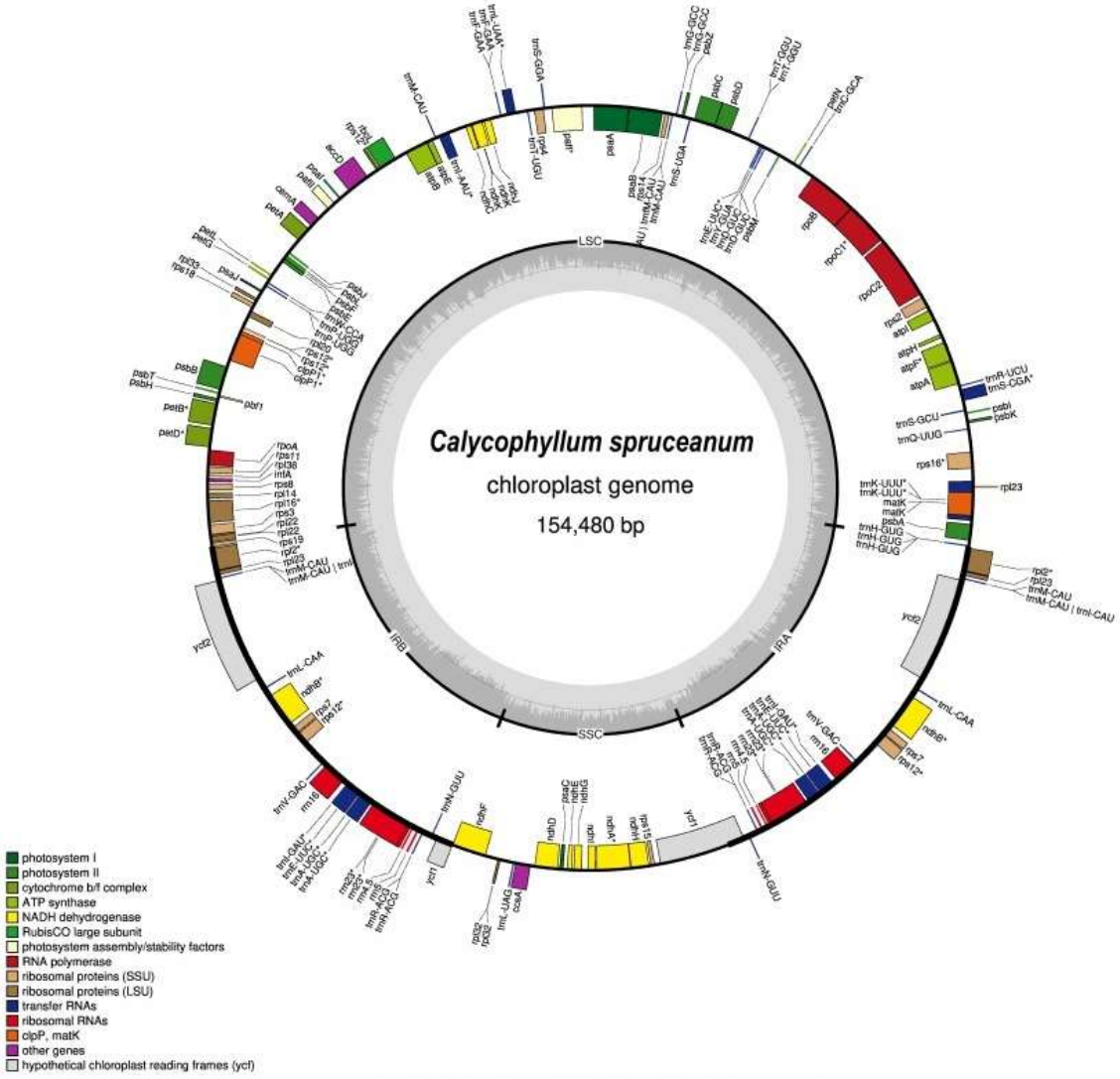

Figure 1. Gene map of *C. spruceanum*. Genes lying outside the outer circle are transcribed in a counter-clockwise direction, and genes inside this circle are transcribed in a clockwise direction. The colored bars indicate known protein-coding genes, transfer RNA genes, and ribosomal RNA genes. LSC, large single-copy; SSC, small single-copy; IR, inverted repeat.

As expected, the duplicated IR of the *C. sprucearum* chloroplast genome resulted in complete duplication of 18 genes: five protein-coding genes such as *rpl2*, *rpl23*, *rps7*, *rps12*, and *ndhB*; seven tRNAs as *trn*I-CAU, *trn*L-CAA, *trn*V-GAC, *trn*I-GAU, *trn*A-UGC, *trn*R-AGC, and *trn*N-GUU; four rRNAs genes as *rrn*23, *rrn*16, *rrn*5, *rrn*4.5 (see Figure 1), and 5′ end of *ycf1*. The SSC region contained 12 protein-coding and one tRNA gene, while LSC region contained 69 protein-coding and 22 tRNAs.

Codon usage analysis identified a total of 26,572 codons in the *C. sprucearum* chloroplast genome. Among all codons, leucine (Leu) was the most abundant amino acid with a frequency of 10.62%, followed by isoleucine (Ile) with a frequency of 8.40%, whereas cysteine (Cys) was less abundant with a frequency of 1.14%. Moreover, only one codon was identified for methionine (Met) and tryptophan (Trp) amino acids. Thirty codons were observed to be used more frequently than the expected usage at equilibrium (RSCU > 1) and 31 codons showed the codon usage bias: (RSCU < 1), and the third positions of the biased codons were A/U (Table 3). Biased codons with the highest values of RSCU were Leu (UUA), Ser (UCU), Gly (GGA), Tyr (UAU), and Asp (GAU).

Table 1. Features of the chloroplast genomes of *C. spruceanum* and six Ixoroideae species.

Genome features	*Calycophyllum spruceanum*	*Coffea arabica*	*C. canephora*	*Emmenopterys henryi*	*Fosbergia shweliensis*	*Gardenia jasminoides*	*Scyphiphora hydrophyllacea*
Genome size (bp)	154,480	155,189	154,751	155,379	154,717	154,921	155,132
SSC length (bp)	18,101	18,137	18,133	18,245	18,230	18,095	18,165
LSC length(bp)	84,813	85,166	84,850	85,554	84,747	85,236	85,239
IRA length (bp)	25,783	25,908	23,834	25,790	25,870	25,795	25,864
IRB length (bp)	25,783	25,943	23,884	25,790	25,870	25,795	25,864
No. of protein-coding genes	87	85	86	87	85	87	88
No. of different rRNA genes	4	4	4	4	4	4	4
No. of tRNA genes	37	38	37	37	36	37	37
%GC content in LSC	35.48	31.28	31.75	31.90	35.5	35.3	31.65
% GC content in SSC	31.89	35.35	35.48	35.48	31.4	31.5	35.49
% GC content in IR	43.14	43.01	43.55	43.26	43.2	43.2	43.17

