Tropical tree genomics:
domestication of the legume tree *Inga* in tropical South America

Nom et contact des encadrants:

Rosalía Piñeiro (INRAE, UMR Biogeco), rosalia.pineiro-portela@inrae.fr

Myriam Heuertz (INRAE, UMR Biogeco), myriam.heuertz@inrae.fr

Laboratoire d'accueil: UMR Biogeco

Lieu du stage: INRAE, UMR Biogeco, 69 route d'Arcachon, 33600 Cestas

Contexte scientifique et objectifs:

This Master subject, aims to conduct research in line with the labex CEBA, http://www.labex-ceba.fr/ on data generated within the project "Improving agroforestry and silvopastoral systems in Latin America by maximising species and genetic diversity of the multipurpose legume Inga" BBSRC-GCRF BB/P022898/1, http://blogs.exeter.ac.uk/ingasystems/the-team/ (PI: Toby Pennington, University of Exeter).

The species-rich legume tree genus *Inga* is widely distributed and common throughout tropical America. *Inga* has an important history of utilization by local communities in the Amazon. Some species, notably *I. edulis* (pacae) and *I. feuillei* (ice-cream bean) have been cultivated since ancient times for their edible sweet-tasting fruits. This Master project aims to unlock the domestication history of *I. feuillei*. Using genome-wide data (Hybrid capture) across the entire distribution of this cultivated species and three wild congeners, we intend to identify its closest wild relatives and locate its areas of domestication in tropical South America.

Inga feuillei, known as pacae, grows only in cultivation on the pacific slopes of South America from Colombia to Chile. It is cultivated in the deserts of coastal Peru, where the large fruits are sold in local markets. The edible part of the fruit is the sweet white pulp that surrounds the seeds within the pods. This crop has been important since pre-Columbian times, as suggested by archaeological remains found in burial sites of north-western Peru between 2700 and 295 years old. Pacae leaves were used to fill mummies' bundles, while pods were placed at the base of mummies. Thanks to the favourable dry conditions, these remains are remarkably well preserved. Prehistoric ceramics with the shape of the fruit, notably Mochica (100-800 AD) and Chimu (1100-1470 AD) pottery also provide evidence of its importance. The fruit is represented as reproductions of natural size, in flat or cylindrical stirrup-spout bottles, or elongated drinking vessels. The first historical record is a reference of the Spanish historian Pedro Pizarro who states that in 1532 the Inca emperor Atahualpa sent a basket full of pacae to Francisco Pizarro, the conqueror of Peru. According to taxonomical treatments, the most likely origin of this species is within sect. Tetragona, which includes three species naturally occurring in the Amazon and in the Amazonian slopes of the Peruvian Andes (*I. striata*, *I. sapindoides*, and *I. macrophylla*).

Genome-wide data have been generated by the host team in order to investigate the origin of this tropical crop from natural tree populations of Amazonian rainforests. In addition to *I. striata*, two other closely related species, *I. sapindoides*, and *I.*

macrophylla will be studied. Sequences of specimens of *I. feuillei* and these three wild relatives will be analysed across their distributional ranges, with special focus on Peru, in order to understand when and where the domestication took place and if there was a single or several domestication events over the centuries. We will also estimate the date of domestication and evaluate whether it is congruent with the oldest archaeological remains around 2000 years old.

The student will gain experience in analysis of targeted capture data and in comparative phylogenetics and population genetics. A reference whole genome sequence of *Inga* is available. Software will include:

HYBpiper https://github.com/mossmatters/HybPiper

Phyluce https://github.com/faircloth-lab/phyluce

BWA https://github.com/lh3/bwa

bcftools https://github.com/samtools/bcftools

RAXML https://github.com/stamatak/standard-RAxML

ASTRAL-II https://github.com/smirarab/ASTRAL

BEAST https://github.com/beast-dev/beast-mcmc

SNAPP https://github.com/BEAST2-Dev/SNAPP

The main goals of this proposal will be subdivided into four scientific objectives:
1. Curation and analysis of genome-wide sequences (Hybrid capture) of collections of *I. feuillei* and its three Amazonian wild relatives *I. striata*, *I. macrophylla* and *I. sapindoides*, with special focus on Peru. Cutting-edge bioinformatics pipelines will be used.
2. Phylogenetic analyses and inference of the date of domestication of *I. feuillei* from its closest wild relative.
3. Assessment of spatial gradients of genetic diversity in order to identify ancestral populations of *I. feuillei*.
4. Depending on time constraints, selection of genes relevant for agriculture will be investigated. In particular, we expect that the domestication of *I. feuillei* implied adaptation to dry conditions, as all three wild relatives occur in rainforest environments.

The project will be supervised locally by M Heuertz and R Pineiro.

Informations additionnelles: For further information on the project and available genetic data, please send an email to the supervisors (we communicate fluently in French, too).

Montant de la gratification: Standard INRAE, ca. 600 EUR/month.

Exigences particulières: Interest in tropical rainforests, history of human settlements in the Amazon and genetic basis of plant domestication. The candidate will gain

experience in bioinformatic analysis, phylogenetics and population genetics of genome-wide DNA sequences.

Date limite de candidature, date de retour sur les candidatures: Applications will be examined until a suitable candidate is found.