



PERÚ

Ministerio  
de Desarrollo Agrario  
y Riego



Instituto Nacional de Innovación Agraria



# Aplicación y desarrollo de herramientas moleculares modernas para el estudio genético especies forestales nativas de la Amazonia peruana

Subdirección de Investigación y Estudios Especiales

Dirección de Desarrollo Tecnológico Agrario

junio, 2022

Neale et al. *Genome Biology* 2013, 14:120  
<http://genomebiology.com/2013/14/6/120>



## OPINION

# Open access to tree genomes: the path to a better forest

David B Neale<sup>1\*</sup>, Charles H Langley<sup>2</sup>, Steven L Salzberg<sup>3</sup> and Jill L Wegrzyn<sup>1</sup>

### Abstract

An open-access culture and a well-developed comparative-genomics infrastructure must be developed in forest trees to derive the full potential of genome sequencing in this diverse group of plants that are the dominant species in much of the earth's terrestrial ecosystems.

**Keywords:** Forest tree genome, Open access, Sequencing, Genomics, Database

### The great diversity found in forest trees

There are an estimated 60,000 tree species on earth, and approximately 30 of the 49 plant orders contain tree species. Clearly, the tree phenotype has evolved many times in plants. The diversity of plant structures, development, life history, environments occupied and so on in trees is nearly as broad as higher plants in general, but trees share the common characteristic that all are perennial and many are very long lived. Because of the sessile nature of plants, each tree must survive and reproduce in a specific environment over the seasonal cycles of its lifetime. This tight association between

# Avances de grupos internacionales



GigaScience, 6, 2017, 1–4

doi: 10.1093/gigascience/giw016  
Advance Access Publication Date: 15 February 2017  
Data Note

## DATA NOTE

### An improved assembly of the loblolly pine mega-genome using long-read single-molecule sequencing

Aleksey V. Zimin<sup>1,2</sup>, Kristian A. Stevens<sup>3</sup>, Marc W. Crepeau<sup>3</sup>, Daniela Puiu<sup>2</sup>, Jill L. Wegrzyn<sup>4</sup>, James A. Yorke<sup>1</sup>, Charles H. Langley<sup>3</sup>, David B. Neale<sup>5</sup> and Steven L. Salzberg<sup>2,6,\*</sup>

<sup>1</sup>Institute for Physical Sciences and Technology, University of Maryland, College Park, MD, <sup>2</sup>Center for Computational Biology, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins School of Medicine, Baltimore, MD, <sup>3</sup>Department of Evolution and Ecology, University of California, Davis, CA, <sup>4</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, <sup>5</sup>Department of Plant Sciences, University of California, Davis, CA and <sup>6</sup>Departments of Biomedical Engineering, Computer Science, and Biostatistics, Johns Hopkins University, Baltimore, MD

\*Correspondence: [salzberg@jhu.edu](mailto:salzberg@jhu.edu)

Received: 26 November 2019 | Revised: 9 December 2019 | Accepted: 9 December 2019

DOI: 10.1111/eva.12902

SPECIAL ISSUE PERSPECTIVE

Evolutionary Applications WILEY

## Forest genomics: Advancing climate adaptation, forest health, productivity, and conservation

Nathalie Isabel<sup>1,2</sup> | Jason A. Holliday<sup>3</sup> | Sally N. Aitken<sup>4</sup>

<sup>1</sup>Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, Québec, Canada

<sup>2</sup>Canada Research Chair in Forest Genomics, Centre for Forest Research and Institute for Systems and Integrative Biology, Université Laval, Québec, Canada

<sup>3</sup>Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA, USA

<sup>4</sup>Centre for Forest Conservation Genetics and Department of Forest and Conservation Sciences, University of British Columbia, Vancouver, Canada

### Correspondence

Nathalie Isabel, Laurentian Forestry Centre, Canadian Forest Service, Natural Resources Canada, Québec, Canada.  
Email: [Nathalie.Isabel@canada.ca](mailto:Nathalie.Isabel@canada.ca)

### Abstract

Forest ecosystems provide important ecological services and resources, from habitat for biodiversity to the production of environmentally friendly products, and play a key role in the global carbon cycle. Humanity is counting on forests to sequester and store a substantial portion of the anthropogenic carbon dioxide produced globally. However, the unprecedented rate of climate change, deforestation, and accidental importation of invasive insects and diseases are threatening the health and productivity of forests, and their capacity to provide these services. Knowledge of genetic diversity, local adaptation, and genetic control of key traits is required to predict the adaptive capacity of tree populations, inform forest management and conservation decisions, and improve breeding for productive trees that will withstand the challenges of the 21st century. Genomic approaches have well accelerated the generation of knowledge of the genetic and evolutionary underpinnings of nonmodel tree species, and advanced their applications to address these challenges. This special issue of Evolutionary Applications features 14 papers that demonstrate the value of a wide range of genomic approaches that can be used to better understand the biology of forest trees, including species that are widespread and managed for timber production, and others that are threatened or endangered, or serve important ecological roles. We highlight some of the major advances, ranging from understanding the evolution of genomes since the period when gymnosperms separated from angiosperms 300 million years ago to using genomic selection to accelerate breeding for tree health and productivity. We also discuss some of the challenges and future directions for applying genomic tools to address long-standing questions about forest trees.

