Check for updates

Ecology and Evolution

WILEY

Ecology and Evolution Open Access

GENETICS NOTES OPEN ACCESS

The Complete Chloroplast Genome of Tornillo (*Cedrelinga cateniformis* Ducke 1922, Fabaceae)

Nora Scarcelli[1] | Cédric Mariac[1] | Marie Couderc[1] | Diana Castro Ruiz[2] | Guillain Estivals[2] | Carlos Alberto Custodio Angulo Chavez[2] | Hector Acho Vasquez[2] | Jhon Gregory Alvarado Reategui[2] | Tony Vizcarra Bentos[2] | Carmen Garcia-Davila[2]

[1]DIADE, Univ Montpellier, Cirad, IRD, Montpellier, France | [2]Instituto de Investigaciones de la Amazonía Peruana (IIAP), Laboratorio de Biología y Genética Molecular (LBGM), Iquitos, Peru

Correspondence: Nora Scarcelli (nora.scarcelli@ird.fr)

Received: 10 February 2025 | **Revised:** 28 March 2025 | **Accepted:** 15 April 2025

Funding: This study was supported by the OBAP (Observatoire de la Biodiversité de l'Amazonie Péruvienne), co-funded by IRD (Institut de Recherche pour le Développement) and IIAP (Instituto de Investigaciones de la Amazonía Peruana).

Keywords: *Cedrelinga cateniformis* | chloroplast genome | Fabaceae | tornillo

ABSTRACT

Tornillo (*Cedrelinga cateniformis* Ducke 1922) is a tropical tree of the Fabaceae family. It is commonly used in the lumber factory and is an interesting substitute to overexploited tropical timber species. We sequenced and assembled the first whole chloroplast genome of Tornillo, using Oxford Nanopore technology. The Tornillo's chloroplast is a circular molecule of 176,700 bp with 138 genes and a classic quadripartite structure. The Inverted Repeats present a huge expansion, combined to a strong reduction of the Short Single Copy. Similar results were previously observed in other species of the tribe Ingeae. A maximum likelihood phylogenetic tree reveals a clear distinction of the Ingeae tribe from other tribes (Acacieae and Mimoseae) of the Fabaceae family.

1 | Introduction

Cedrelinga cateniformis Ducke 1922 is a tropical tree species of the family Fabaceae (Figure 1) commonly known as 'Tornillo' in Peru. Tornillo has a wide ecological distribution in humid tropical, subtropical and dry tropical environments. It usually grows in environments with annual precipitations ranging from 2500 to 3800 mm and average temperatures from 23°C to 38°C. It is reported in the Amazon regions of Ecuador, Peru, Colombia and Brazil, from 120 to 800 m above sea level (Cruz et al. 2020). Due to its rapid growth in natural and managed environments, Tornillo is considered a secondary succession species (Baluarte-Vásquez and Alvarez-Gonzales 2015; Guariguata et al. 2017).

According to FAO (2018), Tornillo is one of the main species used as sawn wood from tropical forests. Its timber potential for various uses in the furniture and construction industry (Haag et al. 2020), as well as its suitability for restoring degraded areas (Rojas Briceño et al. 2020), places it as an economic alternative for cultivation in agroforestry or polyculture systems. Tornillo's wood is easy to work with, providing a good finish, which explains its use in carpentry (Gonçalez and Gonçalves 2001). As it is now considered a substitute for overexploited tropical species (Haag et al. 2020), it is urgent to draw a strategy for its exploitation and conservation. Yet, there are very few studies on its genetic diversity and few genetic resources are available (Cruz et al. 2020). In an attempt to increase these genetic resources, we present here the first complete chloroplast genome of *Cedrelinga cateniformis*.

2 | Materials and Methods

We collected fresh leaves from a single Tornillo tree in the research centre of Jenaro Herrera, IIAP, Peru, at the coordinates

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Ecology and Evolution* published by British Ecological Society and John Wiley & Sons Ltd.

