



Original article

WCSdb: a database of wild *Coffea* species

Romain Guyot^{1,5,*}, Perla Hamon¹, Emmanuel Couturon¹,
Nathalie Raharimalala², Jean-Jacques Rakotomalala²,
Sreenath Lakkanna³, Sylvie Sabatier⁴, Antoine Affouard⁶
and Pierre Bonnet⁴

¹Institut de Recherche pour le Développement, UMR DIADE, Université de Montpellier, 911 Avenue Agropolis, 34394 Montpellier, France, ²FOFIFA, BP 1444, Ambatobe, Antananarivo 101, Madagascar, ³Plant Biotechnology Division, Unit of Central Coffee Research Institute, Coffee Board, Manasagangothri, Mysore 570 006, India, ⁴AMAP, Univ Montpellier, CIRAD, CNRS, INRA, IRD Avenue Agropolis, 34398 Montpellier Cedex 5, France, ⁵Department of Electronics and Automatization, Universidad Autónoma de Manizales, Antigua Estacion del Ferrocarril, 170001 Manizales, Colombia and ⁶INRIA Sophia-Antipolis—ZENITH team, LIRMM—UMR 5506—CC 477, 161 rue Ada, 34095 Montpellier Cedex 5, France

*Corresponding author: Email: romain.guyot@ird.fr Tel: +33467416229; Fax: +33467416330

Citation details: Guyot,R., Hamon,P., Couturon,E., *et al.* WCSdb: a database of wild *Coffea* species *Database* (2020) Vol. 00: article ID baaa069; doi:10.1093/database/baaa069

Received 28 January 2020; Revised 22 June 2020; Accepted 30 July 2020

Abstract

Coffee is a beverage enjoyed by millions of people worldwide and an important commodity for millions of people. Beside the two cultivated species (*Coffea arabica* and *Coffea canephora*), the 139 wild coffee species/taxa belonging to the *Coffea* genus are largely unknown to coffee scientists and breeders although these species may be crucial for future coffee crop development to face climate changes. Here we present the Wild Coffee Species database (WCSdb) hosted by PI@ntNet platform (http://publish.plantnet-project.org/project/wildcofdb_en), providing information for 141 coffee species/taxa, for which 84 contain a photo gallery and 82 contain sequencing data (genotyping-by-sequencing, chloroplast or whole genome sequences). The objective of this database is to better understand and characterize the species (identification, morphology, biochemical compounds, genetic diversity and sequence data) in order to better protect and promote them.

Database URL: http://publish.plantnet-project.org/project/wildcofdb_en

Introduction

Coffee is a beverage enjoyed by millions of people worldwide. It is also an important commodity for millions of small coffee farmers living in tropical countries. Two species are mainly cultivated Arabica (*Coffea arabica*)

and Robusta (*Coffea canephora*). Coffee fields are deeply impacted by climate change and the emergence of diseases (1). Beside cultivated coffee trees, numerous wild coffee species are known to botanists, but largely ignored by agronomists and breeders although these species may be

crucial for future coffee crop development to face climate changes.

Based on morphological data, the phylogenetically closest genus called *Psilanthus* (2) has been recently placed into *Coffea* (3). The genus *Coffea* (broad sense since (3)) includes woody plants belonging to the Rubiaceae family. It comprises 124 species and 17 additional taxa, with a natural distribution covering tropical Africa, Madagascar, Comoros, Mauritius and the Reunion Islands extending to Southern and Southeast Asia and Australasia. However, the two genera mainly differ by the flower morphology with short corolla tube and reproductive organs exerted for *Coffea*, long corolla tube and reproductive organs inserted for *Psilanthus*. They differ also by their natural distribution. The genus *Psilanthus* is present on the African continent, Asia (India, Sri Lanka, tropical and Southeast Asia) and Oceania (Northern Australia) but absent from the islands of the West Indian Ocean (Madagascar, Mascarenes and Comoros). The merging of these two genera introduced a possible source of name confusion as numerous works are focused only on *Coffea sensu stricto*. Another source of species name confusion is the presence of numerous past or modern synonymies. For example, it is generally accepted that *Psilanthus minor* is in fact *P. sapinii* or that *Coffea vaughanii* is *C. myrtifolia*.

Anyway, these 141 identified and classified species/taxa are currently considered for further phylogenetic and molecular analyses (4).

