Diversity and Evolution in tropical rainforest trees: example of Eperua falcata in French Guiana

Louise Brousseau

Collegium Sciences & Technologies
Ecole Doctorale RP2E

Thèse

présentée pour l'obtention du titre de

Docteur de l'Université de Lorraine

en Biologie Végétale et Forestière

par **Louise BROUSSEAU**

Diversité et Evolution des Arbres de Forêt Tropicale Humide: Exemple d'*Eperua falcata* en Guyane française.

Soutenance publique le 10 Décembre de l'an 2013

Membres du jury :

Rapporteurs :	Sophie KARRENBERG	Professeur associé, Université d'Uppsala, Suède
	Santiago GONZALEZ-MARTINEZ	Directeur de recherche, INIA Madrid, Spain
Examinateurs :	François LEFEVRE	Directeur de recherche, INRA Avignon, France
	Benoît PUJOL	Chargé de recherche, CNRS Toulouse, France
	Sylvain DELZON	Chargé de recherche, INRA Bordeaux, France
	Ivan SCOTTI (directeur de thèse)	Directeur de recherche, INRA Kourou, Guyane
	Erwin DREYER (directeur de thèse)	Directeur de recherche, INRA Nancy, France

INRA, UMR EEF (« Ecologie et Ecophysiologie Forestières », Champenoux, France) & INRA, UMR EcoFoG (« Ecologie des Forêts de Guyane », Kourou, Guyane française)

Foreword 4

Acknowledgements 6

SYNTHESIS 10

INTRODUCTION - Evolution in Amazonia 12

1. Short overview of the Amazonian rainforest 12

2. The building of biodiversity in Amazonia 15

3. The maintenance of diversity in Amazonia: a subtle combination of chance and determinism 22

4. Spatial heterogeneity in the Amazonian rainforest 24

5. Tree species model, research questions and study sites 32

PART I - Molecular evolution: population genetics and genomics 38

1. Population evolution 38

2. Population differentiation: the complex interplay between gene flow, selection, and drift. 50

3. Neutral differentiation 52

4. Adaptive differentiation 60

5. Next generation sequencing / genotyping and new opportunities 74

PART II - Phenotypic evolution and quantitative genetics 81

1. Causes of phenotypic variation 81

2. Phenotypic evolution in populations 91

3. Phenotypic differentiation 94

DISCUSSION 105

1. Neutralism and adaptation in *Eperua falcata* 105

2. Open questions and perspectives 111

3. Importance of assessing genetic diversity in a changing world 113

PhD RESULTS & DEVELOPMENT OF BIOINFORMATIC TOOLS 117

Article n°1 - Molecular divergence in tropical tree populations occupying environmental mosaics 119

Article n° 2 - Genome scan reveals fine-scale genetic structure and suggests highly local adaptation in a Neotropical tree species (*Eperua falcata*, Fabaceae) 147

Bioinformatic tools – 'Rngs': A suite of R functions to easily deal with next-generation (454-)sequencing data and post-process assembly and annotation results. 171

1. Introduction to bioinformatics 171

2. Short description of 'Rngs' 172

3. Short overview of the functions 173

4. Detailed description of the functions 173

Article n°3 - High-throughput transcriptome sequencing and polymorphism discovery in four Neotropical tree species. 179

Article n°4 - Highly local environmental variability promotes intra-population divergence of quantitative traits: an example from tropical rainforest trees 197

Article n°5 - Local adaptation in tropical rainforest trees: response of *E. falcata* (Fabaceae) seedling populations from contrasted habitats to drought and to water-logging 227

Preliminary results - Reciprocal transplants 244

BIBLIOGRAPHY 261

Foreword

This PhD was supported by a CJS INRA leading to a two years post-doctoral contract. My research work was carried out in two INRA mixt unit research "EcoFog" (Ecology of Forests of French Guiana) and "EEF" (Forest Ecology and Ecophysiology) and supported by MEDD ECOFOR 'Ecosystèmes tropicaux', PO-FEDER 'ENERGIRAVI' and 'CEBA' (Labex) research programs.

This manuscript was written as a global synthesis containing the state of the art, the main results and a general discussion. I tried to link ecology and population evolution in a global 'ecological genetics' approach. Particular theoretical notion and methods were included into 'green' and 'purple' boxes respectively. A synthetic summary of each main result was integrated in the synthesis through a topic sentence, a short description of the experiment and the main figures.

Complete methods and results are described in research articles, and given a part from the global synthesis. For more readability, figures and tables were included in the main text of research articles rather than giving them a part.

For more simplicity and to avoid redundancies, I merged the bibliographies of synthesis and research papers into a single bibliography.

Acknowledgements

Je remercie chaleureusement toutes les personnes qui m'ont aidée, accompagnée, soutenue et participé à l'aboutissement de cette thèse.

Je tiens cependant à addresser des remerciements tout particuliers ...

A mes directeurs de thèse, Ivan Scotti et Erwin Dreyer,

Qui ont su guider ma réflexion pendant ces trois années, apprivoiser mon énergie et apaiser mes craintes. Ivan, je te remercie de m'avoir transmis tes savoirs avec passion, et de m'avoir offert un enseignement d'une grande qualité scientifique. Je te remercie aussi de l'habileté avec laquelle tu as su m'octroyer une immense liberté, tout en sachant canaliser mes ambitions. Je te remercie Erwin de ton soutien pendant ces trois années de thèse, et de tes précieux conseils en écophysiologie. Je te remercie aussi d'avoir pris le temps de te rendre en Guyane chaque année pour suivre ma thèse et de m'avoir accompagnée chaque fois sur le terrain !

∞

A Damien Bonal,

Pour m'avoir gentiment confié ses données issues des expérimentations de « common garden » et « traitements hydriques » qui auront occupé une part importante de ma thèse et m'ont permis de croiser les approches moléculaires avec des approches écophysiologiques. Je te remercie de tout le soin apporté à répondre à mes questions, suivre avec attention les analyses de données et participer à la rédaction des deux articles associés. Je te remercie aussi de m'avoir accueillie dans ton bureau à Nancy et d'avoir activement participé au nettoyage des données de l'expérimentation de traitements hydriques, analyses et interprétations.

∞

A Caroline Scotti-Saintagne,

Pour avoir encadré mes expérimentations au laboratoire, et participé aux analyses des

AFLPs. Je te remercie aussi des excellents moments bio-info passés ensembles ! Merci aussi de ta douceur, et de ton calme : tu as été d'un grand soutien tout au long de ma thèse, car ta présence au labo suffit à calmer les esprits, à nous calmer tous et ça nous fait du bien !

∞

A Pauline Garnier-Géré,

Sans qui je n'aurais jamais osé ouvrir la boite noire de la génétique à haut débit ! Je te remercie d'avoir guidé l'analyse du transciptome de *l'Eperua falcata* et de m'avoir fait partager tes scripts et ta passion pour la génétique des pops et la génomique. Je te remercie fortement de l'esprit critique avec lequel tu as travaillé à l'article de 454, et de m'avoir enseigné la rigueur scientifique.

∞

A Marie-Pierre Etienne et Jean-Jacques Boreux,

Qui m'ont aidé à apprivoiser les bases de la modélisation bayésienne, à utiliser « OpenBUGS » et « JAGS ». Je vous remercie du précieux temps que vous avez accordé à répondre à mes nombreuses de questions et à vérifier mes scripts.

∞

A l'équipe d'Ecologie des populations d'EcoFoG,

Pour m'avoir accueillie dans leur labo pendant deux stages d'abord, puis pendant mes trois années de thèse, et avec qui j'ai passé d'excellents moments ! Je vous prie d'excuser avec un peu de retard le bazar que j'ai pu mettre au labo, mes oublis, mes maladresses qui ont coûté la vie à quelques verreries… Je remercie tout particulièrement Eliane Louisana et Valérie Troispoux, « colonne vertebrale » du laboratoire, qui assurent pour que l'on puisse travailler dans de bonnes conditions. Leurs minutieuses petites mains inspectent, rangent, réparent, remplacent pour nous permettre à tous de travailler dans la bonne humeur !

Merci aussi Eliane de tes précieux conseils de laboratoire (merci à ta double dé-protéinisation), et merci d'avoir toujours répondu présente dans les moments délicats, quand la centri nous lâche, ou quand les UV vacillent… Merci Saint-Omer de m'avoir accompagné sur le terrain, d'avoir sué avec moi à la mise en place des transplantations réciproques, et d'avoir encadré les étudiants FTH lors des campagnes de mesures. Merci de m'avoir fait partager ta passion de la nature et tes récits forestiers, malgré tous mes efforts pour me perdre, Maskilili ne m'a pas eue ! Merci Alexandra et François pour les bons moments que l'on a passé ensemble en bioinfo, de l'installation de notre poste linux à

l'annotation de nos transcriptomes, on s'est bien éclatés dans le bureau des thésards ! Je remercie aussi Stéphanie Barthe, et repense avec joie à notre séjour à Ottawa, au dix-huitième étage du « Westin » où nous avons passé une excellente semaine scientifique ! A tous les membres du laboratoire, j'ai adoré nettoyer le labo et dégivrer les congélos en votre compagnie ! Je remercie aussi Stéphane Traissac pour son eternelle bonne humeur, ses éclats de rires et coups de poing sur la table ! Merci aussi de m'avoir confié tes étudiants sur le terrain dans le cadre des 3 modules FTH auxquel j'ai eu l'honneur de participer. Je remercie aussi Bruno Ferry (pas vraiment EcoFoG mais néanmoins pillier fondateur et surtout mascotte presque officielle du module FTH) pour m'avoir aidée à l'analyse pédo de mes sites de transplantations ! Je remercie enfin Céline Leroy pour m'avoir prêtée son appareil photo, et avoir pris le temps de participer à l'analyse des photographies hémisphériques

∞

A l'UMR EcoFoG dans son ensemble,

Je vous remercie d'avoir accueilli avec autant de bienveillance mon travail de thèse. Je remercie la direction de l'UMR, Eric Marcon et Marc Gibernau, ainsi que le service d'appui (Josie, Catherine, Sylvie, Carole et Amycelda) pour leur dévouement.

∞

A mes amis et tous ceux qui m'ont bien fait marrer depuis mon arrivée en Guyane, il y a 5 ans déjà…

Je remercie tout d'abord Julien Engel pour son soutien quotidien et son eternel calme. Merci de nous avoir supportés, ma thèse et moi, d'avoir accepté mon côté geek, de m'avoir nourrie tous les jours de bons petits plats préparés avec soin, d'avoir été là ! Je remercie Marion Chartier pour avoir été ma plus fidèle amie Guyanaise : merci pour tes pas de danse (« chtouc, chtouc »), nos sorties endiablées, ton parapluie, et ta robe de touloulou (elle a certes un peu déteint, mais c'est une preuve que j'en ai fait le meilleur des usages : danser… et forcément transpirer !).

Je remercie Seb' pour tous ces bons moments passés depuis mon arrivée en Guyane, les tournées des grands ducs, de l'Yguana au Beau Bourg, du vieux monmartre au Clibberton, les soirées carnaval à n'en plus finir… Je remercie Damien et Isa de nous avoir accueillis à Cayenne les week-ends de désoevrement, et d'avoir animé Noëls et Nouvel ans !

Je remercie aussi les collègues du campus agronomique avec qui je n'ai pas travaillé mais avec qui je me suis bien amusée : merci Benjamin (Leudet) et Olivier, Fabien (Wagner) et

Julia, et merci Guillaume (Odonne) ! Je remercie aussi le staff « dentistes », à commencer par Jaouad et Jenna, Abdel et Dounia, Saïd, Mathieu, et tous les autres ! Merci aussi Gersende et Vincent pour les excellents moments passés ensembles ! Je remercie enfin le quartier du Vieux Bourg, l'Yguana, Roukmout et Gégé !

Mon départ de Guyane approchant, je voudrais vous dire à quel point vous allez me manquer, mais comme dirait Spartacus…

« Je reviendrai, et je serai des millions » !

SYNTHESIS

In this section, I explore the process of evolution in tropical rainforests. I perform a survey of the literature and review the building of diversity in Amazonia, as well as the structuring of genetic diversity in neotropical trees. I also present important notions of population and quantitative genetics and introduce the main methods to study population evolution. This section is composed of two parts, molecular and phenotypic evolution respectively. The main PhD results are summarized into orange boxes.

INTRODUCTION - Evolution in Amazonia

1. Short overview of the Amazonian rainforest

Climate:

The Amazon's climate is influenced by its tropical location: the temperatures are globally stable (27°C in average), precipitations are abundant (2000-4000mm/year), and relative air humidity is high (80 to 90% of saturation). The seasonality is influenced by the latitudinal movements of the inter-tropical convergence zone (the belt of low pressures where the northeast and southeast winds come together causing cumulonimbus) with two main seasons: the dry season (July to November in Guiana), and the rainy season (November to July interspersed by a short dry season during March in Guiana). The intensity of the dry season in Amazonia varies across years, and cyclic intense drought events (due to '*El nino* events') occurred every 4 to 8 years (**Figure 1**).

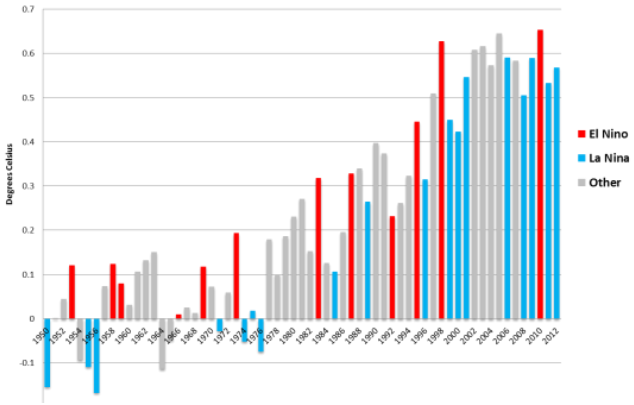


Figure 1 : Annual global temperature anomalies and El nino events (1950-2012). From http://www.ncdc.noaa.gov.

Carbon storage and biodiversity:

The tropical rainforest of Amazonia is one of the most important wilderness areas of the world (Cincotta *et al.* 2000, Anhuf *et al.* 2006, **Figure 2**). Spread over 7.3 million km², Amazonia is home for a luxurious tropical rainforest that covers 9 countries (Brazil, Peru, Bolivia, Colombia, Ecuador, Venezuela, Guyana, Suriname, and French Guiana plus parts

of Venezuela, Brazil and Columbia), **figure 3**. The Amazonian Basin is composed of 43 ecoregions (**figure 3**), in which the Guiana shield (French Guiana, Suriname and Guyana) corresponds to the 'Guianan moist forest' ecoregion.

Tropical rainforests constitute an important store of carbon (about 40% of the total carbon store in the terrestrial biomass, Anhuf *et al.* 2006, ter Seege *et al.* 2006). The total carbon stored by tropical rainforests is estimated to 247 PgC (with an annual net primary production of 17.8 PgC, Field *et al.* 1998, **Figure 4**). Moreover, the tropical rainforest of Amazonia is one of the world's greatest stores of biodiversity (ter Seege *et al.* 2006, Hoorn *et al.* 2010), including insects, mammals, amphibians, and plants (**Figure 5**). In particular, the tropical rainforest of Amazonia is home for ~50000 vascular plant species, among which ~12500 tree species (Hubbell *et al.* 2008). Moreover, in undisturbed forests, tree species diversity may easily reach 100 tree species /ha, as currently observed in French Guiana (**Figure 6**).

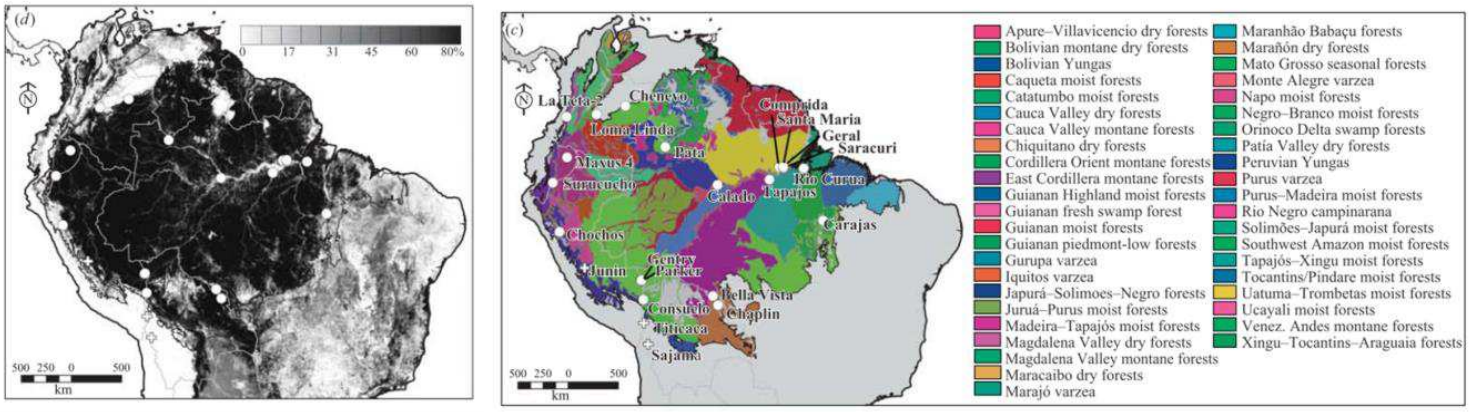


Figure 3 : Tree cover map (per cent) and ecoregions across the Amazon basin. From Mayle & Power 2008.

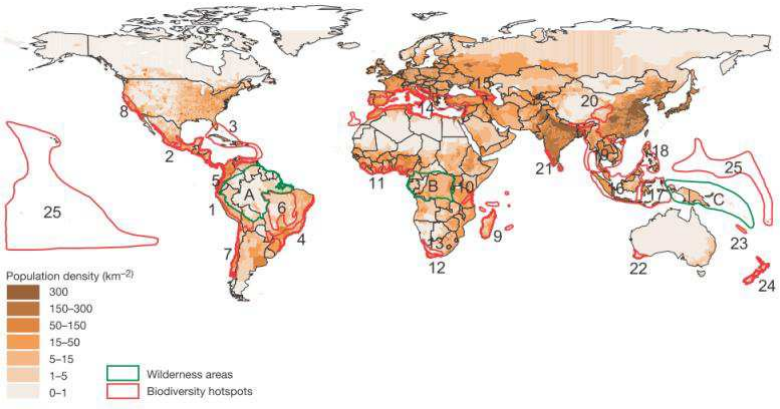


Figure 2 : Major Biodiversity hotspots (red) and wilderness areas (green). From Cincotta *et al.* 2000

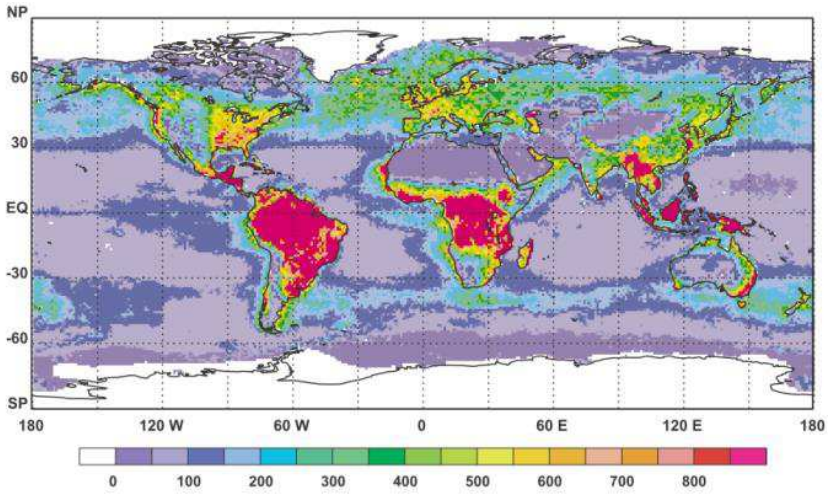


Figure 4: Net primary production (in grams of C per square meter per year). Form Field *et al.* 1998.

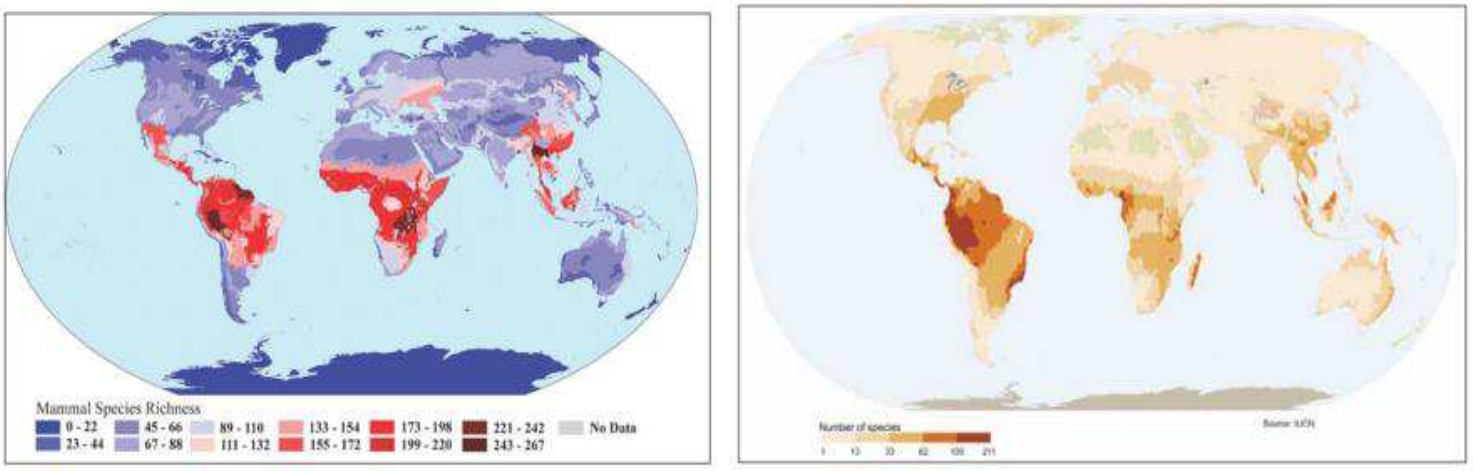


Figure 5: Global diversity of mammal (left) and amphibian (right) species (in number of species). From Olson *et al.* 2001 and the IUCN Red List 2009.

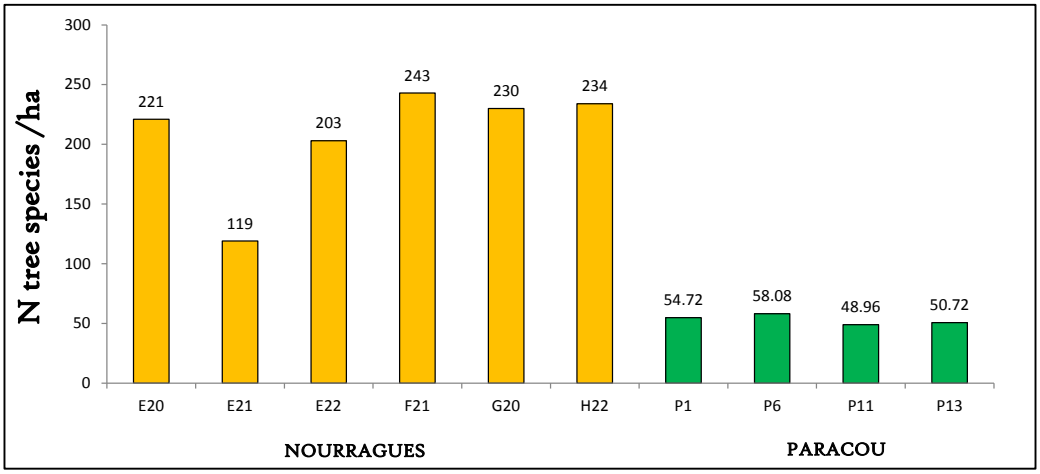
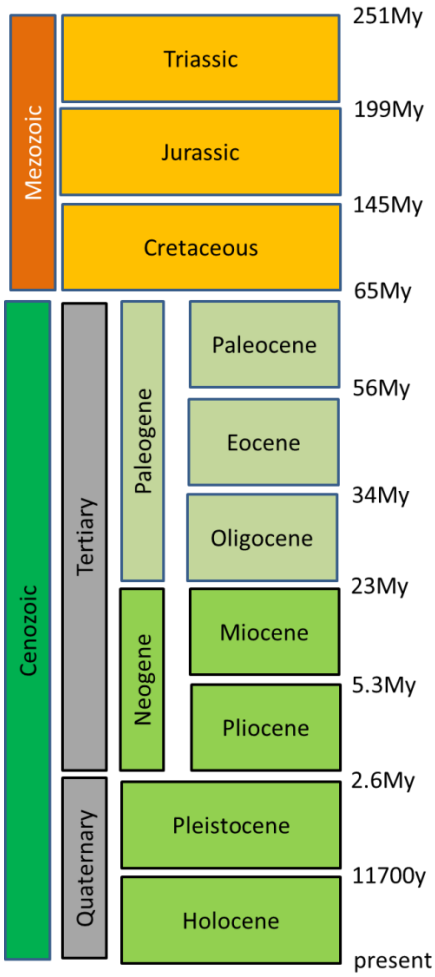


Figure 6: Tree species diversity (number of tree species/ha) in several plots of two experimental devices of French Guiana (Nouragues and Paracou). Data from Paracou and Nouragues inventories (UMR EcoFoG).

2. The building of biodiversity in Amazonia



The building of biodiversity in Amazonia through geological ages has been studied for a long time, and both geological and climatic histories of the Amazonian region allow understanding the present diversity. In particular, **tectonic events** (Andean Uplift during the Tertiary) and **climatic changes** (quaternary glaciations) had a crucial role in **species diversification in Amazonia.**

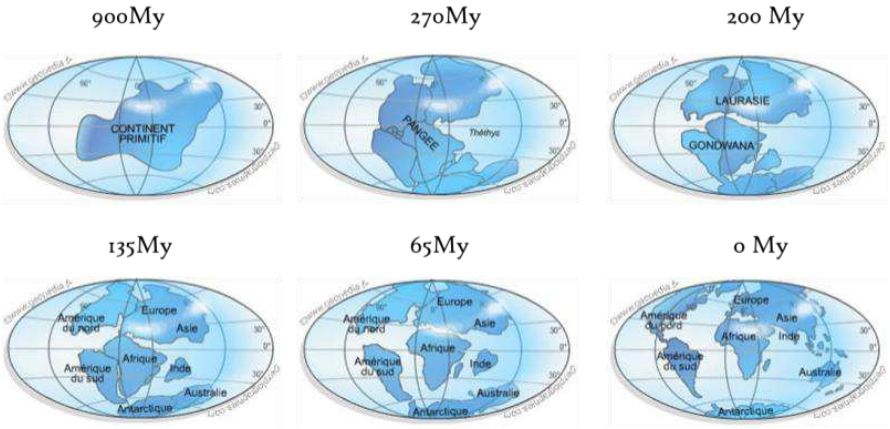


Figure 7: Continental drift. (from http://www.geopedia.fr/derive-continents.htm)

Tertiary - Andean Uplift: (Hoorn *et al.* 2010)

Andes formations began when continents broke-up (from -135 to -100 My before present, **figure 7**). From -65 to -23 My (paleogene), tectonic events in the 'pan-Amazonian' region (the region corresponding to modern Amazonia) formed a sub-andean river system. North and Nort-West of pan-amazonia were submitted to the alternance of fluvial and marine conditions (due to marine introgressions), **figure 8.**

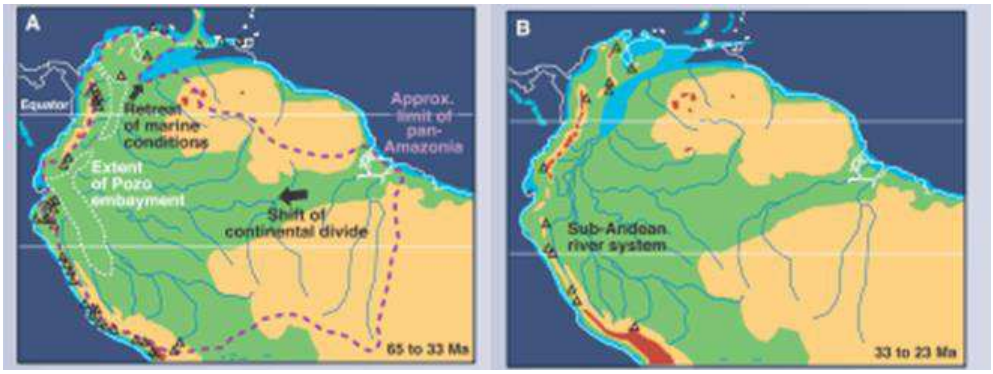


Figure 8: Geologic history of Amazonia from -65My to -23My (from Hoorn *et al.* 2010)

During this period, South America was colonized by xenarthrans, reptilians, and plant groups through the gondwana connection with Australia and Antartica (**figure 9A**), and floral diversity varied with temperatures (with decreases in plant diversity during cool periods).

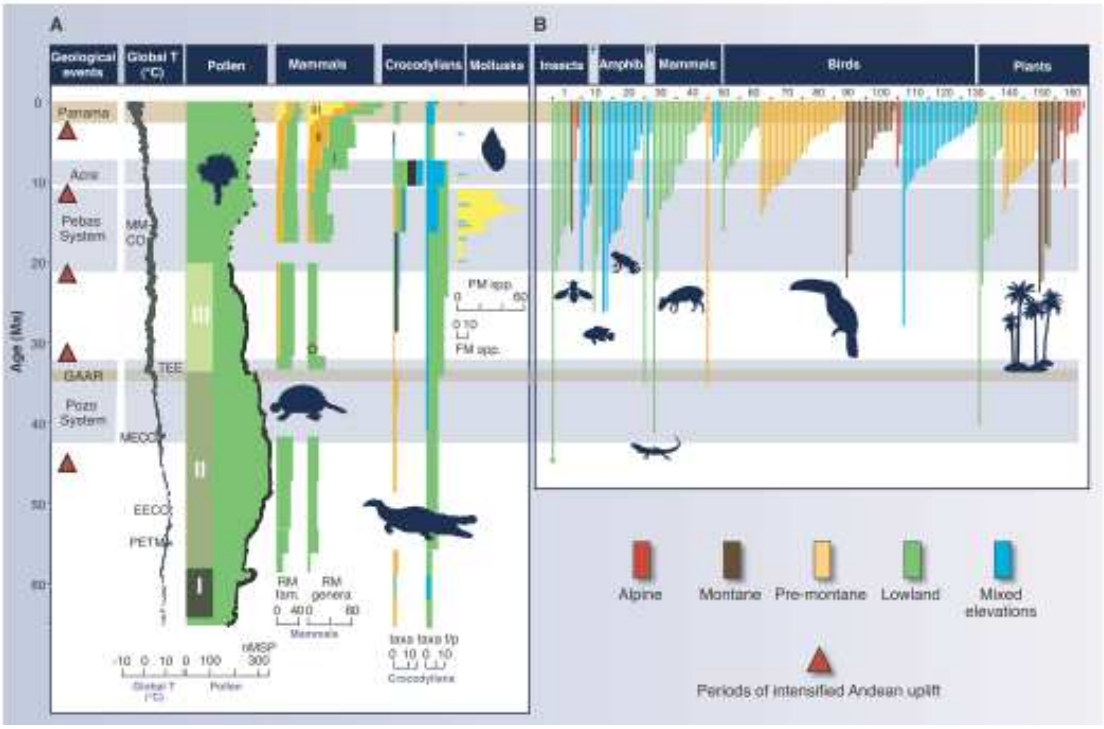


Figure 9: Biotic changes in Amazonia through time (from Hoorn *et al.* 2010).

The first peak of the Andean mountain appeared around ~23 My, and coincides with the diversification of modern montane plant and birds genera (**figure 9B and figure 10**). During Neogene, the coupling of tectonic and climatic processes strongly affected the biodiversity in Amazonia. As mountain raised, rainfall increased along the eastern flank. In parallel, a large wetland of lakes and swamps developed in Western Amazonia (**figure 11C**).

Lake formation was accompanied by mollusk and reptilian diversification (**figure 9**). The Amazonia was thus composed by a wetland and a diverse forest comparable with the modern forest.

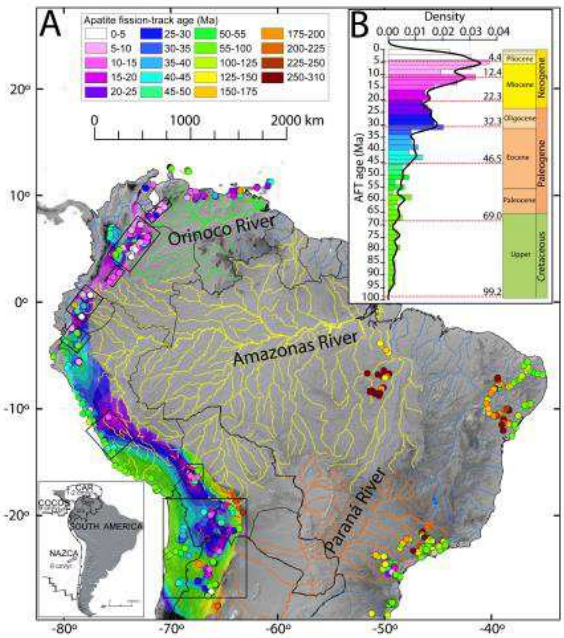


Figure 10: Andean Uplift (from Hoorn *et al.* 2010)

Pollen recorded revealed a peak of plant diversity at -13My at the end of the middle Miocene climatic optimum (-20 to -10 My). From 10 My, Andean uplift phases accelerated and Andean sediments reached the Atlantic (**figure 11D**). Western Amazonia changed from a lacustrine to a fluvial system that corresponds to modern Amazonia river. Accelerations in Andean building induced spectacular radiations of highland plants, and flood-plains became covered by grasses. Moreover, the Andes became a barrier for tropical rainforest trees, because many lowland organisms were unable to disperse across the mountain (Cavers & Dick 2013). The transition from lake to fluvial conditions also affected the diversity of endemic marine animals (mollusks) unable to adapt to new conditions (**figure 9**).

From -7 to -2.5 My, Andean sediment supply created terrestrial conditions in west

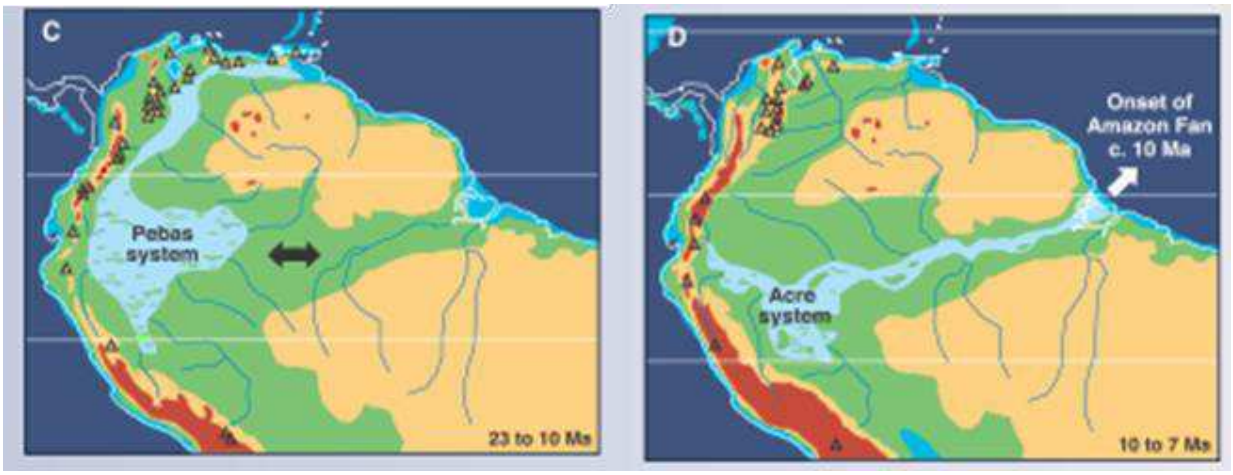


Figure 11: Geologic history of Amazonia from -23My to -7My (from Hoorn *et al.* 2010)

Amazonia (**figure 12**). Until the Pliocene, bats and plants (Malpighiaceae, Fabaceae, Annonaceae, Rubiaceae ...) migrated from boreotropical regions. 3.5 My ago, the closure of the Panama isthmus allowed migration and diversification of taxa from North America.

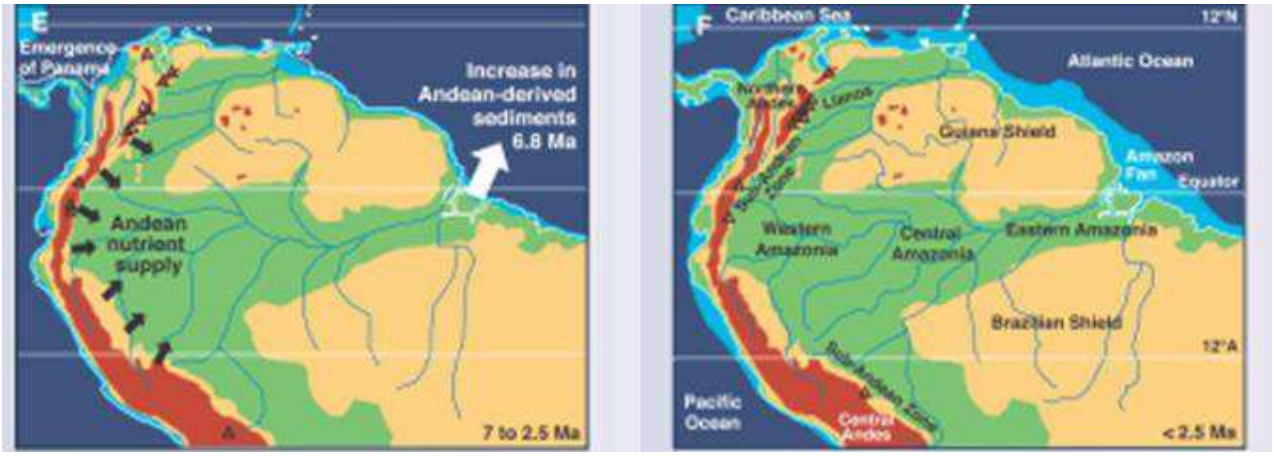


Figure 12: Geologic history of Amazonia from -7My to present (from Hoorn *et al.* 2010)

Quaternary ice ages:

The Pleistocene was characterized by repeated glacial periods, in which the last glacial maximum occurred (LGM) at -20 000 years (**figure 13**). Glacial periods were associated with a decrease in temperature in Amazonia, ranging from 2 to 6°C during the LGM (Broccoli 2000, Anhuf *et al.* 2006), **figure 14**. Cooler periods were also associated with a decrease in precipitations (between 20 and 30% during the LGM, Stute *et al.* 1995, Cowling *et al.* 2001, Anhuf et al. 2006), leading to a drier climate in Amazonia.

Species evolution in Amazonia through ice ages has been strongly debated. Two main hypotheses have emerged: the '**refuge hypothesis**' and the '**species re-association hypothesis**'.

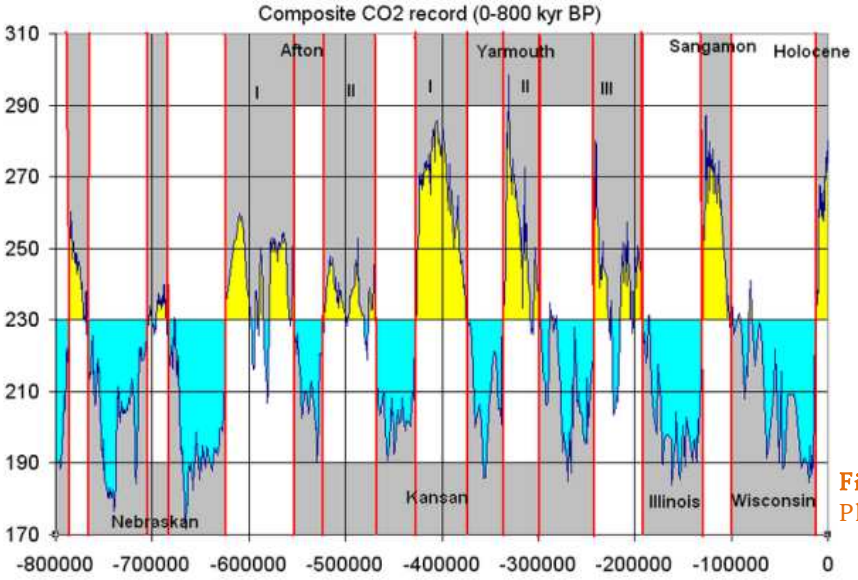


Figure 13: Glacial and interglacial cycles of Pleistocene (from Wikipedia).

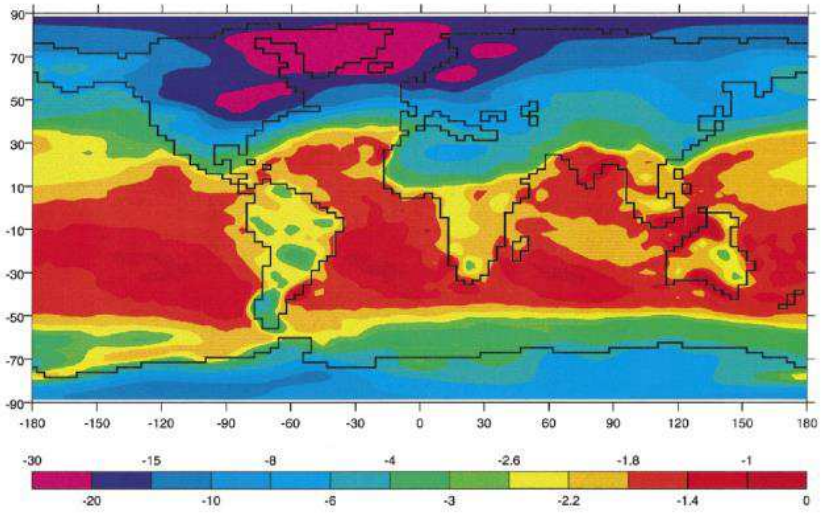


Figure 14: Map of annual mean surface air temperature difference between LGM and modern integrations, in Kelvin (from Broccoli 2000).

The 'refuge hypothesis' suggests that the tropical rainforest was fragmented into refuge islands during dry periods: the Amazonian basin changed into savanna punctuated with isolated patches of tropical rainforest (i.e. movements of the whole plant communities). LAI was probably lower in a large area of the Amazon Basin than today (Cowling 2001). Anhuf *et al.* (2006) used pollen records to map South America and Africa vegetation during the LGM (**figure 15**). They suggest that Amazon evergreen forest was located 200km further south and 300 km further north than the modern forest. More recently, Mayle & Power (2008) described sites that show signs of transition from forest to Savannas during the mid-Holocene. Isolation between these islands would have led to high rates of allopatric speciation and is supposed to be responsible for spatial patterns of species diversity and endemism (Haffer 1969).

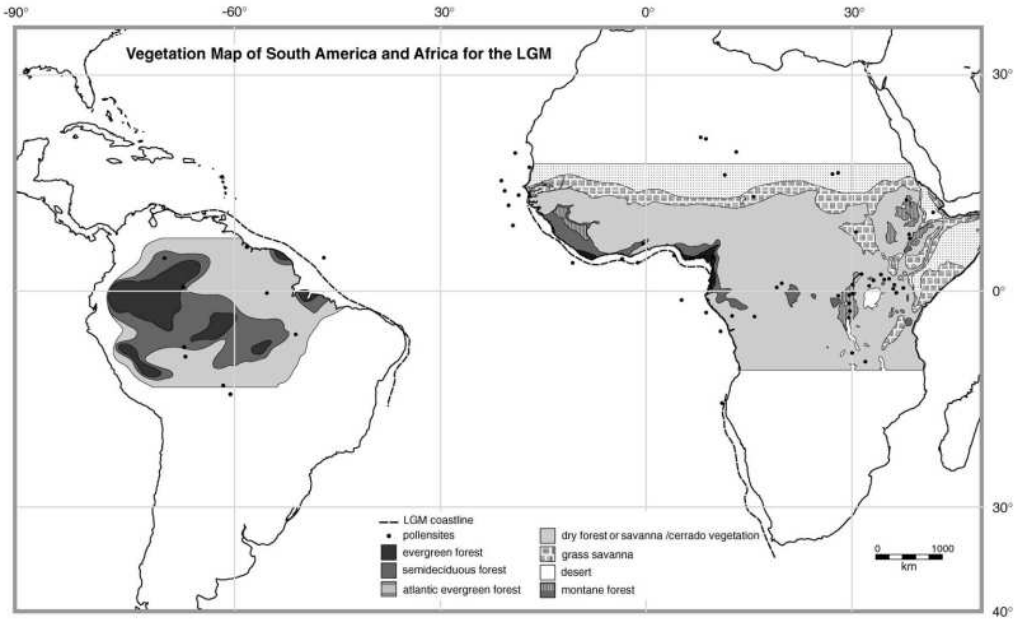


Figure 15: Vegetation map during the LGM (from Anhuf *et al.* 2006): evergreen forests (black), semideciduous forests (dark grey), dry forest or savannah (grey).

The 'species re-association' hypothesis suggests that cooling induced changes in species compositions but not in biomes: because species respond individually to physiological constraints (Collinvaux *et al.* 2000), climate change would have impacted the abundance of species (and species distributions), leading to species re-associations (Bush & De Oliveira 2006). For example, the abundance of mountain taxa would have decreased, while lowland pollen taxa have increased in Peruvian Andes during the LMG (Bush *et al.* 2004). However, this hypothesis suggests that tree cover remained stable. In particular, pollen records show a continuum of forest pollen through the LGM (**figure 16**; Colinvaux *et al.* 2000, Da Silveira Lobo Sternberg 2001), even with an increasing abundance of grass in

many areas that not necessarily traduces a drier climate (Colinvaux *et al.* 2000).

Both hypotheses agree that the forest was different from today, because it experienced transformations in floristic composition during the glaciations.

Much of Neotropical diversity was primarily influenced by Tertiary (Andean Uplift) and Quaternary (climatic changes) events. However, the actual triggers of speciation are probably more complex, involving factors such as adaptation to habitat heterogeneity and biotic interactions.

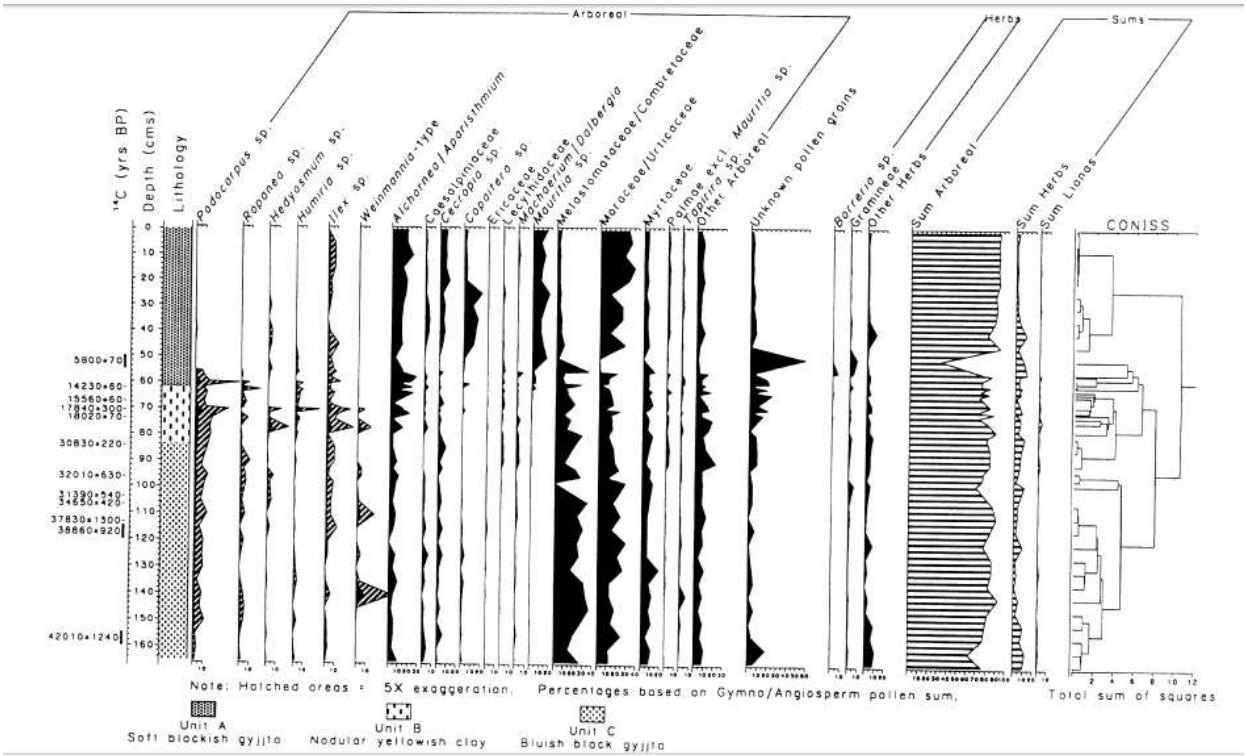


Figure 16: Pollen diagram from a lowland tropical forest in Brazil (from Colinvaux *et al.* 2000).

3. The maintenance of diversity in Amazonia: a subtle combination of chance and determinism

Two main theories are evoked to explain community assembly and the maintenance of high diversity across tropical rainforest landscapes: Neutralism and Determinism. The contrast between the neutralist and the determinist theories of community assembly is quite comparable to the contrast between neutral and adaptive (molecular) evolution of populations.

Under the unified neutral theory of biodiversity (Hubbell 2001), meta-community dynamics is governed by the speciation-extinction equilibrium in which the size of populations changes randomly ('ecological drift'), eventually leading to extinction, and populations exchange individuals according with dispersal distance between them (Ricklefs, 2006). Thus, species assemblages are **random subsets** of the available pool of species able to spread in a given area (Tuomisto & Ruokolainen 1997). Even if this model is often unrealistic (Ricklefs, 2006), it accounts for most of the observed patterns of species abundance in tropical communities, suggesting that neutral process play a crucial role in community assembly (Chaves *et al.*, 2003). From an evolutionary point of view, populations may evolve neutrally (under the combination of random mutation, migration, genetic drift and demographic events). In theory, populations may diverge into separate species if **gene flow is restricted**, either by a biogeographic barrier, or by the geographic distance between populations (also called isolation-by-distance). In such cases of **allopatric speciation**, the probability to observe a given species in a given area is thus a function of the dispersal abilities of the neighborhood populations of this species (Latimer *et al.* 2005). The great diversity observed in Amazonia, by comparison with temperate forests, is commonly explained by differences in speciation-extinction rates that are themselves dependent on the size of the climatically similar area. The main hypothesis is that there is a positive relationship between an ecoclimatic zone and the geographic range size of a species. Subsequently, two main hypotheses could explain the great diversity of the tropics: **'museum'** and **'cradle'** (Chown & Gaston 2000, Mittelbach *et al.* 2007, Arita & Vazquez-Domingez 2008). The 'museum' hypothesis postulates that there is a negative relationship between the geographic range size of a species and its likelihood of extinction. This is because large ranges should buffer species against extinction by reducing the probability of range wide catastrophes and because large population sizes would minimize the chance of extinction due to stochastic reasons. Because large species range sizes are typical of the

tropics, tropics should act as a museum of diversity with low extinction rates with older taxa by comparison with temperate zones. The 'cradles' hypothesis postulates that there is a positive relationship between the geographic range size of a species and the likelihood of its speciation. This is because species with larger ranges are more likely to undergo allopatric speciation resulting from isolations-by-distance or isolations by biogeographic barriers. Tropic may thus be viewed as cradles of diversity, with high speciation rates.

In the 'environmental filtering' theory, species assemblages are controlled by **determinist factors** involving abiotic and biotic interactions (Wright 2002). In particular, **habitat heterogeneity** (Terborgh *et al.* 2002) and local interactions (mainly competition and predation) are commonly evoked as important drivers of diversity in tropical landscapes. Environmental filtering exerted by both abiotic and biotic factors would have led to **niche partitioning** and **habitat specialization** in tropical rainforest trees. From an evolutionary point of view, the evolution of populations and the divergence between species may have been driven by selective pressures exerted by environmental heterogeneity (**sympatric speciation**). Moreover, habitat heterogeneity is associated to disturbance gradients (particularly logging, and tree-fall gaps). Under the **disturbance hypothesis**, species diversity is enhanced by intermediate levels of disturbance, as observed in French Guiana (Molino *et al.* 2001). Another determinist hypothesis evokes **density-dependant mortality** around mother trees. This hypothesis was formulated by Janzen (1970) who observed a decrease in seedlings mortality with the distance to the mother trees, probably due to allelopathic chemical compounds or to density-dependent predation. This process leads to 'gaps of regeneration' around mother trees, allowing the installation of other tree species and preventing mono-specific assembly. However, this process remains poorly understood and documented.

Neutrality and determinism probably act in pair in governing species evolution and assembly structuring (Gravel *et al.* 2006, Jabot *et al.* 2008), and their relative effects probably vary across geographical scales and study areas (Gravel *et al.* 2006, Jabot *et al.* 2008). In the following section, I will focus on spatial heterogeneity in tropical landscapes (and particularly that observed at local scales) without, however, excluding the existence of neutral processes.

4. Spatial heterogeneity in the Amazonian rainforest

Environmental heterogeneity across forests landscapes

At continental and regional scales, both precipitations and the intensity of the dry season are the main causes of climatic variations across the Amazonian forest landscape: while temperatures are quite homogeneous, precipitations show large variations among regions (ranging from 1000 to 3000 mm per year, **figure 17**) with a precipitation gradient that increases from Southeast to Northwest Amazonia (Mayle & Power 2008). Moreover, the intensity of the dry season is more pronounced at the extreme of the gradients, where precipitations are the most abundant. French Guiana also exhibits a large gradient of precipitations that increases from west to east, **figure 18** (Wagner 2011).

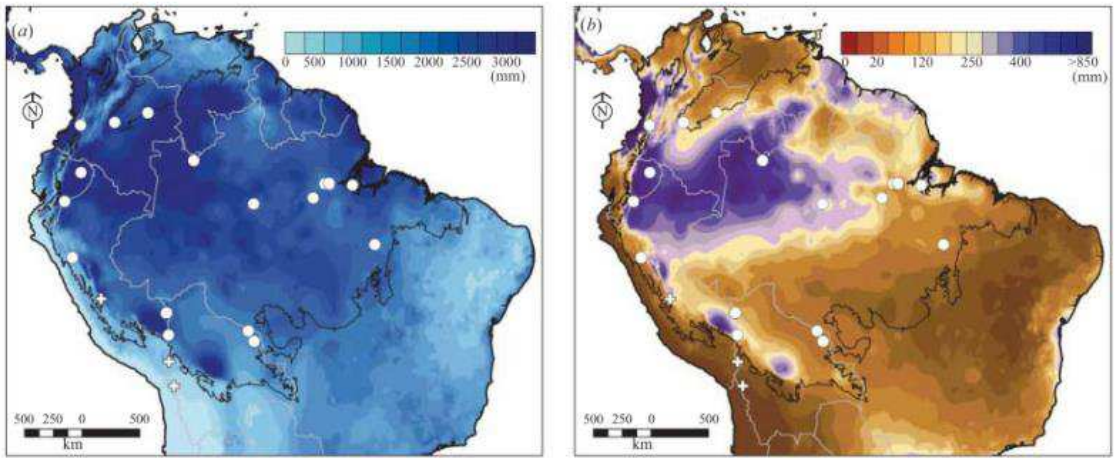


Figure 17: Annual precipitations and precipitations during the driest three months in the Amazonian basin, in mm. From Mayle & Power 2008.

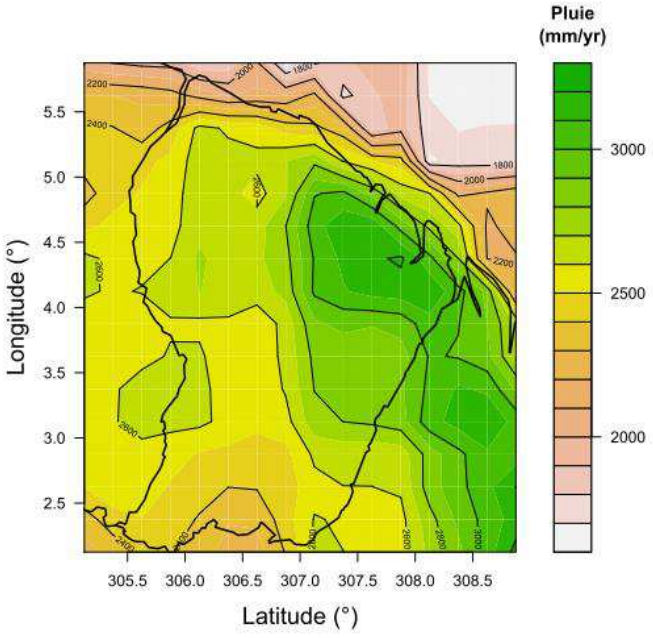


Figure 18: Annual precipitations in French Guiana in mm/yr. From Wagner 2011.

At local scale, large environmental variations are caused by soil factors related to topography (**figure 19**). Despite its apparent homogeneity, the tropical landscape of Amazonia displays complex habitat patchiness due to the alternation of water-logged bottomlands and terra-firma. Local topography causes strong differences in environmental factors (including water, light, and nutrient availability) among local micro-habitats.

In bottomlands, plant communities are established on hygromorphic soils submitted to seasonal or permanent water-logging and frequent flooding events. As in temperate ecosystems, water-logging is a major constraint for tree regeneration and growth. Water-logging decreases the solubility and transfer of o2 in the soils. Due to root and soil microbial respiration, oxygen quickly decreases in soils; leading to hypoxia and accumulation of CO_2 (Ponnamperuma 1972, Kozlowski 1997) that in turn affects root and microbial respiration (Epron *et al.* 2006). Moreover, water-logging leads to production of reactive oxygen species by roots that causes oxidative stress (mainly, H_2O_2 is produced by mitochondria when respiration slow down), Perata *et al.* 2011. In parallel, hypoxia causes a decreases in the root permeability that subsequently affect water and nutrient uptake from the soil, causing stomatal closure and a decrease in photosynthesis (Perata *et al.* 2011). On the contrary, terra-firme (slopes and hilltops) are display ferralitic and well-drained soils allowing important vertical and lateral drainage. Thus, terra-firme soils usually display lower water content than bottomlands. Tree communities, particularly seedlings unable to directly uptake water from the ground water table, may experience seasonal drought stress due to the depletion of water from at least the upper soil layers (Bonal *et al.* 2000, Daws *et al.* 2002, **figure 20**).

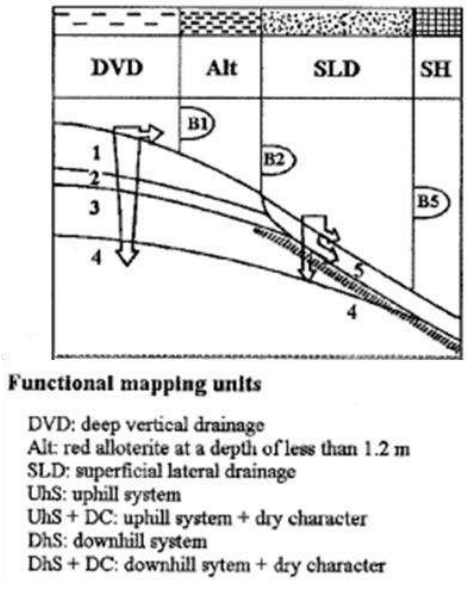


Figure 19: Soil properties along topography gradients (from Sabatier *et al.* 1997). Mainly, DVD=deep vertical drainage, Alt=red alloteriet, SLD= superficial lateral drainage and SH= hydromorphic soil.

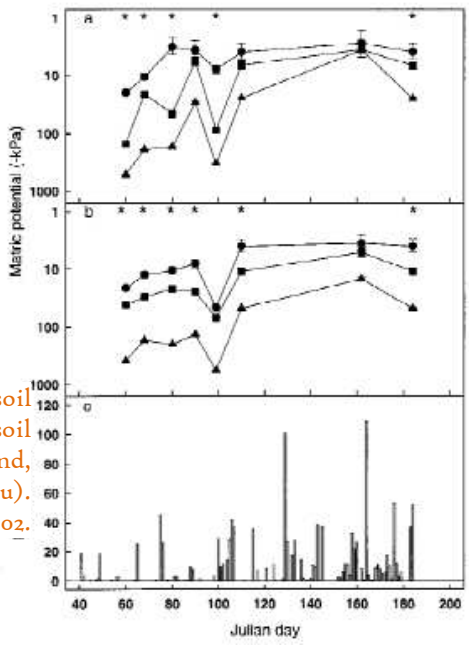


Figure 20: Seasonal variations of soil metric potential in different soil types (circles=bottomland, squares=slopes, triangle=plateau). From Daws *et al.* 2002.

Moreover, soil fertility varies from hilltops to bottomland. Reductions in soil respiration affect nitrogen cycling in bottomlands (Luizao *et al.* 2004) that commonly contain less nitrogen than hilltops or slopes (**figure 21**) but frequently contain more phosphorous than hilltops (Ferry *et al.* 2010). Last, topography gradients are associated with variations in irradiance transmitted below the canopy. As the soil is instable in slopes and water-logged soils, tree-fall gaps occur more frequently in slopes and bottomlands (Marthews *et al.* 2008, Ferry *et al.* 2010), **figure 22**.

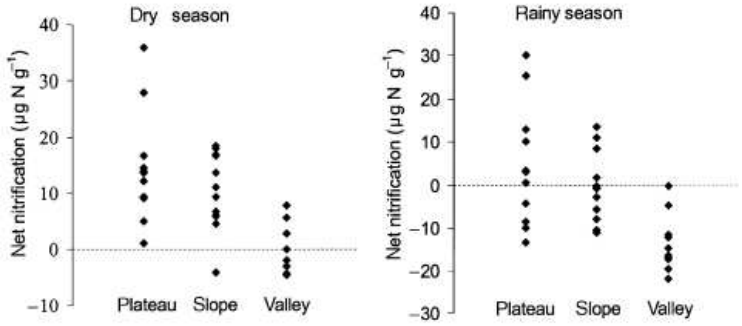


Figure 21: Net nitrification during 10 days of soil incubation in different soil types. From Luizao *et al.* 2004

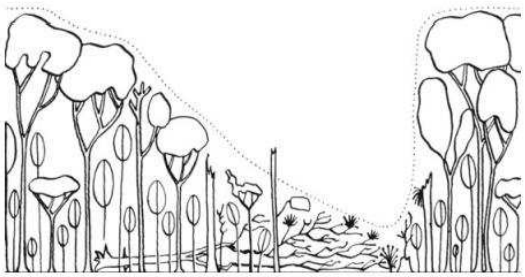


Figure 22: Tree fall gap. From Matthews *et al.* 2008

<u>Consequences of spatial heterogeneity on plant communities:</u>

At regional scale, the structure and composition of plant communities may vary along rainfall gradients, as proposed by numerous studies (Givnish, 1999, Engelbrecht & Kursar, 2003, Condit *et al.* 2004). However, discerning whether adaptive or neutral processes are involved is a complex issue at such large scales. In the particular case of Amazonia, rainfall is supposed to exert a small effect on species diversity, whereas a strong effect of local patchiness is evident (ter Steege & Hammond 2001).

At local scale, large variations of plant community composition and diversity vary along topographic gradients. The most obvious variation of plant communities is the large increase in palm biomass in bottomlands (Kahn 1987, **figure 23**) and variations in tree species composition. Indeed, numerous palm and tree species are significantly associated to a particular habitat-type (Clark *et al.* 1999, Vormisto *et al.* 2004, Baraloto *et al.* 2007). This statement is commonly invoked as a result of adaptive radiations caused by topography leading to niche partitioning and habitat specialization. However, several studies suggested that the majority of species is generalists regarding to local habitat (**figure 24**, Webb & Peart 2000, Valencia *et al.* 2004) and their distribution is probably constrained by dispersal without being influenced by habitat heterogeneity.

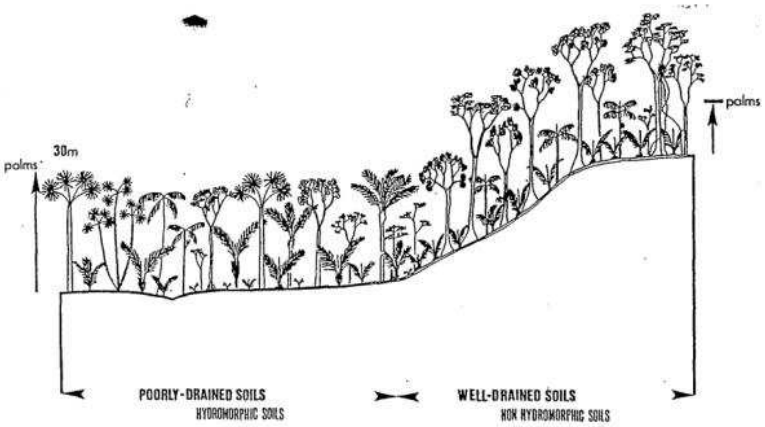


Figure 23: Schematic representation of plant communities along a topography gradient. From Kahn *et al.* 1987.

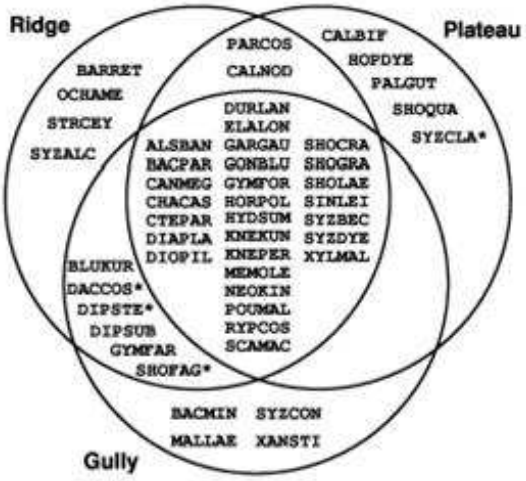


Figure 24: Venn diagram showing associations of tree species to three local habitats (from Webb & Peart 2000). Circle intercepts show species encountered in different habitats.

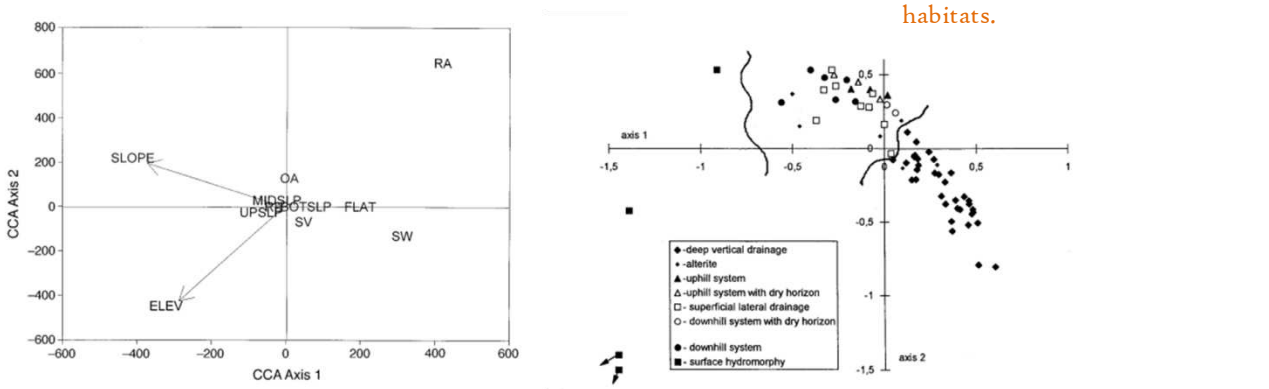


Figure 25: Left: Canonical correspondence analysis for environmental variables: soil types, topographical positions, slope and elevation (from Clark *et al.* 1999); Right: Vegetation ordination after correspondence analysis: symbols indicates different soil types differing in drainage and hygromorphy: (from Sabatier *et al.* 1997)

Several topographic and soil variables are however particularly relevant for explaining tree community composition and structuring (ter Steege *et al.* 1993, Clark *et al.* 1999, Sabatier *et al.* 1997, Kanagaraj *et al.* 2011), including slope, elevation, soil water availability, drainage, and water logging, **figures 25**.

Even if a majority of studies focus on one or several environmental factors or topographic variables, the structure of plant communities probably results from a complex superposition of factors (among which local irradiance, nutrient availability, water-logging and drought). Thus significant habitat-associations are commonly explained by species sensitivity to the underlying constraints: Engelbrecht *et al.* (2005, 2007) and Poorter *et al.* (2008) proposed drought, Paliotto *et al.* (Palmiotto *et al.* 2004) suggested irradiance, Lopez *et*

al. (Lopez & Kursar 2003) proposed both flood and drought, whereas Baraloto *et al.* (Baraloto *et al.* 2005) proposed both nutrients and light. For example, a field experiment revealed a reversal of performance ranking among species between local situations (Baraloto *et al.* 2005), suggesting different degrees of sensitivity to constraints among species. Thus, adaptation to a particular habitat may partly explain the differences in community composition and species abundance among micro-habitat.

Local habitat patchiness is also associated with large variations of tree biomass and functional traits. In bottomlands, tree biomass is lower than in terra-firma (Kahn 1987, Ferry *et al.* 2010), probably because soil instability constrains a more superficial root anchorage and limits tree growth. Moreover, Kraft et al. (Kraft *et al.* 2008) found a significant structuring of functional traits at the community level in Ecuador, which is also consistent with a role of habitat filtering, **figure 26**. Another kind of phenotypic structuring commonly observed in tropical rainforest is the ability of trees to develop morphological particularities, particularly in bottomlands. For example, buttress or stilt roots prevent constraints due to soil instability, whereas adventitious roots, lenticels, and aerenchyma tissues allow partial maintenance of root respiration in water-logged habitats, by allowing oxygen uptake directly from the air and oxygen transport to roots (Kozlowski 1997, Parelle 2010).

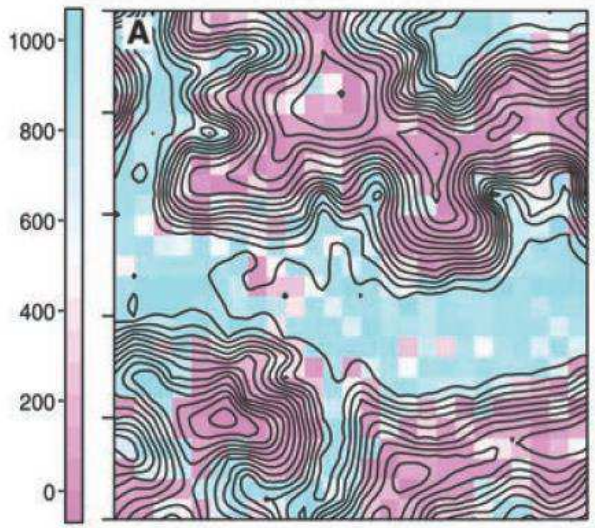


Figure 26: Distribution of SLA (expressed as a deviance from null distribution) in relation with local topography. From Kraft *et al.* 2008.

