

Wild *Coffea* Species: A Modern Genomic Approach to Unravel Variations for Future Cultivated Coffee Improvement [†]

Romain Guyot ^{1,*} , Laura Gonzalez ^{1,2}  and Rickarlos Bezandry ^{1,3} 

¹ Institut de Recherche pour le Développement, UMR DIADE, 911 Ave Agropolis, 34000 Montpellier, France; ln.gonzalez138@uniandes.edu.co (L.G.); richarlos@hotmail.fr (R.B.)

² Systems and Computing Engineering Department, Universidad de los Andes, Cra 1 N 18A, Bogotá 111711, Colombia

³ Faculté des Sciences, de Technologies et de l'Environnement, Campus Universitaire d'Ambondrona, Université de Mahajanga, BP 652, Mahajanga 401, Madagascar

* Correspondence: romain.guyot@ird.fr

[†] Presented at the International Coffee Convention 2024, Mannheim, Germany, 17–18 October 2024.

Abstract: The genetic diversity of wild *Coffea* species holds immense potential for the enhancement of cultivated coffee trees, offering solutions to challenges such as disease resistance, climate adaptability, and yield improvement. This study leverages modern genomic approaches to investigate evolution and variations among wild *Coffea* species. By employing advanced sequencing technologies and comparative analysis, the aim was to understand the molecular evolution of these wild species, identifying key genes and genetic markers which contribute to desirable traits. This research integrates comprehensive genomic data analysis with field studies to provide an understanding of the genetic diversity within wild *Coffea* populations. The aim of this research is to target significant genetic variations that can be harnessed to develop superior coffee cultivars, ensuring sustainability and resilience in the face of changing environmental conditions. This study not only highlights the importance of preserving wild *Coffea* species but also underscores the role of genomic research in unlocking their potential for coffee breeding programs. By bridging the gap between conservation and cultivation, this work paves the way for future innovations in coffee agriculture.

Keywords: wild *Coffea* species; genetic diversity; caffeine metabolism; disease resistance; drought tolerance; climate change adaptation; conservation strategies



Citation: Guyot, R.; Gonzalez, L.; Bezandry, R. Wild *Coffea* Species: A Modern Genomic Approach to Unravel Variations for Future Cultivated Coffee Improvement. *Proceedings* **2024**, *109*, 23. <https://doi.org/10.3390/ICC2024-18165>

Academic Editor: Dirk W. Lachenmeier

Published: 24 July 2024



Copyright: © 2024 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

By 2032, global temperatures are expected to increase by 1.5 °C at the current rate, and a 2 °C increase could be reached by 2057, as reported by Berkeley Earth in its Global temperature report of 2023 [1]. This increase will have significant impacts on the environment, such as increased risks of drought, loss of biodiversity, reduced water availability, and decreased crop yields [2]. Placed among economically important crops, coffee supports the livelihoods of millions of people. Climate variations affect coffee production by increasing biotic and abiotic stresses, resulting in the displacement of cultivation to more favorable areas, particularly at higher altitudes. Currently, two *Coffea* species dominate the market: The first is *C. arabica* L. (or Arabica), an allotetraploid species cultivated for several hundred years, representing 60% of the total coffee production. Originating from Ethiopia and, to a lesser extent, South Sudan, *C. arabica* is a high-altitude plant adapted to temperatures of 18–23 °C, known for its sensitivity to climate variations [3]. For instance, Arabica yields in Tanzania decrease by 137 kg per hectare for every degree increase in night-time temperatures due to climate change [4]. The second is *C. canephora* Pierre ex A. Froehner (or Robusta), a diploid species which has been cultivated since the 1850s, representing about 40% of the market. *C. canephora* shows greater tolerance compared to Arabica, occupying ecological niches with temperatures ranging from 22 °C to 30 °C. However, it has been

established that the optimal growth temperature for Robusta is 20 °C, and temperature variations of 1 °C below or above a range of 16–24 °C result in a significant production loss (14%) [5].