Fourteen years ago, Davis and co-workers (5) revealed that numerous wild coffee species are vulnerable (23 species), endangered (30 species) or seriously threatened (19 species). A recent reassessment (1) confirmed that 60% of them are now threatened with extinction, suggesting a bad prospect for wild coffee species all over the tropical world. Wild coffee species living collections were initiated in the years 1960 in Africa and Madagascar. Today only 55% of them are in such collections: at the research station of the Centre National de Recherche en Agronomie, Divo, Côte d'Ivoire (for African species), at the Centre de Ressources Biologiques (CRB) Bassin-Martin, Reunion island (<http://florilege.arcad-project.org/fr/crb/coffee>; for African, Comorian and Mascarene species; Supplemental data 1) and at the research station of the National Centre for Applied Research for Rural Development (FOFIFA) in Kianjavato, Madagascar (for Madagascan species; Supplemental data 2). The analysis of wild coffee species conserved in living collection revealed large morphological variation such as flower morphology, size and color of fruits, plant height and leaf morphologies, days to fruit maturation and growth habitats and adaptation. In addition to morphology, large variations were observed in terms of seed biochemical compounds involved in the quality of coffee such as caffeine (6), trigonelline, sucrose and mangiferin contents into others (7–9). However, this

diversity is not comprehensively reported so far in any publication or any publicly available database.

The first *Coffea* genome has been published in 2014 and concerned *Coffea canephora* (10). Since this first release, several sequencing data have been published such as the genotyping-by-sequencing data to provide the first resolved phylogeny of the *Coffea* genus (11), the partial sequencing of 16 *Coffea* species (12) and the chloroplast reconstructions and nuclear SNP mining (13). Now the Genus is subjected to intensive genome sequencing (*C. arabica*, *C. eugenoides* and *C. canephora* (14); 82 wild coffee species, unpublished results; Supplemental data 3), allowing research of genes of agronomical interest and genome composition and evolution studies.

To intensify the protection and the conservation of wild coffee species and to promote their use in the search of agronomic genes of interest, more shared information is needed. So far, few databases are dedicated to wild coffee species. The Global Biodiversity Information Facility (GBIF, <https://www.gbif.org/species/2895315>) includes 8462 occurrences with images for 176 species. However, the reported samples include those collected in their natural area and also, those introduced by humans outside their natural distribution and those collected from *ex situ* living collections. The Reunion_Coffea database hosted in GBIF database was published by INRA Antilles-Guyane (15, <https://www.gbif.org/dataset/510b8030-6293-4bd9-812d-c195a9915a74>). This database provides the taxonomic distribution of occurrences (number of accessions per species), date of setting up in Reunion but does not include pictures or associated Supplementary data. Other databases are only focused on genomic data, such as the coffee Genome hub (<http://coffee-genome.org>; Robusta genome only), the TropGeneDB (<http://tropgenedb.cirad.fr/tropgene/JSP/interface.jsp?module=COFFEE>, mainly genetic markers) and the SOL Genomics Network <https://sgn.cornell.edu/search/organisms>).

In this study, we developed a wild coffee species database (WCSdb), hosted by Pl@ntNet (http://publish.plantnet-project.org/project/wildcofdb_en). The general objective of this database is to better understand the species (identification, morphology, biochemical compounds, genetic diversity and sequence data) in order to better protect and promote them. More specifically, the database presents: (i) each species held in collection on the sites of Bassin-Martin, Reunion Island and Kianjavato, Madagascar with a photo gallery of the tree morphology with a total of 597 images; (ii) different detailed information such as synonymy, natural distributions, habitats, architectural, morphological, phenological, biochemistry traits, genetic/genomic data, trait of interest retrieved from the literature and personal observations on living collection and (iii) a general geographical map of the species distribution.

Database construction and content

Data source

Pictures were mainly provided by Emmanuel Couturon for CRB (BassinMartin, La réunion) hosted plants (Supplemental data 1; list of species present at BassinMartin), by Eva N Raharimalala and Perla Hamon for Madagascan species (Supplemental data 2; list of species present at Kianjavato), and Lakkanna Sreenath for *Psilanthus. bengalensis*, *Psilanthus. travancorensis* and *Psilanthus. wightianus*. Links to molecular data were introduced such as genome size (Mbp 1C) from the plant DNA C-value database (<https://cvalues.science.kew.org>) for 57 species, genotyping-by-sequencing (GBS) data

for 77 species (11, <https://datadryad.org/resource/doi:10.5061/dryad.kk71t>), the sequenced and assembled chloroplast genomes for 39 species (13; Guyeux et al. unpublished results) and whole genome deep pair-end Illumina sequencing for 82 species (R. Guyot and Perla Hamon, unpublished results, data available upon request). genotyping-by-sequencing (GBS) and sequencing data information are available on the web site in a separate table and in Supplemental data 3).