2.2. Comparative analysis of genome structure

In order to know the structural characteristics of *C. sprucearum* chloroplast genome (154,480 bp total length), we compared it with other six Ixoroideae species as *Coffea canephora, C. arabica, F. shweliensis, S. hydrophyllacea, E. henryi* and *G. jasminoides*, whose chloroplast genome sizes were 154,751 bp; 155,189 bp; 154,717 bp; 155,132 bp; 155,379 bp; 154,921 bp, respectively (Table 1). Our results showed that gene coding regions were more conserved than the noncoding regions, and the SSC and LSC regions showed more divergence than IRa and IRb regions (Figure S1, Figure 2). Additionally, it was also observed that the intergenic spacers regions between several pairs of genes varied greatly, for example, between *psbA-trnH-GUG, rps16-matK, atpI-atpH, ndhJ-rps4, rbcL-psaI, psaI-petA, ycf11-rps15* and *rpl32-ndhF*. In the coding regions, slight variations in sequence were observed in *matK, rpoC2, rps19 and ycf1* (Figure 2). The identity matrix revealed that the values in the IR region varied between 0.91 to 0.99. The LSC region presented values that fluctuated between 0.90 to 0.97 and the SSC region presented the highest divergence values, ranging from 0.82 to 0.97 (Figure S1). Gene order between *C. spruceanum* and other six Ixoroideae species showed similar patterns, however, greater divergences were found between *C. spruceanum* and *C. canephora.*

Table 2. Genes found in the assembled capirona chloroplast genome.

Category	Function	Genes
RNA genes	Transfer RNA	*trnH-GUG*, *trnK-UUU*[b], *trnQ-UUG*, *trnS-GCU*, *trnG-UCC*[b], *trnR-UCU*, *trnC-GCA*, *trnD-GUC*, *trnY-GUA*, *trnE-UUC*, *trnT-GGU*, *trnS-UGA*, *trnG-GCC*, *trnfM-CAU*, *trnS-GGA*, *trnT-UGU*, *trnL-UAA*[b], *trnF-GAA*, *trnV-UAC*[b], *trnM-CAU*, *trnW-CCA*, *trnP-UGG*, *trnI-CAU* (x2), *trnL-CAA* (x2), *trnV-GAC* (x2), *trnI-GAU*[b] (x2), *trnA-UGC*[b] (x2), *trnR-ACG* (x2), *trnN-GUU* (x2), *trnL-UAG*,
	Ribosomal RNA	*rrn23* (x2), *rrn16* (x2), *rrn5* (x2), *rrn4.5*(x2)
Transcription and translation related genes	Transcription and splicing	*rpoA*, *rpoB*, *rpoC1*[b], *rpoC2*
	Ribosomal protein large subunit	*rpl2*[b] (x2), *rpl14*, *rpl16*[b], *rpl20*, *rpl22*, *rpl23* (x2), *rpl32*, *rpl33*, *rpl36*
	Ribosomal protein small subunit	*rps2*, *rps3*, *rps4*, *rps7* (x2), *rps8*, *rps12*[ac] (x2), *rps11*, *rps14*, *rps15*, *rps16*[b], *rps18*
Photosynthesis	ATP synthase	*atpA*, *atpB*, *atpE*, *atpF*[b], *atpH*, *atpI*
	Photosystem I	*psaA*, *psaB*, *psaC*, *psaI*, *psaJ*
	Photosystem II	*psbA*, *psbB*, *psbC*, *psbD*, *psbE*, *psbF*, *psbH*, *psbI*, *psbJ*, *psbK*, *psbL*, *psbM*, *psbN*, *psbT*, *psbZ*
	Cytochrome complex	*petA*, *petB*[b], *PetD*, *petG*, *petL*, *petN*
	Calvin cycle	*rbcL*
	NADH dehydrogenase	*ndhA*[b], *ndhB*[b] (x2), *ndhC*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*, *ndhJ*, *ndhK*
Miscellaneous group	Translation initiation factor IF1	*infA*
	Acetyl-CoA carboxylase	*accD*
	Maturase	*matK*
	ATP-dependent protease	*ClpP1*[a]
	Inner membrane protein	*cemA*
	Cytochrome c biogenesis	*ccsA*
Other genes	Conserved hypothetical chloroplast ORF	*ycf1*, pseudogene *ycf1*, *ycf2* (x2), *ycf3*[a], *ycf4*, *ycf15* (x2)

[a] Gene containing two introns;

[b] Gene containing a single intron;

[c] Gene divided into two independent transcription units

Table 3. The codon-anticondon recognition pattern and codon usage in the chloroplast genome of *Calycophyllum spruceanum*.

Amino acid	Codon	Count*	RSCU	tRNA	Amino acid	Codon	Count*	RSCU	tRNA
	UUU	961	1.28		Gly(G)	GGU	581	1.27	
Phe(F)	UUC	538	0.72	*trnF-GAA*	Gly(G)	GGC	184	0.4	*trnG-GCC*
Leu(L)	UUA	832	1.77	*trnL-UAA*	Gly(G)	GGA	738	1.62	*trnG-UCC*
Leu(L)	UUG	581	1.24	*trnL-CAA*	Gly(G)	GGG	320	0.7	
Leu(L)	CUU	626	1.33		Pro(P)	CCU	409	1.48	
Leu(L)	CUC	193	0.41		Pro(P)	CCC	224	0.81	
Leu(L)	CUA	416	0.88	*trnL-UAG*	Pro(P)	CCA	326	1.18	*trnP-UGG*
Leu(L)	CUG	174	0.37		Pro(P)	CCG	150	0.54	
Ile(I)	AUU	1111	1.49	*trnI-CAU*	Thr(T)	ACU	523	1.55	
Ile(I)	AUC	425	0.57	*trnI-GAU*	Thr(T)	ACC	255	0.75	*trnT-GGU*
Ile(I)	AUA	697	0.94		Thr(T)	ACA	417	1.23	*trnT-UGU*
Met(M)	AUG	628	1	*trnM-CAU, trnfM-CAU*	Thr(T)	ACG	157	0.46	
Val(V)	GUU	537	1.47		Ala(A)	GCU	632	1.82	
Val(V)	GUC	181	0.49	*trnV-GAC*	Ala(A)	GCC	234	0.67	
Val(V)	GUA	548	1.5	*trnV-UAC*	Ala(A)	GCA	388	1.12	*trnA-UGC*
Val(V)	GUG	198	0.54		Ala(A)	GCG	136	0.39	
Cys(C)	UGU	226	1.49		Tyr(Y)	UAU	787	1.61	
Cys(C)	UGC	77	0.51	*trnC-GCA*	Tyr(Y)	UAC	189	0.39	*trnY-GUA*
Stop	UGA	19	0.66		Stop	UAA	47	1.62	
Trp(W)	UGG	462	1	*trnW-CCA*	Stop	UAG	21	0.72	
Arg(R)	CGU	348	1.29	*trnR-ACG*	His(H)	CAU	476	1.5	
Arg(R)	CGC	103	0.38		His(H)	CAC	158	0.5	*trnH-GUG*
Arg(R)	CGA	380	1.41		Gln(Q)	CAA	729	1.51	*trnQ-UUG*
Arg(R)	CGG	132	0.49		Gln(Q)	CAG	235	0.49	
Arg(R)	AGA	487	1.81	*trnR-UCU*	Asn(N)	AAU	990	1.55	
Arg(R)	AGG	163	0.61		Asn(N)	AAC	291	0.45	*trnN-GUU*
Ser(S)	UCU	605	1.75		Lys(K)	AAA	1034	1.47	*trnK-UUU*
Ser(S)	UCC	331	0.96	*trnS-GGA*	Lys(K)	AAG	373	0.53	
Ser(S)	UCA	405	1.17	*trnS-UGA*	Asp(D)	GAU	887	1.64	
Ser(S)	UCG	202	0.59		Asp(D)	GAC	196	0.36	*trnD-GUC*
Ser(S)	AGU	400	1.16		Glu(E)	GAA	1038	1.51	*trnE-UUC*
	AGC	127	0.37	*trnS-GCU*	Glu(E)	GAG	334	0.49	