With global changes, a reduction of nearly 50% in the cultivated coffee area is projected by 2050 [3]. Four of the five main coffee producers in the world (Brazil, Vietnam, Colombia, and Indonesia) are expected to see a profound decline in the size and quality of their best coffee-growing areas. For instance, projections indicate a reduction of about 20 to 60% in the areas currently suitable for coffee cultivation in southeastern Brazil [6]. In 2022, the Brazilian state of Paraná experienced a 40% drop in production due to extreme climatic variations [7].

2. The Diversity of Wild *Coffea* Species

The genus *Coffea* currently comprises 130 recognized species, but this number could increase up to 141, considering the undescribed species in various collections. Most of these species are allogamous, originating from Africa, Madagascar, the Comoros, the Mascarene Islands, and Australasia, and are classified into major phylogeographic groups [8,9] (Eucoffea, Mozambicoffea, Baracoffea, Mascarocoffea, Mascarenes islands, Ex Psilanthus Africa, and Ex Psilanthus Asia; Figure 1). A dozen of them have been or were consumed in the past, then abandoned for various reasons. Examples of these include *C. liberica*, *C. dewevrei*, *C. stenophylla* [10], *C. mauritiana*, *C. congensis*, *C. racemosa*, *C. zanguebariae*, *C. bengalensis*, *C. travancorensis*, *C. wightiana*, *C. eugenioides*, and *C. humblotiana*.