# Esquema de un proyecto de secuenciamiento

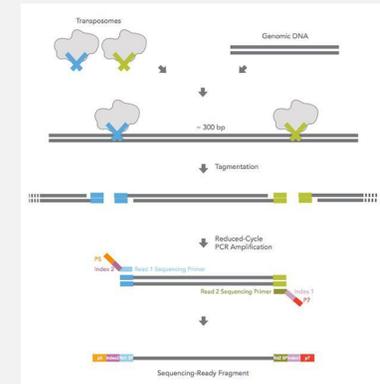
Obtención de la muestra



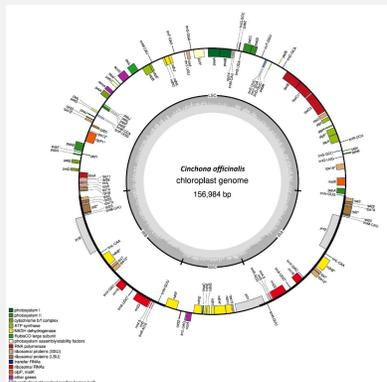
Purificación el ADN genómico



Fragmentación del ADN genómico



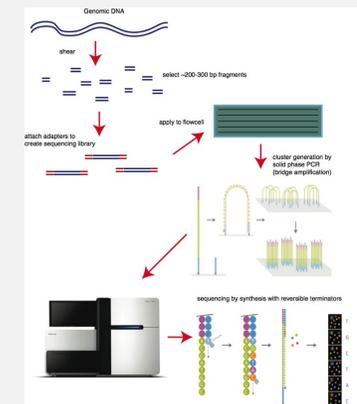
Anotación del genoma

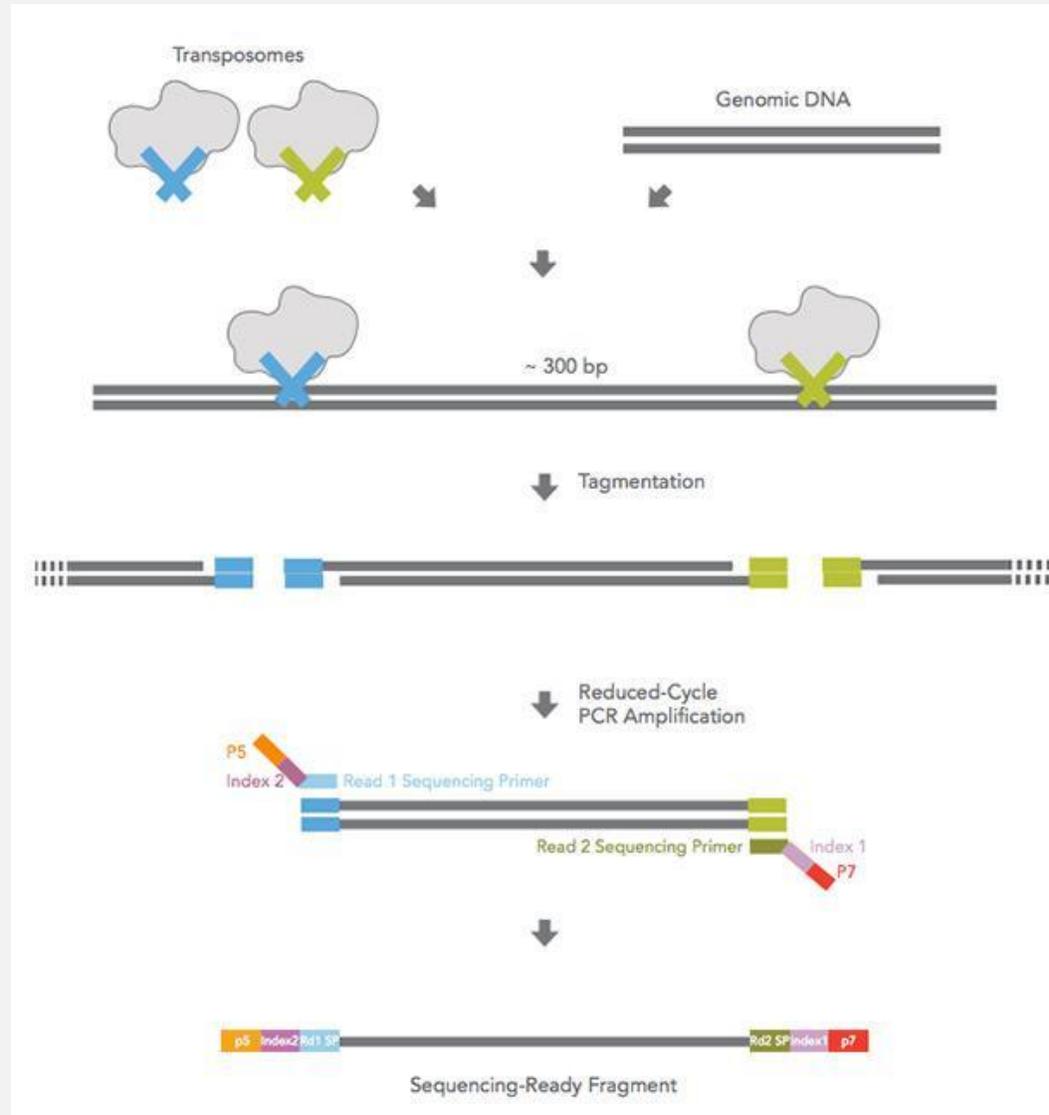


Ensamblaje de las secuencias obtenidas



Secuenciamiento de los fragmentos en la biblioteca







PERÚ

Ministerio de Desarrollo Agrario y Riego



Instituto Nacional de Innovación Agraria

**Siempre con el pueblo**



**Lugar de secuenciamiento:** EEUU

**Plataforma:** Illumina, 1 x 150 bp

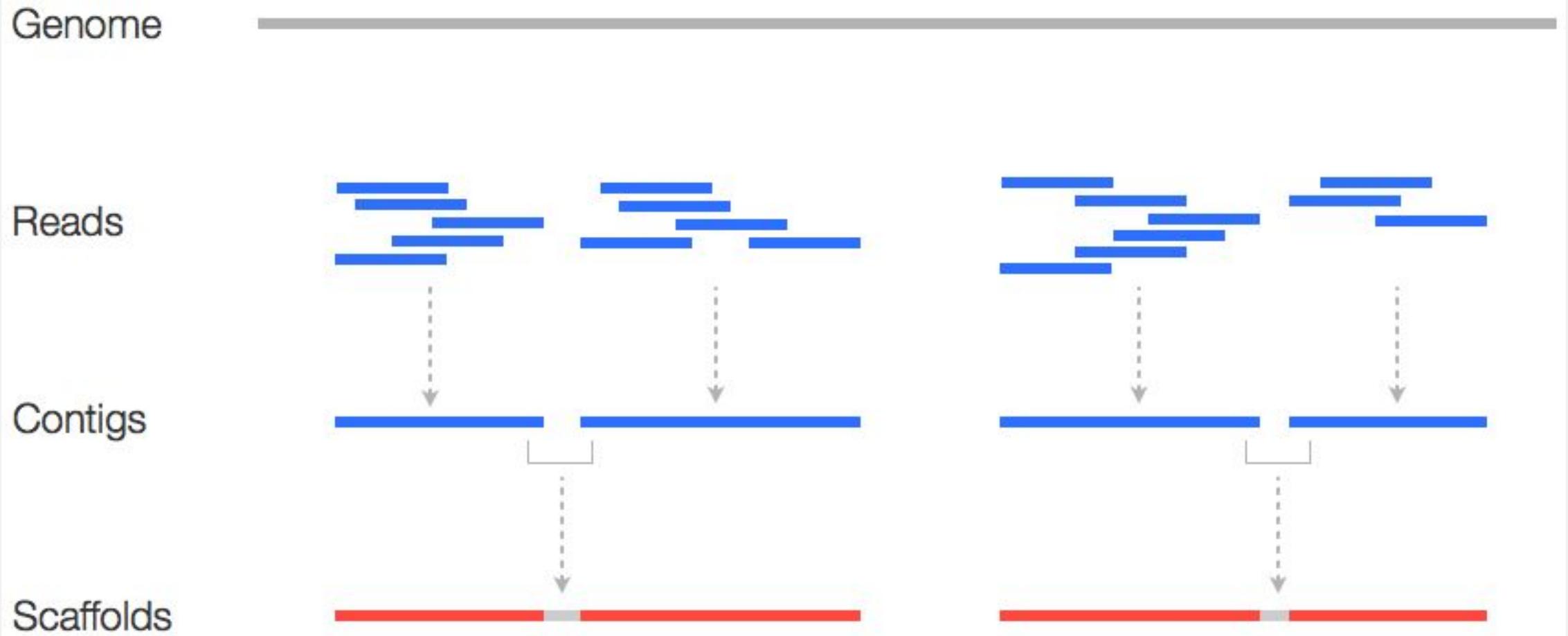
**Estado:** en análisis, en servidor INIA, UNALM.

```

csaldana@bioinformatica:~/olivo/getorganelle
login as: csaldana
csaldana@bioinformatica.lamolina.edu.pe's password:
Last login: Wed Jun 15 15:27:07 2022 from 200.123.20.129
[csaldana@bioinformatica ~]$ cd
[csaldana@bioinformatica ~]$ ls
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1.fq.gz  mapping_quina.pbs          quast_all_more_300      quast_spades
assemble.sh  masurca_assemble         quast_allresults       richard
cat.pbs      masurca_bt.fasta        quast_allresults_test  SOAPdenovo2
cinchona    masurca_cinchona.fasta  quast_cinchona.pbs    soapdenovo2_cinchona_133.scafSeq
gzip.e98127  ncbi                     quast_masurca         soapdenovo2_cinchona_80.scafSeq
gzip.e98139  olivo                    quast_results         soapdenovo2_cinchona_85.scafSeq
gzip.o98127  Q1_R1_001.fastq         quast_richard_bt      soapdenovo2_cinchona_93.scafSeq
gzip.o98139  Q1_R2_001.fastq         quast_richard.pbs     Spades2ndRun
jellyfish   quast_all                quast_soapdenovo2     spades_cinchona.fasta
[csaldana@bioinformatica ~]$ cd cinchona/
[csaldana@bioinformatica cinchona]$ ls
data  logs  masurca.pbs  masurca.pbs.tar.gz  results  scripts
[csaldana@bioinformatica cinchona]$ cd ..
[csaldana@bioinformatica ~]$ cd olivo
[csaldana@bioinformatica olivo]$ ls
aligments
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_fastqc.zip
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1.fq
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_paired_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_paired_fastqc.zip
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_trimomatic_paired.fq.gz
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_unpaired.fq.gz
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_fastqc.zip
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2.fq
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_paired_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_paired_fastqc.zip
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_trimomatic_paired.fq.gz
[csaldana@bioinformatica olivo]$ cd getorganelle/
[csaldana@bioinformatica getorganelle]$ ls
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1.fq.gz_trimming_report.txt
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_val_1_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_1_val_1_fastqc.zip
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2.fq.gz_trimming_report.txt
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_val_1_fastqc.html
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_val_2_fastqc.zip
GeSeqJob-230522.gff3
[csaldana@bioinformatica getorganelle]$
AL_OE_32_CSFP220009433-1a_HK5GFDSX3_L3_2_unpaired.fq.gz
geseq
getorganelle
get_organelle_olivo.pbs
gzip_olivo.e97921
gzip_olivo_novoplasty.e98079
gzip_olivo_novoplasty.e98080
gzip_olivo_novoplasty.e98081
gzip_olivo_novoplasty.o98079
gzip_olivo_novoplasty.o98080
gzip_olivo_novoplasty.o98081
gzip_olivo_novoplasty.pbs
gzip_olivo.o97921
gzip_olivo.pbs
gzip_olivo_trimm.e98069
gzip_olivo_trimm.o98069
gzip_olivo_trim.pbs
gzip.pbs
nohup.out
novoplasty
plastome_oliv
refseq_olivo.fasta
results
trim_olea.log
trim.pbs
trim.pbs.save
wget.pbs
GeSeqJob-240522.gff3
GeSeqJob-240522_novoplasty}.gff3
GeSeqJob_240522_ref_lnovoplasty.gff3
GeSeqJob_getorganelle_ref_1.gff3
GeSeq_novoplasty_olivo_Contig1_GFF3.gff3
getorganelle_olivo.log
getorganelle_trim_olivo.log
nohup.out
plastome_oliv
plastome_olivo
raw.reads
trim_olea.log

```

# Ensamblaje de genomas



# Ensamblaje de genomas

## Evaluación de los ensamblajes : *C. officinalis*

Softwares para ensamblar:

- ✓ Getorganelle
- ✓ NOVOPlasty
- ✓ SOAPdenovo
- ✓ SPAdes
- ✓ MaSuRCA

El N50 es la longitud del contig más corto que ordenando los contigs del mayor al menos alcanza el 50 % del tamaño del genoma total.