The entire forest dynamics vary along topographic gradients: canopy opening events created by frequent tree-fall gaps are also proposed as a driver of diversity in meta-communities (Schnitzer 2001, Robert 2003), by allowing establishment of light-demanding pioneer species and thus, creating patches of regenerations in the middle of mature communities composed by a majority of shade-tolerant tree species (Denslow et al. 1987, Schnitzer 2001, Ferry *et al.* 2010). Quesada *et al.* (2009) categorized forest dynamics according to a function of disturbance from soils, see **figure 27**.

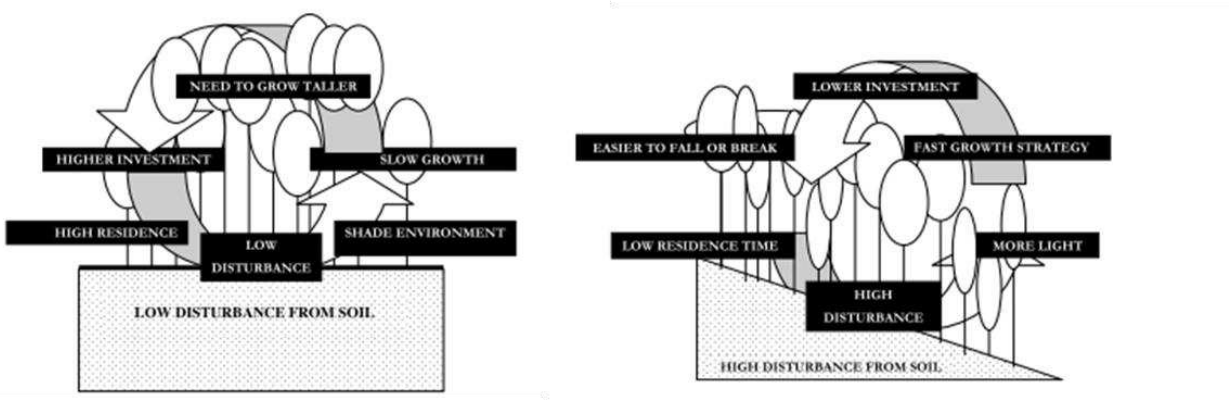


Figure 27: Variations in forest dynamics in relation with soil type in Amazonia. From Quesada *et al.* 2009.

Consequences of spatial heterogeneity on population evolution & species divergence

As quickly evoked previously ('3.The maintenance of diversity in Amazonia'), population evolution is driven by a combination of neutral (mutation, recombination, genetic drift, migration, reproduction, demography) and adaptive (natural selection) processes. Populations may diverge into new species, either due to isolation-by-distance that may be caused by populations isolation into refuges or biogeographic barriers (**allopatric speciation**), or by **local adaptation** to habitat heterogeneity (**sympatric speciation**). However, the drivers of populations evolution and speciation processes in tropical rainforest trees are poorly known, partly because the boundaries of species are often confused, and many species are organized in species complexes, with incomplete reproductive isolation between species and cryptic species (Cavers & Dick 2013).

At regional scale, many phylogeographic analyses revealed patterns of genetic divergence structured by the biogeographic history of the species, and mainly dispersal constraints that occurred during tertiary and quaternary. For example, *Jacaranda copaia* is widespread in the Amazon basin and comprises two sub-species: one subspecies widespread from Central America to Bolivia and another one distributed in the Guiana shield. In a recent study, Scotti-Saintagne *et al.* (2012) showed that the geographical patterns of genetic diversity in these two *Jacaranda copaia* sub-species were largely shaped by Pleistocene climatic changes that isolated ancestral species into refuges, with a center of diversification in Central Amazonia probably due to a secondary contact zone. Moreover, the absence of cross-Andean disjunction suggested that the Andean uplift was not a barrier to dispersal, probably because *Jacaranda copaia* is a wind-dispersed pioneer species, favored by canopy gaps and disturbances, and able to tolerate relatively dry conditions. Another example is

provided by the *Carapa* species complex (Duminil *et al.* 2006, Scotti-Saintagne *et al.* 2012). Scotti-Saintagne *et al.* suggested that the biogeographic history of two *Carapa* species was a combination of tertiary and quaternary events, including Pliocene Andean uplifts, and then late Miocene development of Amazon drainage, but was also influenced by hybridization and introgressions during the Quaternary, **figure 28**.

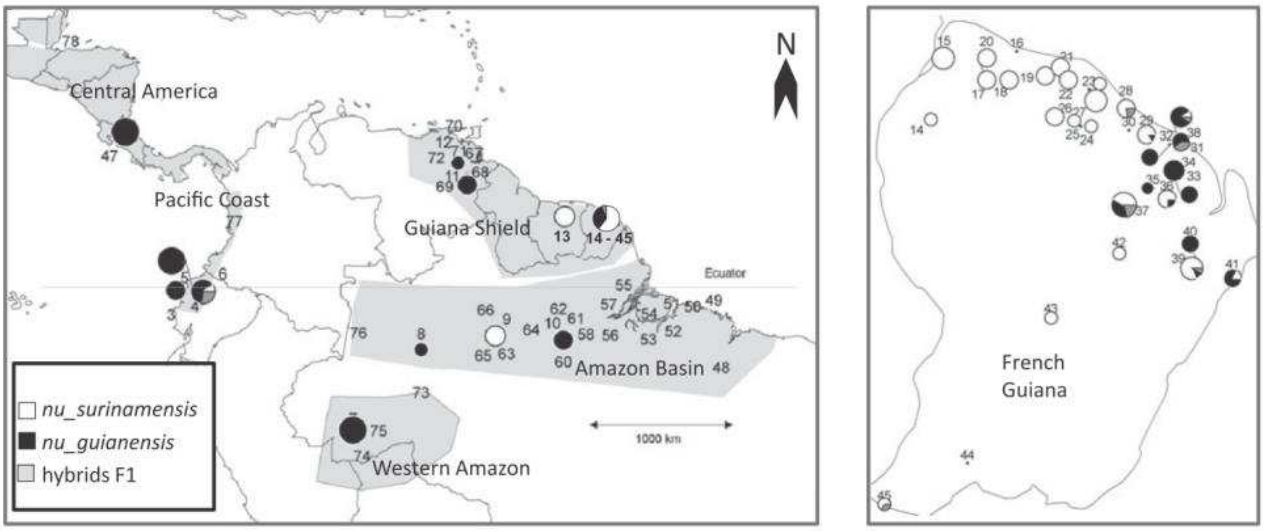


Figure 28: Bayesian clustering analysis for the tree Genus Carapa in the Neotropics (from Scotti-Saintagne *et al.* 2012). Maps indicate the structuring of genetic diversity at continental and regional scales.

In an original study (Fine & Kembel 2010), Fine *et al.* evoked the large influence of specialization to habitat type in driving the phylogenetic divergence between species. They analyzed the phylogenetic structure of Amazonian communities involving 1972 taxa across habitat types in Peru (white-sands that were widespread before Andean uplift and terra-firme forests composed by Cretaceous sediments that were laid down during Miocene). They compared the relative effects of habitat type and geographic distances between communities on the phylogenetic distances between taxa. They concluded that both dispersal limitation and habitat specialization influenced species divergence in tropical forests,

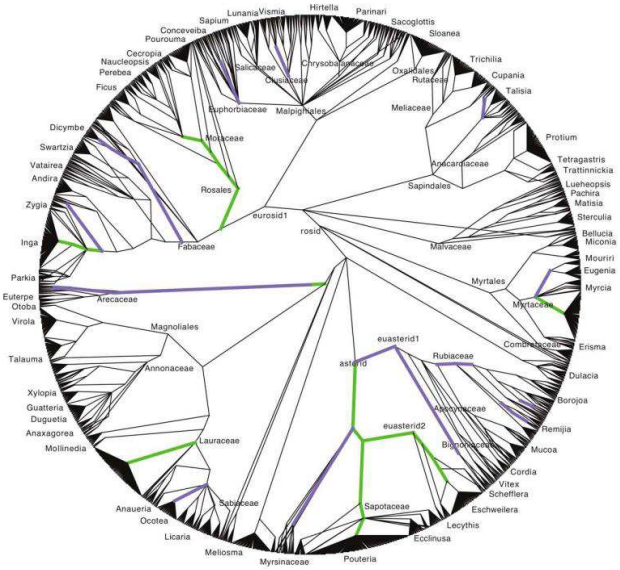


Figure 29: Phylogenetic tree linking 1972 taxa in Amazonia. Thick lineages indicate lineages containing more descendant taxa associated to terra-firme (green) and wite-sand (blue) habitats than expected by chance. (Fine & Kembel 2010)

but the effect of habitat specialization was greater than distance between communities, **figure 29**. They remained, however, cautious about the age of divergence: both biogeographic history of habitat types and recent *in situ* adaptive radiations governed by habitat heterogeneity would be involved in clade divergence.

Taken together, these results reveal that the **biogeographic history of species is often insufficient to catch all the processes that structured the genetic diversity and induced speciation in Amazonian landscapes. In particular, more recent specialization to constraints would also be involved in species evolution and divergence, particularly at local scale.**

At local scale, several studies revealed strong evidence of habitat specialization among closely related species. Baraloto *et al.* analyzed the distribution of four pairs of species from the same genus and observed divergent local habitat-associations between closely-related species (Baraloto *et al.* 2007). They proposed that specialization to local habitat may explain patterns of adaptive radiation in many tree genera. Similarly, Tuomisto *et al.* (Tuomisto 2006) observed strong evidence of niche specialization to local edaphic constraints (soil texture, soil cation content, inundation) between species of the *Polybotrya* genus in northwestern Amazonia.

Even if numerous studies evoked the influence of local variations in shaping the genetic diversity of tropical plants and in driving sympatric speciation, no study yet provided molecular evidences of local adaptation at intra-specific level in Amazonia. In temperate and boreal plant communities, local adaptation has been largely investigated and provides a wide range of examples: local adaptation to altitudinal gradients (Savolainen 2011), to water-logging (Parelle *et al.* 2010) etc... (see section 'Molecular evolution'). In tropical rainforests, however, the relative influence of local adaptation and neutral processes in structuring the genetic diversity over short spatial scales remains largely misunderstood and requires much attention, particularly in the current context of climate change.

5. Tree species model, research questions and study sites

In this study, I address the question of population evolution at local scale within continuous populations of a dominant tree species widespread in French Guiana: *Eperua falcata* (a complete description of the species is given page 35). I addressed two main questions:

1) **How is the genetic diversity of *Eperua falcata* structured in the forest landscapes of French Guiana?**
2) **Which evolutionary drivers are relevant to explain the structure of genetic diversity at local scale?**
3) **Does local adaptation contribute to structure the genetic diversity at local scale within continuous populations?**

I analyzed the patterns of genetic diversity distribution within continuous forest landscapes of French Guiana through a global approach integrating both ecophysiological (phenotypic) and population genetics (molecular) approaches that are treated separately.

Figure 30 (page 34) provides a complete overview of the methods, the specific questions and future prospects.

Molecular evolution:

The section 'Molecular evolution' aims at (i) analyzing patterns of genetic differentiation among local habitats, (ii) identifying which evolutionary drivers structure the local genetic diversity of *Eperua falcata*, and (iii) testing for local adaptation by (iiia) detecting outlier loci under diversifying selection among local habitats and (iiib) estimating the extent of (divergent) natural selection in the genome of *Eperua falcata*. This section involves two main approaches:

- a candidate gene approach in which targeted genes of known function (potentially involved in adaptive genetic differentiation among local habitats) were sequenced: aquaporins, catalase, farnesyltransferase, etc...
- a genome-scan approach in which I genotyped a large number of (anonymous) AFLP markers spread over the genome.

The candidate gene approach was developed during the PhD of Delphine Audigeos. I participate to this work during my Master degree by developing genetic markers and by contributing to genetic analyses. The AFLPs approach was set-up during this PhD.

In parallel to population genetics, I worked on creating a large database of *Eperua falcata* expressed sequences (cDNA) that were sequenced by 454-pyrosequencing prior to this PhD. I realized the bioinformatics assembly and post-processed it to characterize genes and identify polymorphism. Such a database will be useful for further high-throughput re-sequencing or genotyping of candidate loci.

The different results obtained are detailed in the research articles, but the main results are summarized into this synthesis ('orange boxes'). The prospects of the study are discussed in the section 'Discussion'.

Phenotypic evolution:

In the section 'Phenotypic evolution', I analyzed (i) whether functional traits are (inherently) structured by local habitats, and (ii) whether habitat patchiness may have shaped tree sensitivity to environmental constraints (with a particular focus here on water stresses, including drought and water-logging). This section involves three experiments:

1- a provenance test under controlled and non-limiting conditions ('common garden'),
2- a provenance test under constraining conditions in which different water treatments were applied (drought and water-logging),
3- a reciprocal transplant experiment in natural conditions.

The two first experiments were designed, and their realization supervised by D. Bonal & I. Scotti from 2006 to 2008. The reciprocal transplant experiment was set up in 2011 at the beginning of this PhD. I designed and set up this third experiment (seed sampling, sowing, and seedlings transplant), and followed seedling growth from 2011 to 2013.

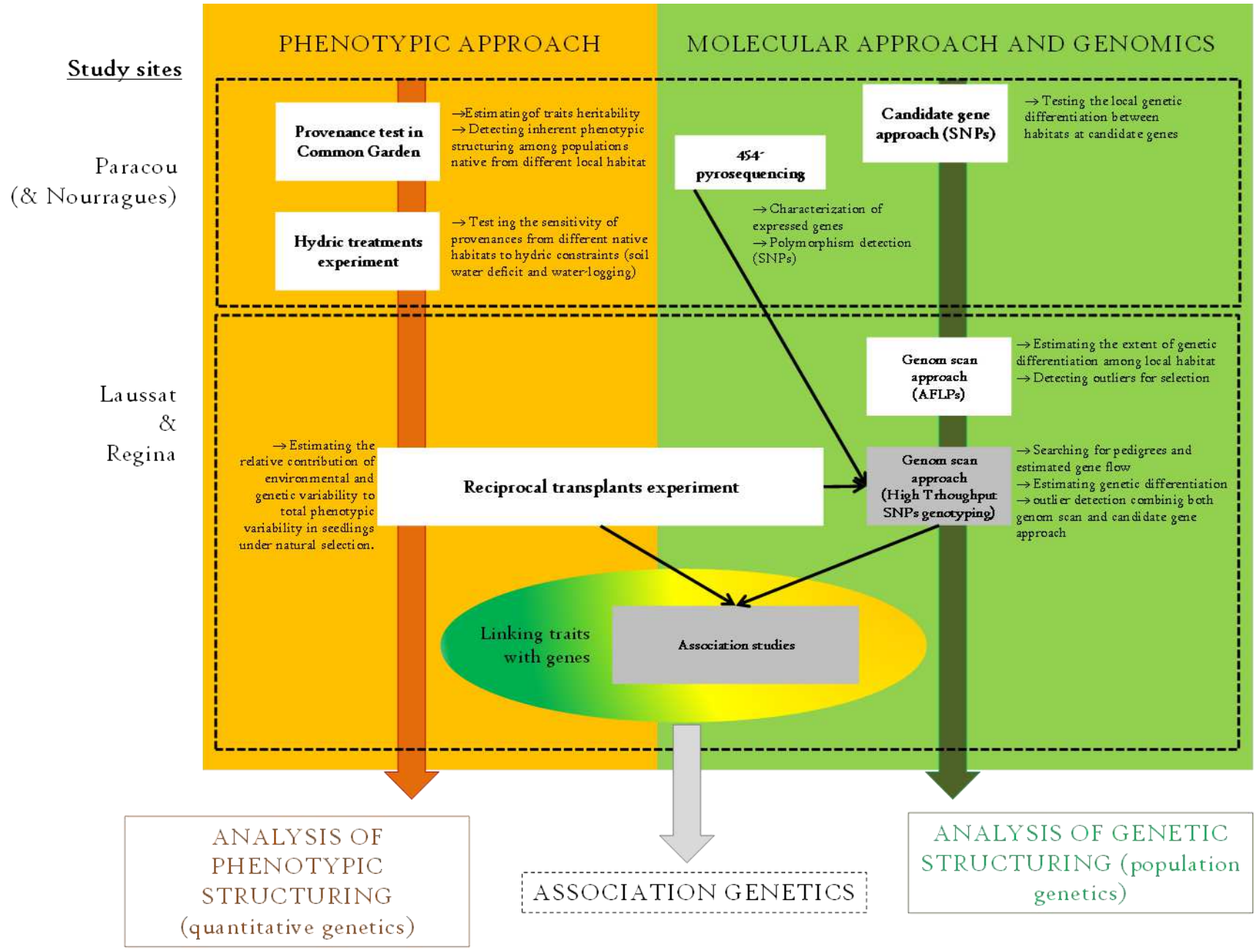


Figure 30: Complete overview of the methods with their specific questions and future opportunities (in grey boxes).

Eperua falcata (Aublet.)

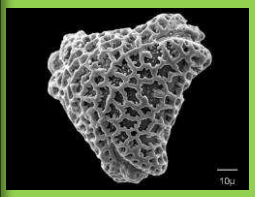
Taxonomy: Fam. Fabaceae, Subfam. Caesalpinioideae

Diversity:

The genus *Eperua* comprises about fifteen tree species in the Amazonia, but only three are present in French Guiana: *E. falcata*, *E. grandiflora*, and *E. rubiginosa*. *Eperua falcata* is the most common species of French Guiana

Continental distribution:

E. falcata is widespread in the Guiana shield. Its native distribution covers the whole Guiana shield (French Guiana, Suriname, Guyana) plus the North of Brazil and Venezuela.



Local distribution and spatial dynamics:

E. falcata is a generalist tree species able to colonize both water-logged and terra-firme habitats. However, it is more abundant in water-logged bottomland (Baraloto *et al.* 2006), while *E. grandiflora* is restricted to hilltops (Barthes 1991). The third species (*E. rubiginosa*) is mainly encountered along rivers, but it has already been observed on well-drained ferralitic soils. *E. falcata* has an aggregative behavior (Bariteau 1992), and often exhibits high population densities.

Phenology and Reproduction:

E. falcata flowers and fruits during athe the end of the dry season (September to March), Cowan 1975. It is probably both self-compatible and outcrossing, even if differential ripening of the anthers and stigmas would limit selfing. Large pollen grains (size > 100μm) are probably not dispersed by wind. Recurrent observations of bats visiting flowers suggest that the species is chiropterophilous (bat-mediated pollination). *E. falcata* is autochorous: heavy seeds are dispersed at short distance around mother trees by explosive pod dehiscence. Autochorous seed dissemination results in very restricted seed dispersal (Hardy *et al.* 2006), thus explaining its aggregative distribution (Bariteau 1992).

Successional status and physiology:

E. falcata is an evergreen canopy-dominant tree species which often emerges above the canopy. It is a '**fast-growing late successional species**' (Bonal *et al.* 2007): it displays lower carbon assimilation rates, leaf nitrogen and SLA than early-successional species, but higher SLA and leaf nitrogen than slow-growing late-successional species. As it emerges above the canopy, it is considered as a **shade hemitolerant** species, a group displaying higher water use efficiency than heliophilic or shade tolerant species that is considered as an adaptive trait to high evaporative demand prevailing in the emerging tree crowns. Because emergent trees are commonly not shaded by other trees and because *E. falcata* reaches large circumferences, it displays high sapflow densities (Granier *et al.* 1996). It is well **tolerant to drought**: assimilation balance of adult trees remains positive under moderate to severe drought (Bonal & Guehl 2001) and leaf physiology is not affected by seasonal soil drought (Bonal *et al.* 2000). It displays an **anisohydric behavior** in relation to soil drought: trees are deep rooted (with tap roots below 3 m, Bonal *et al.* 2000) and the stomatal conductance of seedlings diaplay a limited sensitivity to drought (Bonal & Guehl 2001).

Research advantages:

-Widespread in French Guiana, high population densities

-Easy to identify thanks to its characteristic even-pinnate sickle-shaped leaflets.

- Conciliatory with both genetic analysis and shade-house experiments.

Study sites:

Four study sites were used along the different approaches and experiments, **figure 31**. Three were located on the coast of French Guiana (Laussat, Paracou and Regina), whereas the site of Nourragues was the most continental. The study sites display large differences in water-regime, with and annual mean precipitations ranging from 2500 (Laussat) to 4000 (Regina) mm/year.

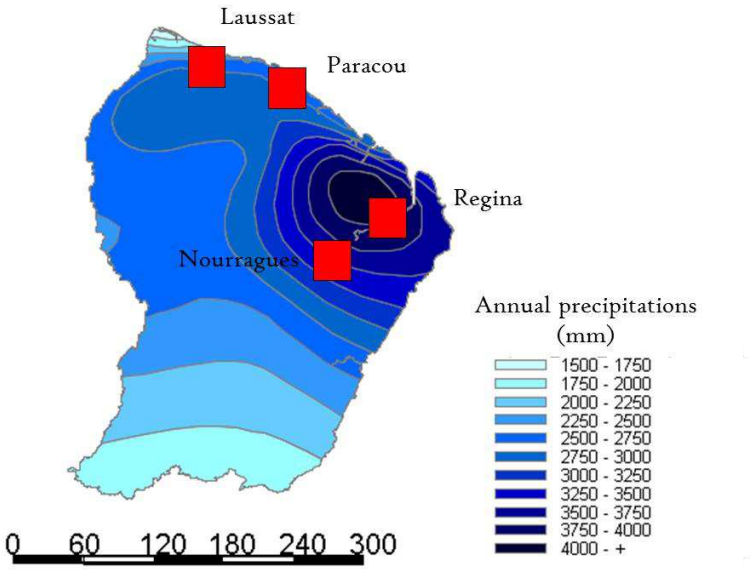


Figure 31: Location of the study sites.

Site	Location	p° (mm/year)	Area (ha)	*Eperua falcata* population density	Coexistence with other *Eperua* species
Laussat	Coastal (W)	2500	4.3	48.1	No
Paraou	Coastal (center)	2700	6.25	42.7	*E. grandiflora*
Regina	Coastal (East)	4000	6.7	29.9	*E. rubiginosa*
Nourragues	Continental (E)	3000	1	NA	NA

The different plots cover different habitat types, from bottomland to terra-firme, but display several topographic differences. Laussat and Nourragues experimental plots are composed by a permanently water-logged bottomland and a large plateau of low elevation and low slope (at Laussat, elevation ranges between 20 and 60 meters). Paracou is composed of a seasonally water-logged bottomland surrounded by two hilltops and separated by moderate slopes. At Regina, topography is more complex, leading to a habitat patchiness

composed by smaller patches more finely juxtaposed. Regina is composed by high and thin hilltops bordered by important slopes, with elevations ranging from 40 to 100 meters. A complex hydrological network carries water toward a seasonally water-logged bottomland submitted to frequent flooding events during the rainy season. In spite of these differences, the soil properties of local habitats are quite similar between sites: all bottomlands are characterized by hydromorphic soils with a large accumulation of organic matter whereas terra-firme are characterized by ferralitic soils undergoing important drainage due to their sandy texture, probably leading to soil water deficits the dry season.

PART 1 - Molecular evolution: population genetics and genomics

1. Population evolution

Evolution starts with the existence of genetic polymorphisms in the genome of organisms due to mutations: substitution ("single nucleotide polymorphism" or SNPs), insertions-deletions, copy number variation ("simple sequence repeat" or SSR). Mutations occurred randomly during meiosis and are transmitted to the progeny **by Mendelian inheritance**. Moreover, crossing over contributes to break **linkage disequilibrium** between two physically linked loci, and creates new combinations of alleles (genotypes) considering the two loci simultaneously.

In a population of infinite, and thus constant, size (i.e. no genetic drift, no demographic changes), if reproduction is panmictic among individuals (i.e. mating is random) and if there is no selection, the population is at the **Hardy-Weinberg equilibrium**: allelic and genotypic frequencies remain stable across generations. For a bi-allelic locus, homozygotes (A/A), (a/a) and heterozygotes (A/a) occur in proportion p^2, q^2, and 2pq respectively; where p and q correspond to allelic frequencies for the alleles (A) and (a), **figure 32**. The expected heterozygosity under Hardy-Weinberg equilibrium is thus equal to 2pq or $1-(p^2+q^2)$. This index is called **Nei's diversity index** and may be extended to multi-allelic loci, given: $h = 1 - \sum f(allelic)^2$.

On the contrary, the future of mutations in populations may vary across generations, driven by evolutionary drivers: genetic drift, gene flow, and natural selection. I also include mating system as well as the demographic history of populations as drivers of evolution.

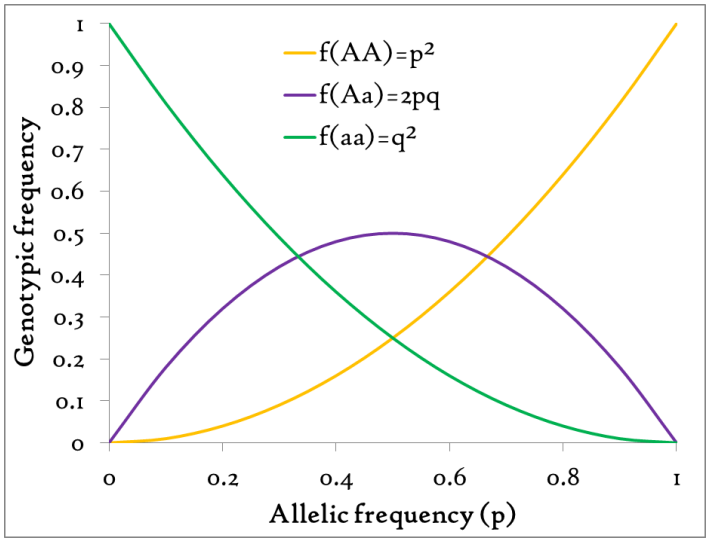


Figure 32: Genotypic frequencies as a function of allelic frequencies in a population at the equilibrium. "p" and "q" refer to allelic frequencies for the alleles (A) and (a) respectively, where q=1-p.

Mating (selfing and inbreeding)

Mating between organisms within populations is an important point to understand population evolution, particularly in plants in which selfing is common and because plants are immobile and thus more susceptible to be affected by consanguinity. Mating affects genotypic frequencies, by decreasing heterozygosis (i.e. the proportion of heterozygotes in the population), without affecting allelic frequencies. Selfing drastically affect genetic diversity, by decreasing the frequency of heterozygotes across generations. In a population where all individuals are 100% selfing, heterozygosis decreases in proportion 1/2 each generation, **figure 33**.

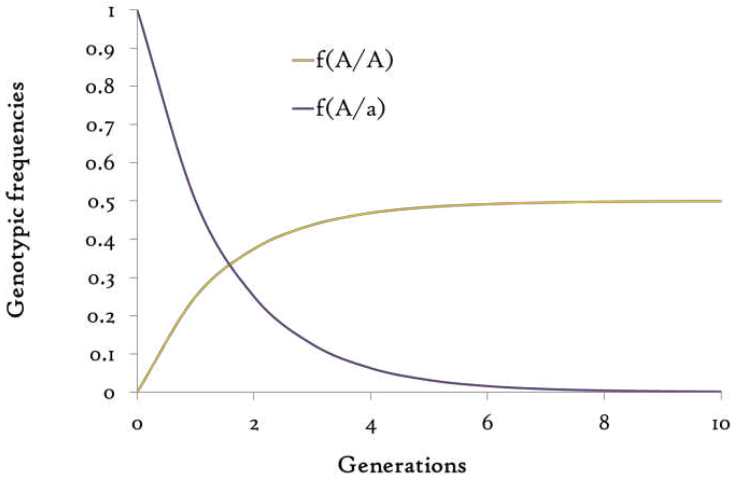


Figure 33: Variations in genotypic frequencies across generations in a theoretical 100% selfing population of constant population size (each organism produces one descendant each generation).

Inbreeding also affects heterozygosis, depending on the inbreeding coefficient (Fis). One definition of the inbreeding coefficient is the difference between the expected (under Hardy Weinberg equilibrium) and the observed heterozygosis divided by the expected heterozygosis:: $Fis = (He - Ho)/He$ (Hartl & Clark 2007). A positive *Fis* indicates a deficit in heterozygotes due to inbreeding, whereas a negative value indicates an excess of heterozygotes. In populations, inbreeding affect genotypic frequencies, given $f(A/A)=p^2+pq*Fis$, $f(A/a)=2pq-2pq*Fis$ and $f(a/a)=q^2+pq*Fis$, **figure 34**.

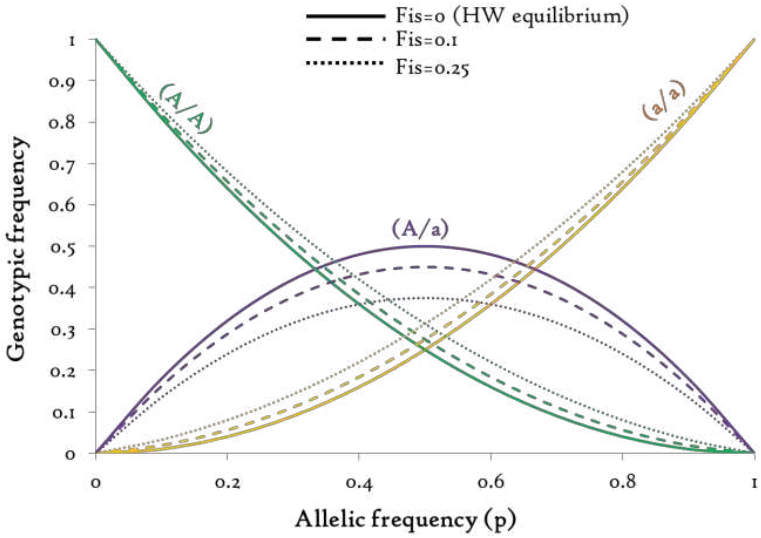


Figure 34: Genotypic frequencies as a function of allelic frequency under the HW equilibrium (Fis=0) and in populations submitted to inbreeding (Fis=0.1 and Fis=0.25).

BOX 1 – THEORY

Apprehending genetic diversity from AFLP markers

"Amplified fragment length polymorphism" (AFLPs) is a powerful method to detect polymorphisms in populations, by allowing the analysis of numerous markers spread in the genome very quickly with a limited cost. The technique consists in digesting the genome with one or several enzymes (frequently two enzymes) and by amplifying digested fragments by PCR. The amplification of a fragment produces a band by genotyping, whereas the absence of band traduces a polymorphism that prevents enzyme clipping at this site (Vos *et al.* 1995). Thus, AFLPs are poorly informative dominant markers: even if the absence of a band necessary traduces a homozygote (0/0), the presence of a band confounds homozygotes (1/1) and heterozygotes (1/0). Thus, estimating allelic and genotypic frequencies in populations from AFLPs requires either the assumption that the population is at equilibrium, or a prior knowledge about the inbreeding coefficient (Fis) in populations estimated from other kinds of molecular markers (such as SNPs).

For each marker j, the frequency of homozygotes (0/0) is estimated by:

$f(00)_j = q^2 + pq * Fis$

where p and q expresse the frequency of the allele (1) and (0) respectively, with p=1-q, leading to:

$f(00)_j = (1 - Fis) * q_j^2 + (Fis * q_j) \leftrightarrow (1 - Fis) * q_j^2 + \left(Fis * q_j\right) - f(00)_j = 0$

Thus, $q_j = \frac{-Fis+\sqrt{\Delta_j}}{2*(1-Fis)}$ with $\Delta_j = Fis^2 - [4 * (1 - Fis) * \left(-f(00)_j\right)]$

$N(01)_j = 2 * N_j * q_j - 2 * N(00)_j$ where N_j corresponds to the number of phenotypes available for this marker (with removal of missing values).

This method was applied for estimating genotypic frequencies from the AFLP dataset obtained during this PhD (**Article n°2**). A mean *Fis* was estimated from genes sequenced during the candidate gene approach (**Article n°1**).

In the species-rich tropical rainforest, numerous tree species are monoecious and occur at low population densities. This observation originally led botanists to predict that tree species should be highly self-fertilizing and inbred. However, recent investigations revealed that dioecy is consistently more frequent in tropical than in temperate trees (>20% of tropical tree species, Ward *et al.* 2005, Dick *et al.* 2008), while estimates of outcrossing revealed that tropical tree species are mainly outcrossing (Ward *et al.* 2005), **figure 35**. High outcrossing rates, even in hermaphrodic species, may be a result of incompatibility mechanisms preventing selfing, and inbreeding depression (i.e. the fitness of selfed seedling is lower than the fitness of outbred seedlings, see the following paragraph on 'natural selection'). However, mixed mating remains frequent in several species. Outcrossing depends on the balance between pollen dispersal (see paragraph on 'gene flow') and distance

between crowns. Thus, selfing is favored in populations of very low density, whereas outcrossing is favored by high population density. However, in species with an aggregative distribution -as it is the case in *Eperua falcata*- mating would occur principally among neighbors, leading to local inbreeding between trees (Dick *et al.* 2003).

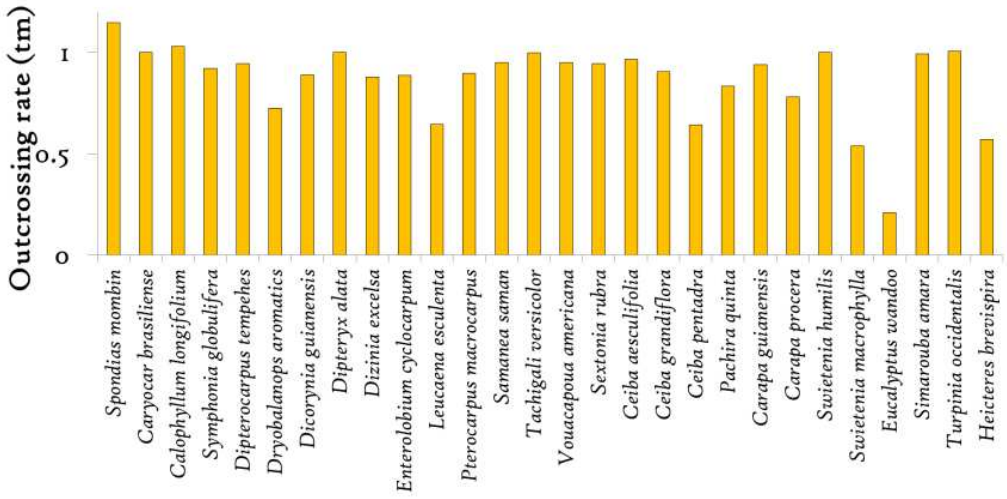


Figure 35: Estimated outcrossing rates (tm) in several tropical tree species (From Ward *et al.* 2005, Hardy *et al.* 2006, Dick *et al.* 2008 and all references within).

Genetic drift

In a **finite population**, the random sampling of gametes causes variations in allelic frequencies across generations. Genetic drift may be modeled by the Wright-Fisher model in which each generation is constructed by random sampling from a pool of gametes. Alleles frequencies vary randomly across generations, leading either to allele fixation (p=1) or to allele loss (p=0). In small populations, allelic frequencies show strong variations across generations, and allelic fixation or loss occurs more quickly than in large populations, **figure 36**. In general, trees are characterized by high fecundity (by comparison with animals) leading to large population sizes (Petit & Hampe 2006). Thus, genetic drift is supposed to be low in continuous tree populations. However, low population densities encountered in numerous tropical trees, as well as frequent asynchronism of flowering among trees of a given species, may reinforce genetic drift (Ward *et al.* 2005, Dick *et al.* 2008).

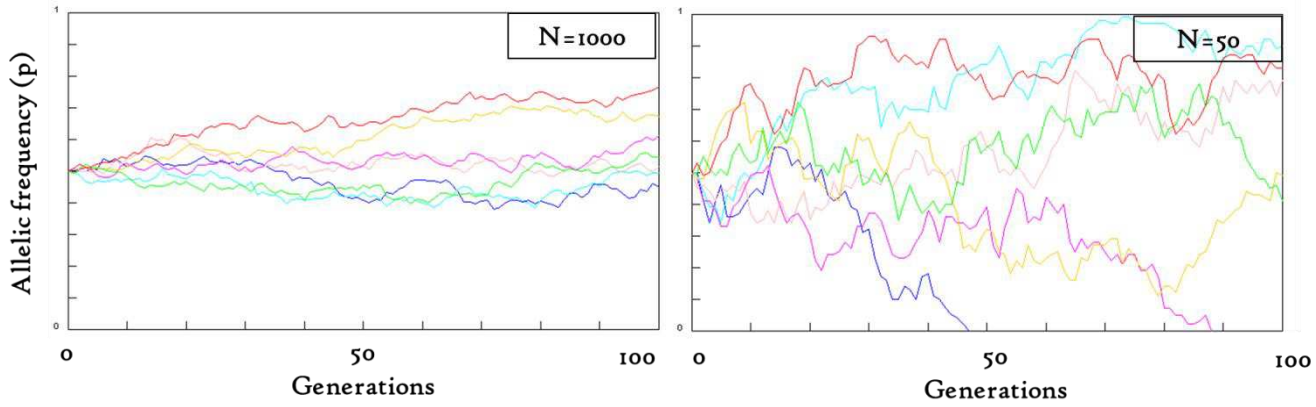


Figure 36: Allelic frequencies (p) for 7 loci under the Wright-Fisher model simulated using the simulation engine available at: http://darwin.eeb.uconn.edu/simulations/drift.html

Gene flow

Gene flow refers to the movements of genes within or between populations. In plants, gene flow occurs through movements of haploid gametes (**pollen flow**), and diploid zygotes (**seed dispersal**). Moreover, gene flow is not only a function of dispersal, but also the success of migrants in different habitats (i.e. natural selection directly impacts the 'realized' gene flow). Gene flow is commonly estimated through paternity and maternity tests or indirectly inferred from the analysis of the fine-scale genetic structure of populations (see section 'Neural genetic differentiation'). Trees display high levels of gene flow in both temperate and tropical ecosystems (Petit *et al.* 2006, Savolainen *et al.* 2007), and pollen flow is globally higher than seed dispersal in the latter (Dick *et al.* 2008), **figure 37**.

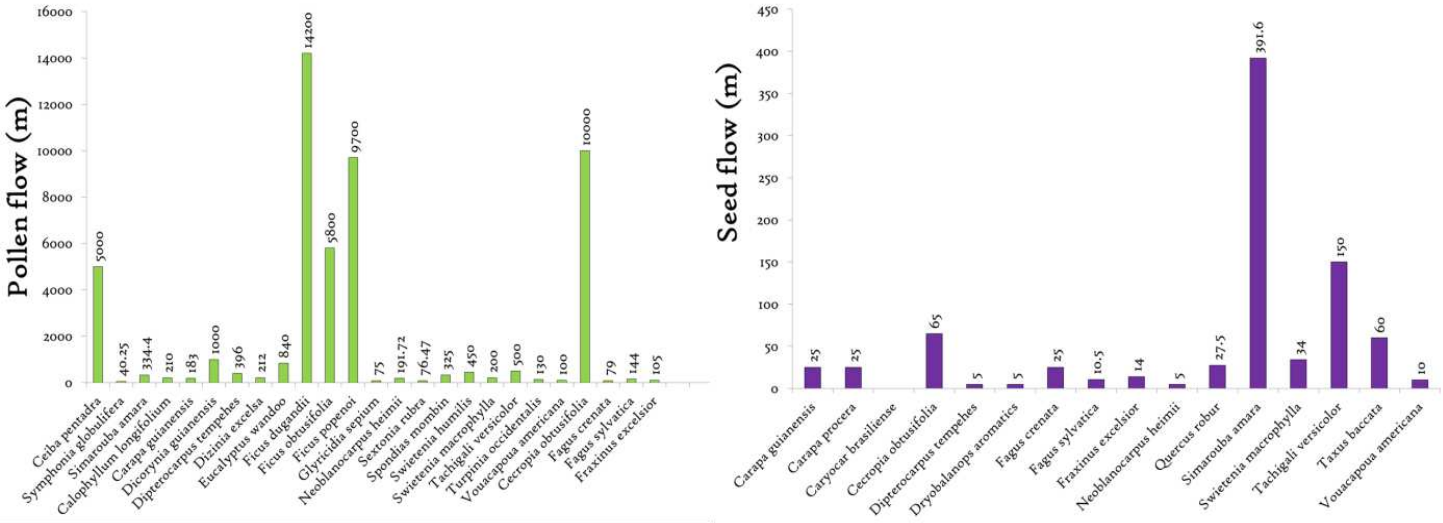


Figure 37: Estimates of pollen flow and seed dispersal in several temperate and tropical tree species (realized from Ward *et al.* 2005, Hardy *et al.*2006, Petit & Hampe 2006, Dick *et al.* 2008, and references within).

Pollen flow is high in tropical tree species, even in animal-pollinated species, and ranges from 200m to 19km (Ward *et al.* 2005). Indeed, tropical tree species are mainly

animal-pollinated (70% of the species), because high air humidity and frequent precipitations prevent wind-pollination. One of the rare examples of wind-pollination is provided by the pioneer species from the genus *Cecropia*, able to disperse pollen to several kilometers (6-14 km in *C. obtusifolia*, Kaufman *et al.* 1998), **figure 38**. Other examples of long-distance pollen flow are provided by the bat-pollinated *Ceiba pentandra*, able to disperse pollen up to 18 km, and the wasp-pollinated species from the *Ficus* genus able to spread pollen from 6 to 15 km (Nason *et al.* 1998). However, long-distance pollen flow is not the norm for all tree species, and the extent of gene flow may be modulated by population density and species behavior as evoked in the 'mating' paragraph. Moreover, habitat fragmentation may increase pollen flow, suggesting that tropical tree species would be more adaptable to forest fragmentation than expected. In *Swietenia humilis*, White *et al.* (White *et al.* 2002) reported that pollen flow was 10 times larger in fragmented habitats than previous results in undisturbed populations. In the same way, Dick *et al.* (Dick *et al.* 2003) found strong differences in pollen dispersal between undisturbed (mean = 212 m) and fragmented habitats (mean = 1509 m) in the African tropical tree *Dizinia excelsia*.

Figure 38: *Cecrobia obtusifolia* (Urticaceae)

Contrary to pollen flow, **seed dispersal** occurs principally at local scale in tropical rainforests and is commonly below 100 m, with a maximum at ~400 m in *Simarouba amara* (Hardesty *et al.* 2006). Hardy *et al.* (Hardy *et al.* 2006) found a relation between seed dispersers and total gene flow (including both pollen and seed dispersal): tree species dispersed by monkeys or birds have more effective gene flow than trees dispersed by gravity (as it is the case for *Eperua falcata and E. grandiflora*), wind or rodents. They also suggested that limited seed dispersal would indirectly affect pollen dispersal by increasing local population densities. Moreover, rare events of extreme long-distance dispersal have already been reported and, even if rare, such extreme dispersal may be involved in the colonization of new areas. For example, a cross-Atlantic dispersal event would have allowed the

Figure 39: *Ceiba pentandra* (Malvaceae) called 'fromager' (literally 'the cheese maker') due to its soft wood.

colonization of Africa from the neotropics by *Ceiba pentandra* (Dick *et al.* 2007), **figure 39**.

Natural selection

Natural selection acts on genetic diversity through fitness-related phenotypic traits. Contrary to genome-wide neutral processes, natural selection acts on targeted genes involved in fitness-related traits, and affects the frequency of alleles in a population across generations. Fitness may be defined as the property for a genotype to survive and produce a fertile progeny. Mathematically, fitness is the ratio between the number of descendant produced by a given genotype and that produced by the genotype with the greater fitness. For a bi-allelic locus, fitness is called W_{AA}, W_{Aa} and W_{aa} for the genotypes (AA), (Aa), and (aa). Genotypic frequencies at the following generation is thus (Hamilton 2009):

$$f(genotype)_{t+1} = \frac{f(genotype)_{t+1} \times W(genotype)_t}{\bar{W}}$$

where $\bar{W}$ traduces the marginal fitness or the frequency-weighted relative fitness of genotypes: $\bar{W} = f(AA)_t \times W(AA) + f(Aa)_t \times W(Aa) + f(aa)_t \times W(aa)$

Thus, the future for an allele under selection may be easily guessed. Variations in the frequencies of the allele (A) depend on the fitness of the different genotypes, **figure 40**. The frequency of the allele (A) increases after selection if homozygotes (AA) are favored but decreases if homozygotes (aa) are favored, **figure 41**. When heterozygotes have the greatest or the lowest fitness, variations of allelic frequencies must be either positive or negative, depending on allelic frequency before selection. If homozygotes have equal fitness, selection will lead to equilibrating allelic frequencies around 0.5 (if no drift). However, natural selection may also impact neutral loci (leading any advantage or disadvantage to the different genotypes) because of a physical linkage between them (**'Hitchiking'**).

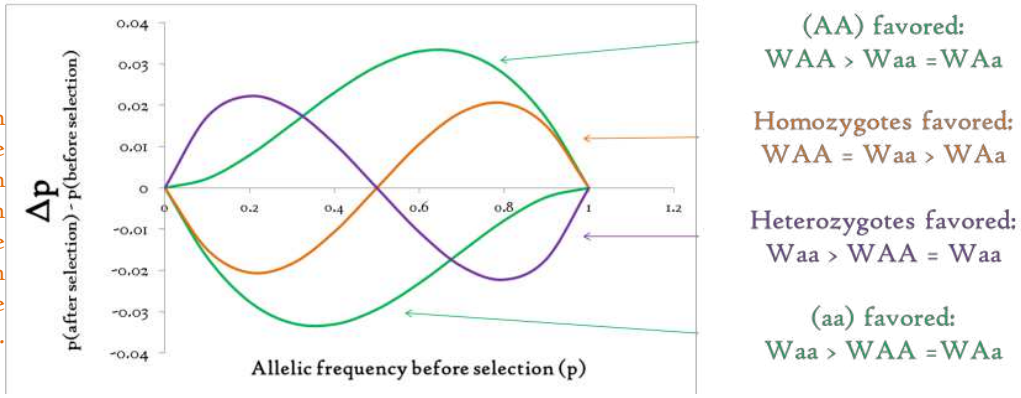


Figure 40: Variation in the frequency of the allele (A) after selection ('Delta p') as a function of the frequency of the allele (A) before selection ('p') and the relative fitness of the genotypes.

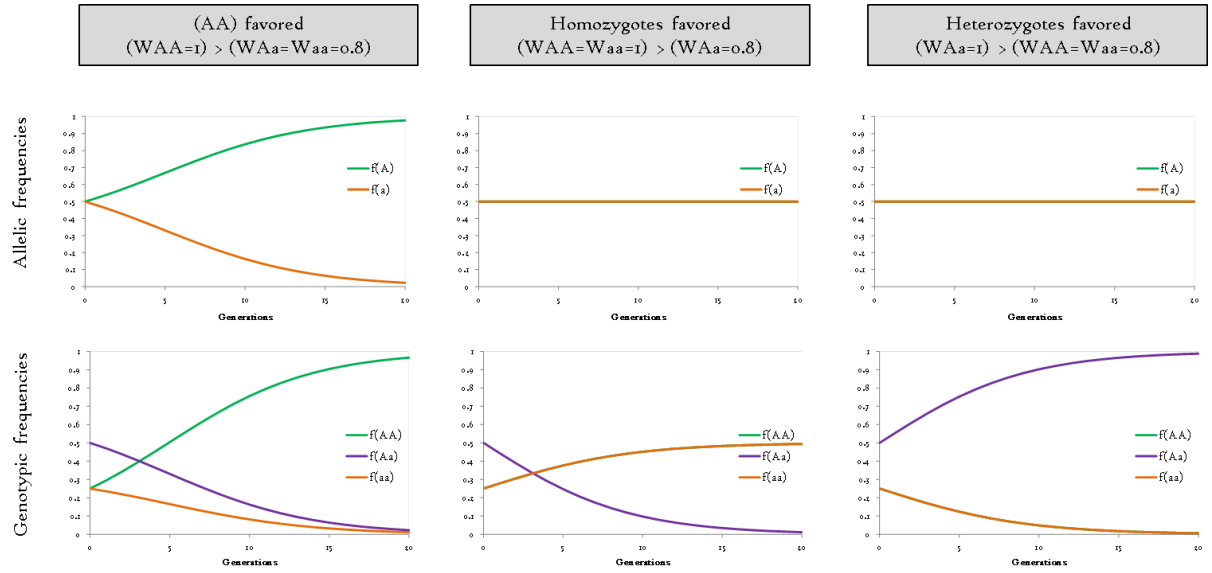


Figure 41: Variations of allelic and phenotypic frequencies across generations *under selection (without genetic drift)*

Three kinds of natural selection may be distinguished. **Positive selection** favors an advantageous allele that will increase in frequency across generations until fixation, **figure 42.** Under selection, allele fixation is expected to occur quickly than with drift only. **Negative (or purifying) selection** eliminates deleterious mutations until their complete disappearance. Both positive and negative selection leads to an excess of rare alleles at a polymorphic locus by comparison with neutral expectations. **Balancing selection** favors several alleles of equal contributions, leading to an excess of alleles in intermediate frequencies than expected under neutrality. The **figure 42** shows a conceptual allele frequency spectrum at a single locus that represents the patterns of allelic frequencies at a locus submitted to natural selection by comparison with the expected pattern under neutrality of this locus.

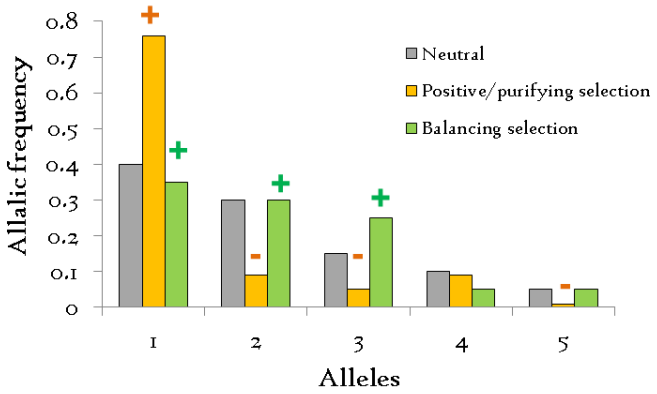


Figure 42: Conceptual allele frequency spectrum at a multi-allelic locus under neutrality and under selection.

Traditionally, two main approaches allow searching for footprints of natural selection: '**candidate gene**' and '**genome scan**' approaches. In the former, footprints of natural selection are searched at individual loci using single-locus selection tests, classically based on analyzing levels of diversity or allele-frequency-spectrums (for example Tajima's test), **figure 43**. Targeted genes are empirically chosen based on prior knowledge or assumptions. These genes may either be quantitative trait loci (QTLs, i.e. loci involved in variations of phenotypic traits) or genes encoding for proteins involved in a candidate metabolic pathway or biological process (and eventually expressed in large amount in response to particular constraints). The latter (genome scan) involves the analysis of numerous molecular markers, with no necessary known function. It starts from the hypothesis that the majority of polymorphisms in the genome is selectively neutral (**box 2**) and that genetic diversity at a locus submitted to selection would be different from the global genetic diversity apprehended overall genome. Tests for selection based on genome scans allow identifying outliers by characterizing the (neutral) distribution of particular statistics among loci, mainly linkage disequilibrium, synonymous/non-synonymous ratio of mutations, or differentiation (Fst), **figure 43**. This approach is probably being the most popular because it allows identifying footprints of natural selection free from neutral processes with genome-wide effects (such as demographic changes). Moreover, next generation sequencing, genotyping, and re-sequencing technologies are going to merge these two approaches, as they provide genetic information about large numbers of loci of known function (see part 3. 'Next-generation sequencing, genotyping and new opportunities').

Demography

Test category	Signature detected	Limitations
Level of diversity	Unusually low or high genetic diversity around the selected locus	High sensitivity to demographic assumptions
Site frequency spectrum (SFS) based-test	Modification in the relative proportions of low and high frequency mutations in the selected region	High sensitivity to demographic assumptions. High rate of false positives
Linkage disequilibrium (LD)	A rise in frequency of long haplotypes created by the increased LD around the selected region	Spurious signal of selection created by population structure. LD levels decrease rapidly after selective sweep is complete
Synonymous/nonsynonymous mutations	Differences between the ratio of nonsynonymous to synonymous polymorphism and nonsynonymous to synonymous divergence	Cannot distinguish between past and current selection. Slightly deleterious mutations inflate polymorphism. Spurious signal of selection with population expansion and bottlenecks if there are slightly deleterious mutations
Population differentiation	Increased or decreased population differentiation of a genomic region relative to the rest of the genome	Hierarchical genetic substructure creates false positives. Importance of the sampling scheme

Figure 43: Overview of the methods for detecting selection (Siol, *et al.* 2010).

The genetic structure of populations is also influenced by past (eg. ice ages), recent (eg. pre-colombian occupation), or current demographic changes. In French Guiana, Stephanie Barthe (2012) found large demographic changes concordant with past climatic changes (**figure 44**), but she did not detect demographic variations concordant with pre-colombian human occupation, Barthe 2012. She also found that several tree species may have different demographic history among regions, as it is the case for *Vouacapoua americaana*.

Demography attracts a particular attention, not only in biogeographic but also in

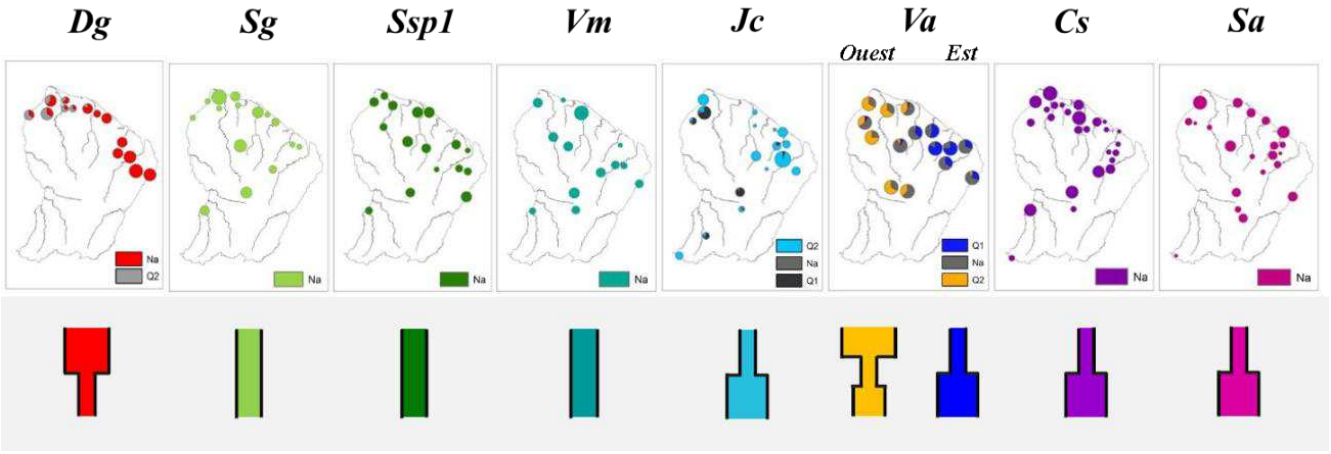


Figure 44: Past demographic events in several species of French Guiana. From Stéphanie Barthe (2012). Maps indicate the structuring of genetic diversity at regional scale, schemes indicate which demographic scenario were experienced by the populations of different species (constant size, bottleneck, or expansion).

adaptation studies. Indeed, the demographic history of populations mimics the effects of natural selection at a given locus. A population expansion commonly produces an excess of rare alleles, and may be confounded with positive and purifying selection. On the contrary, a population decrease ('bottleneck') produces an excess of alleles in intermediate frequencies and may be confounded with balancing selection. That is why genome-scan approaches are highly powerful to identify targets of natural selection, because they identify loci free from genome-wide neutral processes such as past demographic changes.

BOX 2 – THEORY

Adaptive and Neutral evolution: from Darwin to Kimura

Charles Darwin (1809-1882) was one of the first to accept that species can change, originally from the observation that species phenotypic traits vary geographically in function of the environment they inhabit. He was the first to propose that favorable variations tend to be preserved, whereas unfavorable ones tend to be destroyed through selection. The hypothesis of adaptation was evoked by Darwin while **Gregor Mendel** (1822-1884) was discovering the principles of inheritance, and Darwin's theory lacked, at this time, a satisfactory theory of heredity. During the 20th century, Ronald Fisher, J. Haldane and Sewal Wright showed that natural selection operates through Mendelian inheritance, reconciling Mendelism and Darwinism in the 'synthetic theory of evolution' or 'modern synthesis'. Neo-Darwinism became widely accepted, even if most attention was focused on adaptation. Several authors, however, introduced other evolutionary forces as drivers of evolution: The Wright-Fisher model of random fixation introduced the notion of genetic drift (**figure 45**), while Dobzansky and Huxley introduced gene flow as they suggested that geographically separated populations would evolve into new species.

At the end of the 20th century (1983) **Moto Kimura** properly wrote the neutral theory of evolution (Kimura 1983). He proposed that evolution is in majority driven by neutral processes, mainly mutation and drift. His theory is based on the idea that the majority of polymorphisms are selectively neutral: while deleterious mutations are quickly eliminated by purifying selection, numerous mutations affect un-coding regions (introns) or numerous mutations affect coding regions without affecting the amino acid encoded (silent mutation) or without affecting the protein function (conservative mutation).

However, he didn't completely exclude the impact of natural selection, as he proposed that only several highly deleterious mutations are eliminated by natural selection. This argument was used to explain why:

(1) Non-coding DNA regions (introns) generally accumulate mutations more easily than coding ones (exons).

(2) Proteins constitutive for cells never mutate and mutations rarely affect the active sites of proteins (implying that natural selection prevents all deleterious mutations in genes encoding constitutive proteins or in protein active sites).

Now, it is widely accepted that the majority of polymorphisms observed in the genome is selectively neutral. However, it is not excluded that several loci would be modeled by natural selection. More, the conciliation between adaptive and neutral theory has led to the emergence of a new method for identifying selected loci and apprehending the extent of natural selection: the genome scan approach.

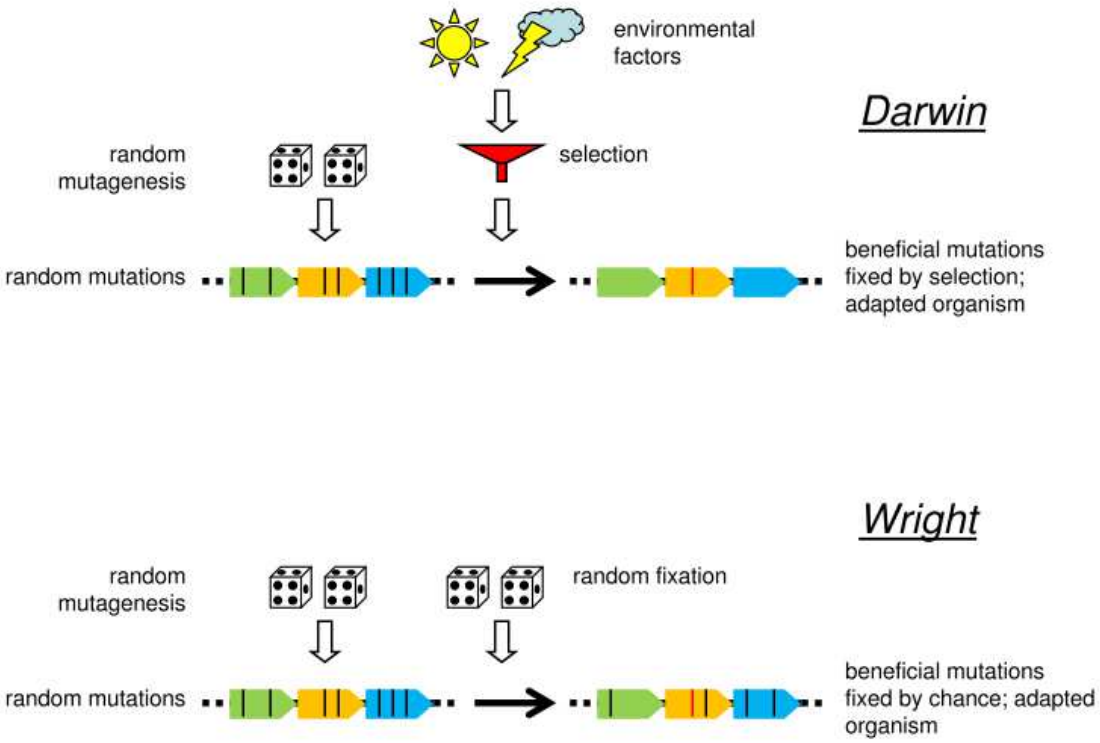


Figure 45: Evolution according with Darwin and Wright. From Koonin & Wolf 2009.

2. Population differentiation: the complex interplay between gene flow, selection, and drift.

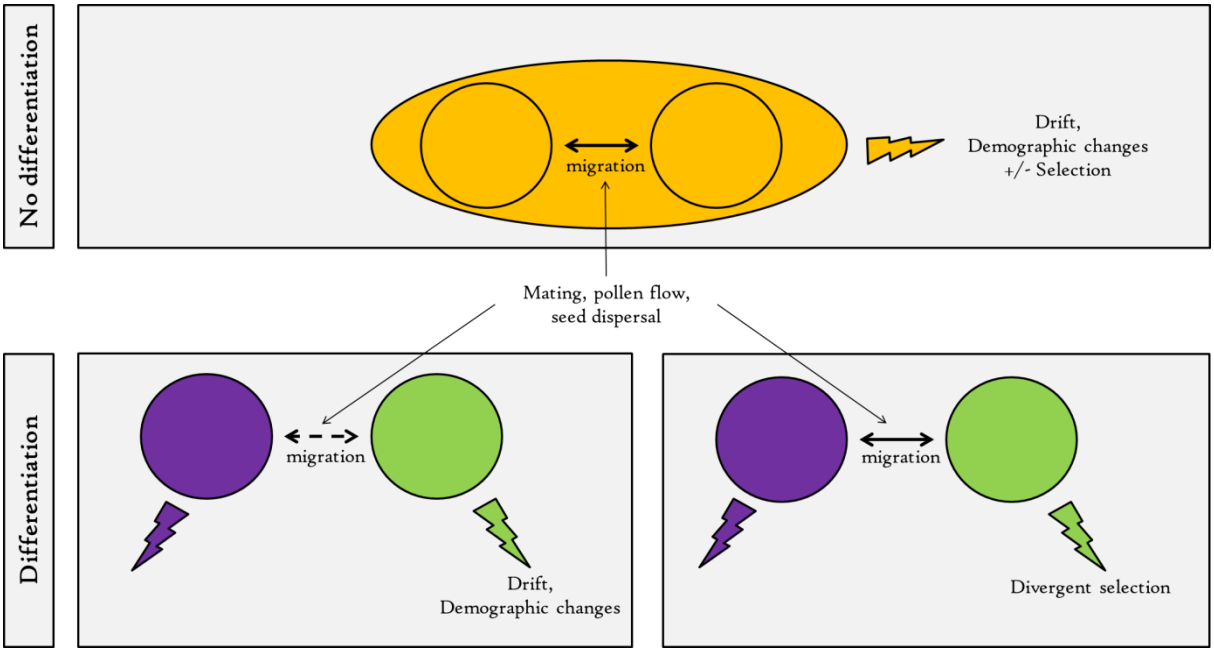


Figure 46: Conceptual framework of population differentiation.

Genetic differentiation between populations (also called 'demes') results from the subtle interplay between evolutionary drivers, mainly gene flow, drift, and natural selection, **figure 46**. Migration (pollen flow and seed dispersal in the case of trees) tends to homogenize the genetic diversity among populations (Lenormand 2002, Bolinick & Nosil 2007). Thus, populations connected by an extensive gene flow in the absence of strong disruptive or directional selection are expected to be poorly differentiated: the entire population is thus submitted to drift, demographic changes and stabilizing selection, and these processes act similarly in the whole population, Ridley 2003, Hartl & Clark 2007, Hamilton 2009.

When migration is restricted between demes, they may diverge into sub-populations through the action of demographic events (if the sub-populations experienced different demographic history) and genetic drift (depending on the effective size of each sub-population): differentiation depends in this case on the couple **migration-drift** (see following paragraph 'Neutral differentiation').

However, gene flow does not necessarily prevent differentiation. In the particular case of **divergent selection** caused by habitat heterogeneity, natural selection may drive genetic differentiation in spite of low distances between sub-populations, because propagules from a particular habitat are unable to establish in others. This particular case is

called **local adaptation**, and is widely documented in both animal and plant species, Kawecki & Ebert 2004, Leimu & Fisher 2008, Savolainen 2007 b. Here, I define **"local adaptation" as the genetic divergence that occurs over shorter distances than potential gene flow due to divergent selection among contrasted habitats.**

To quantify the extent of population subdivisions, S. Wright (1921) defined the **fixation index** (Hartl & Clark 2007). This index expresses the reduction in heterozygosity expected (under Hardy-Weinberg equilibrium) at any level of a population hierarchy relative to another. In a hierarchical model of population subdivision, let H_S define the average Nei's index (heterozygosity expected under Hardy-Weinberg) within sub-populations, H_R the average Nei's index within regions, and H_T the Nei's index of the total study area. Three different **Wright's F-statistics** allow quantifying the extent of differentiation, **figure 47** (Hartl & Clark 2007, Excoffier *et al.* 2009).

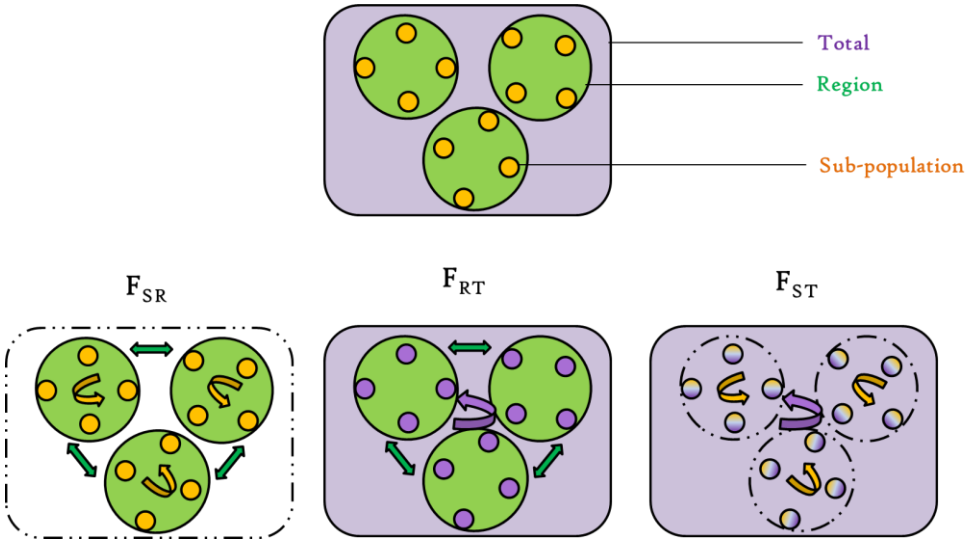


Figure 47: Wright's F-statistics of differentiation.

- F_{SR}: Differentiation between sub-populations within regions relative to differentiation between regions: $F_{SR} = \frac{H_R - H_S}{H_R}$ [1]
- F_{RT}: Differentiation between regions relative to the diversity in the total population: $F_{RT} = \frac{H_T - H_R}{H_T}$ [2]
- F_{ST}: Differentiation between sub-populations within regions relative to the diversity in the total population: $F_{ST} = \frac{H_T - H_S}{H_T}$ [3]

These three indexes are linked by the relation: $(1 - F_{ST}) = (1 - F_{SR})(1 - F_{RT})$ [4]

Comparing F_{RT} and F_{SR} allows assessing if there is more variation among regions (as measured by F_{RT}) than there is among sub-populations within regions (F_{SR}).

3. Neutral differentiation

Neutral genetic differentiation is commonly modeled under the couple **migration-drift** using migrations models (**figure 48**), among which the **island model** is probably the most popular.

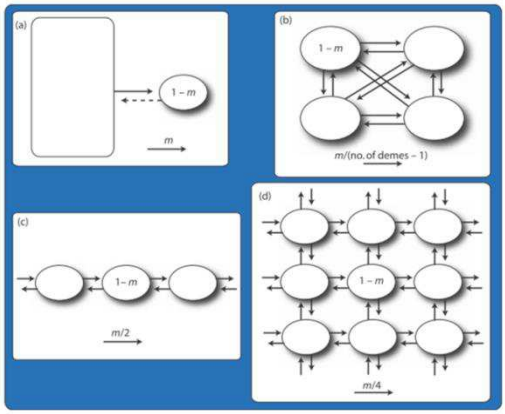


Figure 48: Classic models of population subdivision. (a) Continent-island model, (b) island model, (c) and (d) stepping-stone model in one or two dimensions. (From Hamilton 2009)

In the island model, a large population is split into many sub-populations (demes), and migration is assumed to be symmetrical between demes (the proportion of migrants from each deme into each other is thus 1/d, where d is the number of demes). Under the island model, the differentiation between sub-populations is explained as a function of Ne (effective size of the meta-population) and m (the migration load between demes): $F_{ST} \approx \frac{1}{1+4\times Ne\times m}$. It is thus intuitive that FST decreases as the migration rate (m) increases. FST also increases when Ne decreases, as a result of genetic drift within sub-populations. To illustrate the relationship between FST, Ne and m, I simulated theoretical populations using the 'Easypop' software (Balloux 2001), **figure 49**. In large populations, differentiation decreases when migration increases, leading to an absence of differentiation for m=0.05. Small populations are differentiated even for m=0.05 (Fst close to the Fst obtained without

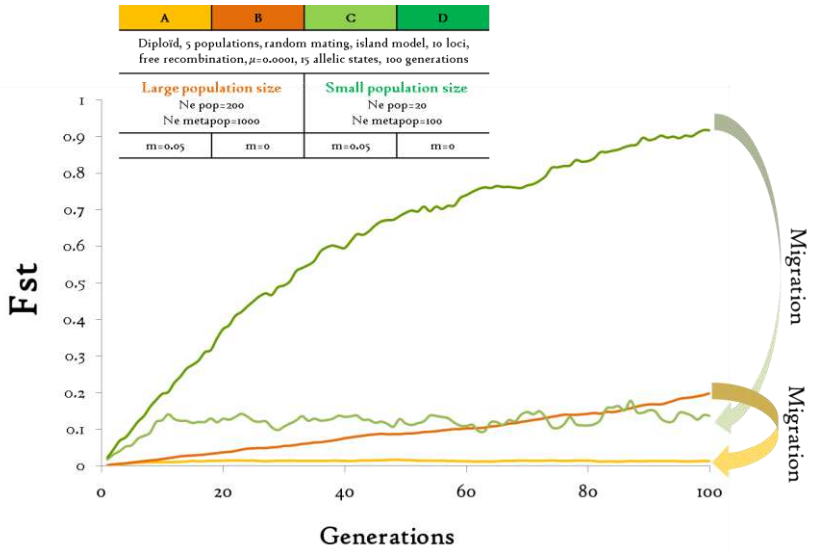


Figure 49: Differentiation between sub-populations obtained under different scenarios of effective population size and migration rate using the program 'Easypop' (Balloux 2001).

gene flow in large populations after 100 generations), and Fst bursts in the complete absence of migration (close to 1 after 100 generations), because isolated sub-populations evolve independently through strong genetic drift.