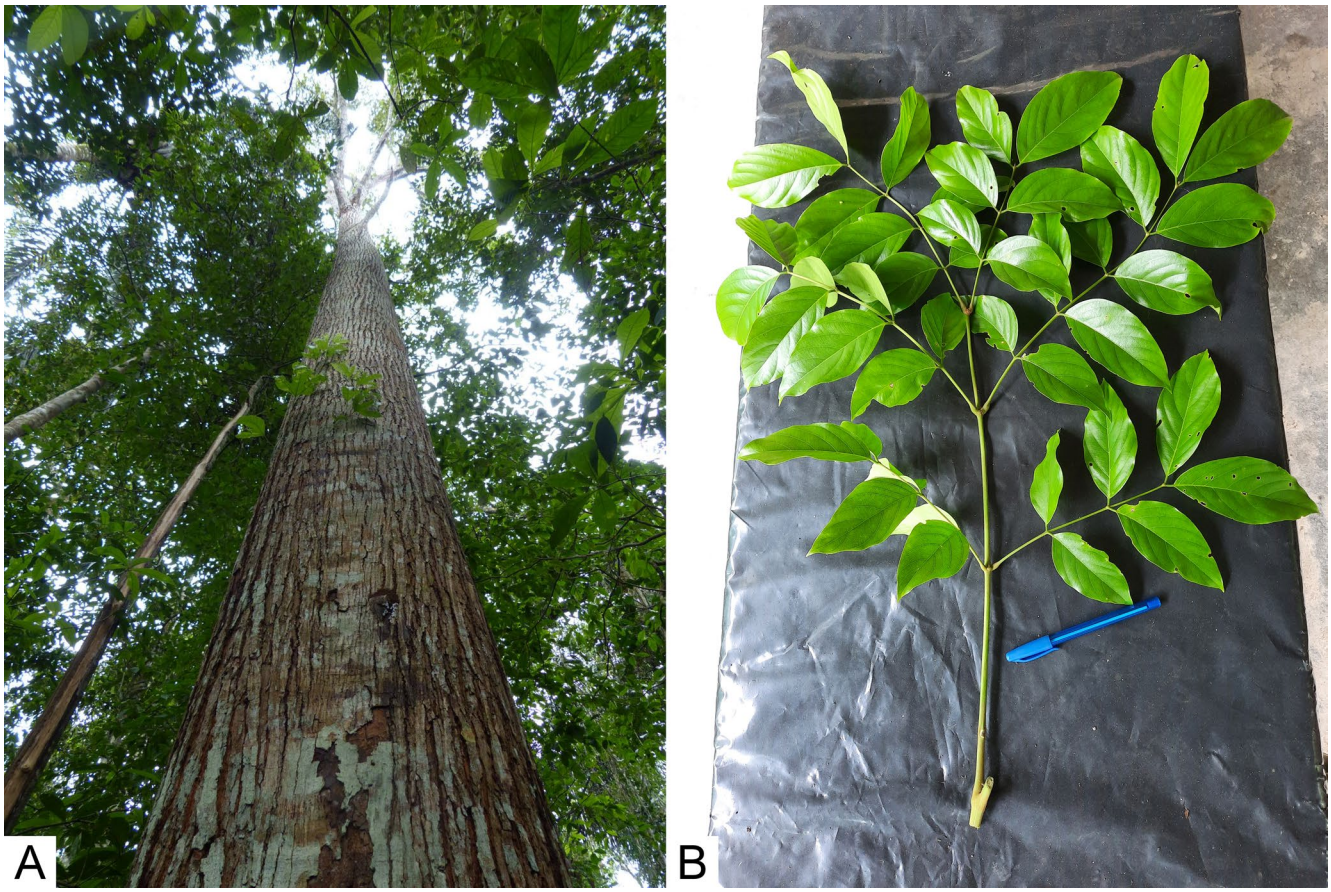

FIGURE 1 | Tornillo tree (A) and details of young leaves (B). Photos taken by T. Vizcarra Bentos in the IIAP research centre of Jenaro Herrera.

−4°53′59.4024 N/−73°38′55.7016E (WGS84). The voucher MERP930 of the plant is available in the Herbario Herrerense (Data S1. Herbarium HH, headquarters Iquitos, contact Dr. Dennis del Castillo Torres, ddelcastillo@iiap.gob.pe).