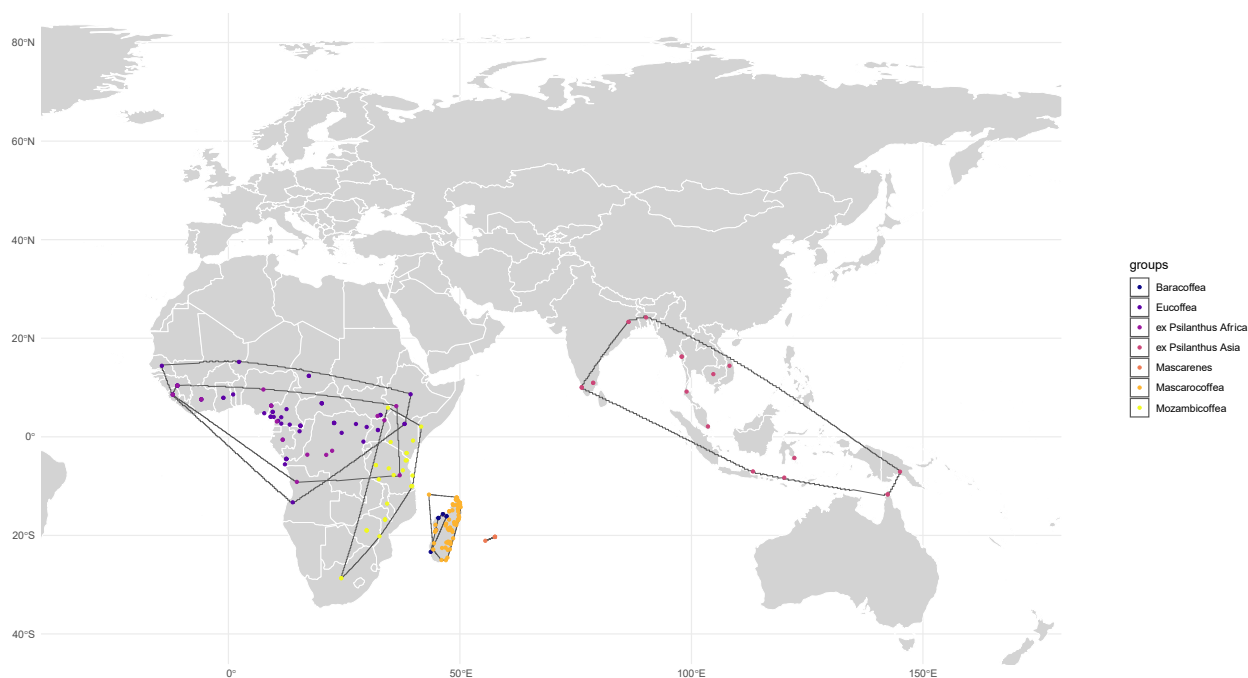


Figure 1. Distribution of wild *Coffea* species. The following phylogenetic groups are represented by colored areas: Baracoffea, Eucoffea, ex Psilanthus genus from Africa, ex Psilanthus genus from Asia, Mascarenes islands, Mascarocoffea, and Mozambicoffea.

While the organoleptic quality of some wild species is low or unknown, they exhibit exceptional environmental adaptation characteristics. For example, they show resistance to coffee leaf rust (*Hemileia vastatrix*) in *C. liberica*, leaf miner (*Perileucoptera coffeella*) in *C. racemosa*, drought tolerance in Baracoffea, *C. stenophylla*, and *C. racemosa*, highly variable fruit maturation times (from less than 4 months to over 14 months) [11], and agronomic and biochemical traits of interest such as the absence of caffeine in the seeds [8]. However, there are major threats to these wild species, and at least 60% of them are endangered due to anthropogenic activities reducing their habitat. This includes species with potential interest for the future improvement of cultivated coffee trees in response to climate change.

About 60% of species are conserved in ex situ collections such as in Divo (Centre National de Recherche Agronomique (CNRA), Ivory Coast), La Réunion (Centre de Ressources Biologiques (CRB) Coffea), Kianjavato (National Centre of Applied Research and Rural Development (FOFIFA), Madagascar), Londrina (Instituto Agronômico do Paraná (IAPAR), Brazil), and Campinas (Instituto Agronômico de Campinas (IAC), Brazil).

3. Characteristics of Interest of Wild *Coffea* Species

Wild *Coffea* species exhibit remarkable adaptive characteristics. They can occupy contrasting ecological niches. The best example to date features species from Madagascar, which represent half of the known species in the genus. A total of 66 species are native to the Comoros and Madagascar and occupy a wide variety of environments and climates, showing enormous phenotypic diversity. Based on phylogenetic analyses, a recent diversification (~8 Myr) was proposed for these species [8]. The most environmentally adapted group is undoubtedly the “Baracoffea” alliance, a group of nine species occupying the sandy soils of the dry deciduous forests of western Madagascar [12], growing in a hot and dry climate. Morphologically, these species share unique characteristics compared to other coffee trees, such as terminal inflorescences, the appearance of a quaternary axis (short shoot), a mixed vegetative axis (monopodial growth “trunk, branches twigs” and sympodial growth “short shoot”), rhythmic growth, deciduous leaves, and short fruiting times [13] (Figure 2). The traits responsible for their tolerance to much drier and hotter climates compared to their sister clade need to be studied and elucidated to understand this adaptation. This research could potentially inspire breeding efforts to improve the drought tolerance of cultivated coffee varieties. Unfortunately, most of the coffee species from Madagascar are classified by the International Union for Conservation of Nature (IUCN) [14] as Endangered or Critically Endangered. Thus, both in situ and ex situ conservation efforts are necessary to protect these species.



Figure 2. *Coffea ambongensis*, a Baracoffea species from the Northeast Madagascar environment. Picture by Rickarlos Bezandry.

4. Genomic Research and Genome Sequencing

To understand genome dynamics and the relationship between genetic diversity and adaptation, genomics has become a powerful tool thanks to the possibility of long-read

sequencing solutions and powerful bioinformatics tools, enabling the high-quality assembly of reference genomes as well as the structural and functional annotation of genes and mobile elements. However, to truly understand genome dynamics, it is essential to capture the diversity of a group of individuals from a species or multiple species, selected precisely based on their phenotype or adaptation. Comparing these individuals at the genomic level will provide deeper insights into their adaptive traits and evolutionary processes.