Dick *et al.* (Dick *et al.* 2008) reviewed the estimates of differentiation in both temperate and tropical tree species and found that tropical tree species display a larger genetic differentiation among populations (mean Fst = 0.177) than their counterparts in the temperate zone (mean Fst = 0.116). Strong genetic differentiation in tropical trees is probably the result of mixed mating systems, restricted seed dispersal and high levels of local inbreeding. Moreover, they suggested that the extent of differentiation is independent to the canopy stature of tree species (understory, canopy, emergent), **figure 50**.

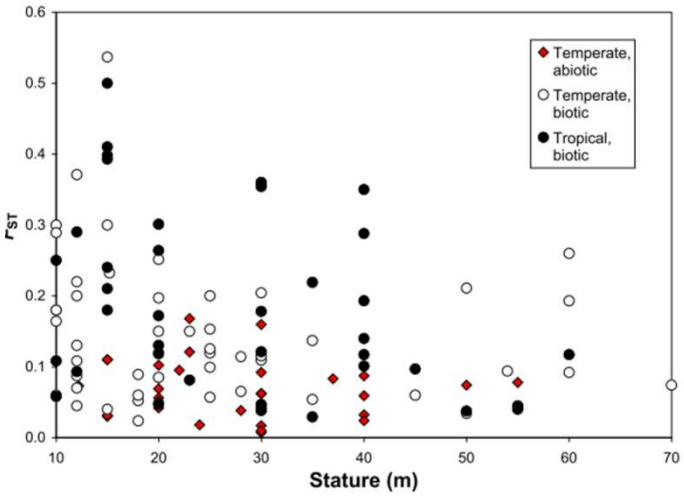


Figure 50: Fst estimates in tropical and temperate zone trees (Dick, 2008).

At fine spatial scales, a restricted gene flow due to limited seed dispersal, as it is frequently the case in tropical rainforest trees, may cause a spatial genetic structuring over short geographical scales, even in populations of large size, because progenies are geographically grouped (Dick *et al.* 2008, Hardy *et al.* 2006). Moreover, mating among neighbors in aggregative tree species causes local inbreeding that reinforces the spatial structuring, leading to strong genetic divergence over short spatial scales **figure 51**, Hamilton

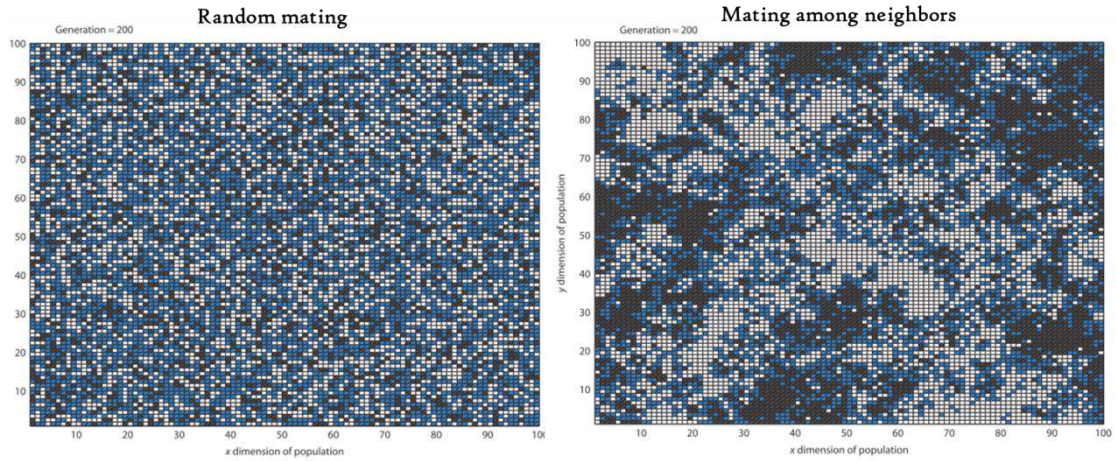


Figure 51: Mating among neighbors causes spatial clumping of genotypes and therefore clumping of allele frequencies (from Hamilton 2009). Genetic structuring after 200 generations under random mating (left) and mating among neighbors (right).

2009. Thus, assessing fine-scale spatial genetic structure (SGS, **box 3**) is one of the most popular methods to assess whether neutral processes drive the genetic differentiation in continuous areas (Hardy *et al.* 2006).

BOX 3 - METHODS

Fine-scale Spatial Genetic Structure (SGS)

Spatial Genetic Structure (SGS) analyses allow test **isolation-by-distance** hypothesis at short geographical scales. Fine-scale SGS is characterized by a decrease of relatedness (Kinship coefficient) between individuals with distance. The following **figure 52** describes how SGS is assessed in wild populations and how gene flow is estimated (Hardy *et al.* 2006).

Significant SGS is often interpreted as resulting from dispersal limitation, because seed dispersal may produce fine scale genetic structure even if pollen flow is long distance, because sibs are aggregated Dick *et al.* 2008.

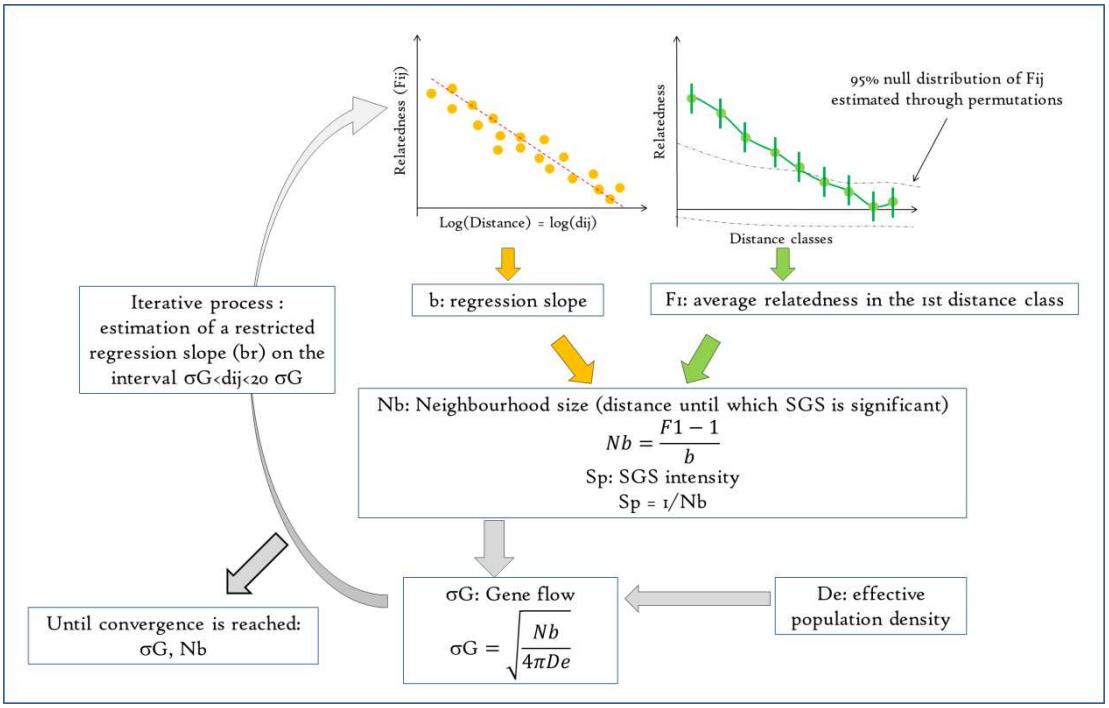


Figure 52: Method for analyzing fine-scale SGS and estimating gene flow.

Significant fine-scale genetic structuring is common in both temperate and tropical zones. In temperate tree populations for example, significant SGS had been reported in *Fagus crenata* (Oddou-Muratorio *et al.* 2010), *Fagus sylvatica* (Vornam*et al.* 2004, Jump *et al.* 2006, Oddou-Muratorio *et al.* 2010, Jump & Penuelas 2012), *Quercus petraea* and *Q. robur* (Streiff *et al.* 1998). The tropical zone also provides numerous examples of fine-scale genetic structuring in both dominant and pioneer tree species. I presented an overview of the main published papers in the following table. The most spectacular spatial structure was found in *Aucoumea klaineana* (Gabon) populations that display a significant relatedness between individuals up to 5 kilometers. A significant structure was also found in the insect-pollinated *Eperua grandiflora*, for which gene dispersal was estimated to ~320 meters in spite of its heavy seeds dispersed by gravity (Hardy *et al.* 2006).

Species	Family	Location	SGS	References
Aucoumea klaineana	Bursrseracea	Gabon	* (5km)	Born *et al.* 2008
Carapa guianensis	Meliaceae	Brazil	* (100m)	Cloutier *et al.* 2006, Cloutier *et al.* 2007
Carapa procera	Meliaceae	French Guiana	low	Doligez & Joly 1996, Hardy *et al.* 2006
Caryocar brasiliense	Caryocaraceae	Brazil	* (8m)	Collevatti *et al.* 2001, Collevatti *et al.* 2010
Cecropia obtusifolia	Moraceae	Mexico	*	Kaufman *et al.* 1998
Chrysophyllum sanguino lentum	Sapotaceae	French Guiana	*	Hardy *et al.* 2006
Dicorynia guianensis	Fabaceae	French Guiana	* (160m)	Latouche-Hallé *et al.* 2004, Cavers *et al.* 2005, Hardy *et al.* 2006
Dipteryx alata	Fabaceae	Brazil	ns	Collevatti *et al.* 2010
Eperua falcata	**Fabaceae**	**French Guiana**	*** (30-56m)**	**THIS STUDY**
Eperua grandiflora	Fabaceae	French Guiana	*	Hardy *et al.* 2006
Glyricidia sepium	Fabaceae	Guatemala	* (50m)	Dawson *et al.* 1997
Neoblanocarpus heimii	Dipterocarpaceae	Malaysia	low	Konuma *et al.* 2000
Sextonia rubra	Lauraceae	French Guiana	* (400m)	Cavers *et al.* 2005, Veron *et al.* 2005, Hardy *et al.* 2006, Cloutier *et al.* 2007
Swietenia macrophylla	Meliaceae	Costa rica	* (100-110m)	Lowe *et al.* 2003, Cavers *et al.* 2005
Symphonia globulifera	Clusiaceae	French Guiana	* (180-200m)	Aldrich & Hamrock 1998, Degen *et al.* 2004, Cavers *et al.* 2005, Hardy *et al.* 2006
Tibouchina papyrus	Maleastomataceae	Brazil	* (3m)	Collevatti *et al.* 2010
Vouacapoua americana	Fabaceae	French Guiana	* (100-300m)	Dutech *et al.* 2002, Hardy *et al.*2006

Several authors have already reviewed the processes of fine-scale genetic structuring

due to gene flow in tropical forests (Hardy *et al.* 2006, Ward *et al.* 2005, Dick *et al.* 2008). Taken together, it is now possible to draw a global scheme of the processes driving neutral differentiation over short spatial scales in tropical rainforests, **figure 53**.

During this PhD, I investigated the spatial genetic structure of *Eperua falcata* to assess whether the genetic diversity in wild populations may be influenced by neutral forces, mainly limited gene flow and local inbreeding. A synthetic summary of the main results is presented in the following page, while complete results are given in the **article n°2**.

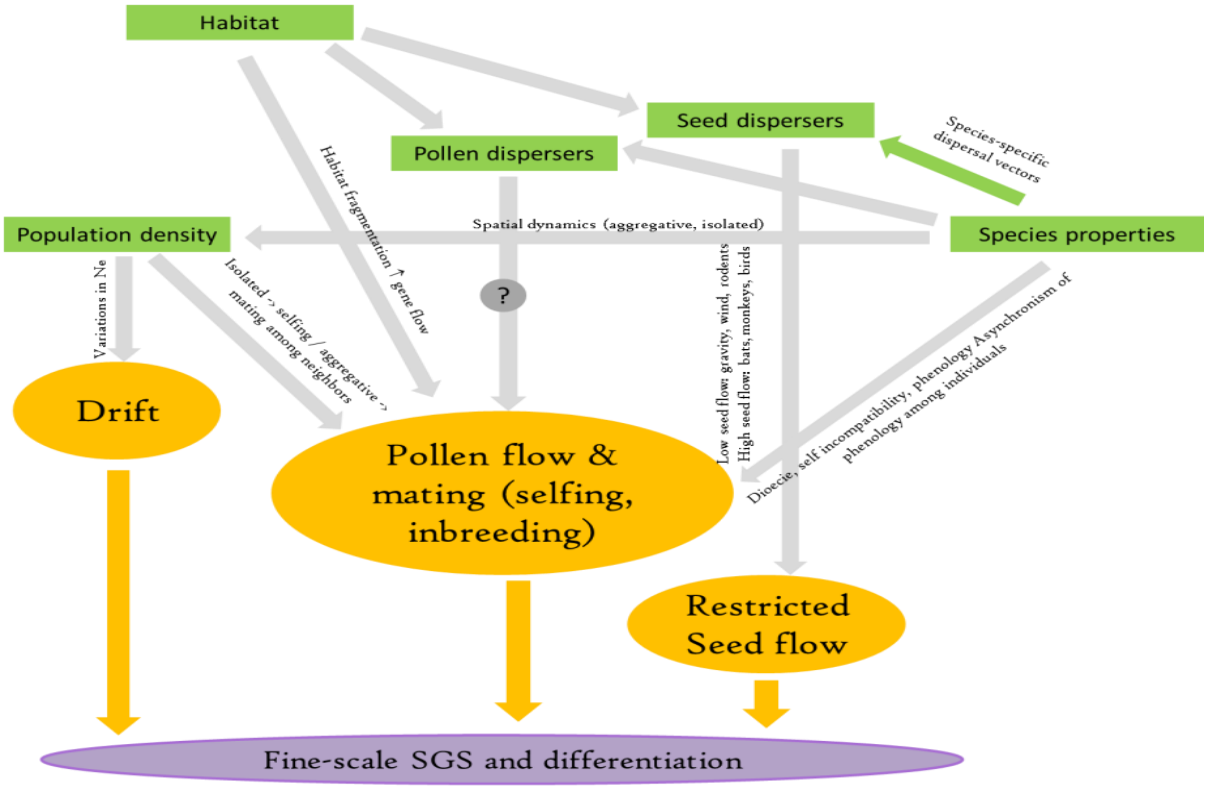


Figure 53: Process of neutral genetic differentiation and local structuring in trees.

Summary of PhD results:

Limited gene dispersal and local inbreeding causes the (neutral) genetic differentiation over short spatial scales in *Eperua falcata* (see article n°2).

I investigated the spatial genetic structuring in two populations of *E. faclcata* spread across different local habitats types (Laussat & Regina) using AFLPs data (1196 markers), **figure 54**.

Spatial autocorrelations revealed significant relatedness between individuals until 30 and 56 meters for Regina and Laussat respectively, **figure 55**. Gene flow varied between 45.7 (Laussat) to 64.4 meters (Regina). Surprisingly, gene flow estimated here was lower than gene flow in the congeneric species *Eperua grandiflora* (323 meters, Hardy *et al.* 2006) in spite of the heavy seeds of the latter.

Eperua falcata is an aggregative species that generally occurs in high population densities. Mating among neighbors probably influences high levels of local inbreeding in this species. Seeds are dispersed through an explosive dehiscence of pods around crowns of maternal trees, leading to a highly restricted gene dispersal and strong genetic structuring over short spatial scales due to clumping of the progenies, **figure 56**.

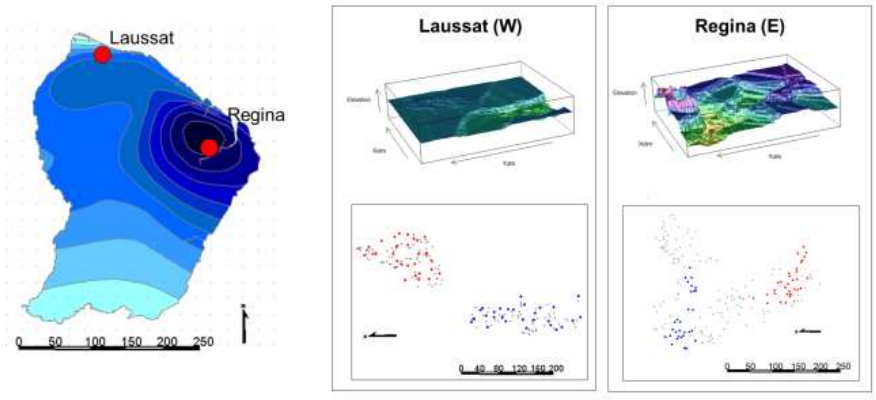


Figure 54: Study sites locations, local topography and Eperua falcata distribution (red=sampled trees from terra-firma, blue=sampled trees from bottomland).

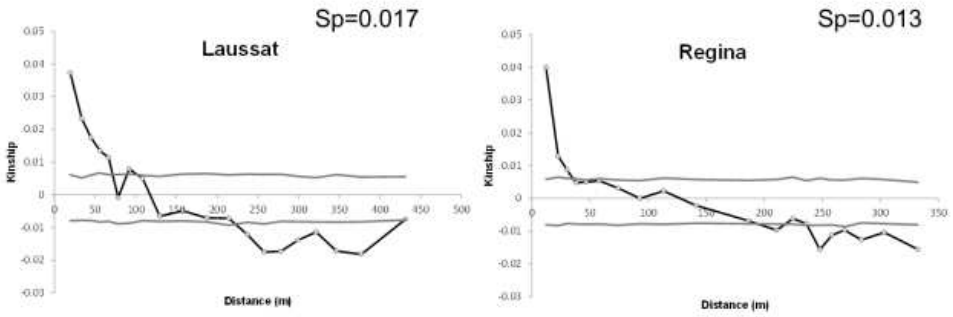


Figure 55: Spatial autocorrelations using Kinship relatedness coefficients based on 1711 and 1810 tree pairs in Laussat and Regina respectively.

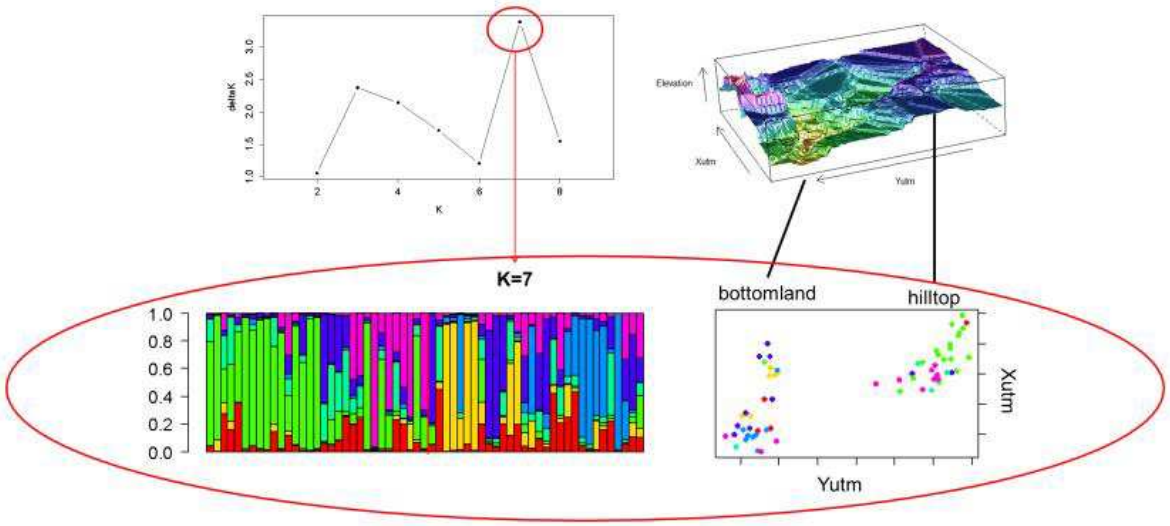


Figure 56: Blinded analysis of genetic structuring in Regina. The most probable number of genetic clusters was found for K=7 clusters (maximum deltaK) that probably corresponds to different progenies or clusters of related trees. The different clusters are geographically clumped, suggesting the existence of neutral structuring over short geographical scales due to restricted dispersal probably reinforced by local inbreeding.

4. Adaptive differentiation

Methodological considerations

In populations spread across contrasted habitats, natural selection may drive genetic differentiation in spite of low distances between the sub-populations. Adaptation to local environment has been observed experimentally in many organisms, but the genetic basis of local adaptation remains poorly known. Two main approaches are commonly used to identify molecular footprints of **divergent selection**:

- **'Genetic-environment associations' (GEAs)** search for significant relationships between (quantitative) environmental variables and allelic frequencies, often using **candidate genes** (Bierne *et al.* 2011).
- **'Fst-based methods'** uses **genome scans** to identify **'outlier' loci** for which the observed differentiation between (discrete) populations is different from the overall (and supposed selectively neutral) genetic differentiation. A central tenet of Neo-Darwinism is that evolution of adaptive traits involves allelic substitutions with small effects for a large number of loci. However, experimental studies revealed that the number, size and distribution of such genomic regions varies substantially among studies: several studies have provided cases in which adaptation is attributable to a small number of genes with large effects, while other studies demonstrated adaptations for a large number of genes of small effects. Indeed, the extent of natural selection is dependent on the genetic architecture of the selected traits, and genome-scans at molecular levels allow assessing the extent of natural selection without information about the phenotypic traits targeted by natural selection (Storz 2005).

Because I used Fst-based methods for studying local adaptation in *Eperua falcata*, I devote more methodological attention to Fst-based selection tests than to other methods such as GEAs.

Two kinds Fst-based methods can be distinguished:

- **Coalescent methods** use coalescence to simulate populations and draw an expected joint distribution of Fst vs. heterozygosity that may be directly compared with observations **(box 4)**.
- **Bayesian methods** partition observed Fst into a population- and a loci-specific component **(box 5)**. These coefficients are directly inferred through a Bayesian modeling approach.

Fst-selection tests of genome-scan data have several advantages, by comparison for example with QTL mapping (Storz 2005): (1) Genome scans may be applied to natural populations of any species (not restricted to species that can be crossed in the lab. (2) While QTL mapping in crossed lines typically found loci with large effects, molecular genome scans are capable of identifying loci that have experienced a weak selection (loci with small effects), Stapley *et al.* 2010. Indeed, natural populations result from long-time selective pressures. The cumulative effects of small selections over long times can produce a signal detectable by genome-scans. (3) They allow identifying selected loci without having information about the selected traits. (4) They are robust to a range of non-equilibrium situations. (5) Even if the original coalescent method proposed by Beaumont & Nichols (1996) is sensitive to bottlenecked populations, Bayesian methods (Beaumont & Balding 2004) are robust to many demographic scenarios (Beaumont 2005).

However, these methods have also several limits, mainly because they may detect several false-positives (type I error), or fail to detect true-positive (false-negative or type II errors). Type I errors (false-positive) -that is probably a more serious risk than type II errors- may have different causes. First, correlated allele frequencies due to co-ancestry may inflate the differentiation (Fst) under the island model. That is why, the Bayesian method developed by Foll & Gaggiotti (Foll & Gaggiotti 2008) uses a model of genetic differentiation with co-ancestry (inspired from Falush *et al.* 2003) that allows admixture between lineages. Moreover, genetic incompatibility (i.e. the intrinsic incompatibility between genetic groups) that constitutes an intrinsic barrier of gene flow may also inflate differentiation. Last, the use of a classical island model in the case of hierarchically structured sub-populations may also bias the analysis (because migration between demes of a same region is expected to be larger than migration between regions). That is why, Excoffier *et al.* (Excoffier *et al.* 2009, Excoffier & Lischer 2010) implemented the software 'Arlequin' with a hierarchical island model able to deal with such hierarchical designs.

Independently, Narum & Hess (2011) and Vilas *et al.* (Vilas *et al.* 2012) tested the power of different methods by simulating populations with a known number of genes under selection. They found that Bayesian methods ('BAYESCAN' software) perform more efficiently than coalescent ones ('FDITST' and 'DFDIST' software), Vilas *et al.* 2012. Moreover, 'Arlequin' (coalescent method under a hierarchical island model) produced more type I and type II errors than 'FDIST2' (coalescence under an island model) and 'BAYESCAN', Narum & Hess 2011. Last, the Bayesian method produces the lowest number

of Type I errors, Narum & Hess 2011. In spite of these limits, these methods remain the best way to identify targets of divergent selection in wild populations, and crossing results from different approaches is recommended.

BOX 4 - METHODS

Fst-based selection tests: COALESCENT methods

Coalescent methods draw an expected distribution of Fst under particular migration and mutation models through coalescent simulations **figure 57**. One of the most popular migration model is the hierarchical island model, in which sampled trees are organized into D demes from K groups that exchange genes. Each deme is made of N diploid organisms. At each generation, each deme (d) receives a proportion m1 of the 2N copies from other demes, or m1/(d-1) from each other demes. Similarly, each group (k) receives a proportion m2/(k-1) of gametes from each of the other groups (m1>>m2). In 1991, Slatkin (Slatkin & Vloem 1991) expressed Wright's F-statistics as a function of coalescent time:

$$Fsr = \frac{t_1 - t_0}{t_1}\,;\, Frt = \frac{t_2 - t_1}{t_2}\,;\, Fst = \frac{t_2 - t_0}{t_2}$$

where to is the mean coalescence time of two genes from the same deme, t1 is the mean coalescence time of two genes from the same group and t2 the mean coalescence time from different groups:

$$t0 = 2kdN\,;\, t_1 = \frac{k(d-1)}{2m1} + 2kdN\,;\, t_2 = \frac{k(d-1)}{2m1} + \frac{k-1}{2m2} + 2kdN$$

That leads to:

$$Fsr = \frac{1}{1+4Nm1\frac{d}{d-1}}\,;\, Frt = \frac{1}{1+4Nd\frac{k}{k-1}m2+(d-1)\frac{k}{k-1}m1}\,;\, Fst \approx \frac{1}{1+4Nd\frac{k}{k-1}m2}$$

Observed Fst is primarily estimated using the estimate proposed by Weir & Cockerham (Weir & Cockerham 1984), that is very close to Wright's F-statistics:

$$Fst = \frac{f0 - f1}{1 - f1}$$

where (1-fo) is the average pairwise difference between all possible pairs of genes within pops (Beaumont & Nichols 1996), or the average heterozygosity within pops (Excoffier *et al.* 2009), and (1-f1) is the average pairwise difference between all possible pairs of genes between pops, i.e. the probability that two genes from different populations are different. Measured Fst values are used to estimate the parameters N, m1 and m2 of the hierarchical island model (Excoffier *et al.* 2009). The model is then used to simulate populations and draw the expected (neutral) joint distribution of Fst VS h1, where h1 is estimated as $h1 = \frac{h0}{1-Fst}$: $Fst = \frac{f0-f1}{1-f1} \leftrightarrow Fst = \frac{(1-h0)-(1-h1)}{h1} \leftrightarrow h1 = \frac{h0}{1-Fst}$

2.5% and 97.5% quantiles of the expected Fst vs h1 distribution are estimated through a Kernel density approach and corresponds to the neutral expectations, also called 'neutral envelop'. Confronting the observed data to this neutral envelop allows the direct identification of outliers: loci that are above the neutral envelop are under diversifying selection, whereas genes under the neutral envelope are under uniform selection.

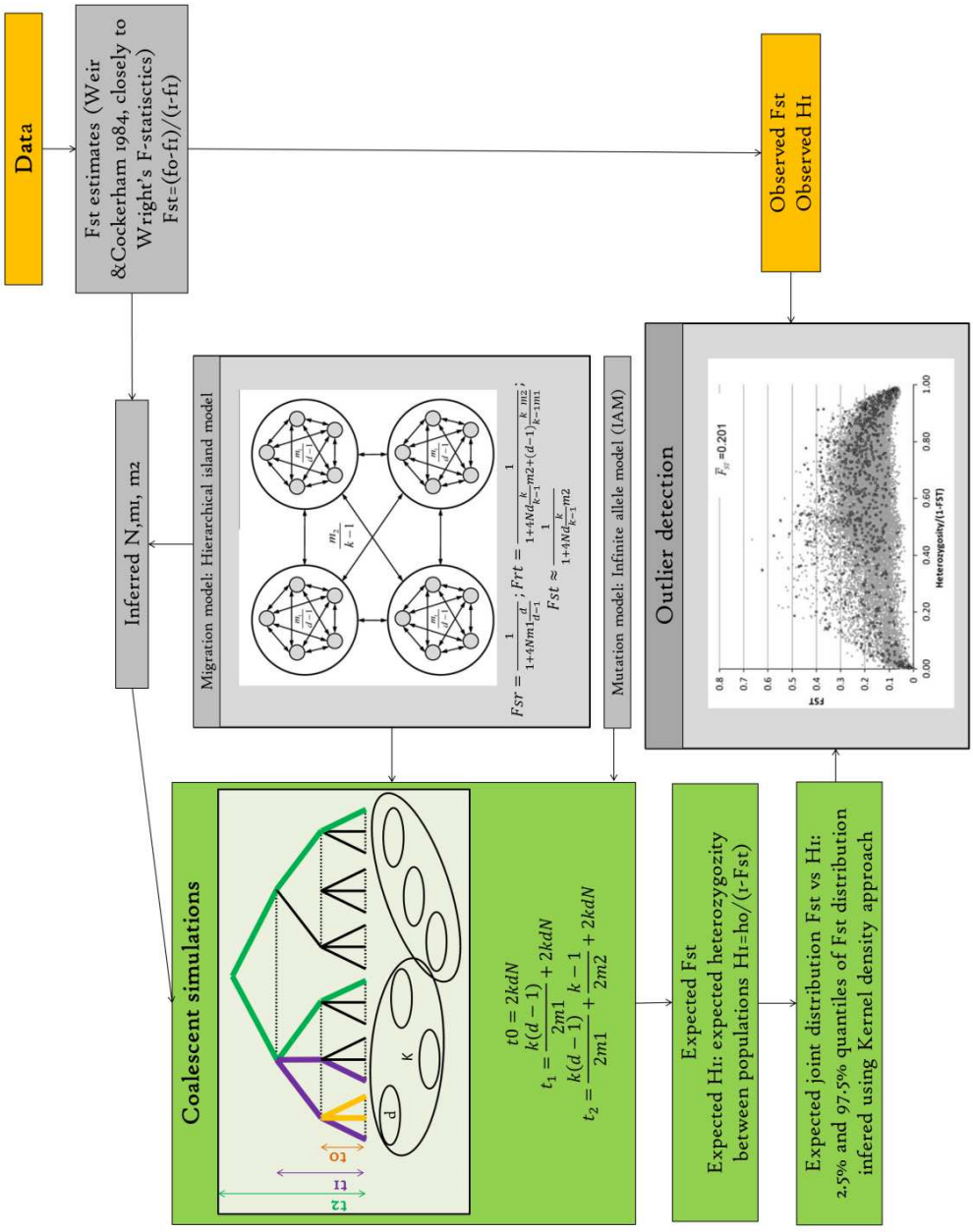


Figure 57: Principle of Fst-based selection tests using coalescent methods.

BOX 5 - METHODS

Fst-based selection tests: BAYESIAN methods

Bayesian methods use a model of population differentiation as originally proposed by Balding & Nichols (Balding & Nichols 1995), in which an ancestral population is split into J sub-populations and $Fst_{i,j}$ refers to the differentiation between the ancestral and the observed population j for the locus I, **figure 58.** Balding & Nichols proposed that allelic frequencies of each derived population may be drawn from a Dirichlet distribution of parameters $\theta_{i,j}$ and $p_{i,k}$:

$$p_{i,j} \sim Dir(\theta_{i,j} p_{i1}, \dots, \theta_{i,j} p_{iK})$$

pi,j refers to the allelic frequency at locus i in the population j

pi,k the frequency of the allele k at locus i in the ancestral population

$$\theta_{i,j} = \frac{1 - Fst_{i,j}}{Fst_{i,j}}$$

And the frequency for the entire set of allele is: $p = \prod_{i=1}^{I} \prod_{j=1}^{J} p_{i,j}$

Instead of inferring I*J Fi,j coefficient, Balding *et al.* (Balding *et al.* 1996) proposed to partition Fst into a locus (α_i) and population-specific (β_j) components, originally as:

$$Fst_{i,j} = \frac{1}{1 + \alpha_i + \beta_j} \leftrightarrow \frac{1 - Fst_{i,j}}{Fst_{i,j}} = \alpha_i + \beta_j$$

In 2008, Foll & Gaggiotti (Foll & Gaggiotti 2008) generalized this model to be applicable to dominant (AFLPs) markers and implemented Bayescan software. They use:

$$\theta_{i,j} = \frac{1 - Fst_{i,j}}{Fst_{i,j}} \leftrightarrow 1/\theta_{i,j} = \frac{Fst_{i,j}}{1 - Fst_{i,j}} \leftrightarrow log(1/\theta_{i,j}) = lo\,g\left(\frac{Fst_{i,j}}{1 - Fst_{i,j}}\right) = \alpha_i + \beta_j$$

$$\theta_{i,j} = e^{-(\alpha_i + \beta_j)}$$

Two alternative models are then calibrated using Bayesian modeling: a model M1 excluding the coefficient α_i, and a model M2 including α_i. The best model is selected using the bayes factor:

$Bayes\ factor = BF = \frac{P(data/M2)}{P(data/M1)}$.

If the coefficient α_i is retained for a locus i, thus the loci is submitted to selection: $\alpha_i > 0$ indicates a positive selection at the population level (corresponding to divergent selection at the meta-population level), while $\alpha_i < 0$ indicates balancing selection (uniform selection among sub-populations).

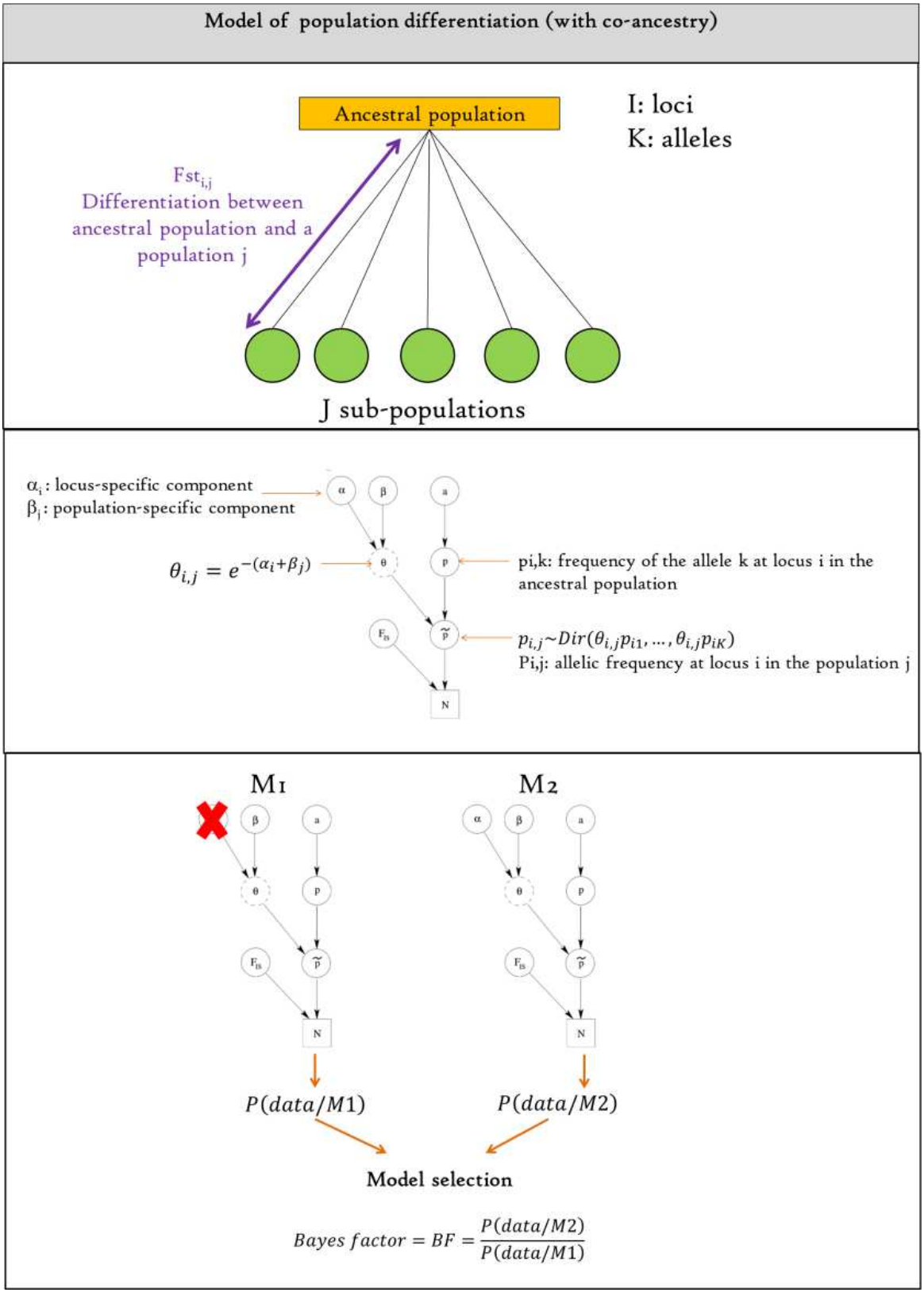


Figure 58: Principle of Fst-based selection tests using Bayesian methods.

Evidences of adaptation in temperate and tropical plant populations

All methods confounded, the literature provides numerous molecular evidences of adaptive divergence in plant species, mainly across broad climatic gradients.

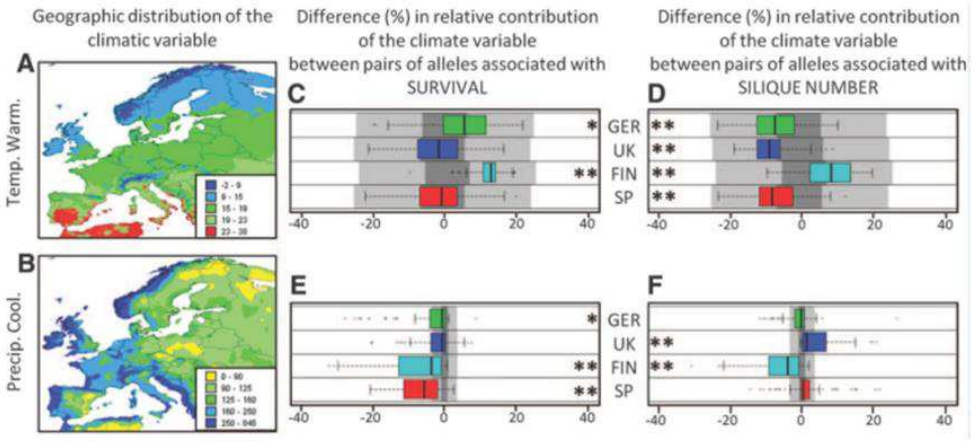


Figure 59: Influence of temperature and precipitation on the distribution of alleles linked to fitness (Fournier-level *et al.* 2011).

In *Arabidopsis thaliana*, a genome-wide association study revealed that fitness-related loci (growth and fruit production) exhibit signatures of local adaptation linked to climatic variables across Europe (Fournier-level *et al.* 2011), **figure 59**. In black spruce (*Picea mariana*), several genes involved in growth, response to constraints (cold and drought) show patterns of differentiation concordant with diversifying selection among both climatic and precipitation partitioning in Québec (Prunier *et al.* 2011). In *Pinus pinaster*, Eveno *et al.* (Eveno *et al.* 2007) analyzed the structure of genetic diversity across the maritime pine range for SNPs within genes candidates for drought stress tolerance. Several were identified as 'outliers' probably under diversifying selection. In Lobolly pine (*Pinus taeda*), several genes involved in responses to biotic and abiotic constraints were structured by aridity in the United States (Eckert *et al.* 2010). Similarly, Richardson *et al.* (Richardson *et al.* 2009) found that 70% of the genetic variations (obtained from anonymous AFLPs) is explained by climate in *Pinus monticola* inhabiting the west coast of USA.

Altitudinal gradients also provide molecular evidence of local adaptation. In white spruce, Namroud *et al.* (Namroud *et al.* 2008) found patterns of genetic differentiation concordant with divergent selection among populations of different elevations for genes involved in flowering time, oxidative stress and nitrogen uptake. In the coastal Catalonian montains, Jump *et al.* (Jump & Penuelas 2006) detected significant variation in gene frequencies related to temperatures in *Fagus sylvatica*.

However, the question of plant adaptation to environmental conditions is highly neglected in tropical rainforests: there is, up to now, **few study dealing with adaptation in tropical trees**. Moreover, the great majority of studies that provides molecular evidence of adaptation in trees of temperate zones are focusing on broad climatic gradients acting at large spatial scales. On the contrary, only few studies have provided evidence of adaptation to local constraints (such as edaphic constraints among micro-habitats or local biotic constraints). Burgarella *et al.* (Burgarella *et al.* 2012) detected footprints of diversifying selection for taxol-related genes (involved in defense against predators) in *Taxus baccata* in Spain mountains. They suggested that local selective pressures exerted by predators and host-enemy co-evolution would have led to genetic divergence among uplands. In an original study, Manel *et al.* analyzed patterns of adaptation in a mountain plant (*Arabis alpina*) across geographical scales (Manel *et al.* 2010). Surprisingly, they found a higher proportion of loci of ecological relevance (Fst-based outliers) at local scale. At regional scales, temperature and precipitations were identified as the major drivers of allele distribution, but it was less clear at local scale in which environmental variations are characterized by topography-related variations rather than climatic ones. They suggested that there may be two different types of adaptive responses acting on A. *alpina*: a site-specific local adaptation (caused by topography-related variations) and a more general adaptive response at larger geographic scales (caused by large climatic gradient, including both temperatures and precipitations).

In this study, I used both candidate genes (based on SNPs markers) and genome scan (based on anonymous AFLPs markers) to test for adaptation in the neotropical tree *Eperua falcata* over very short geographical scales (hundreds meters), see below and articles n° 1 and 2.

Summary of PhD results:

In *Eperua falcata*, both candidate genes and genom scan approaches revealed footprints of divergent selection among local habitats (see articles n°1 and n°2).

During this PhD, two main approaches were carried out to identify footprints of natural selection driven by local habitat heterogeneity in *Eperua falcata*. The **candidate gene approach** involved trees inhabiting the sites of Paracou and Nourragues, whereas the **genome scan approach** involved the sites of Laussat and Regina.

In the first approach (**article n°1**), a collection of candidate genes for divergent selection combined with several genes of unknown function was sequenced to identify Single nucleotide polymorphisms (SNPs). A coalescent method (under an island model) revealed that several genes were probably submitted to divergent selection between water-logged bottomlands and well-drained terra-firma, among which two genes encoding proteins involved in plant response to stresses: a catalase that is involved in the response to oxidative stresses frequently experienced during water-logging, and the farnesyl-transferase which plays a role in the regulation of stomatal conductance, **figure 60**. On the contrary, genes encoding aquaporins were either neutral (Paracou) or submitted to uniform selection (Nourragues).

In the second approach (**articles n°2**), a large panel of anonymous markers (1196 AFLPs) was genotyped to estimate the extent of divergent selection in the genome of *Eperua falcata* in the study sites of Laussat and Regina. Both Coalescent (hierarchical island model) and Bayesian methods were used. Both methods revealed that global differentiation among local habitats was very low (ranging between 0.02 and 0.04 depending on the method used) but concordant with the average Fst estimated from candidate genes. The Coalescent method detected 21 outliers under uniform selection, while the Bayesian procedure (which is more stringent than coalescent methods, Narum & Hess 2011, Vilas *et al.* 2012) detected from one to three outliers depending on the dataset used (all regions or by-site analysis). Eleven of the detected outliers show similar patterns of genotypic frequency among local habitats in the two study sites, and are concordant with the hypothesis some alleles are favored or exclude by a particular habitat, **figure 61**.

These results are consistent with genetic diversity for outlier loci having been driven by divergent selection. Under the hypothesis that the 1196 analyzed AFLPs were uniformly distributed in the genome (Strasburg *et al.* 2011) and thus well representative of the entire genome, it is possible to estimate the extent of divergent selection among local habitats: between 0.2 and 0.9%.

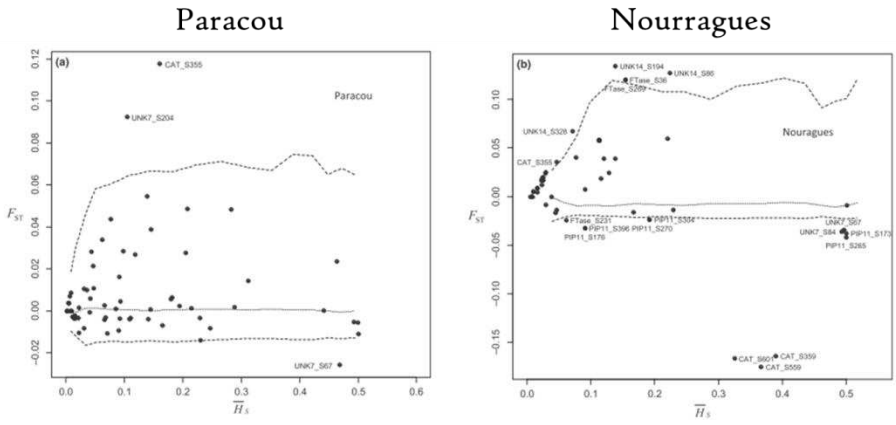


Figure 60: Fst Vs Heterozygosity distribution. Points indicate observed values whereas lines represents 95% neutral envelop estimated through coalescent method.

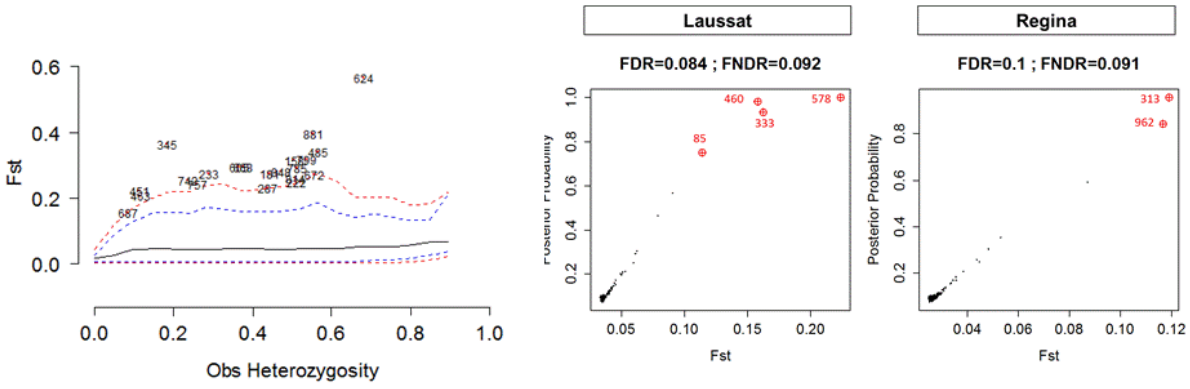


Figure 61: Fst-based outlier detection using both coalescent (left) and bayesian (right) approaches using AFLPs data (from Laussat and Regina study sites)

I performed a literature survey for a relationship between the geographical scale of the different studies and the proportion of outliers detected through Fst-based methods, by reviewing a number of published studies searching footprints of divergent selection using Fst-based methods at both inter- and intra-specific levels across a large range of animal and plant models. The following table summarizes them, with the following abbreviations:
For the column 'Model': A=animal, P=plant
For the column 'Marker': A=AFLPs, Al=Allozymes, M=microsatellite, S=SNPs
For the column 'Method": B=Bayesian, C=Coalescent, O=Other

Biological model	Model	Study scale	Marker	Method	Outliers (%)	Reference
Ovis aries	A (mammal)	4300km	M	O	18.95	Joost *et al.* 2007
Ovis aries	A (mammal)	4300km	M	C	28.57	Joost *et al.* 2007
Populus tremula	P (tree)	4000km	M	C	17.14	De Carvalho *et al.* 2010
Hylobius abietis	A (insect)	3800km	A	O	13.25	Joost *et al.* 2007
Hylobius abietis	A (insect)	3800km	A	C	4.82	Joost *et al.* 2007
Picea abies	P (tree)	3000km	A, M, S	C	1.97	Achéré *et al.* 2005
Gasterosteus aculeatus	A (fish)	2200km	M	B	8.77	Makinen *et al.* 2008
Phytomyza glabricola	A (insect)	2100km	A	C	11.11	Scheffer *& Hawthorne* 2007
Pinus taeda	P (tree)	2000km	S	C	0.78	Eckert *et al.* 2010
Pseudotsuga menziesii	P (tree)	2000km	S	O	6.61	Eckert *et al.* 2009
Pinus monticola	P (tree)	1800km	A	C	12.12	Richardson *et al.* 2009
Heliantus annus, H. debilis	P	1600km	M	C	3.41	Scascitelli *et al.* 2010
Cryptomeria japonica	P	1400km	S	C	3.38	Tsumura *et al.* 2007
Heliantus annus, H. petiolaris	P	1300km	M,S	C	4.1	Yatabe *et al.* 2007
Heliantus annus, H. petiolaris	P	1300km	S	B	3.3	Strasburg *et al.* 2009
Neochlamisus bebbinae	A (insect)	950km	A	C	10.51	Egan *et al.* 2008
Neochlamisus bebbinae	A (insect)	950km	A	C	4.03	Egan *et al.* 2008
Neochlamisus bebbinae	A (insect)	950km	A	C	5.15	Egan *et al.* 2008
Neochlamisus bebbinae	A (insect)	950km	A	C	1.12	Egan *et al.* 2008
Neochlamisus bebbinae	A (insect)	950km	A	C	8.72	Egan *et al.* 2008
Capra hircus	A (mammal)	900km	S	C	11.11	Pariset *et al.* 2009
Picea glauca	P (tree)	750km	S	C	3.56	Namroud *et al.* 2008

Arabis alpina	P	730km	A	O	10	Manel *et al.* 2010
Picea mariana	P (tree)	700km	S	B	0.26	Prunier *et al.* 2011
Picea mariana	P (tree)	700km	S	O	1.17	Prunier *et al.* 2011
Picea mariana	P (tree)	700km	S	B	0.13	Prunier *et al.* 2011
Picea mariana	P (tree)	700km	S	O	1.43	Prunier *et al.* 2011
Salmon Trutta	A (fish)	600km	M	C	2.7	Meier *et al.* 2011
Peromyscus maniculatus	A (mammal)	500km	Al	C	8.33	Storz & Dubach. 2004
Populus alba, P. tremula	P (tree)	500km	various	C	35.48	Lexer *et al.* 2010
Crassostrea virginica	A	300km	A, R	C	1.33	Murray & Hare 2006
Crassostrea virginica	A	300km	A, R	C	0	Murray & Hare 2006
Arabidopsis halleri	P	200km	A	B	3.99	Meyer *et al.* 2009
Arabis alpina	P	150km	A	O	2.95	Manel *et al.* 2010
Arabis alpina	P	150km	A	O	9.45	Poncet *et al.* 2010
Arabis alpina	P	150km	A	O	7.39	Poncet *et al.* 2010
Arabis alpina	P	150km	A	O	2.55	Poncet *et al.* 2010
Coregonus clupeaformis	A (fish)	130km	A	O	3.18	Campbell & Bernatchez 2004
Rana temporaria	A	100km	A	C	1.53	Bonin *et al.* 2006
Littorina saxatilis	A	70km	A	C	3.3	Galindo *et al.* 2009
Zostera marina	P	50km	M	C	12	Oetjen & Reusch 2007
Quercus robur, Q. petraea	P (tree)	36km	various	C	12.08	Scotti-Saintagne *et al.* 2004
Diabrotica virgifera	A (insect)	32km	A	C	1.19	Miller *et al.* 2007
Littorina saxatilis	A	26km	A	C	4.9	Wilding *et al.* 2001
Viola cazorlensis	P	23km	A	B	2.44	Herrera & Bazaga 2008
Arabis alpina	P	20km	A	O	2.55	Manel *et al.* 2010
Arabis alpina	P	20km	A	O	15.66	Manel *et al.* 2010
Arabis alpina	P	20km	A	O	17.18	Manel *et al.* 2010
Fagus sylvatica	P (tree)	10km	A	C	0.39	Jump & Penuelas 2006
Howea belmoreana, H. fosteriana	P (palm)	10km	A	C	1.46	Savolainen *et al.* 2006
Timema cristinae	A (insect)	5km	A	C	14.61	Nosil *et al.* 2008
Eperua falcata	**P (tree)**	**0.5km**	**A**	**C**	**1.76**	**THIS STUDY (Article n°2)**
Eperua falcata	**P (tree)**	**0.5km**	**A**	**B**	**0.26**	**THIS STUDY (Article n°2)**
Eperua falcata	**P (tree)**	**0.1km**	**S**	**C**	**3.51**	**THIS STUDY (Article n°1)**

This overview of literature suggests that the proportion of outliers is slightly higher in inter-specific comparisons (involving closely-related plant species) than in intra-specific ones, and revealed a large variability among biological models (all scales confounded), **figure 62**.

The relationship between geographical distance and proportion of outliers is not

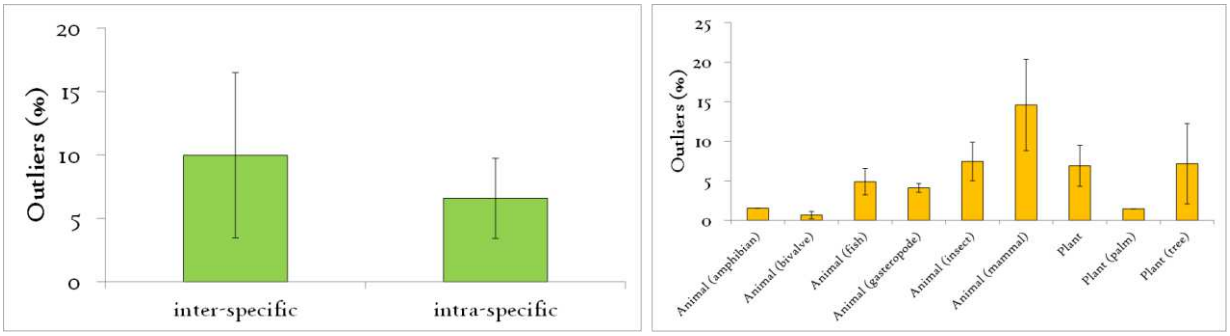


Figure 62: Variations in the proportion of outlier between inter- and intra-specific levels and between biological models (drawn from papers quoted in the table above).

clear and certainly non-linear. A regression with the logarithm of distances (**figure 63**, left) suggests that the proportion of outliers is lower at intermediate geographical distances. However, this trend disappeared when including *Eperua falcata* studies at very short spatial scale (right). Last, Soto-Cerda *et al.* (2013) provided the most extreme case (not included in the plots) as no outlier was detected in the plant species *Linum usitatissum* in a world-wide analysis; suggesting thus the probable absence of relationship between geographical scales and the extent of adaptation in the genome of organisms.

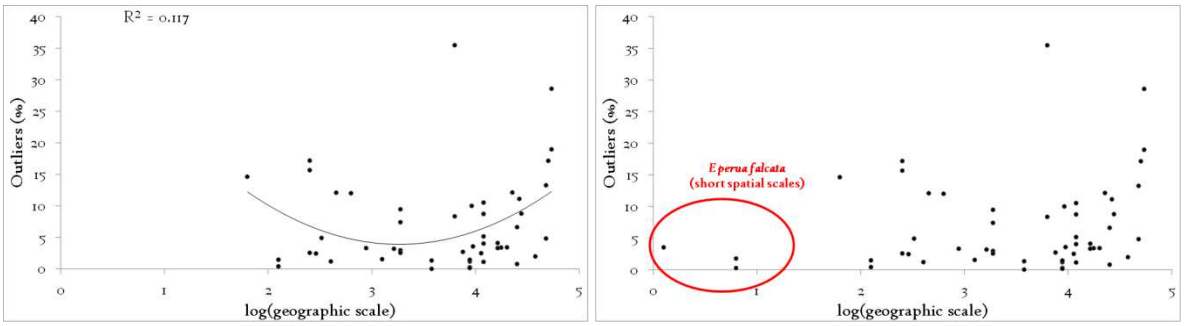


Figure 63: Relationship between the proportions of outliers detected and geographical scales.

Even if comparing the number of outliers detected between studies becomes common (Luikart *et al.* 2003, Nosil & Ortiz-Barrientos 2009, Strasburg *et al.* 2012), such comparisons should be taken with caution. Indeed, the different studies use different kinds of molecular markers, and we can expect that the proportion of outliers would be more frequent in candidate genes approaches than in genome-scans involving anonymous markers. Molecular markers also differed in their mutation rates, making the comparison

between microsatellites and SNPs difficult, for example. Last, the method used for detecting loci under selection may also bias the comparison, as Bayesian methods are known to be more robust to demographic scenarios and to detect fewer false-positives (Narum & Hess 2011, Vilas *et al.* 2012). Thus, the literature is, up to now, not sufficiently rich to properly realize such surveys, but the current popularity of Fst-based methods will probably provide sufficient examples in a close future.

5. Next generation sequencing / genotyping and new opportunities

Modern evolutionary ecology is currently progressing rapidly because of advances in genomics technologies (Stapley *et al.* 2010). Next-generation technologies refer to the panel of new sequencing/genotyping technologies, such as Roche 454-pyrosequencing, Illumina (High-Seq), Illumina micro-arrays etc...

Next-generation sequencing (NGS) attracts a particular attention because it is more cost-effective than classical Sanger sequencing (Morozova *et al.* 2009), given the large amount of DNA sequenced, such as transcriptomes and more recently, completes genomes. Complementary DNA (cDNA synthetized from mRNA by reverse transcription) or genomic DNA (gDNA) is sequenced by fragments of varying size depending on the technologies (named 'reads'). The sequenced fragments need to be assembled into contigs, either using a 'de novo' assembly method or by mapping against a reference. Next-generation sequencing proved to be useful for **gene characterization**, **gene expression profiling** and for **identifying polymorphisms**, such as SNPs (Lister *et al.* 2009). Blast and functional annotations allow the characterization of contigs (called 'unigenes' once characterized). Blast allows confronting assembled sequences to public databases, while functional annotation allow determining in which biological processes the protein encoded is involved. Gene expression profiling allows a quantification of RNA

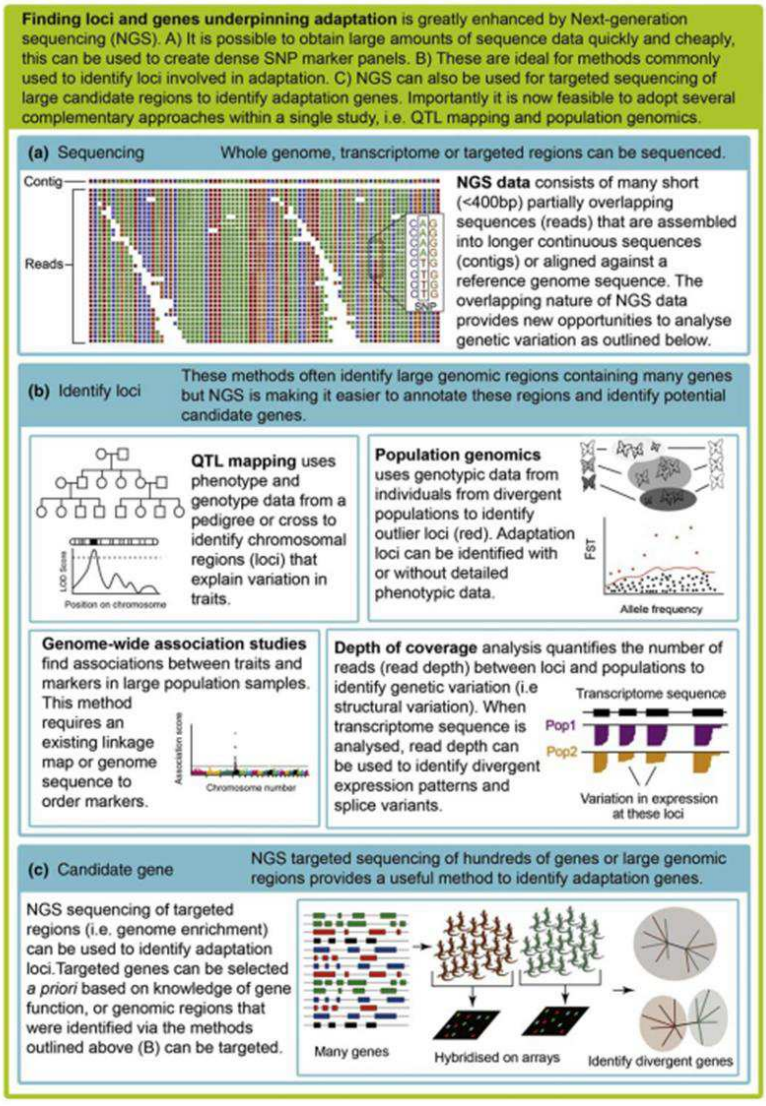


Figure 64: Common methods used to identify loci in non-model organisms and how they have been improved by NGS (from Stapley *et al.* 2010).

expressed in organs and tissues. Last, and when next-generation sequencing allow a sufficient depth (i.e. number of reads for a single site that equals to the number of sequencing repetitions for this site), it is possible to identify polymorphisms, (such as SNPs) giving valuable information for both evolutionary biologists and quantitative geneticists (Ganal *et al.* 2009, Rafalski 2002, Picoult-Newberg *et al.* 2011, Tassel *et al.* 2008), **figure 64**. For example, next-generation sequencing technologies allow **selecting good candidate genes** for adaptations studies: polymorphic genes involved in biological processes of interest or (possibly non-annotated) polymorphic genes expressed in high levels in response to particular constraints. They enable to track genetic loci under selection for adaptation in **non-model organisms**.

Candidate genes may be high-throughput sequenced or their polymorphisms genotyped, as next-generation sequencing also allows the sequencing of targeted sequences (by-capture sequencing, re-sequencing) and the simultaneous genotyping of large amounts of targeted SNPs.

Thus, novel sequencing technologies are going to extend genome scan approaches, by providing better coverage of transcriptomes or genomes, and adaptation studies using 'whole-genome scans' will probably appear in the next years. Moreover, NGS provides a very large amount of characterized genes by comparison with AFLPs that suffer from the major limitation that outliers are anonymous, **figure 65**.

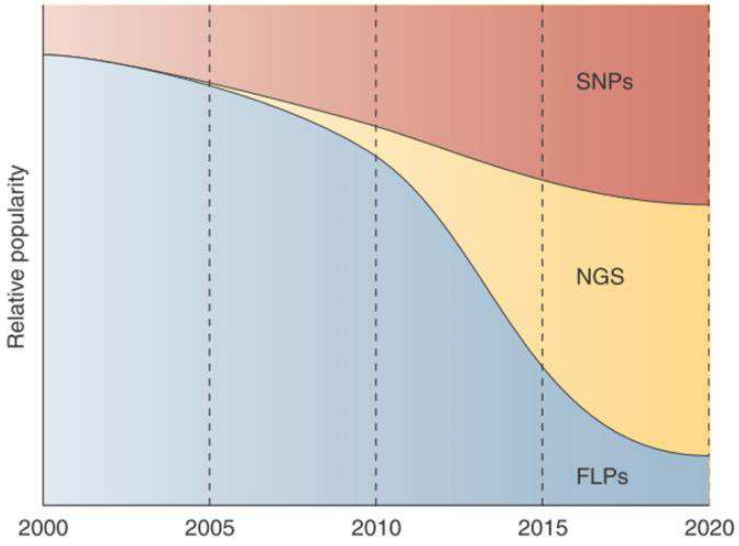


Figure 65: Subjective view of the importance of genotyping strategies for non-model organisms (Seeb *et al.* 2011).

However, next-generation technologies have several disadvantages (Stapley *et al.* 2010):

(1) The first limitation is informatics. Due to the large amount of data produced, data manipulation is complicated and post-treatment requires automation of all steps of analyses. However, the domain suffers from a lack of 'user-friendly' tools (hardware, software, algorithms).

(2) They are less accurate than Sanger sequencing as they produce more errors of base calls that can result in false-positive polymorphisms (De Pristo *et al.* 2011). They require a careful post-processing assembly and cleaning to be properly interpretable, and the integration of individual base quality is required even if it is often neglected.

(3) Short-reads can be difficult to assemble '*de novo*' (i.e. without reference), particularly in whole genome sequencing of non-model organisms due to the large amount of repeated DNA in the genome.

(4) Distinguishing a real polymorphism in a single gene versus a genetic variation between two duplicated genes is challenging. Applying stringent assembly criteria would limit the risk of false-positive discovery, but may result in an under-estimation of the true diversity as several false-negative would be excluded.

(5) Obtaining complete coverage of the transcriptome can be difficult due to the disparity in expression of different genes and between tissues.

(6) Population genomics uses pooled samples to minimize sequencing costs and directly infer allelic frequencies (Futschik & Schlotterer 2010, Turner *et al.* 2010).

(7) Population genomics commonly involve population genetics models that gained complexity over the past 10 years. However, the majority of the widespread population genetics approaches (such as those involving genome scans) would be complicated in the case of large datasets provided by NGS, as they are often time-consuming even in small datasets (Nielsen *et al.* 2005).

During this PhD, I analyzed the transcriptome of seedlings of *Eperua falcata* (plus three other species without interest in this manuscript) sequenced by 454-pyrosequencing, providing the first example of NGS application to non-model tropical tree species. It allowed the massive characterization of potential candidate genes (**article n°3**). I wrote a complete suite of R scripts that deal with next generation data and facilitate their manipulation and post-processing (including assembly cleaning and SNP detection). The complete suite of R scripts is described in the section '**PhD Results & Jobs**' and will be soon packaged into R.

Summary of PhD results:

High-throughput transcriptome sequencing allowed the massive characterization of expressed genes and polymorphism (SNPs) discovery (see article n°3).

I analyzed the transcriptome of 4 neotropical tree species widespread in French Guiana. Total mRNA were extracted from leaves, stems and roots of two vigorous seedlings per species, and converted into cDNA. cDNA libraries were sequenced by 454-pyrosequencing.

In *E. falcata*, 153,551 reads (out of (224554) were assembled into 23390 contigs. I characterize 16159 unigenes that returned a blast result with an e-value below 10-25. After contaminant removal (removal of contigs that never blasted into a green plant species among their 10-top hits), 15664 unigenes remained and 11240 were annotated (i.e. the protein encoded was associated to particular biological processes).

I analyzed transcription profiles within each organ (leaves, stems, roots) and I used a permutation test to identify biological processes particularly relevant in a particular organ (biological processes represented by contigs over-expressed in a particular organ, based on the number of reads that brought specific organ-tagged within each contig). In *E. falcata*, I identified between 7 (leaves) and 26 (roots) biological processes over-represented in that organ, **figure 66**.

Prior to polymorphism detection, I cleaned the assemblies following a stringent procedure:

-Individual bases were masked using several criterions, including minimum allele number, minimum allele frequency, depth, and individual base quality, **figure 67**.

-Sites (assembly columns) composed by masked bases and deletions only were removed.

A total of 5713 SNPs were identified, among which

- 2657 high quality SNPs (substitutions of only two variants) for a transition/transversion ratio Ti/tv= 1.66.
- 2992 insertion-deletion
- 64 SNPs with more than two variants

At last, 1283 contigs were polymorphic (only 5.5% of assembled contigs, mainly because numerous contigs lacked a sufficient depth for searching SNPs), for a total SNP density of 0.95 per 100 bp.

Even if the true diversity that may be encountered in wild population is probably higher than the polymorphism detected from two seedlings (4 gametes), this database provides useful information for future investigations. In particular, it provides a large panel of candidate genes (expressed genes of known function). Several of these candidates will be high-throughput genotyped or re-sequenced in seedlings currently transplanted in wild conditions (reciprocal transplants) and in their adult trees. It will allow extending the present study by including pedigree analyses, association genetics studies, and by expanding the genome scan approach for testing selection by including SNPs contained in genes of known function (see '**Discussion**').

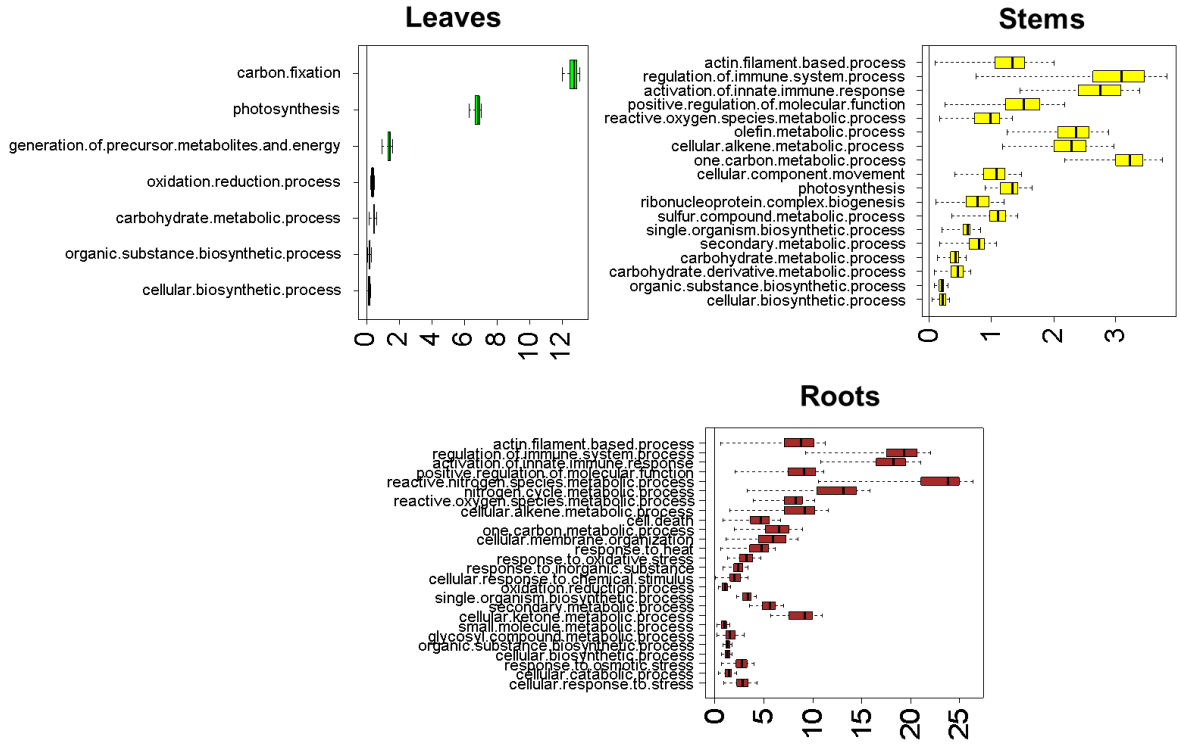


Figure 66: Difference between observed and randomized mean contig expression (number of reads) within each biological process (level 4) and organ over 1000 permutations. Only significant biological processes are represented.