*Frequency of usage of each codon in 26,572 codons in 87 protein-coding genes RSCU, relative synonymous codon usage

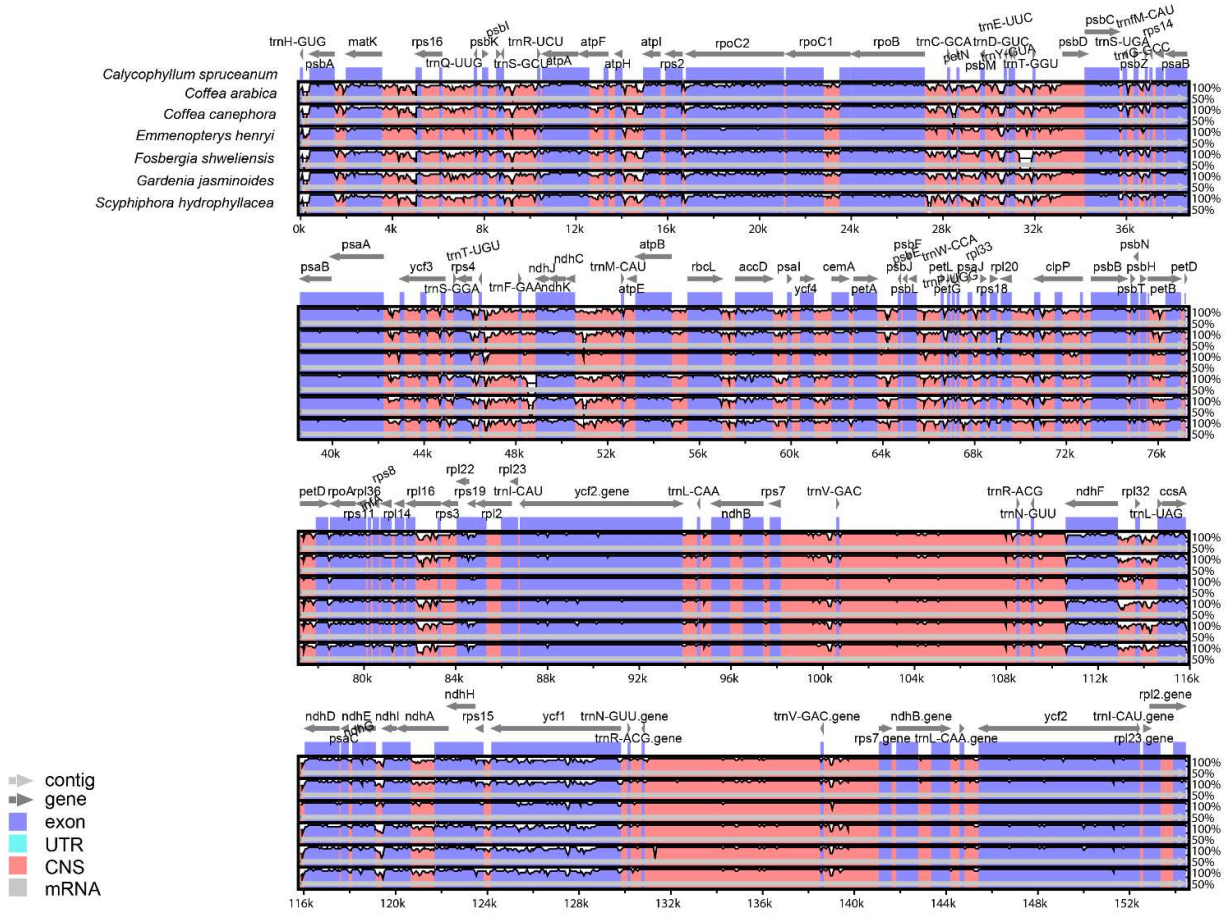

Figure 2. mVISTA identity plot comparing the seven Ixoroideae plastid genomes considering *C. spruceanum* as a reference. The top line shows genes in order (transcriptional direction indicated by arrows). The y-axis represents the percent identity within 50–100%. The x-axis represents the coordinate in the chloroplast genome. Genome regions are color-coded as protein-coding (exon), tRNAs or rRNAs, and conserved noncoding sequences (intergenic region). The white block represents regions with sequence variation between two species.

2.3. SSR Loci Identified in Ixoroideae cp Genomes

The analysis of SSRs distribution within *C. sprucearum* chloroplast genome revealed a total of 41 SSRs, being the most abundant the mononucleotide repeats (29) followed by dinucleotides (5). Additionally, SSRs with trinucleotides repeats (3) and tetranucleotides repeats (4) motifs in these genomes were identified in less quantity. The number of SSRs identified for *C. arabica*, *C. canephora*, *F. shweliensis*, *S. hydrophyllacea*, *E. henryi, and G. jasminoides* was variable (43, 38, 42, 52, 46 and 30, respectively). All these species presented the highest number of SSRs for A/T mononucleotides, and for AT/TA dinucleotides (Figure 3). Only *F. shweliensis* and *S. hydrophyllaceae* presented SSRs with pentanucleotide repeats, and even *S. hydrophyllaceae* has SSRs with hexanucleotide repeats. On the other hand, we detected that the SSRs were not only mainly found in the non-coding regions (*psbA-trnH-GUG, rps16- matK, atpI-atpH, ndhJ-rps4, rbcL- psaI)*, but also in coding regions, such as *rpoC2* and ycF2, *ndhF* ,*ndhG* and *matK,* .Also we detected in less quantity SSR located in tRNA sequences (Figure 4).