Tamaño total: 17 000 pb

Ordenar de mayor a menor:



7000 pb + 5000 pb = 12 000 pb más del 50 %  
entonces N50 = 5000



# Evaluación de los ensamblajes : *C. officinalis*

[View in Icarus contig browser](#)

All statistics are based on contigs of size  $\geq 500$  bp, unless otherwise noted (e.g., "# contigs ( $\geq 0$  bp)" and "Total length ( $\geq 0$  bp)" include all contigs).

Worst Median Best  Show heatmap

Statistics without reference	scaffolds	primary.genome.scf	graph_sample_21.scafSeq	graph_sample_33.scafSeq	graph_sample_55.scafSeq	graph_sample_77.scafSeq
# contigs	416 465	346 820	782	9183	318 708	379 967
# contigs ( $\geq 0$ bp)	3 527 329	435 110	4 327 145	16 266 622	4 137 867	2 956 483
# contigs ( $\geq 1000$ bp)	169 356	192 342	26	216	160 743	185 233
# contigs ( $\geq 5000$ bp)	21 507	7630	0	0	26 541	30 156
# contigs ( $\geq 10000$ bp)	7473	278	0	0	9197	12 499
# contigs ( $\geq 25000$ bp)	1293	4	0	0	1286	2797
# contigs ( $\geq 50000$ bp)	225	0	0	0	158	612
Largest contig	191 184	37 893	2169	4144	123 902	327 284
Total length	691 579 176	518 943 592	486 122	5 632 667	666 903 301	827 470 957
Total length ( $\geq 0$ bp)	1 221 838 118	553 900 976	555 537 602	2 142 489 304	1 197 821 335	1 284 578 653
Total length ( $\geq 1000$ bp)	519 542 226	407 593 443	31 177	277 357	556 932 484	691 982 223
Total length ( $\geq 5000$ bp)	238 362 856	49 533 458	0	0	283 063 503	386 999 983
Total length ( $\geq 10000$ bp)	142 431 775	3 430 129	0	0	163 247 899	264 947 191
Total length ( $\geq 25000$ bp)	51 582 512	137 737	0	0	47 547 755	119 471 582
Total length ( $\geq 50000$ bp)	15 959 553	0	0	0	10 416 065	45 974 060
N50	2551	1854	589	586	3766	4292
N75	1003	1088	536	534	1413	1422
L50	52 379	82 827	329	3926	38 203	35 934
L75	168 500	174 714	547	6452	112 768	125 048
GC (%)	32.92	33.41	40.36	30.84	32.87	32.96
<b>Mismatches</b>						
# N's	2 964 105	509 428	25 630	162 220	59 884 702	30 467 600
# N's per 100 kbp	428.6	98.17	5272.34	2879.99	8979.52	3682.01

# Ensamblaje de genomas

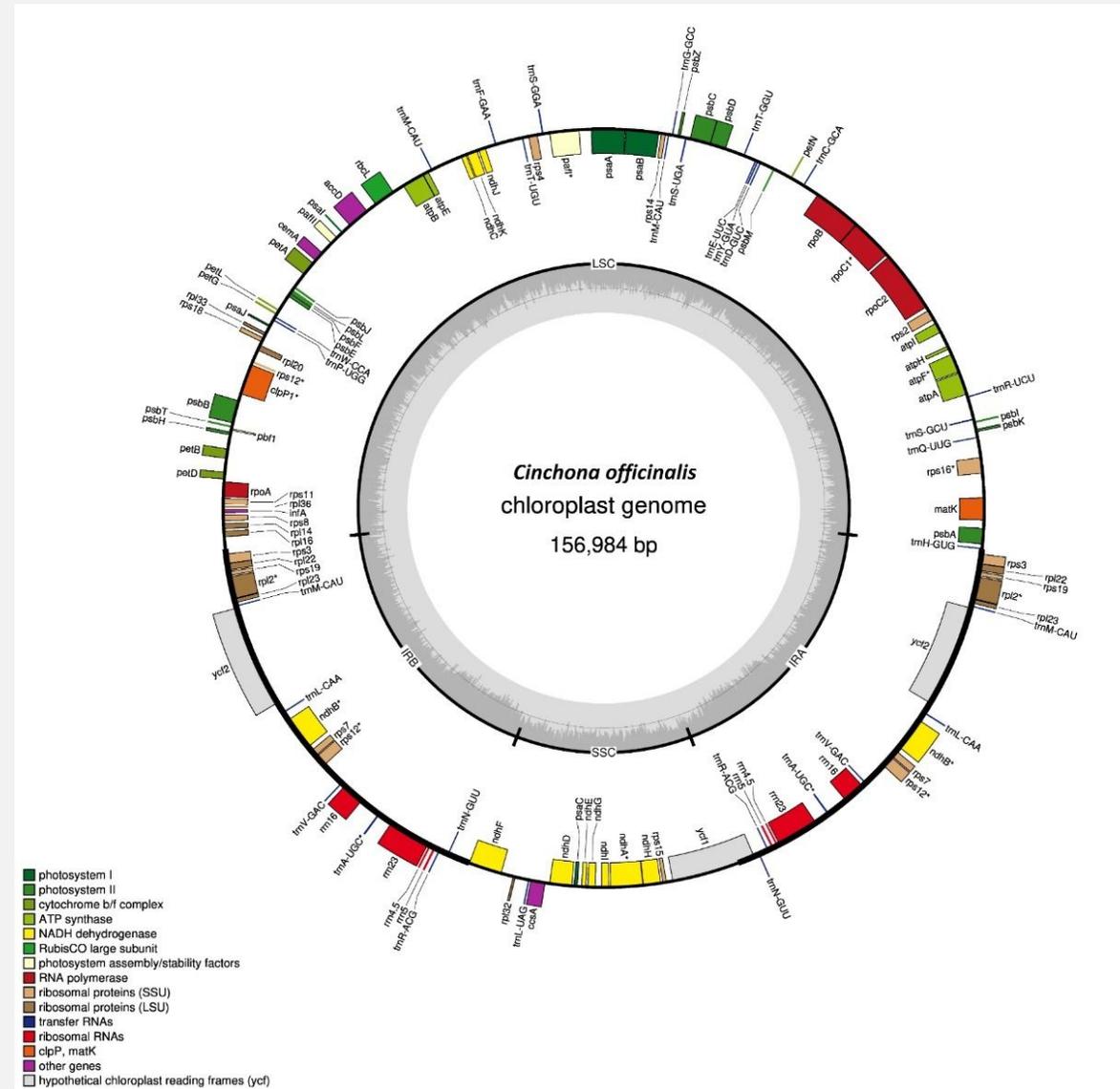
Scaffold: Utilizando un genoma de referencia se ordenan los contigs



# Anotación de genomas

Softwares para anotar:

- ✓ GeSeq
- ✓ PGA
- ✓ Augustus
- ✓ A5
- ✓ Prokka





PERÚ

Ministerio  
de Desarrollo Agrario  
y Riego



Instituto Nacional de Innovación Agraria



# Trabajos en ejecución

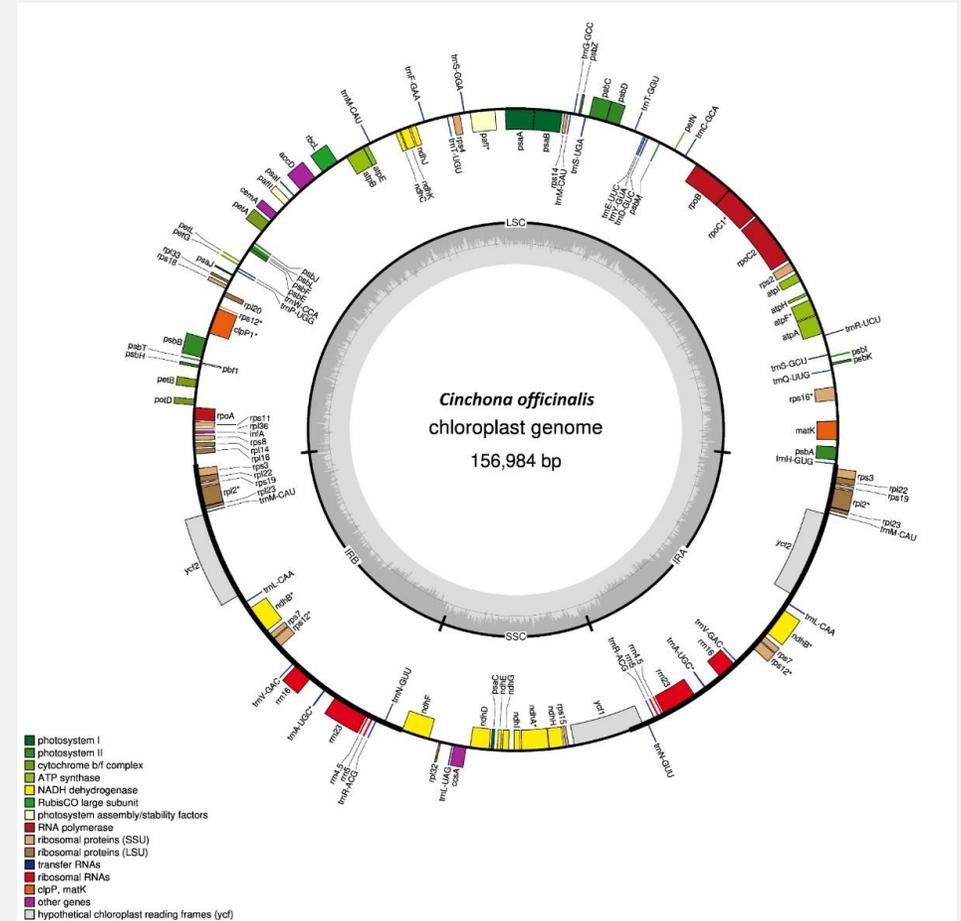
- Árbol de la quina (*Cinchona officinalis*)
- Capirona (*Calycophyllum spruceanum*)
- Ulcumano (*Retrophyllum rospigliosii*)
- Olivo (*Olea europaea*)
- Pepino dulce,
- Ají arnaucho
- Zapallo loche



Quina, árbol bandera del Perú

# Árbol de la quina (*Cinchona officinalis*)

- Muy pocos estudios en todas la áreas
- Genoma disponible en: <https://www.ncbi.nlm.nih.gov/nuccore/MZ151891.1>
- Primer trabajo a nivel molecular para esta especie



- Ensamblaje y anotación de genoma nuclear draft de quina.
- Se estimó el tamaño del genoma de la quina: **0.6 Gbp**

 **data** 

Article 1

## Finally, out of the dark: the first draft genome assembly of the national tree of Peru (*Cinchona officinalis*, Rubiaceae) 2 3

Carla L. Saldaña <sup>1</sup>, Richard Estrada<sup>1</sup>, Julio C. Chávez Galarza <sup>2</sup>, Juan C.4 Guerrero-Abad <sup>3</sup>, Héctor V. Vásquez <sup>1</sup>, Jorge L. Maicelo <sup>1</sup>, Carlos I. Arbizu <sup>1, 2</sup>. 4 5

<sup>1</sup> Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Perú; carla18317@gmail.com (C.L.S.); hvasquez@inia.gov.pe (H.V.V.); jmaicelo@inia.gov.pe (J.L.M.) 6 7 8

<sup>2</sup> Universidad Nacional de Barranca, Jr. José Gálvez 557, Barranca; jochavezgalarza@gmail.com (J.C.C.-G.) 9

<sup>3</sup> Dirección de Recursos Genéticos y Biotecnología, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; jguerreroa@inia.gov.pe (J.C.G.-A) 10 11

\* Correspondence: carbizu@inia.gov.pe 12

**Abstract:** *Cinchona officinalis* is a forest tree species native to the South American Andean zone. In Peru, it is considered an emblematic plant resource of the country's flora because it symbolizes the plant diversity of the Peruvian territory. Its economic value is due to the content of quinine, an alkaloid widely recognized to prevent and control malaria. Currently quina is considered a neglected forest species as genetic and genomic resources for this species are still limited. We here report the first draft genome assembly for the national tree of Peru. A single *C. officinalis* tree was selected to be sequenced from the Cajamarca region (620058.000S, 7903027.800W). High-quality genomic DNA was extracted from fresh leaves by the CTAB method, with minor modifications. The whole genome was sequenced using a shotgun paired-end 150 strategy on the Illumina HiSeq 2500 platform. We generated approximately 494.5 million raw paired-end reads and a GC content of 35.97%. To estimate the genome size, low-quality regions of raw PE and universal adapters were removed, and the genome size of *C. officinalis* was subsequently estimated at 0.6 Gbp where the main peak lied at the k-mer depth of 100. This genome also presents a 0.996 % of heterozygosity. The genome data provided here is expected to contribute to a better understanding of the genetics of this species, as well as important molecular pathways that could be crucial for the biology, management and promoting its appropriate management and genomics-assisted breeding. 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28

Citation: Saldaña, C.L.; Chávez-Galarza, J.C.; Guerrero-Abad, J.C.; Vásquez, H.V.; Maicelo, J.L.; Arbizu, C.I. Finally out of the dark: the first draft genome assembly of the national tree of Peru (*Cinchona officinalis*, Rubiaceae). *Data* 2022, 7, Firstpage-Lastpage. <https://doi.org/10.3390/xxxxx>

# Capirona (*Calycophyllum spruceanum*)



Article

## Genetic Diversity and Population Structure of Capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon Revealed by RAPD Markers

Carla L. Saldaña<sup>1</sup>, Johan D. Cancan<sup>1</sup>, Wilbert Cruz<sup>1</sup>, Mirian Y. Correa<sup>2</sup>, Miriam Ramos<sup>3</sup>, Eloy Cuellar<sup>3</sup> and Carlos I. Arbizu<sup>1,4\*</sup>

- 1 Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; carla18317@gmail.com (C.L.S.); clg.johan@gmail.com (J.D.C.); wilbertcruz2002@gmail.com (W.C.)
- 2 Facultad de Ciencias Biológicas, Universidad Nacional Pedro Ruiz Gallo, Juan XXIII No 391, Lambayeque 14013, Peru; miryacormell5@gmail.com
- 3 Facultad de Ciencias Forestales, Universidad Nacional Agraria La Molina, Av. La Molina s/n, Lima 15024, Peru; miriam.piaforestal@gmail.com (M.R.); eloycuellar@lamolina.edu.pe (E.C.)
- \* Correspondence: carbizu@inia.gob.pe

**Abstract:** Capirona (*Calycophyllum spruceanum* Benth.) is a tree species of commercial importance widely distributed in South American forests that is traditionally used for its medicinal properties and wood quality. Studies on this tree species have been focused mainly on wood properties, propagation, and growth. However, genetic studies on capirona have been very limited to date. Currently, it is possible to explore genetic diversity and population structure in a fast and reliable manner by using molecular markers. We here used 10 random amplified polymorphic DNA (RAPD) markers to analyze the genetic diversity and population structure of 59 samples of capirona that were sampled from four provinces located in the eastern region of the Peruvian Amazon. A total of 186 bands were manually scored, generating a 59 × 186 presence/absence matrix. A dendrogram was generated using the UPGMA clustering algorithm, and, similar to the principal coordinate analysis (PCoA), it showed four groups that correspond to the geographic origin of the capirona samples (LBS, Irazola, Masisea, Inapari). Similarly, a discriminant analysis of principal components (DAPC) and STRUCTURE analysis confirmed that capirona is grouped into four clusters. However, we also noticed that a few samples were intermingled. Genetic diversity estimation was conducted considering the four groups (populations) identified by STRUCTURE software. AMOVA revealed the greatest variation within populations (71.56%) and indicated that variability among populations is 28.44%. Population divergence ( $F_{st}$ ) between clusters 1 and 4 revealed the highest genetic difference (0.269), and the lowest  $F_{st}$  was observed between clusters 3 and 4 (0.123). RAPD markers were successful and effective. However, more studies are needed, employing other molecular tools. To the best of our knowledge, this is the first investigation employing molecular markers in capirona in Peru considering its natural distribution, and as such it is hoped that this helps to pave the way towards its genetic improvement and the urgent sustainable management of forests in Peru.