Web interface, usage

Pl@ntNet Publish is an IT platform dedicated to the dissemination of botanical data focused on taxa or specimen

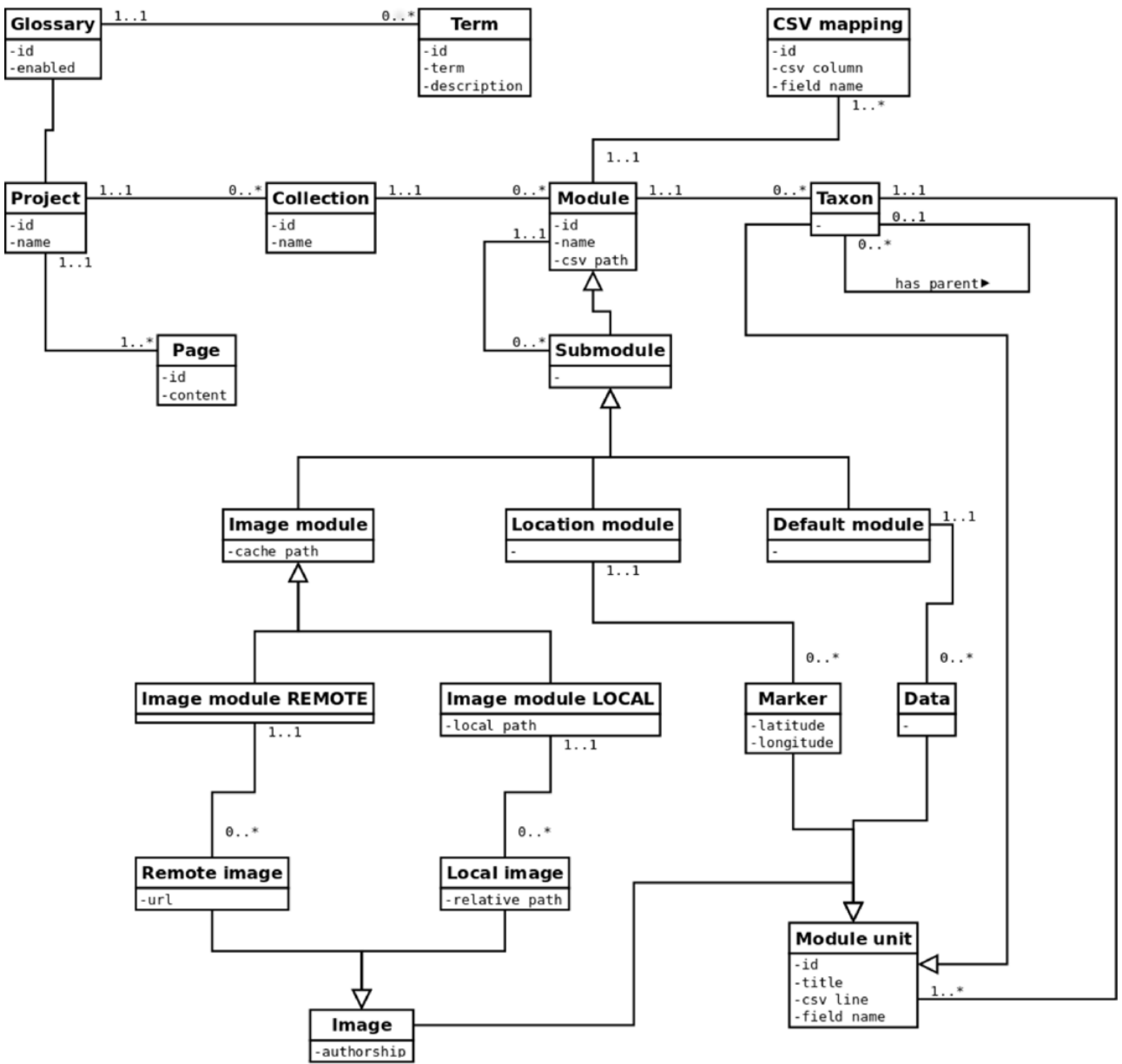


Figure 1. Schematic representation of the PI@ntNet Publish UML class diagram.