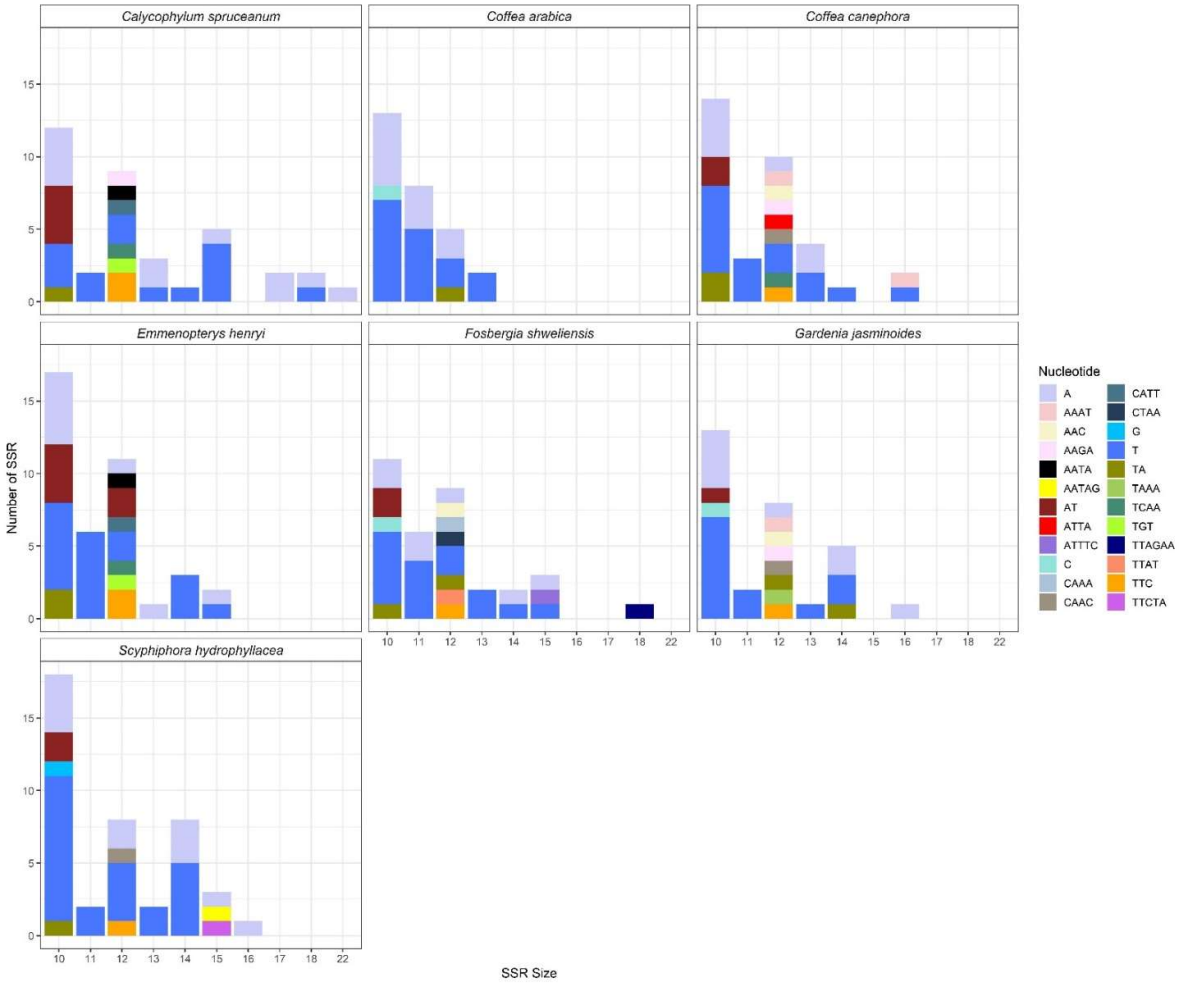

Figure 3. Analysis of simple sequence repeat (SSR) distribution in *C. spruceanum* and six other Ixoroideae species. The X-axis shows the size of SRR, the Y axis shows the number of SSR. The colored bars indicate the different motifs within SSRs.

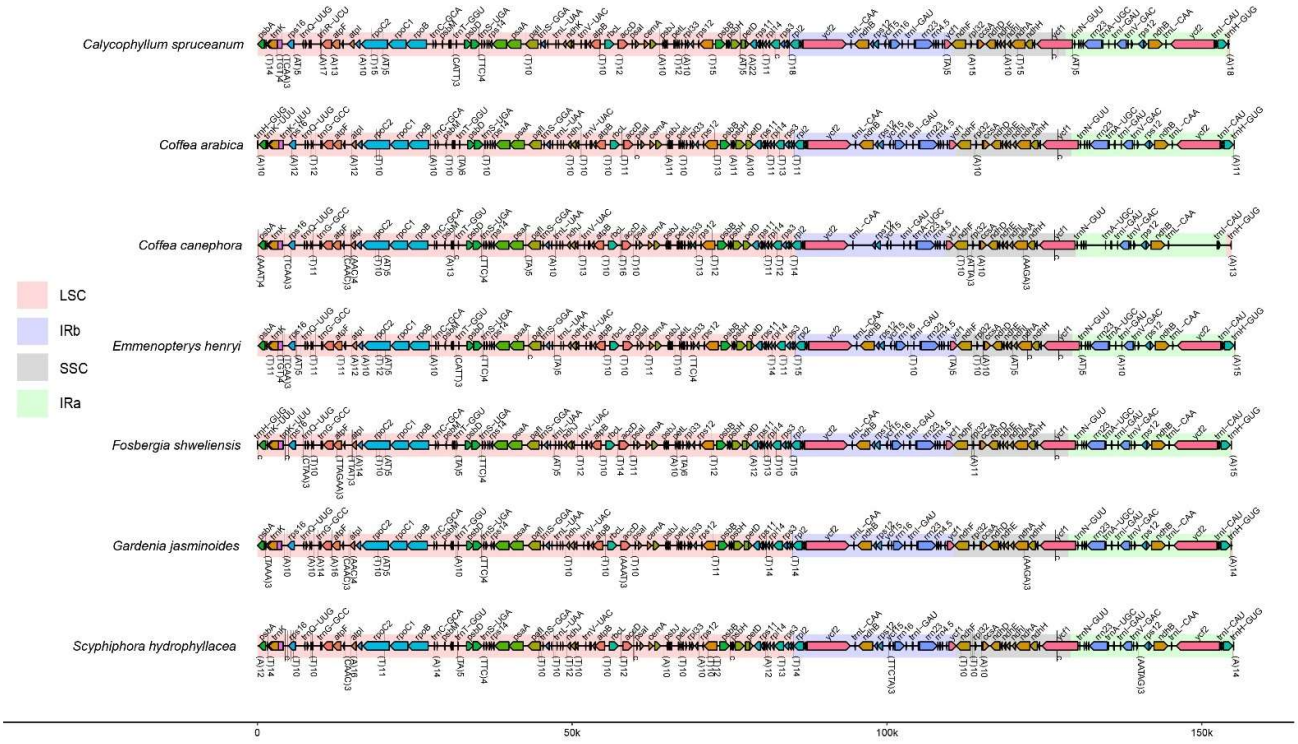

Figure 4. Comparison of the genome structure and location of the simple sequence repeat (SSR) of seven Ixoroideae cp genomes, with *C. spruceanum* as a reference.