Integrative genomic approaches such as synteny, collinearity, and pangenome (i.e., the complete set of genes present in all strains of a species) have been applied to many agro-nomically important species. These studies have demonstrated the potential of diversity, particularly in the gene pools present in Crop Wild Relatives (CWRs), for the improvement of cultivated species. It is known that CWRs can mitigate the impact of climate change because their genetic composition confers greater tolerance to drought and other abiotic and biotic stresses.

Applied to the knowledge of plant genomes, comparative genomic approaches allow for testing hypotheses on biosynthetic pathways and conducting comparative approaches. Sequencing the genome of *Coffea canephora* has enabled the analysis of the caffeine biosynthetic pathway. Three methylation steps are necessary to produce caffeine from xanthosine, involving the same family of genes: *N*-methyltransferases (NMTs). To understand the absence of caffeine in Malagasy *Coffea* species, the genome of *C. humblotiana*, a wild and endangered, naturally caffeine-free species from the Comoros, was sequenced. A comparative analysis at the *NMT* gene location between the genomes of *C. canephora* and *C. humblotiana* revealed the absence of the caffeine synthase gene converting theobromine into caffeine, likely due to a deletion [15]. This absence appears to be the main cause of the absence of caffeine. This result is intriguing, as it suggests that *C. humblotiana* is derived from an ancestor which acquired the biosynthetic pathway and then lost it. This study perfectly illustrates the potential of genomics to elucidate traits and provide targets for species improvement.

5. The “Bridges *Coffea*” Project

Advances in sequencing techniques and genome assembly pipelines have enabled the development of large-scale projects aimed at generating high-quality reference genomes, contributing to a better understanding of species diversity and evolution. Establishing reference genomic resources is also crucial for biodiversity conservation issues, having propelled major groundbreaking initiatives such as the Earth Biogenome Project and its pan-European branch: the European Reference Genome Atlas (ERGA). Data from these initiatives and others have allowed pangenome-wide studies, enabling researchers to move from individual-level to species-level analysis of genomic diversity and identify key genes involved in adaptation to contrasting environments.

The pangenome is classically based on the conservation of orthologous genes present in all individuals constituting the “core genome” and those present in at least one individual representing the “dispensable genome”. Dispensable genes are often enriched in functions involved in responses to biotic and abiotic stresses and are, therefore, related to the adaptation of individuals or species to the environment. In addition to the presence-absence of genes, pangenomes can incorporate other sources of variation important in the onset of certain phenotypes and in the adaptation of species, such as structural variations opening new avenues of study. The success of developing pangenomes in a species or group of species depends on several important factors, such as the choice of individuals, the approach used for pangenome assembly, and the quality of genomes, variation detection, and annotations. Recently, major initiatives such as the pangenomics of wild grape, watermelon, tomatoes, and potatoes, aimed at sequencing known CWRs, began [16–19].

To identify traits of interest in wild coffee species better adapted to contrasting climates which could replace cultivated species or be used in improvement programs via interspecific hybridization, an international project was launched to study diversity at the genetic and genomic levels: the “Bridges_*Coffea*” project [20]. Indeed, current knowledge is insufficient to fully understand the evolution of genome structure, genetic content, the

impact of transposable elements, and structural variations in coffee trees due to a lack of genomic resources.

The goal of this project is to begin addressing this situation by leveraging the development of new sequencing technologies and state-of-the-art bioinformatic tools for genome assembly, comparative genomics, and pangenome integration. The aim is to sequence *Coffea* genomes in collections worldwide. These represent about 60% of the species known to date. This project is part of a broader program that has been conducted for several years among partners and concerns the knowledge of the structural genome diversity of the *Coffea* genus. The results of the project will have major implications for understanding the evolution of the structural characteristics of a tropical tree genus, the evolutionary pattern of biosynthetic pathways, and the association between the function of genes in the “shell” and “cloud” genomes and their role in environmental adaptation. The authors’ long-term ambition is to provide open “omic” resources and collaborative platforms to promote the use of wild coffee species in future crop improvement programs and their protection and conservation in their natural environment in Africa and Madagascar and in ex situ collections.