levels. It is based on Symfony (PHP) and MongoDB and allows its users to manage data publication spaces: descriptive texts, search forms, taxonomic data, visual data, geographic data, etc. (Figure 1). Data can be uploaded via comma-separated values (CSV) files and are then available through responsive Web pages dedicated to the portal created by users. Data exploration is also possible via a RESTful Web services (JSON). Pl@ntNet Publish was initially developed in 2014, with the support of Agropolis Fondation and has been adapted to various cases studies, such as Herbarium collections, Regional taxonomic checklists, information for Weed and Invasive species management, into others. Pl@ntNet Publish was deposited in 2015 at the Agency for the Protection of Programs (IDDN.FR.001.320007.OOO.R.C.2015.000.31235).

The main tab or species tab shows information for 124 accepted species and 17 taxa (Figure 2). Three species has duplicated entries, according to their distribution or variety: *C. mauritiana* for populations from Mauritius and Reunion islands, for *Coffea liberica* (variety *liberica* and variety *dewevrei*), and for *Psilanthus bengalensis* (variety *bababudanii* and variety *bengalensis*). In the species tab, we decide to keep the nomenclature *Coffea* and *Psilanthus* for the species to avoid any confusion. This database also shows species according to their botanical section or geographic origin: Eucoffea, Mascarocoffea, Mascarenes, Baracoffea (Madagascan species endemic to the western coast), *Psilanthus* from Africa and *Psilanthus* from Asia. The species table allows the user to select the species to examine. In total, for each species, 30 data fields were

Pl@ntNet Home Collections Search English

Wild Coffea species database en

The general objective of this database is to improve knowledge of wild coffee species in order to better protect and enhance them. More specifically, we:

- gather for each species, the available information from referenced or non-referenced literature and personal observations on their natural distributions, habitats, morphological, phenological, biochemical, genetic/genomic characteristics, etc.) in order to make the published data accessible.
- present specifically the wild species in living collection on the sites of Bassin-Martin, Reunion Island and Kianjavato, Madagascar (photo gallery)

Collections

Bassin-Martin collection (Biological Resources Center of Bassin-Martin)

Kianjavato collection

Details

Coffee species

C. kianjavatensis J.-F. Leroy Mascarocoffea Madag...

Coffee trees (*Coffea* genus) are woody plants belonging to the Rubiaceae family. Within this family and in particular the Coffeae tribe, the situation of the genus *Psilanthus* appears problematic because during long time it was considered as "false coffee trees". As such, they are nearly absent from the world's main living collections. *Psilanthus*, present on the African continent, Asia (India, Sri Lanka, tropical and South-East Asia) and Oceania (Northern Australia) but absent from the islands of the West Indian Ocean (Madagascar, Mascarenes and Comoros), constitute the closest phylogenetically to *Coffea* (Robbrecht and Manen, 2006). Its status established by Hooker (1873), confirmed by Chevalier (1947) and Leroy (1981), accepted by Stoffelen (1998), Khumar and Sreenath (2008) has been refuted by Davis et al (2011). Thus, according to these latter, coffee trees are represented by the 103 species of the genus *Coffea* (including the group of Baracoffea, endemic to the west coast of Madagascar) and the 21 *Psilanthus* species (see distribution map); the new genus being called *Coffea* (possible source of confusion).

Details

Bassin-Martin
6 5

Kianjavato
5 5

species
143 557 197

Figure 2. Home page of the wild coffee species database (http://publish.plantnet-project.org/project/wildcofdb_en).