2.4. Phylogenetic Inference of C. spruceanum

In this study, 19 species belonging to Rubiaceae and one outgroup (*Lonicera hispida*, Caprifoliaceae) were employed to infer their phylogenetic relationships using complete chloroplast genome sequences. The phylogenetic topology revealed well-supported monophylies for subfamilies Rubioideae, Cinchonoideae, and Ixoroideae. Maximum likelihood (ML) bootstrap support (BS) were very high, 16 nodes had 100% bootstrap values, and only one presented 85%. *C. spruceanum* was placed within subfamiliy Ixoroideae, and with 100% BS revealed to be a sister species of *Emmenopterys henryi* (Figure 5). This phylogenetic tree were consisted with traditional taxonomy of Rubiaceae family, suggesting that the chloroplast genome sequences can effectively resolve relationships of species.

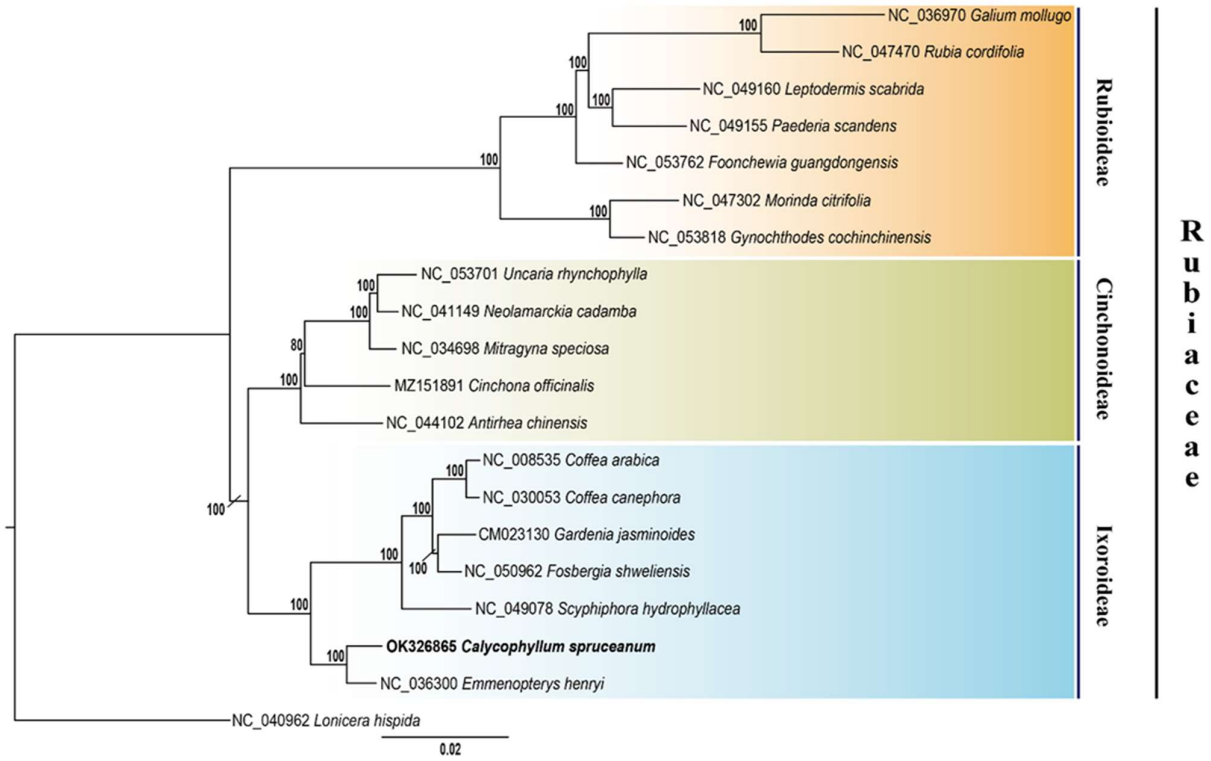

Figure 5. The maximum likelihood (ML) phylogenetic tree of the Rubiaceace family based on chloroplast genome sequences. Values along branches correspond to bootstrap percentages. The position of capirona (*C. spruceanum*) is indicated in black text. *Lonicera hispida* was set as the outgroup.

3. Discussion

Capirona (*Calycophyllum spruceanum*) is an economically important forest species belonging to the Ixoroideae subfamily, within the Rubiaceae family. Until very recently, only a few complete chloroplast genome sequences for the Ixoroideae subfamily have been deposited to GenBank, with the very first being that of *Coffea canephora* in 2016. Nevertheless, with the development of Next Generation Sequencing (NGS), the chloroplast (cp) genome of most species of Ixoroideae subfamily has been obtained [2,15–17], however, to date, cp genome of members of the genus *Calycophyllum* remained unknown. Thus, in the present study we sequenced for the first time the *C. spruceanum* chloroplast genome (accession number: OK326865.1) and compared with other members of the sub family Ixoroideae that are closely related. The *C. spruceanum* cp genome is in agreement with the characteristics of most angiosperm species in structure and gene content. In fact, the complete cp genome of *C. spruceanum* was 154,480pb, similar to other Ixoroideae genomes [15,17], with a quadripartite structure (LSC, SSC and two IR regions) which is a common characteristics in higher plants [11]. The annotation of *C. spruceanum* cp genome predicted 87 protein-coding genes (CDS), and similar patterns of protein-coding genes are also present in other Rubiaceae plants [16]. Similar to other studies [18,19], there were three genes (*rps12*, *clpP1* and *ycf3*) that included two intron regions in the cp genome of capirona. It has been demonstrated gene *clpP1* (caseinolytic protease P1) is essential for plant development [20]. Moreover, Boudreau et al. [21]demonstrated that gene *ycf3* are required for the accumulation of the photosystem I (PSI) complex.

The GC content in the IR region was much higher than in the LSC and SSC region in *C. spruceanum* cp genome, probably due to the presence of eight ribosomal RNA (rRNA) genes in this region, which is consistent with previous analyses in other Ixoroideae [15,16], and in other angiosperms cp genomes [14,22,23]. The IR (A/B) region has

always been considered consistent and stable in the cp genome, and it is also common in the evolution of plants with contraction or expansion events in the border region [16]. Also, these results suggest that the cp genome in this subfamily had rather conserved genome organization [15,16]**.** We identified that in the seven sequences of the cp genome are some highly divergent regions including *psbA-trnH-GUG*, *rps16- matK*, *atpI-atpH*, *ndhJ-rps4*, *rbcL- psaI*, *psaI-petA*, *ycf1- rps15*, and *rpl32-ndhF*. These variable regions could be used for the development of molecular markers for DNA barcoding and phylogenetic studies in species of the Ixoroideae subfamily. Interestingly, *C. canephora* presents higher divergence values when compared with the other six species (Figure S1). In addition, *ycf1* gene presented the greatest differentiation, suggesting that it is useful for providing phylogenetic resolution at the species level, as demonstrated for genus *Pinus* and *Daucus* [25].