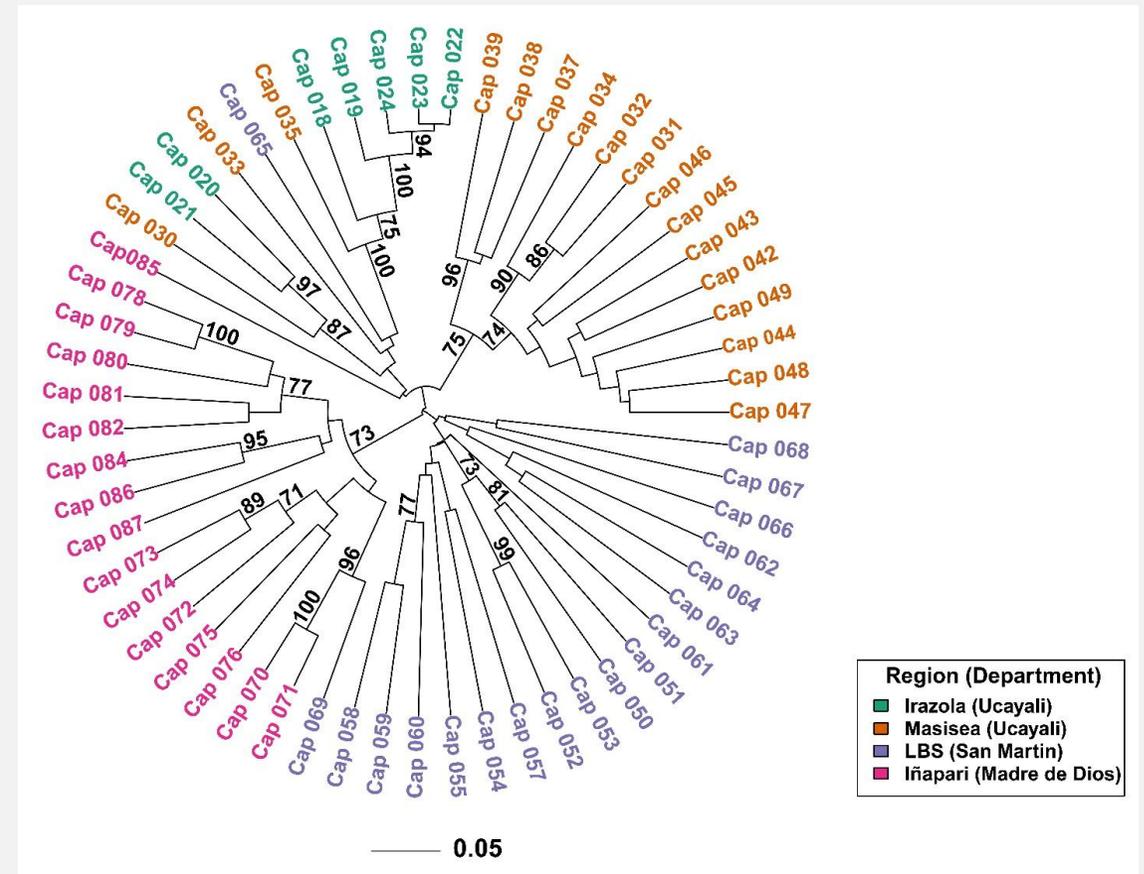


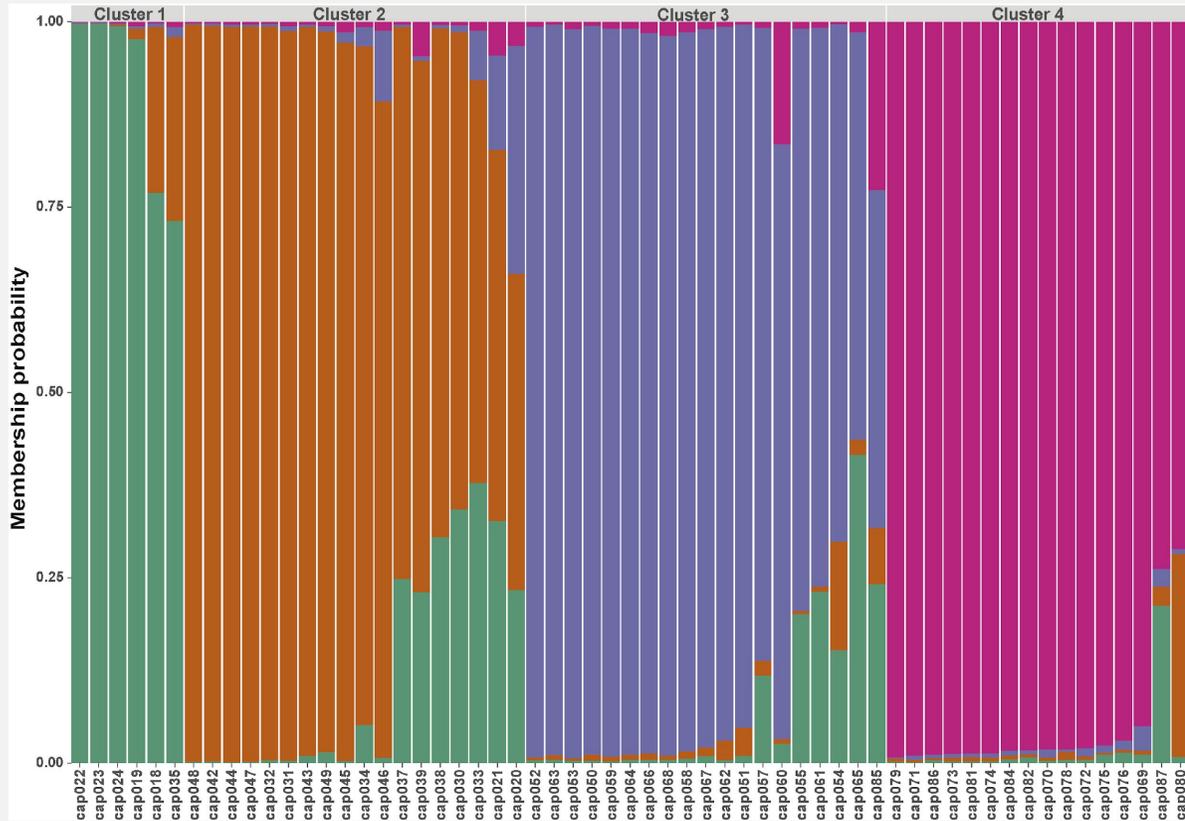
**Citation:** Saldaña, C.L.; Cancan, J.D.; Cruz, W.; Correa, M.Y.; Ramos, M.; Cuellar, E.; Arbizu, C.I. Genetic Diversity and Population Structure of Capirona (*Calycophyllum spruceanum* Benth.) from the Peruvian Amazon Revealed by RAPD Markers. *Forests* **2021**, *12*, 1125. <https://doi.org/10.3390/f12081125>

Academic Editors: Carol A. Loopstra and Timothy A. Martin

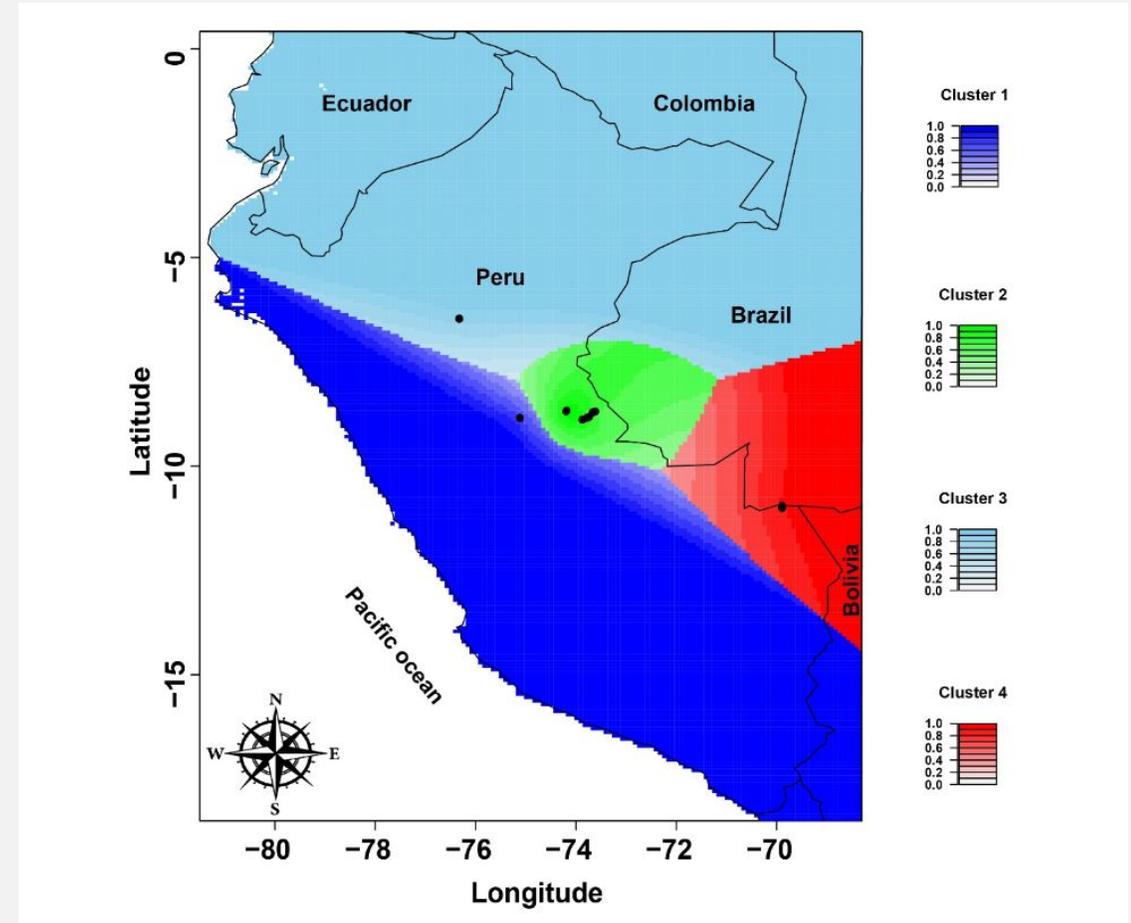
Received: 14 July 2021  
Accepted: 20 August 2021  
Published: 22 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.