6. Conclusions

The results of the “Bridges_Coffea” project will have major implications for understanding the evolution of structural features of the *Coffea* genus, the evolutionary patterns of biosynthetic pathways, and the relationship between gene function and adaptation to environmental constraints. The long-term goal is to provide open genomic resources and collaborative platforms to promote the use of wild coffee species in future breeding programs and ensure their protection and conservation in their natural habitats in Africa and Madagascar and in ex situ collections.

Author Contributions: Conceptualization, R.G.; methodology, L.G. and R.B.; formal analysis, R.G. and R.B.; writing—original draft preparation, R.G.; writing—review and editing, R.G., L.G. and R.B.; and funding acquisition, R.G., L.G. and R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Agence National de la Recherche (ANR), with grant ANR-23-CE20-0047-01, The Agropolis Foundation grant n° 2202-204 and the Rufford foundation, with grant 39692-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge the Institut Français de Bioinformatique (IFB, funded by ANR, ANR-11-INBS-0013) for supporting the bioinformatic analyses and Maia Lejbowicz for her assistance in ensuring compliance with the Nagoya protocol.

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

References

1. Rohde, R. Global Temperature Report for 2023. In *Berkeley Earth*. 12 January 2024; Available online: <https://berkeleyearth.org/global-temperature-report-for-2023/> (accessed on 14 July 2024).
2. Raza, A.; Razzaq, A.; Mehmood, S.; Zou, X.; Zhang, X.; Lv, Y.; Xu, J. Impact of Climate Change on Crops Adaptation and Strategies to Tackle Its Outcome: A Review. *Plants* **2019**, *8*, 34. [CrossRef] [PubMed]
3. Bunn, C.; Läderach, P.; Ovalle Rivera, O.; Kirschke, D. A bitter cup: Climate change profile of global production of Arabica and Robusta coffee. *Clim. Chang.* **2015**, *129*, 89–101. [CrossRef]
4. Craparo, A.C.W.; Van Asten, P.J.A.; Läderach, P.; Jassogne, L.T.P.; Grab, S.W. *Coffea arabica* yields decline in Tanzania due to climate change: Global implications. *Agric. For. Meteorol.* **2015**, *207*, 1–10. [CrossRef]
5. Kath, J.; Byraredy, V.M.; Craparo, A.; Nguyen-Huy, T.; Mushtaq, S.; Cao, L.; Bossolasco, L. Not so robust: Robusta coffee production is highly sensitive to temperature. *Glob. Chang. Biol.* **2020**, *26*, 3677–3688. [CrossRef] [PubMed]