Figure 67: Raw and cleaned assemblies. "N" indicates individual bases masked by the procedure.

PART II - Phenotypic evolution and quantitative genetics

Studying adaptation based on (quantitative) phenotypic traits attracts a particular attention because natural selection sorts phenotypic variations, thus optimizing the mean fitness of populations. In wild populations, many traits show phenotypic variations among individuals (genotypes) and among habitats, because both genetic factors and environmental conditions affect the phenotype expressed by a genotype in a given environment.

1. Causes of phenotypic variation

Even if several traits show discrete variations, most traits are continuously distributed for two reasons: most traits have a complex genetic architecture involving more than one locus, and they are also affected by other sources of phenotypic variations (environmental variations, maternal effects). Phenotypic variations may thus be partitioned into **genetic** and **environmental** factors, according to a classical linear model:

$$\boldsymbol{P} = \mu + \boldsymbol{G} + \boldsymbol{E} + \boldsymbol{I_{GxE}} + \boldsymbol{\varepsilon}$$

where G represents the global phenotypic differentiation among genotypes, E the global effect of environment (i.e. phenotypic plasticity common to all genotypes), and I_{GxE} the genotype-by-environment interactions. Significant I_{GxE} show that different genotypes are differentially affected by the environment (i.e. genetic divergences among genotypes causes variations in phenotypic plasticity), **figure 68**. Genetic factors may describe the genetic divergence between genotypes, progeny arrays (families), provenances or species.

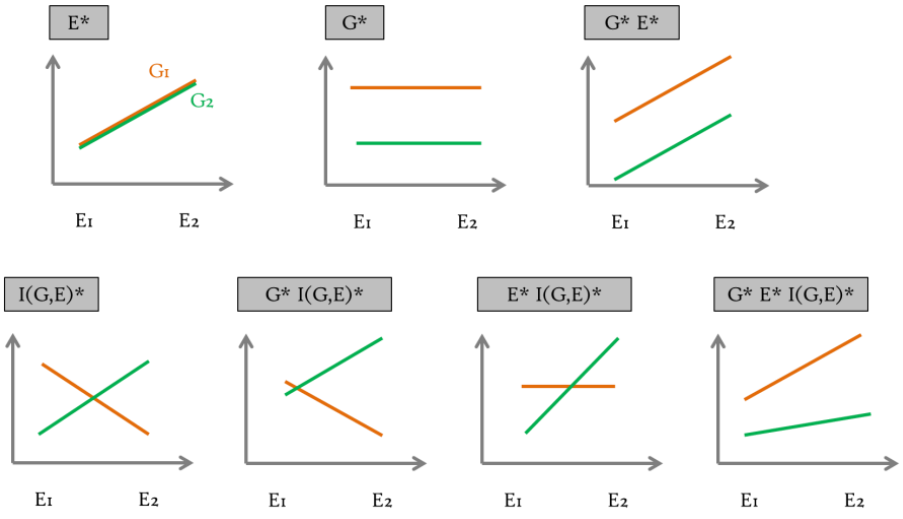


Figure 68: Partition of phenotypic value into genetic and environmental factors (plus their interactions). Plots show different possible patterns of phenotypic value (ordinate) of two genotypes (red and yellow) within two environments (E1 and E2 in abscise). Significant factors are given at the top of each plot.

1.1 Genetic variation:

Since the discovery of Mendelian inheritance, the genetic basis of phenotypic expression is evident. Originally, Mendel discovered that the proportion of (discrete) phenotypes after breeding may be predicted in the case of traits controlled by a limited number of loci. Given a character controlled by a single locus, the crossing of two homozygotes (A/A) produces 100% of homozygotes (A/A) with the same phenotype than parents. The crossing of two heterozygotes (A/B) produces a progeny composed by 25% of homozygotes (A/A) that display the same phenotype than the parental homozygote (A/A), 25% of homozygotes (B/B) that display the same phenotype than the parental homozygote (B/B), and 50% of heterozygotes (A/B) that display either an intermediate phenotype (**additivity**) or the same phenotype than one of the parent (**complete dominance**).

Numerically, phenotypic variations for quantitative traits in a given environment may be expressed using a linear model restricted to genetic factors:

$$\boldsymbol{P} = \mu + \boldsymbol{G} + \boldsymbol{\varepsilon}.$$

Genetically-based phenotypic variations caused by one of the underlying genes may thus be partitioned into **additivity**, **dominance** and **epistasis** effect: $G = A + D + I$, leading to: $\boldsymbol{P} = \mu + \boldsymbol{A} + \boldsymbol{D} + \boldsymbol{I} + \boldsymbol{\varepsilon}$.

Considering two alleles (A) and (B) for a given locus, **additivity** simply describes how allelic states may affect phenotypic values apprehended among homozygotes genotypes: it may be numerically assessed by the relation $|a| = |P_{AA}\text{-}P_{aa}| / 2$. **Dominance** refers to the interactions between alleles of the locus that may influence the phenotypic value of heterozygotes: in the case where there is no dominance, heterozygotes display an intermediate phenotype corresponding to the mean

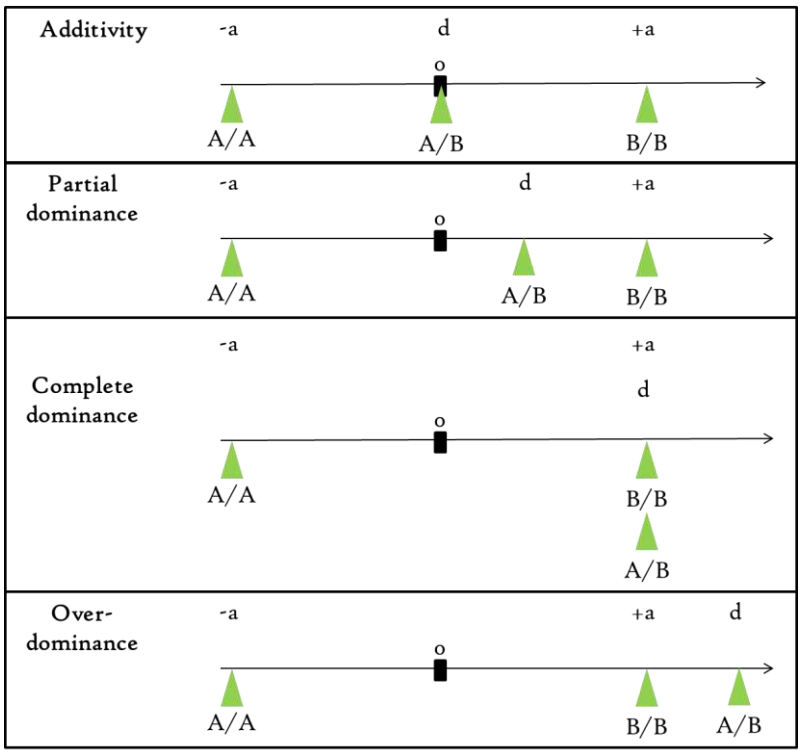


Figure 69: Gene action on the phenotypic value expressed for a given trait: a=additivity, d=dominance.

phenotypic value of the two homozygotes. In the case of dominance, on the contrary, heterozygotes phenotype deviates from the intermediate phenotype, **figure 69**. Last, epistasis corresponds to the interactions with other loci controlling the phenotypic trait.

Genetically-driven phenotypic variations are commonly quantified though **heritability**, **box 6**.

BOX 6 – THEORY

Heritability

Heritability measures the proportion of observable differences in a trait between individuals that is due to genetic differences. Many methods allow estimating traits heritability, among which parent-offspring regression and variance partitioning.

Using **parent-offspring regression**, heritability is defined as the slope of the regression between mid-parents (mean phenotypic value of the two parents) and offspring phenotypic values, **figure 70**. When the slope is high, offspring have phenotypic value close to their parents', suggesting that a high proportion of phenotypic variance is passed down from parents to offspring (Fernandez & Miller 1995).

Heritability is also defined as the proportion of total phenotypic variance that is attributable to variations in additive (narrow-sense heritability) or total genetic values (broad-sense heritability). Heritability may also be estimated by partitioning total phenotypic variations into genetic and environmental sources of variations using linear models.

Broad-sense heritability (h^2_B) is thus defined as the fraction of phenotypic variance attributable to genetic factors without distinction between additive effects, dominance or epistasis, while narrow-sense heritability is defined as the fraction of total variations attributable to additive genetic variance (h^2_N).

$$\boldsymbol{\sigma^2_P = \sigma^2_G + \sigma^2_E + \sigma^2_{GxE} + \sigma^2_{res}} \qquad \boldsymbol{\Rightarrow h^2_B = \sigma^2_G/\sigma^2_P}$$

$$\boldsymbol{\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_I + \sigma^2_E + \sigma^2_{GxE} + \sigma^2_{res}} \qquad \boldsymbol{\Rightarrow h^2_N = \sigma^2_A/\sigma^2_P}$$

When using related individuals instead of clones of the different genotypes, additive genetic variance may be estimated through the relation: $\boldsymbol{\sigma^2_A = \frac{1}{r} * \sigma^2_G}$ where r is the relatedness coefficient (r) (Cotteril 1987). Thus, $\boldsymbol{h^2_N = \frac{\sigma^2_G}{r*\sigma^2_P}}$ where r = 1/2 for full-sibs, and r = 1/4 for half-sibs.

Variance component are thus estimated using classical variance analyses (ANOVA) that calculates variance components as mean squares of each factor (sum of squares of each factor divided by the number of degree of freedom) or directly inferred using Bayesian modeling.

It has to be noticed that there is not a unique value of heritability for a species and a given traits, because heritability may vary among environments (additive variance estimated in a single environment may vary between environmental conditions when gene-by-environment interactions are significant) and among ontogenetic stages, **figure 71**.

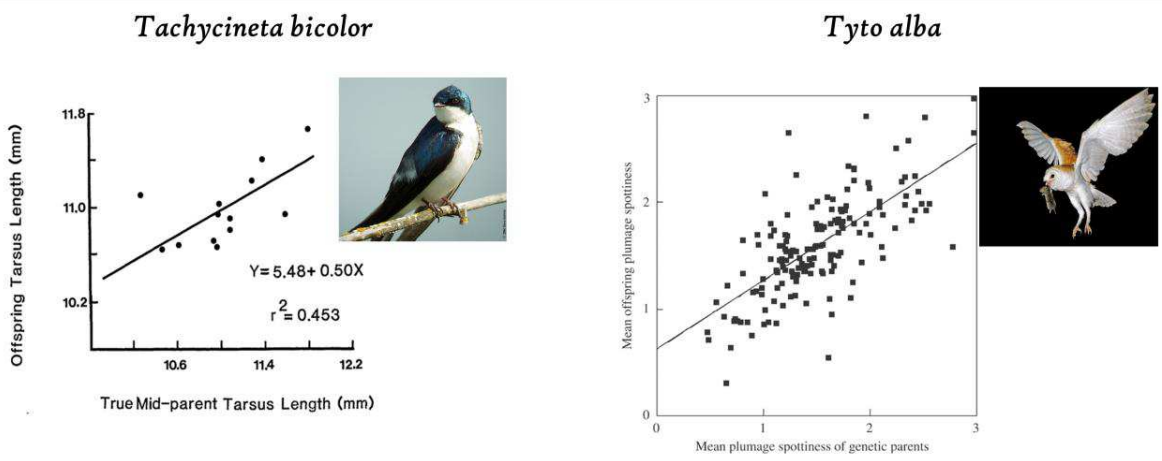


Figure 70: Parents-offspring regression for body size and plumage coloration in two bird species (from Wiggins 1989, Roulin & Dijkstra 2003).

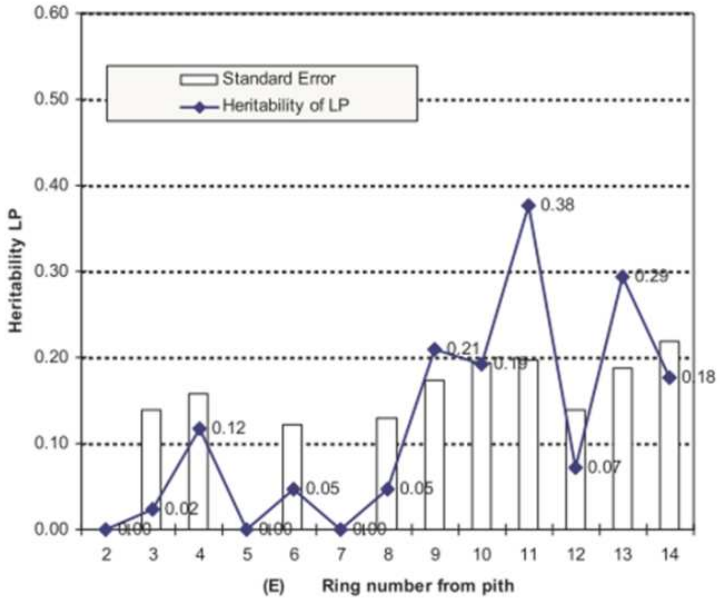


Figure 71: Age trends in individual heritability for latewood proportion in Pinus radiata (from Zamudio *et al.* 2005).

Tree species commonly display significant differences in phenotypic values between genotypes and progenies, suggesting that many traits are commonly heritable (i.e. a significant part of phenotypic variation is due to genetic factors). They commonly display high heritability for morphometric traits (growth traits, wood properties, and leaf traits, **figure 72**) but lower heritability for fitness-related traits (Cornelius 1994, Visscher *et al.* 2008), as natural selection negatively affects heritability by reducing additive genetic variance (see section '**Phenotypic evolution**'). Cornelius (Cornelius 1994) reviewed 67 studies and found evidence of higher heritability for height and straightness than for diameter and volume.

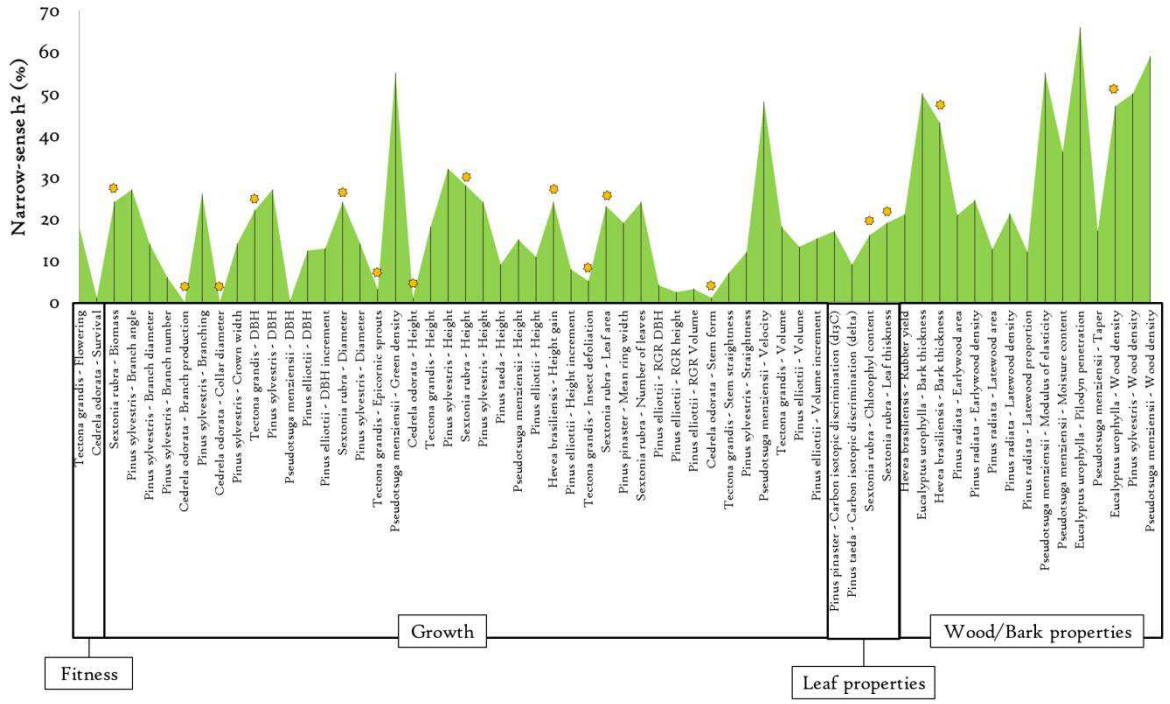


Figure 72: Narrow-sense heritability estimates for fitness, growth, leaf and wood properties in several temperate and tropical tree species. Orange stars indicate tropical tree species.
References: Hodge & White 1992, Haapanen *et al.* 1997, Weir & Borralho 1997, Hannrup *et al.* 1998, Brendel *et al.* 2002, De Souza Goncalves *et al.* 2005, Zamudio *et al.* 2005, Johnson & Gartner 2006, Baltunis et al. 2008, Callister & Collins 2008, Ward *et al.* 2008, Scotti *et al.* 2010.

In this study, I investigated whether functional traits were divergent between progeny arrays of *Eperua falcata* in controlled and non-limiting conditions, and I measured the extent of phenotypic variations due to the relatedness between seedlings progenies. A synthetic summary of the results is presented below; complete results and discussion are described in the **article n°4**.

Summary of PhD results:

A common garden experiment reveals large phenotypic variations among seedlings progeny arrays for growth and leaf traits (see article n°4).

A common garden experiment was used to measure the extent of genetically-driven phenotypic variations for growth (stems dimensions, biomass accumulation and allocation) and leaf traits (leaf composition, leaf structure, leaf photosynthesis and carbon isotopic discrimination) in two *Eperua* species: *E. falcata* and *E. grandiflora*. 44 progeny arrays (harvested around mother trees) grew 24 months in a shade-house in non-limiting conditions.

For all studied traits, significant differences between progenies were detected. Large variations in seedlings growth between progeny arrays led to high σ^2_G/σ^2_P for growth traits (biomass accumulation, height, diameter), **figure 73**. Growth traits displayed higher σ^2_G/σ^2_P than leaf trait, probably because they are complex traits that integrate all properties of individual meristems, leaves and branches. We did not estimated narrow-sense heritability for two reasons:

(1) we did not know the true proportion of full- and half- sibs in the progeny arrays,

(2) we were unable to exclude the fraction of σ^2_G caused by (non-genetic) maternal effects.

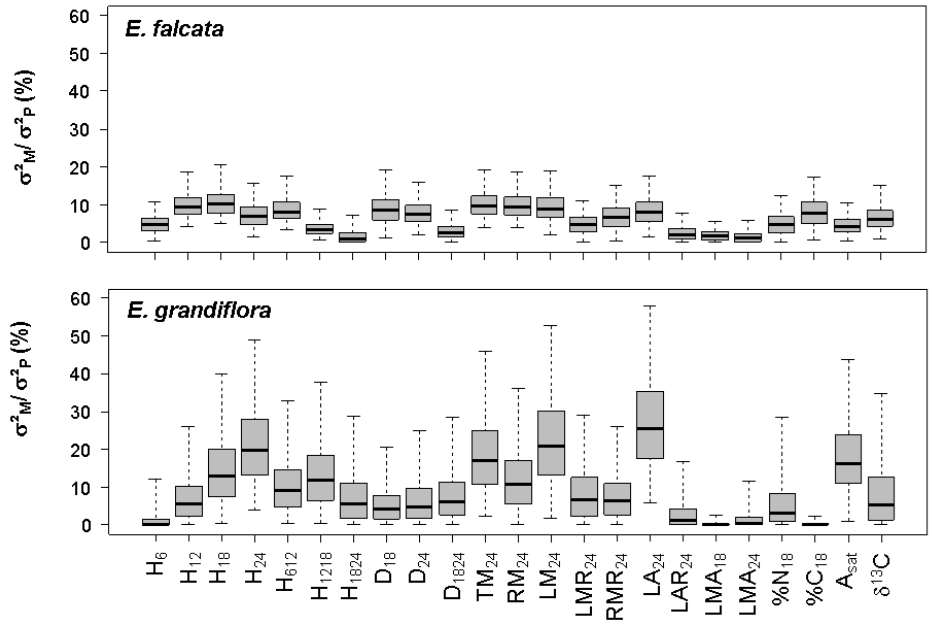


Figure 73: Estimation of σ^2M/σ^2P from seedlings of two Eperua species with a Bayesian analysis of variance (boxplot indicates 2.5%, 25%, 50%, 75% and 97.5% posterior credible intervals).

However, statistical quantitative genetics consider the genome as a 'black box' (it does not require any information about the identity of the genes involved in the phenotypic variations) and it cannot address important questions concerning the genetic architecture of phenotypic traits: What genes affect a given trait? Where are they located in the genome? What is the mode of gene action (additivity, dominance, epistasis)? A particular technique of quantitative genetics, called 'QTL mapping' locates the genomic regions that affect quantitative traits, by (1) creating a genetic map covering the entire genome by crossing divergent populations, most often inbred lines, and (2) searching for associations between allelic states at mapped loci and phenotypic values ('Association genetics'). QTL mapping often reveals the complex genetic architecture of many phenotypic traits. However, this technique is out of the scope of this manuscript.

1.2. Environmental variation:

Environmentally-driven phenotypic variations are closely related with the concept of **phenotypic plasticity**. Plasticity is defined as the ability of a genotype to produce different phenotypes (Pigliucci *et al.* 2006). In other words, it refers to the ability of an organism to alter its physiology, morphology, and life-history traits in response to the conditions it experiences (Nussey *et al.* 2007). Plasticity is also called **acclimation** or **accommodation** depending on its reversibility and the kind of changes: acclimation is often used to describe reversible physiological changes while accommodation often describes non-reversible morphological changes. The function describing the change in a genotype's phenotypic value across an environmental gradient is called '**reaction norm**' (Nijhout 2003, Sarkar & Fuller 2003, Nussey *et al.* 2007).

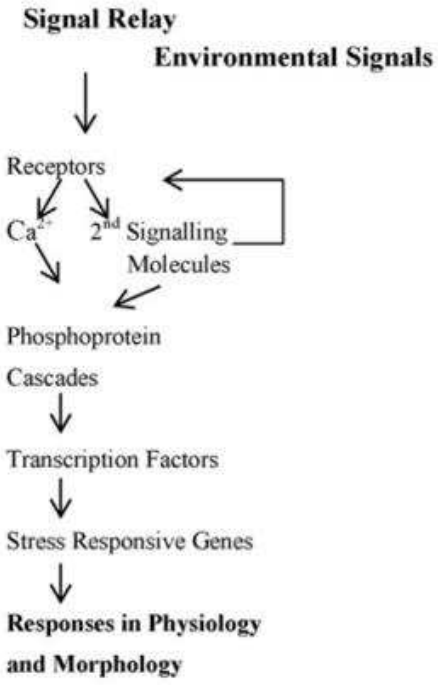


Figure 74: Framework model for signal transduction of stress in plants (from Shao *et al.* 2007).

Two kinds of phenotypic plasticity may be distinguished: passive plasticity refers to the passive reduction in growth due to environmental stresses, whereas active plasticity requires a specific signal perception-transduction system allowing an organism to respond by changing its development (Van Kleunen & Fischer 2005). Active plant response to

stresses involves complex mechanisms with many signaling molecules, transcription factors and stress-responsive genes and proteins, **figure 74 & 75**. Schliting & Smith (Schlichting & Smith 2002) distinguished molecular and phenotypic plasticity: the former refers to the ability of a genotype to produce different levels or isoforms of transcriptome and proteome, while the latter is the ability of a genotype to produce different phenotypes resulting from all molecular changes, **figure 76**.

Class of target	Examples	Possible mode(s) of action
Osmoprotectants	Amino acids (proline, ectoine) Dimethyl sulfonium compounds (glycine betaine, DMSP) Polyols (mannitol, D-ononitol, sorbitol) Sugars (sucrose, trehalose, fructan)	Osmotic adjustment; protein/membrane protection; reactive (OH·) scavenging
Reactive oxygen scavengers	Enzymatic (catalase, Fe/Mn superoxide dismutase, ascorbate peroxidase; glutathione cycle enzymes: glutathione S-transferase, glutathione peroxidase; gamma-glutamylcysteine synthetase, alternative oxidase) Non-enzymatic (ascorbate, flavones, carotenoids, anthocyanins)	Detoxification of reactive oxygen species
Stress proteins	Late embyogenesis abundant proteins	Unknown, protein stabilization, water binding/ slow desiccation rates; chaperones; protein/ membrane stabilization; ion sequestration
Heat shock proteins	Various heat-, cold-, salt-shock proteins in several subcellular compartments	Reversal/prevention of protein unfolding; translational modulation
Ion/proton transporters	High-affinity K^+ transporter; low-affinity K^+ channels; plasma membrane, pre-vacuolar, vacuolar and organellar proton ATPases and ion transporters (H^+/ATPase; Na^+/H^+ antiporters)	K^+/Na^+ uptake and transport; establishment of proton gradients; removal and sequestration of (toxic) ions from the cytoplasm and organelles
Membrane fluidity	Fatty acid desaturases	Increased amounts of dienoic and fluidity; chilling tolerance
Water status	Aquaporins or water channels (solute facilitators: urea, glycerol, CO_2, possibly others and including ions); CO_2 concentration	Regulation of AQP amount differentially in tonoplast and plasma membrane; regulation of membrane location; stomatal behavior
Signaling components	Homologs of histidine kinases (AtRR1/2); MAP kinases (PsMAPK, HOG); Ca^{2+}-dependent protein kinases; SNF1/kinases; protein phosphatases (ABI1/2); CNA/B signaling systems; Ca^{2+} sensors (SOS3); inositol kinases	Ca^{2+}-sensors/phosphorylation mediated signal transduction
Control of transcription	Transcription factors: EREBP/AP2 (DREB, CBF); zinc finger TF (Alfin 1); Myb (AtMyb2, CpMyb10)	Upregulation/activation of transcription
Growth regulators	Altered biosynthetic pathways or conjugate levels for abscisic acid, cytokinins and/or brassinosteroids	Changes in hormone homeostasis

ABI, abscisic-acid-insensitive; AP2, APETELA2; AQP, aquaporin; AMPK1, AMP-activated protein kinase; AtMyb, *Arabidopsis thaliana* myeloblastosis (helix-loop-helix) transcription factor; AtRR1, *A. thaliana* two-component response regulators; CBF, C-repeat/DRE binding factor; CNA/B, calcineurin A/B; CpMyb, *C. plantagineum* myeloblastosis (helix-loop-helix) transcription factor; DMSP, dimethylsulfoniopropionate; DREB, dehydration-responsive element (DRE) binding protein; EREBP, ethylene-responsive element binding protein; HOG, high osmolarity glycerol; PsMAPK, *Pisum sativum* mitogen-activated protein kinase; SNF1, sucrose non-fermenting 1; TF, transcription factor.

Figure 75: The hierarchy of gene expression underlying phenotypic plasticity (from Schlichting & Smith 2002).

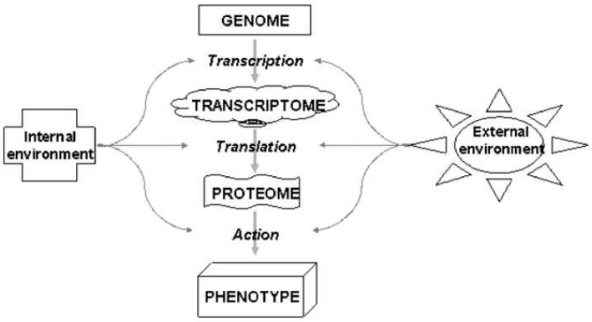


Figure 76: The complexity of stress adaptation (from Cushman & Bonhert 2000).

1.3. Gene-by-environment interactions:

In a set of individuals, the global effect of the environment (E) catches the mean phenotypic change of the different genotypes, while gene-by-environment interactions (I_{GxE}) expresses that the different genotypes display different phenotypic plasticity. The fact that different genotypes display different reaction norms in response to constraints suggest that plasticity may be heritable. **Figure 77** for example shows different effects of flooding on net photosynthesis rates of 13 hybrid poplar clones.

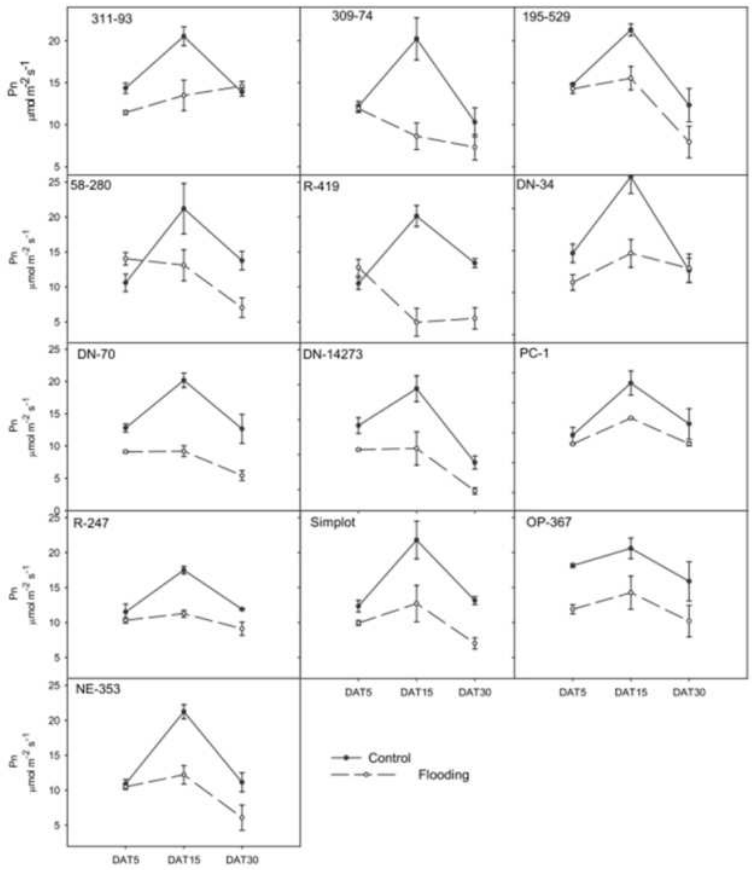


Figure 77: Net photosynthesis rates (μmol.m^{-2}.s^{-1}) of 13 hybrid poplar clones under control and flooding conditions (from Guo et al. 2011).

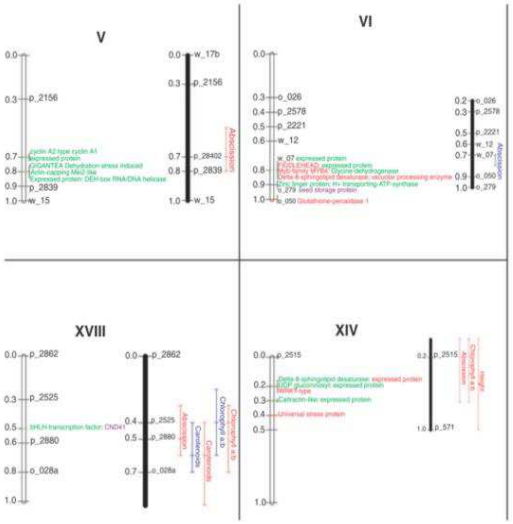


Figure 78: Genetic map locating QTL loci involved in response to drought in *Populus deltoides* and *P. trichocarpa* species (from Street *et al.* 2006).

Because proteins are encoded by genes, proteins involved in phenotypic plasticity may potentially be mutated causing differences in the extent of morphological and physiological responses between genotypes.

QTL mapping helps identifying such genes. For example, Parelle *et al.* (Parelle *et al.* 2010) reviewed the studies that located QTL involved in the tolerance to soil water-logging in many plant species. In the same way, Street *et al.* (2006) identified numerous QTL involved in the response to drought in *Populus*, **figure 78**.

At the population level, significant provenance-by-environment interactions are

commonly interpreted as a result of local adaptation (see section '2. Phenotypic evolution in populations').

1.4. Maternal effects

Genetic variation must be distinguished from maternal effects that are defined as the causal influence of the maternal genotype or phenotype on the offspring phenotype (Wolf & Wade 2009). Maternal effects have themselves both genetic and environmental components. For example, seed stored reserve compounds may influence growth of the progeny. In trees, this maternal effect is influenced by environmental effects, such as resource availability and successional status of mother trees (understory, emergent), but also by genetic effects, as different mother trees may produce seeds of different quality. Seed properties are also influenced by the ontogeny of mother trees, as seed production may vary across lifetimes. That is why seed mass is commonly included as co-factor in linear models of phenotypic value decomposition.

Epigenetic inheritance is also a major maternal effect. The term 'epimutation' refers to the methylation of coding DNA that prevents its transcription into mRNA. Epigenetic changes may be induced by particular environmental conditions and are transmitted from mothers to their progeny. However, such changes in DNA structure have to be distinguished to 'true mutations', as 'epimutations' do not affect the DNA sequence.

Because of their environmental control, maternal effects may be viewed as a trans-generational phenotypic plasticity, or a reaction norm that extends across generations (Mousseau & Fox 1998).

2. Phenotypic evolution in populations

Contrary to neutral processes, natural selection affects genetic diversity by filtering genotypes across generations according with their fitness that is modulated by the phenotypic value of adaptive traits. An **adaptive trait** may be defined as being **variable**, **heritable** and **functional** (Howe & Bruner 2005). A 'Functional trait' is any morphological, physiological, or phenological trait that influences plant fitness (Geber *et al.* 2003). Thus, it refers to a broad range of individual-level and organ-level traits, **figure 79**. In some cases, phenotypic plasticity for fitness-related traits may also be adaptive. Indeed, phenotypic plasticity is highly important in plants because they are immobile and migration toward a more favorable environment requires the establishment of a new population (Shao *et al.* 2007). Phenotypic plasticity is thus primordial for plants to cope with environmental heterogeneity and phenotypic plasticity for some traits may be linked with plant fitness (Van Kleunen & Fischer 2005). However, the developmental cost of active responses to stresses prevents the appearance of Darwinian monsters with infinite plasticity (Pigliucci 2001). Considering phenotypic plasticity as a functional trait underlies that the relationship between plasticity and fitness is probably non-linear (increasing plasticity would be beneficial until a limit above which increasing costs of the plasticity alter plant fitness). For these reasons, the outcome of evolution is often a reduction in plasticity (called 'genetic assimilation'), except in changing environments in which phenotypic plasticity may be selected for and conserved across generations.

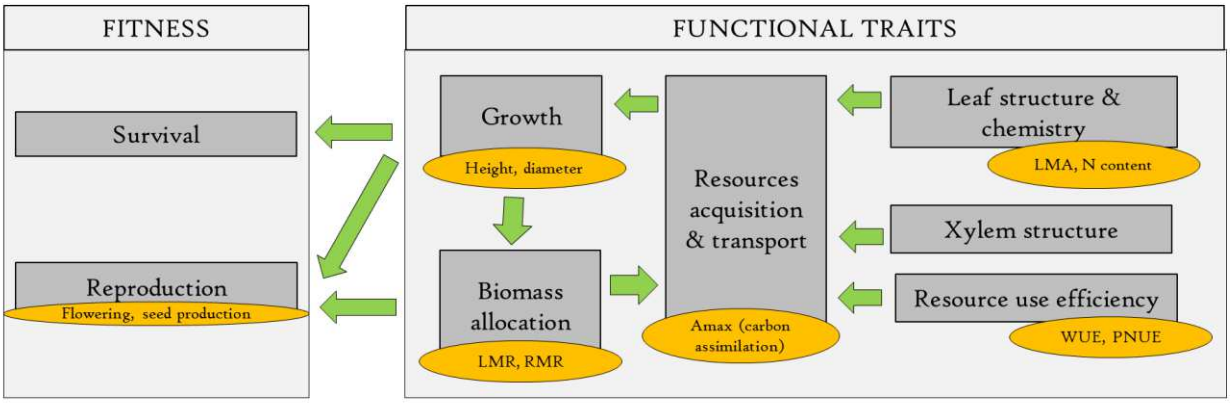


Figure 79: Short overview of major functional traits in plants.

The evolution of the distribution of phenotypic values in populations under selection depends on the relationship between trait and fitness. Commonly, functional traits are related to fitness through a polynomial model:

$$W_i = aP_i^2 + bP_i + c,$$

where W represents the individual fitness and P the individual phenotype. The sign of 'a' determines the kind of selection: **directional** when a=0, **stabilizing** when a<0 and **diversifying** when a>0.

I simulated the evolution of population mean phenotypic values for a functional trait under natural selection using a simplified model of phenotypic evolution: the functional trait is controlled by a single locus (with 4 alleles) assuming no genetic drift and constant population size (N=10000), **figure 80.**

(1) Each genotype is characterized by a phenotypic value called 'genotypic value' (G_i). The model assumes additivity without dominance between alleles.

(2) Individual phenotypic values (P_i) are drawn from a normal law of sd=0.5: $P_i \sim N(G_i, \sigma^2_{res})$, where σ^2_{res} is the residual variability, among which phenotypic plasticity.

(3) Individual fitness is estimated according as: $W_i = aP_i^2 + bP_i + c$,

(4) The next generation is produced according with the mean fitness of the different genotypes.

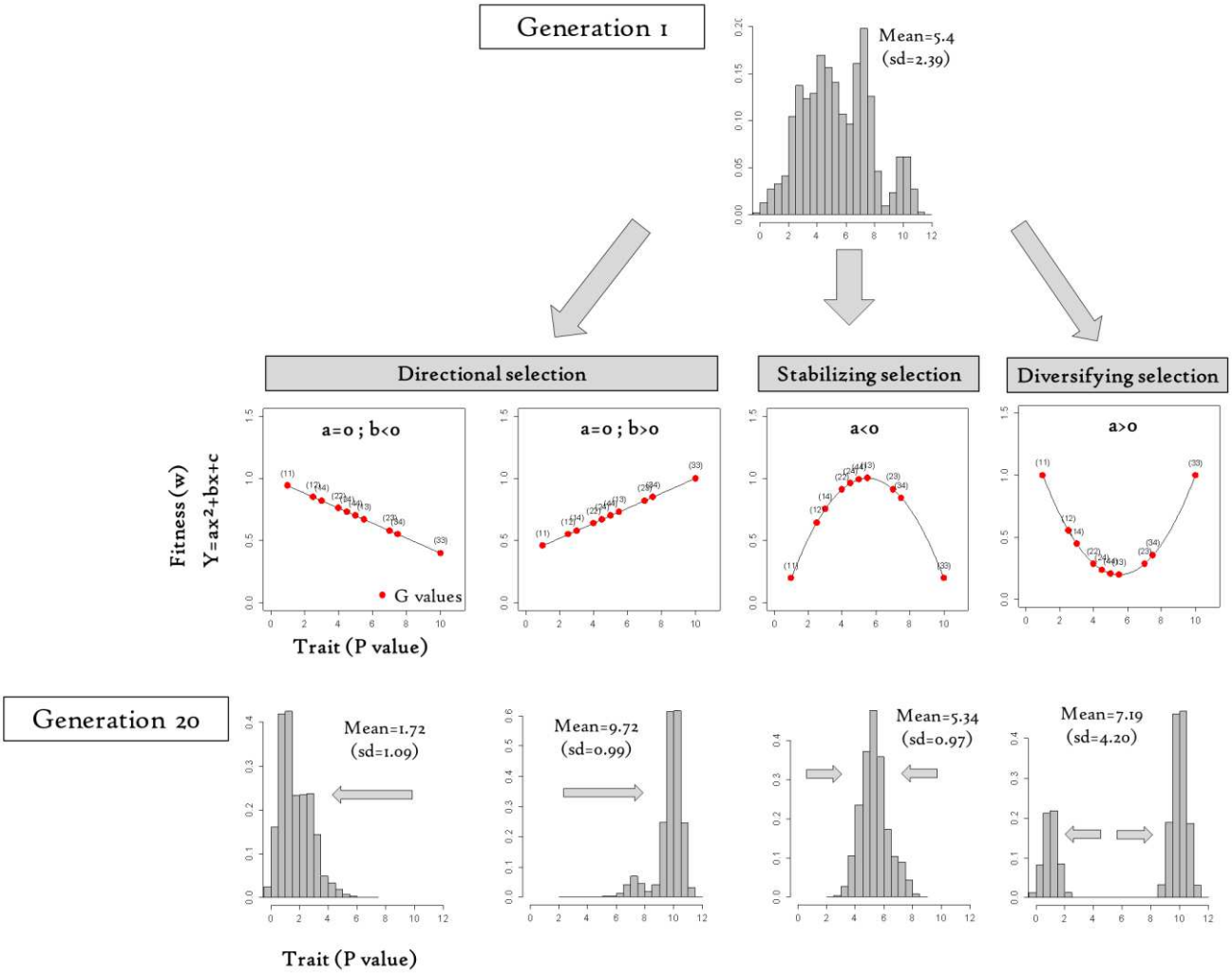


Figure 80: Phenotypic evolution under three kinds of natural selection: directional, balancing and diversifying. The first line describes the relationship between the phenotypic value of a trait (P) and the fitness (W). The second line shows the distribution of phenotypic value after 20 generations.

Despite its simplicity, the model is able to illustrate how natural selection would drive the mean phenotypic value in populations. **Directional selection** shifts the population mean toward low or high phenotypic values and reduces the phenotypic variance in the population. **Stabilizing selection** results in a population with the same population mean than the original population but with restricted phenotypic variance while **diversifying selection** increases phenotypic variance by favoring genotypes with extreme phenotypic values.

3. Phenotypic differentiation

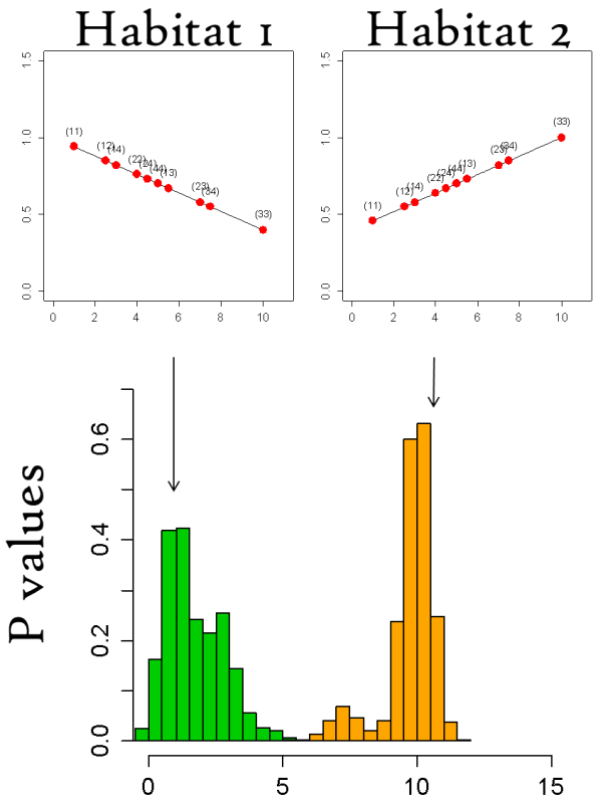


Figure 81: Phenotypic divergence among habitats submitted to divergent selection (i.e. different directional selection within populations).

If the relationship between traits and fitness varies among habitats, natural selection drives the genetic structuring of functional traits and results in an inherent phenotypic differentiation between populations submitted to **divergent selective pressures**, **figure 81**.

Because natural selection commonly structures phenotypic traits and because current patterns of phenotypic differentiation result from past evolution, **analyzing patterns of phenotypic trait distribution in populations distributed along environmental gradients** is one of the most efficient ways to study **local adaptation** with phenotypic traits. It requires however to distinguish the genetic and environmental sources of phenotypic variation through specific experiments such as **provenance tests** and **reciprocal transplant experiments** (**box 7**) as a phenotypic differentiation between sub-populations observed *in situ* may result from different plastic response without implying any genetic differentiation (and then no local adaptation).

The structuring of phenotypic traits across natural landscapes is largely documented in both woody and non-woody plant populations: the literature is rich with examples of adaptations based on phenotypic traits that involve a wide range of environmental gradients (altitude, latitude) or specific constraints (temperature, water availability, light, pathogens).

BOX 7 - THEORY

Studying patterns of phenotypic divergence from provenance tests and reciprocal transplant experiment

Several experiments may help apprehending genetically-driven phenotypic variability. They consist in submitting different genetic groups (mainly populations also called 'provenances' or 'demes') to common conditions or to a panel of environmental conditions.

Provenance tests in common garden aim at quantifying the extent of genetically-driven phenotypic variations in a single environment. In these experiments, phenotypic variation is attributable to genetic variability among provenances, according with the model $P=\mu+G+res$. Common gardens are often realized under non-limiting conditions that allow the expression of genetically-driven phenotypic divergence among genetic groups, even if these conditions differ from the natural conditions encountered in the wild.

Even if provenance tests in common (often non-limiting) conditions allow inferring patterns of genetically-driven phenotypic variations, they cannot conclude about the implication of local adaptation in patterns of phenotypic differentiation among populations.

Both **provenance tests in multiple conditions** and **reciprocal transplants** aim at dissociating genetic and environmental sources of phenotypic variations, according to the model: $P=\mu+G+E+I_{GXE}+res$. These experiments allow estimating the relative influence of environmental and genetic factors on total phenotypic variability. Mainly, they allow distinguishing the global differentiation between genetic groups over all environmental conditions (G) from the differentiation in the mean response to constraints (including both passive and active plasticity) among the genetic groups (I_{GxE}). 'Genes-by-environment' interactions are particularly important when studying local adaptation, as we would expect significant I_{GxE} for fitness-related traits when different populations are locally adapted to the environmental conditions.

Provenance tests in multiple (and often constraining) conditions allow testing the sensitivity of provenance to specific constraints by targeting one or several environmental constraints. However, these experiments cannot test all environmental factors encountered in nature (with all their interactions), leading to non-generalizable results. **Reciprocal transplants**, on the contrary, aim at testing local adaptation *in situ*, even if identifying the environmental factors involved become highly difficult as natural gradients are commonly associated with variations in numerous factors that may be inter-correlated.

Two main approaches are commonly involved for interpreting I(GxE) and to infer the contribution of local adaptation:

- '**Local versus Foreigner**' emphasizes the comparison between populations (or 'provenances') within habitats: local populations are expected to show a higher fitness than foreign demes.
- '**Home versus Away**' emphasizes the comparison of a population's fitness across habitats.

However, these propositions are not equally relevant for testing local adaptation. Indeed, the 'local vs foreigner' criterion addresses the efficiency of divergent selection relative to other evolutionary processes, whereas 'home vs away' confounds the effects of divergent selection with environmental effects due to habitat quality for example (Kawecki 2004).

Adaptation across latitudinal gradients

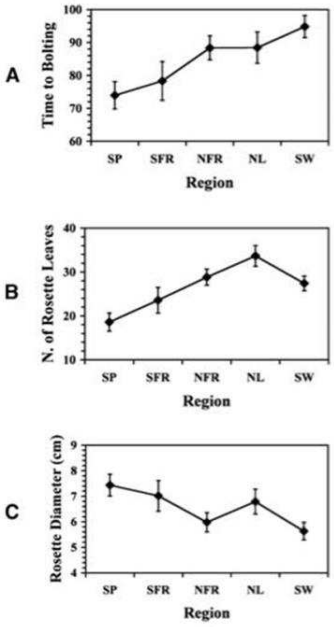


Figure 82: Variations for vegetative traits across regions of origins (from Banta et al. 2007).

In the biological model *Arabidopsis thaliana*, Banta *et al.* (2007) revealed significant differences among populations originating from different latitudes in Europe for several vegetative traits (bolting time, number of rosette leaves and rosette diameter). Their results are largely consistent with adaptation as they found an ordered gradient of phenotypic differentiation according with the latitude of origin, **figure 82**. Moreover, they found a significant gene-by-environment interaction (region of origin X growth chamber differing in light photoperiod) for a fitness trait (number of fruits). However, interactions patterns did not reveal a clear adaptive advantage of seedlings that naturally experienced a given photoperiod compared to others, making the result hard to interpret ecologically. In *Picea sitchensis*, Mimura & Aitken (2010) found significant differences among provenances originating from different latitudes of the Pacific coast of North America for bud set, seedlings biomass and growth period, **figure 83**. They also found significant gene-by-environment interactions (region of origin X growth chamber simulating temperature and photoperiod of the different provenances sites) for height increment and growth rate increment concordant with the 'local VS foreign' criterion.

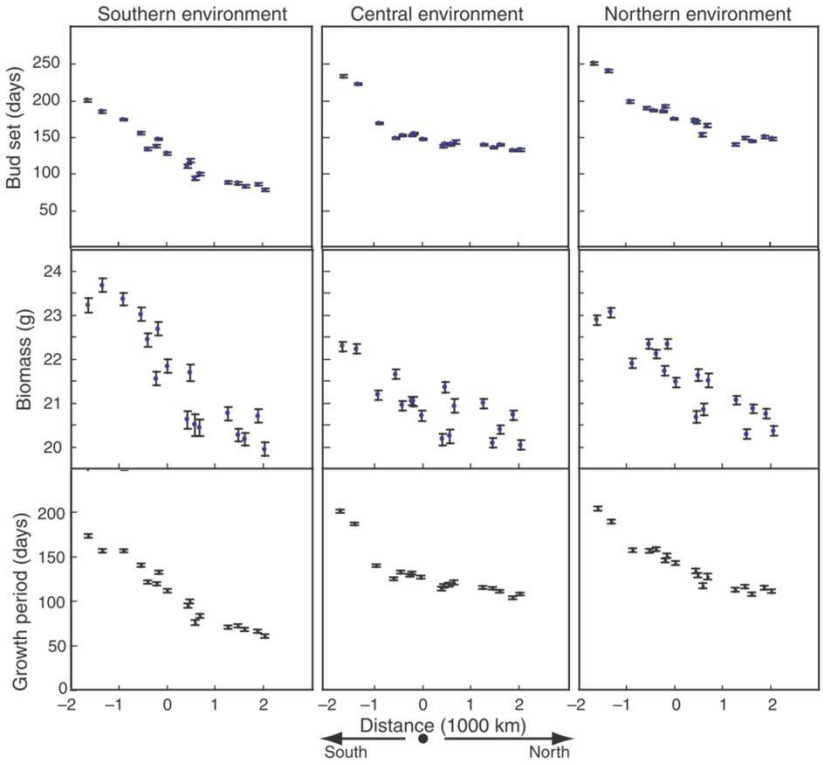


Figure 83: Phenotypic clines in selected growth response variables in three chamber environments by population of provenance (from Mimura & Aitken 2010).

Adaptation across altitudinal gradients

Altitudinal gradients also provide great examples of local adaptation in small plants. Gonzalo-Turpin & Hazard (2009) used a reciprocal transplant experiment to study local adaptation in the mountain plant *Festuca eskia*. They found significant differences in survival rate, growth traits (height, diameter), leaf traits (LDMC, SLA), and reproductive traits (reproductive output allocation, spike number, seed weight) among provenances originating from different altitudes of the Pyrenees mountains: plants from low altitudes displayed lower survival, higher height and diameter, smaller SLA and higher reproductive fitness than plants from high altitudes. They suggested that selection favored higher SLA at high elevation (that subsequently increased their efficiency of light interception and carbon gain). Producing low-cost leaves at high elevation would permit plants to overcome constraints of short-growing season. Moreover, they found significant gene-by-environment interactions for survival and seed weight: plants from middle and high altitudes appeared well adapted to their environment according to 'local vs foreign' criteria for survival, **figure 84**, while plants from low altitudes growing at their home altitude produced heavier seeds than the others.

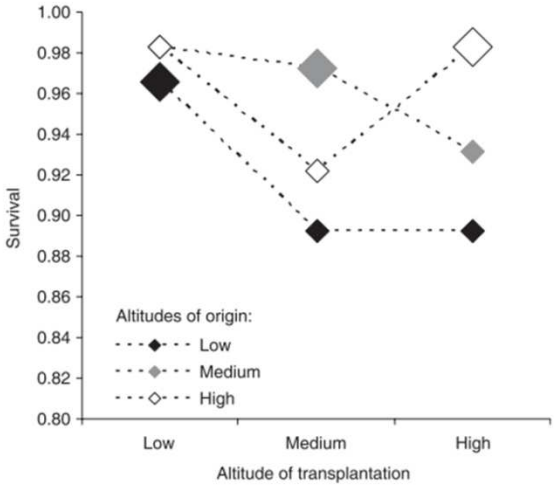


Figure 84: Effect of altitude in transplanted *F. eskia* from three different altitudes (from Gonzalo-Turpin & Hazard 2009).

In a similar experiment, Byars *et al.* (2007) provided evidence of local adaptation in *Poa hiemata*. They found significant differences among provenances for circumference and leaf length, with larger stem circumference and shorter leaf lengths in plants originating from high altitudes. They suggested that these traits have undergone past directional selection even if the exact reason why shorter leaves and wider circumference were selected for at high elevations is not obvious. They also found significant gene-by-environment interactions for survival: genotypes tended to survive better at the same altitude from which they originated, suggesting a fitness advantage for populations growing at their home site.

Tree populations also provide clear examples of inherent phenotypic differentiation along altitudinal gradients. In the Gymnosperm *Picea abies*, Oleksyn *et al.* (1998) found significant relationship between phenotypic traits and the elevation of the native provenances for growth (Height and DBH increment), biomass allocation and carbon assimilation using a common garden experiment, **figure 85**. Significant relationships were also found between these traits and the mean annual temperature of the provenance sites, suggesting that the phenotypic structuring may be partly adaptive and driven by temperatures.

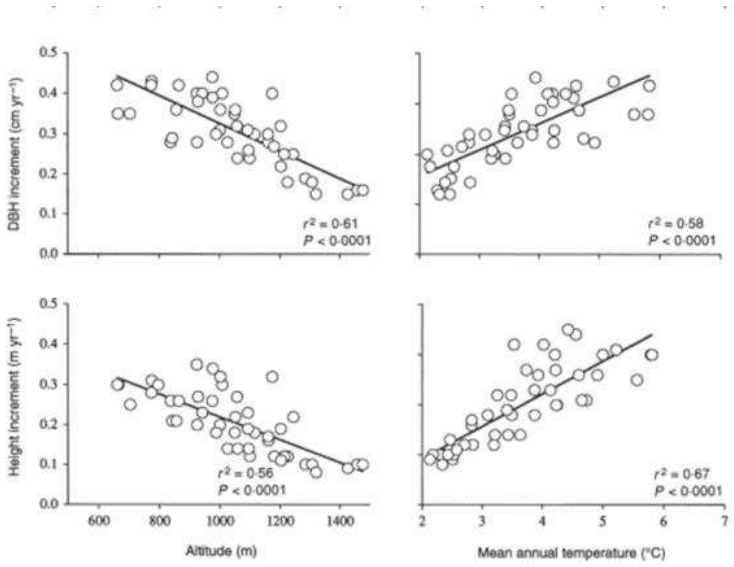


Figure 85: Relationship between altitude of provenance and seedling growth in a common garden (from Oleksyn *et al.* 1998)

Even if studies based on natural gradients provide clear examples of local adaptation, the ecological interpretation of the observed variation in traits is often complicated as natural gradients are complex and associated with changes in many biotic and abiotic factors (moisture, temperature, exposure, wind, soil conditions, competition, predation etc...) that cannot be isolated from each other.

Adaptation to drought

In *Impatiens capensis*, Heschel & Riginos (2005) revealed significant differences in water use efficiency (WUE) as well as in stomatal conductance (gs) and leaf size among provenances when submitted to drought: they found that populations originating from dry sites decreased their stomatal conductance to a larger extent than populations from wet sites when submitted to soil water deficit by comparison with a well-watered treatment. Moreover, plants from dry sites had smaller leaves in well-watered conditions but equal leaf size than plants from wet sites when submitted to soil water deficit. Last, plants from dry populations flowered earlier and produced more flowers and fruits than plants from wet sites whatever the soil water availability. They suggested that it may be adaptive for *I. capensis* to maximize carbon assimilation through early-flowering for plants originating from dry sites. Similarly, Rajakura *et al.* (2003) observed that races of *Lasthenia californica*

from dry sites were able to maintain reproductive fitness under low water availability, suggesting they were quite well adaptation to soil water deficits.

In the tree species *Pinus pinaster*, Nguyens-Queyrens & Bouchet-Lannat (2003) found that the negative relationship between relative water content and osmotic adjustment in needles varied between provenances originating from sites differing in annual rainfall, **figure 86**. They suggested that the different populations probably developed divergent strategies, some limiting water loss by stomatal closure (wet provenances), and others favoring water circulation with the help of an integrated whole-plant strategy of which osmotic adjustment represents one mechanism (dry provenances).

In *Fagus sylvatica*, Meier & Leuschner (2008) revealed that populations from drier sites of Germany allocated more carbon to roots and displayed a larger fine root turnover.

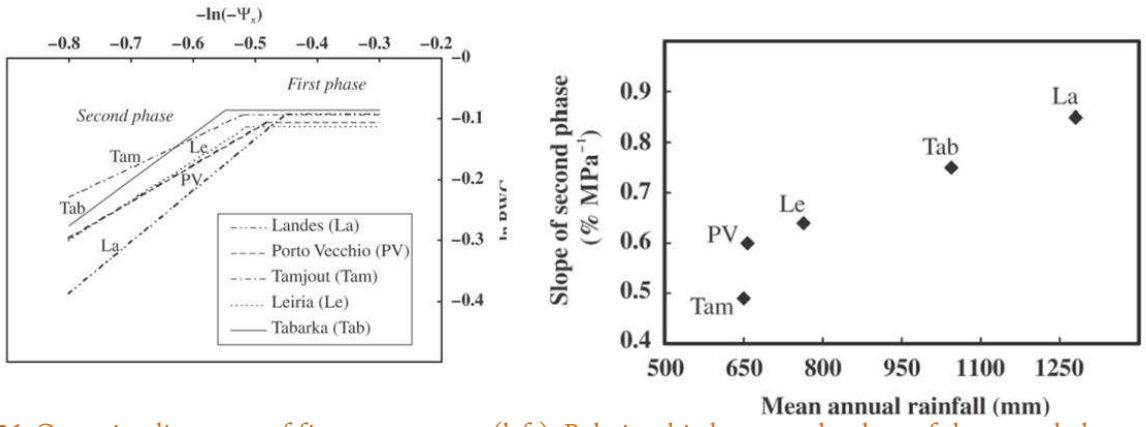


Figure 86: Osmotic adjustment of five provenances (left). Relationship between the slope of the second phase of the osmotic adjustment curves and mean annual rainfall at the site of origin (right). From Nguyen-Queyrens & Bouchet-Lannat 2003.

Adaptation to other constraints (light, herbivores)

In the Mediterranean Fagaceae *Quercus coccifera*, Balaguer *et al.* (2001) revealed that populations of the Iberian Peninsula differed in their phenotypic plasticity in response to irradiance for nutrient content and partitioning, leaf size, leaf area ratio and for crown architecture. These differences suggested ecotypic differentiation toward a lower phenotypic plasticity in the most homogeneous irradiance environment (forest by comparison with garrigue and rock provenances).

In *Quercus rubra* populations occupying a Missouri oak-hickory forest, seedlings showed less herbivore damage when planted at the site of the maternal plant, **figure 87** (Sork *et al.* 1993).

Adaptation in tropical rainforests?

The literature is however poorer in tropical than in temperate ecosystems. Several studies have already reported a structuring of phenotypic traits among trees originating from different areas (*Guazuma crinita* (Rochon *et al.* 2007) and *Calycophyllum spruceanum* (Sotelo Montes *et al.* 2007) in Peruvian Amazon, *Swietenia macrophylla* (Wightman *et al.* 2007) and *Cedrela odorata* (Ward *et al.* 2008) in the Yucatan Peninsula of Mexico) but the aim of these studies was mainly to select varieties with interesting wood properties. By contrast, only few studies addressed the question of local adaptation in driving phenotypic divergence across forest landscapes in tropical areas.

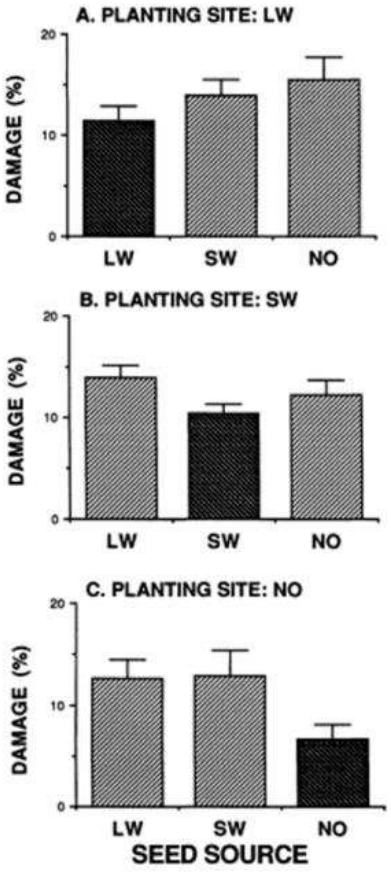


Figure 87: Percentage of leaf damage depending on population origin and transplant sites (from Sork *et al.* 1993)

In *Eucalyptus marginata* populations established in Australia, O'Brien *et al.* (2007) found that trees from low rainfall sites had smaller stem diameter. They suggested that lower growth may be a strategy to prevent drought stress. Moreover, they found that seedlings from high rainfall sites had poorer survival in drier sites than seedlings originating from these sites, suggesting that adaptation to drought may be involved. In *Cedrela odorata* established in Costa rica, Navarro *et al.* found that seedlings from the dry areas were taller, had higher diameter and had higher leaflets than those from wet sites. Rapid growth would facilitate plant survival during the dry season after short wet periods. In the tree species *Parapiptadenia rigida* established in Brazil, Silva et al. (2010) found evidence of ecotypic differentiation in relation to flooding for root properties, aerenchyma formation, growth recovery after flooding and leaf production. In the Lamiaceae *Aegiphila sellowiana*, Medri *et al.* (2011) found that plant surviving to flood were genetically distinct from plant not surviving.

In this study, I used provenance tests in both controlled (shadehouses) and wild conditions (reciprocal transplants) to by decompose phenotypic variations into genetic and

environmental factors and to assess whether growth and leaf traits were structured in relation to local environmental patchiness in *Eperua falcata* (see below for an overview of the main results, complete results and discussion are describe in the article n°4).

Summary of PhD results:

Local habitat patchiness is associated with a strong genetic divergence for phenotypic traits in seedlings growing in non-limiting conditions (see article n°4).

As described previously, we used a common garden experiment in non-limiting conditions to study phenotypic differentiation within continuous populations occupying different habitats for two congeneric, sympatric, and ecologically divergent tree species (*Eperua falcata* and *E. grandiflora*, Fabaceae). We tested (a) whether conspecific populations growing in different habitats diverge at functional traits and (b) whether they diverge in the same way as congeneric species having different habitat preferences.

In both species, seedling populations native of different habitats displayed phenotypic divergence for several traits (including seedling growth, biomass allocation, leaf chemistry, photosynthesis and carbon isotope composition), **figure 88**. This may occur through heritable genetic variation or other maternally inherited effects. Our results indicate that mother trees from different habitats transmit divergent trait values to their progeny, and suggest that local environmental variation selects for different trait optima even at a very local spatial scale. Traits for which differentiation within species follows the same pattern as differentiation between species indicate that the same ecological processes underlie intra- and interspecific variation.

Seedling native from different native habitats are equally affected by drought and water-logging constraints (see article n°5).

In parallel of the common garden experiment in non-limiting conditions, a set of seedlings was submitted to six months of highly constraining hydric conditions: severe drought and water-logging. We hypothesized that local heterogeneity may have driven a divergence in seedlings sensitivity to hydric constraints between sub-populations coming from different habitat types.

The results revealed a significant effect of provenance (as already observed in non-limiting conditions) as well as strong effects of the treatment common to all provenances. However, no differences between provenances were detected in constraining conditions for any growth or leaf trait.

For example, both constraints affected seedling growth (by decreasing seedlings growth rate and total biomass in the case of drought), induced a shift in biomass allocation to leaves by decreasing seedlings LMR, and induced a change in leaf mass per area that increased in a greater degree in response to water-logging than in response to drought. While drought induced no change in RMR, water-logging induced a significant decrease in biomass allocation to roots, probably because of the death of the ancient root system that was replaced by adventitious roots (**figure 89**).

However, no differences between provenances were detected in constraining conditions for any recorded traits, thus revealing that the seedlings coming from different habitat types were equally affected by drought and water-logging. These results suggest the genetically-driven phenotypic differentiation between the provenances is not a result of local adaptation to hydric conditions. This experiment does not allow, however, to completely exclude the influence of local adaptation in driving the genetic structuring between local conditions, as (i) micro-habitats differ not only in hydric conditions, but also in a variety of other environmental factors including many abiotic and biotic factors and (ii) the constraints exerted may have been too severe, and we lack information about the reaction norm of the different provenance to each constraint.

Dissecting genetic and environmental sources of *in situ* phenotypic variations is going to be assessed through a reciprocal transplant experiment (see 'Preliminary results of reciprocal transplants').

In a third time, we realized a reciprocal transplant experiment involving the two study sites of Laussat and Regina to test the local adaptation hypothesis in wild conditions. We sampled seeds from two habitat types (bottomland and terra-firme) in the two study sites and transplanted young seedlings in all sites and micro-habitat conditions. This experiment was set up at the beginning of this PhD and will be followed until 2015.

Up to now, significant effects of provenances and transplant sites were detected on seedling growth at both regional and local scales, but any provenance-by-transplant interaction was detected. Subsequently, we detected any difference between local and foreigners in the different transplant sites. However, the seedlings grow very slowly in the wild by comparison with those placed in shade houses, and they are probably too young to detect clear effects. It is thus too early to interpret properly this experiment and reciprocal

transplant experiments classically require more than three years to provide sufficient divergences and significant results.

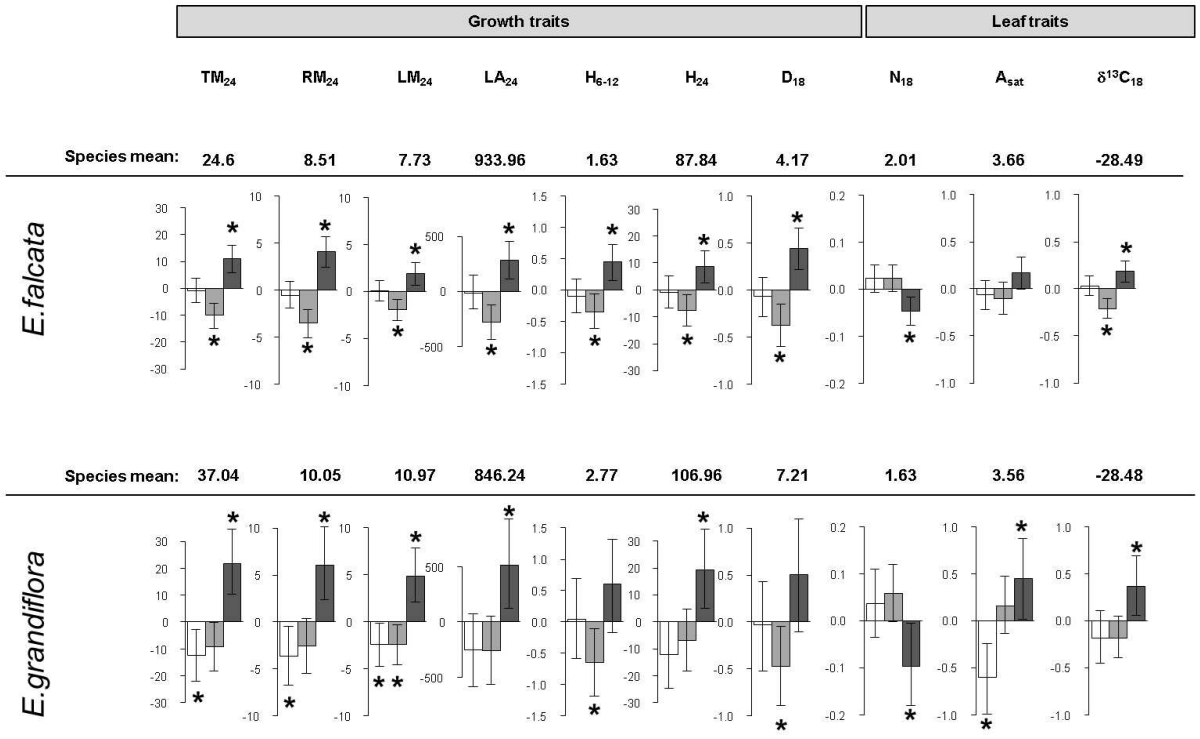


Figure 88: Phenotypic differentiation among habitat types for growth, biomass allocation and leaf traits for two species sampled at Paracou. Bayesian departures of each group from the global mean are shown as boxes; error bars show the 95% credible interval of the estimated parameters. For each plot: left box=hilltop, middle box=slope, right box=bottomland.

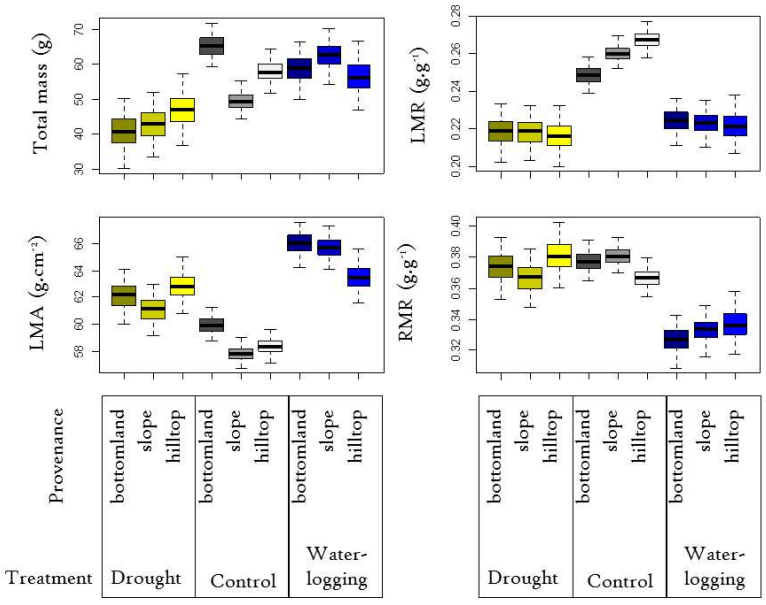


Figure 89: Bayesian estimates (with 95% credible intervals) of phenotypic values displayed by the different provenances (habitat types) in the different hydric conditions (drought, control, water-logging) for a subset of the recorded traits: total biomass, leaf mass per area (LMA), leaf mass ratio (LMR) and root mass ratio (RMR).