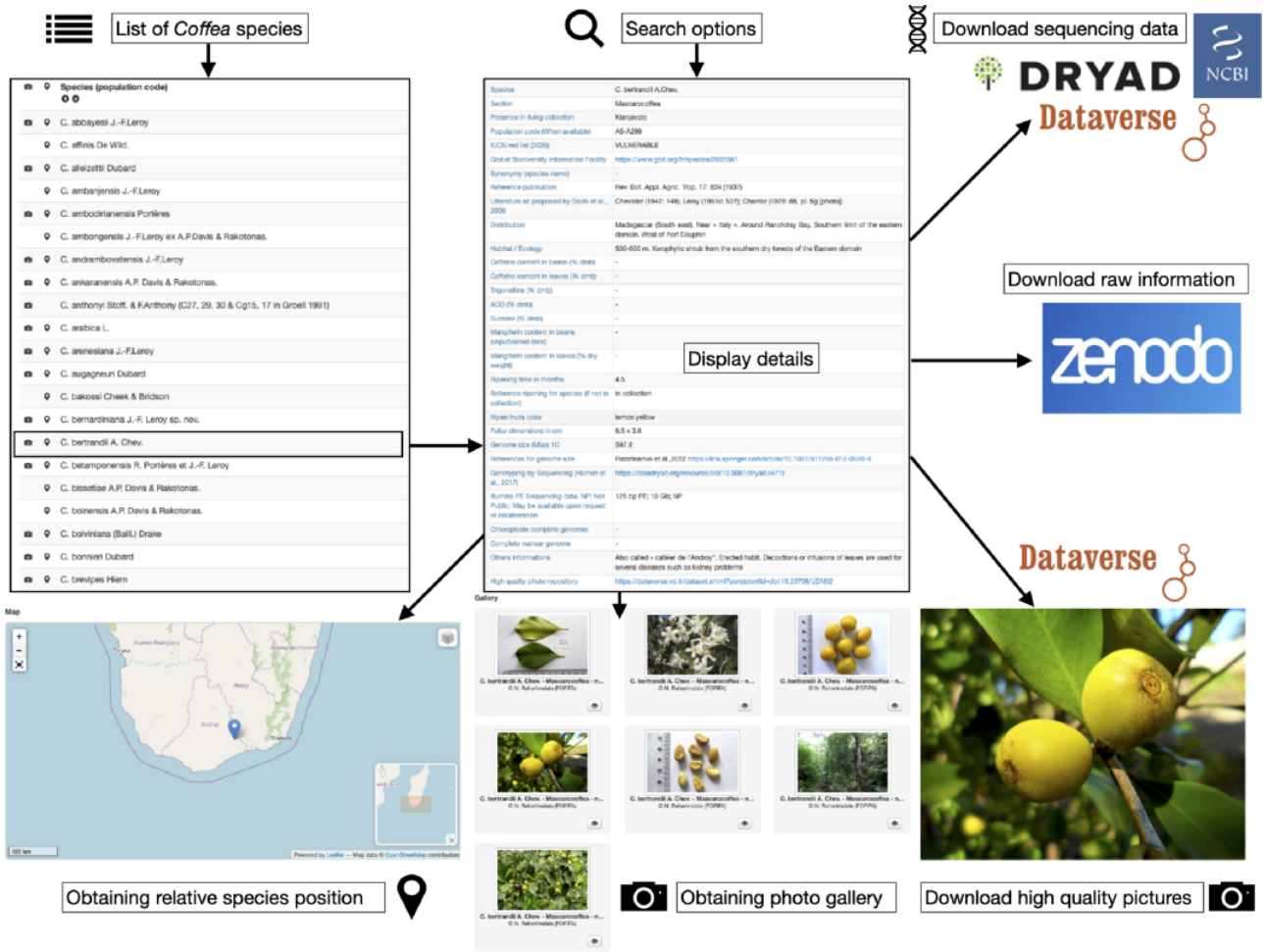


Figure 3. Flowchart showing the database use and outputs.

completed into a detailed table such as species name, section, living collection, synonymy and reference, distribution, habitat, caffeine content in bean and leaves, sucrose, mangiferin, ripening time and color of mature fruits, foliar dimension, genome sizes, sequencing data, and plastid genomes into others (Figure 3). The fields are described in the web site (section 'How to use this site'). The detailed table contains also references and the links to access to the molecular data when available (GBS data, nuclear genome sequences, plastid genomes and raw sequencing data).

In total, 201 localization entries were linked with coffee species, representative approximate area where the species were found when collected in the wild (mainly between years 1960 and 1980). With the detailed table, a photo gallery is available for 84 species. The photographs depict the morphology of beans, leaves, mature fruits, flowers, and the overall tree, allowing a better identification. In total, 551 photographs are available.