6. Tavares, P.S.; Giarolla, A.; Chou, S.C.; Silva, A.J.P.; Lyra, A.A. Climate change impact on the potential yield of Arabica coffee in southeast Brazil. *Reg. Environ. Chang.* **2018**, *18*, 873–883. [CrossRef]
7. Produção de café no Paraná deve ser 40% menor do que previsto, diz estimativa do Deral—Centro do Comércio de Café do Estado de Minas Gerais. 7 March 2022. Available online: <https://cccmg.com.br/producao-de-cafe-no-parana-deve-ser-40-menor-do-que-previsto-diz-estimativa-do-deral/> (accessed on 14 July 2024).
8. Hamon, P.; Grover, C.E.; Davis, A.P.; Rakotomalala, J.-J.; Raharimalala, N.E.; Albert, V.A.; Sreenath, H.L.; Stoffelen, P.; Mitchell, S.E.; Couturon, E.; et al. Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species. *Mol. Phylogenet. Evol.* **2017**, *109*, 351–361. [CrossRef] [PubMed]
9. Guyot, R.; Hamon, P.; Couturon, E.; Raharimalala, N.; Rakotomalala, J.-J.; Lakkanna, S.; Sabatier, S.; Affouard, A.; Bonnet, P. WCSdb: A database of wild *Coffea* species. *Database* **2020**, *2020*, baaa069. [CrossRef] [PubMed]
10. Davis, A.P.; Mieulet, D.; Moat, J.; Sarmu, D.; Haggard, J. Arabica-like flavour in a heat-tolerant wild coffee species. *Nat. Plants* **2021**, *7*, 413–418. [CrossRef] [PubMed]
11. Rimlinger, A.; Raharimalala, N.; Letort, V.; Rakotomalala, J.-J.; Crouzillat, D.; Guyot, R.; Hamon, P.; Sabatier, S. Phenotypic diversity assessment within a major ex situ collection of wild endemic coffees in Madagascar. *Ann. Bot.* **2020**, *126*, 849–863. [CrossRef] [PubMed]
12. Davis, A.P.; Rakotonasolo, F. A taxonomic revision of the baracoffea alliance: Nine remarkable *Coffea* species from western Madagascar. *Bot. J. Linn. Soc.* **2008**, *158*, 355–390. [CrossRef]
13. Bezandry, R.; Dupeyron, M.; Gonzalez-Garcia, L.N.; Anest, A.; Hamon, P.; Ranarijaona, H.L.T.; Vavitsara, M.E.; Sabatier, S.; Guyot, R. The evolutionary history of three Baracoffea species from western Madagascar revealed by chloroplast and nuclear genomes. *PLoS ONE* **2024**, *19*, e0296362. [CrossRef] [PubMed]
14. The IUCN Red List of Threatened Species. In *IUCN Red List of Threatened Species*. Available online: <https://www.iucnredlist.org/en> (accessed on 14 July 2024).
15. Raharimalala, N.; Rombauts, S.; McCarthy, A.; Garavito, A.; Orozco-Arias, S.; Bellanger, L.; Morales-Correa, A.Y.; Froger, S.; Michaux, S.; Berry, V.; et al. The absence of the caffeine synthase gene is involved in the naturally decaffeinated status of *Coffea humblotiana*, a wild species from Comoro archipelago. *Sci. Rep.* **2021**, *11*, 8119. [CrossRef] [PubMed]
16. Bozan, I.; Achakkagari, S.R.; Anglin, N.L.; Ellis, D.; Tai, H.H.; Strömvik, M.V. Pangenome analyses reveal impact of transposable elements and ploidy on the evolution of potato species. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2211117120. [CrossRef] [PubMed]
17. Cochetel, N.; Minio, A.; Guarracino, A.; Garcia, J.F.; Figueroa-Balderas, R.; Massonnet, M.; Kasuga, T.; Londo, J.P.; Garrison, E.; Gaut, B.S.; et al. A super-pangenome of the North American wild grape species. *Genome Biol.* **2023**, *24*, 290. [CrossRef] [PubMed]
18. Li, N.; He, Q.; Wang, J.; Wang, B.; Zhao, J.; Huang, S.; Yang, T.; Tang, Y.; Yang, S.; Aisimutuola, P.; et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat. Genet.* **2023**, *55*, 852–860. [CrossRef] [PubMed]
19. Wu, S.; Sun, H.; Gao, L.; Branham, S.; McGregor, C.; Renner, S.S.; Xu, Y.; Kousik, C.; Wechter, W.P.; Levi, A.; et al. A *Citrullus* genus super-pangenome reveals extensive variations in wild and cultivated watermelons and sheds light on watermelon evolution and domestication. *Plant Biotechnol. J.* **2023**, *21*, 1926–1928. [CrossRef] [PubMed]
20. BRIDGES_COFFEA. Available online: <https://www.bridgescoffea.org/> (accessed on 14 July 2024).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.