We identified simple sequence repeats (SSRs), also kwon as microsatellites in *C. spruceanum*. They are powerful molecular marker, and play an important role in genetic diversity, population structure, evolutionary studies chloroplast genome rearrangement and recombination process [26–28] due its abundant polymorphism, high stability, codomimant inheritance and ease of use [29]. In addition, SSR have been widely applied as molecular markers because of their unique uniparental inheritance [30,31]. In total, 41 perfect SSRs were detected in *C. sprucearum* cp genome distributed in the LSC, SSC, and IR regions with strong A/T bias. Similarly, previous studies also revealed that the non-coding region contained more SSRs than the coding regions [14,16]. Also our results are comparable to those of several previous studies showing that SSRs in cp genomes are highly rich in polythymine (poly T) or polyadenine (polyA) as reported for cp genomes of other plants species [32–34]. On the other hand, repeats containing tandem cytosine (C) and guanine (G) were limited. Our results are in agreement with other studies that report microsatellites markers for other Ixoroideae species such as *C. arabica*, *C. canephora*, and *E. henryi* [16]. However, we disagree with the results obtained by Wang et al [15] for *G. jasminoides*. With the identification of the SSR in the cp genome of *C. spruceanum*, we will be able to evaluate the polymorphism at the intraspecific level, as well as to evaluate the genetic diversity between and within the populations of *C. spruceanum*. These markers could also be used to aid in the selection and characterization of genotypes plus they are suitable for the development of a modern genetic improvement and conservation program.

Codon usage bias is a known phenomenon that occurs in a wide variety of organisms, and apparently, the major cause for selection on codon bias is that some preferred codons are translated more efficiently [35]. As reported for other chloroplast genomes of plants [36], our study revealed the preference in the use of synonymous codons, and the RSCU values of 30 codons resulted in > 1 with biased codons in the third positions for A/T, which may be originated by a composition bias for a high A/T ratio [34]. These results are in accordance with other studies, where the codon usage preference for A/T is found in most other land plant chloroplast genome [37]. Gene expression and molecular evolution system of *C. spruceanum* may be elucidated by conducting research on its codon usage.

The rapid progress in the field of chloroplast genetics and genomics has been facilitated by the advent of high-throughput sequencing technologies. Chloroplast genomes have many features like small size (120 – 160 Kbp), high copy number, generally conservative nature [38] that make them useful for phylogenetic studies, resolving evolutionary relationships within phylogenetic clades especially at low taxonomic levels [25,39,40]. Our entire plastid analysis of Rubiaceae provided a highly supported topology of the family, as reported by Bremer and Eriksson (2009) [41] using five chloroplast regions by Bayesian analysis. Besides, similar to their work, it was possible to obtain very high bootstrap support (BS) for the three subfamilies (Cinchonoideae, Rubioideae, Ixoroideae) clades. The availability of the completed *C. spruceanum* chloroplast genome

allowed us to confirm the phylogenetic position of this forest tree species and understand the phylogeny among Rubiaceae. With 100% BS, *C. sprucearum* was placed as sister species to *Emmenopterys henryi* within Ixoroideae subfamily, confirming its classification within Condomineae Tribe, as suggested by previous studies based on a reduced number of genes and morphological data [1,42]. However, employing additional members of the subfamily Ixoroideae as well as nuclear genome sequences would provide more evidence to accurately infer the evolution history of *Callycophyllum*.

4. Materials and Methods

4.1. Plant Materials and Genomic DNA Extraction

A single capirona tree was selected to be sequenced from San Bernardo Research Station of INIA, located in Madre de Dios department (2°41′8.66" N / 69° 22′49.8" E / 227.2 m.a.s.l) in the Peruvian Amazon. A branch with flowers was collected and deposited at the Scientific Collection of the Herbarium of Universidad Nacional Mayor de San Marcos (UNMSM), under the voucher number No 324323. Total genomic DNA was extracted from fresh leaves by CTAB method with minor modifications [43], the quality was evaluated on a 1% agarose gel and the quantification was performed by fluorescence.

4.2. DNA Sequence and Genome Assembly

High-quality genomic DNA was used to construct libraries. Pair-end clean reads were obtained by PE 150 library and the Illumina HiSeq 2500 platform. Adapters and low-quality reads were removed using Trim Galore [44]. We used clean data to assemble the chloroplast genome using the GetOrganelle v1.7.2 pipeline[45], in which SPAdes v3.11.1 [46], bowtie2 v2.4.2 [47] and BLAST+ v2.11[48] were employed.

4.3. Annotation and Analysis of C. spruceanum chloroplast DNA Sequence

The annotations of the protein-coding genes (PCGs), transfer RNAs (tRNAs) and rRNA genes from *C. spruceanum* chloroplast genome were performed using webserver Geseq [49] by comparing to all available plastid genomes in NCBI of Ixoroideae associated to this server and curated manually. The codon usage analysis was carried out with MEGA X software [50]. The architecture of *C. sprucearum* chloroplast genome was visualized using OGDRAW 1.3.1[51].

4.4. Comparative Analysis of Genome Structure

The Shuffle-LAGAN mode of the mVISTA online program (http://genome.lbl.gov/vista/ mvista/) [52] was used to compare the sequence similarity of the complete chloroplast genome of *Callycophyllum spruceanum* with other six species of Ixoroideae sub family (Table 1). The annotated *C. sprucearum* chloroplast genome generated in this work was used as reference.

4.5. Chloroplast Genome Analysis

SSRs within the *C. spruceanum* chloroplast genome were searched using the MISA software [53]. The criteria of SSR research were set as follows: The minimum numbers of repeats for mononucleotide, dinucleotides, trinucleotides, tetranucleotides, pentanucleotides and hexanucleotides were 10, 5, 4, 3, 3, 3, respectively [16]. A plot with the structure and location of the SSRs in the seven cp genomes analyzed in this study was generated using the *genoPlotR* and *gggenomes* packages in the R software v4.0.2. The codon usage, frequency and relative synonymous codon usage (RCSU) of the *C. spruceanum* cp genome was analyzed using MEGA X software [36]. The parameters used were by default.

4.6. Phylogenetic Analyses

To gain an insight into the phylogenetic location of *C. sprucearum*, a maximum-likelihood (ML) tree was constructed with 1,000 nonparametric bootstrap replicates using

RAxML v8.2.11 software [54] under GTR+GAMMA nucleotide substitution model of evolution. The complete chloroplast genome of *C. sprucearum* was compared and aligned with other 19 chloroplast genomes obtained from Genbank by the MAFFT v7.475 software [55]. Seven species from Rubioideae, five species from Cinchonoideae, and six species from Ixoroideae were included in the analysis. *Lonicera hispida* (Caprifoliaceae) was considered as outgroup.

5. Conclusions

Through this study we seek to contribute to understanding of chloroplast structure and the determination of phylogenetic relationships. Here, we first reported the complete chloroplast genome sequence of a forests tree species *C. spruceanum* and a comparative analysis of six Ixoroideae cp genomes to reveal the genome features. We identified 41 SSRs that can be used for population genetics and evolutionary studies. The genome structure, genes order and content were found to be much conserved with all species except with *Coffea canephora*. Both the LSC and SSC regions were more divergent than IR region in the chloroplast genome of *C. spruceanum*, with the two most variable regions (*PsbA-rps16*) found in the IR regions. Furthermore, the phylogenomic analysis based on whole cp genomes generated a ML tree with the same topologies as previously reported by other researchers, consolidating the taxonomical position of *C. spruceanum* species within the Ixoroideae subfamily and Condomineae Tribe. These results provided important information on the genome organization, gene content, and structural variation of capirona and other Ixoroideae cp genomes. On the same way, our work provided important clues and tools for phylogenetic analysis in the Ixoroideae subfamily, which could be useful for evolutionary and population studies.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Figure S1: Matrix of identities of the 4 regions (LSC, SSC, IRa, IRb) between each of seven chloroplast genomes.