DISCUSSION

1. Neutralism and adaptation in *Eperua falcata*

The study of species evolution and genetic diversification in tropical rainforests remain a vast and difficult topic. The combination of an immense diversity of species, the few number of studies available and the difficulties in studying species and populations in such diverse (and sometimes hostile) environment has largely limited the comprehension of the mechanisms involved. However, more research efforts have been made during the twenty past years and help understanding how species evolve in such particular environments.

Recent biogeographic and phylogeographic studies have investigated the building of biodiversity in Amazonia through geological ages, mainly through the study of past processes of colonization of species from other continents and of species diversification caused by orogenic and climatic (ice ages) changes (eg. Hoorn *et al.* 2010, Scotti-Saintagne *et al.* 2012, Duminil *et al.* 2006). Moreover, several community ecology studies have observed profound changes in forest community structure and compositions in relation to environmental conditions (eg. Kahn 1987, Kraft *et al.* 2008); and have highlighted the probable influence of adaptive radiation in speciation processes (eg. Baraloto *et al.* 2007). In particular, many tree genera are composed by species differing in their ecological preferences to particular environmental conditions, among which local habitat patchiness caused by topography attracts a particular attention.

Even if the processes causing the spatial genetic structuring in tree populations are well documented in temperate zones, tropical rainforests suffer from a lack of knowledge about the process of populations evolution and the genetic structuring of tree populations at intra-specific level (at a level of genetic differentiation more recent than the divergence between species), Savolainen *et al.* 2007. Many studies have already provided evidence of genetic differentiation in temperate plant populations caused by both neutral and adaptive processes involving a large variety of environmental gradients and subjacent ecological factors such as climatic factors (temperature, precipitation), edaphic factors (soil properties among which soil water availability, nutrients), and biotic interactions (competition, predation, mutualism etc...), Caisse & Antonovics 1978, Gonzalez-Martinez et al. 2006, Savolainen *et al.* 2007, Leimu & Fischer 2008, Siol *et al.* 2008, Savolainen *et al.* 2011, Strasburg

et al. 2011, Le Corre & Kremer 2012. In the tropical rainforest of Amazonia, many studies have already addressed the question of neutral differentiation caused by restricted gene flow and local inbreeding (Ward *et al.* 2005, Hardy *et al.* 2006, Dick *et al.* 2008), but no study has yet integrated both neutral and adaptive aspects of populations evolution at intra-specific level in Amazonia.

In this study, I used a common tree species of the Guiana shield (*Eperua falcata*) to study how evolution has structured the genetic diversity in Neotropical forest landscapes. Both molecular and phenotypic approaches provide evidence of **a strong genetic structuring over very local spatial scales** (only several hundred of meters) due to **a combination of neutral** (mainly limited seed dispersal and probably local inbreeding) **and adaptive processes** (driven by environmental factors associated local habitat patchiness).

Based on more than one thousand AFLPs loci (**article n°2**), a **genome-scan approach** revealed a strong fine-scale genetic structuring (SGS) with a strong relatedness between adult trees closer than few dozens of meters to each other in the study sites of Laussat and Regina. This result was corroborated by the blind analysis of genetic structuring within sites, revealing that related trees are geographically grouped and the probable clumping of progeny arrays, as it is the case in numerous Neotropical tree species (Ward *et al.* 2005, Hardy *et al.* 2006, Dick *et al.* 2008). Moreover, gene dispersal distances were estimated to be very low in these two populations (~46 and ~64 meters in Laussat and Regina respectively). Thus, a combination of a restricted gene flow with high local densities may have driven neutral genetic differentiation at very local scales (hundreds of meters) in these two populations. We can also easily imagine that such clumping would be reinforced by local inbreeding, as mating would occur mainly among neighbors in populations of high population densities. This structuring of genetic diversity results in a significant, albeit small, genetic differentiation among local habitat types in these two study sites (Fst~0.03 in Regina and Fst~0.04 in Laussat).

Even if the major part of the genetic differentiation among local habitats (bottomlands and terra-firme) may be attributable to neutral processes, the genome scan approach revealed that several loci (between 0.3% and 1.8%) may be however structured by selective processes associated with variations in topography and soil properties. In particular, many outliers displayed similar patterns of band frequency variations among

local habitats in the two study sites. Moreover, local adaptation may contribute to explain the very low estimated gene flow, as all genotypes are not necessarily able to establish in all micro-habitats. The proportion of outliers detected here is however very low, probably because local adaptation has a minor contribution in governing the genetic differentiation over short spatial scales. It is however possible that the footprints of natural selection detected here do not catch the entire extent of divergent selection among local habitats. In some cases, when many loci are involved in local adaptation, and when natural selection models the genetic correlations between loci rather than fixing alleles at individual loci, the differentiation at selected loci ('Fst_{QTL}') is close to the neutral genetic differentiation estimated overall loci (Fst) and results in a lack of power to detect outliers using Fst-based selection tests (Le Corre & Kremer 2012). Because local adaptation may act on complex traits (i.e. traits controlled by a complex genetic architecture), it is possible that we fail to detect many selected loci. However, the main limit of this approach is that the molecular markers analyzed are anonymous and cannot provide any information about the genetic role potentially played by outlier loci.

A **candidate gene approach** (**article n°1**) revealed that divergent selection among local habitats affect SNPs within two genes of functional importance (a farnesyl-transferase and a catalase) and within one locus of unknown function in the study sites of Paracou and Nouragues (Audigeos 2010, Audigeos *et al.* 2013). The identity of the genes submitted to divergent selection is an important piece of information for understanding which factors are involved in the process of local adaptation. The farnesyltransferase gene is a negative regulator of abscissic acid signal transduction in guard cell that is a major hormone involved in stomatal closure during drought stress (Cutler *et al.* 1996, Schroeder *et al.* 2001). The catalase, on the contrary, may be related to oxidative stresses induced by hypoxia (Willekens *et al.* 1997, Mittler 2002, Blokhina *et al.* 2003): as soil aeration and root respiration decrease during water-logging, and hydrogen peroxide is produced by mitochondria and accumulated into root cells. The catalase contributes to detoxify the hydrogen peroxide through the reaction: $2H_2O_2 \rightarrow O_2+2H_2O$. However, this study involved too few loci to prove adaptation of *Eperua falcata* sub-populations to particular soil water constraints and associated stresses (mainly drought and hypoxia), and would be completed by a study including many other candidate genes.

Even if both genome scan and candidate gene approaches have several limits, they converge toward the idea that local adaptation to constraints associated with topography

and soil properties contributes to structure a fraction of the genome of *Eperua falcata*. These results motivate further investigations in the genetics of divergence. The analysis of the transcriptome by 454-pyrosequencing (**article n°3**) allowed the identification and the characterization of thousands of expressed genes, among which some are suspected to be polymorphic. This database provides a valuable source of candidate genes for developing a genome-scan approach integrating a large panel of genes of known function using next-generation sequencing or genotyping technologies.

The local genetic structuring in forest landscapes observed at the molecular level was also evident at the phenotypic level.

A **common garden experiment in non-limiting conditions** (**article n°4**) revealed that the native habitat of the mother trees explained a significant fraction of phenotypic differentiation for the majority of growth and leaf-traits. Even if seed mass was taken into account, we cannot completely exclude that a part of these divergences may be due to other maternal effects. It remain however probable that a part of them could be truly genetics and thus result from the action of local evolutionary processes. This study did not allow affirming that the inherent phenotypic differentiation observed here was driven by natural selection rather than by other (and neutral) evolutionary processes. There are however some indications that it may be:

(1) The plot we studied (in the experimental site of Paracou) is an environmental mosaic where a single bottomland is bordered by two slopes and two hilltops. It is poorly probable that random neutral processes would have driven such mirror-like (and thus symmetrical) phenotypic differentiation on both sides of the bottomland.

(2) The different micro-habitats are potentially connected by gene flow, as the distance between closed habitats is lower than the expected gene flow. It is thus poorly probable that the different sub-populations would have experienced different histories, considering both their demography and the genetic drift.

(3) The patterns of phenotypic differentiation were surprisingly similar within the two *Eperua* species. In particular, I have chosen a hierarchical model for estimating the effect of native habitat within the two species independently rather than testing for a global effect of native habitat. The fact that two species unrelated by gene flow display close patterns of phenotypic differentiation preferentially suggest the influence of local adaptation rather

than neutral processes. However, we cannot exclude that maternal effects not taken into account (such as seed quality) and common to both species would have contributed to model similar patterns within the two species.

Last, the growth conditions during experiment were very different to natural conditions encountered by seedlings in the wild, and we may not conclude about the maintenance of such phenotypic differentiation *in situ*.

A second experiment in which seedlings native from different micro-habitats were experimentally submitted to drought and water-logging (**article n°5**) revealed that different provenances were equally sensitive to severe constraints: their growth rate decreased when submitted to constraining conditions but resulted in similar growth rate between provenances. This result revealed the absence of differential adaptation to drought and water-logging of the different provenances, suggesting that the genetically-driven phenotypic differentiation between the provenances is not a result of local adaptation to hydric conditions or to hydric conditions alone. Because micro-habitats differ not only in hydric constraints but also in a variety of abiotic (soil and litter chemical composition, light) and biotic factors (community composition, predation), we can suppose that other environmental factors not tested here (or a combination of them) may have caused the phenotypic differentiation between sub-populations observed in non-limiting conditions. Moreover, this experiment do not allow to completely exclude the influence of hydric constraints in driving local adaptation, as the constraints exerted may have been too severe, and we lack information about the reaction norm of the different provenance to each constraint. Indeed, we observed two extreme states corresponding to non-limiting conditions and extreme drought and water-logging after six months of stress. However, without information about the evolution of growth at intermediate levels of constraints, we cannot properly conclude that seedlings from different native habitats are

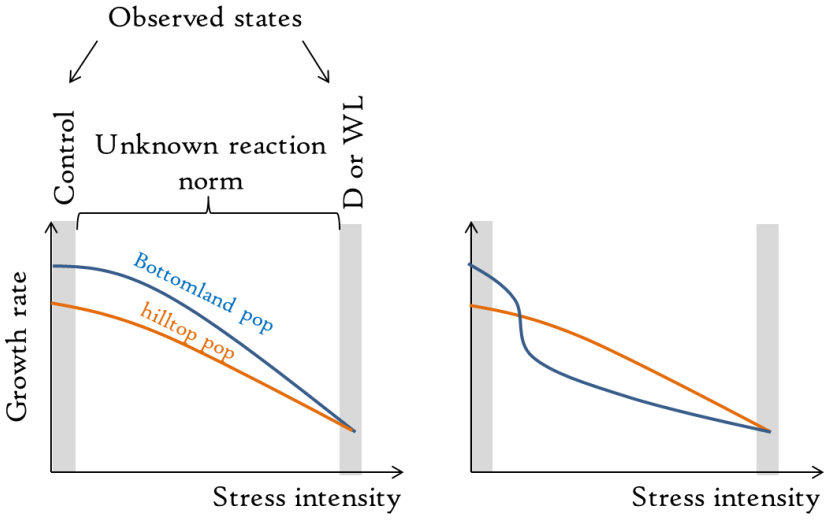


Figure 90: Possible reaction norms leading to the observed states.

equally sensitive to hydric constraints, **figure 90.**

Reciprocal transplants will helps testing the hypothesis of local adaptation in natural conditions (**see 'Preliminary results'**). Up to now, the experiment is however too young to be discussed as the seedlings grow very slowly and remain very small two years after sowing. A significant, albeit small, effect of the local provenance was detected for some traits (mainly survival), but any gene-by-environment interactions was detected at early developmental stages. Moreover, any differences between local and foreigners were detected in each transplant conditions as the seedlings remain very small and the effects not clear. This experiment will however be followed until 2015, and the span of the recorded traits will be extended to many other growth (total biomass accumulation and allocation to leaves stems and roots) and leaf traits (chlorophyll content, carbon and nitrogen content, leaf thickness).

Taken together, these results suggest that the genetic diversity of *Eperua falcata* is structured at very local scale (in the order of several hundreds of meters), mainly by neutral processes. However, local habitat patchiness and associated divergent selective pressures may contribute to enhance the genetic structuring over very short spatial scales.

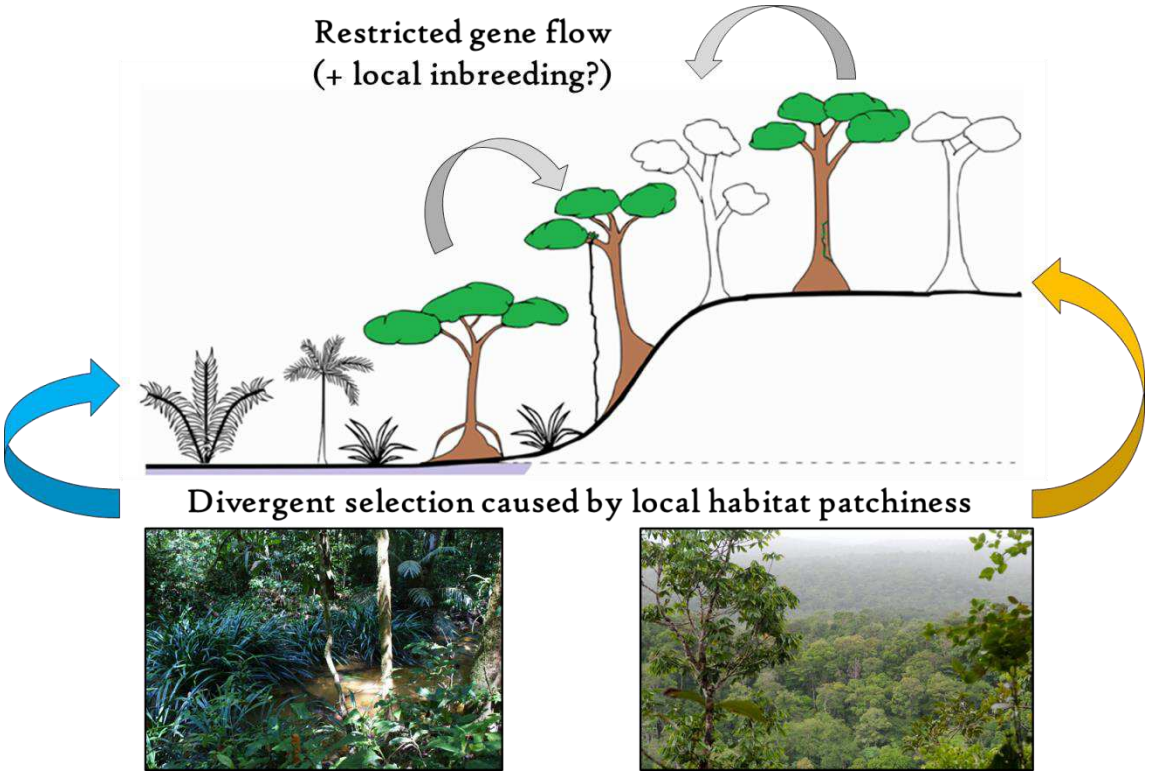


Figure 91: The local structuring of genetic diversity in *Eperua falcata*.

2. Open questions and perspectives

This study helped understanding how evolution operates in Neotropical tree populations over short spatial scales, in the order of several hundred of meters. Several questions remain however widely opened:

(1) How restricted gene flow and local inbreeding govern the spatial dynamics of *Eperua falcata* populations (demography) and the structuring of genetic diversity over short spatial scales?

(2) Have the different demes (sub-populations inhabiting different micro-habitats) and populations (inhabiting the different study sites) experienced the same demographic history or not?

(3) What is the real extent of local adaptation in the whole genome of *Eperua falcata*?

(4) Do different populations adapt to the same agent of selection (i.e. selective environmental factors) in a same way or local adaptation involves different traits and genes depending on the study site?

(5) Is local adaptation responsible for the phenotypic differentiation observed in seedling in non-limiting conditions?

(5a) What is the genetic architecture of the structured traits?

(5b) Do these genes show footprints of natural selection or not?

(6) Is there a phenotypic differentiation also visible on adult trees? (And how to cope with populations composed by trees of different ages with different ontogenic histories?)

All these questions motivate further investigations, on both adult tree populations and the recruited seedlings.

Some of these questions will be completed through two years of post-doctorate. As the reciprocal transplants will be followed until 2015, the experiment will be coupled with advanced genetic investigations based on both adult trees and transplanted seedlings. In particular, I plan to use NGS technologies to re-sequence hundreds of genes (or genotype SNPs within) chosen among the unigenes described from 454 analysis, (**article n°3**). This work will be useful for:

- Searching the pedigrees of the transplant seedlings and search for father and mother trees among the whole populations of adults.
- Estimating the realized gene flow from one generation (the adult trees that fructified

in 2011) to another (the seedlings recruited in 2011) in Laussat and Regina populations.

- Measuring the extent of inbreeding in the generation of seedlings.
- Developing an association genetics approach that will aim at identifying the genes controlling the phenotypic traits studied.

In parallel to the reciprocal transplant experiment, I used the HiSeq Illumina technology to sequence the whole-genome of 40 adult trees inhabiting Laussat and Regina study sites (10 trees within each micro-habitat and site). The assembly of the very short reads obtained will be facilitated by the already available transcriptome (either by mapping the short-reads on the assembled transcriptome of by realizing a hybrid assembly combining both long reads from 454-pyrosequencing and short reads from Illumina technology). This experiment will help me to estimate the genetic differentiation between micro-habitats and the real extent of divergent selections in the whole genome of *Eperua falcata*. As the assembled unigenes will be blasted and annotated, it will lead to more precise conclusions about the identity of the genes targeted by natural selection in the two populations of Laussat and Regina.

3. Importance of assessing genetic diversity in a changing world

Contemporary changes and predictions

As it is the case in numerous regions of the world, the Amazon basin is going to experience strong climatic changes. Since the mid-1970s, all tropical forests regions have experienced a warming at a mean rate of 0.26±0.05°C per decade, **figure 92** (Malhi & Wright 2004). In the particular case of French Guiana, temperatures show a similar trend, with an increase of about 0.25°C per decade (Wagner 2011). Since 1970s, precipitation appears to have declined in tropical forest regions at a rate of 1.0±0.8% per decade without, however, a significant trend in Amazonia. However, Li *et al.* (2008) revealed a significant increase in the frequency of dry events for the period 1970-1999, while Arias *et al.* (2011) have reported a decrease in cloudiness with an increase in solar irradiance in Amazonia for the period 1987-2007. Moreover, the majority of models from the IPCC scenarios expect a significant decrease in precipitation in the Amazon region, and a drier climate for the 21th century (Johns *et al.* 2003, Burke *et al.* 2006).

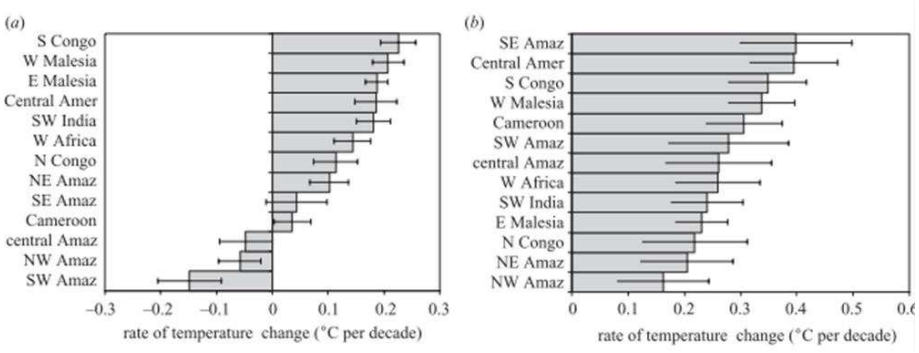


Figure 92: Rate of temperature change (°C per decade) in each tropical forest subregions (a) for the period 1960-1998; (b) for the period 1976-1998. (from Malhi & Wright 2004)

Understanding the limits of populations persistence

Climate-based species distribution models predict a redistribution of tree species in the world during the next century (Aitken *et al.* 2008). These models postulate that environmental conditions are the primary determinant of species distributions: species future range distributions are predicted by projecting the present ecological niches of the species on maps representing future climate scenarios. However, these models are often unrealistic as they take into account neither the potential of species and their populations to evolve, nor their true dispersal abilities with an evolutionary point of view.

Thus, understanding if and how species would be able to overcome rapid climate changes require understanding the limits of population persistence (Chevin *et al.* 2010, Hoffman & Sgro 2011).

Environmental changes threaten populations' persistence, because they affect

populations' size, leading to strong genetic drift. Combined together, the reduction of population size reinforced by genetic drift erodes the genetic diversity of populations ('bottleneck effect'), leaving them more vulnerable to changes. Two mechanisms may allow populations to avoid extinction: **migration** and **adaptation** (Aitken *et al.* 2008, Chevin *et al.* 2010). Thus predicting the ability of populations to overcome environmental changes requires assessing the ability of populations to adapt locally to new conditions, or to migrate toward other favorable areas.

Adaptation to new conditions would primarily depend on the strength of selection exerted by both abiotic (climate change) and biotic factors (such as inter-specific competition with species that recently colonized the area). Secondly, adaptation to new conditions would depend on both (1) the available genetic diversity for climate-related functional traits in the population and (2) the ratio between the rate of environmental change and the rate of adaptation that include both the fecundity and the generation time of the species considered. Tree populations commonly display large genetic variations for functional traits as they display large heritability values. Moreover, the genetic diversity is often spatially structured by local evolutionary processes (such as **local adaptation**) that contribute to maintain high levels of genetic diversity in forest landscapes (Dirzo & Raven 2003, Kawecki & Ebert 2004). However, directional selection is subsequently supposed to affect the genetic diversity for such climate-related genes, leaving populations more vulnerable to future changes (Jump & Penuelas 2005). Last, inter-specific hybridization, which is common in tropical rainforests, is also of major importance, as such hybridizations may produce new genotypes with higher fitness than parental species (Hufford *et al.* 2003, Aitken *et al.* 2008).

In this study, the glasshouse experiments revealed that the majority of growth and leaf traits vary significantly between seedlings progenies, leading to high maternal family-to-total variance ratios (σ^2_M/σ^2_P) and probably high heritability values in the congeneric species *Eperua falcata* and *E. grandiflora*. Moreover, both neutral and adaptive processes contribute to structure the genetic diversity over short spatial scales in Amazonia, and thus to maintain high levels of genetic diversity in large areas. These results suggest the existence of an extensive genetic variability for phenotypic traits in widespread natural populations of *Eperua* that would be beneficial for future populations' adaptation to contemporary climate change.

Migration toward a new area (i.e. the establishment of a new population in a new area in the case of plant species) requires that the distance to another favorable habitat would be lower than seed dispersal. It depends both on the species properties to realize long-distance dispersal and on the fragmentation of the habitat: the establishment of a new population in a new area requires that a favorable environment of sufficient area would be available in the limits of the populations' dispersal abilities. Moreover, the colonization of a new area requires a sufficient number of founders (Hufford *et al.* 2003), because a new population composed by too few founders (and thus with a poor genetic diversity) would be submitted to strong genetic drift and thus may not establish durably.

Analysing the fine-scale genetic structuring in *Eperua falcata* revealed that gene flow is highly restricted (less than hundreds of meters), that is lower previous estimates of gene flow in the congeneric *E. grandiflora*. Such limitations of gene flow would be critical for this species to colonize new areas in the current context of rapid climate changes. However, this study did not investigate the process of rare long-distance gene flow in *E. falcata*.

Last, **phenotypic plasticity** may also be of major importance, as it may contribute to buffer populations against extinction. According to Pigliucci (Pigliucci 2001), phenotypic plasticity may be viewed as a proximate cause of developmental change. The ability of a species to develop phenotypic plasticity (i.e. the ability of some individuals composing the population to overcome environmental changes by altering either their physiology or their morphology) may prevent local extinction of the populations, at least at short term before adaptation. Moreover, once small founder populations are established in a new area, the phenotypic plasticity of founders may allow for the persistence of the newly established populations (Aitken *et al.* 2008).

Submitting *E. falcata* seedlings to soil water content constraints revealed that growth was significantly affected by both drought and water-logging by comparison with non-limiting conditions. Moreover, soil constraints induced changes in leaf traits suggesting that *E. falcata* is able to develop plastic response to environmental constraints, among which soil water depletion. In addition to previous studies revealing that *E. falcata* is well tolerant to drought (Bonal *et al.* 2011), this study revealed its ability to develop phenotypic plasticity that may contribute to prevent local extinctions in the context of global warming in Amazonia.

Even if studying each species of the Amazonia remain impossible, more research efforts -at least on the more abundant tree species- should help expecting the capacity of neotropical trees to deal with rapid climate change. In particular, studying the process of evolution in tropical forest landscape allows:

- ➔ estimating the ability of species to migrate toward new habitats and their migration rate,
- ➔ estimating the extent of phenotypic variations for traits involved in plant response to stresses (mainly drought, **figure 93**) and assessing the extent of genetic diversity for the underlying climate-related genes.

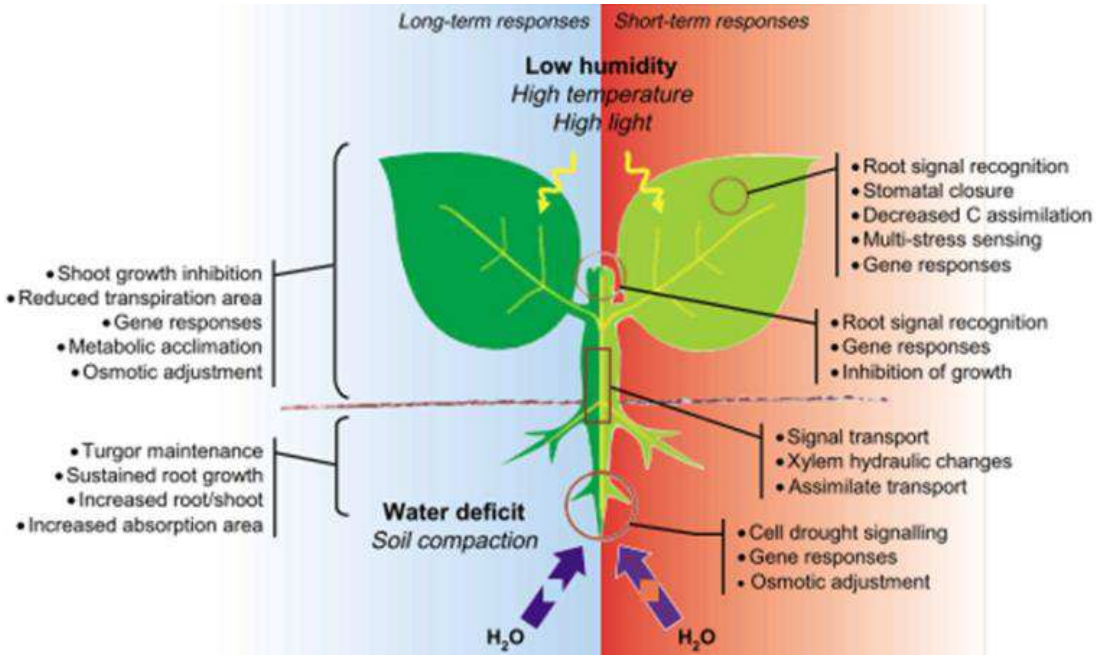


Figure 93: Mechanisms of plant response to drought stress. (from Chaves *et al.* 2003)

PhD RESULTS & TOOLS

This section contains the complete results of this PhD. They are organized in three parts: molecular evolution, genomics and phenotypic evolution. PhD results are formatted as research articles.

Molecular evolution

Article n°1

Molecular divergence in tropical tree populations occupying environmental mosaics.

D. Audigeos, L. Brousseau, S. Traissac, C. Scotti-Saintagne & I. Scotti.

(Published in *Journal of Evolutionary biology*, 2013)

Article n°2

Fine-scale genetic structure and local adaptation in a neotropical tree species of Amazonia (*Eperua falcata*, Fabaceae).

L. Brousseau, M. Foll & I. Scotti

(*in prep.*)

Genomics

Bioinformatic tools

'Rngs': A suite of R functions to easily deal with next-generation (454-) sequencing data and post-process assembly and annotation results.

L. Brousseau, C. Scotti-Saintagne

(will be compiled into an R package)

Article n°3

High-throughput transcriptome sequencing and polymorphism discovery in four Neotropical tree species.

L. Brousseau, A. Tinaut, C. Duret, T. Lang, P. Garnier-Géré & I. Scotti

(Submitted to *BMC Genomics*)

Phenotypic evolution

Article n°4

Highly local environmental variability promotes intra-population divergence of quantitative traits: an example from tropical rainforest trees.

L. Brousseau, D. Bonal, J. Cigna, I. Scotti.

(Published in *Annals of Botany*, 2013)

Article n°5

Local adaptation in tropical rainforest trees: response of *Eperua falcata* (Fabaceae) seedling populations from contrasted habitats to drought and to water-logging.

L. Brousseau, I. Scotti, E. Dreyer, D. Bonal

(*in prep.*)

Preliminary results

Reciprocal transplants

Article n°1 - Molecular divergence in tropical tree populations occupying environmental mosaics

Delphine AUDIGEOS*, Louise BROUSSEAU*, Stéphane TRAISSAC†, Caroline SCOTTI-SAITAGNE* & Ivan SCOTTI*

* INRA UMR 0745 EcoFoG ('Ecologie des forêts de Guyane'), Campus Agronomique, Kourou, French Guiana, (France)

† AgroParisTech-ENGREF UMR 0745 EcoFoG ('Ecologie des forêts de Guyane'), Campus Agronomique, Kourou, French Guiana, France

Published in: Journal of Evolutionary Biology (2013) **26**: p 529-544 (doi: 10.1111/jeb.12069).

Abstract

Unveiling the genetic basis of local adaptation to environmental variation is a major goal in molecular ecology. In rugged landscapes characterized by environmental mosaics, living populations and communities can experience steep ecological gradients over very short geographical distances. In lowland tropical forests, interspecific divergence in edaphic specialization (for seasonally flooded bottomlands and seasonally dry terra firme soils) has been proven by ecological studies on adaptive traits. Some species are nevertheless capable of covering the entire span of the gradient; intraspecific variation for adaptation to contrasting conditions may explain the distribution of such ecological generalists. We investigated whether local divergence happens at small spatial scales in two stands of *Eperua falcata* (Fabaceae), a widespread tree species of the Guiana Shield. We investigated Single Nucleotide Polymorphisms (SNP) and sequence divergence as well as spatial genetic structure (SGS) at four genes putatively involved in stress response and three genes with unknown function. Significant genetic differentiation was observed among sub-populations within stands, and eight SNP loci showed patterns compatible with disruptive selection. SGS analysis showed genetic turnover along the gradients at three loci, and at least one haplotype was found to be in repulsion with one habitat. Taken together, these results suggest genetic differentiation at small spatial scale in spite of gene flow. We hypothesize that heterogeneous environments may cause molecular divergence, possibly associated to local adaptation in *E. falcata*.

Introduction

Environmental gradients – the more or less continuous spatial variations of biotic and abiotic conditions – and environmental patchiness produce spatially variable selective pressure on biological populations, inducing their genetic diversification through local adaptation (Antonovics 1971, Linhart & Grant 1996, Kawecki &Ebert 2004, Savolainen *et al.* 2004, Fine *et al.* 2005, Hedrick 2006, Namroud *et al.* 2008). Correlation between environmental variables and frequencies of adaptive genetic variants has been repeatedly observed, and such patterns have generally been interpreted as signatures of selection forcing genetic pools to adjust to local environment (Storz & Kelly 2008, e.g. Hedrick 2006). The observation of adaptive genetic divergence between populations occupying different parts of an environmental gradient is therefore suggestive of the action of disruptive selection in favor of local adaptation.

The study of how genetic diversity is coupled with environmental gradients rests on solid theory, and stems from the rather intuitive idea that genetic turn-over can be quantified through changes in allele frequencies, and that if a gradient influences allele frequencies, then an association should be present between the two (Epperson 2003). Although with considerable refinement, this approach is the base of all studies of ecological-genetic gradients (Bergmann 1978, Ingvarsson *et al.* 2005, Joost *et al.* 2007, Ingvarsson 2008, Eckert *et al.* 2009, Coop *et al.* 2010, Eckert *et al.* 2010, Fournier-Level *et al.* 2011, Montesinos-Navarro *et al.* 2011, Hancock *et al.* 2011, Chen *et al.* 2012), including those of populations inhabiting contrasting habitats.

Conventionally, the effect of environmental gradients has been sought at scales that go from regional to continental (Achere *et al.* 2005, Tsumura *et al.* 2007, Eveno *et al.* 2008, Namroud *et al.* 2008), implicitly assuming that at shorter scales migration will systematically overwhelm selection. Nevertheless, there are reasons to think that disruptive selection acts even at very local scales. Even in the absence of selection gradients, genetic relatedness tends to be spatially structured in plant populations (and particularly in trees) because of preferential dispersal in the close neighborhood. Limitations to dispersal can therefore reinforce differential, spatially structured disruptive selection. Conversely, moderate levels of gene flow may increase the rate of adaptation, by enabling the emergence of novel multilocus genotypes and by exposing alleles to multiple environments, thus facilitating the action of selective filters (Goudet *et al.* 2009, Kremer & Le Corre, 2012). Finally, most plant populations produce a large excess of seeds and seedlings each season,

which should set the stage for very strong selection, even if it is partially confounded by random processes. The existence of local adaptation in spite of gene flow has been reported at the very short spatial scale in animals (Storz, 2005), inartificial plots for outcrossing wind-pollinated annual plants (Freeland *et al.* 2010) but also on larger scales for wind-pollinated trees (Savolainen *et al.* 2007, Eveno *et al.* 2008, Eckert *et al.* 2009, Eckert *et al.* 2010). Jump & Penuelas (2005) have reviewed proofs that intra-population genetic variation for traits and genes related to response to climatic gradients exists in plant species. Their analysis rests on a long tradition of studies on local adaptation to patchy or continuously varying environments, of which clear examples are found in annual plants at both the landscape (Angert & Schemske, 2005, Manel *et al.* 2010, Poncet *et al.* 2010) and within-population scale (Schmitt & Gamble 1990). For instance, local adaptation has been identified at the molecular level for tree species within a range of less than 3 km (Jump *et al.* 2006), and parapatric or sympatric speciation for palms has likely occurred on a single 12 km^2 island (Savolainen *et al.* 2006, Babik *et al.* 2009). Thus, even for long-lived organisms, such as trees and palms, it is possible to observe genetic divergence at a very local scale, in spite of the (real or expected) presence of recurrent gene flow among environmental patches or portions of the gradient. It is therefore legitimate to ask whether locally variable selection contributes to the diversification of sub-populations and to the build-up and maintenance of genetic diversity and adaptive potential in tree species.

With the development of genomic methods, several strategies for testing the association of Expressed Sequence Tags (EST), Single Nucleotide Polymorphisms (SNPs) or anonymous markers with traits and/or eco-logical preferences (association mapping; population genomics) have been introduced (Luikart *et al.* 2003, Neale & Savolainen 2004, Gonzalez-Martinez *et al.* 2006, Eckert *et al.* 2009). These methods usually require a priori information that may not be easily accessible for nonmodel taxa (Luikart *et al.* 2003), while enabling gene-level selection studies without prior knowledge about the relationship of phenotype to genotype or the precise function of candidate loci (Storz 2005, Vasemagi & Primmer 2005). Higher (or lower)-than-expected levels of divergence among populations at a given locus is then taken as suggestive of disruptive (or stabilizing) selection (Beaumont & Nichols 1996, Luikart *et al.* 2003, Beaumont & Balding 2004, Storz 2005, Gonzalez-Martinez *et al.* 2006, Riebler *et al.* 2008). This strategy can be applied at the genome level, when extensive genomic information is available, or to sets of candidate genes (Phillips 2005, Wright & Gaut 2005) when a particular ecological and physiological process is

targeted.

When environmental factors are spatially structured, for example in the case of habitat patches or gradients, the study of Spatial Genetic Structure (SGS) can also help testing the association of genotypes and environ-mental conditions. SGS can result from a variety of processes, including spatially structured selection and limited dispersal (Condit *et al.* 1996, Clark *et al.* 1998, Plotkin *et al.* 2000). It is therefore necessary to distin-guish the relative role of the different evolutionary forces (Heywood 1991; Manel *et al.* 2003, Vekemans & Hardy 2004). In structured environments, the distribution of genotypes relative to habitat gradients can be compared with the overall distribution of genotypes (or to a null distribution). Specifically, at loci under diver-gent selection, it is expected that turnover of alleles is steeper along the gradient than in any other direction (and between ecologically contrasted zones than between randomly drawn zones; Oden & Sokal, 1986). Landscapes with abrupt habitat changes occurring over short spatial scales and with an alternation of eco-logically divergent habitat patches provide a suitable opportunity for the study of the strength of selective forces leading to local adaptation. Seasonally flooded lowland forests of the Guiana shield occur in a rugged landscape characterized by small creeks alternating with small hills, where edaphic (i.e. related to soil characteristics) conditions can vary steeply from bottomlands to the top of hills and hillocks, resulting in environmental mosaics (Baraloto & Couteron 2010). Therefore, forest tree populations growing in this region provide the opportunity to test the occurrence of local adaptation phenomena. Habitat specialization has been repeatedly observed in tropical trees (Plotkin *et al.* 2000, Harms *et al.* 2001, Lopez & Kursar 2003, Palmiotto *et al.* 2004, John *et al.* 2007). Several studies have tested responses to edaphic constraints in trees from the Guiana shield (Baraloto *et al.* 2005, 2006, 2007) and analyzed the interspecific variability of traits related to edaphic stress response (Bonal *et al.* 2000, Bonal & Guehl 2001, Coste *et al.* 2005, Bonal *et al.* 2007; Scotti *et al.* 2010), but the presence of intraspecific local adaptation in species occupying several habitats (and its possible genetic base) have never been tested. The present work focuses on populations of *E. falcata*, a common tree species of the Fabaceae family growing in relatively dense clusters of up to several hundreds of trees and densities of up to 40 stems (diameter at breast height > 10 cm) per hectare. Seed dispersal is barochore and pollination is mostly performed by bats. *E. falcata* was found to be significantly positively associated with flooded forest (Collinet 1997, Baraloto *et al.* 2005, 2007). However, distribution maps show that it can occur on a large spectrum of edaphic conditions, up to hilltops, thus

showing a somewhat generalist behavior and therefore potential for local adaptation.

The present study focuses on three main edaphic habitats occupied by this species: bottomlands, seasonally flooded by heavy rainfall during the rainy season; slopes, with thin soil and highly variable soil water content and terra firme plateaus, with deep, well-drained soil possibly prone to drought during the dry season (Wright 1992).We have analyzed genetic diversity in a set of seven genes of which four have a known function related to response to hypoxic stress (Catalase), drought stress (Farnesyltransferase) and plant water balance (two aquaporins; Audigeos *et al.* 2010) and three were randomly drawn from an EST library obtained from seedlings from one of these papers' study areas. We assessed SGS and performed multilocus scans for genetic differentiation at small spatial scale (~ 6 ha), in two forest plots presenting environmental patchiness. We tested local differentiation of *E. falcata* populations as a function of variation of edaphic conditions and found loci potentially undergoing disruptive selection.

Materials and methods

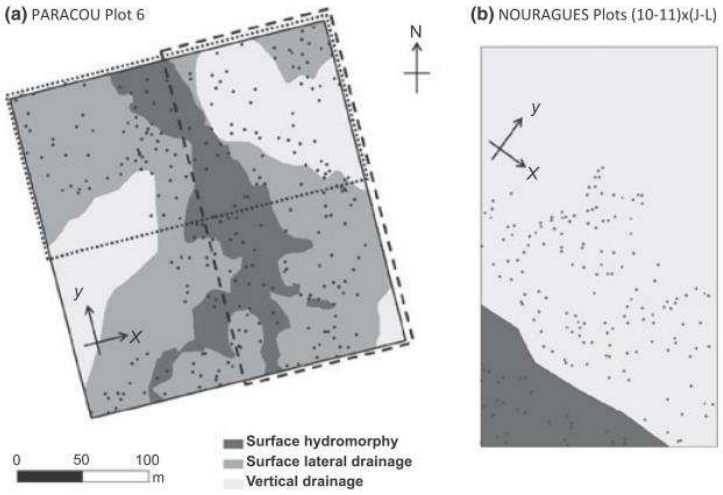


Figure 1 *E. falcata* cartography on drainage map for each sampling site: Paracou Plot 6 (A) and Nouragues Plots 10J, 10K, 10L, 11K, 11J, 11L (B). Light grey represents VD, medium grey represents lateral drainage (SLD) and dark grey represents surface hydromorphy (SH). The axes used in each plot for directional autocorrelation analyses are shown. Dashed line: Paracou 'Eastern Half'. Dotted line: Paracou 'Northern half'.

Sampling and DNA isolation

Trees were sampled from two forest inventory plots in French Guiana, one with an environmental mosaic (Paracou, **Fig. 1a**) and one with two homogeneous patches of strongly contrasted environments (Nouragues, **Fig. 1b**). The Paracou experimental station, 50 km from Kourou (5 ° 15'N, 52 ° 55'W), is formed by 15 plots of 6.25 ha (Gourlet-Fleury *et al.* 2004). The Nouragues research station (4 ° 05'N, 52 ° 41'W) has an area of over 100 ha, subdivided in 1-ha square plots. Both sites represent relatively accessible but undisturbed forest areas. At both sites, the study area was subdivided based on a discrete categorization (Ferry *et al.* 2010) of environmental (and particularly edaphic) conditions (**Fig. 1**): vertical drainage (VD) corresponding to terra firme forests, with deep soil rarely undergoing drought and never undergoing flooding; surface lateral drainage (SLD), which could experience drought during the dry season due to its very variable soil water-content and surface hydromorphy (SH) corresponding to seasonally flooded forest with soil saturated with water to the surface, undergoing hypoxic stress during the rainy season. This environmental partition loosely corresponds also to topography classes like plateau, slope and bottomland respectively. In Paracou, Plot 6 (of an area of 6.25 ha) was chosen. The three environmental conditions described above are represented in this plot (**Fig. 1a**). In Nouragues, six contiguous 1-ha plots crossed by a river (10J, 11J, 10K, 11K, 10L, 11L), for a total area of 6 ha, approximately equivalent to Paracou's Plot 6. Only the VD and SH ecological conditions were found at the Nouragues study site (**Fig. 1b**). The majority of Nouragues individuals are found in VD zones and a few in SH zones. Cambium was collected from 440 *E. falcata* trees with a diameter larger than 10 cm at breast height: 258 in Paracou (48, 172

and 38 in the SH, SLD and VD zones respectively) and 182 in Nouragues (29 and 153 in the SH and VD zones respectively) and DNA was extracted following a CTAB method (Doyle & Doyle 1987, Colpaert *et al.* 2005).

Amplicon choice, sequencing and polymorphism detection

To study evolutionary processes responsible of genetic diversity in *E. falcata*, we analyzed sequences with a putative role in the response to stresses related to edaphic conditions and sequences with unknown function. Of the seven loci used in the study, two were aquaporin gene fragments, PIP1.1 and PIP1.2 characterized in a previous study (Audigeos et al., 2010), whereas the other nuclear fragments were obtained by sequencing clones from both cDNA and genomic libraries. The five additional sequences included: a fragment of the gene coding for Catalase (CAT), involved in the response to oxidative stress caused by flooding; a fragment of the gene coding for Farnesyltransferase (FTase), involved in the abscissic acid (ABA) metabolic pathway; a DNA fragment coding for a hypothetical protein (HYP5) and two ESTs with unknown function (UNK7 and UNK14). The chosen loci represent a mix of candidates for a putative role in the response to edaphic constraints (Cat, FTase), genes with a housekeeping function in plant water balance (PIPs) and 'randomly drawn' gene functions (HYP5, UNK7, UNK14).Although we cannot assume that any of these sequences are 'neutral' in the general sense, we have no special reason to think that they undergo disruptive selection in this particular habitat gradient, with the notable exception of CAT and FTase. Therefore, we consider the sampled gene panel as representative of the general behavior of the transcriptome with respect to this particular gradient. Moreover, we decided to make use only of EST sequences because other kinds of regions, such as anonymous genomic sequences or microsatellites, may have different molecular properties (e.g. different substitution rates, nucleotide composition and linkage disequilibrium levels for the former, different mutation model and rates for the latter), making the data set inhomogeneous from the evolutionary point of view. By restraining our analysis to only one kind of sequence type, we have tried to avoid any bias in data interpretation that may arise from comparisons between data sets with different underlying structures. The libraries were obtained using the Lambda ZAP II kit (Stratagene, La Jolla, CA, USA) following the manufacturer's protocol. About 200 clones were sequenced and their putative function assigned based on their comparison with public databases using BLASTn and BLASTx. A set of 47 sequences with length between 300 and 600base pairs

(bp) was selected, including: proteins with known function, retrotransposons, hypothetical proteins or sequences without match in public databases. Primers were designed using PRIMER - BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/) and OLIGO - CALC (http://www.basic.northwestern.edu/biotools/oligocalc.html). Preliminary tests for amplification were conducted on two individuals. Fifteen fragments produced specific ampli-cons; PCR and sequencing on 16 individuals were performed to evaluate sequence quality and polymorphism level. Five chosen fragments plus the two aquaporin genes were then sequenced in all samples. Haplotypes have been deposited in GenBank under accession numbers JQ801740 – JQ801745 (**Table 1**).

Table 1 Description of the fragments and their amplification conditions. Locus code: short name used throughout the text to indicate the locus; Accession numbers: EMBL/GenBank accession numbers; BlastX: closest BlastX match for each sequence in the EMBL/GenBank data base; Function: function of the closest BlastX match; Putative role in stress: role in the response to environmental stress, if known; Primers: primer pair for the amplification of each fragment; Annealing temperature: temperature used for annealing in PCRs (see Methods for details).

Locus code	Accession numbers	Protein prediction	BlastX	Function	Putative role in stress	Primers (5'→3')	Annealing temperature
CAT	JQ801740* JQ801745†	Catalase	AAR84578	Convert hydrogen peroxide in water and oxygen	Oxidative stress	F : TCCAGCTTCCTGTCAATGC R : ACAACGCACATGGCACAC	64 °C
FTase	JQ801744	Putative farnesyltransferase alpha subunit	XP002534116	Add a farnesyl group to the –SH of the cysteine	ABA stress signalling pathway	F : GCCCACCCTGAGAATGAAAG R : TGCCTGAACCTGAAAACAAG	55→52 °C(D)
HYP5	JQ801741	Hypothetical protein	EEF45947	Unknown	Unknown	F : AATGCAATGGACCTTGAGC R : TTCATGAAACGTGATCAACC	55→52 °C‡
PIP1.1	FJ807642	Putative aquaporin PIP1	ABD63904	Plasma membrane water channel	Water-balance	F : CCCAGCAGTGACCTTCG R : AACCAAGAACACAGCGAACC	64→57 °C‡
PIP1.2	FJ807646	Putative aquaporin PIP1	ABR68794	Plasma membrane water channel	Water-balance	F : CAACCCGGCTGTGACC R : GCCAAATGGACCAAGAACAC	64→57 °C‡
UNK7	JQ801742	Unknown	–	Unknown	Unknown	F : GACCGGAACAGTAATTCGTTG R : ATTTCGCTAAAAAGGCCTGC	64→61 °C‡
UNK14	JQ801743	Unknown	–	Unknown	Unknown	F : GTATTGGGGGTATTCTCCGC R : GCTGCCACTTCATGTGACC	64→61 °C‡

*Partial cds (coding sequence).
†Partial genomic sequence.
‡Touchdown PCR: annealing temperature decreases from the highest to the lowest temperature over the first seven cycles (see Methods for details).

PCRs were carried out in a 15 l μL volume containing15 ng of DNA, 1x Taq buffer, 2 mM MgCl 2 , 0.25 mM of each dNTP, 0.3 U Taq polymerase (all products from New England Biolabs) and 0.5 μM of each primer. An initial denaturation at 94 ° C for 10 min was followed by 35 cycles of (45 s at 94°C; 20 s at the annealing temperature shown in **Table 1**; 1 min 30s at 72 °C) and a final extension at 72 °C. PCR products were purified with EXOSAP-IT (USB Corporation). Sequencing reactions were performed with the BigDye®Terminatorv3.1 cycle sequencing kit (Applied Biosystems) in a total volume of 10 μL containing 0.5 μL of Big Dye,1.5 μL of Buffer, 1 μL of 2 μM primer, 4 μL of cleaned-up PCR product and 3 μL of milli-Q water. All fragments were sequenced in both directions. Sequencing reactions were then purified by ethanol purification and sequence data were obtained on an ABI 3130xlcapillary sequencer (Applied Biosystems). Base calling and contig assembly were done using CODONCODE ALIGNER V 2.0.1 (Codoncode Corporation, Dedham, MA, USA). All polymorphisms were visually checked. As DNA sample were diploid, the identification of haplotypes (i.e. sequence variants) for individuals with more than one SNP was performed using PHASE (Stephens *et al.* 2001, Stephens & Donnelly 2003) implemented in DNASP V 5 (Librado & Rozas,2009) to produce two haploid sequences per individual.

Data analyses

We performed our analyses at the 'site' level (Paracou and Nouragues) and at the 'habitat' level (VD, SLD and SH) within a site. We use the terms throughout the article: 'amplicon' to refer to sequenced PCR fragments; 'haplotype' for the different amplicon sequence variants and 'SNP' for each polymorphic site (including indels).

- *Molecular diversity and differentiation*

Nucleotide diversity of each amplicon was estimated by both θs (Watterson 1975), based on the number of segregating sites and θπ (Nei, 1987), based on the average number of pairwise nucleotide differences between sequences in a sample. Haplotype diversity Hd (Nei 1987) was also calculated for each amplicon. Analyses of diversity were conducted in DNASP V 5 (Librado & Rozas 2009). Linkage disequilibrium among amplicons was estimated only for haplotypes occurring with > 5%frequency, using a likelihood ratio test (Slatkin & Excoffier 1996) as implemented in ARLEQUIN V 3.5 (Excoffier & Lischer, 2010). Linkage disequilibrium (LD) within amplicons and departure from Hardy-Weinberg equilibrium were tested on a contingency table of observed vs. predicted genotype frequencies using a modified Markov-chain random walk algorithm as described by Guo & Thompson (1992) and implemented in ARLEQUIN. LD was tested with 5% significance before and after applying the sequential Bonferroni correction for multiple testing. We also computed the LD descriptive statistic r^2 (Hill & Robertson 1968), as it summarizes both recombination and mutation history and it is less sensitive to sample size than other common LD statistics, such as D' (Flint-Garcia *et al.* 2003). r^2 was calculated on SNPs using DNASP V5 and statistical significance of r^2 was computed with a one-tailed Fisher's exact test and applying Bonferroni corrections for multiple testing. The decay of LD with physical distance was estimated using nonlinear regression of LD between SNPs (r^2) onto their distance in base pairs (Remington *et al.* 2001; Ingvarsson 2005). The expected value of r^2 under drift-recombination equilibrium, taking mutation into account, was computed according to Hill & Weir (1988). The genetic structure of populations was investigated by the analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) implemented in ARLEQUIN V 3.5. AMOVA was estimated among sites and among environments within site, for each amplicon as Nst (Pons & Petit 1996) and for all amplicons as Fst (Weir & Cockerham 1984).

- *Detection of 'outlier' loci*

Departures from the standard neutral model of molecular evolution were investigated by two different methods: the frequentist method described by Beaumont & Nichols (1996) and the more refined Bayesian method described in Beaumont & Balding (2004). To compare the results obtained with the two methods, we assigned confidence levels of 99% and 90% for FDIST 2 and BAYESFST. The use of these two significance thresholds confers comparable false discovery rates to the two methods (Beaumont & Balding 2004). Identification of polymorphisms carrying a possible signature of natural selection ('outlier' loci) was first performed with the FDIST2 program, which uses the summary-statistics approach described in Beaumont & Nichols (1996) and further developed in Beaumont & Balding (2004).Twenty-thousand coalescent simulations were performed with three and two populations of 50 individuals for Paracou and Nouragues respectively. Because sample size was unequal between sub-populations at each site, and because only one sample size can be entered as a parameter in FDIST 2, we also ran the analyses with three populations, sample size 170 and two populations, sample size 150, for Paracou and Nouragues respectively; this corresponds to the largest sample size for each site. The numbers of populations and samples to simulate were chosen to model as closely as possible the populations that have been analyzed at each site. Expected F_{st} for simulations was determined as the mean of observed F_{st} values. To comply with the assumption of independence of loci required for the estimation of population diversity and divergence, three independent subsets of 21 SNPs (three per amplicon) with zero pairwise LD were used to compute F_{st} 's. This led to three independent simulations, each of which is based on 21 statistically independent loci; as these are statistically uncorrelated, we consider them as being effectively independent loci, although they come from a restricted number of physical genome locations. The neutral envelop was constructed for each simulation at the 99% confidence level. A single envelop was obtained by selecting, in each diversity bin computed by the algorithm, the most conservative F_{st} values (i.e. the largest upper bound and the smallest lower bound). Loci with a F_{st} value exceeding the upper limit of the neutral envelop conditional on heterozygosity were considered as potentially under divergent selection. The Bayesian inference method implemented in the BAYESFST program (Beaumont & Balding 2004) was also used to identify genes under selection. This algorithm relies on a Bayesian model to identify locus-specific population divergence between samples, by implementing a Metropolis-Hastings Markov Chain Monte Carlo (MCMC) process based

on the likelihood of allele counts. It has the advantage of disentangling locus effect (α_i), population effect (β_j) and optional interaction between locus and population effects (γ_{ij}). A positive value of α_i indicates the presence of disruptive selection at the locus, whereas a negative value suggests balancing selection. The γ_{ij}'s also have an interpretation in terms of selection: a large positive γ_{ij} could indicate a potentially advantageous mutation that would be locally adapted in a particular population (Beaumont & Balding 2004). Default prior distributions were used to generate 10 000 parameter series and convergence was checked using the CODA package of R version 2.10.1. Outlier values for α_i and γ_{ij} were identified setting the confidence level at 90%.

- *Spatial analyses*

We tested whether the distribution of genotypes was likely to have arisen by chance, given the spatial structure of stems and habitats, using a method adapted from Harms *et al.* (2001). We compared the relative abundance of each haplotype in each habitat to its expectation under the null hypothesis of random distribution of haplotypes. The null distribution of each haplotype's relative abundance was simulated by 10 0000torus-translations of stem locations to conserve their spatial pattern. The limits of the neutral confidence interval were defined as values excluding 5% of the highest and lowest values. If the relative frequency of a genotype, determined from the true habitat map, wa soutside the confidence interval, then it was considered to be statistically associated with the habitat (if the frequency had a positive value) or dissociated from the habitat (negative value). Habitat association of haplotypes and genotypes for each amplicon and of SNPs was tested for each site. Spatial genetic structure was assessed at plot scale using directional spatial autocorrelation analyses (Epperson 2003) of the pairwise kinship coefficient between individuals (f_{ij}; Loiselle *et al.* 1995), which was computed for haplotypes and individual SNPs. Calculations were performed by SPAGEDI (Hardy & Vekemans 2002). Kinship coefficient values were computed for a set of nine 20 m-wide distance intervals (from 0to 180 m) and the significance of the slope of f_{ij} as a function of geographical distance was tested based on the permutation procedure implemented in SPAGEDI with10 000 permutations. Significance of negative slopes (indicating that genetic similarity decreases with geo-graphical distance) was tested at one-tailed a = 5%with Bonferroni correction for multiple tests. Directional autocorrelation was performed by taking into account all and only the pairs of points connected by a segment aligned in the desired direction, with a tolerance of *pi*/12 radians on each side. The matrix of distances for suitable pairs of points

was computed using an R script written for this purpose and available from the Authors. Autocorrelation was performed: (a) for Paracou: along the (orthogonal) X and the Y axes indicated in **Fig. 1a**, and omni directionally, for the whole plot as well as for its Northern its Eastern halves (**Fig. 1a**; these sub-plots were included in the analyses because eye inspection of the landscape revealed that they contained a habitat gradient along one of the axes); (b) in Nouragues, along the X and Y axes indicated in **Fig. 1b** and omnidirectionally. The Y axis corresponds to the presumed direction of the environ-mental gradient for all cases except the Northern half of the Paracou plot, where the presumed cline direction is the X axis (for the whole plot in Paracou, despite the ruggedness of the pattern, the proportion of points sampled in the VD condition steadily increases along the Y axis; **Fig. 1a**). SGS was conservatively considered as anisotropic (i.e. strength of autocorrelation varied between directions) when the slope of f_{ij} values with distance was significant in one, but not the other, of the mono-directional tests, although some degree of autocorrelation is expected to occur in all directions due to neutral processes, such as limitations to seed and pollen dispersal.

Results

DNA polymorphism

Sequence polymorphism data obtained from the seven EST loci for the two experimental sites are shown in **Table 2.** The total number of polymorphic sites per amplicon ranged from 6 for HYP5 to 31 for FTase and the number of haplotypes per amplicon ranged from 12for PIP1.2 to 36 for FTase.

Table 2 DNA polymorphism for seven amplicons across the two sampling sites. *N*: number of analysed haploid sequences, L: sequence length, S_T: number of polymorphic sites, h: number of haplotypes; H_d: Nei's gene diversity computed on haplotype frequencies; θ_π, θ_W: estimates of population-level diversity based on the average number of pairwise differences per site between sequences and on the number of segregating sites respectively.

Amplicon	*N*	*L*	S_T	h	H_d	θ_π(x10^3)	θ_W (x10^3)
CAT	309	657	8	15	0.567	1.53	1.93
FTase	578	397	31	36	0.637	3.12	11.26
HYP5	642	361	6	10	0.618	2.12	2.36
PIP1.1	456	552	17	23	0.796	4.45	4.60
PIP1.2	664	525	7	12	0.394	1.11	1.88
UNK7	482	417	8	13	0.645	3.18	2.84
UNK14	622	412	11	24	0.502	2.30	3.83

The average nucleotide diversity θ_π across polymorphic fragments was 0.00254 and varied from 0.00111 for PIP1.2 to 0.00445 for PIP1.1.The average of θ_s is higher (0.0041) than h p; values ranged from 0.00188 for PIP1.2 to 0.01126 for FTase, which had the highest number of SNPs and haplotypes. In this amplicon, the great majority of SNPs are non-synonymous (18), one was triallelic and one was a heterozygous singleton coding for a termination codon that shortens the protein sequence of the last 10 amino acids.

Linkage disequilibrium

We did not find any clear evidence of tight linkage dis-equilibrium among amplicons. Three loci (CAT, PIP1.1and PIP1.2) showed significant linkage disequilibrium after Bonferroni correction: CAT and PIP1.1 were associated in Nouragues; PIP1.1 and PIP1.2 in Paracou. As none was in LD in both Nouragues and Paracou, all loci were considered as independent. Between 2% (FTase) and 40% (HYP5) of within-amplicon linkage disequilibrium tests between SNP loci were significant at P<0.01. Decay of linkage disequilibrium within amplicons was rapid (**Fig. 2**). Nonlinear fitting of the squared correlation of allele frequencies r^2 as a function of distance between SNPs showed expected values of ~ 0.10 at100 bp (determination coefficient for the

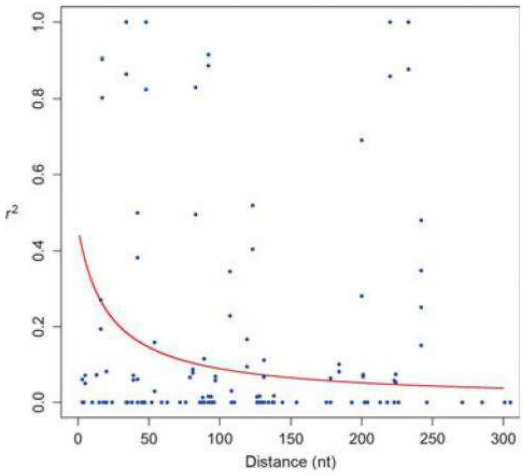


Figure 2 Plot showing the squared correlation of allele frequency (r^2) as a function of physical distance between sites for seven amplicons in *E. falcata*. A nonlinear fitting was performed using Equation 1 (Remington *et al.*, 2001).

fitted modelR2 = 3.5%). It has to be noted that the locus with the highest density of SNPs (FTase) also shows the lowest level of linkage disequilibrium, thus excluding the possibility that high levels of polymorphism are due to co-amplification of two different loci (which would cause strong linkage disequilibrium between variants belonging to the two loci).

Population structure

Population differentiation analyses are presented in **Table 3**. The average level of genetic differentiation between sites was very low (F_{st} = 0.010), but significantly different from zero ($P<10^{-5}$). N_{st} values at the amplicon level ranged from -0.01 to 0.03, with three amplicons having a value significantly different from zero: CAT, FTase and UNK14. The level of global genetic differentiation among habitats within Paracou was quite similar, with an overall F_{st} significant value of 0.01. N_{st} values at the amplicon level ranged from 0.00to 0.04, with two amplicons showing significant differentiation: CAT and FTase. The situation is similar for pairwise comparisons between environments, with significant divergence for CAT in three cases and for FTase in two cases. In Nouragues, the mean level of genetic differentiation was null. N_{st} values varied between -0.10 and 0.10 among amplicons, with two amplicons (FTase andUNK14) showing significant positive values.

Table 3 Results of the analysis of molecular variance (AMOVA). Genetic differentiation (*F*-statistics) at the haplotype level for each amplicon (N_{ST}) and at the multilocus level (F_{ST}) for different hierarchical levels (between sites and among and between environments).

		Paracou vs. nouragues	Paracou				Nouragues
			Global	Pairwise VD vs. VLD	VD vs. SH	VLD vs. SH	Pairwise VD vs. SH
Locus	Statistic	*F*-statistic	*F*-statistic	*F*-statistic	*F*-statistic	*F*-statistic	*F*-statistic
CAT	N_{ST}	0.03*	0.04*	0.06*	0.06*	0.02*	−0.12
FTase	N_{ST}	0.02*	0.02*	0.04*	0.03*	0.00	0.10*
HYP5	N_{ST}	0.00	0.00	0.01	0.01	0.00	0.01
PIP1.1	N_{ST}	0.00	0.00	0.01	0.01	0.00	0.00
PIP1.2	N_{ST}	0.00	0.01	0.01	0.00	0.00	−0.01
UNK7	N_{ST}	−0.01	0.00	−0.01	0.00	0.00	−0.02
UNK14	N_{ST}	0.01*	0.00	0.01	−0.01	0.00	0.05*
All loci	F_{ST}	0.01*	0.01*	0.01	0.02*	0.01	0.00

Significant values (α = 5%) are indicated by an asterisk.

Outlier detection

The summary-statistic simulation method implemented in FDIST 2 identified two SNPs of 74 and six of 60 as outliers showing footprints of disruptive selection at the99% confidence level in Paracou and Nouragues respectively (**Fig. 3**). The outliers found for the Paracou site belong to two amplicons (CAT and UNK7). Outlier detection by pairs of habitats in Paracou (**Supplementary Fig. S1**) shows that the results obtained in the global

analysis are mainly due to divergence between VD and SH. The six outliers detected in Nouragues belong to three amplicons (CAT, FTase and UNK14). One SNP (CAT_S355) was a significant outlier at both sites (outlier detection based on simulations with larger samples sizes provided much more liberal results; **Supplementary Fig. S2**). The more robust Bayesian method, implemented in BAYESFST, provided different results at a comparable 90% confidence level: no SNP was significantly different from neutral expectations. However, the SNPs detected as significant by the coalescent-based method showed the highest α_i values with the Bayesian method.

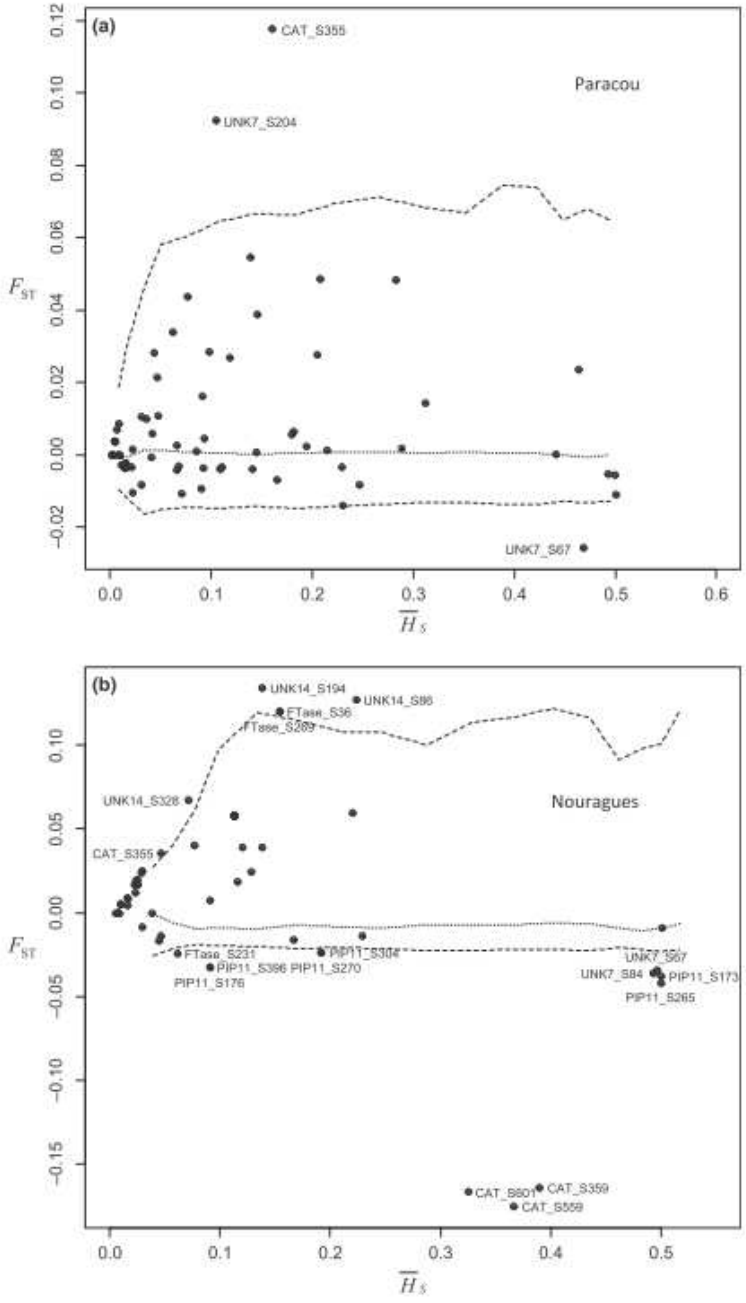


Figure 3 Distribution of observed F_{ST} values for each locus as a function of its average within-population heterozygosity ($\overline{H}_S$). The simulated median (dotted line) and 99% neutral envelop confidence limits (dashed lines), obtained by coalescent simulation are shown. The names of loci lying outside the neutral envelop are displayed. (a) Paracou (b) Nouragues. Note that the scale on the *y* axis differs between the two plots.

Spatial genetic structure

Torus translation tests detected 20 significant independent habitat associations (not counting for tightly linked SNP loci): nine in Nouragues, 11 in Paracou, of which 13 for haplotypes and seven for SNPs, of 454 tests (4%) at two-tailed α = 5%; one remained significant after Bonferroni correction (**Table 4** and **Suppl. Table 1**).

Table 4 Torus-translation tests for habitat associations on the two study plots.

	Habitat association						
Paracou	VD+	VD−	SLD+	SLD−	SH+	SH−	Total
Haplotypes [115]	5	0	1	1*	1	0	8 (1*)
SNPs [149]	1	1	0	1	0	0	3
Nouragues							
Haplotypes [75]	2	3	NA	NA	0	0	5
SNPs [115]	3	1	NA	NA	0	0	4
Total	11	5	1	2 (1*)	1	0	20 (1*)

Six associations with SLD and with VD, as well as one with SH, were detected (**Suppl. Table 1**), along with six cases of repulsion with SLD and one with VD (no case of repulsion with SH was detected). The most frequent haplotype at the FTase locus (h1) showed strong association with SH and repulsion with SLD in Paracou, and was associated with VD in Nouragues, together with two SNPs of the same gene. In Paracou, one PIP1.1 haplotype (h15) was associated with SLD and PIP1.2's most common haplotype (h1), as well as one PIP1.2 SNP (S145), were associated with VD. The repulsion between FTase haplotype H1 and SLD in Paracou was the only significant test left after Bonferroni correction. Directional and omnidirectional autocorrelation was tested for each individual SNP, and for all amplicons at the haplotype level, at the two sites. After Bonferroni correction, 26 autocorrelograms (of a total of 894, or3%), involving six SNPs and four amplicons, showed a significant negative slope at the α=5% threshold (**Suppl. Table 2**). In eight cases (**Fig. 4**, **Suppl. Table 2**), there was significant autocorrelation along the direction of the gradient (Y axis for all tests except for Paracou, Northern Half), but not for the direction orthogonal to the gradient. Two of these tests involved amplicon UNK14 in Nouragues, for one SNP (UNK14_194) and for the whole amplicon; four involved amplicon HYP5 in Paracou, two for one SNP (HYP5_160) and two for the whole amplicon; two involved amplicon CAT in Paracou, one for a SNP (CAT_299) and one for the whole amplicon.