**Análisis de estructura poblacional de capirona. Saldaña et al. 2021**



**Ubicación geográfica de los clusters. Saldaña et al. 2021**

**genes** MDPI

## Unlocking the Complete Chloroplast Genome of a Native Tree Species from the Amazon Basin, Capirona (*Calycophyllum spruceanum*, Rubiaceae), and Its Comparative Analysis with Other Ixoroidae Species

Carla L. Saldaña <sup>1</sup>, Pedro Rodríguez-Grados <sup>1,2</sup>, Julio C. Chávez-Galarza <sup>1</sup>, Shefferson Feljoo <sup>3</sup>, Juan Carlos Guerrero-Abad <sup>4</sup>, Héctor V. Vásquez <sup>1</sup>, Jorge L. Malcoto <sup>1</sup>, Jorge H. Jhonson <sup>5,6</sup> and Carlos I. Arbizu <sup>1,\*</sup>

**Abstract:** Capirona (*Calycophyllum spruceanum* Benth.) belongs to subfamily Ixoroidae, one of the major lineages in the Rubiaceae family, and is an important timber tree. It originated 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 the 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 PE150 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 one pseudogene. A total of 41 simple sequence repeats (SSR) of this cp genome were divided into mononucleotides (29), dinucleotides (5), trinucleotides (3), and tetranucleotides (4). Most of these repeats were distributed in the noncoding regions. Whole chloroplast genome comparison with the other six Ixoroidae species revealed that the small single copy and large single copy regions showed more divergent than inverted regions. Finally, phylogenetic analyses resolved that *C. spruceanum* is a sister species to *Entmeopteryx hetero* and confirms its position within the subfamily Ixoroidae. 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; phylogenetics

Check for updates

Genes 2022, 13, 113. <https://doi.org/10.3390/genes13010113>

Academic Editor: Marcial Escobedo

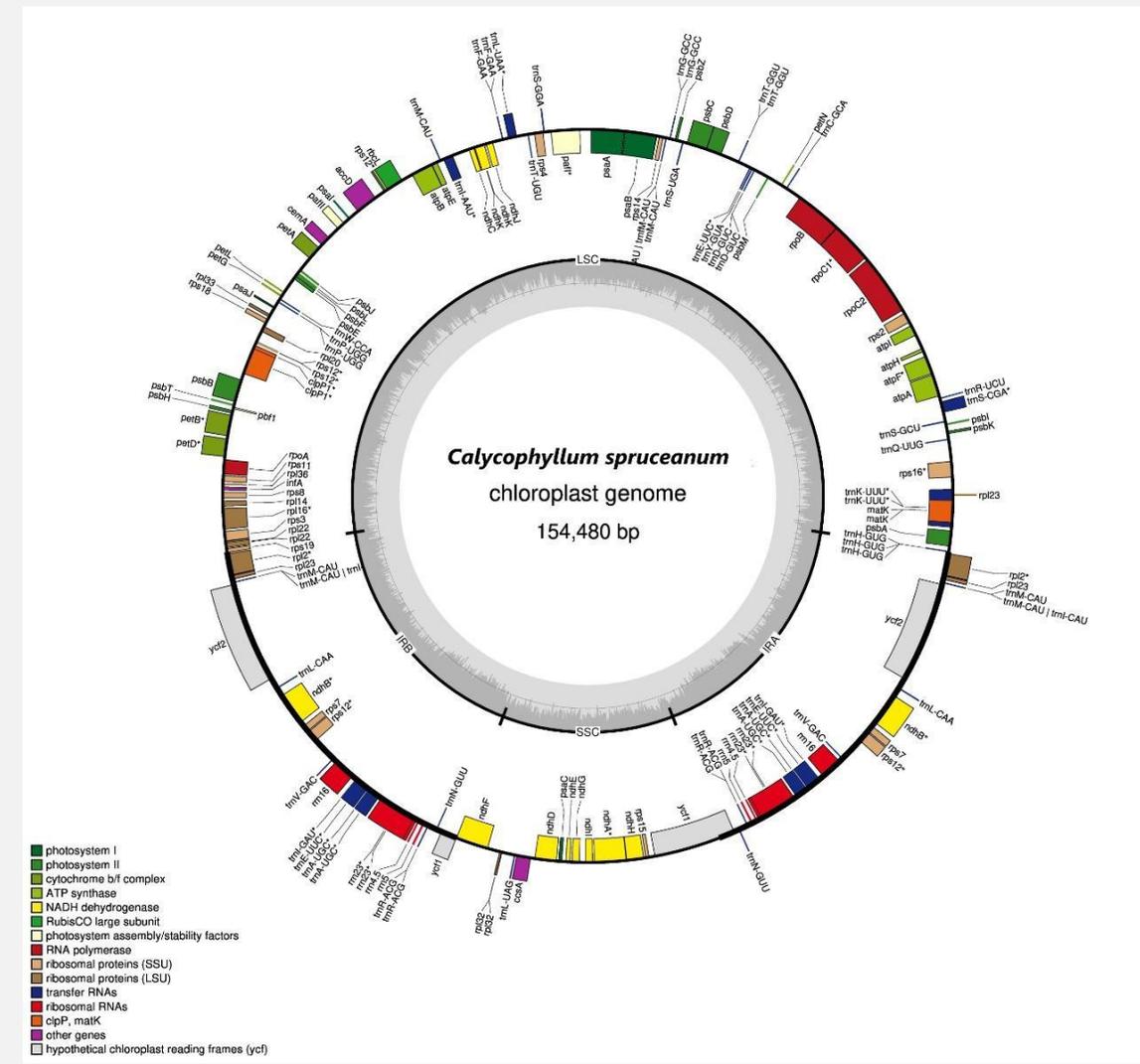
Received: 30 November 2021  
Accepted: 7 January 2022  
Published: 7 January 2022

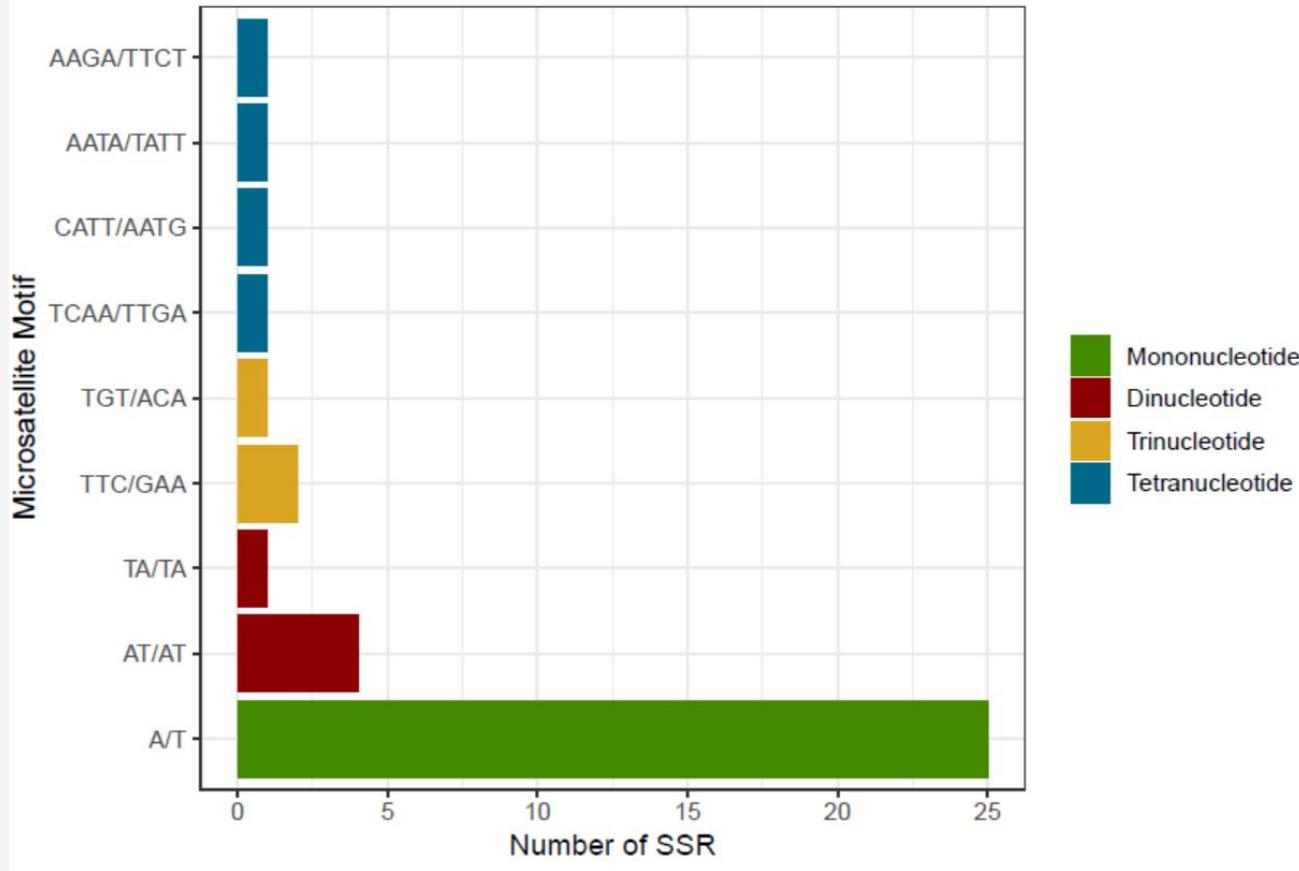
**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