We extracted DNA of high molecular weight following Scarcelli et al. (2024). We then constructed a single library using the ligation sequencing kit DNA SQL-LSK110 (Oxford Nanopore Technology), following the constructor's recommendations. We sequenced the library on an R9.4 flowcell, using a MinION Mk1B. The base calling was done using the high accuracy model of Minknow 7.1.4 + d7df870c0 with the configuration file dna_r9.4.1_450bps_sup.cfg. Only high-quality reads (Q-score > 10) were kept for the following steps.

We used the pipeline ptGAUL 1.0.5 (Zhou et al. 2023) to generate the whole chloroplast genome assembly. This pipeline first discarded non-chloroplast reads, based on alignment on a chloroplast reference using Minimap2 (Li 2018). We used *Albizia julibrissin* (NC_058305.1) and *Senegalia senegal* (NC_045513.1) for reference because, at that time, these two species were the closest species to *C. cateniformis* with a whole chloroplast available. The pipeline then discarded short reads (< 1000 bp) and reduced the coverage to 50X (Data S2). Finally, the pipeline used Flye (Kolmogorov et al. 2019) to generate the whole genome and Racon (Vaser et al. 2017) to polish it. We annotated the whole chloroplast genome using GeSeq (Tillich et al. 2017), with *Albizia odoratissima* (NC_034987.1) as reference and default parameters. Finally, we manually checked the annotations using Geneious Prime 2023.2.1. We observed 38 missing bases in ORF (0.021% of total sequence length), all in homopolymer regions. This type of sequencing error is common for ONT sequencing and well documented (Delahaye and Nicolas 2021). We therefore included 38 N to correct these missing bases.

We generated a phylogenetic analysis with 11 species of the Fabaceae family (Table 1). Species were chosen to span all the

TABLE 1 | List of species used to perform the phylogeny.

Species	Accession	Tribe
Cedrelinga cateniformis	PQ868097	Ingeae
Albizia julibrissin	NC_058305.1	Ingeae
Samanea saman	NC_034992.1	Ingeae
Pararchidendron pruinosum	NC_035348.1	Ingeae
Sphinga acatlensis	NC_047398.1	Ingeae
Mimosa pudica	NC_042921.1	Mimoseae
Leucaena trichandra	NC_028733.1	Mimoseae
Prosopis cineraria	NC_049133.1	Mimoseae
Senegalia senegal	NC_045513.1	Acacieae
Vachellia nilotica	NC_045514.1	Acacieae
Citrus limon	NC_034690.1	Rutaceae

20457758, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ece3.71355 by Cochrane Peru, Wiley Online Library on [14/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tribes of the Mimosoideae sub-family, except Mimozygantheae because no complete chloroplast had been published at that time. Within each tribe, species were chosen at random, except for *Albizia julibrissin* and *Senegalia senegal*, which were already chosen for the chloroplast assembly. Briefly, we aligned whole chloroplast genomes with MAFFT 7.505 (Katoh and Standley 2013). Then we performed a Maximum Likelihood phylogenetic analysis with RAxML 8.2.12 (Stamatakis 2014), using the GTRGAMMA substitution model, *Citrus limon* (NC_034690.1) as outgroup and 100 bootstraps.

All code lines used to perform analyses are available in Data S3.

3 | Results and Discussion

The complete chloroplast genome of Tornillo (*Cedrelinga cateniformis*) was assembled as a circular molecule of 176,700 bp with a classic quadripartite structure (Figure 2). The Large Single Copy (LSC) spanned 92,034 bp and contained 32.9% GC. The Short Single Copy (SSC) was only 5010 bp with the lowest GC content (28.7%) while the two inverted Repeats (IRa and IRb) spanned 39,828 bp each, with the highest GC content (38.5%).