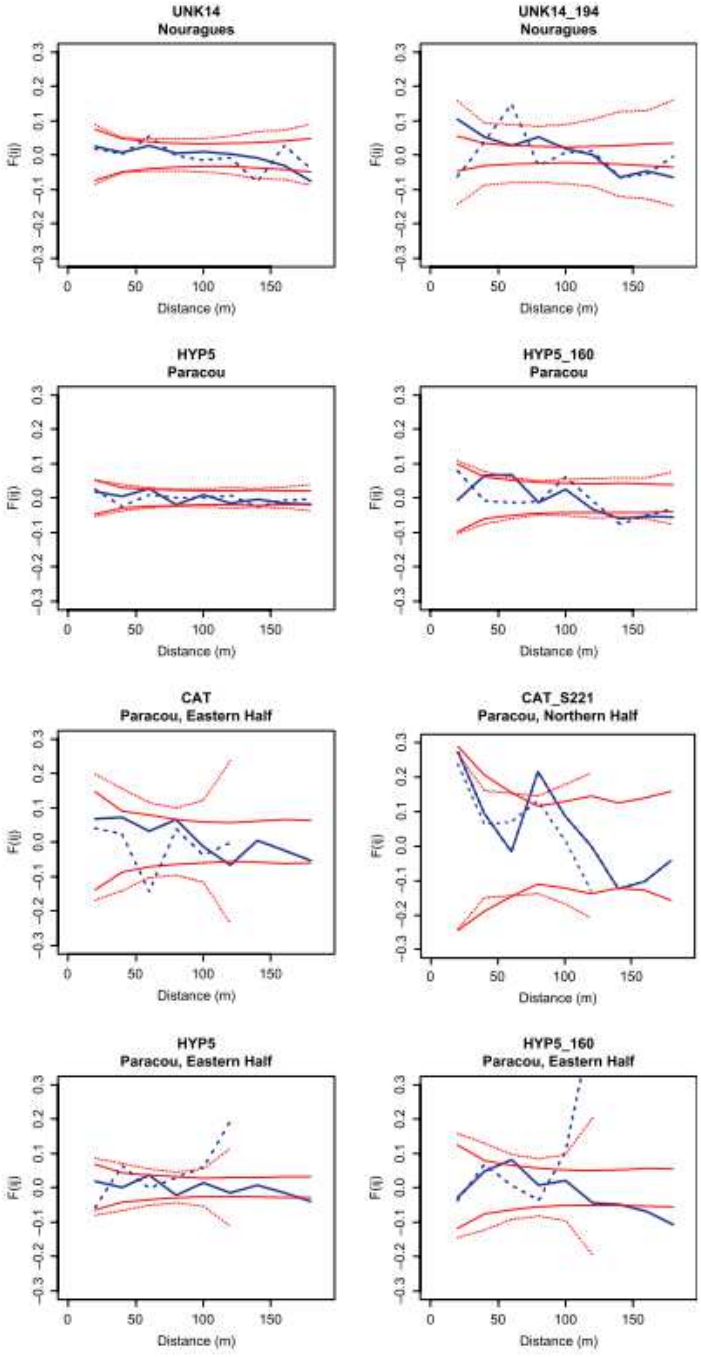


Figure 4 Directional autocorrelograms of estimated kinship coefficient f_{ij} (Loiselle *et al.*, 1995) for all tests were significant in the direction of the gradient and nonsignificant in the orthogonal direction. Thick lines: observed values. Thin lines: upper and lower 95% neutral confidence limits. Solid lines: plot of (significant) autocorrelation values obtained along the expected gradient (Y axis for all plots except for Paracou, Northern half, for which the gradient is along the X axis). Dotted lines: plot of (nonsignificant) autocorrelation values obtained along the direction orthogonal to the gradient. Left panes: haplotype-level autocorrelograms. Right panes: SNP-level autocorrelograms.

Discussion

The results presented here show patterns of genetic differentiation associated with micro-geographical habitat variations at fine spatial scale in populations of *E. falcata*. These results were obtained with four independent methods and suggest that divergent selection may be strong even between sub-populations belonging to a continuous population. The results obtained for each amplicon and each kind of analysis are summarized in **Table 5**. Analysis of molecular variance (**Table 3**) shows that the Catalase and Farnesyltransferase genes can reach very high levels of sub-population divergence within a plot (several SNPs displayed strongly negative F_{st} values in Nouragues).

Table 5 Summary of significant results for all amplicons, all statistics and all sites. Amplicon/Haplotype/SNP level indicates whether the analysis was performed, respectively, for all haplotypes of an amplicon, for each of its haplotypes or for each of its SNPs. For amplicon-level tests, only the names of the populations having shown significant tests are listed. In toroid tests, 'ass' and 'rep' indicate whether a given haplotype or SNP variant was found to be in association (ass) or in repulsion (rep) with a given habitat.

	AMOVA	Outlier detection	Toroidal permutation tests		Directional spatial autocorrelation	
Amplicon	Amplicon level	SNP level	Haplotype level	SNP level	Amplicon level	SNP level
Catalase (CAT)	Paracou	Paracou: CAT_S355 Nouragues: CAT_S355	–	–	Paracou Eastern half	Paracou Northern half: CAT_S221
Farnesyltransferase (FTase)	Paracou Nouragues	Nouragues: FTase_S36, FTase_S269	Paracou: h1/SH (ass), h1/SLD (rep)* Nouragues: h1/VD (ass)	Paracou: FTase_S242/SLD (rep) Nouragues: FTase_S36/VD (ass), FTase_S269/VD (ass)	–	–
PIP1.1	–	–	Paracou: h15/VLD (ass)			
PIP1.2	–	–	Paracou: h1/VD (ass)	Paracou: PIP1.2_S145/VD (ass)		
Hypothetical Protein 5 (HYP5)	–	–	Paracou: h10/VD (ass) Nouragues: h2/VD (ass), h8/VD (rep), h11/VD (rep)	Paracou: HYP5_S267/VD (rep) Nouragues: HYP5_S160-S201/VD (ass)	Paracou Paracou Eastern half	Paracou: HYP5_S160 Paracou Eastern half: HYP5_S160
Unknown amplicon 7 (UNK7)	–	Paracou: UNK7_S204	Paracou: h1/VD (ass)	–	–	–
Unknown amplicon 14 (UNK14)	Nouragues	Nouragues: UNK14_S86, UNK14_S194, UNK14_S328	Paracou: h4/VD (ass), h7/VD (ass) Nouragues: h9/VD (rep)	Nouragues: UNK14_S86/VD (rep)	Nouragues	Nouragues: UNK14_S194

SNPs that appear as significant in at least two independent analyses are shown in bold. The asterisk (*) indicates the only toroid test that remained significant after Bonferroni correction.

Because the SH population is much smaller than the VD, both demographically and in surface, this may imply that the lower portion of the VD population is more similar to SH, at neutral loci, than it is similar to the upper part of VD, due to neutral SGS. This would have the consequence of generating negative F_{st}, i.e. closer relatedness between alleles from different populations than from the same population). Coalescent-based outlier detection methods revealed eight SNPs under disruptive selection belonging to four amplicons (although none was significant with the more conservative Bayesian approach). Directional autocorrelation identified three SNPs (and the amplicons they belong to) as significantly associated to the expected direction of the gradient (although none was associated to the orthogonal direction). Finally, allele (or genotype)-by-habitat association

tests obtained by torus permutation identified 20significant tests at the two-tailed 5% threshold; one of these remained significant after Bonferroni correction. **Table 5** shows that at least six SNP variants, haplotypes or amplicons turned out to be significant in at least two independent analyses. The Catalase amplicon showed significant results (mostly in Paracou) both at the amplicon and at the SNP variant levels; one SNP of the Catalase amplicon (CAT_S355) was a significant outlier in both populations. Farnesyltransferase displayed significant results at all levels and in both plots, with two SNP variants showing significant results in outlier detection and in torus permutations. The latter analysis, both at the haplotype and at the SNP level, indicates that generally the detected variants are less represented than expected in drought-prone SLD habitats: all significant tests show either association with VD or SH, or repulsion with SLD; the only results that remains significant after Bonferroni correction is the repulsion of haplotype 1 and SLD in Paracou. HYP5 also shows a SNP variant with clear trends of habitat association, as well as the UNK14 amplicon. These results suggest that forces behind the differentiation between sub-populations are very strong even at short spatial distances, and that these forces are structured by variation in habitat rather than by neutral dispersal processes. The processes underlying the observed divergence occur over distances in the order of few hundreds of meters – well within gene dispersal distances predicted for the genus (Hardy *et al.* 2006). Therefore, it is likely that at least part of the observed differentiation is caused by disruptive selection (Linhart & Grant 1996). On one hand, our findings support the idea that environmental heterogeneity generates genetic heterogeneity within populations. On the other hand, the contrast between results observed in Nouragues and Paracou suggests that the contrasts we have studied are of different kinds. The structure of the gradients may differ between the two plots, as suggested by their differences in topography. Moreover, and more generally, it is likely that environmental conditions, other than the limited set of edaphic properties that we have taken into account, differentiate habitats in the two sites. Differences among the results obtained with the three methods suggest that each captured different aspects of the distribution of genetic diversity. For instance, both the outlier detection and torus permutation tests stress the idea of differences in gene frequencies between (sub)populations, but the latter also takes into account the spatial distribution of genotypes; moreover, autocorrelation rests on the explicit spatial layout of pairwise individual relatedness, while ignoring population-level distributions (except for the determination of neutral envelopes). Thus, the three methods may be able to detect different

patterns, which in turn may be the result of different dynamic processes: outlier detection methods stress the quantitative difference between the effects of selection and drift on divergence between groups; torus-based tests also compare groups, but stress departures from random distribution of individual variants; autocorrelation methods detect departures from the random distribution of individual relationships and tests continuous turnover of genotypes. As our analyses are based on seven loci only, a possible source of incoherence among results obtained with the three methods may also lie in limited robustness. Seven loci certainly are far from providing a satisfactory representation of the whole transcriptome, let alone of the genome. Even without the ambition to evaluate genome-level processes, our study nevertheless proves that genetic divergence can be detected at the within-stand level, at least for some loci. Moreover, the robust-ness of each of the three methods used here resides (i) in the number of SNPs (not ESTs) for outlier detection, (ii) in the number of genotypes per locus (not in the number of loci, which are analyzed individually) for torus-translation and (iii) autocorrelation analyses. For the latter analysis, it is not uncommon to obtain results from data sets containing between five and 10loci (Collevatti & Hay 2011; Oddou-Muratorio *et al.* 2010).The partial incoherence shown by the results suggests a pattern of moderate divergence affecting multiple loci, occurring at the micro-geographical scale in relation with habitat conditions. It is important to underline that the diffuse signal of divergence that we observe must not be interpreted as straightforward indication of disruptive selection acting upon the observed loci. Other mechanisms, such as isolation by adaptation (Nosil *et al.* 2008), genomic hitch-hiking (Via & West 2008) or partial restrictions to mating (e.g. by environ-mentally cued flowering time differences) may produce moderate levels of divergence at neutral loci. Such divergence is observed against a background of overall weak but diffuse SGS patterns (**Suppl. Table 2**: omnidirectional autocorrelation is significant for 'all loci' in three cases of four), probably caused by limited pollen and seed dispersal, as already observed at Paracou in a closely related species (Hardy *et al.* 2006). Population structuring is not, however, strong enough to prevent long-term genetic mixing, as shown by the rate of decay of intragenic linkage disequilibrium. The pattern shown in **Fig. 2** indicates that historical genetic mixing at the population (and species) level is globally as intense as in other angiosperm tree species (e.g. Ingvarsson 2008) and at least as intense as in most conifers (Brown *et al.* 2004; Gonzalez-Martinez *et al.* 2006, Heuertz *et al.* 2006). Current mixing at the stand level appears to be relatively intense, because only a minority of loci showed significant spatial

autocorrelation and because F_{st} values between sub-populations were overall small. This is not inconsistent with the possibility that the observed divergence is caused by selection, because moderate levels of gene flow may facilitate divergence, as indicated by theoretical predictions of divergence with gene flow (Goudet *et al.* 2009, Kremer & Le Corre, 2012). Moreover, it is actually possible that the weak but detectable back-ground spatial genetic structure contributes to create the conditions for divergence to operate: preferential mating between spatially close trees would tend to enrich sub-populations with locally adapted genotypes, thus enhancing the outcome of ecological filtering and facilitating sub-population divergence. This hypothesis can be put to test by building spatially explicit, individual-based models describing simultaneously pollination, seed dispersal and selection in divergent habitat patches (e.g. by building combinations of models for dispersal)and selection in continuous environmental patches (Debarre & Gandon 2010) and for pollen and seed dispersal (Klein & Oddou-Muratorio 2011). The indication of the action of diversifying processes, observed in *E. falcata*, motivates further studies in the genetics of divergence, that will need to take advantage of population genomic approaches (Luikart *et al.* 2003) now accessible to nonmodel species in general (Ekblom & Galindo 2011) and to trees in particular (Gonzalez-Martinez *et al.* 2006). To take advantage of the wealth of data that can be produced by genomic approaches, these studies will need to be matched by breakthroughs in the modeling of processes of divergence with gene flow. The combination of theoretical advances and large data sets will permit to disclose the mechanisms underlying patterns of ecological-genetic divergence such as those demonstrated in *E. falcata*, and perhaps ultimately provide the key to the understanding of the maintenance of reservoirs of adaptive variation in natural populations.

Authorship

DA, CSS, IS, contributed to experimental conception and setup; DA, ST, CSS, IS, contributed to samplings trategy choice and sampling; DA, LB contributed to marker development and sequence data collection; DA, LB, ST, CSS, IS, contributed to data analyses. All authors contributed to the writing of the manuscript.

Acknowledgments

The authors gratefully acknowledge the CIRAD field assistants for their contribution to data collection (M. Baisie, A. Etienne, K. Ficadici, F. Kwasie, K. Martinus, O. N'Gwete, P. Naisso, P. Pétronelli and R. Santé) and A. Jolivot for her work in data checking and data management. The authors wish to thank Saint-Omer Cazal for DNA extractions and Valérie Troispoux for her helpful technical assistance. Part of the experiments presented in the present publication was performed at the Genotyping and Sequencing facility of Bordeaux (grants from the Conseil Régional d'Aquitaine no 20030304002FA and 20040305003FA and from the European Union, FEDER no 2003227). This work was funded by the EU-funded INCO 'SEEDSOURCE' project, by the French Ministry of Ecology and Sustainable Development 'ECOFOR –ECOSYSTEMES TROPICAUX' program and by the EU-funded PO-FEDER 'ENERGIRAVI' program. The Authors also wish to thank Olivier Hardy and Pau-line Garnier-Géré for discussions on the methods and results reported here and for critically reading earlier versions of the manuscript, and Lino Ometto as well as two anonymous reviewers for carefully and critically reading the submitted manuscript.

Article n°1 – Supplementary text

Description of genes and polymorphisms for genes with known functions

Catalase is related to detoxification during oxidative stress (Mittler 2002; Willekens et al. 1997) which may be induced by hypoxia or anoxia during the rainy season in SH environment (Blokhina et al. 2003) or by drought during the dry season in VD environment (Moran et al. 1994). Farnesyltransferase is related to the ABA signaling pathway (Cutler et al. 1996; Pei et al. 1998) which could be induced by several biotic and abiotic stresses including drought (Raghavendra et al. 2010; Wilkinson& Davies 2002). PIP 1.1 and 1.2 are members of the multi-gene aquaporin family, involved in transmembrane water transport, and belong to the "plasmamembrane" (P) subfamily, located in cell membranes. The molecular effects of SNPs with significant habitat structure are the following: CAT_s221 and CAT_s355 are located in intronic regions; FTase_s36 is a nonsynonymous mutation, replacing a polar and positively charged amino acid (Arginin) by a polar and uncharged amino acid (Glutamine); FTase_s242 is a nonsynonymous tri-allellic SNP, the most frequent amino acid is Aspartic acid (polar and negatively charged) replaced by Asparagin (polar and uncharged) or Histidin (polar and positively charged); FTase_s269 is a synonymous mutation; PIP1.2_s145 is a non-synonymous mutation replacing a Tyrosine with a Cysteine (both polar aminoacids).

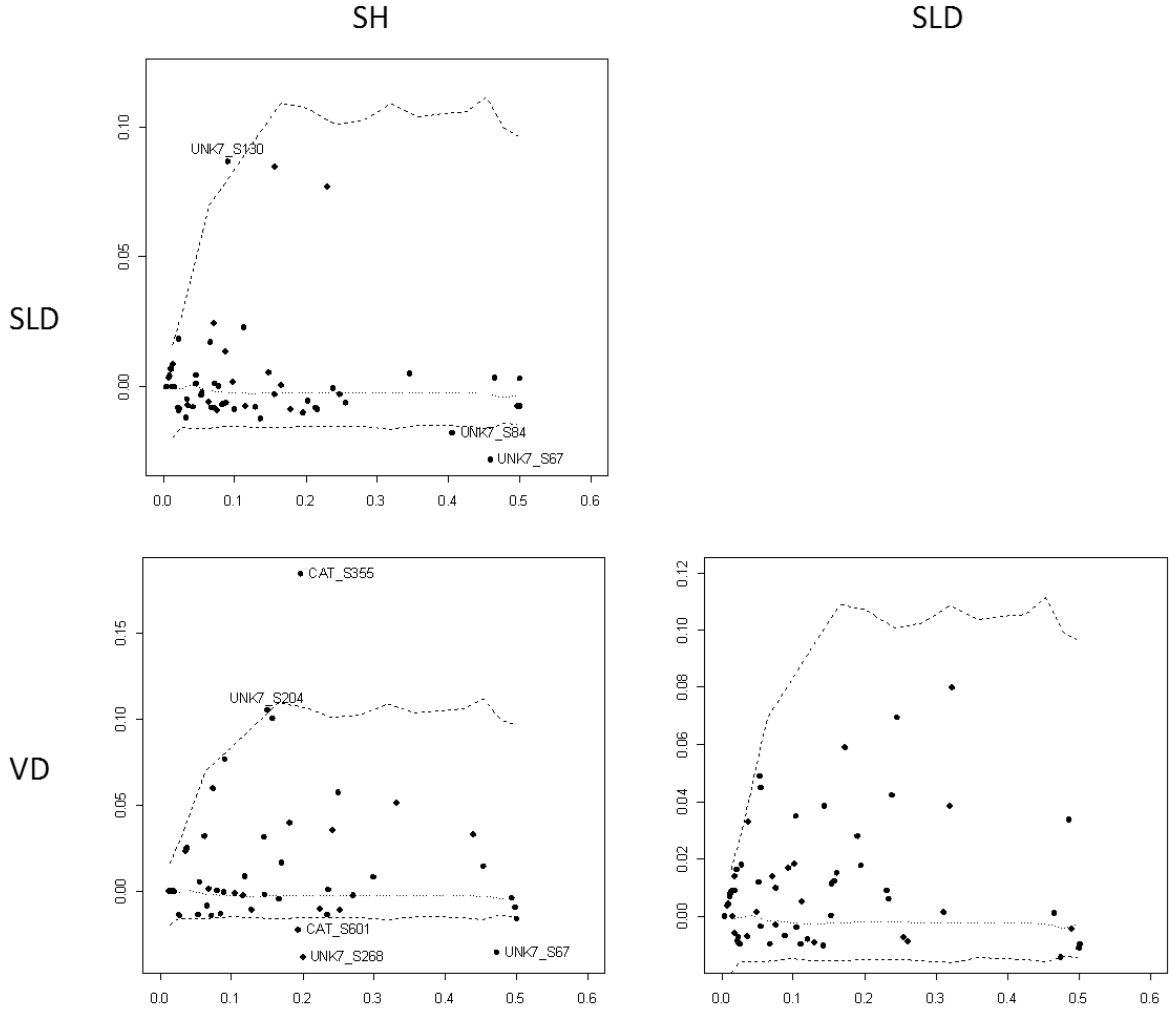


Figure S1: Distribution of observed F_{ST} values for each locus (y axis) as a function of its heterozygosity ($\overline{H}_S$) (x axis) for pairs of habitats at the Paracou site (see Figure 1). The simulated median (dotted line) and 99% neutral envelop confidence limits (dashed lines), obtained by coalescent simulation, are shown. The names of loci lying outside the neutral envelop are displayed. SH: surface hydromorphy; SLD: surface lateral drainage; VD: vertical drainage. Note that the scale on the y axis differs between plots.

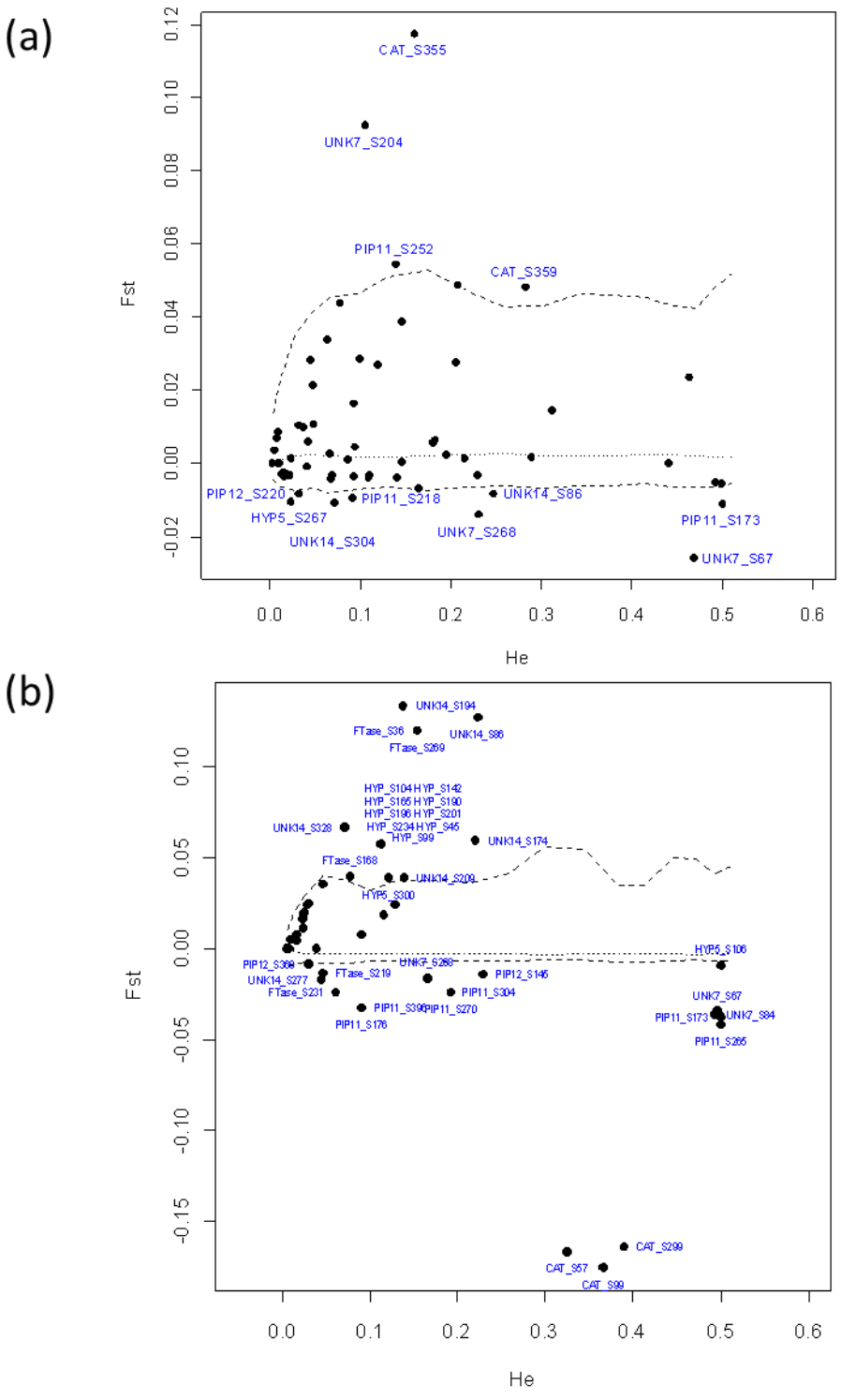


Figure S2: Outlier detection tests (a) for Paracou with sample size N = 170 and (b) for Nouragues with sample size N = 150

Article n°1 - Supplementary tables

Table S1 : Results of toroidal permutation tests: list of individual SNPs and haplotypes having shown at least one significant habitat association. Type: type of locus / variant being tested: H = single haplotypes within locus; S = SNP loci (the association of most frequent variant to habitats is shown). '+' = association and '-' = repulsion (two-tailed α = 5%). Tests significant after Bonferroni correction are marked by '*'.

		Plot				
		Paracou			Nouragues	
Type	haplotype/locus	VD	SLD	SH	VD	SH
H	Ftase_h1	ns	-*	+	+	ns
S	Ftase_S36	ns	ns	ns	+	ns
S	Ftase_S242	ns	-	ns	ns	ns
S	FTase_S269	ns	ns	ns	+	ns
H	HYP5_h2	ns	ns	ns	+	ns
H	HYP5_h8	ns	ns	ns	-	ns
H	HYP5_h10	+	ns	ns	ns	ns
H	HYP5_h11	ns	ns	ns	-	ns
S	HYP5_S160	ns	ns	ns	+	ns
S	HYP5_S165	ns	ns	ns	+	ns
S	HYP5_S190	ns	ns	ns	+	ns
S	HYP5_S201	ns	ns	ns	+	ns
S	HYP5_S267	-	ns	ns	ns	ns
H	PIP1.1_h15	ns	+	ns	ns	ns
H	PIP1.2_h1	+	ns	ns	ns	ns
S	PIP1.2_S145	+	ns	ns	ns	ns
H	UNK14_h4	+	ns	ns	ns	ns
H	UNK14_h7	+	ns	ns	ns	ns
H	UNK14_h9	ns	ns	ns	-	ns
S	UNK14_S86	ns	ns	ns	-	ns
H	UNK7_h1	+	ns	ns	ns	ns

Table S2: Slopes of spatial autocorrelation plots. Only results for loci whose tests provided at least one negative, significant slope are displayed. f_{ij}: interval (minimum, maximum) of kinship observed in omnidirectional tests. OMNI: omnidirectional autocorrelation; X: autocorrelation in the X direction; Y: autocorrelation in the Y direction. *d*: autocorrelation plotted against linear distance; *ln(d)*: autocorrelation plotted against logarithm of distance. The number of tests carried out for SNPs is indicated in parentheses for each plot (the total does not correspond to the total number of polymorphism because some loci were monomorphic at one or the other site and because some loci were excluded from the analyses due to missing data). Loci that show directional correlation along the Y axis, but not along the X axis, are indicated in bold.

Locus	f_{ij}	OMNI		X		Y	
		d	ln(d)	d	ln(d)	d	ln(d)
SNP LEVEL							
Nouragues (61)							
UNK14_194	**-0.065,0.103**	**-0.001**	**-0.070**	**ns**	**ns**	**-0.001**	**-0.010**
Paracou (75)							
ALL LOCI	-0.005,0.018	-0.00006	-0.007	ns	ns	ns	ns
CAT_S221	-0.007,0.124	-0.0005	-0.051	ns	ns	ns	ns
FTase_330	-0.027,0.076	-0.0005	-0.052	ns	ns	ns	ns
HYP5_160	**-0.024,0.036**	**-0.0004**	**-0.035**	**ns**	**ns**	**-0.0006**	**-0.048**
UNK14_387	-0.005,0.077	-0.0003	-0.030	ns	ns	ns	ns
Paracou, Northern half (71)							
CAT_S221	**-0.11,0.21**	**-0.0015**	**-0.111**	**-0.0018**	**-0.181**	**ns**	**ns**
FTase_S330	-0.077,0.065	-0.001	-0.074	ns	ns	ns	ns
Paracou, Eastern half (63)							
ALL LOCI	-0.020,0.029	-0.0001	-0.0145	ns	ns	ns	ns
CAT_S355	-0.071,0.202	ns	-0.0878	ns	ns	ns	ns
HYP5_160	**-0.065,0.068**	**-0.0011**	**-0.0678**	**ns**	**ns**	**-0.0011**	**-0.0819**
AMPLICON LEVEL							
Nouragues (7)							
ALL LOCI	-0.018,0.016	-0.0001	-0.008	ns	ns	ns	ns
UNK14	**-0.032,0.044**	**-0.0004**	**-0.028**	**ns**	**ns**	**-0.0006**	**-0.040**
Paracou (7)							
ALL LOCI	**-0.006,0.018**	**-0.00007**	**-0.008**	**ns**	**ns**	**-0.00008**	**-0.009**
FTase	-0.015,0.033	-0.00007	-0.011	ns	ns	ns	ns
HYP5	**-0.015,0.023**	**-0.0001**	**-0.014**	**ns**	**ns**	**-0.0001**	**-0.018**
Paracou, Northern half							
-	-	-	-	-	-	-	-
Paracou, Eastern half							
ALL LOCI	**-0.008,0.021**	**ns**	**ns**	**ns**	**ns**	**ns**	**-0.015**
CAT	**-0.054,0.073**	**ns**	**ns**	**ns**	**ns**	**-0.0005**	**-0.052**
HYP5	**-0.032,0.032**	**-0.0004**	**-0.0263**	**ns**	**ns**	**-0.0003**	**-0.027**

Article n° 2 - Genome scan reveals fine-scale genetic structure and suggests highly local adaptation in a Neotropical tree species (*Eperua falcata*, Fabaceae)

Louise BROUSSEAU[1,2], Matthieu FOLL[3], Ivan SCOTTI[1]

1. INRA, UMR 'Ecologie des Foreˆts de Guyane', Campus agronomique, BP 709, 97387 Kourou cedex, French Guiana

2. INRA, UMR 'Ecologie et Ecophysiologie Forestières', 54280 Champenoux, France

3. Computational and Molecular Population Genetics (CMPG), Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, CH-3012 Bern, Switzerland

Abstract

Populations undergoing divergent ecological constraints may diverge genetically due to the effect of directional selection. The outcome of divergence processes depends on the balance between selection, drift and gene flow. If selection is sufficiently strong, it can overcome the blurring effects of the other two forces, and population divergence can be observed at the phenotypic and molecular level. Genome scans can reveal loci under divergent selection and permit to estimate the portion of the genome involved in divergence. Although genome scan approaches are now widespread, they have never been applied to megadiverse tropical rainforests and to conditions where ecological divergence occurs at very short spatial distances ('highly local' processes, where environmental turnover occurs well within the range of gene flow).

We have applied and AFLP-based genome scan to population of the Neotropical tree, *Eperua falcata* (Fabaceae) in the Guiana Shield, where it grows in dense stands that cross the boundaries between starkly contrasting habitats such as seasonally or permanently flooded swamps and well-drained plateaus. We have found that, despite the short spatial distances and the presence of gene flow, habitat-structured subpopulations diverge at a substantial number of loci. Simulation analyses show that the observed levels of divergence are compatible with strong directional selection. Intense selective processes may therefore maintain genetic and phenotypic variability within rainforest tree populations; such adaptive diversity may constitute the fuel that feeds the great diversity harbored by these communities.

Introduction

Environmental heterogeneity influences the distribution of plant genetic diversity across habitat types. Forest trees provide numerous examples of adaptation to environmental variations at both phenotypic and molecular levels (Savolainen *et al.* 2007). Several provenance tests performed in common gardens and reciprocal transplants have revealed that tree populations undergo phenotypic divergence under the pressure of ecological gradients and contrasts (Petit & Hampe 2006). Such patterns of phenotypic divergence are often interpreted as a result of divergent selection driven by environmental heterogeneity that may be caused by biotic and abiotic factors. At the molecular level, numerous studies reported footprints of divergent selection in the genome of forest trees among habitats using both genome scans (Jump *et al.* 2006) and candidate genes approaches (Eveno *et al.* 2008, Audiegeos *et al.* 2013).

Amazonian lowland rainforests are characterized by complex habitat patchiness, whereby environmental variations occurs at a very small spatial scale. The succession of waterlogged bottomlands and well-drained *terra firme* contributes to explain the maintenance of high tree diversity in such forests. The structure of tree communities strongly differs among habitat types, with variations in tree and palm biomass (Khan 1987, Ferry *et al.* 2010), and differentiation in some phenotypic traits (Kraft *et al.* 2008). It has been suggested that divergent selective pressures among habitat types may have driven niche differentiation and specialization of trees to local conditions: significant habitat associations within species complexes are supposed to result from adaptive radiations along topography gradients (Baraloto & Couteron 2010). At the population level, a recent study has revealed footprints of divergent selection between local populations, occupying distinct habitats, for genes putatively involved in plant responses to environmental stresses (catalase, farnesyl-transferase, Audigeos *et al.* 2013) in the canopy tree species *Eperua falcata*. Genetic differentiation was accompanied by weak but consistent phenotypic divergences for growth and leaf physiology at the seedling stage (Brousseau *et al.* 2013). These results suggest that adaptive phenomena may be widespread and may affect a substantial fraction of the genome, even when divergence occurs at highly local scales, in conditions in which gene flow may easily erase the effects of weak selective forces.

The evolutionary mechanisms driving the large diversity of tropical rainforests are still poorly understood, and to date genome scan approaches have not been applied to the study of the extent of habitat-driven adaptive differentiation in any tropical rainforest

ecosystem.

Genome scans allow screening the genome to detect locus-specific signatures of population divergence, which are taken as suggestive of natural selection (Storz 2005). Because most of the genome is supposed selectively neutral (Kimura 1985), loci identified in genome-wide analyses, whose differentiation is higher than the average genome-wide estimate, can be interpreted as being under divergent selection. This conclusion must be taken with caution, however, because excess divergence does not automatically mean selection at or near the divergent locus (Nosil *et al.* 2008, Excoffier *et al.* 2009, Hermisson 2009, Bierne *et al.* 2011), but the identification of divergence outlier loci is suggestive, nevertheless, of non-neutral differentiation processes, if departures from neutral demographic and spatial patterns can be excluded (Bierne *et al.* 2013). AFLP markers (Vos & Bleeker 1995) are particularly effective in such genome-wide analysis, because they are cost-intensive and universal (Bonin *et al.* 2007, Meudt & Clarke 2007), although they are dominant and thus require prior genetic knowledge (such as inbreeding coefficient, F_{IS}) for proper interpretation.

Here, we investigated whether sub-populations growing in contrasting environments display signatures of potential divergent selection at the genome level. We did this in two populations of a canopy tree species (*Eperua falcata*) of the coastal Guiana Shield, and we used a genome scan approach involving AFLP markers to test the hypothesis. *Eperua falcata* is a canopy-dominant tree species widely distributed in French Guiana. Because of its generalist behaviour relative to environmental heterogeneity and of its high population densities, it is a good model to analyse the genetic structuring among local habitats within continuous tree populations over short geographical scales. We detected 0.3% and 1.8% outlier markers, and simulations based on the actual levels of divergence observed in the populations show that such loci may be under selection of moderate strength.

Materials and methods

Study sites and sampling

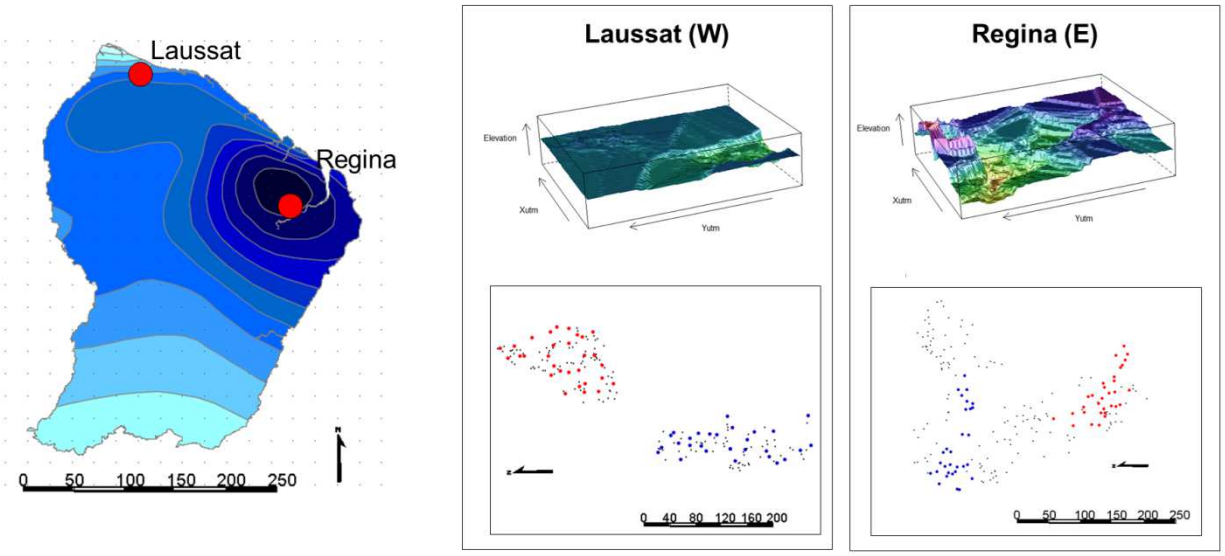


Figure 1: Geographical and topographic situation of the study sites. Coloured dots: trees sampled for genotyping.

Our study includes two populations of *Eperua falcata* established in the costal shield of French Guiana: Laussat (W) (X: 5°28'37''; Y: -53°34'36'') and Régina (E) (X: 4°18'44'' ; Y: -52°14'6''), (**Fig. 1**). They are submitted to contrasted levels of precipitations, with a mean annual precipitation of 2500 mm and 3500 mm (in years 2010 and 2011) respectively.

Both sites harbour different habitat types, from a bottomland to *terra firme*, but differ in landscape raggedness. At Laussat a permanently water-logged bottomland lies next to a plateau of low elevation (trees elevation at this site ranges between17 and 60 meters). At Régina, a seasonally water-logged bottomland with flooding events lies at the foot of hilltops of higher-elevation plateaus with steeper slopes (trees elevation ranges between 43 and 109 meters). At both sites, bottomlands are characterized by hygromorphic soils with a large accumulation of organic matter up to a depth of 1 metre. On the contrary, *terra firme* have well-drained ferralitic soils rich in iron oxides with a sand-clay texture allowing free vertical drainage.

All trees of diameter at breast height (d.b.h.) > 20 cm dbh were mapped in a continuous area of 6.7 ha in Régina, and in two areas of 2.5ha and 1.8ha, one in the bottomland and one in the plateau, in Laussat (**Fig. 1**). Population density varied between 29.9 adult trees/ha and 48.1 trees/ha in Régina and Laussat, respectively. In each study site, two groups of 30 trees, one for each habitat, were sampled for genetic analysis.

Molecular methods

Fresh leaves were sampled and frozen at -80°C. Genomic DNA was extracted from leaves using a CTAB protocol (Doyle & Doyle 1987, Colpaert *et al.* 2005), and each sample was extracted twice independently. Amplified fragment length polymorphisms (AFLPs) profiling was performed on all of the 240 samples according to the protocol of Vos & Bleeker (1995). DNA was digested using *Pst*I and *Mse*I restriction enzymes. Restriction fragments were amplified using two selective PCRs with respectively one and three selective nucleotides. Fifteen primer combinations were analyzed (see **Table 1**). The whole protocol was applied to each duplicate of all samples to obtain a complete replicate for each individual.

Peak profiles were scanned using PeakScanner v1.0 and the bin set was created using RawGeno v2.0. The complete method of AFLP scoring and data cleaning is available in supplementary methods. After data cleaning, 1196 markers were retained for further analysis.

Table 1: List of Primer used for AFLP profiling.

Combination	Pst+1	Pst+3	Mse+1	Mse+3
1	Pst+A	Pst+ACA	Mse +T	Mse+TAA
2		Pst+ATT		
3		Pst +AAC		
4		Pst +ATA		
5		Pst+ACA		Mse+TAG
6		Pst+ATT		
7		Pst +AAC		
8		Pst +ATA		
9	Pst+T	Pst+TAA	Mse +C	Mse+CAA
10		Pst+TAG		
11	Pst+A	Pst+ACA		Mse+CAA
12		Pst+ATA		
13		Pst+ACA		Mse+CAT
14		Pst+ATT		
15		Pst+ATA		

General statistical analysis

- *Linkage disequilibrium analysis*

Pairwise statistical disequilibrium between 0/1 AFLP scores at pairs of markers was estimated, within each study site, based on two-way contingency tables using a Fisher exact test.

- *Estimation of inbreeding (F_{IS}) coefficients*

We used SNPs data from a previous study (Audigeos *et al.* 2013) to estimate inbreeding coefficients (F_{IS}). We used ARLEQUIN v3.5.1.2 (Excoffier & Lischer 2010) to compute observed and expected heterozygosity in two study sites at SNP markers for which differentiation (F_{ST}) between local habitats were not significant. For each site, the mean F_{IS} (across loci) revealed an excess of heterozygotes varying from -0.207 to -0.089. We used a mean Fis of -0.14 for the present genetic analysis.

- *Genetic structure and spatial genetic structure analysis (SGS)*

A Bayesian clustering analysis was performed on AFLP data using STRUCTURE v2.3.4 (Pritchard et al. 2000, Falush *et al.* 2007) both at the regional and the local scale. The analyses were performed with the "admixture model" and "correlated allelic frequencies" settings. A burn-in of 10000 iterations was followed by 10000 iterations. Twenty runs were performed for all K (number of clusters) values from K=1 to K=9. The true number of genetic groups was assessed *a posteriori* using the method proposed by Evanno *et al.* (2005).

Fine-scale genetic structuring and gene dispersal were assessed using the spatial autocorrelation method based on kinship coefficients, developed by Hardy & Vekemans (1999) and implemented in SPAGeDi v1.3 (Hardy & Vekemans 2002). Within each site, the spatial autocorrelation of the kinship coefficient (F_{ij}) was analysed over twenty evenly spaced distance classes between 0 and 500 m. 95% *null* confidence intervals were obtained through 1000 random permutations of individuals among geographical locations. Neighbourhood size (N_b) and gene dispersal (σ_g) with prior knowledge about population densities in the study site were also estimated using SPAGeDi. The slope b of the regression of relatedness (F_{ij}) against geographic distance (d_{ij}) were also computed, along with their standard error estimated by Jackniffing over loci, and allowed the quantification of SGS intensity: $S_p=b/(F_{(1)}-1)$ where $F_{(1)}$ is the average kinship coefficient between individuals separated by distances belonging to the first distance class.

- *Assignment of "synthetic" AFLP genotypes*

Expected heterozygote frequencies within each study site and local habitat were computed based on the inbreeding coefficient estimated from SNPs data for each sub-population, by solving the equations relating the inbreeding coefficient and allele and genotype frequencies: for each marker j, with f indicating relative frequencies and N indicating absolute frequencies:

$$f(00)_j = (1 - F_{IS}) * q_j^2 + (F_{IS} * q_j) \leftrightarrow (1 - F_{IS}) * q_j^2 + \left(F_{IS} * q_j\right) - f(00)_j = 0$$

where 'o' is the "absence of peak" allele, and q is the relative frequency of the 'o' allele;

solving for q:

$$q_j = \frac{-Fis+\sqrt{\Delta_j}}{2*(1-Fis)} \text{ with } \Delta_j = Fis^2 - [4 * (1 - Fis) * (-f(00)_j)]$$

The expected absolute frequency of heterozygotes is:

$N(01)_j = 2 * N_j * q_j - 2 * N(00)_j$ where N_j is sample size for this marker. Finally, $N(01)_j$ fragment-carrying samples were randomly selected and assigned the heterozygote (01) genotype. The remaining fragment-carrying samples were assigned the dominant (11) genotype.

- *Detection of loci under selection*

Adaptive divergence within populations inhabiting contrasting habitats was assessed based on two F_{ST}-based approaches: (a)

we used the coalescent-based FDIST method (Excoffier & Foll 2009) implemented in ARLEQUIN v3.5.1.2 (Excoffier & Lischer 2010). This method simulates samples under a hierarchical island model by the coalescent and compares the observed genetic differentiation for each locus to the *null* distribution obtained from simulations. (b) we used the Bayesian method implemented in BAYESCAN to detect outliers (Foll & Gaggiotti 2008). The method relies on a logistic regression model that partitions the genetic differentiation at each locus within each population into two components: a population-specific component (beta) common to all loci, and a locus-specific component (alpha); if the latter is significantly non-zero, this is interpreted as a departure from neutrality at the locus.

We looked for loci under selection between local habitats within each study sites and between the two local habitats without distinction of the study site (local habitat partitioning).

Results

Blind Bayesian clustering

At regional scale, three clusters were detected with a maximum ΔK for K=3 (**Fig. 2**). The first cluster included all individuals from Regina, the second cluster included individuals from the plateau of Laussat ant the third cluster included trees inhabiting the bottomland of Laussat. For K=2, the two groups corresponding to the two study sites were retrieved.

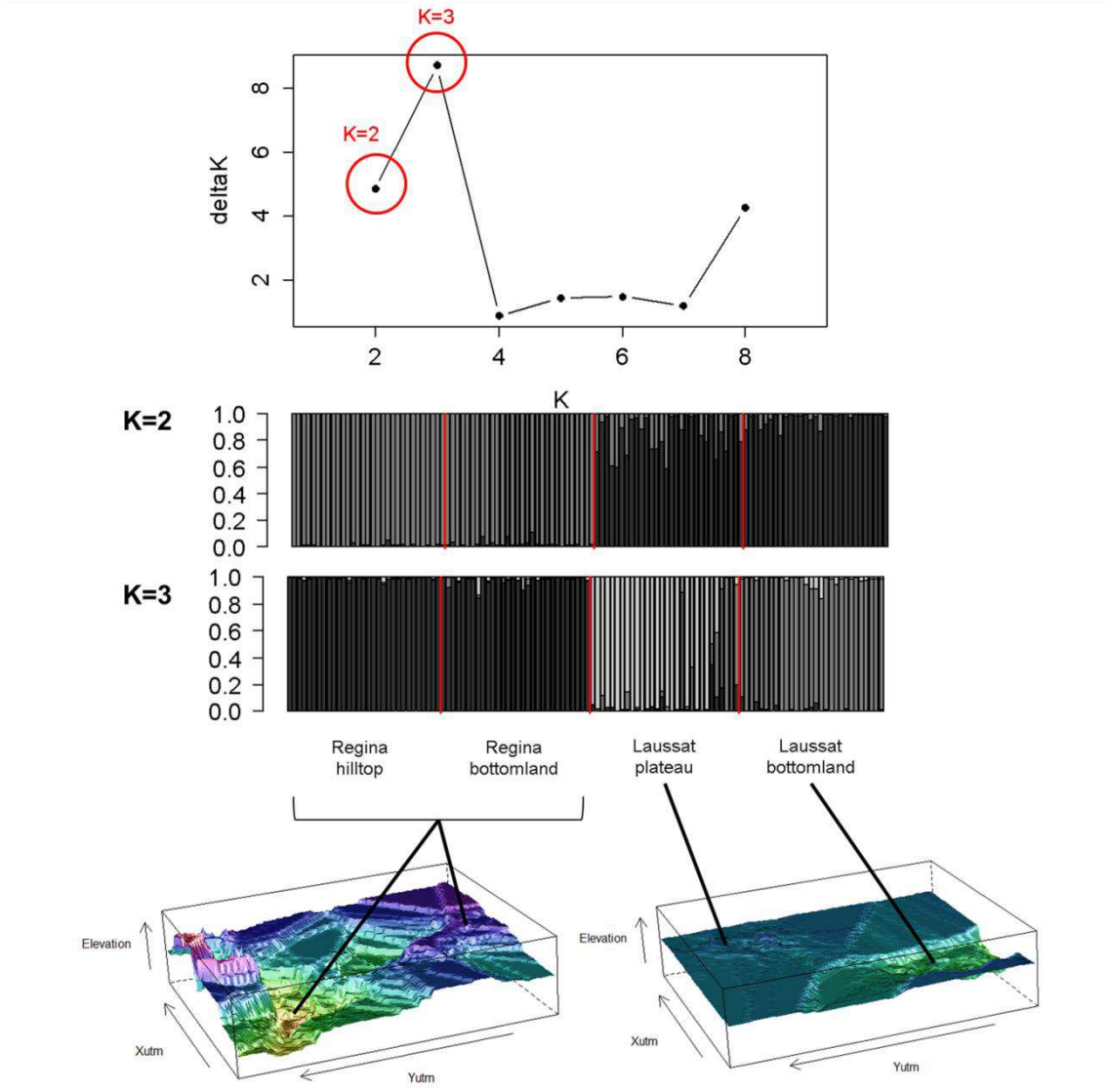


Figure 2: Bayesian clustering analysis on the whole data set. Upper pane: Δ*K* values. Lower pane: individual α values for *K*=2 and *K*=3.

Within Laussat, two peaks for ΔK were detected, for K=2 and K=6 (**Supplementary Fig. S1**). At K=2, inferred clusters are geographically grouped in agreement with local habitat

structuring and with the spatial subdivision of the sample, except for 5 trees belonging to the hilltop and that were assigned to the same cluster as individuals from bottomland. At K=6, two 'major' groups detected (based on the number of individuals assigned to) were in accordance with local habitats, whereas the remaining clusters included several individuals from both habitats.

At Regina, a maximum ΔK was found for K=7 (**Supplementary Fig. S2**). Individuals assigned to the different clusters were geographically grouped. A second peak of ΔK was found at K=3: one cluster of trees from the bottomland, one cluster of trees from the hilltop, and a large cluster of trees scattered across the site. As in Laussat, individuals were associated to a cluster in agreement with local habitat structuring and spatial sample subdivision for K=2.

Fine-scale SGS Analysis and gene dispersal estimation

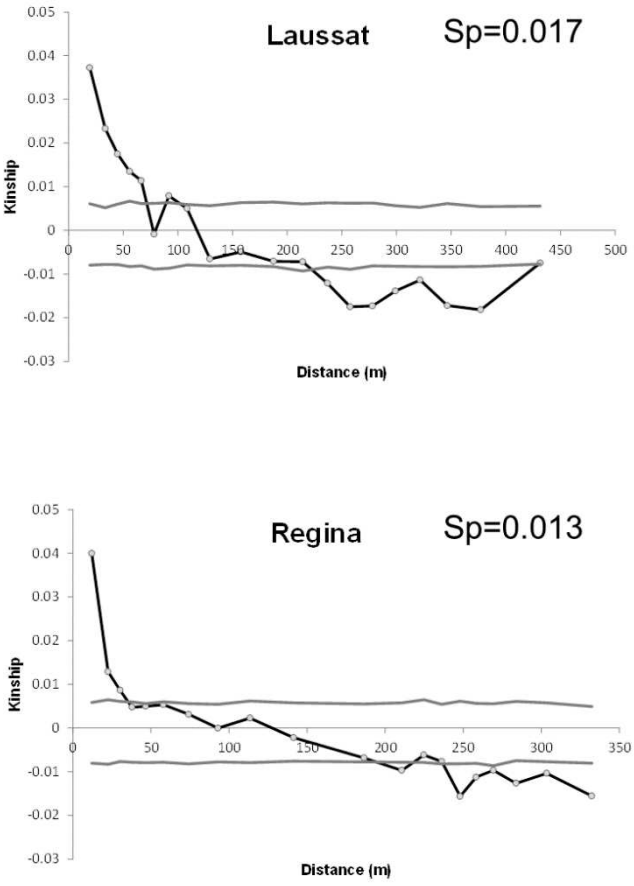


Figure 3: Spatial genetic structure analysis based on all AFLP markers.

Spatial genetic structuring was assessed by estimating relative relatedness between 1711 pairs of individuals in Regina and 1810 pairs in Laussat. The mean number of pairs by distance class was 86 in Laussat and 92 in Regina. Significant SGS was detected in both sites (**Fig. 3**), with kinship declining with increasing geographical distance. In Laussat, spatial autocorrelation was significantly positive until 56 m, and it became significantly negative from 230 m onward. In Regina, autocorrelation was positive and significant until 30 m and became negative and significant after 250 m. Gene dispersal was estimated at 45.7 and 64.39 m in the two sites respectively (**Table 2**).

Table 2: SGS and gene dispersal parameters estimated by SpaGeDi

		Laussat	Regina
SGS parameter estimates	b (SE)	-0.016(0.001)	-0.012 (0.001)
	F1 (SE)	0.037 (0.003)	0.04 (0.032)
	Sp	0.017	0.013
Gene dispersal parameter estimates	D	0.005	0.003
	Nb (SE)	65.62 (21.02)	78.51 (14.21)
	σg (SE)	45.7 (7.33)	64.4 (5.82)

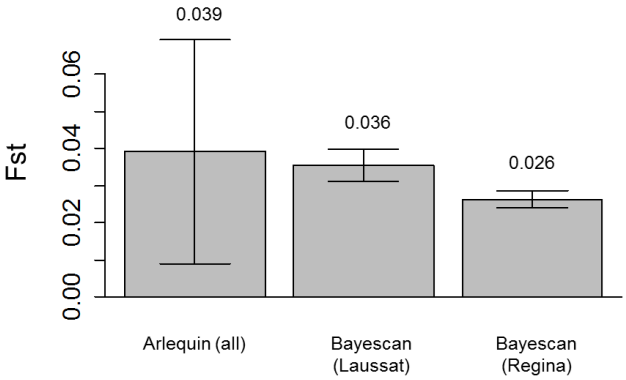


Figure 4: F_{ST} values for all comparisons (boxes) with their confidence intervals (bars).

Outlier detection

Overall F_{ST} was 0.039 among the four sub-populations, 0.036 and 0.026 between habitats within site for Laussat and Régina, respectively (**Fig. 4**).

A total of 24 loci were detected as outliers being under divergent selection in at least one analysis and are summarized in **Table 3**.

Table 3: Summary of outliers detected in at least one analysis.

Marker number	Coalescent method	Bayesian method		Remark
		Within Laussat	Within Regina	
86	ns	*	ns	
158	*	ns	ns	
181	*	ns	ns	
222	*	ns	ns	band is more frequent in hilltop at both sites
233	*	ns	ns	band is more frequent in hilltop at Regina, band is absent in bottomland at Laussat
287	*	ns	ns	
313	ns	ns	*	band is more frequent in bottomland at both sites
345	*	*	ns	band is present at Laussat bottomland only
451	*	ns	ns	band is present at Laussat bottomland only
463	*	ns	ns	band is present at Regina hilltop only
485	*	*	ns	band is more frequent in bottomland at both sites
605	*	ns	ns	band is more frequent in bottomland at Regina, band is absent in hilltop at Laussat
624	*	*	ns	band is more frequent in hilltop at Regina, band is absent in bottomland at Laussat
668	*	ns	ns	band is more frequent in hilltop at both sites
672	*	ns	ns	band is more frequent in hilltop at both sites
687	*	ns	ns	band is present at Regina bottomland only
742	*	ns	ns	band is more frequent in bottomland at both sites
757	*	ns	ns	band is more frequent in bottomland at lausssat, band is absent in hilltop at Regina
785	*	ns	ns	band is more frequent in bottomland at both sites
799	*	ns	ns	
814	*	ns	ns	band is more frequent in hilltop at both sites
848	*	ns	ns	
881	*	ns	ns	
962	ns	ns	*	

Under the coalescent model, the between-sites outlier search (Fct) detected 16 loci with excess divergence (1.34 %) and 53 loci (4.43%) with a divergence deficit ($P \leq 0.01$). Tests involving sub-populations within regions (Fst) based on the hierarchical island model detected 21 loci with significantly large divergence (1.75%) and 31 loci (2.60%) with significantly small divergence ($P \leq 0.01$) (**Fig. 5**).

The Bayesian analysis within each site detected four outliers (M86, M345, M485 and M624, FDR=0.084 and FNDR=0.092) at Laussat, and two outlier (M313 and M962, FDR=0.01 and FNDR=0.091) at Regina (**Fig. 6**). Fisher exact tests on AFLP score frequencies revealed significant statistical linkage between the markers M345, M485 and M624 in Laussat (P=0.03 between M345 and M485, P=0.0003 for M345/M624 and P=0.0004 for M485/M624), **Fig. S5**. Three outliers detected by the Bayesian analysis in Laussat (M345, M485 and M624) were also detected by the coalescent approach at a p-value ≤ 0.01. Many outliers (12/1196) showed a similar pattern of band presence (phenotype '1') frequency variations between local habitats in the two study sites. For M222, M233, M624, M668, M672 and M814, the band frequency was higher in hilltop than in bottomland in both sites. For M313, M485, M605, M742, M757 and M785, the band frequency was higher in bottomland than in hilltop in both sites. M345, M451, M463 and M687 were detected as outliers but an AFLP band was only present in one habitat of one of the study sites and may not be considered as a 'true outlier'.

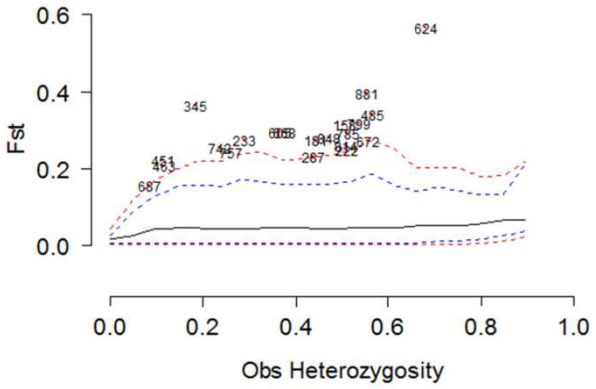


Figure 5: Results of the coalescent outlier search. Blue dashed line: 95% neutral envelop; red dashed line: 99% neutral env

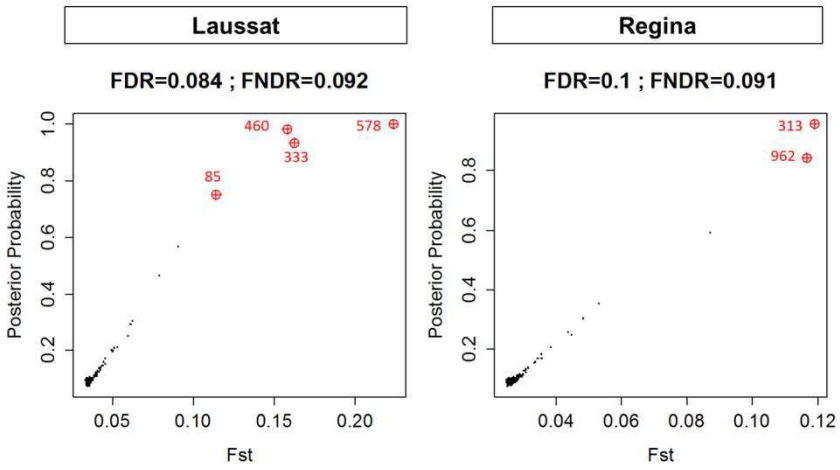


Figure 6: Results of the Bayesian outlier search.

Discussion

Bayesian clustering

At the regional scale, the large genetic differentiation between study sites (K=2) can easily be explained by isolation by distance. Nevertheless, the most likely number of clusters was K=3, with one group for Regina, and two groups for Laussat, one from the bottomland and one from the plateau. This suggests a larger differentiation between groups inhabiting contrasting habitats, or belonging to distinct spatial groups, in Laussat than in Regina, as also revealed by estimates of within-site F_{ST} values (**Fig. 4**). Within both sites with K=2, individuals from different local habitats formed separate clusters. This result suggests overall restriction to genetic mixing at short distances, with further genetic subdivision, as shown by ΔK peaks at K=3 and K=7 in Laussat and Regina respectively.

Fine-scale SGS and gene dispersal

To evaluate the role of neutral processes in shaping within-population genetic structure, we investigated the fine-scale genetic structuring over all loci within each study sites. Kinship coefficients decreased quickly with geographical distances in the two study sites as expected under the isolation-by-distance model: significant relatedness between individuals became non-significant at 56 and 30 meters for Laussat and Regina respectively. The structure was globally flat, with kinship values never exceeding 0.04 above or below the population average, indicating that the spatial distribution of relatedness is relatively uniform (as a term of comparison, kinship values are 10-fold larger for Bayesian outliers; **Supplementary Fig. S1**). Similar SGS patterns have already been observed in several temperate (Leonardi & Menozzi 1996, Streiff *et al.* 1998, Vekemans & Hardy 2004, Vornam *et al.* 2004, Jump & Penuelas 2007, Chybicki *et al.* 2011, Hampe *et al* 2010, Oddou-Muratorio *et al.* 2010, Jump *et al.* 2012) and tropical tree species (Stacy *et al.* 1996, Doligez & Joly 1997, Konuma *et al.* 2000, Dick *et al.* 2003, Lowe *et al.* 2003, Cloutier *et al.* 2006, Born *et al.* 2008, Collevati *et al.* 2010), including in the Guiana shield (Dutech *et al.* 2002, Degen *et al.* 2004, Latouche-Halé *et al.* 2004, Cavers *et al.* 2005, Hardy *et al.* 2006).

Significant, albeit weak, spatial genetic autocorrelation is commonly explained by a restricted gene flow by seed and pollen, because it can lead to significant genetic differentiation within continuous stands (Cavers *et al.* 2005). Even if the strength of SGS may be influenced by some AFLPs submitted to divergent natural selection across habitats (Jump *et al.* 2012), significant spatial autocorrelations observed over all loci within each site

is more likely to be caused by neutral processes In tropical rainforests, gene dispersal is commonly restricted to short distances, as observed in numerous studies (Ward *et al.* 2005, Hardy *et al.* 2006, Dick *et al.* 2008, and all references within). In *E. falcata*, gene dispersal estimates (σ_g ranging from 45.7 to 64.4 meters depending on the study site) are concordant with the hypothesis of limited gene flow and local inbreeding, as in other tropical and temperate trees (Heuertz *et al.* 2003, Oddou-Muratorio *et al.* 2010). Nevertheless, a previous study (Audigeos *et al.* 2013) found similar patterns in this species, but showed that genetic turnover occurs at shorter distances in the direction of environmental gradients than within environmentally homogeneous patches. This suggests that at least part of the observed SGS is linked to environmental filters.

Outlier detection

F_{ST} values found between sub-populations within sites were rather large (between 2.6% and 3.9%), considering the small distance of the sampled groups at each site (up to 200 m). This suggests the presence of mechanisms inducing divergence at a highly local scale. The most likely candidates are neutral processes (drift, inbreeding, restricted gene flow, and demographic events); differentiation outliers are suggestive of the action of additional evolutionary processes, such as various forms of selection, that may make these loci depart from the neutral average distribution (Beaumont & Balding 2004, Foll & Gaggiotti 2008, Excoffier *et al.* 2009); alternatively, they may indicate the presence of some other indirect mechanisms inducing genetic divergence, that may or may not be directly related to environmental filters (Excoffier *et al.* 2009, Hermisson 2009, Bierne *et al.* 2013). Outlier tests based on a differentiation index (F_{ST}) are robust to inter-locus variations, and theoretical models show that footprints of natural selection persist longer kept in differentiation indices (F_{ST}) than in intra-population estimators of genetic diversity (Storz 2005). F_{ST}-based methods are also supposed to be robust to many demographic scenarios (Beaumont 2005, Bonin *et al.* 2006), partly because demographic events affect the genome in a homogeneous manner (Eveno *et al.* 2007). However, the inclusion of bottlenecked populations may bias the method (Storz 2005), and the degree to which these tests are robust to demography has not been fully explored (Nielsen *et al.* 2005).

Both the coalescent and the Bayesian method detected outliers at the very local scale we studied (between sub-populations separated by few hundred meters at most). Three outliers were detected by both methods (M345, M485 and M624). Moreover, many markers

(M222, M233, M313, M485, M605, M624, M668, M672, M742, M757, M785 and M814) showed a similar pattern of band frequency variations between local habitats in the two study sites; markers 345, 451, 463 and 687 may not been considered as a 'true outlier', as described in the Results section. There were fewer outliers with excess divergence between sites than between sub-populations within site. This suggests that factors driving divergence among regions are not necessarily stronger than those occurring between local sub-populations growing in different habitats. As the effects of dispersal limitation can only increase with distance, it seems unlikely that this kind of neutral process be stronger locally than at the regional level, suggesting that highly local divergence may be due to selective forces. In this study we have found a lower proportion of outliers than in Audigeos *et al.* (2013) in the same species at two different sites. However, both the kind and the number of marker used differed between the two studies, as Audigeos *et al.* (2013) focused on few hundred SNPs markers, by using a combination of candidate genes and anonymous loci, that may have been enriched for loci undergoing divergent selection. Yet, the proportion of outliers found in our study was surprisingly high when considering the very local scale studied here. Scans for outlier detection at varying geographical scales in a variety of biological models (including animals and plants, both aquatic and terrestrial) are abundant in the literature. **Table 4** displays a survey of such studies.

We found 21 outliers (1.8%) with the coalescent method and 6 outliers (0.5%) with

Table 4: Detection of outliers for selection in several studies varying in biological models and geographical scales. Model: A/P = animal/plant; W/T = aquatic/terrestrial; d = geographical distance (Km) (e: elevation); Method: C: coalescence; B: Bayesian; SS: summary statistics (software used, in parentheses: BF: BAYEFST; BS: BAYESCAN; D: DFDIST; F: FDIST2; FS: FstSNP); N: Number of markers; P: proportion of outliers (e: elevation; B: Bayesian; C: coalescent; SS: summary statistics; T: temperature gradient; R: rainfall gradient)

Species	Model	d	Marker	Method	N	P	Citation
Salmo trutta	A,W	600	SSRs	C (D)	74	2.7%	Meier *et al.* (2011)
Gasterosteus aculeatus	A,W	2200	SSRs	B (BF)	57	8.7%	Makinen *et al.* (2008)
Zostera marina	P,W	50	SSRs	C (F)	25	4%	Oetjen &Reusch (2007)
Capra hircus	A,T	900	SNPs	C (F)	27	11%	Pariset *et al.* (2009)
Peromyscus maniculatus	A,T	200; 0.5 (e)	Isozymes	C (F)	12	0%; 8.3% (e)	Storz & Dubach (2004)
Linum usitatissum L	P,T	Worldwide	SSRs	C (F); B (BS)	150	0%	Soto-Cerda & Cloutier (2013)
Picea mariana	P,T	700	SNPs	SS (FS); B (BF)	768	T: 1.2%(SS); 0.26%(B) R:1.4% (SS)0.13%(B)	Prunier *et al.* 2011
Fagus sylvatica	P,T	10	AFLPs	C (D)	254	0.39%	Jump *et al.* (2006)
Littorina saxatilis	A,T	70	AFLPs	C (D)	2356	0.033%	Galindo *et al.* (2009)
Rana temporaria	A,W/T	100	AFLPs	C (D)	392	1.5%	Bonin *et al.* (2006)
Timema	A,	5	AFLPs	C (D)	534	14.6%	Nosil *et al.* (2008)
Eperua falcata	P,T	0.5	SNPs	C (F)	57	3.5%	Audigeos *et al.* (2013)

the Bayesian method. 3 outliers (0.3%) were detected by both methods. Among all outliers, 12 (1%) showed a similar pattern of band frequency variations between local habitats in the two study sites. These estimates are of the same order of magnitude as those detected in studies comparing populations separated by larger distances, suggesting that the same processes that occur with a larger degree of spatial separation in other species may occur at very short distances in *E. falcata*. How should the observed divergence be interpreted? It is customary to interpret directly outliers as loci under directional selection or tightly linked to loci under selection; however, this logic has been criticised (Bierne *et al.* 2011, Bierne *et al.* 2013, Hermisson 2009) and the methods shown to be sensitive to genetic structure and demography (Excoffier *et al.* 2009). To assess how confident we can be in weighing the role of selection in the generation of outliers, we need to check our biological system against possible departures from the theoretical model underlying the tests. Bierne *et al.* (2013) list exhaustively the possible assumption violations: (a) departures from Wright's island model: we have sampled continuous populations, which can confidently considered as equivalent to a continuous island model (Hardy & Vekemans 1999); (b) variations in recombination rate around loci: although this cannot be excluded, we have used entirely anonymous markers of a single type, which lets us think that there should be no systematic bias; (c) selective sweeps: this cannot be excluded and we have no information about it; (d) cryptic hybrid zones: this is rather unlikely, given that the sub-populations we have sampled from different habitats belong to a well-defines botanical and genetic species that has a clustered distribution, with each cluster covering both habitats; (e) pervasive selection: although this cannot be excluded *a priori*, it seems highly unlikely that the AFLP markers used here are massively selected; (f) genome-wide effects of genetic incompatibilities: surely we cannot exclude all effects of some restriction to mating between trees from different habitats (beyond plain restriction of dispersal); for example, differences in resource availability (e.g. water may be available at different dates in different habitats) may cause shifts in flowering phenology, so that trees from the same habitat may mate more frequently; or flowering traits, influencing pollinator behaviour, may be correlated to other physiological traits, which may be in turn ecologically divergent. In both cases, there would be pre-zygotic barriers to hybridisation, which would fall in the 'coupling effect' category described by Bierne *et al.* (2011). However, although flowering phenology is irregular, we have not detected time shifts between habitats at the same site for *E. falcata* and flower trait dimorphism has so far not been reported. In conclusion, with the exception of point (c), it

seems unlikely that our experimental design violates in any major way the assumptions of "selection outlier" detection methods. Therefore, we tend to think that the observed divergence is caused by some form of directional or disruptive selection. In *E. falcata*, a common garden experiment has further revealed intrinsic differences in seedling growth and leaf physiology between subpopulations from divergent habitats (Brousseau *et al.* 2013), which suggests some form of adaptive divergence for complex traits. The selective agency behind the observed divergence needs to be proven experimentally and functionally, in particular by showing that the putatively selected polymorphisms control adaptive traits; if it is the case, one should expect that the selected loci have major effects on traits, because in the case of the classical polygenic model it is not expected that quantitative trait loci underlying traits under divergent selection be more divergent than neutral loci (Kremer & Le Corre 2012, Le Corre & Kremer 2012). The observation of patterns of divergence at the highly local scale studied here indicates that major evolutionary events can occur even within continuous populations.

Acknowledgements

We thank Saint-Omer Cazal and Valérie Troispoux for technical assistance and Caroline Scotti-Saintagne for her valuable advice with both data acquisition and analyses. The project was funded by the LABEX CEBA.

Article n°2 - Supplementary methods

Supplementary methods I - AFLP scoring

- *Reading of Peak profiles*

We used PeakScanner to read peak profiles within the range 50-500 bp.

- *Thresholds definition*

We analyzed peaks profiles in both negative controls and sample profiles in the whole analysis window (50-500 bp) for defining detection thresholds. For each combination, we analyzed all peaks contained in negative controls and defined "lim.max.0" as the 95% quantile of peaks height within negative control. Thus, we suggest that the 5% of higher peaks in negative control may be "true peaks" due to contaminations, while 95% of peaks contained in negative controls may be considered as "true" background noise.

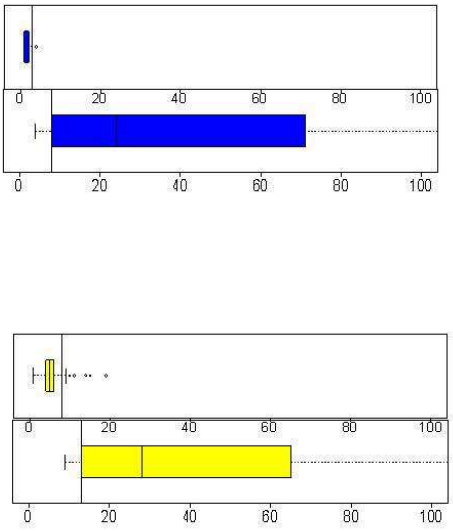


Figure: Peaks distribution and thresholds in both negative controls (top) and samples profiles (bottom).