Conclusions and prospects

The WCS database represents the first comprehensive information about wild coffees species, largely unknown to coffee scientists and breeders. The information collected from the literature and from living collections at Kianjavato, Madagascar and Bassin-martin, La Réunion, may help researchers working in the preservation of coffee species, geneticists and breeders working with trait or genes of interest and improvement of cultivated species (tolerance to drought, resistances to diseases and pest, increase the quality of beans) or breeders motivated to re-cultivate forgotten species adapted to climate changes or adapted to specific habitats. More information will be integrated into the WCS database in the future, such as the availability of raw genomic sequences and assembled genome sequences, new characterized species, and new photo gallery.

Supplementary Data

Supplementary data are available at Database Online.

Funding

We thank the respective institutions IRD, FOFIFA, and CIRAD for support and funding.

References

1. Davis, A.P., Chadburn, H., Moat, J. *et al.* (2019) High extinction risk for wild coffee species and implications for coffee sector sustainability. *Sci Adv.*, 5, eaav 3473.
2. Robbrecht, E. and Manen, J.-F. (2006) The major evolutionary lineages of the coffee family (Rubiaceae, angiosperms). Combined analysis (nDNA and cpDNA) to infer the position of *Coptosapelta* and *Luculia*, and supertree construction based on *rbcL*, *rps16*, *trnL-trnF* and *atpB-rbcL* data. A new classification in two subfamilies, *Cinchonoideae* and *Rubioideae*. *Syst. Geography Plants*, 76, 85–146.
3. Davis, A.P., Tosh, J., Ruch, N. *et al.* (2011) Growing coffee: *Psilanthus* (Rubiaceae) subsumed on the basis of molecular and morphological data; implications for the size, morphology, distribution and evolutionary history of *Coffea*. *Bot. J. Linn. Soc.*, 167, 357–377.
4. Couturon, E., Raharimalala, N.E., Rakotomalala, J.J. *et al.* (2016) *Wild Coffee-trees: A Threatened Treasure in the Heart of Tropical Forests !* Association Biodiversité, Ecovalorisation et Caféiers ed.
5. Davis, A.P., Govaerts, R., Bridson, D.M. *et al.* (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot. J. Linn. Soc.*, 152, 465–512.
6. Hamon, P., Rakotomalala, J.J., Akaffou, S. *et al.* (2015) Caffeine-free species in the genus *Coffea*. In: Preedy V. (editors) *Coffee in Health and disease prevention*. San Diego: Academic Press, 39–44.
7. Campa, C., Ballester, J.F., Doulbeau, S. *et al.* (2004) Trigonelline and sucrose diversity in wild *Coffea* species. *Food Chem.*, 88, 39–43.
8. Campa, C., Doulbeau, S., Dussert, S. *et al.* (2005) Qualitative relationship between caffeine and chlorogenic acid contents among wild *Coffea* species. *Food Chem.*, 93, 135–139.
9. Campa, C., Mondolot, L., Rakotondravao, A. *et al.* (2012) A survey of mangiferin and hydroxycinnamic acid ester accumulation in coffee (*Coffea*) leaves: biological implications and uses. *Ann. Bot.*, 110, 595–613.
10. Denoeud, F., Carretero-Paulet, L., Dereeper, A. *et al.* (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science*, 345, 1181–1184.
11. Hamon, P., Grover, C.E., Davis, A.P. *et al.* (2017) Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (*Coffea*) and insights into the evolution of caffeine content in its species: GBS coffee phylogeny and the evolution of caffeine content. *Mol. Phylogenet. Evol.*, 109, 351–361.
12. Guyot, R., Darré, T., Dupeyron, M. *et al.* (2016) Partial sequencing reveals the transposable element composition of *Coffea* genomes and provides evidence for distinct evolutionary stories. *Mol. Genet. Genomics*, 291, 1979–1990.
13. Guyeux, C., Charr, J.C., Tran, H.T.M. *et al.* (2019) Evaluation of chloroplast genome annotation tools and application to analysis of the evolution of coffee species. *PLoS One*, 14, e0216347.
14. Mueller, L., Strickler, S., Domingues, D. *et al.* (2015) Towards a better understanding of the *Coffea arabica* genome structure. In: *Proceedings of the 25th International Conference on Coffee Science*. ASIC, pp 42–45. ISBN: 978-2-900212-24-0
15. Couturon, E. and Dussert, S. (2014) *Reunion_Coffea*. INRA Antilles-Guyane. Occurrence dataset. <https://doi.org/10.15468/jdrsoe> accessed via GBIF.org 4 June 2019, date last accessed.