Gene annotation recovered 112 unique genes, comprising 78 protein coding genes, 30 tRNA and four rRNA. 26 genes were duplicated in the IRs: *ndh*A, *ndh*B, *ndh*D, *ndh*E, *ndh*G, *ndh*H,

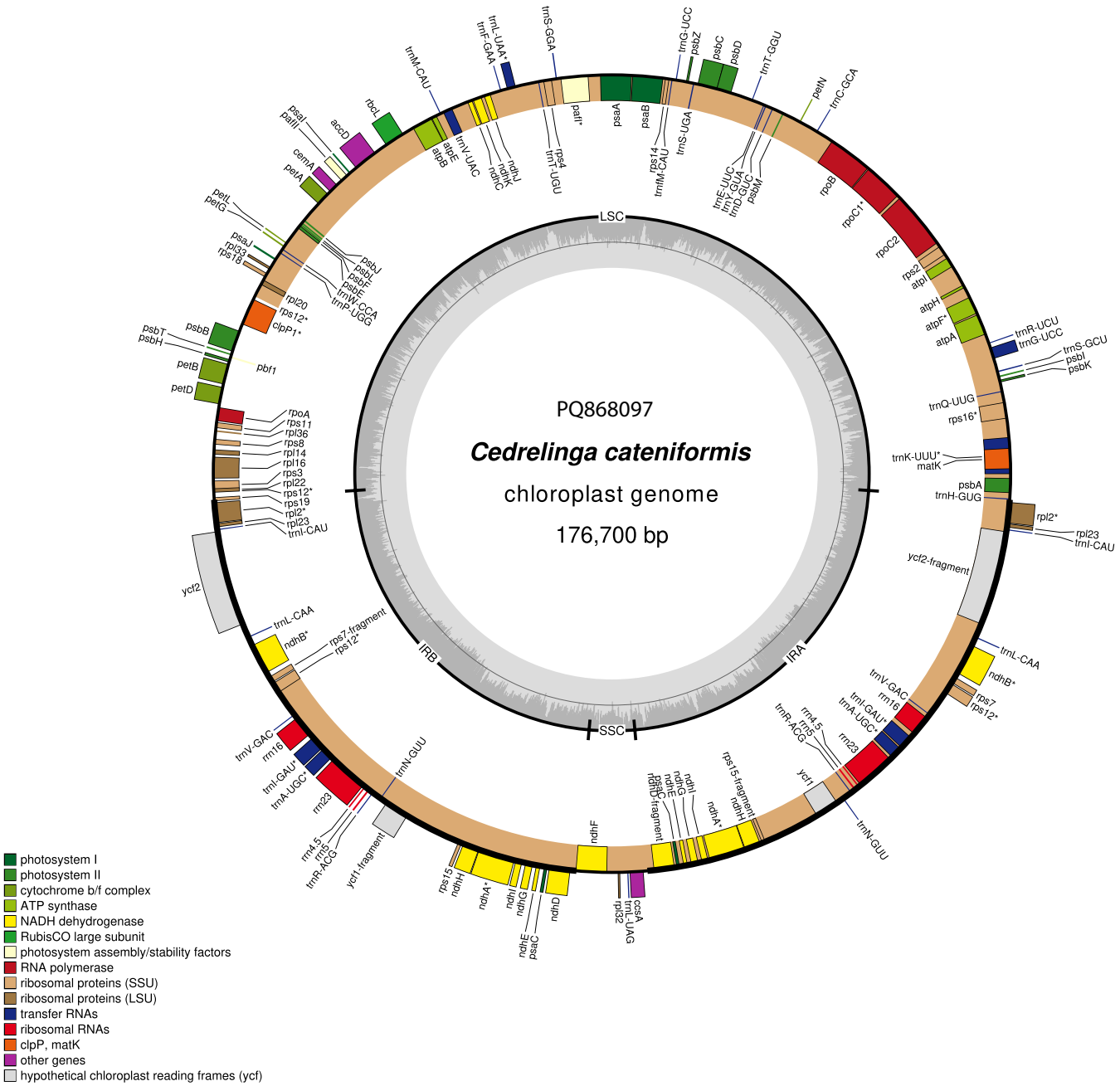

FIGURE 2 | Genome map of *Cedrelinga cateniformis* chloroplast PQ868097, drawn using ORGDRAW (Greiner et al. 2019). Inner genes are translated in a clockwise direction; outer genes are translated counterclockwise. The inner circle represents the GC content. Genes marked with an asterisk contain at least one intron.