Then, we analyzed the distribution of all peaks higher than "lim.max.0" in sample profiles and defined a "lim.min.1" threshold corresponding to 25% of the distribution of peaks height outside background noise. These thresholds will be used to score "peak presence" in the further steps.

- *Binset definition*

Binset was defined using RawGeno (Arrigo 2009) with parameters: maximum=2bp, minimum bin width=1bp, range=50-500bp. We used "lim.min.1" as threshold for bin design. We didn't use the "replicate" option because we wanted to analyze the replicates of each sample independently before doing consensus. Binset was manually corrected and exported. Last, we searched for peaks within bins and used the intensity of each sample within each bin for data cleaning and scoring steps.

- *Data Pre-cleaning*

We pre-cleaned data by removing, for each combination, samples for which the two replicates were not available due to problems during genotyping (off-scale size standard or profile of bad quality).

- *Scoring and consensus*

We assigned “0”, “N”, or “1” to each peak according with the criterions:

Peak ≤ lim.max.0 : “0”

lim.max.0 < Peak < lim.min.1 : “N”

Peak ≥ lim.min.1 : “1”

and defined consensus as follow:

0/0 or 0/N : “0” (we considered the phenotype “0” if the replicates displayed two peaks within background noise or one peak within background noise and a “small peak” of intensity lower than lim.min.1).

1/1: “1”

1/N : “N” or “1” if a small peak outside blank (“N”) was supported by a peak of high intensity in the replicate (peak intensity within the 0.5 upper quantile of peaks intensity distribution within samples, see above).

0/1: “*” (“*” indicated a mismatch)

N/N : “$” : (“$” indicated an ambiguity, i.e. the two replicates displayed two peaks outside background noise but intensity lower than lim.min.1)

- *Data post-cleaning*

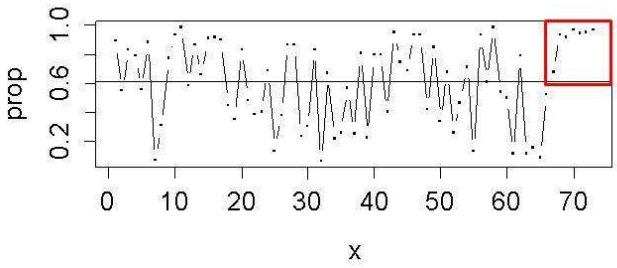


Figure : Masking step

Data were post-cleaned by eliminating markers for which a peak of intensity higher than lim.min.1 was found in a least one negative control (contaminant).

We also remarked that peak intensity decreased within the 50-500bp window and that it was variable among profiles. It resulted in numerous 0 and missing data (including “N”, “*” and “$”) in profile tails that would result in assigning false-“0” to an absence of peak. To avoid it,

We masked all “0” at the end of each profile (“N”) until a true peak (noted “1”) was found

We removed all markers displaying any “1” or any “0”.

We removed the last markers for which the proportion of “0” plus missing values (including “N”, “*” and “$”) were higher than the mean proportion of “0” plus missing values in bins of the whole dataset.

Last, we removed all markers that did not display at last 15 “true values” (true “0” or true “1”) for each site and local habitat.

- Data fusion

Last we merged datasets from the different combinations. Samples that were absent from a combination (because one of the two replicates failed) were noted NA (by opposition with missing values due to post-treatment "N", "*", and "$").

Supplementary methods II – Genotypes allocations

For each sub-population inhabiting each local habitat, we estimated the frequency of homozygote (11) and heterozygotes (01) under the different values of Fis using the observed frequency of individuals displaying the phenotype [0] that necessarily corresponds to the genotype (00).

For each marker j,

$f(00)_j = (1 - Fis) * q_j^2 + (Fis * q_j) \leftrightarrow (1 - Fis) * q_j^2 + \left(Fis * q_j\right) - f(00)_j = 0$ where q traduces the frequency of the allele (0).

Thus,

$q_j = \frac{-Fis+\sqrt{\Delta_j}}{2*(1-Fis)}$

With $\Delta_j = Fis^2 - [4 * (1 - Fis) * \left(-f(00)_j\right)]$

Thus,

$N(01)_j = 2 * N_j * q_j - 2 * N(00)_j$ where N_j corresponds to the number of phenotypes available for this marker (with removal of missing values)

Last, we assigned the genotype (00) for individuals displaying the phenotype [0]. We also randomly assigned (01) or (11) for individuals displaying the phenotype [1] with respect with the expected number of heterozygotes (N_{01}) and homozygotes (N_{11}) estimated.

Article n°2 - Supplementary figures

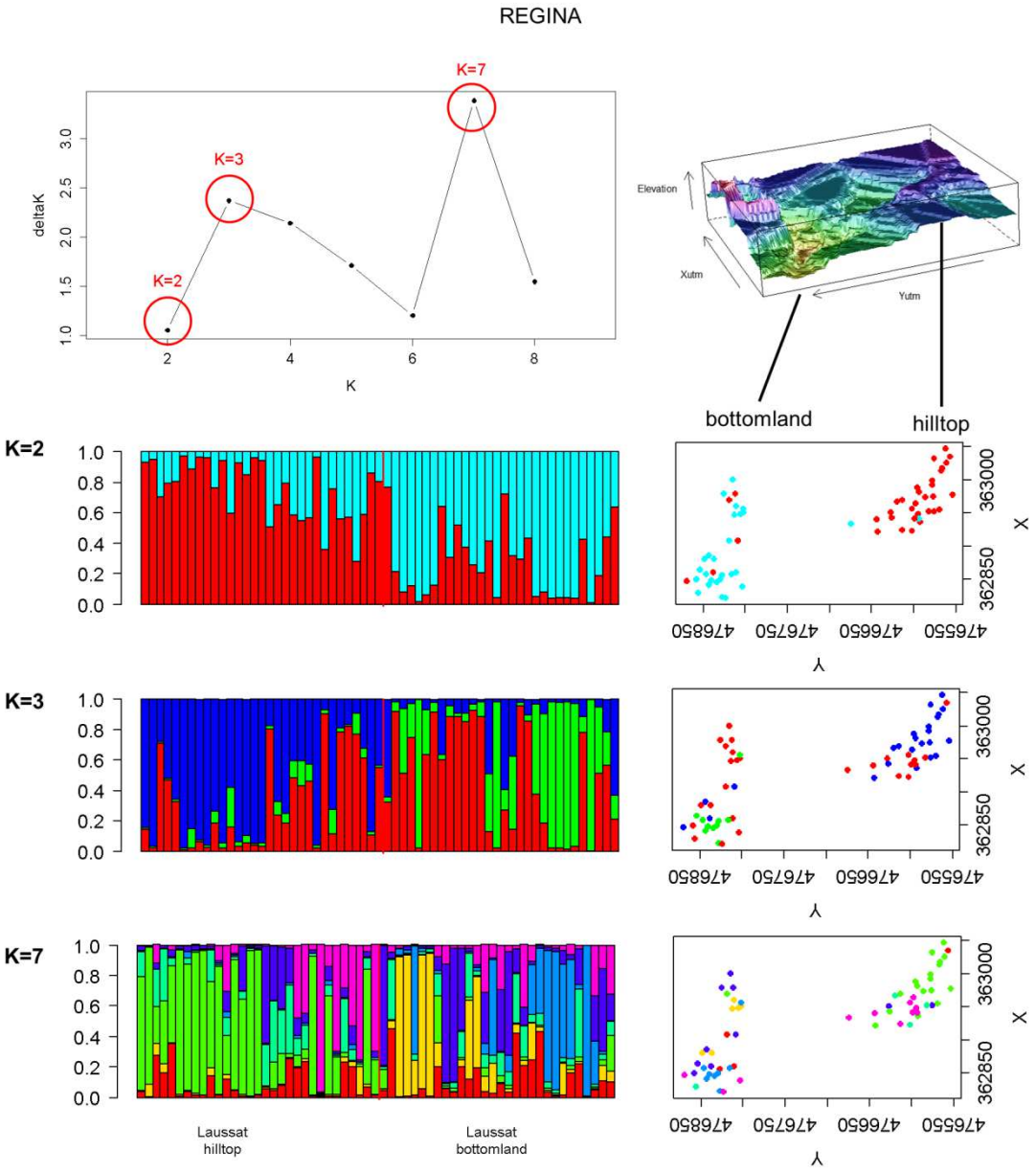


Figure S1: Bayesian clustering analysis on the Régina data set. Left upper pane: ΔK values. Left lower pane: individual α values for $K=2$ and $K=6$. Right pane: geographical distribution of individuals belonging to the main clusters (see text).

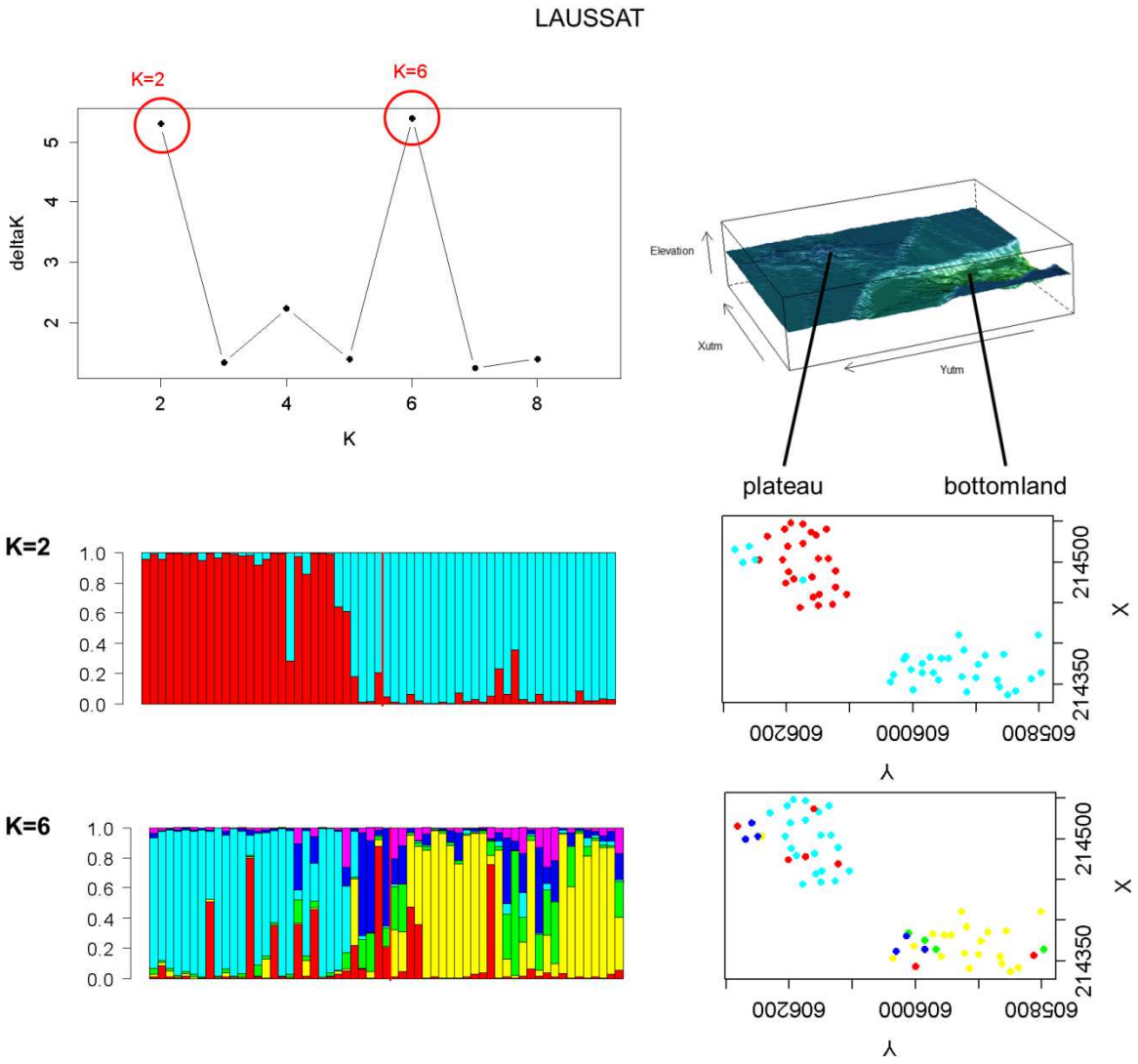


Figure S2: Bayesian clustering analysis on the Laussat data set. Left upper pane: ΔK values. Left lower pane: individual α values for K=2, K=3 and K=7. Right pane: geographical distribution of individuals belonging to the main clusters (see text).

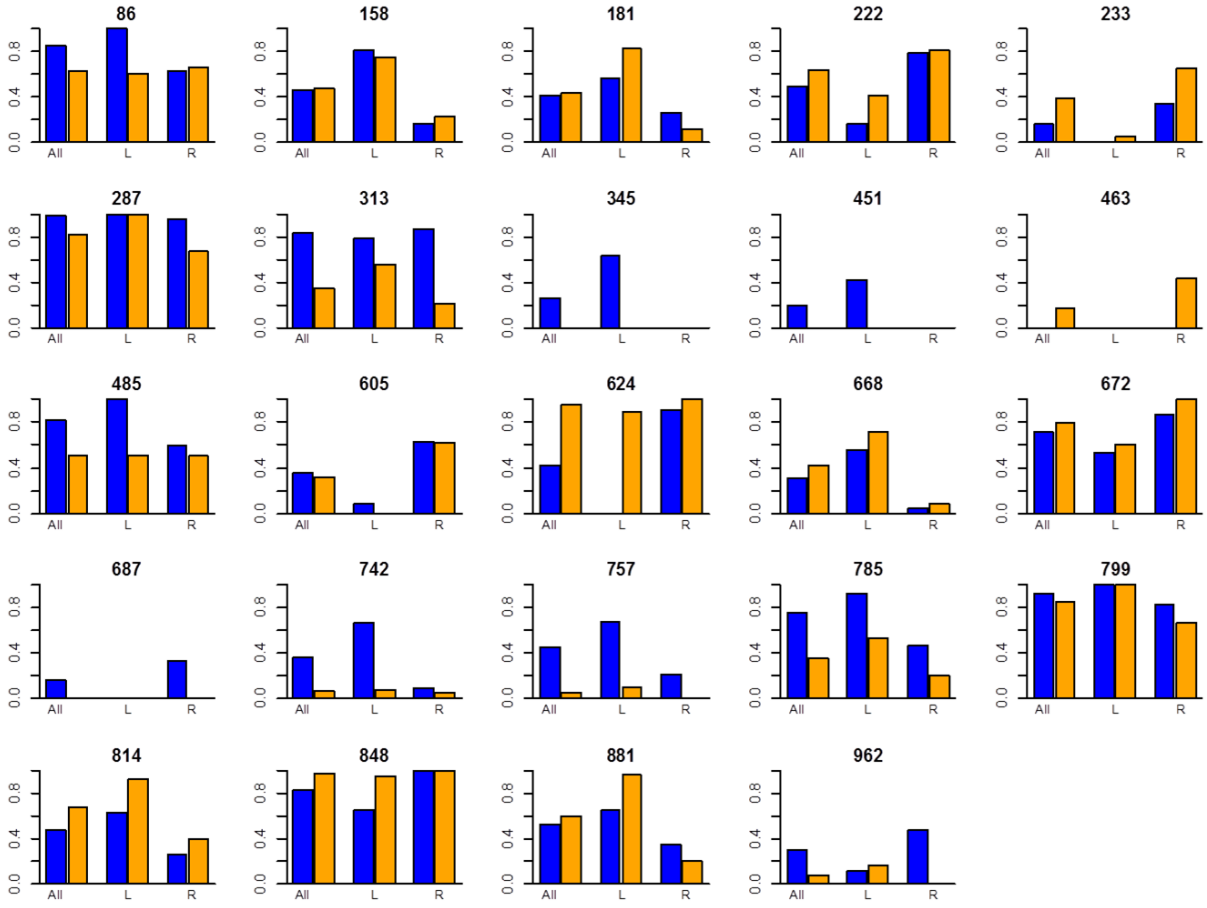


Figure S3: Band presence (phenotype '1') frequency for the global data set ('All'), Laussat ('L') and Régina ('R'). Blue bars: bottomlands; orange bars: hilltops

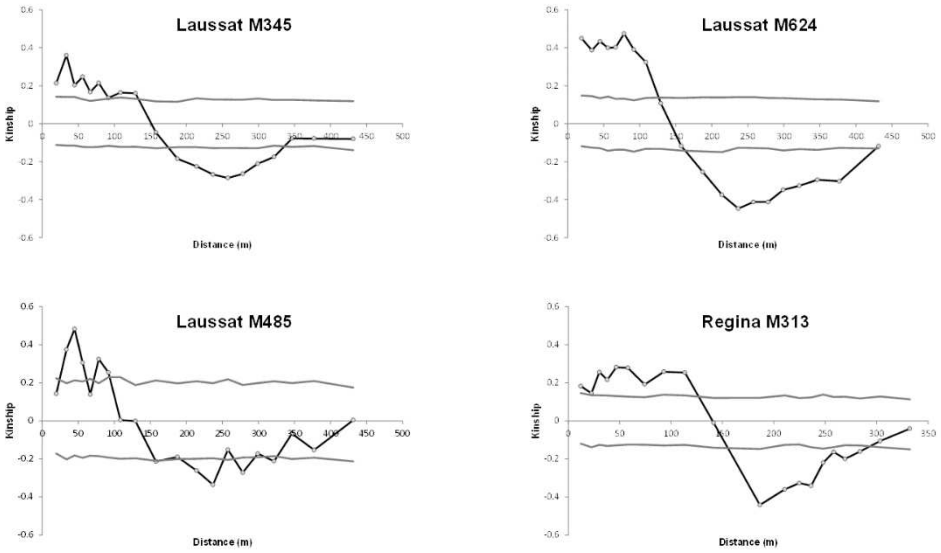


Figure S4: Spatial genetic structure analysis for Bayesian outliers.

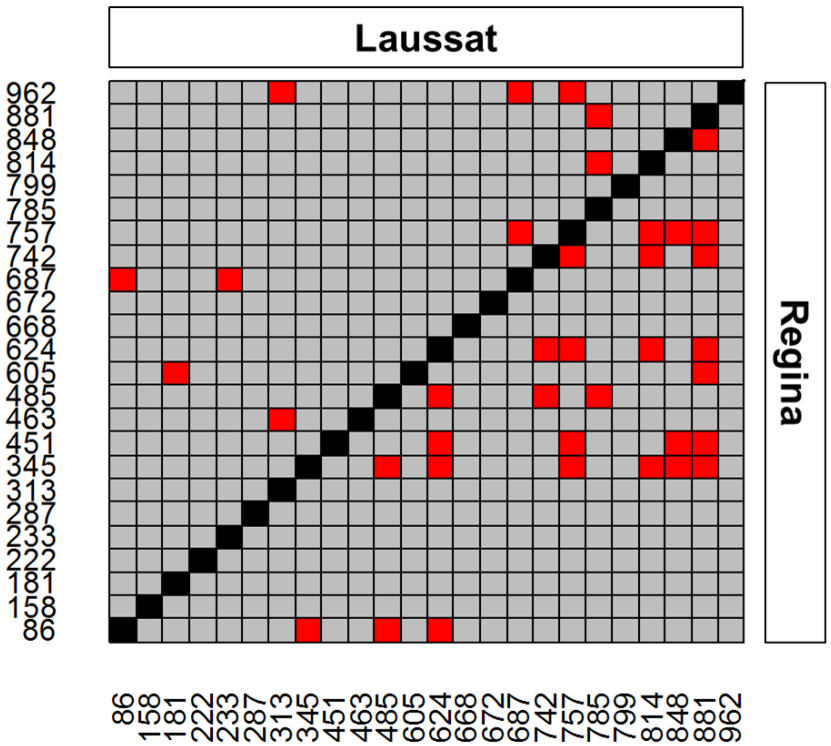


Figure S5: Linkage disequilibrium between all pairs of outliers within each study site. A "red" polygon indicates a significant p-value ($p<0.05$), whereas a "grey" polygon indicates a non-significant p-value ($p\geq0.05$).

Bioinformatic tools - 'Rngs': A suite of R functions to easily deal with next-generation (454-)sequencing data and post-process assembly and annotation results.

1. Introduction to bioinformatics

Next generation sequencing technologies are now able to produce large amount of genetic data with reduced cost and time. However, dealing with such datasets needs to be automatized and requires specific tools often difficult to use.

Assembly. Next-generation sequencing allows sequencing large amounts of DNA templates, but produces small sequenced fragments (reads) that need to be assembled into contigs. In 454-pyrosequencing for example, DNA libraries are sequenced by fragments of 200 to 500 bases, and each portion of the genome/transcriptome may be sequenced one or several times (**Fig. 1**). Numerous software packages allow the assembly of NGS reads into contigs, among which the most popular are probably MIRA, Newbler, and CAP3.

Figure 1: Example of assembled data from 454 pyrosequencing.

Blast and functional annotations. Once reads are assembled into contigs, blast and functional annotations allow the characterization of unigenes among contigs. Blast compares the assembled sequence of each contig to public databases (including encoded proteins, dna sequences, and ESTs sequences). In a second step, functional annotation allows identifying the biological processes and metabolic pathways in which each unigene is involved, by assigning "Gene Ontology terms" (http://www.geneontology.org/) to contigs that returned a blast result in the previous step. A widespread tool software used to blast and annotate plant assemblies is B2G (Conesa & Gotz 2008).

SNPs discovery. Genetic analyses start with the existence of polymorphisms in DNA

sequences. Thus, identifying polymorphisms in characterized unigenes is a major feature of evolutionary biology, quantitative genetics and QTL mapping. However, identifying 'true' polymorphisms remain challenging due to higher error rates in next-generation sequencing than classical Sanger sequencing of targeted sequences. It requires the filtering of assembly mismatches through the use of informative statistics. Numerous softwares automatize SNP detection, but the large majority of them remain complex, incomplete (without taking into account for individual bases qualities within the different reads) or expensive. The package presented here is largely inspired from SeqQual software.

2. Short description of 'Rngs'

"Rnsg" is a suite of R scripts developed during this PhD that allows to easily post-process NGS data from 454-pyrosequencing. Functions are coded in the widespread R language to be freely available and modifiable by users. Because R is a simplified programming language, it is more accessible for evolutionary biologists than other languages, such as python or perl.

This package contains functions that help dealing with NGS data and outputs from different widespread softwares packages (**Fig. 2**). The functions are organized into three categories: "assembly", "blast & annotation", and "SNP discovery".

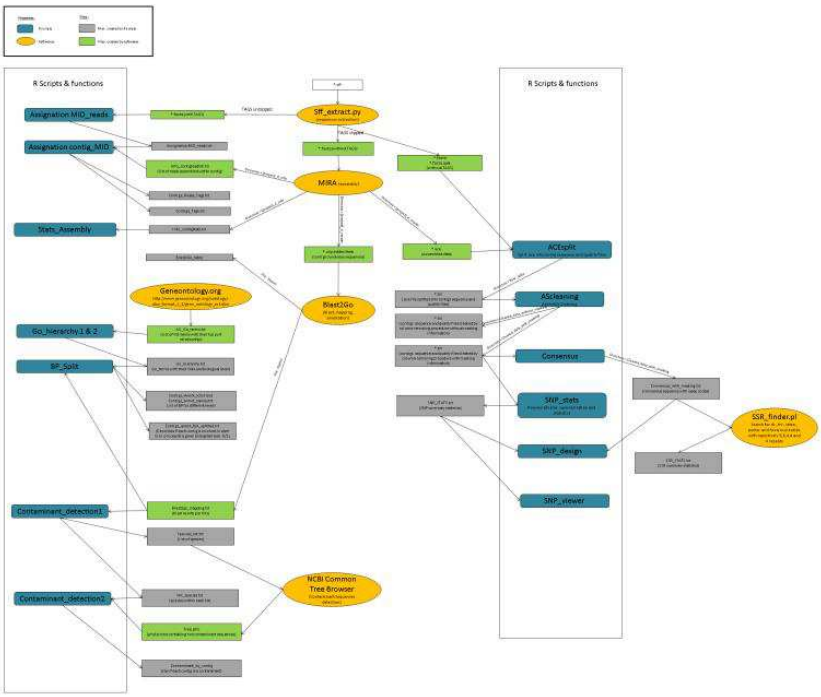


Figure 2: Overview of the 'Rngs' functions and their links with commoly-used softwares

Category	function	Brief Description	options	Inputfile(s)	Outputfile(s)
Assembly	Assignation_MID_reads.function	Search nucleotidic tags in read sequences.	"tags" : nucleotidic tags to search / "lim": widow size (bases) to search tags	.fasta (raw data)	Assignation_MID_reads.txt
	Assignation_Contig_MIDS.function	Count the number of reads that brought each tag in the different assembled contigs	"lim": minimal nuber of reads bringing each tag (within each contig) to be accepted for statistics / "lim2: minimum number of tags per contig accepted for statistics	Assignation_MID_reads.txt / info_contigreadlist.txt (MIRA output)	Contigs_Reads_Tags.txt / Contigs_Tags.txt
Blast & Annotation	Go_hierarchy.function & Go_hierarchy2. function	Format Gene ontology database into matrix		Gene_ontology_database.txt (freely available in the Gene Ontology Website)	Go_hierarchy.txt and Go_hierarchy2.txt
	BP_split.function	Extract annotation fom B2G, explore B2G database to find all has part relatioships between GO-terms / splits information into a binary matrix	"nContigs": the total nuber of assembled contigs contained in B2G project / "project_id" : project identifiant (character)/ "selected_BL" : biological level to use / "threshold" : blast e-value threshold to keep both blast and annotation for the contig	Go_hierarchy2.txt / Blast2Go_mapping.txt (edited by B2G)	Contig_Annot_id.txt / Contig_Annot_names.txt / Contigs_annot_BL_splitted.txt
	Contaminant_detection1.function	Extract a non-redundant list of species from blast results		Blast2Go_mapping.txt (edited by B2G)	Species_list.txt / Hit_species.txt
	Contaminant_detection2.function	Search contaminant contigs		Hit_species.txt / Tree.phy	Contaminant_by_contig.txt
SNP detection	ACEsplit	Split .ace into contig sequence and quality files		.fasta (raw sequence data with tags clipped) / .qual (raw quality data with tags clipped) / .ace (assembly)	*.txt (.ace splitted into assembled sequence and quality matrices)
	Quality_anaysis	Analyzed the distribution of base quality in the assembly		*.txt	Quality_by_contig.txt
	Ascleaning	Assembly cleaning	man : minimum allele number / maf: minimum allele frequency / mq: minimum quality	*.txt	*_cleaned.txt (cleaned assembled sequence and quality matrices)
	Consensus	Consensus edition (IUPAC code)		*_cleaned.txt	consensus_with_masking.txt
	SNP_stats	SNP statistics	min_dist : minimum distance to another SNP / min_depth : minimum depth	*_cleaned.txt	SNP_stats.txt
	SNP_design	SNP design	nBase_Before: & nBase_After : numbe of bases in the flanking region	consensus_with_masking.txt / SNP_stats.txt	SNP_design.txt
	SNP_viewer	SNP vizualization by a PCA based on several statistics		SNP_stats.txt	.tiff (figure)
	Contig_depth	estimate min, max, average contig depth and depth at each base of the contigs	min_depth : minimum depth	*_cleaned.txt	Contig_depths.txt

4. Detailed description of the functions

'Assignation_MID_reads.function':

Raw data from NSG are commonly delivered into .fasta and .fasta.qual file formats (or fastq that combined both .fasta and .fastq information) containing the sequences and base qualities of each read. In NGS, identifying individuals or DNA pools of populations requires the addition of nucleotide tags to DNA libraries. This function allows identifying which tag is brought by each read, by searching for strict nucleotide patterns in the beginning of each read sequence (contained in the .fasta file). The window size for tag detection is defined by the user, and may vary from the strict length of the searched tags to the whole read. Then, reads assembly requires the prior clipping of tags that may disturb reads assembly. Looking at the number of reads bringing an identified tag across different window size may help deciding how many bases to clip: when increasing window size does not affect the number of reads carrying an identified tag, applying longer clips would contribute to loss a part of 'true sequence' contained in the reads and to reduce reads length inappropriately.

'Assignation_contig_MID.function':

This function counts the number of reads that brought each tag in the different contigs. This function was developed to deal with the output files provided by MIRA assembler (Chevreux *et al.* 2004) but will be extended shortly to deal with several other assemblers. This function merges the "read-tags association matrix" edited by the 'Assignation_MID_reads.function' with the "contig-reads association matrix" edited by MIRA (named "infos_contigreadlist.txt"). The function exports two .txt files: "Contigs_Reads_Tags.txt" contains full information about the tags brought by each read within each contig, and "Contigs_Tags.txt" summarized the number of reads that brought each tag within each contig.

'Go_hierarchy.function' , 'Go_hierarchy2. Function' & 'BP_split.function'

'Go_hierarchy' and 'Go_hierarchy2.' functions allow formatting Gene-ontology databases (that contains all GO-terms with their *has part* relationships) into a tab-delimited matrix named 'Go_hierarchy.txt'.

The function 'BP_split' extracts annotation results from B2G mapping table (named ('Blast2Go_mapping.txt'), and searches for all GO-terms associated for all biological levels

in the matrix-formatted Gene Ontology database. Two tables summarize which GO-terms are associated to each contig at different biological level: one contains GO-term identifiers ('Contig_Annot_id.txt') and one contains GO-term names ('Contig_Annot_names.txt'). The function also allows splitting annotation for a given biological level into binary matrix (0/1) with contigs in lines and biological processes for a given level in columns. Several options need to be specified: the total number of contigs contained in the B2G project, the project identifier, the desired biological level of analysis, the blast e-value threshold to use.

'Contaminant_detection1.function' & 'Contaminant_detection2.function'

'Contaminant detection 1' extracts a non-redundant list of species from the B2G mapping table (named 'Blast2Go_mapping.txt') that contains the species to which the different blasted sequences belong to ('Species_list.txt'). This non-redundant list of species may be imported into 'NCBI Common Tree Browser' (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi) to search phylogenetic relationship between the species. In particular, exporting only nodes corresponding to "green plants" allow automatizing research for possible contaminants based on blast results (to export in .phylip file named 'Tree.phy'). The function edits a matrix containing the list of species associated to each contigs ('Hit_species.txt').

'Contaminant detection 2' identified probable contaminant contigs by identifying contigs for which any of the top blast result corresponds to a green plant. It exports a matrix that resume if each contig is a contaminant or no.

'ACEsplit.function'

.ace file format is a universal format for assembled data from NGS delivered by the most popular assemblers. It contains numerous information about the assembled contigs, in particular:

- The list of reads belonging to each contig, including:
    - Reads sequence
    - Read position within the assembled contig (beginning and end)
    - Information about bases masked by the assembler (not taken into account during assembly)
- Contig consensus sequences
- Contig consensus base qualities (average base quality for each base of the contig).

However, the .ace file format is difficult to use with this format. Moreover, the .ace file excludes individual base qualities (the quality of each base of each read composing each contig). This information is, however, of major importance when searching for SNPs due to higher base calling error rates in NGS than in classical Sanger sequencing. Thus, base quality of each read need to be recovered from raw quality data contained in (tags clipped) .fasta.qual file. This function splits .ace and re-assembles contigs into matrix, and attributes base quality scores for each base of each read. The results are edited in text files stored in a directory name 'Raw data', with two text files by contig: one contains the sequence matrix of assembled reads, and one contains quality scores matrix.

'AScleaning.function'

Prior to SNP detection, assembled contig sequences and qualities need to be cleaned to minimize 'false-SNP' discovery. 'AScleaning' allows cleaning assemblies based on several statstics including singletons, minimum allele frequency, and minimum base quality. The function screens all mismatches of the assembly, and follows a simple procedure:

(1) The function masks alternative bases composed by a single read (singletons)
(2) The function masks alternative bases if their allele frequency is lower than a defined threshold (for example 0.1)
(3) The function masks all bases with a quality lower than a defined threshold (for example 20)
(4) Last, the function remove bases (matrix columns) containing only indel ('-') and masked bases ('N'). This final step reduces contig length.

'SNPsearch.function', 'Consensus.function', 'SNP_design.function' & 'SNP_viewer

Both 'SNPsearch' and 'Consensus.function' screen each base of the contigs.

'Consensus.function' edits the consensus sequence of the 'cleaned' contigs (with IUPAC codes).

'SNPsearch' searches for SNPs in the cleaned assembly and summarizes them by several statistics (Contig, SNP position, variants, absolute and relative allele frequencies, minimum allele frequency, maximum allele frequency, depth at the base, and distance to another SNP).

These two functions lead to two other functions:

- 'SNP_viewer.function' allows the visualization of SNPs through a principal component analysis (PCA) based on several statistics
- SNP_design.function' designs SNPs for submitting them for high-throughput genotyping. SNPs variants are indicated under brackets and separated by "/".SNPs flanking regions are designed based on IUPAC consensus edited by the 'Consensus' function. An option allows defining the length of flanking regions to design. SNPs with insufficient flanking regions (within ends of contig) are automatically discarded. Another option allows avoiding SNPs close to another by specifying the desired minimum distance (bases) between two targeted SNPs.

<u>Other supplementary functions</u>

Two other functions are also available:

'Quality_anaysis.function' analyzes the distribution of bases quality in the whole assembly before cleaning. In particular, it may help assessing the global quality of sequencing.

'Contig_depth.function' analyses the minimum, maximum, and average contig depths, as well as the depth at each base of each contig.

Article n°3 - High-throughput transcriptome sequencing and polymorphism discovery in four Neotropical tree species.

Louise BROUSSEAU[†§], Alexandra TINAUT[§], Carline DURET[§], Tiang LANG[§§], Pauline GARNIER-GERE[£], and Ivan SCOTTI[§*]

[§] INRA, UMR 0745 EcoFoG, Campus agronomique BP 709, F-97387 KOUROU Cedex, France

[†] INRA, UMR 1137 EEF, allée de l'Arboretum, 54280 CHAMPENOUX, France

[§§] Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Mengla, Yunnan, 666303, China

[£] INRA, UMR1202 BIOGECO, F-33610, Cestas, France; Univ. Bordeaux, BIOGECO, UMR1202, F-33400, Talence, France

Abstract

The Amazonian rainforest is predicted to suffer from ongoing environmental changes. Despite the need to evaluate the impact of such changes on tree genetic diversity, we almost entirely lack genetic resources. In this study, we analysed the transcriptome of four tropical tree species (*Carapa guianensis*, *Eperua falcata*, *Symphonia globulifera* and *Virola michelii*) with contrasting ecological features, belonging to four widespread botanical families (respectively Meliaceae, Fabaceae, Clusiaceae and Myristicaceae). We sequenced cDNA libraries from three organs (leaves, stems, and roots) using 454 pyrosequencing. We have developed an R and bioperl-based bioinformatic procedure for de novo assembly, gene functional annotation and marker discovery. SNP discovery takes into account single-base quality values as well as the likelihood of false polymorphism as a function of contig depth and number of sequenced chromosomes. Between 17103 (for *Symphonia globulifera*) and 23390 (for *Eperua falcata*) contigs were assembled. We identified between 6885 (for *Symphonia globulifera*) and 12878 (for *Virola surinamensis*) high-quality SNPs. The resulting overall SNP density was comprised between 1.3 (*C. guianensis*) and 1.46 (*V. surinamensis*) SNP/100bp. These newly identified polymorphisms are a first step towards acquiring much needed genomic resources for tropical tree species.

Introduction

The Amazonian rainforest of Northern South America hosts one of the greatest pools of terrestrial biodiversity, including very large tree species diversity (Hubbel *et al.* 2008, Hoorn *et al.* 2010, Hawkins *et al.* 2011). In forest genetics, most efforts have so far focused on temperate and boreal tree species. While ongoing anthropogenic climate change is suspected to deeply affect the stability of Neotropical rainforests (Phillips *et al.* 2009), tropical tree species genetic resources and adaptive potential are still poorly known (Savolainen *et al.* 2007), although data for at least some species are now available (Audigeos *et al.* 2010, Audigeos *et al.* 2013). Identification of polymorphisms and robust estimates of tropical tree species' standing genetic diversity are thus needed to evaluate the vulnerability to environmental changes of populations and their ability to endure them (Jump *et al.* 2008, Scotti 2010).

A thorough assessment of tropical tree species' genetic diversity requires large amounts of genomic data and informative molecular markers (Aitken *et al.* 2008, Stapley *et al.* 2010). Single-nucleotide polymorphisms (SNPs) have become the most popular genome-wide genetic markers (Seeb *et al.* 2011) and are increasingly used to characterize potentially adaptive genetic variation (e.g. Schlotterer 2002, Eveno *et al.* 2008, Eckert *et al.* 2010).

High-throughput sequencing and genotyping methods are paving the way to genomic studies in non-model species (Ellegren 2008, Allendorf *et al.* 2010, Seeb *et al.* 2011). Indeed, advances in next-generation sequencing (NGS) techniques allow cost-effective parallel sequencing of millions of sequences and are now an efficient route for generating very large genetic data collections. Thus, NGS provides a valuable starting point for identifying molecular markers in non-model species (Hayes *et al.* 2007, Seeb *et al.* 2011).

While assembling whole-genome sequence reads without a reference sequence can be very complex and in the best cases incomplete, transcriptome sequencing constitutes an efficient alternative in information-poor organisms since it avoids dealing with a large amount of repetitive sequences (usually outside the transcriptome; Pop & Salzberg 2008). Transcriptomes also include a large number of loci with known or predictable functions (Bouck & Todd 2006, Emrich *et al.* 2007) and have been applied to comparative genomics (Vera *et al.* 2008), marker discovery (Novaes *et al.* 2008), and population genomic studies (Namroud *et al.* 2008).

Among common NGS techniques, the Roche 454-pyrosequencing technology is the one producing on average the longest reads (Wicker *et al.* 2006, Weber *et al.* 2007, Emrich *et*

al. 2007), which makes de novo assembly easier in non-model species without prior genomic resources (e.g. in Eucalyptus grandis (Novaes *et al.* 2008), in Cucurbita pepo (Blanca *et al.* 2011) and in Maruca vitrata (Margam *et al.* 2011)). This technique also permits to identify allelic variants by aligning assembled ESTs from different haplotypes (Barbazuk *et al.* 2007) and is commonly used for transcriptome analysis (gene expression profiling by mRNA identification and quantification; Morozova & Marco 2008).

We describe the transcriptome and its polymorphism in four widespread Neotropical tree genera chosen to represent different botanical families, ecological properties and patterns of local and range distribution (see Materials and Methods).

The objectives of the present study are (i) to describe the transcriptomes of these four tropical genera, (ii) to compare expression profiles among organs (leaves, stems and roots), and (iii) to identify expressed single nucleotide polymorphisms (SNPs).

Material and methods

Study species and sampling

The four species studied (*Symphonia globulifera* (L. f.) (Clusiaceae); *Virola surinamensis* ((Rol. ex Rottb.) Warb.); *Carapa guianensis* (Aubl.) (Meliaceae); *Eperua falcata* (Aubl.) (Fabaceae)) are characterized by contrasted ecological requirements, spatial structure and seed dispersal strategies (**Table 1**). For each species, we collected about ten seeds from three different sampling sites: Paracou (5°16'20''N; 52°55'32''E) for *E. falcata* and *V. surinamensis*, Matiti (5°3'30''N; -52°36'17''E) for *S. globulifera*, and Rorota (4°51'32''N; -52°21'37''E) for *C. guianensis*. Seeds germinated and grew during twelve months in a greenhouse under non-limiting light and water conditions as described in Baraloto *et al.* (2007). Two vigorous seedlings of each species were selected for transcriptome analyses. Plant material was sampled from three organs: leaves, stems and roots.

Table 1: Species description.

Species name	Range	Ecology - light	Ecology-soil	Spatial population structure	Seed dispersal
Carapa guianensis	Neotropics [35]	Light-responsive [30]	Indifferent [30]	Non-aggregated [31]	gravity, rodents [36]
Eperua falcata	Guiana Shield [37]	Shade tolerant [30]	Mostly seasonally flooded [30]	Aggregated [38]	gravity [37]
Symphonia globulifera	Neotropics, Paleotropics [29]	Shade tolerant [30]	Seasonally flooded [30]	Non-aggregated [31]	gravity, vertebrates [32]
Virola surinamensis	Neotropics [33]	Light-responsive [30]	Seasonally flooded [30]	Non-aggregated [31]	large vertebrates [34]

cDNA library preparation and sequencing

Total RNA from each fresh sample was extracted using a CTAB protocol as described by Le Provost et al. (Le Provost *et al.* 2003). mRNAs were converted to double stranded cDNA using Mint cDNA synthesis kit (Evrogen) according to the manufacturer's instructions.

For each species, cDNA libraries from the different organs (leaves, stems and roots) were identified by a specific molecular identifier (MID) tag. Samples from the same organ of different conspecific individuals were pooled for sequencing (MID1 = leaves, MID2 = stems, MID3 = roots). Libraries of the different species were sequenced separately (one run per species) according to a standard Roche-454 protocol (Myer 2008). The raw data were

submitted to the ENA database (study number: PRJEB3286; http://www.ebi.ac.uk/ena/data/view/display=html&PRJEB3286) and given the accession numbers ERS177107 through to ERS177110.

Assembly and functional annotation

The bioinformatic flowchart includes the following steps (**Fig. 1**).

For each species, .sff files were extracted into .fasta, .fasta.qual and .fastq files using the '.sff extract' script available at http://bioinf.comav.upv.es/sff_extract/). The extraction was made both with and without clipping of read ends. Adaptor and MID sequences were identified in .fasta files (with unclipped ends) by searching exact motifs of MID1, MID2 and MID3 in the first twenty bases of each read.

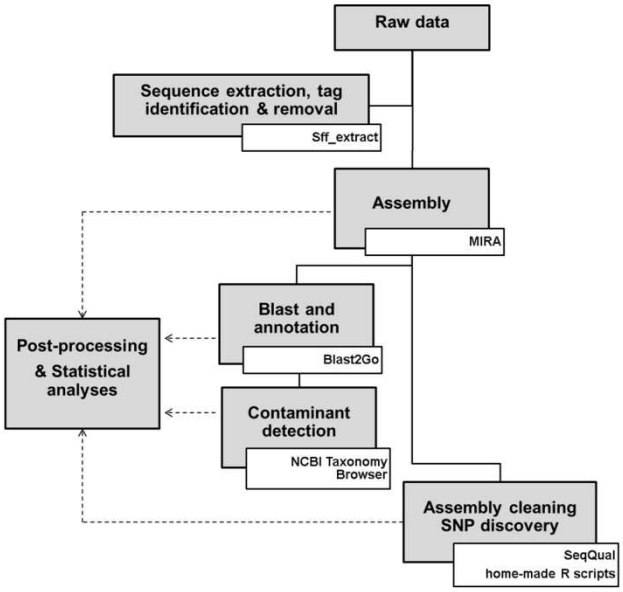


Figure 1: Bioinformatics flowchart

Reads were de novo assembled into contigs using MIRA v.3.4.0 that allows much flexibility with a large range of parameters (Chevreux *et al.* 2004) and has been used efficiently for transcriptome assemblies (Kumar & Blaxter 2010). We applied the "accurate" mode (with 'job' arguments: 'de novo, est, accurate') to limit the assembly of paralogous genes.

Assembled contig consensus sequences were submitted to Blast2Go analysis (http://www.blast2go.de/b2ghome) that allows large-scale blasting, mapping and annotation of novel sequence data particularly in non-model plant species (Conesa & Gotz 2008). Consensus showing Blast results with low e-value (10^{-25}) and valid functional annotation were submitted to public databases (in process). We realized a semi-automated search for contaminants by verifying the organism identity of each blast hit as follows: NCBI Taxonomy CommonTree Browser (http://www.ncbi.nlm.nih.gov/Taxonomy/CommonTree/wwwcmt.cgi) was searched with a non-redundant list of species extracted from B2G. Among the ten hits with the lowest e-values (below 10^{-25}), contigs having at least one hit with a sequence from a genus

belonging to the "green plant" node of the generated tree were further considered as non-contaminants, with contigs with no “green plant” genus sequence hits being treated as contaminants and excluded. Studied contigs were then assigned the best informative functional annotations from plant species hits, provided that their e-value was smaller than 10^{-25}.

The B2G analysis also allowed the matching of each contig with one or several biological processes (either “1” for an association or “0” for an absence of association with processes considered here at levels 3 and 4). Level 3 processes were ranked by their number of contigs and the cumulative distribution of the number of contigs was inspected. Moreover, considering that a contig’s number of reads is a rough estimator of the level of expression of the corresponding gene (Weber *et al.* 2007, Torres *et al.* 2008, Frias-Lopez *et al.* 2008, Craft *et al.* 2010), we used the number of reads belonging to contigs associated to particular level 4 biological processes versus all other processes to identify processes with remarkable expression levels in the different organs considered in each species. These particular processes were compared qualitatively among organs in a second step. To identify those processes, we used a permutation analysis as follows:

(1) For each organ, and for each level 4 biological process, both the number of reads per contig was recorded, and its association (“1”) or not (“0”) for that particular process.

(2) The observed average number of reads across all contigs associated to this biological process was computed; and this statistic was considered as an estimator of the average expression level of all genes involved in that biological process (contigs with zero counts were excluded).

(3) Then, the number of reads per contig was permuted at random 1000 times among contigs, under the assumption that there was no particular association (“1”) to the targeted process. At each permutation, the average number of reads per contig associated to the targeted biological process was computed again.

(4) The thousand averages obtained by permutation provided a null distribution of average read counts per contig within that biological process.

(5) If, for that biological process, the observed average read count per contig was larger than 95% of the average values obtained by permutation, then the group of genes associated to that biological process was considered as over-expressed, and consequently the biological process was considered functionally important for that organ.

Because a contig may be associated to different biological processes, steps (ii)-(v)

above were performed for each biological process separately, acknowledging the fact that some of these tests are therefore not all independent.

SNP discovery

Assemblies were post-processed using both bioperl scripts from the SeqQual pipeline (Lang et al. in preparation), and home-made R scripts that followed various steps of filtering the data by integrating a number of quality criteria. Both the SeqQual and R scripts are available on request from the authors. The different steps of the procedure used were as follows:

- *Splitting .ace assembly files and linking to quality*

Assembled contig sequence files were extracted from the .ace files given by MIRA and linked to their original base quality scores contained in the .fasta.qual files

- *Assembly cleaning*

Nucleotide differences were screened in assembled contigs and particular bases were masked according to several criterions:

- being a singleton
- being a variant with a frequency lower than 0.1 (see also 'Computing SNP statistics and post-filtering' below).
- having a quality value (PHRED score equivalent) lower than 20 for polymorphic sites (i.e. incorrect base call probability of 1/100).

Following this 'masking step', a 'cleaning step' removed all positions (i.e. corresponding to one base) of the assembled contigs that contained only indel and masked bases. This last step is particularly relevant for 454 data where false insertions due to homopolymers were very common and drastically affect contig consensus, hampering further re-sequencing and SNP design for genotyping. Consensus (using IUPAC codes) were edited from cleaned assembled data and used both for estimating the total transcriptome length obtained and for identifying quality SNPs for submission to databases.

- *Computing SNP statistics and post-filtering*

All potential SNP contained in the cleaned assemblies were used to build a summary statistics table (number of occurrences and frequency of the different variants, depth, mean quality, minimum allele frequency (maf)). This table was used to identify the highest quality SNPs a posteriori (without affecting assembly and consensus) for further SNP

design and larger scale genotyping. In particular, we chose to avoid:

- SNPs adjacent to each other, because they are likely to be assembly artefacts (You *et al.* 2011).

- SNPs with lower-than-expected frequencies based on the number of gametes sequenced. With two genotypes, four different gametes were sequenced and the probability of having a variant was 0.25 at minimum. The following rationale can be applied to any number of gametes N. The probability of observing a particular number of times (or fewer) the minority variant (1/2N) follows a binomial distribution. The probability of observing the variant exactly t times out of x reads is computed as $p(t) = \binom{x}{t} p^t q^{x-t}$ and the probability of observing it t times or fewer is given by $\sum_0^t p(t)$. All polymorphisms that were present in a configuration with a cumulative probability P < 0.05 (e.g. 3 variants among 29 reads) were considered as false positives and were discarded. In roughly half the case, these configurations had already been excluded based on the below 0.1 frequency rule (see 'Assembly cleaning'). In the other half of the cases, additional configurations where variant frequencies ranged from 0.1 to 0.15 but which had a probability below 5% could also be excluded, therefore increasing the overall likelihood of the detected variants.

- SNPs having a depth lower than 8X, which can be considered as a stringent criteria, given the 20 quality score for each base, a minimum SNP frequency of 2/8= 0.25 here (since singletons have been previously excluded), and the fact that this configuration has a probability of 0.31 based on the binomial distribution rationale, which is well above the 5% threshold chosen before.

Following the filtering steps described above, SNPs were counted and their density per base was computed as the total number of polymorphisms (including SNPs at contig ends that passed the quality and singleton filters) divided by the total number of bases where the depth was at least 8 reads. Numbers of transitions, transversions, and deletions were also reported.

Results and Discussion

Assembly

Sequence data were obtained from all tissues and species except *S. globulifera*, for which root cDNA library preparation failed. Between 167140 and 248145 reads were obtained per species. More reads were associated with roots than with stems or leaves (**Table 2**). This is likely due to technical artefacts such as a more efficient RNA extraction and/or cDNA amplification from roots than from other organs, and a lower RNA extraction yield in leaves due to high concentrations of secondary metabolites.

Table 2: Partitioning of reads among different organs (leaves, stems, roots) in each species cDNA library (*C. guianensis*, *E. falcata*, *S. globulifera* and V.surinamensis). Under brakets, the number of assembled reads.

Number of reads	***Carapa guianensis***	***Eperua falcata***	***Symphonia globulifera***	***Virola surinamensis***
From leaves [MID1]	63016 [43334 (28%)]	17421 [11417 (9%)]	49894 [32190 (30%)]	31526 [22077 (11%)]
From stems [MID2]	47100 [29720 (19%)]	28362 [18088 (14%)]	110373 [66874 (66%)]	41435 [28284 (14%)]
From roots [MID3]	132030 [77052 (50%)]	175551 [100909 (76%)]	7 [2 (0%)]	141948 [89918 (72%)]
No of reads without tag	5999 [3435 (2%)]	3260 [1799 (1%)]	6866 [4367 (4%)]	4314 [2691 (2%)]

Between 103433 (*S. globulifera*) and 153551 (*C. guianensis*) reads were successfully assembled into contigs and between 17103 and 23390 contigs were obtained, depending on the species (**Table 3**). These figures are close to the average number of contigs commonly obtained in similar studies (Kumar & Blaxter 2010, Blanca *et al.* 2011, Sloan *et al.* 2012) and suggest reasonable transcriptome coverage from the data if we assume that the number of contigs approximates the species' unigenes. Average contig length varied between 414 bp (*E. falcata*) and 523 bp (*C. guianensis*) (**Supplementary Fig. S2**).

Table 3: Assembly results.

	Carapa guianensis	***Eperua falcata***	***Symphonia globulifera***	***Virola surinamensis***
Number of reads	248145	224554	167140	219223
Number of assembled reads	153551 (61.9%)	132213 (58.9%)	103433 (61.9%)	142970 (65.2%)
Number of contigs	21770	23390	17103	21070
Total length (bp)	11393209	9688583	7743116	9725915
Average length per contig (bp)	523.3445	414.219	452.734	461.6001
Average number of reads per contig	7.05333	5.65244	6.047	6.785447

A large number of contigs was solely associated to roots for the three species (**Fig. 2**), particularly in E. falcata (61% of contigs from roots only, compared to 29% and 37% for *C. guianensis* and *V. surinamensis*). This probably resulted from the predominance of root-tagged reads (MID3, **Table 2**). In contrast, contigs exclusive to stems and leaves were in much lower proportions in the three species with root data, varying from 4% to 7% for stems, and 3% to 12% for leaves (**Fig. 2**).

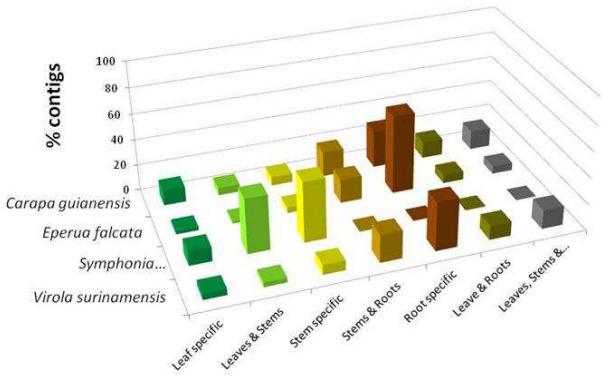


Figure 2: Number of contigs associated with each organ (leaves, stems, roots) (Note: sequencing from *S. globulifera* roots failed). *Carapa* = *Carapa guianensis*; *Eperua* = *Eperua falcata*; *Symphonia* = *Symphonia globulifera*; *Virola* = *Virola surinamensis*. L, S and R indicate contigs specific to Leaf, Stem and Root, respectively; combinations of symbols correspond to contigs occurring in multiple organs.

Functional annotation

Functional annotation based on BlastX and gene ontology analyses allowed classifying contigs into biological processes. A majority of contigs returned a Blast hit result with e-values below 10-25 (**suppl. Fig. S3**) for *C. guianensis* (79%), *E. falcata* (69%), *S. globulifera* (74%) and *V. surinamensis* (70%), but only between 48.1% (*E. falcata*) and 64.1% (*C. guianensis*) had functionally informative annotations (**Table 4**). Less than 3.1% of the characterized contigs were identified as contaminants for any species (1.58%, 3.06%, 2.92% and 0.29% in *C. guianensis*, *E. falcata*, *V. surinamensis* and *S. globulifera* respectively). After removing contaminants, from 12603 (*S. globulifera*) to 16912 unigenes (*C. guianensis*) with an e-value lower than 10^{-25} were retained, that covered 4.65 Mbp (in S. globulifera) to 7.75 Mbp (in *C. guianensis*) of the total transcriptome across species (**Table 4**).

The analysis of the cumulative distribution of contigs associated to each biological process (level 3) revealed that 50% of the contigs belonged to around 12% of the same biological processes across species (**Fig. 3**). This corresponds to 8 biological processes (out of 70 for C. guianensis and *V. surinamensis*) and 9 biological processes (out of 73 and 75 for *E. falcata* and *S. globulifera* respectively).

Table 4: BlastX statistics across species, performed on consensus sequences from MIRA.

	Carapa guianensis	***Eperua falcata***	***Symphonia globulifera***	***Virola surinamensis***
No of sequences that did not return any blast result	4586 (21.1%)	7231(30.9%)	4463 (26.1%)	6384 (30.3%)
No of blasted sequences [No sequences after contaminant removal]	17184 (78.9%) [16912]	16159 (69.1%) [15664]	12640 (73.9%) [12603]	14686 (69.7%) [14545]
No of mapped sequences	15879 (72.9%)	13629 (56.3%)	11639 (68.1%)	13000 (61.7%)
No of annotated sequences	13962 (64.1%)	11240 (48.1%)	10164 (59.4%)	11073 (52.6%)
Total assembly length without contaminant (bp) [Total length of blasted sequences after removal of contaminant and sequences with e-values>10^{-25}]	11266552 [7746737]	9501561 [4789056]	7728777 [4748202]	9666680 [5887279]

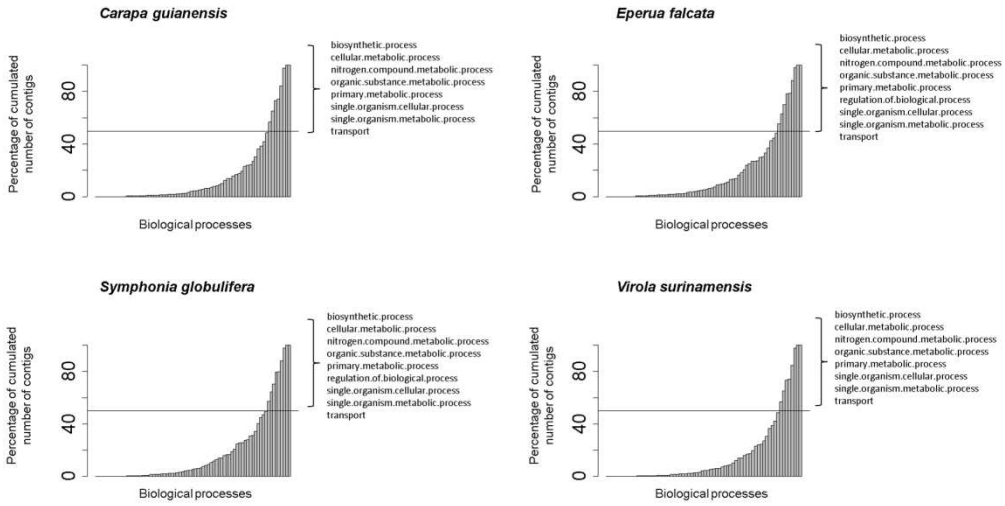


Figure 3: Cumulative percentage of contigs annotated by biological process (level 3). Only non-contaminant contigs with an e-value lower or equal to 10^{-25} were retained for the analysis.

Permutation analyses allowed us to identify biological processes (level 4) showing a significantly higher occurrence of contigs for a given organ, that could be interpreted as a higher expression of genes belonging to that process in that organ (**Fig. 4**).

In leaves, between five (*V. surinamensis*) and ten (*C. guianensis*) biological processes stood out (**Fig. 4** left column), eight of them being identified more than one species. Not surprisingly, biological processes related to photosynthesis and carbon cycle in leaves appear in this group ('carbohydrate metabolic process' , 'carbon fixation', 'generation of precursor metabolites and energy', 'nitrogen cycle metabolic process', 'organic substance biosynthetic process', 'oxidation reduction process', 'photosynthesis', 'response to radiation').

In stems, we detected between eight (*S. globulifera*) and twenty-five (*V. surinamensis*) relevant biological processes (**Fig.4** middle column), fifteen of them being shared among different species. At least a subset of these processes ('cellular biosynthetic process', 'cellular component movement', 'organic substance biosynthetic process', 'organic substance catabolic process', 'secondary metabolic process') are potentially related to cell differentiation events that occur during wood formation.

In roots, between seven (*C. guianensis*) and twenty-six (*E. falcata*) biological processes appeared as particularly relevant, eleven being shared by different species. They reflect two main functions of roots: water and nutrient acquisition ('response to inorganic substance', 'response to 'organic substance transmembrane transport') and response to stresses caused by soil constraints, which fall in two classes: (a) soil water depletion (e.g.

'response to osmotic stress') which frequently occurs in tropical rainforests during the dry season; (b) oxidative stresses caused by soil hypoxia, to which the processes 'reactive oxygen species metabolic process', 'response to oxidative stress', and 'response to oxygen containing compound' are related; flooding-induced hypoxia is particularly frequent in water-logged bottomlands.

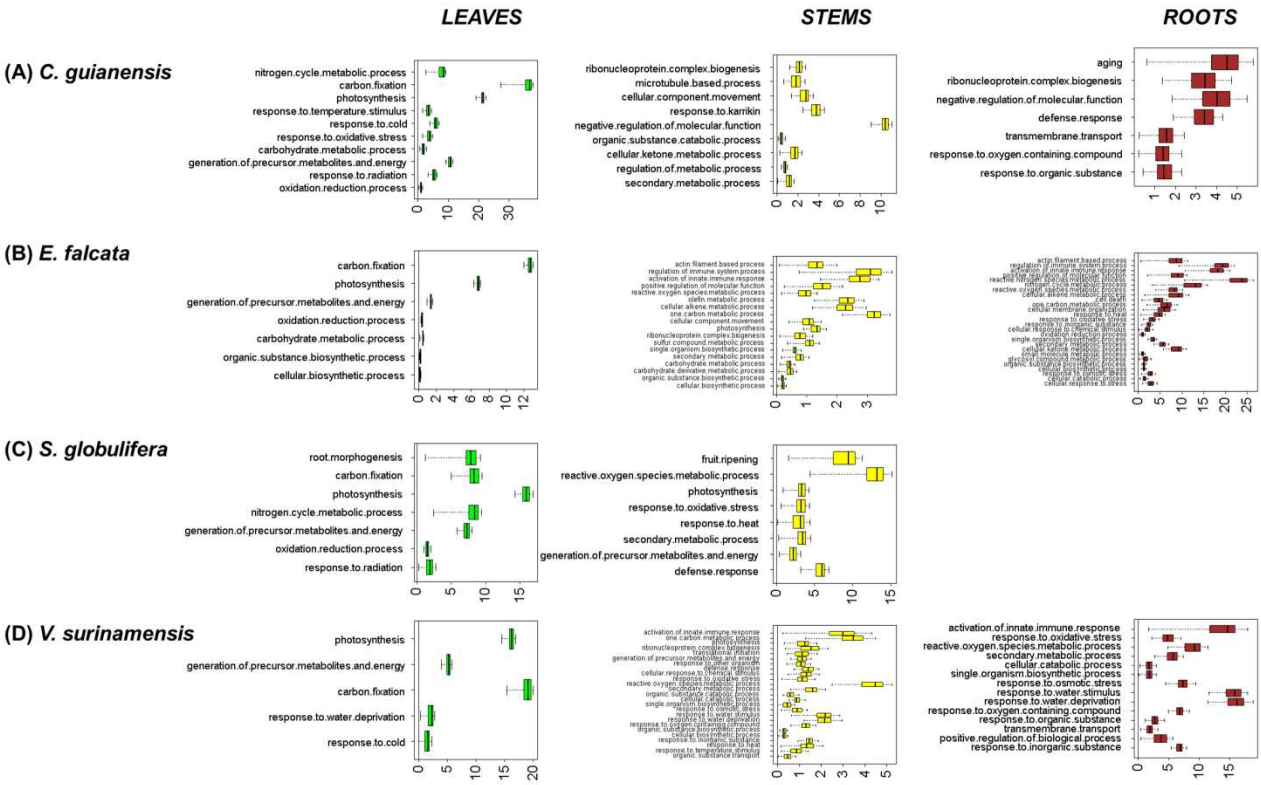


Figure 4: Differences between observed and randomized mean contig number of reads within each organ from 1000 permutations of the number of reads per contig across all contigs for each biological processes (level 4) showing significant over-expression in each organ.

(A) *C. guianensis*: Leaves (nitrogen cycle metabolic process, carbon fixation, photosynthesis, response to temperature stimulus, response to cold, response to oxidative stress, carbohydrate metabolic process, generation of precursor metabolites and energy, response to radiation, oxidation reduction process); Stems (ribonucleoprotein complex biogenesis, microtubule based process, cellular component movement, response to karrikin, negative regulation of molecular function, organic substance catabolic process, cellular ketone metabolic process, regulation of metabolic process, secondary metabolic process), Roots (aging, ribonucleoprotein complex biogenesis, negative regulation of molecular function, defense response, transmembrane transport, response to oxygen containing compound, response to organic substance);
(B) *E. falcata*: Leaves (carbon fixation, photosynthesis, generation of precursor metabolites and energy, oxidation reduction process, carbohydrate metabolic process, organic substance biosynthetic process, cellular biosynthetic process), Stems (actin filament based process, regulation of immune system process, regulation of innate immune response , activation of innate immune response, positive regulation of molecular function, reactive oxygen species metabolic process, olefin metabolic process, cellular alkene metabolic process, one carbon metabolic process, cellular component movement, photosynthesis, ribonucleoprotein complex biogenesis, sulfur compound metabolic process, single organism biosynthetic process, secondary metabolic process, carbohydrate metabolic process, carbohydrate derivative metabolic process, organic substance biosynthetic process, cellular biosynthetic process), Roots (actin filament based process, regulation of immune system process, activation of innate immune response, positive regulation of molecular function, reactive nitrogen species metabolic process, nitrogen cycle metabolic process, reactive oxygen species metabolic process, cellular alkene metabolic process, cell death, one carbon metabolic process, cellular membrane organization, response to heat, response to oxidative stress, response to inorganic substance, cellular response to chemical stimulus, oxidation reduction process, single organism biosynthetic process, secondary metabolic process, cellular ketone metabolic process, small molecule metabolic process, glycosyl compound metabolic process, organic substance biosynthetic process, cellular biosynthetic process, response to osmotic stress, cellular catabolic process, cellular response to stress);
(C) *S. globulifera*: Leaves(root morphogenesis, carbon fixation, photosynthesis, nitrogen cycle metabolic process, generation of precursor metabolites and energy, oxidation reduction process, response to radiation), Stems (fruit ripening, reactive oxygen species metabolic process, photosynthesis, response to oxidative stress, response to heat, secondary metabolic process, generation of precursor metabolites and energy, defense response);
(D) *V. surinamensis*: Leaves (photosynthesis, generation of precursor metabolites and energy, carbon fixation, response to water deprivation, response to cold), Stems (activation of innate immune response, one carbon metabolic process, photosynthesis, ribonucleoprotein complex biogenesis,translational initiation, generation of precursor metabolites and energy, response to other organism, defense response, cellular response to chemical stimulus, response to oxidative stress, reactive oxygen species metabolic process, secondary metabolic process, organic substance catabolic process, cellular catabolic process, single organism biosynthetic process, response to osmotic stress, response to water stimulus, response to water deprivation, response to oxygen containing compound, organic substance biosynthetic process, cellular biosynthetic process, response to inorganic substance, response to heat, response to temperature stimulus, organic substance transport), Roots (activation of innate immune response, response to oxidative stress, reactive oxygen species metabolic process, secondary metabolic process, cellular catabolic process, single organism biosynthetic process ,response to osmotic stress, response to water stimulus, response to water deprivation, response to oxygen containing compound, response to organic substance, transmembrane transport, positive regulation of biological process, response to inorganic substance);
(Note: sequencing from *S. globulifera* roots failed).

rRNA intron-encoded homing endonucleases were very abundant in the *E. falcata* assembly (581 unigenes against 43, 39 and 17 unigenes in *C. guianensis*, *S. globulifera* and *V. surniamensis* respectively). In *E. falcata*, these unigenes comprised between two and 920 reads with a mean of 15.3 (s.d.=69.77). Among them, fourteen had more than 100 reads, and 74 had between 10 and 100 reads.

Homing endonucleases from group I introns are self-splicing genetic elements or parasitic genes mostly found in organellar genomes (Cho *et al.* 1998, Burt & Koufopanou 2004). Among contigs that showed BLAST hits with rRNA-intron-encoded homing endonucleases in E. falcata, 69 were polymorphic and contained from 1 to 18 SNPs with many haplotypes (Yahara *et al.* 2009). High transcription levels of such elements, combined with the high numbers of mutations that they have accumulated, suggests a massive but ancient genome invasion event (Yahara *et al.* 2009, Nystedt *et al.* 2013) in the *E. falcata* genome compared to the other three species. The evolutionary implications of transfers of such elements remain poorly understood, because of their 'super-Mendelian' inheritance (such elements may be both vertically and horizontally transmitted; Koufopanou *et al.* 2002), and because they have no known function (Yahara *et al.* 2009).

SNP discovery

It has been shown that relaxed criteria for in silico SNP selection from next-generation sequencing data or previous EST databases lead to high failure rates in subsequent high-throughput SNP genotyping (Huse *et al.* 2007, De Pristo *et al.* 2011). We have applied a stringent filtering process based on data quality and a probabilistic argument in order to decrease the frequency of false SNPs. SNP depth was significantly reduced after the first masking steps: original depth was on average 1.31-1.53 times from ~20 to ~23 across species on average per contig to a final average depth of ~16 to ~17 for the retained SNPs, depending on the species (**Suppl. Fig. S1**). Between 4434 (for *S. globulifera*) and 9076 (for *V. surinamensis*) were retained after all the filtering steps had been applied (**Table 5**). Between

Table 5: SNP detection.

	Carapa guianensis	*Eperua falcata*	*Symphonia globulifera*	*Virola surinamensis*
Total length with depth ≥ 8X after assembly cleaning (bases)	956876	603897	499694	862357
Before Post-filtering based on Binomial test				
N SNPs	**10615**	**7084**	**5447**	**10897**
N polymorphic contigs	1716 (7.88%)	1299 (5.55%)	987 (5.77%)	1752 (8.32%)
SNP density (/100bp)	1.11	1.17	1.09	1.26
N polymorphisms with 2 variants	10420 (98.16%)	6968 (98.36%)	5362 (98.44%)	10757 (98.72%)
N transitions	2655	1875	1090	2182
N transversions	1699	1155	779	1474
Ti/Tv	1.56	1.62	1.40	1.48
N indel	6066	3938	3493	7101
N polymorphisms > 2 variants	195 (1.84%)	116 (1.64%)	85 (1.56%)	140 (1.28%)
After Post-filtering based on Binomial test				
N SNPs	**8646**	**5713**	**4434**	**9076**
N polymorphic contigs	1706 (7.83%)	1283 (5.5%)	979 (5.72%)	1746 (8.29%)
SNP density (/100bp)	0.90	0.95	0.89	1.05
N polymorphisms with 2 variants	8534 (95.70%)	5649 (98.89%)	4388 (98.96%)	8981 (98.95%)
N transitions	2380	1657	989	1989
N transversions	1488	1000	681	1310
Ti/Tv	1.60	1.66	1.45	1.52
N indel	4666	2992	2718	5682
N polymorphisms > 2 variants	112 (1.13%)	64 (1.12%)	46 (1.04%)	95 (1.05%)

5.5% (E. falcata) and 8.3% (*V. surinamensis*) of contigs contained at least one SNP (**Table 5**).The great majority of polymorphisms (between 95.7% in *C. guianensis* and 99% in *S. globulifera*) were bi-allelic, with a majority of indels (**Fig. 5**). The transition/transversion ratio (Ti/Tv) varied between 1.5 and 1.7, lower than those observed in other exome assemblies (De Pristo *et al.* 2011). Estimated SNP density across polymorphic contigs varied between 0.89 per 100 bp (*C. guianensis*) and 1.05 per 100 bp (*V. surinamensis*) (**Table 5**). However, these estimates are based on only four haploid genomes per species, so the overall species-wide molecular diversity is likely to be larger. These estimates are, however, in the same order of magnitude as observed in other studies: Parchman *et al.* (2010) reported between 0.6 to 1.1 SNPs per 100 bp in *Pinus taeda*, depending on the stringency of their filtering criteria.