Author Contributions: Conceptualization, Carla Saldaña and Carlos I. Arbizu; Formal analysis, Carla Saldaña, Julio Chávez-Galarza and Carlos I. Arbizu; Funding acquisition, Juan Carlos Guerrero-Abad, Héctor Vásquez, Jorge L. Maicelo and Carlos I. Arbizu; Methodology, Carla Saldaña, Pedro Rodriguez-Grados, Julio Chávez-Galarza and Shefferson Feijoo; Project administration, Jorge L. Maicelo and Jorge H. Jhoncon; Resources, Shefferson Feijoo, Héctor Vásquez, Jorge L. Maicelo and Jorge H. Jhoncon; Supervision, Héctor Vásquez and Carlos I. Arbizu; Validation, Pedro Rodriguez-Grados and Juan Carlos Guerrero-Abad; Visualization, Shefferson Feijoo, Juan Carlos Guerrero-Abad, Jorge L. Maicelo and Jorge H. Jhoncon; Writing – original draft, Carla Saldaña, Julio Chávez-Galarza and Carlos I. Arbizu; Writing – review & editing, Carla Saldaña and Carlos I. Arbizu.

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 Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government, with grant number CUI 2449640. C.L.S. was supported by PP0068 "Reducción de la vulnerabilidad y atención de emergencias por desastres".

Acknowledgments: We would like to thank Ivan Ucharima, Cristina Aybar and Erick Rodriguez for supporting the logistic activities in the laboratory.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kainulainen, K.; Razafimandimbison, S.G.; Bremer, B. Phylogenetic relationships and new tribal delimitations in subfamily Ixoroideae (Rubiaceae). *Bot. J. Linn. Soc.* **2013**, *173*, 387–406.

2. Ly, S.N.; Garavito, A.; De Block, P.; Asselman, P.; Guyeux, C.; Charr, J.C.; Janssens, S.; Mouly, A.; Hamon, P.; Guyot, R. Chloroplast genomes of Rubiaceae: Comparative genomics and molecular phylogeny in subfamily Ixoroideae. *PLoS One* **2020**, *15*, 1–21.
3. Russell, J.R.; Weber, J.C.; Booth, A.; Powell, W.; Sotelo-Montes, C.; Dawson, I.K. Genetic variation of *Calycophyllum spruceanum* in the Peruvian Amazon Basin, revealed by amplified fragment length polymorphism (AFLP) analysis. *Mol. Ecol.* **1999**, *8*, 199–204.
4. Sears, R.R. New Forestry on the Floodplain: The Ecology and Management of *Calycophyllum spruceanum* (Rubiaceae) on the Amazon Landscape. *Dr. Diss. Columbia Univ. EEUU* **2003**, 274.
5. Tauchen, J.; Lojka, B.; Hlasna-Cepkova, P.; Svobodova, E.; Dvorakova, Z.; Rollo, A. Morphological and genetic diversity of *Calycophyllum spruceanum* (Benth) K. Schum (Rubiaceae) in Peruvian Amazon. *Agric. Trop. Subtrop.* **2011**, *44*, 212-218.
6. Saldaña, C.L.; Cancan, J.D.; Cruz, W.; Correa, M.Y.; Ramos, M.; Cuellar, E.; Arbizu, C.I. Genetic diversity and population structure of capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon revealed by RAPD markers. *Forests* **2021**, *12*, 1–12.
7. Dávila-Lara, A.; Affenzeller, M.; Tribsch, A.; Díaz, V.; Comes, H.P. AFLP diversity and spatial structure of *Calycophyllum candidissimum* (Rubiaceae), a dominant tree species of Nicaragua's critically endangered seasonally dry forest. *Heredity*. **2017**, *119*, 275–286.
8. Gray, M.W. The evolutionary origins of organelles. *Trends Genet.* **1989**, *5*, 294–299.
9. Howe, C.J.; Barbrook, A.C.; Koumandou, V.L.; Nisbet, R.E.R.; Symington, H.A.; Wightman, T.F.; Fray, R.; Leaver, C.J.; Walker, J.E.; Gray, J.C.; et al. Evolution of the chloroplast genome. *Philos. Trans. R. Soc. B Biol. Sci.* **2003**, *358*, 99–107.
10. Fu, P.C.; Zhang, Y.Z.; Geng, H.M.; Chen, S.L. The complete chloroplast genome sequence of *Gentiana lawrencei* var. *farreri* (Gentianaceae) and comparative analysis with its congeneric species. *PeerJ* **2016**, *2016*, 1–15.
11. Wicke, S.; Schneeweiss, G.M.; dePamphilis, C.W.; Müller, K.F.; Quandt, D. The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Mol. Biol.* **2011**, *76*, 273–297.
12. Daniell, H.; Lin, C.S.; Yu, M.; Chang, W.J. Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biol.* **2016**, *17*, 1–29.
13. Jansen, R.K.; Raubeson, L.A.; Boore, J.L.; DePamphilis, C.W.; Chumley, T.W.; Haberle, R.C.; Wyman, S.K.; Alverson, A.J.; Peery, R.; Herman, S.J.; et al. Methods for obtaining and analyzing whole chloroplast genome sequences. *Methods Enzymol.* **2005**, *395*, 348–384.
14. Raman, G.; Park, S.J. The complete chloroplast genome sequence of the *Speirantha gardenii*: Comparative and adaptive evolutionary analysis. *Agronomy* **2020**, *10*.
15. Wang, W.; Shao, F.; Deng, X.; Liu, Y.; Chen, S.; Li, Y.; Guo, W.; Jiang, Q.; Liang, H.; Zhang, X. Genome surveying reveals the complete chloroplast genome and nuclear genomic features of the crocin-producing plant *Gardenia jasminoides* Ellis. *Genet. Resour. Crop Evol.* **2021**, *68*, 1165–1180.
16. Zhang, Y.; Zhang, J.W.; Yang, Y.; Li, X.N. Structural and comparative analysis of the complete chloroplast genome of a mangrove plant: *Scyphiphora hydrophyllacea* Gaertn. f. and related Rubiaceae species. *Forests* **2019**, *10*.
17. Geng, Y.; Li, Y.; Yuan, X.; Luo, T.; Wang, Y. The complete chloroplast genome sequence of *Fosbergia shweliensis*, an endemic species to Yunnan of China. *Mitochondrial DNA Part B Resour.* **2020**, *5*, 1796–1797.
18. Arbizu, C.I.; Ferro-Mauricio, R.D.; Chávez-Galarza, J.C.; Guerrero-Abad, J.C.; Vásquez, H. V.; Maicelo, J.L. The complete chloroplast genome of the national tree of Peru, quina (*Cinchona officinalis* L., Rubiaceae). *Mitochondrial*

DNA Part B Resour. **2021**, *6*, 2781–2783.