20457758, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ece3.71355 by Cochrane Peru, Wiley Online Library on [14/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

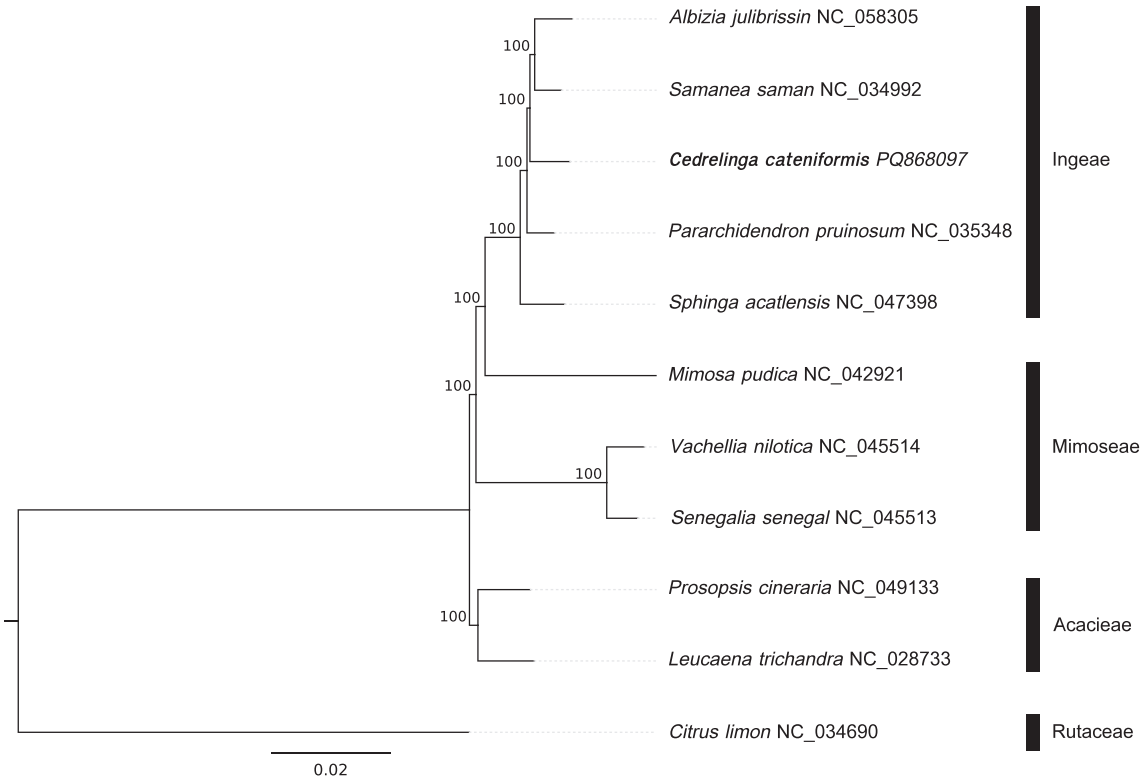

FIGURE 3 | Phylogenetic tree generated by RAxML with bootstrap values indicated on nodes.

*ndh*I, *psa*C, *rpl*2, *rpl*23, *rps*7, *rps*12, *rps*15, *ycf*1, *ycf*2, *trn*A-UGC, *trn*I-CAU, *trn*I-GAU, *trn*L-CAA, *trn*N-GUU, *trn*R-ACG, *trn*V-GAC, *rrn*16, *rrn*23, *rrn*4.5 and *rrn*5. The trans-splicing gene *rps*12 showed a complex but classical structure (Data S4): exon 1 was located in the LSC, while exons 2 and 3 were duplicated and inverted in the IRs. This structure produced two transcripts, one with exon 1 and 2/3 from IRa and the other with exon 1 and 2/3 from IRb.

