A molecular passport for the rosewoods of Madagascar

Simon Crameri[1,4], Edgard Espinoza[2], Tahiana Ramananantoandro[3], Porter P. Lowry II[4], Alex Widmer[1]
*[1]Institute of Integrative Biology, ETH Zürich, CH-8092 Zürich, [2]National Forensics Laboratory, US Fish &
Wildlife Service, Ashland (OR), USA, [3]Ecole Supérieure des Sciences Agronomiques (ESSA), Antananarivo,
Madagascar, [4]Missouri Botanical Garden, St. Louis (MO), USA*

Key words
Rosewood, *Dalbergia*, forensic timber identification, NIRS, DART-ToF MS, Single nucleotide polymorphisms

Aim of the study
This study was performed to develop molecular fingerprints of *Dalbergia* species from Madagascar (tropical hardwoods known as rosewoods and subjected to unsustainable and often illegal logging), based on the analysis of (i) wood chemical composition, (ii) wood physical properties, and (iii) DNA variation; (iv) build reference databases containing the results of these methods, and (v) develop a software application that queries these databases for rapid and accurate identification of Malagasy *Dalbergia* species of potential economic interest.

Material and methods
There are c. 73 currently recognized species (55 currently published, 18 submitted or in prep.), of which at least 58 grow to exploitable dimensions (20 m height and/or 20 cm trunk diameter). We applied **direct analysis in real time, time-of-flight mass spectrometry (DART-ToF-MS)** to heartwood samples (1101 individuals, 55 species) of Malagasy *Dalbergia*. We also tested **near-infrared spectroscopy (NIRS)** using a Bruker benchtop device (779 individuals, 57 species) and a handheld micro-NIRS (399 individuals, 46 species) to an overlapping set of individuals. We subjected subsets of all datasets (taxa represented by ≥5 samples) to eight **machine learning** techniques (FDA, KNN, LDA, MLR, NNET, PLS-DA, RF and SVM). Training sets comprising 67% of the data and 10-fold cross-validation were used for hyperparameter tuning. A test set with the remaining 33% including all modelled taxa was used to estimate the resulting model's prediction success. We developed an **R shiny web application** containing reference databases and prediction models for NIRS and DART data, respectively. We further developed 7 **PCR primer pairs** for amplification and Sanger DNA sequencing in laboratories with basic equipment. PCR primers were designed for discrimination of morphologically similar and closely related species of the Malagasy *Dalbergia* sections Chlorocarpa and Maritima, respectively.

Results and significance
The 7 **PCR primers** were successfully amplified in a high fraction of samples and will facilitate multi-locus genetic analyses in Madagascar. **DART-ToF-MS** was identified as the most accurate tested method to identify heartwood samples, achieving 85% correct predictions on test samples. **Micro-NIRS** was found to be of limited use for species identification (49% correct predictions on test samples). The classification performance could not be significantly improved using **Bruker NIRS** data, despite a much higher spectral range (800–2300 nm vs. 950–1350 nm) and resolution (2200 vs. 113 distinct wavelength absorbances) compared to micro-NIRS. The best-ranking **machine learning techniques** were Kernel Partial Least Squares Discriminant Analysis (PLS-DA) for DART-ToF-MS data, and Linear Discriminant Analysis (LDA) for NIRS data, respectively. The **R shiny web application** (https://mbgmada.shinyapps.io/rosewoodid/) allows for user-friendly and interactive model performance assessment, classification (taxon prediction), visualization, and outlier analysis of available and novel NIRS and DART-ToF-MS data, and follows a modular design that allows for integration of updated classification models, and other types of data such as SNPs.

Related publications

Crameri, S.; Fior, S.; Zoller, S.; Widmer, A. (2022) A target capture approach for phylogenomic analyses at multiple evolutionary timescales in rosewoods (*Dalbergia* spp.) and the legume family (Fabaceae). Molecular Ecology Resources 22(8):3087–3105. https://doi.org/10.1111/1755-0998.13666

Crameri, S.; Phillipson, P. B.; Rakotonirina, N.; Wilding, N.; Andriamiarisoa, R. L.; Lowry II, P. P.; Widmer, A. (2022) Taxonomic Studies on Malagasy *Dalbergia* (Fabaceae). III. Two New Species from Southeastern Madagascar and an Emended Description of the Rosewood Species *Dalbergia maritima*. Systematic Botany 47(2): 397–416. https://doi.org/10.1600/036364422X16512564801614

Rakotonirina, N.; Phillipson, P. B.; Crameri, S.; Wilding, N.; Lowry II, P. P.; Rakouth, B., Razakamalala, R. (in review). Taxonomic studies on Malagasy *Dalbergia* L. f. (Fabaceae). IV. A new species from central and southern Madagascar, and a narrowed circumscription for *D. emirnensis* Benth. Novon.

Ramanantsialonina, R. N.; Crameri, S.; Sandratriniaina, N. A.; Wiemann, M. C.; Hermanson, J. C.; Rakouth, B.; Ravaomanalina, B. H. (2022) Comparative wood anatomy of 16 Malagasy *Dalbergia* species (Fabaceae) using multivariate techniques. IAWA J 0(0): 1–28. https://doi.org/10.1163/22941932-bja10105

Wilding, N.; Phillipson, P. B.; Crameri, S.; Andriambololonera, S.; Andriamiarisoa, R. L.; Andrianarivelo, S. A. F.; Bernard, R.; Rakotonirina, N.; Rakotovao, C.; Randrianaivo, R. I.; Razakamalala, R.; Lowry II, P. P. (2021a). Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). I: Two new species from northern Madagascar, and an emended description for *D. manongarivensis*. Candollea 76: 237–249. https://www.research-collection.ethz.ch/handle/20.500.11850/496970

Wilding, N.; Phillipson, P. B.; Crameri, S. (2021b) Taxonomic studies on Malagasy *Dalbergia* (Fabaceae). II: A new name for *D. mollis* and the reinstatement of *D. chermezonii*. Candollea 76: 251–257. https://www.research-collection.ethz.ch/handle/20.500.11850/496971

Wilding, N.; Phillipson, P.B.; Crameri, S. (submitted) Taxonomic studies on Malagasy *Dalbergia* L.f. (Fabaceae) V. Eight new large tree species and notes on related Malagasy species. Candollea.

Project 3.20.01

Project duration February 2021 – November 2022