19. Ren, W.; Guo, D.; Xing, G.; Yang, C.; Zhang, Y.; Yang, J.; Niu, L.; Zhong, X.; Zhao, Q.; Cui, Y.; et al. Complete chloroplast genome sequence and comparative and phylogenetic analyses of the cultivated *Cyperus esculentus*. *Diversity* **2021**.
20. Kuroda, H.M.P. The plastid clpP1 protease gene is essential for plant development. *Nature.* **2003**, *425*, 30–33.
21. Boudreau, E.; Takahashi, Y.; Lemieux, C.; Turmel, M.; Rochaix, J. The chloroplast ycf3 and ycf4 open reading frames. *EMBO J.* **1997**, *16*, 6095–6104.
22. Raman, G.; Park, V.; Kwak, M.; Lee, B.; Park, S.J. Characterization of the complete chloroplast genome of A*rabis stellari* and comparisons with related species. *PLoS One* **2017**, *12*, 1–18.
23. Yang, J.B.; Yang, S.X.; Li, H.T.; Yang, J.; Li, D.Z. Comparative chloroplast genomes of *Camellia* species. *PLoS One* **2013**, *8*, 1–12.
24. Olsson, S.; Grivet, D.; Cid Vian, J. Species-diagnostic markers in the genus *Pinus*: Evaluation of the chloroplast regions *matk* and *ycf1*. *For. Syst.* **2018**, *27*.
25. Spooner, D.M.; Ruess, H.; Iorizzo, M.; Senalik, D.; Simon, P. Entire plastid phylogeny of the carrot genus (*Daucus*, Apiaceae): Concordance with nuclear data and mitochondrial and nuclear DNA insertions to the plastid. *Am. J. Bot.* **2017**, *104*, 296–312.
26. Provan, J.; Powell, W.; Hollingsworth, P.M. Chloroplast microsatellites: New tools for studies in plant ecology and evolution. *Trends Ecol. Evol.* **2001**, *16*, 142–147.
27. Dong, W.; Liu, H.; Xu, C.; Zuo, Y.; Chen, Z.; Zhou, S. A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: A case study on ginsengs. *BMC Genet.* **2014**, *15*, 1–8, doi:10.1186/s12863-014-0138-z.
28. Nybom, H.; Weising, K.; Rotter, B. DNA fingerprinting in botany: Past, present, future. *Investig. Genet.* **2014**, *5*, 1–35.
29. Khayi, S.; Gaboun, F.; Pirro, S.; Tatusova, T.; El Mousadik, A.; Ghazal, H.; Mentag, R. Complete chloroplast genome of *Argania spinosa*: Structural organization and phylogenetic relationships in Sapotaceae. *Plants* **2020**, *9*, 1–15.
30. Varshney, R.K.; Sigmund, R.; Börner, A.; Korzun, V.; Stein, N.; Sorrells, M.E.; Langridge, P.; Graner, A. Interspecific transferability and comparative mapping of barley EST-SSR markers in wheat, rye and rice. *Plant Sci.* **2005**, *168*, 195–202.
31. Li, B.; Lin, F.; Huang, P.; Guo, W.; Zheng, Y. Development of nuclear SSR and chloroplast genome markers in diverse *Liriodendron chinense* germplasm based on low-coverage whole genome sequencing. *Biol. Res.* **2020**, *53*, 1–12.
32. Liu, H.Y.; Yu, Y.; Deng, Y.Q.; Li, J.; Huang, Z.X.; Zhou, S.D. The chloroplast genome of *Lilium henrici*: Genome structure and comparative analysis. *Molecules* **2018**, *23*, 1–13.
33. Biju, V.C.; P.R., S.; Vijayan, S.; Rajan, V.S.; Sasi, A.; Janardhanan, A.; Nair, A.S. The Complete Chloroplast Genome of Trichopus zeylanicus , And Phylogenetic Analysis with Dioscoreales . *Plant Genome* **2019**, *12*, 190032.
34. Kuang, D.Y.; Wu, H.; Wang, Y.L.; Gao, L.M.; Zhang, S.Z.; Lu, L. Complete chloroplast genome sequence of *Magnolia kwangsiensis* (Magnoliaceae): Implication for DNA barcoding and population genetics. *Genome* **2011**,
35. Hershberg, R.; Petrov, D.A. Selection on codon bias. *Annu. Rev. Genet.* **2008**, *42*, 287–299.
36. Dong, F.; Lin, Z.; Lin, J.; Ming, R.; Zhang, W. Chloroplast genome of rambutan and comparative analyses in Sapindaceae. *Plants* **2021**, *10*, 1–15.
37. Yu, X.; Zuo, L.; Lu, D.; Lu, B.; Yang, M.; Wang, J. Comparative analysis of chloroplast genomes of five *Robinia* species: Genome comparative and evolution analysis. *Gene* **2019**, *689*, 141–151.

38. Wolfe, K.H.; Li, W.H.; Sharp, P.M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 9054–9058.
39. Spalik, K.; Downie, S.R. Intercontinental disjunctions in Cryptotaenia (Apiaceae, Oenantheae): An appraisal using molecular data. *J. Biogeogr.* **2007**, *34*, 2039–2054.
40. Du, Y.P.; Bi, Y.; Yang, F.P.; Zhang, M.F.; Chen, X.Q.; Xue, J.; Zhang, X.H. Complete chloroplast genome sequences of *Lilium*: Insights into evolutionary dynamics and phylogenetic analyses. *Sci. Rep.* **2017**, *7*, 1–10,
41. Bremer, B.; Eriksson, T. Time tree of Rubiaceae: Phylogeny and dating the family, subfamilies, and tribes. *Int. J. Plant Sci.* **2009**, *170*, 766–793.
42. Bremer, B.; Jansen, R.K.; Oxelman, B.; Backlund, M.; Lantz, H.; Kim, K-J. More characters or more taxa for a robust phylogeny--Case Study from the Coffee Family. *Syst. Biol.* **1999**, *48*, 413-435.
43. Doyle, J.J.; Doyle, J.L. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem*. Bulletin **1987**, *19*, 11–15.
44. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2013**, *17*, 10–12.
45. Jin, J.J.; Yu, W. Bin; Yang, J.B.; Song, Y.; dePamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* **2018**, *21*, 241.
46. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477.
47. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359,
48. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinformatics* **2009**, *10*, 1–9.
49. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq - Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11.
50. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
51. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64.
52. Frazer, K.A.; Pachter, L.; Poliakov, A.; Rubin, E.M.; Dubchak, I. VISTA: Computational tools for comparative genomics. *Nucleic Acids Res.* **2004**, *32*, 273–279.
53. Beier, S.; Thiel T.; Münch T.; Scholz, U, and Mascher, M. MISA-web: a web server for microsatellite prediction. *Bioinformatics* **2017**, *33*, 2583–2585.
54. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313.
55. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780.