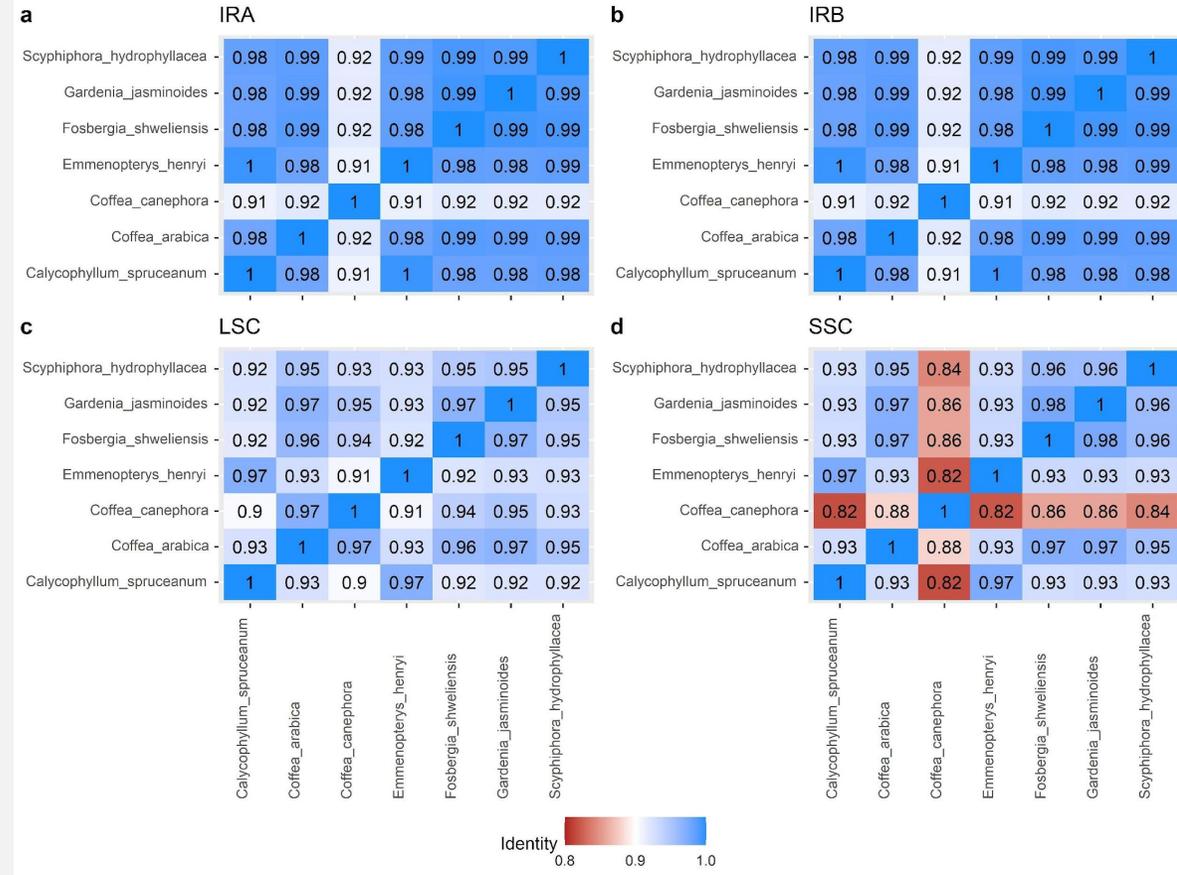
IMPACT FACTOR 4.096

Covered in: PubMed





Identificación de marcadores moleculares SSR. Saldaña et al. 2022. *Genes* 13: 113.



Matriz de identidades. Saldaña et al. 2022. *Genes* 13: 113.



PERÚ

Ministerio  
de Desarrollo Agrario  
y Riego

Instituto Nacional de Innovación Agraria



## Calycophyllum spruceanum chloroplast, complete genome

GenBank: OK326865.1

[FASTA](#) [Graphics](#)[Go to:](#) 

LOCUS OK326865 154480 bp DNA circular PLN 07-NOV-2021  
DEFINITION Calycophyllum spruceanum chloroplast, complete genome.  
ACCESSION OK326865  
VERSION OK326865.1  
KEYWORDS .  
SOURCE chloroplast Calycophyllum spruceanum  
ORGANISM [Calycophyllum spruceanum](#)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliopsida; eudicotyledons; Gunneridae;  
Pentapetalae; asterids; lamiids; Gentianales; Rubiaceae;  
Ixoroideae; Condamineae; Calycophyllum.  
REFERENCE 1 (bases 1 to 154480)  
AUTHORS Saldana,C., Arbizu,C., Rodriguez-Grados,P., Chavez-Galarza,J.,  
Feijoo,S., Guerrero-Abad,J.C. and Maicelo-Quintana,J.  
TITLE Unlocking the complete chloroplast genome of capirona  
(Calycophyllum spruceanum Benth., Rubiaceae), a native tree species  
from the Amazon basin  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 154480)  
AUTHORS Saldana,C., Arbizu,C., Rodriguez-Grados,P., Chavez-Galarza,J.,  
Feijoo,S., Guerrero-Abad,J.C. and Maicelo-Quintana,J.  
TITLE Direct Submission  
JOURNAL Submitted (28-SEP-2021) Direccion de Desarrollo Tecnologico  
Agrario, Instituto Nacional de Innovacion Agraria, Av. La Molina

<https://www.ncbi.nlm.nih.gov/nucleotide/OK326865>

# Estudio sobre capirona presentado en el PAG 2022

The screenshot shows the website for the International Plant & Animal Genome Conference XXIX. The header includes the conference logo, the title "PLANT & ANIMAL GENOME CONFERENCE XXIX", the tagline "The Largest Ag-Genomics Meeting in the World.", and the dates "January 8-12, 2022" in "San Diego, CA". The website URL "www.intlpag.org" is also present. On the left, there is a navigation menu with options: Home, Happening Now, Search, Browse by Type, Browse by Day, Authors, Poster Categories, New Workshops for XXIX!, Town & Country Resort Map, and Technical Support. At the bottom of the menu, there are two time zone options: "Meeting Time - PST" and "My Time - GMT-5". The main content area displays the abstract for paper PE0142, titled "The Complete Chloroplast Genome of a Native Tree Species from the Amazon Basin, Capirona (*Calycophyllum spruceanum* Benth., Rubiaceae), and Comparative Analysis with Other Ixoroideae Species".

My Time

INTERNATIONAL PLANT & ANIMAL GENOME CONFERENCE XXIX

January 8-12, 2022  
San Diego, CA

www.intlpag.org

PLANT & ANIMAL GENOME CONFERENCE XXIX

The Largest Ag-Genomics Meeting in the World.