The chloroplast structure of *Cedrelinga cateniformis* was similar to that of other species of the tribe Ingeae. However, compared to the other species of the Fabaceae family, all the species of the tribe Ingeae analysed here (*Cedrelinga cateniformis*, *Albizia julibrissin*, *Samanea saman*, *Pararchidendron pruinosum* and *Sphinga acatlensis*) present a huge expansion of the IRs, combined with a strong reduction of the SSC (Data S5): genes *rps*15, *ndh*H, *ndh*A, *ndh*I, *ndh*G, *ndh*E, *psa*C and *ndh*D, located in the SSC in the Fabaceae family, are duplicated and located in the IRs in the tribe Ingeae. This phenomenon was previously reported for seven species of the tribe Ingeae: *Pararchidendron pruinosum*, *Samanea saman*, *Acacia dealbata*, *Albizia odoratissima*, *Archidendron lucyi*, *Faidherbia albida*, *Inga leiocalycina* and *Pithecellobium flexicaule* (Wang et al. 2017). According to Wang et al. (2017), this expansion could be linked to an extremely AT-rich region 100 bp upstream of the IR/SSC junction. We observed a similar pattern for the five Ingeae species analysed here, with a mean of 93% of AT in the 100 bp upstream of the IR/SSC junction.

A Maximum Likelihood phylogenetic tree was performed with five species of the tribe Ingeae and six species of another tribe of the Fabaceae family (Table 1; Figure 3). The phylogenetic tree obtained was well supported and fully resolved. As expected, the five Ingeae species clustered together with high confidence and are well separated from species of other tribes (Acacieae and Mimoseae). A similar result was previously observed using a combination of plastid and nuclear markers (Ferm et al. 2021).

4 | Conclusion

Despite its importance in the wood industry, little genetic resource is available for the species *Cedrelinga cateniformis*. Here we used Oxford Nanopore long-read technology to reconstruct the first complete chloroplast sequence of the *Cedrelinga* genus. This will help monitor the species diversity and help set conservation and use plans.

Author Contributions

Nora Scarcelli: conceptualization (equal), formal analysis (equal), investigation (equal), methodology (equal), writing – original draft (lead), writing – review and editing (lead). **Cédric Mariac:** conceptualization (equal), investigation (equal), methodology (equal), validation (equal), writing – review and editing (equal). **Marie Couderc:** conceptualization (equal), investigation (supporting), methodology (equal), writing – review and editing (supporting). **Diana Castro Ruiz:** investigation (equal), writing – review and editing (supporting). **Guillain Estivals:** investigation (equal), writing – review and editing (supporting). **Carlos Alberto Custodio Angulo Chavez:** investigation (equal), writing – review and editing (supporting). **Hector Acho Vasquez:** investigation (equal), writing – review and editing (supporting). **Jhon Gregory Alvarado Reategui:** methodology (equal), writing – review and editing (supporting). **Tony Vizcarra Bentos:** methodology (equal),

20457758, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ece3.71355 by Cochrane Peru, Wiley Online Library on [14/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

writing – review and editing (supporting). **Carmen Garcia-Davila:** investigation (supporting), writing – review and editing (supporting).

Acknowledgements

The authors acknowledge the ISO 9001 certified IRD i-Trop HPC (South Green Platform https://bioinfo.ird.fr/) at IRD Montpellier for providing HPC resources.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI under the accession no. PQ868097. The associated BioProject, SRA and Bio-Sample numbers are PRJNA1208262, SRR32023740 and SAMN46171921 respectively.

References

Baluarte-Vásquez, J. R., and J. G. Alvarez-Gonzales. 2015. "Modelamiento del Crecimiento de Tornillo *Cedrelinga catenaeformis* Ducke en Plantaciones en Jenaro Herrera, Departamento de Loreto, Perú." *Folia Amazónica* 24, no. 1: 21–32. https://doi.org/10.24841/fa.v24i1.57.

Cruz, W., C. Saldaña, H. Ramos, R. Baselly, J. C. Loli, and E. Cuellar. 2020. "Genetic Structure of Natural Populations of Cedrelinga Cateniformis 'tornillo' From the Oriental Region of Peru." *Scientia Agropecuaria* 11, no. 4: 521–528. https://doi.org/10.17268/sci.agropecu.2020.04.07.