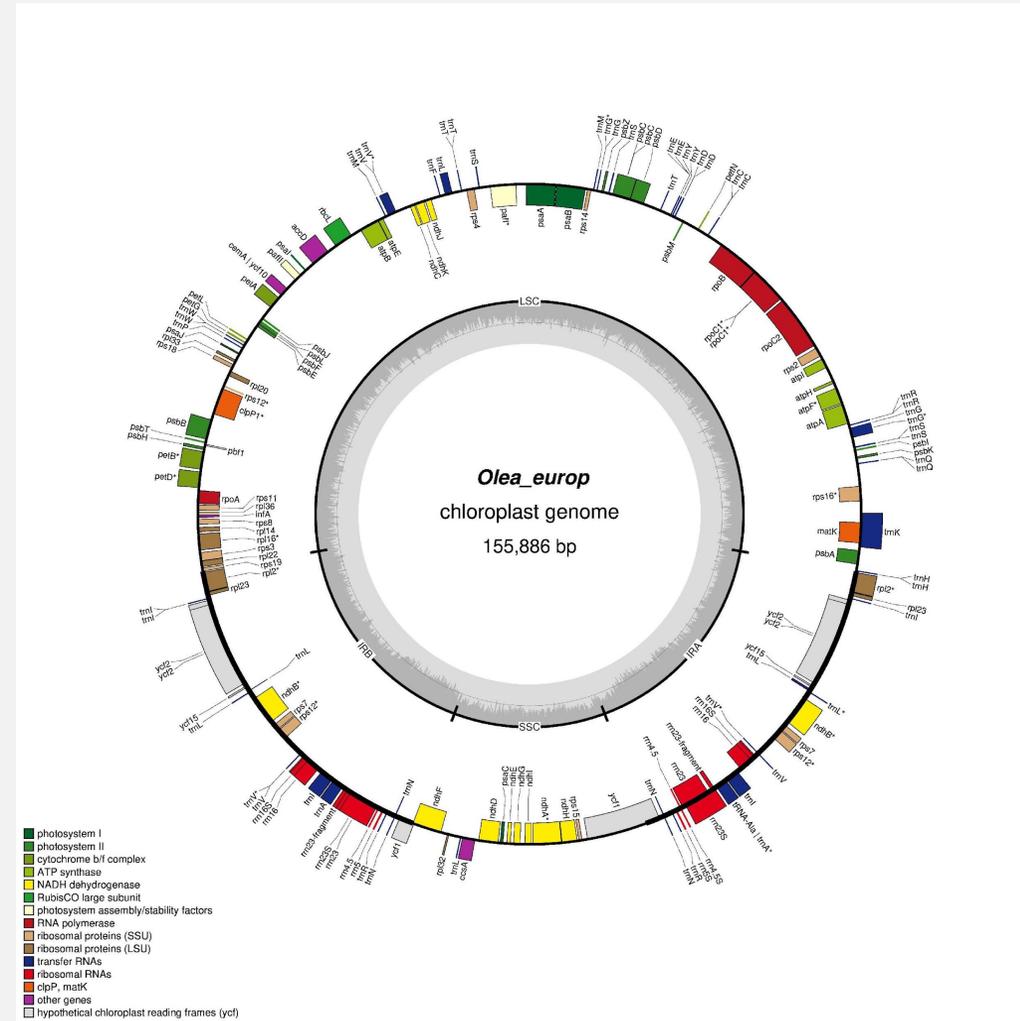
## PE0142 - The Complete Chloroplast Genome of a Native Tree Species from the Amazon Basin, Capirona (*Calycophyllum spruceanum* Benth., Rubiaceae), and Comparative Analysis with Other Ixoroideae Species

### Abstract

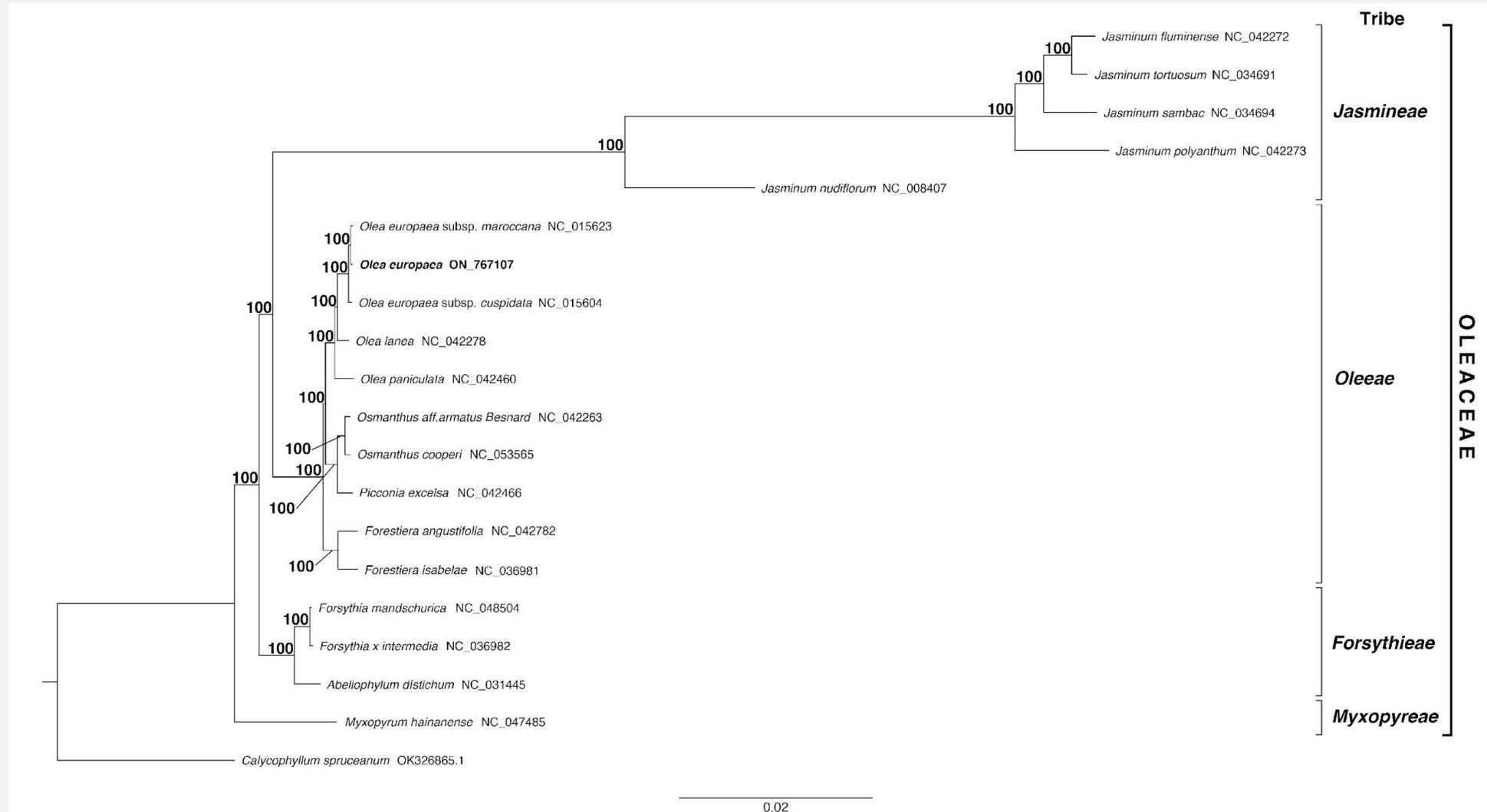
Capirona (*Calycophyllum spruceanum* (Benth.), Rubiaceae), is an important timber tree, with origin in Amazon Basin and has widespread distribution in Bolivia, Peru, Colombia, and Brazil. In this study, we obtained the first complete chloroplast (cp) genome of capirona from department of Madre de Dios 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 chloroplast (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 an 18,101 bp small single-copy (SSC) region, 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 pseudogen. A total of 29 simple sequence repeats (SSR) of this cp genome were divided into 10-nucleotide (10), 11-nucleotide (2), 12-nucleotide (2), 13-nucleotide (4), 14-nucleotide (1), 15-nucleotide (5), 17-nucleotide (2), 18-nucleotide (2), 22-nucleotide (1). Finally, phylogenetic analysis resolved that capirona 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 *Calycophyllum spruceanum* and has been compared with six species of the subfamily Ixoroideae, providing valuable information for genetic and evolutionary studies in the genus *Calycophyllum* and beyond.

Meeting Time - PST  
My Time - GMT-5

# Olivo (*Olea europaea*)



Genoma cloroplastidial de *Olea europea*. Saldaña et al. Sometido



Árbol filogenético de *Olea europaea*. Saldaña et al. Sometido



PERÚ

Ministerio  
de Desarrollo Agrario  
y Riego



Instituto Nacional de Innovación Agraria



# Próximas investigaciones

- Lupuna (*Ceiba pentandra*) – Universidad de Ucayali
- Bolaina (*Guazuma crinita*) - Universidad Nacional de San Martín
- Shihuahuco



PERÚ

Ministerio  
de Desarrollo Agrario  
y Riego



Instituto Nacional de Innovación Agraria



Inclusión del INIA en el 10KP: 10,000 Plant Genomes Project y en otros consorcios adicionales



# 10KP

Beta

HOME

STRATEGY

TIMELINE

SPECIES DATABASE

PRIORITY LIST

USER CENTER

# Equipo de trabajo





PERÚ

Ministerio de Desarrollo Agrario y Riego



Instituto Nacional de Innovación Agraria

Siempre con el pueblo



Universidad Nacional Agraria  
**LA MOLINA**