Delahaye, C., and J. Nicolas. 2021. "Sequencing DNA With Nanopores: Troubles and Biases." *PLoS One* 16, no. 10: e0257521.

FAO. 2018. "La Industria de la Madera en el Perú, Identificación de las Barreras y Oportunidades Para el Comercio Interno de Productos Responsables de Madera." Proveniente de Fuentes Sostenibles y Legales en las MYPES del Perú. 178.

Ferm, J., B. Ståhl, N. Wikström, and C. Rydin. 2021. "Phylogeny of the Ingoid Clade (Caesalpinioideae, Fabaceae), Based on Nuclear and Plastid Data." BioRxiv. https://doi.org/10.1101/2021.11.23.469677.

Gonçalez, J., and D. M. Gonçalves. 2001. "Valorization of Two Brazilian Timbers Cedrelinga Catenaeformis e Enterolobium SHOMBURGKII." *Revista Científica Do Laboratório de Produtos Florestais* 71: 69–74.

Greiner, S., P. Lehwark, and R. Bock. 2019. "OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes." *Nucleic Acids Research* 47: W59–W64.

Guariguata, M. R., J. Arce, T. Ammour, and J. L. Capella. 2017. "Las Plantaciones Forestales en Perú: Reflexiones, Estatus Actual y Perspectivas a Futuro." *Documento Ocasional* 169: 1–40.

Haag, V., G. Koch, E. Melcher, and J. Welling. 2020. "Characterization of the Wood Properties of *Cedrelinga cateniformis* as Substitute for Timbers Used for Window Manufacturing and Outdoor Applications." *Maderas. Ciencia y Tecnología* 22, no. 1: 23–36.

Katoh, K., and D. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780. https://doi.org/10.1093/molbev/mst010.

Kolmogorov, M., J. Yuan, Y. Lin, and P. Pevzner. 2019. "Assembly of Long Error-Prone Reads Using Repeat Graphs." *Nature Biotechnology* 37, no. 5: 540–546. https://doi.org/10.1038/s41587-019-0072-8.

Li, H. 2018. "Minimap2: Pairwise Alignment for Nucleotide Sequences." *Bioinformatics* 34, no. 18: 3094–3100.

Rojas Briceño, N. B., D. A. Cotrina Sánchez, E. Barboza Castillo, et al. 2020. "Current and Future Distribution of Five Timber Forest Species in Amazonas, Northeast Peru: Contributions Towards a Restoration Strategy." *Diversity* 12: 305.

Scarcelli, N., C. Garcia Davila, M. Couderc, et al. 2024. "The Complete Chloroplast Genome of Marupa (*Simarouba amara* Aubl., Simaroubaceae)." *Ecology and Evolution* 14, no. 7: e11688.

Stamatakis, A. 2014. "RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies." *Bioinformatics* 30, no. 9: 1312–1313.

Tillich, M., P. Lehwark, T. Pellizzer, et al. 2017. "GeSeq—Versatile and Accurate Annotation of Organelle Genomes." *Nucleic Acids Research* 45, no. W1: W6–W11.

Vaser, R., I. Sović, N. Nagarajan, and M. Šikić. 2017. "Fast and Accurate De Novo Genome Assembly From Long Uncorrected Reads." *Genome Research* 27, no. 5: 737–746.

Wang, Y. H., X. J. Qu, S. Y. Chen, D. Z. Li, and T. S. Yi. 2017. "Plastomes of Mimosoideae: Structural and Size Variation, Sequence Divergence, and Phylogenetic Implication." *Tree Genetics & Genomes* 13, no. 2: 41. https://doi.org/10.1007/s11295-017-1124-1.

Zhou, W., C. E. Armijos, C. Lee, et al. 2023. "Plastid Genome Assembly Using Long-Read Data." *Molecular Ecology Resources* 23, no. 6: 1442–1457.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

20457758, 2025, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/ece3.71355 by Cochrane Peru, Wiley Online Library on [14/05/2025]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License