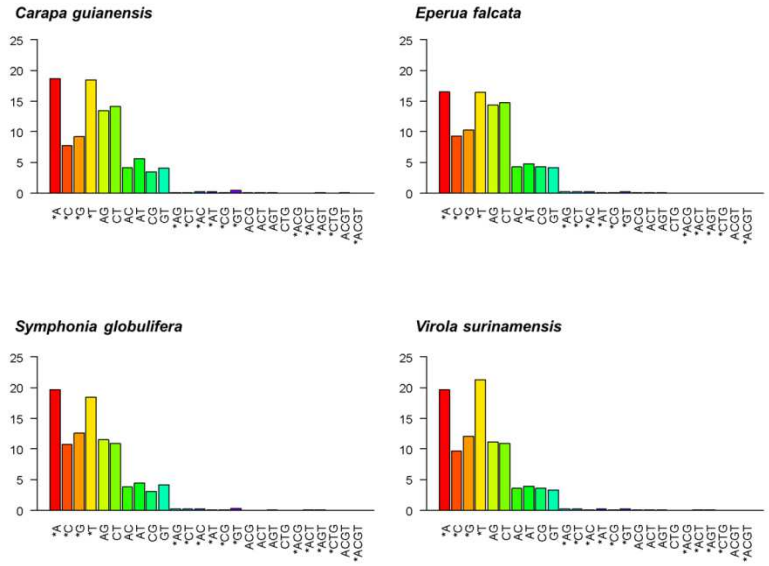


Figure 5: Proportion of SNPs depending on allelic patterns ('*' indicates deletions).

Transcriptome polymorphism and its usefulness in population genetics studies

Next-generation sequencing, allowing massive de novo acquisition of molecular data, provides a range of new potential applications for evolutionary and ecological-genetic studies in non-model species. High-throughput SNP data have indeed shown their potential for inferences about demographic and adaptive processes in natural populations (Eckert *et al.* 2010, Nielsen *et al.* 2005, Nielsen *et al.* 2009, Li & Wolfgang 2006, Siol & Barett 2010, Turner *et al.* 2010, Fournier-Level *et al.* 2011, Hancock *et al.* 2011). However, this assumes that the identified polymorphisms are of high quality, which is why we have tried to accomplish here. The genomic resources obtained here will trigger new exciting fields of research on tropical biodiversity. Providing a catalogue of putative functions for genomic regions with a high potential diversity will help identifying useful candidate genes for further resequencing or SNP genotyping (Lister *et al.* 2009, Morozova *et al.* 2009, Helyar *et al.* 2011). These genes belong to a large range of biological processes, including growth, reproduction,

light and nutrient acquisitions, as well as plant response to biotic and abiotic stresses. Focusing on genes potentially involved in adaptive processes in Neotropical forest tree species will permit to test hypotheses about evolutionary processes underlying genome evolution and the build-up of biological diversity in tropical forest ecosystems.

Acknowledgements

We thank the GENOTOUL platform who performed the sequencing. We also thank Valérie Léger (INRA, UMR 1202 BioGeCo) who helped designing and setting up the laboratory protocols. This project was supported by PO FEDER ENERGIRAVI and the MEDD-ECOFOR "Ecosystèmestropicaux" program.

Authors' contributions

IS designed the experiment. CD realized the experiment. LB wrote R scripts for bioinformatics analyses. TL and PGG wrote bioperl scripts for bioinformatics analyses and worked on adapting SeqQual for 454 data. LB, FA and AT performed bioinformatic analyses with the help of IS and PGG. LB post-processed data from bioinformatics analysis and realized statistical analyses with the help of PGG. All authors wrote the paper.

Article n°3 - Supplementary figures

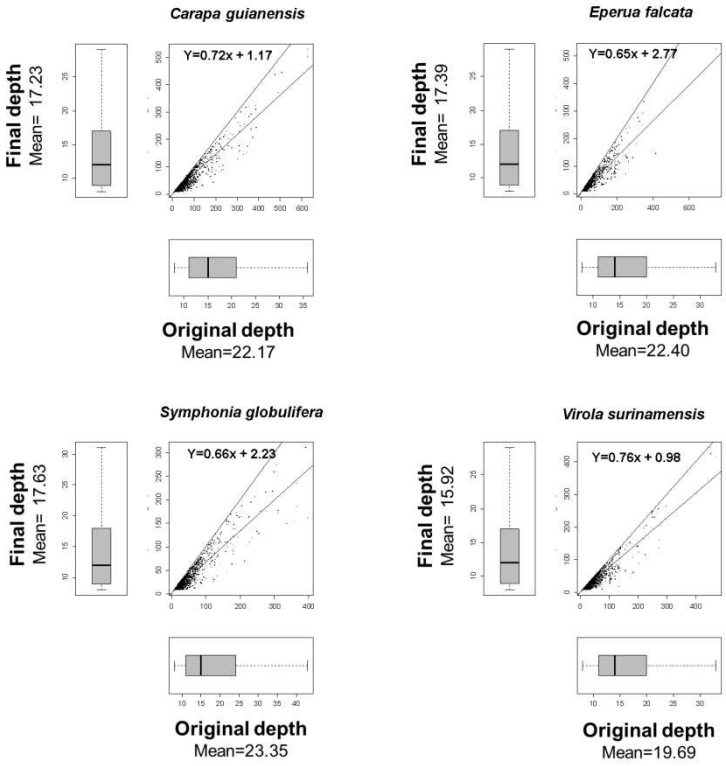


Figure S1: Representation of SNP depths before and after the masking procedure.

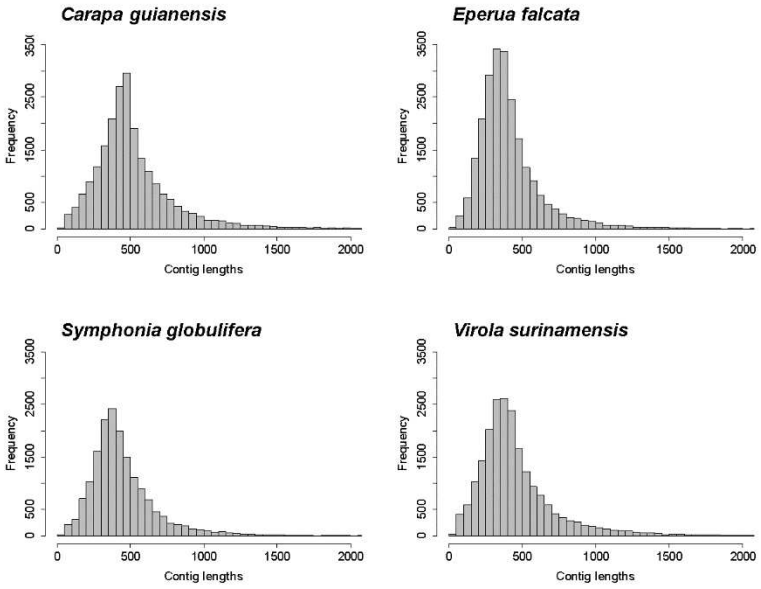


Figure S2: Distribution of contig lengths within each assembly.

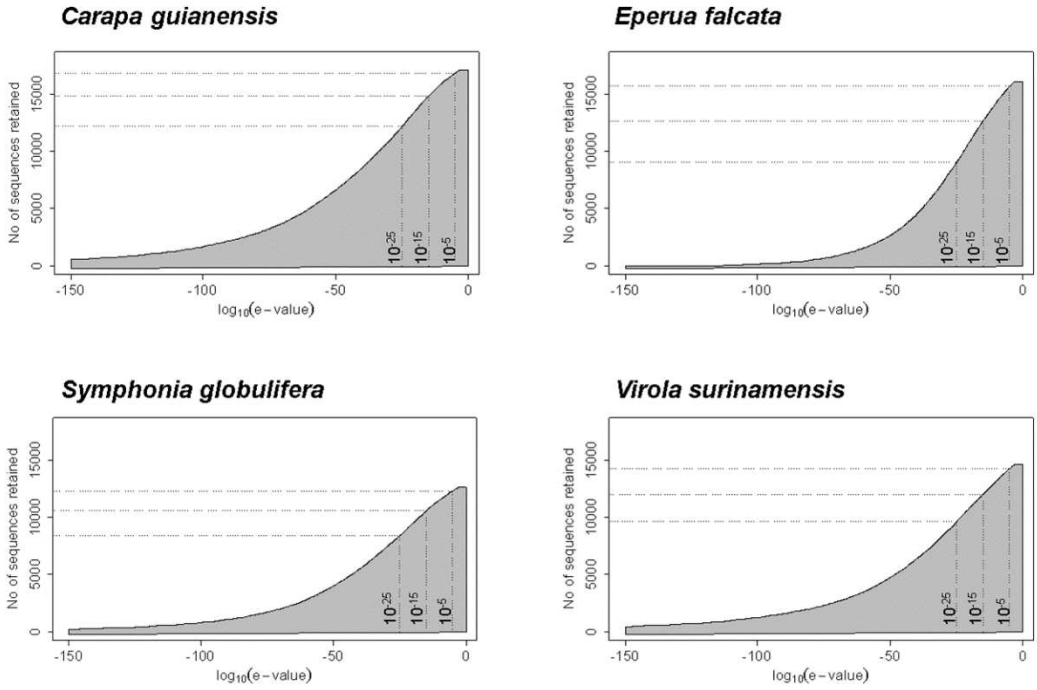


Figure S3: Number of contigs returning a blast result using different e-value thresholds: 10^{-5}, 10^{-10}, 10^{-15}, 10^{-20} and 10^{-25}.

Article n°4 - Highly local environmental variability promotes intra-population divergence of quantitative traits: an example from tropical rainforest trees

Louise BROUSSEAU[1,2], Damien BONAL[1,2,3], Jeremy CIGNA[1], Ivan SCOTTI[1]

1. INRA, UMR « Ecologie des Forêts de Guyane », Campus agronomique, BP 709, 97387 Kourou cedex, French Guiana

2. INRA, UMR « Ecologie et Ecophysiologie Forestières », 54280 Champenoux, France.

3. Université de Lorraine, UMR « Ecologie et Ecophysiologie Forestières », 54280 Champenoux, France.

Published in: Annals of Botany (2013) **112**: p 1169-1179 (doi: 10.1093/aob/mct176)

Abstract

• In habitat mosaics, plant populations face environmental heterogeneity over short geographical distances. Such steep environmental gradients can induce ecological divergence. Lowland rainforests of the Guiana Shield are characterized by sharp, short-distance environmental variations related to topography and soil characteristics (from water-logged bottomlands on hydromorphic soils to well-drained terra firme on ferralitic soils). Continuous plant populations distributed along such gradients are an interesting system to study intra-population divergence at highly local scales. In this study, we tested (a) whether conspecific populations growing in different habitats diverge at functional traits and (b) whether they diverge in the same way as congeneric species having different habitat preferences.

• We studied phenotypic differentiation within continuous populations occupying different habitats for two congeneric, sympatric, and ecologically divergent tree species (*Eperua falcata* and *E. grandiflora*, Fabaceae). Over 3000 seeds collected from three habitats were germinated and grown in a common garden experiment, and twenty-three morphological, biomass, resource allocation and physiological traits were measured.

• In both species, seedling populations native of different habitats displayed phenotypic divergence for several traits (including seedling growth, biomass allocation, leaf chemistry, photosynthesis and carbon isotope composition). This may occur through heritable genetic variation or other maternally inherited effects. For a subset of traits, the intraspecific

divergence associated with environmental variation coincided with interspecific divergence.

• Our results indicate that mother trees from different habitats transmit divergent trait values to their progeny, and suggest that local environmental variation selects for different trait optima even at a very local spatial scale. Traits for which differentiation within species follows the same pattern as differentiation between species indicate that the same ecological processes underlie intra- and interspecific variation.

Introduction

Environmental variation occurring at the local scale creates complex habitat patchiness which has been found to contribute to shaping the great diversity observed in tropical rainforests (Ricklefs, 1977, Wright 2002, Vincent *et al.* 2011). A common explanation for these diversity patterns is the divergence of preferences for edaphic conditions among tree species, as repeatedly shown throughout the Neotropics (e.g. ter Steege *et al.* (1993), Sabatier *et al.* (1997), Clark *et al.* (*1999*), Valencia *et al.* (2004), Baraloto *et al.* (2007), John *et al.* (2007), Kanagaraj et al. (2011)). Community-level differences in functional traits have been found to underlie such differences (Kraft et al., 2008): for instance, Lopez & Kursar (2003) and Engelbrecht *et al.* (2007) showed that divergence in species distribution between hilltops and bottomlands are determined by variations in tolerance to drought and water-logging.

It has been shown that bottomland, slope, and hilltop habitats actually differ in many ways that may explain their impact on forest community composition. Generally speaking, water availability in lowland tropical forests is strongly associated with topography and soil characteristics (Sabatier *et al.* 1997). Large variations occur in soil drainage and moisture between hilltops, slopes and bottomlands (Clark *et al.* 1999, ter Steege *et al.* 1993, Webb and Peart 2000). Bottomland soils are subject to frequent periods of flooding and undergo cyclical changes in O_2 availability that strongly affect the metabolism of root tissues and thus tree establishment and growth (Kozlowski 1997, Perata *et al.* 2011, Ponnamperuma 1972). In contrast, thin soils on slopes undergo lateral drainage, which increases their susceptibility to water shortage during dry periods (Sabatier *et al.* 1997). Finally, hilltops are usually characterized by deep soils and display deep vertical drainage, with strong seasonal variations in soil water availability (Sabatier *et al.* 1997). Beside differences in water availability constraints, these habitats also differ in nutrient content, with lower nitrogen and higher phosphorus content in bottomlands than on plateaus (Ferry *et al.* 2010, Luizao *et al.* 2004). Moreover, soil respiration decreases in bottomlands as root biomass and soil carbon content decreases (Epron *et al.* 2006). These variations in soil characteristics have an additional impact on forest dynamics, with slopes and bottomlands exhibiting more frequent light gaps than hilltops and therefore higher irradiance reaching the understory (Ferry *et al.* 2010).

The widespread links between gradients of soil properties and species-specific habitat preferences suggest that ecological specialisation has recurrently arisen through

evolutionary processes such as adaptation and species divergence (Endler, 1977, Schluter, 2001, Rundle & Nosil 2005, Savolainen *et al.* 2007). Evolutionary dynamics may play a major role in the build-up of lowland rainforest community diversity, and the role of genetic diversity (including sensu lato both allelic and gene expression variability) in ecological processes has been widely acknowledged (Ford 1964, Randall Hughes *et al.* 2008). In other words, if ecological sorting of functional traits has occurred across different habitats and has led to the emergence of ecologically different species, it is sensible to expect that such processes are also currently occurring within species. Therefore, in species with continuous stands growing in different, contiguous habitats, we should be able to observe "highly local" intra-specific divergence (sensu Salvaudon *et al.* 2008) between subpopulations submitted to divergent local environmental conditions; moreover, we expect that divergence between intraspecific subpopulations growing in different habitats should co-occur with divergence between species with different ecological preferences for those habitats. Here, we use the term 'highly local' to characterise patterns observed at scales for which environmental turnover occurs at shorter distances than gene flow (i.e. the average distance between patches of different habitat types is shorter than the average gene dispersal distance, implying that gene flow occurs among different habitats).

Tree populations in general are known to harbour large amounts of heritable variation for several putatively adaptive characters (Cornelius 1994, González-Martínez *et al.* 2006); Neotropical rain forest trees are no exception (Scotti et al., 2010, Navarro et al., 2004). If adaptation contributes to divergence between sub-populations occupying different habitats, these sub-populations should be differentiated at potentially adaptive traits (sensu Howe & Brunner 2005). The goal of the present study was therefore to test whether populations of tree species growing as continuous stands across different habitats could be subdivided into habitat-associated sub-populations displaying phenotypic divergence for such traits (i.e. divergence caused by differentiation in (multi-locus) gene frequencies, by maternal effects or by inheritance of stable gene expression patterns ("epigenetic inheritance")). The test was performed in two congeneric rainforest tree species of the Guiana Shield (*Eperua falcata* and *E. grandiflora*), that display partially divergent habitat preferences (Sabatier *et al.* 1997, Baraloto *et al.* 2007) but occur, even in low abundance, in multiple habitat types. In *Eperua* species, gene flow is expected to be restricted – mainly due to heavy seeds – but still intense at the distances considered here (estimate of mean parent-offspring distance for *E. grandiflora*: 166-343 m; Hardy *et al.* 2006). In spite of such dispersal

distances, a recent study, performed partly on the same populations as those studied in the present paper (Audigeos *et al.* 2013), has shown that molecular divergence occurs (in *E. falcata*) at a highly local scale for genes involved in response to soil water content-related stress, against an overall background of no genetic differentiation at other loci.

The specific questions asked in this study about phenotypic divergence in these two congeneric species are: (i) Do seedlings from different local habitats diverge phenotypically? (ii) Are patterns of intraspecific phenotypic divergence similar to those observed at the interspecific level?

Material and methods

Study species

E. falcata and *E. grandiflora* are abundant in the Guiana Shield, and grow sympatrically in different but partially overlapping habitats. This allowed us to compare intraspecific and interspecific patterns of divergence in the same phylogenetic context and ecological background. *E. falcata* (Aubl.) (Fabaceae) has a preference for seasonally water-logged bottomlands, whereas *E. grandiflora* (Aubl.) Benth (Fabaceae) is mostly restricted to hilltops and slopes (Baraloto *et al.* 2007). The two species differ in several morphological and functional traits, but their seedlings display similar degrees of tolerance to drought or hypoxia under controlled conditions (Baraloto et al., 2007), indicating that they are potential generalists for soil water conditions, at least at the younger life stages. Both species are bat-pollinated (Cowan 1975) and disperse their heavy seeds by explosive dehiscence and gravity at short distances of a few meters (Forget 1989). Gene dispersal distance is about 150-350 m for *E. grandiflora* (Hardy *et al.* 2006) and probably similar for *E. falcata* (O. Hardy, pers. comm.), well beyond the size of the habitat patches studied here. Data from nuclear genetic markers (Audigeos *et al.* 2013) suggest that *E. falcata* is allogamous with no significant selfing.

Study site

The experiment was performed in Plot 6 at the Paracou forest inventory site (5°18'N, 52°53'W) (Gourlet-Fleury *et al.* 2004) located in an undisturbed forest in coastal French Guiana, South America. The sampling area covers 9 ha and is characterized by a rugged landscape formed by the alternation of 40-50 m-high hills, slopes, and bottomlands, varying in soil drainage type and water table depth (Gourlet-Fleury *et al.* 2004). In such a habitat mosaic, variations occur on geographical distances of the same order of magnitude as pollen and seed dispersal but do not occur monotonically (i.e. there is no continuous gradient in a given spatial direction). Three habitat types have been identified in the study area (**Suppl. Fig. S1**) based on elevation, soil drainage, and water-logging characteristics (Ferry *et al.* 2010): "Bottomlands" (B) with hydromorphic soils and a water table between 0 and 60 cm in depth depending on the season (**Suppl. Fig. S1**); "Slopes" (S) with surface drainage conditions, and a water table consistently below 100 cm; "Hilltops" (H) with deep soils, deep vertical drainage and a water table consistently below 150 cm.

Seed sampling

Two-hundred and sixty-seven *E. falcata* trees and 67 *E. grandiflora* trees were identified in the study area. Operators visited the plot at least three times a week in February-March 2006, 2007 and 2008 to hand collect seeds on the ground from 44 fruiting trees. The choice of the mother tree set was based on several considerations: (a) tree fertility; (b) balanced sampling from all habitats; (c) non-overlapping tree crowns. Pairwise distances between same-habitat fruiting trees were not statistically smaller than between trees in different habitats (**Fig. 1**).

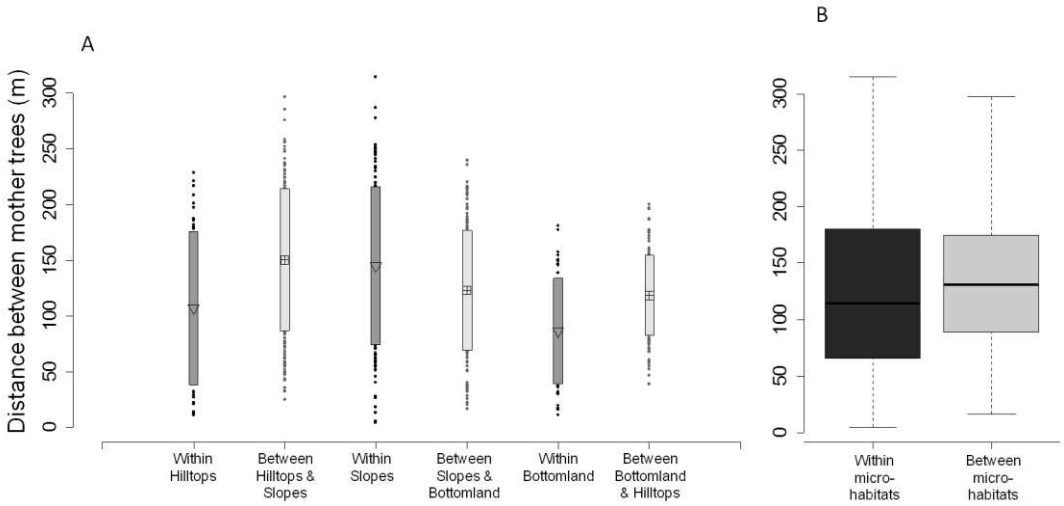


Figure 1: A) Pair-wise spatial distances between mother trees within and between micro-habitats. Boxes show the standard deviation of each group. B) Boxplots showing the distribution of pair-wise spatial distance between mother trees within and between all micro-habitats.

Seeds collected for our experimental study were assigned to the same habitat as their mother tree, thus forming three different native habitat types ("B", "S" and "H"). When crowns of conspecific trees overlapped, seeds were collected at opposite sides of the crown. Each seed was assigned to a maternal family corresponding to its mother tree. A total of 3122 seeds were collected over the three seed production years.

Glasshouse Experiment

The seeds were weighed and laid down in germination boxes that were filled with a substrate made of river sand which was kept damp using an automatic sprinkler system. Germination success rate was about 60% for both species. Two months after germination, the seedlings were transplanted into individual 12-l pots filled with a mixture of sand and an A-horizon soil (30/70 v/v), then transferred to a glasshouse. The A-horizon had been collected in the same plot as the seeds and contained about 1.4 - 1.9 g kg-1 of nitrogen (Ferry

et al. 2010).

About 4% of the seedlings died before transfer to the glasshouse. The remaining seedlings were grown in the glasshouse for 24 months, until the study ended; then they were harvested. The 1637 seedlings (**Supp. Table S1**) were randomly assigned to each of 103 16-plant blocks. Each block contained four seedlings from each of four randomly drawn maternal families, so that each family was combined randomly with a different set of other families in each of the blocks in which it was represented (see **Suppl. Method 1** for details). The seedlings were placed under non-limiting conditions, which prevented both drought and hypoxia (expected to occur in the field on hilltops/slopes and in bottomlands, respectively; see above). Moreover, seedlings grown in the glasshouse experienced higher light levels and milder competition than in natural conditions, favouring optimal growth. A layer of neutral shade-cloth was used to reduce irradiance received to about 13% of full sun (maximum photosynthetic photon flux density ≈ 300 μmol m-2 s-1) to simulate solar radiation levels received by seedlings in gap openings. Seedlings were watered 2-3 times per week to maintain the substrate close to field capacity (≈ 0.25 m3 m-3). The pots were fertilised every six months (5 g complete fertiliser per pot, 12/12/17/2 N/P/K/Mg). Pots were distributed in the glasshouse following an incomplete randomized block layout (for the details of the experimental design, see **Suppl. Method 1**).

One-thousand-six-hundred-and-thirty-seven seedlings survived until month 24. For measures taken at 24 months, the sample used in the present study was restricted to 656 seedlings of *E. falcata* and 170 seedlings of *E. grandiflora* (**Suppl. Tables S1** and S2), since two thirds of the seedlings grown in this experiment were set aside for a companion experiment involving different soil water content treatments.

Phenotypic traits

We recorded twenty-three functional traits (**Table 1**) related to plant growth, biomass allocation, leaf structure and leaf physiology (photosynthetic capacity and carbon isotope composition). These traits are commonly used as proxies of plant fitness in general (Kraft *et al.* 2008) and their ecological significance as proxies of fitness in seedlings has been established by several studies (Wright *et al.* 2004, Cornelissen *et al.* 2003, Westoby *et al.* 2002).

Table 1: List of abbreviations and units of phenotypic traits.

Growth and biomass allocation		
Seedling dimensions :		
H_6	Height at 6 months	cm
H_{12}	Height at 12 months	cm
H_{18}	Height at 18 months	cm
H_{24}	Height at 24 months	cm
H_{612}	Elongation rate from 6 to 12 months	cm.month^{-1}
H_{1218}	Elongation rate from 12 to 18 months	cm.month^{-1}
H_{1824}	Elongation rate from 18 to 24 months	cm.month^{-1}
D_{18}	Diameter at 18 months	mm
D_{24}	Diameter at 24 months	mm
D_{1824}	Radial growth rate from 18 to 24 months	mm.month^{-1}
Biomass and allocation :		
TM_{24}	Total dry mass at 24 months	g
RM_{24}	Root dry mass at 24 months	g
LM_{24}	Total leaf dry mass at 24 months	g
LA_{24}	Total leaf area at 24 months	cm^2
LMR_{24}	Leaf / total mass ratio at 24 months	g.g^{-1}
RMR_{24}	Root / total mass ratio at 24 months	g.g^{-1}
LAR_{24}	Leaf area / total biomass ratio at 24 months	cm^2.g^{-1}
Leaf traits		
LMA_{18}	Leaf mass / area ratio at 18 months	g.m^{-2}
LMA_{24}	Leaf mass / area ratio at 24 months	g.m^{-2}
$\%C_{18}$	Carbon content in leaves at 18 months	%
$\%N_{18}$	Nitrogen content in leaves 18 months	%
A_{sat}	Light-saturated carbon assimilation rate at 18 months	µmol m^{-2} s^{-1}
$\delta^{13}C$	Carbon isotope composition of leaves at 18 months	per mil

Plant height and stem diameter at collar were measured every six months. Net CO_2 assimilation rate under saturating irradiance (Asat , µmol m-2 s-1) was recorded in vivo at 18 months on one leaf per plant with a portable photosynthesis system (CIRAS1, PP-Systems, Hoddesdon, UK) operating in open mode and fitted with a Parkinson leaf cuvette, under the following microclimate: ambient air CO_2 concentration = 380 µmol mol-1; photosynthetic photon flux density = 600 ± 20 µmol m−2 s−1; vapour pressure deficit = 1.0 ± 0.5 kPa; ambient air temperature = 28.7 ± 2.0 °C. Full stabilization was obtained after about 3-5 minutes. Measurements were conducted between 9:00 am and 1:00 pm to avoid mid-day depression of photosynthesis. After gas exchange measurements, six to eight mature and fully expanded leaflets were collected per plant close to the top of the stem. Fresh leaf area was then measured in the laboratory with an area meter (Li-2100, Licor, Lincoln, Nebraska). The leaves were subsequently dried to constant weight at 60°C for about three days, then finely ground to measure carbon (C) and nitrogen (N) content and carbon isotope composition (δ13C, ‰) as a surrogate for intrinsic water-use efficiency (WUEi; Farquhar et al., 1982). Elemental and isotopic analyses were conducted on a sub-sample of about 1 mg of dry leaf powder with an isotopic ratio spectrometer (Delta-S Finnigan Mat, Bremen, Germany). Leaf mass to area ratio (LMA, g m-2) was calculated as the ratio of dry mass to leaf area.

At 24 months, all the plants were harvested and the leaves, stems, and roots were separated for biomass measurements. Total leaf area was measured with the same area meter as above. All three compartments were dried at 60°C to constant weight for about 3-4 days and then weighed. Leaf area to total biomass ratio (LAR, m² g-1) was obtained by dividing the total leaf area of a given plant (LA) by its total dry weight. Leaf mass ratio

(LMR, g g-1) and root mass ratio (RMR, g g-1) were calculated as the ratio of leaf or root dry mass to total plant dry mass (**Table 1**). Growth rates for height and diameter growth between two dates were calculated as $\Delta P/\Delta t=(Pt_2-Pt_1)/(t_2-t_1)$, where P indicates the phenotypic value and t1, t2 the times of the two different measurements.

Linear model of character variation

We fitted a classical linear model for the partition of individual phenotypic values, including species, native habitat, maternal family, year of seed collection and seed mass as sources of trait variation in a hierarchical framework. To produce unbiased estimates of progeny and native habitat type effects, inter-annual variation and seed mass effects were used as cofactors in the model, as they capture, at least partially, environmental effects mediated by maternal allocation to seeds, and thus represent "maternal effects" related to resource availability (Rice *et al.* 1993, Leiva and Fernández-Alés 1998, González-Rodríguez *et al.* 2012). Our hierarchical framework allowed us to estimate the effects of each habitat type for each species, and the effect of each maternal family in each native habitat and each species. The linear model for all traits is as follows:

$$Yijklm = \mu + \alpha j + \beta k + \gamma kl + \tau klm + (\varphi k \times \text{Seed mass}i) + \varepsilon ijklm \quad (1)$$

where Yijklm is the phenotypic value of the i-th individual, μ the global mean, αj the effect of the j-th year of seed sampling and cultivation, βk the effect of the k-th species, γkl the effect of the l-th native habitat type within the k-th species, τklm the effect of the m-th progeny within the l-th native habitat within the k-th species, φk the regression coefficient between trait value and seed mass in the k-th species, seed massi is the fresh mass of the i-th seed and $\varepsilon ijklm$ the residual variation of the i-th individual.

Model parameters and effects were estimated in a Bayesian framework (see **Suppl. Methods 2** for details) using the WINBUGS ® software (Lunn *et al.* 2000). Bayesian methods can easily accommodate for unbalanced / incomplete experimental designs (Browne & Draper 2006) (erratic seed output (**Suppl. Tables S1 and S2**) made a balanced design impossible in our study).

Conventional hypothesis testing of the significance of effects can be performed using the 95% posterior distribution of effects (Song 2007). In this context, credible intervals are treated as the Bayesian analogs of confidence intervals: an estimated parameter has 95% of chance to be within the credible interval (Ellison 1996): parameters for which zero falls outside the credible interval are considered significantly different from zero. The statistical

consequences of multiple testing were evaluated by computing the Bayesian analogue of False Discovery Rate (FDR; Benjamini & Hochberg, 1995, Miranda-Moreno *et al.* 2007).

Bayesian estimation of maternal family variance effects

We computed the ratio of maternal family variance (which include truly genetic, epigenetic and possibly non-genetic maternal effects, and which we summarise as σ^2M) to total phenotypic variance (σ^2P). To estimate variances, we used a reduced version of linear model (1) restricted to family variations within each species. Phenotypic values were broken down as follows:

$Yijm = \mu + \alpha j + \tau m + \varepsilon ijm$ (2),

and the ratio of maternal family variance to total variance was estimated as:

$\sigma^2M / \sigma^2P = \sigma^2\tau / \sigma^2Y$.

Maternal family effects were estimated by fitting a quantitative-genetic hierarchical model by a Bayesian inference method of variance partitioning (**Suppl. Methods 3**). This simplified model was preferred to the full model to compute variance components because (i) it is designed to directly estimate variance components, thus saving computation time and (ii) the maternal family-level component (σ^2M) we wished to obtain included all sources of among-family variation, including habitat, but not include species effects (each species is treated separately).

Phenotypic correlations between traits

We estimated phenotypic correlations both at the individual (seedling) and at the maternal family level, using observed individual phenotypic values of seedlings and Bayesian estimates of maternal family values, respectively. The latter were computed as the sum of all sensu lato 'genetic' factors from Equation (1): $Y'klm = \mu + \beta k + \gamma kl + \tau klm$.. The sum of these factors conveys the mean phenotypic value of each progeny free from seed mass and year effects (which represent 'environmental maternal effects'). Phenotypic correlations were calculated using Pearson's coefficient. Significance at two-tailed $\alpha = 0.05$ was tested by the cor.test function in R (R Development Core Team 2008). False Discovery Rate (FDR; Benjamini & Hochberg 1995) was computed for all correlation matrices.

Results

At the intra-specific level, native habitat had a significant effect on eighteen out of twenty-three traits in *E. falcata*, and fifteen out of twenty-three in *E. grandiflora* (**Fig. 2 and Suppl. Tables S3 to S6**). Both species displayed significant variation among native habitats for growth traits (including height and diameter; height and diameter growth rates; total, root and leaf mass): seedlings from bottomlands grew faster and produced more biomass than those from slopes and hilltops. Growth rate varied significantly among native habitats at early stages in both species, but this effect vanished after twelve and eighteen months for *E. grandiflora* and *E. falcata* respectively. In both species, δ13C, leaf area and leaf mass were larger, and N content smaller, in seedlings from bottomlands than from the other two habitats. *E. falcata* seedlings from bottomlands showed lower LAR, but higher LMA, than those from slopes and hilltops. For E. *grandiflora*, Asat was higher in bottomland seedlings than in those from hilltops. We did not find any significant variation in RMR among native habitats. We estimated the expected rate of false positives (false discovery rate) as 0.8% with a single test alpha = 5% as used here.

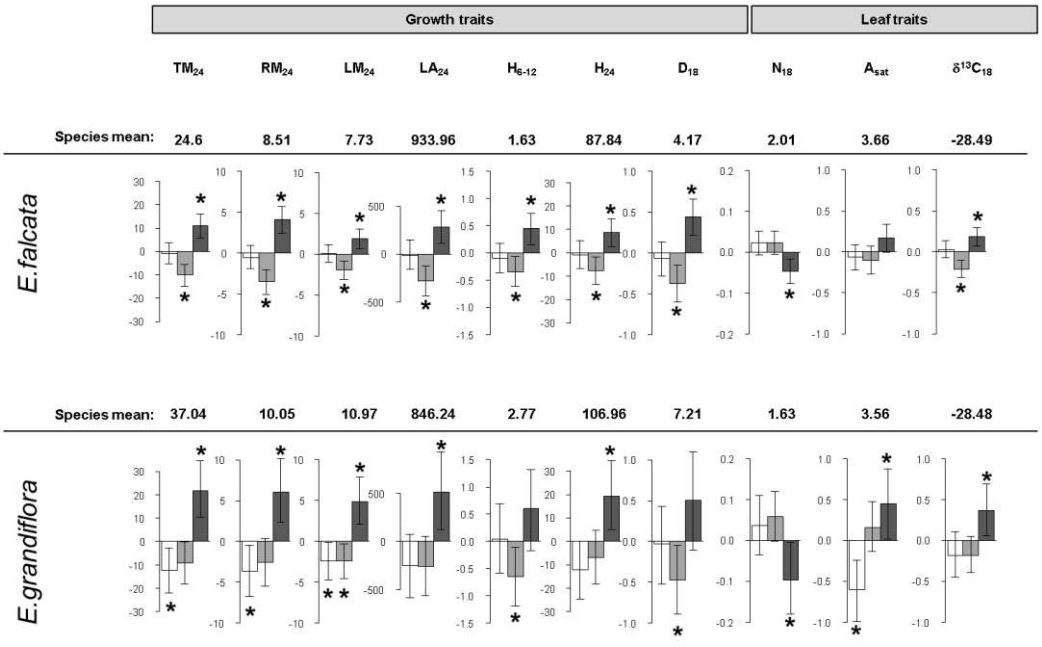


Figure 2: Phenotypic differentiation among habitat types for growth, biomass allocation and leaf traits for two species (Upper pane: Eperua falcata; lower pane: Eperua grandiflora) sampled at Paracou, French Guiana. Bayesian estimates of departures of each group from the global mean are shown as boxes; t-bars show the 95% Bayesian credible interval of the estimated parameters. Figures above each plot provide the within-species trait means, which correspond to the zero value in the plots. Units for each trait are provided in Table 1. For each plot: left box: hilltop; middle box: slope; right box: bottomland. Stars indicate a significant effect of habitat type.

The two species displayed significant differences for a subset of the recorded traits (**Suppl. Tables S3 to S6**): *E. falcata* seedlings had significantly smaller stems, higher LA and LMR, higher %N and lower LMA than *E. grandiflora*. No difference was detected for growth rate, biomass accumulation, Asat or δ13C. Nine traits (LMR24, LAR24, D18, D24, H6, H12, H18, LMA24 and %N18; **Suppl. Tables S3-S6**) had significant differences at both the intra- and interspecific level. For these traits, intraspecific trends ran contrary to the interspecific ones (**Fig. 3, Suppl. Tables S3-S6**): that is, the overall direction of change between same-species hilltop and bottomland subpopulations was contrary to the change between hilltop-preferring *E. grandiflora* and bottomland-preferring *E. falcata*. None of the traits showing significant differences among hilltop and bottomland subpopulations also showed significant differences in the same direction between hilltop-preferring *E. grandiflora* and bottomland-preferring *E. falcata*. Four traits (RMR24, D1824, H1218 and Asat; **Suppl. Tables S3-S6**) showed such a trend, but for none of them were effects significant both at the species and at the subpopulation level. Cofactors representing maternally transmitted environmental effects (year of fruit set and seed mass) also influenced several traits (**Suppl. Tables S3 to S6**).

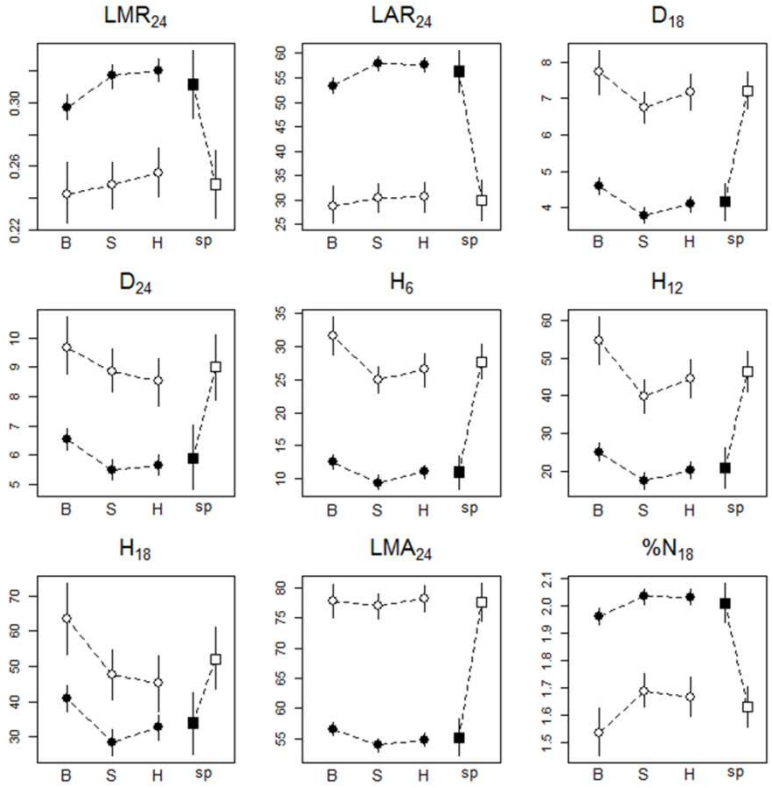


Figure 3: Comparison of the direction of trait value change between habitats (within species) and between species for traits with significant differences both at the intraspecific and at the interspecific level (see Supplementary Tables S3-S6 for raw results). Trait names and units as described in Table 1. B = bottomlands; S = slopes; H = hilltops; sp = species-level values. Black symbols: Bayesian posterior medians for Eperua falcata; white symbols: Bayesian posterior medians for Eperua grandiflora. Vertical lines: Bayesian 95% credible intervals. Non-overlapping credible intervals between two values imply significant differences.

The maternal family effect (which is obtained independently from native habitat effect described above) was significant for all traits in both species (**Suppl. Tables S3 to S6**). Ratios of maternal family-to-total variance (σ^2M / σ^2P)

ranged between 1.2 % (H1824; 95% credible interval (c.i.0.95) = 0.007-7.2 %) and 10.1 % (H18; c.i.0.95 = 4.9-20.5 %) in *E. falcata*, and from 0.02% (LMA18; c.i.0.95 = 0.00003-2.61 %) to 25.4 % (LA24; c.i.0.95 = 6-58 %) in *E. grandiflora* (**Fig. 4 and Suppl. Table S7**). Credible intervals were larger in *E. grandiflora* than in *E. falcata* (**Suppl. Table S7**) probably due to differences in sample size (**Suppl. Table S1**).

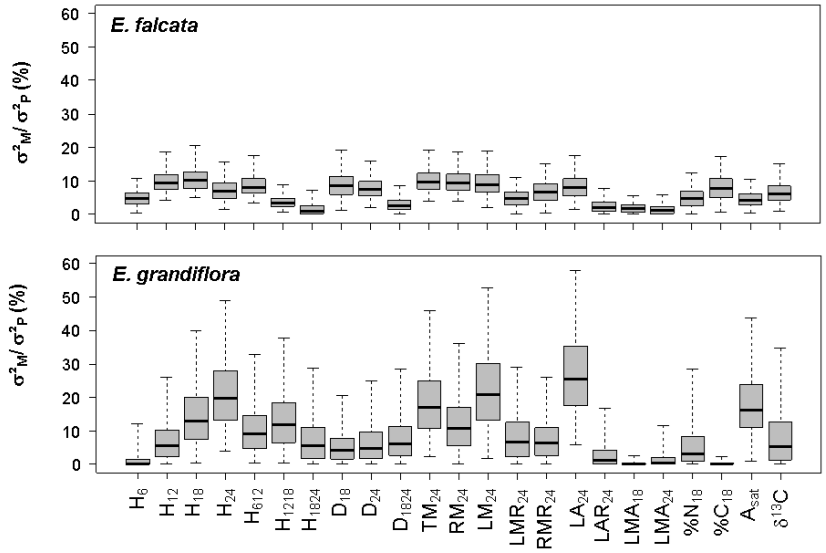


Figure 4: Boxplots of Bayesian posteriors of σ^2M/σ^2P for all traits for Eperua falcata (upper pane) and Eperua grandiflora (lower pane). Trait names as described in Table 1.

Correlation matrices were very similar between the two species (**Suppl. Fig. S3 and S4**). Most traits showed significant correlation at the individual seedling level (raw phenotypic data), but not at the maternal family level (capturing maternally inherited effects on traits). Seedling-level and progeny-level trait correlation matrices, if both significant, always had the same sign; we did not observe any significant family-level correlation without matching significant seedling-level correlation. At the seedling level, two main correlation groups emerged: dimensions, biomass and leaf traits (**Table 1 and Suppl. Fig. S3 and S4**) were tightly correlated; allocation traits were all negatively correlated with the remaining traits and had a mixed pattern of correlation to each other. Leaf mass per area (LMA) was somewhat intermediate, showing both positive and negative correlations with dimension, leaf and biomass traits and positive correlation with RMR. At the maternal family level, traits such as Asat and δ13C retained their positive correlation with biomass traits (but not with dimension traits) and their negative correlation with allocation traits; the latter globally retained their negative correlation with all other traits and the positive correlation between LAR and LMR (although fewer correlations were significant in *E. grandiflora* than in *E. falcata*). The FDR was smaller than 2% for all matrices for both the 5% and the 1% significance threshold (**Suppl. Fig. S3 and S4**).

Discussion

Divergence among sub-populations and maternal families was apparent for several traits, indicating the presence of maternally inherited variability in both species, in agreement with existing estimates of quantitative trait diversity in wild tree populations (Cornelius 1994, Scotti *et al.* 2010, Coutand *et al.* 2010).

After removal of environmentally derived maternal effects (as described by seed mass and year of fructification), native habitat explained a significant fraction of phenotypic differentiation for several leaf- or plant-level traits. These effects are relatively small (**Suppl. Fig. S2**) but significant, which is quite surprising, considering the small spatial scale at which they occur. A subset of these traits may show divergence between sub-populations only because they are correlated with traits that are involved in some adaptively meaningful divergence (Lande & Arnold 1983). The analysis of phenotypic correlations at the progeny level actually reveals that twenty of the thirty-three traits (61%) showing some degree of divergence are correlated to at least another divergent trait. Because maternal family level correlations were estimated on mean maternal family phenotypic values (which do not include seed mass and year-of-production effects), the correlations between traits is likely driven by several factors (including epigenetic effects, pleiotropy, and physical QTL linkage), which we cannot break apart with the current data set.

Nine traits (**Fig. 3**) displayed divergence both between species and between subpopulations within species. For all these traits, the intraspecific patterns ran opposite to the interspecific one. This suggests that intraspecific trait distributions may be unimodal functions of environmental variables with peak positions that differ between species ("reaction norm shift": Figure 3; figure 5 in Albert et al., 2010, Crispo, 2007). In such conditions, if the span of environmental conditions sampled is limited relative to the extent of such unimodal distributions, one may observe the kind of patterns reported here, with intraspecific trends contrary to interspecific ones (Albert *et al.* 2010). Four additional traits (RMR24, D1824, H1218 and Amax18; **Suppl. Tables S3-S6**) had monotonic intraspecific trends that were concordant with interspecific ones, but without significant effects at either the species or the population level, or both. These results show that, at least for a relatively large subset of traits (9 out of 23, or 39%), it is possible to detect intraspecific variation for those traits showing interspecific variation along the same environmental gradients. This is in agreement with the hypothesis that the differentiation processes currently affecting within-population diversity may be the same as the ones that caused species divergence,

although our observations require confirmation by functional-ecological experiments.

The maternally transmitted component of both trait divergence and trait correlations may have multiple origins:

(a) Environmentally driven maternal effects (i.e. variation in resource availability transmitted to seedlings through seed resources) can influence seedling growth (González-Rodríguez *et al.* 2011, González-Rodríguez *et al.* 2012); in our study, these were controlled through modelling of the effect of both seed mass and year of seed set, which are estimated separately from maternal family effect; therefore we suggest that these effects should be negligible in our estimation of sensu lato genetic factors, although some cases of maternal background X environmental effects have been reported (Rice *et al.* 1993, González-Rodríguez *et al.* 2011).

(b) "Epigenetic" maternal effects (mainly due to the transient transmission of gene expression states through the embryo) can contribute to similarity of traits within maternal families, thus inflating maternal family effects. Epigenetic inheritance has been proven to occur in trees (Rix *et al.* 2012), although its overall impact on trait variance was negligible. It is not possible to estimate the importance of such effects in our study, and they can clearly contribute to trait divergence among maternal families from different native habitats, if mother trees transmit environmentally induced gene expression states to their progeny. These variations in epigenetic state may have an adaptive meaning, if epigenetically inherited trait values confer higher fitness in the maternal habitat.

(c) Truly heritable (additive and non-additive) genetic effects may also contribute to trait divergence, and also have an adaptive meaning, for the same reasons as in (b). Two arguments let us think that "truly genetic" effects may account for at least part of the observed divergence between sub-populations. First, the same *E. falcata* adult tree population used for the present study displayed molecular-genetic divergence between habitats for genes involved in response to stresses related to soil water content (Audigeos *et al.* 2013); this supports the possibility that genetic structuring can occur in these populations. Secondly, we have shown that there are significant phenotypic differences between maternal families within habitats. If habitat-driven differentiation were only caused by epigenetic effects related to environmental differences, variation between same-habitat maternal families should be negligible, which is not the case in our results. Traits that had large maternal family variance components (σ^2M/σ^2P) in our study (e.g. height and biomass traits; leaf area; **Fig. 4**) often also showed high heritability in other tropical or

temperate tree species (Vásquez & Dvorak 1996, Hodge *et al.* 2002, Carnegie *et al.* 2004, Navarro *et al.* 2004, Scotti-Saintagne *et al.* 2004, Costa e Silva *et al.* 2005, Sotelo-Montes *et al.* 2007, Callister & Collins 2008, Ward *et al.* 2008, Scotti *et al.* 2010), suggesting that a non negligible part of the phenotypic divergence among maternal families may be due to true genetic factors; it has to be noted that heritability estimates are generally obtained at the species or at the whole-population level, without considerations for environmental subdivision, and therefore our σ^2M/σ^2P estimates are properly comparable to previous studies. Finally, it has been proven that plant populations can show genetic divergence at functional traits even if they are potentially connected by migration (Hovenden & Vander Schoor 2004, Byars *et al.* 2007) or have been shown to undergo strong gene flow (Gonzalo-Turpin & Hazard 2009).

Whatever the mechanistic base of phenotypic divergence between sub-populations from different native habitats, how likely is it that these differences have arisen because of neutral processes, e.g. to spatial genetic structure (due to local inbreeding)? Our study plot is a 300m-sided square, and the largest possible distance between trees is approximately 425 m, within *Eperua* gene dispersal distance (Hardy *et al.* 2006); gene flow is thus possible between the different habitat types. Moreover, seeds were sampled in a habitat mosaic, and mother trees inhabiting a same habitat type are not on average closer than trees inhabiting different habitats (**Fig. 1**). Thus, neutral divergence induced by neutral spatial genetic structure seems unlikely.

Several studies on plants have shown divergence in adaptive traits along environmental gradients (Kawecki & Ebert 2004, Carlson *et al.* 2011), particularly with respect to edaphic factors and water-logging conditions (Silva *et al.* 2010). The existence of sensu lato heritable traits showing highly local divergence between sub-populations suggests that local adaptation at short geographical distances may occur (Ehrlich & Raven 1969, Schemske 1984, Jump *et al.* 2006, Turner *et al.* 2010) in presence of gene flow, which is precisely the sense given by Kawecki (2004) to the term "local adaptation". Conditions for highly local adaptation are not unlikely in tropical rainforests, based on evidence about local species distribution (ter Steege & Hammond, 2001) and the association between functional traits and habitats (Baraloto *et al.* 2005) over short spatial scales (< 50 meters) (Kraft *et al.* 2010).

Functional considerations can help the interpretation of the observed differences among seedlings native from different habitats. A higher productivity of seedlings from

bottomlands as compared to the other two habitats is consistent with larger leaf area and higher Asat, since these seedlings are therefore able to assimilate more carbon, use it to synthesise more biomass, and eventually allocate it to growth. This is consistent with the results of previous studies revealing a trend towards increasing growth performances from drier to wetter habitats (Russo *et al.* 2005, Kariuki *et al.* 2006, Sanchez-Gomez *et al.* 2006, Ferry *et al.* 2010). LMR and LAR were slightly lower in *E. falcata* seedlings from bottomlands, suggesting that they invest more biomass in roots and stems than in leaves. This is consistent with frequent water-logging events that drastically reduce O_2 availability in the soil and decrease hydraulic conductivity of roots, with consequences similar to those of drought (Ponnamperuma 1972). Lower LMR and LAR would also contribute to reducing water loss through a lower leaf area per unit of plant mass (Poorter & Markesteijn 2008). Higher LMA in bottomland seedlings also permits a reduction of water loss through the reduction of transpiring leaf area at the leaf level (Poorter *et al.* 2009). In parallel, higher investment in root biomass would enhance water capture ability during dry periods as well as root O_2 absorption during wet periods. Furthermore, bottomland seedlings of both species display higher water use efficiency (i.e. less negative $\delta 13C$) than slope or hilltop seedlings, which means that, during photosynthesis, they use less water for the same amount of CO_2 assimilation (Farquhar *et al.* 1982). This trade-off in water and carbon use at leaf level is an efficient strategy when soil water resources are limiting (e. g. Ehleringer & Cooper 1988) not only on hilltops but also in the bottomlands (Baraloto *et al.* 2007). Finally, variations of N content are well identified as a determinant of photosynthetic capacities (Reich *et al.* 1994), as revealed by the strong correlations between leaf nitrogen and Asat. In natural conditions, leaf nitrogen and foliar N:P ratios are known to be highly dependent upon soil chemical properties (Townsend *et al.* 2007), and the dependence of Amax to N is expected to be stronger in N-limiting habitats than in P- or Ca- limiting habitats. Bottomlands have higher N content and lower P content than hilltop habitats (Ferry *et al.* 2010, Luizao *et al.* 2004), and we observe here lower %N in bottomland than in hilltop seedlings. This suggests that the faster-growing bottomland seedlings, which also have higher photosynthetic rates, have lower nitrogen content, contrary to what is expected – at the interspecific level – according to the World Leaf Economic Spectrum (Donovan *et al.* 2011).

Conclusion

We detected phenotypic divergence for growth and physiological traits occurring over very short spatial distances within a habitat mosaic. This suggests that large reservoirs of within-species adaptive potential are maintained by trait filtering caused by niche partitioning and habitat associations (Russo *et al.* 2005, Kraft *et al.* 2010), and possibly by local adaptive processes. Species displaying such variation may respond more easily to environmental changes through micro-evolution (by being able to react adaptively to the expected impact of global change), if at least part of the variation is heritable or is caused by adaptive plasticity. It is worth remembering that epigenetic (maternal) effects can be considered as heritable in the broad-sense (Klironomos *et al.* 2013, Bossdorf *et al.* 2008). The mechanisms underlying such local intra-specific divergence may also turn out to play a major role in the generation of the outstanding diversity in tropical forest ecosystems and, more generally, to be a fundamental mechanism in the maintenance of trait variation in natural populations.

Funding information

This work was supported by the French MEDD ministry "Ecofor – Ecosystèmes forestiers" program and by the EU-funded PO-FEDER 2007-2013 ENERGIRAVI program.

Acknowledgements

We thank Saintano Dufort, Laetitia Ruelle, Jean Weigel, Jocelyn Cazal, Saint-Omer Cazal, Valérie Troispoux, Caroline Duret, Ann-Magret Amui-Vedel, Vincent Vedel and the CIRAD Kourou forest technical team for seed sampling, plant culture, and phenotypic measurements. We thank Jean-Jacques Boreux and Marie-Pierre Etienne for their advice on Bayesian modelling. We thank Erwin Dreyer, Jérôme Chave and Christian Lexer who carefully and critically read this manuscript, and Lynda Delph for insightful discussions. Louise Brousseau was supported by a "Contrat Jeune Scientifique" assistantship from INRA.

Article n°4 – Supplementary methods

Method 1 - Design of the incomplete randomized block experiments

Each block was made of sixteen seedlings from four different maternal families (four seedlings per family). To obtain this design, we proceeded as follows. Each block was randomly assigned to a given position in the glasshouse. Next, sets of four families were randomly assigned to blocks, then seedlings from each family were assigned to each block containing that family, and finally the sixteen seedlings belonging to a block were randomly assigned positions within the block. Seedlings were submitted to daily and seasonal natural variations of irradiance. They were maintained in non-limiting water conditions (i.e., soil water content close to field capacity, i.e. around 0.20 m3 m-3) throughout the experiment (i.e. 24 months) by watering the pots every second or third morning. Homogeneity of the environmental conditions in the glasshouse (i.e. air temperature, air humidity, radiation) was tested twice a year over a 3 week's period. Air temperature and humidity (average 28.6 ± 2.2 °C and 72.7 ± 8.6%, respectively) were recorded at three different locations in the glasshouse using a temperature and relative humidity probe (HMP45, Yaisala, Helsinky, Finland) connected to a CR10X datalogger (Campbell Scientific Inc., Logan, UT, USA). Photosynthetic photon flux density (PPFD) was measured above each block using a linear PAR ceptometer (AccuPar, Decagon Devices, Pullman, WA, USA) and compared with incident photosynthetic photon flux density outside the glasshouse. This allowed calculating a value of relative irradiance for each block and an average relative irradiance in the glasshouse, which was about 14.3 ± 2.3 % over the study period. To avoid any competition for light among the plants, pots were occasionally turned or their position swapped within the block if necessary to minimise vertical overlap of leaves between seedlings.

Method 2 - Bayesian model of phenotypic value decomposition

Phenotypic differences among species, habitats-of-provenance and maternal families were detected using a hierarchical linear model including seed mass effects and the three levels of genetic divergence (species, provenance, and families), as shown in the 'conceptual model' figure:

For all individuals "i":

$$Yijklm \sim N(mean_{ijklm}, \tau res)$$

$$mean_{ijklm} = \mu + \alpha j + \beta k + \gamma kl + \tau klm + (\varphi k \times Seed\ massi)$$

"Y" corresponds to the individual value for the phenotypic character. "τres" corresponds to residual precision (1 / "within-groups" variance) of phenotypic variations. The term "α" corresponds to the effect of the different years of seed sampling and culture, "β" corresponds to species effect, "γ" corresponds to provenance effect within each species (we allow each species to respond differently to soil provenance), and "τ" corresponds to the effect of maternal family (in terms of mother tree identity) within each species and soil of provenance. Then, the term "φ" is the regression coefficient between trait value and seed mass. We also defined one parameter per species in order to allow for divergent effects of seed mass variation in the two species. This coefficient may be null, suggesting that intra-specific variations in seed mass does not affect phenotypic variations.

Prior definition:

All parameters were sorted using non-informative priors:

τres ~ Gamma(0.0001,0.0001)

μ ~ N(0.00001,0.00001)

for all year "j" : αj ~ N(0.00001,0.00001)

for all species "k" : βk ~ N(0.00001,0.00001)

for all provenance "l" (within each species k): γkl ~ N(0.00001,0.00001)

for all maternal families "m" (within each provenance l and species k): τklm ~ N(0.00001,0.00001)

for all species "k" : φk ~ N(0.00001,0.00001)

The model was made identifiable by defining constraint Σ αj = 0 for each factor. Model was computed using 1 000 000 iterations with a burning of 100 000 and a thinning of 500. Parameters were estimates with 95% credible interval.

A parameter with 95% credible interval not overlapping 0 indicates that the phenotypic value of the group diverges to the phenotypic mean with a probability of 95%. Two groups identified by the same component (e.g. two habitats within the same species, for a given trait) are considered as different if their 95% credible intervals do not overlap.

Method 3 - Estimation of σ^2_M/σ^2_P with a Bayesian two-ways analysis of variance

Heritability was estimated at intra-specific level by estimation of 'among-family' exact precision.

For all individuals,

$$Trait_{ijm} \sim N(mean\ ijm, \tau res)$$

$$mean_{ijm} = \mu + \alpha j + \tau m$$

For all years of sampling and culture,

$\alpha j \sim N(0, \tau year)$

For all maternal families,

$\tau m \sim N(0, \tau progeny)$

Priors definition

$\mu \sim N(0.00001, 0.00001)$

$\tau res \sim Gamma(0.0001, 0.0001)$

$\tau years \sim Gamma(0.0001, 0.0001)$

$\tau progeny \sim Gamma(0.0001, 0.0001)$

We computed σ^2M/σ^2P where σ^2M is the genetic variance "among groups" (inverse of "among groups" precision τM), σ^2P the total phenotypic variance (sum of the inverse of "among families" precision τM plus the inverse of "among years" precision $\tau years$, plus the inverse of residual precision τres).

Model was computed using 500 000 iterations with a burning of 2000 and a thinning from 20 to 50 depending on the different traits.

Article n°4 – Supplementary figures

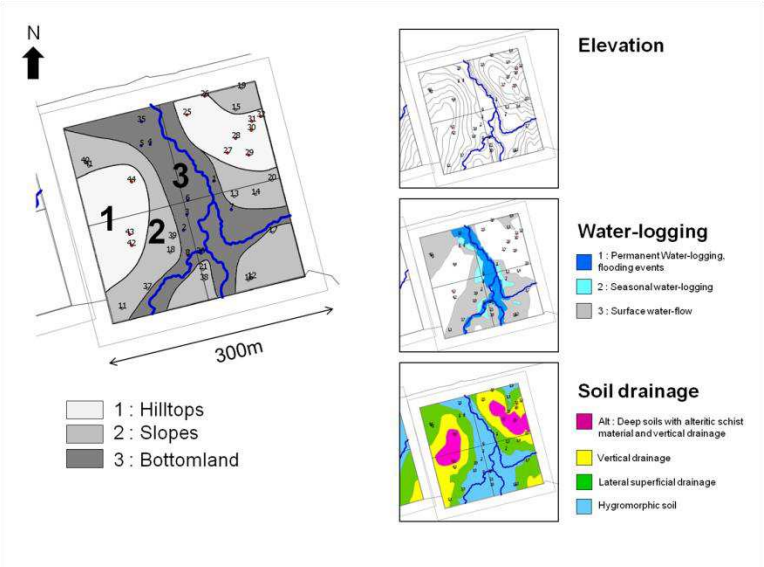


Figure S1: Three habitats were defined according to elevation, water-logging and soil drainage conditions. Sampled mother trees are indicated by red circles and numbers.

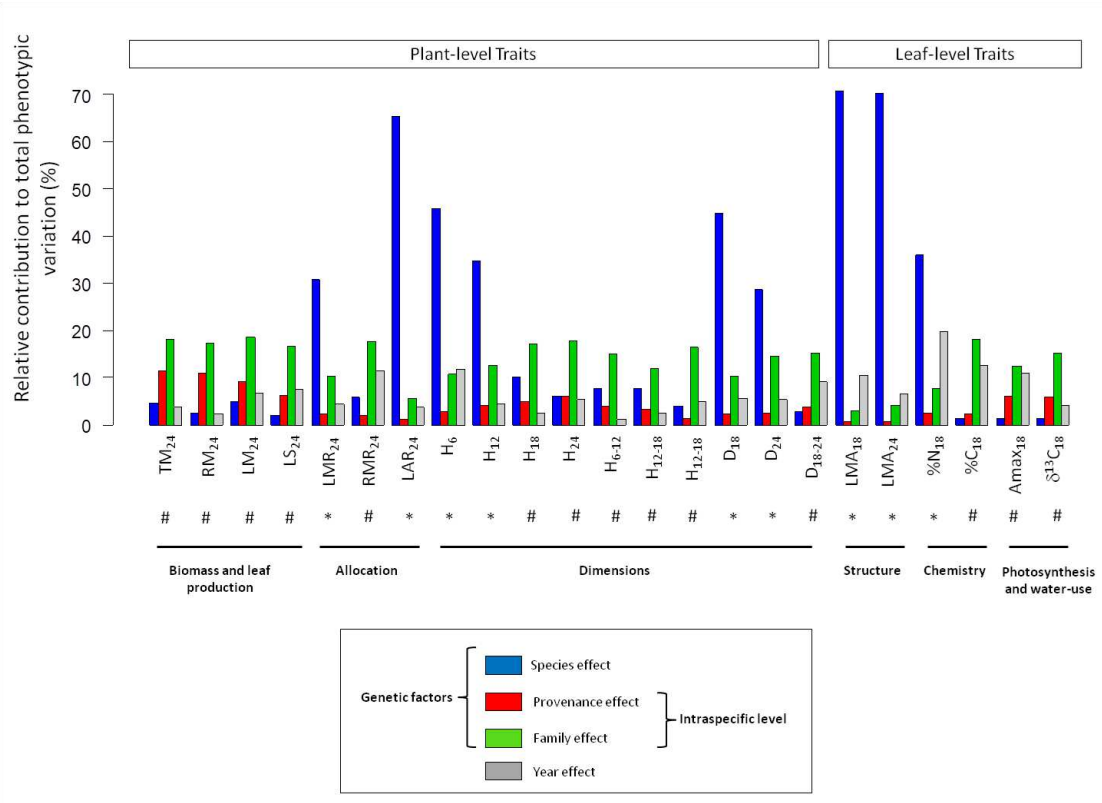


Figure S2: Relative size of different effects on trait variability (seed mass effect is not displayed here and therefore the bars do not sum to 100%).

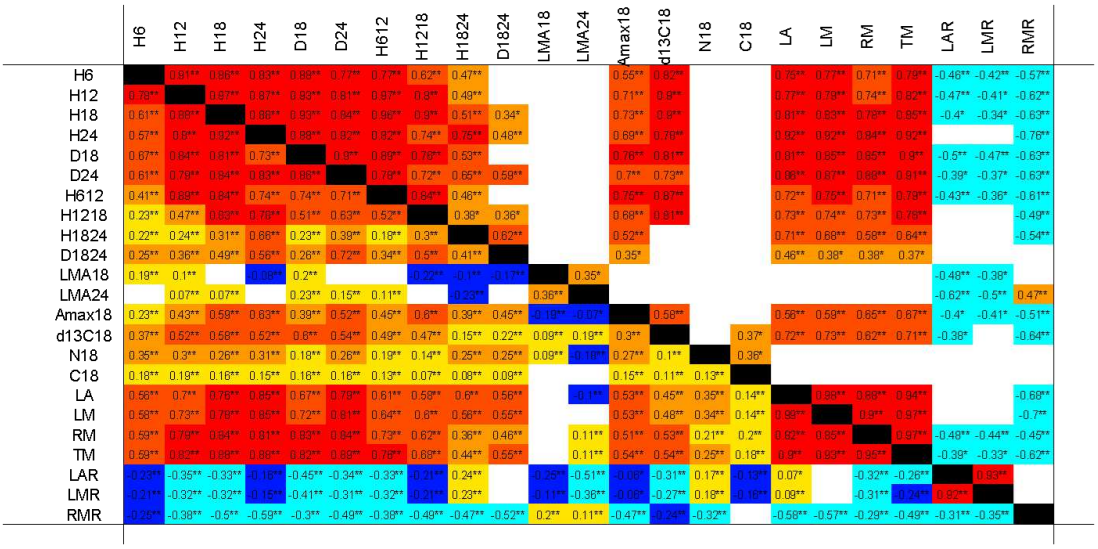


Figure S3: Pearson's correlation between phenotypic traits at seedling and maternal family level for *Eperua falcata*. Colours indicate the sign and strength of significant correlations (blue = negative correlation; red = positive correlation; deeper colours indicate stronger correlation). Empty cells correspond to non-significant correlations (α = 5%). Significance levels: * = 5%; ** = 1% or less. FDR: seedling level: 0.07% (α = 5%) and 0.01% (α = 1%); maternal family level: 0.4% (α = 5%) and 0.06% (α = 1%).

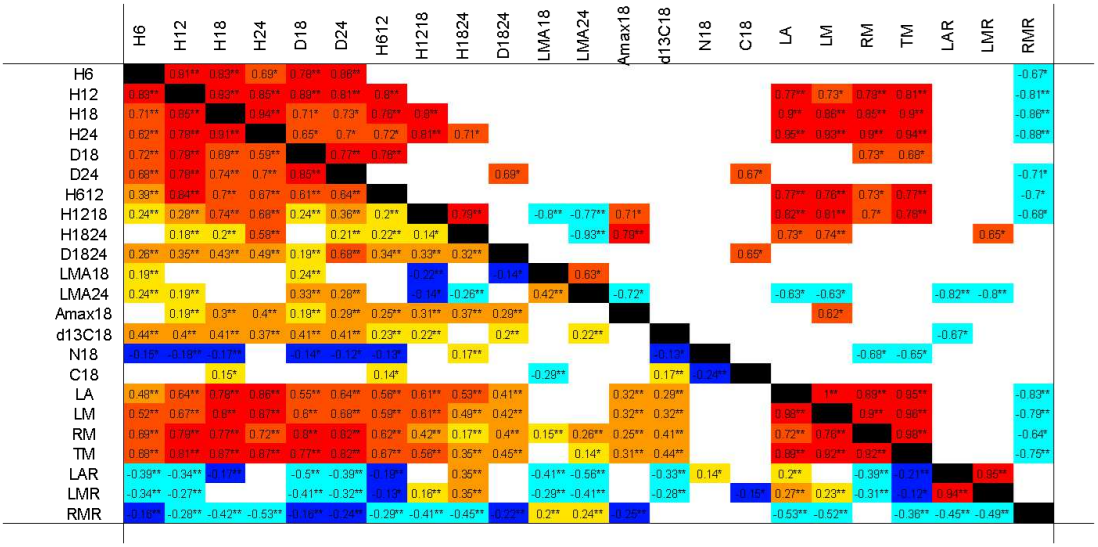


Figure S4: Pearson's correlation between phenotypic traits at seedling and maternal family level for *Eperua grandiflora*. Colours indicate the sign and strength of significant correlations (blue = negative correlation; red = positive correlation; deeper colours indicate stronger correlation). Empty cells correspond to non-significant correlations (α = 5%). Significance levels: * = 5%; ** = 1% or less. FDR: seedling level: 0.2% (α = 5%) and 0.04% (α = 1%); maternal family level: 1.2% (α = 5%) and 0.3% (α = 1%).

Article n°4 – Supplementary tables

Table S1: Sampling size. Nprogeny indicates the number of mother trees (maternal families) and nseedlings the number of seedlings used in the analysis.

	Provenance	$N_{progeny}$	$n_{seedlings}$ (18 months)	$n_{seedlings}$ (24 months)
E. falcata	Bottomland	10	425	180
	Slope	14	538	274
	Hilltop	10	400	202
E. grandiflora	Bottomland	2	58	26
	Slope	5	155	100
	Hilltop	3	61	44
	Total	44	1637	826

Table S2: Sampling size for each mother tree and year of fructification.

		Mother tree	2006	2007	2008
E.falcata	Bottomland	1	0	7	0
		2	60	0	0
		3	0	20	48
		4	0	12	0
		5	0	12	0
		6	40	0	28
		7	47	0	0
		8	0	12	48
		9	0	12	0
		10	0	52	27
	Slope	11	0	16	0
		12	18	0	0
		13	21	0	0
		14	5	0	0
		15	0	17	0
		16	0	7	0
		17	18	8	48
		18	57	0	0
		19	0	52	0
		20	45	0	24
		21	12	0	0
		22	0	20	31
		23	0	30	38
		24	0	51	20
	Hilltop	25	0	0	44
		26	20	0	0
		27	53	0	0
		28	59	0	0
		29	0	12	0
		30	0	12	48
		31	0	48	23
		32	0	36	0
		33	0	7	0
		34	0	38	0
E.grandiflora	Bottomland	35	0	6	0
		36	0	0	52
	Slope	37	0	0	20
		38	0	0	64
		39	0	8	20
		40	0	12	0
		41	0	31	0
	Hilltop	42	0	0	38
		43	0	0	15
		44	0	8	0

Table S3: Parameters estimated by the Bayesian analysis of phenotypic value decomposition for seedlings biomass and energy allocation.

			Total Mass 24			Root Mass 24			Leaf mass 24			Leaf surface 24			LMR 24			RMR 24			LAR 24		
			2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5	
Year effect	⊖2006		-13.940	7.777		-4.217	2.758		-3.954	1.178		-565.5	45.7		-3.42E-02	1.99E-03		8.31E-03	5.37E-02	*	-8.6	-1.431	*
	⊖2007		0.535	16.230	*	-0.780	4.262		1.181	4.891	*	237.8	684.4	*	2.93E-03	2.79E-02	*	-4.99E-02	-1.71E-02	*	0.3759	5.314	*
	⊖2008		-12.930	1.274		-3.461	1.101		-3.450	-0.094	*	-439.5	7.8		-1.15E-02	1.12E-02		-1.34E-02	1.62E-02		-0.2466	4.232	
Species effect	⊖ *E. grandiflora*		-7.664	19.560		-3.691	5.052		-1.657	4.775		-506.0	395.3		-5.26E-02	-1.02E-02	*	-4.48E-02	1.21E-02		-17.36	-8.969	*
	⊖ *E. falcata*		-19.560	7.664		-5.052	3.691		-4.775	1.657		-395.3	306.0		1.02E-02	5.26E-02	*	-1.21E-02	4.48E-02		8.969	17.36	*
Habitat-of-provenance effect	*E. falcata*	⊖ Hilltop	-5.135	3.778		-1.903	0.959		-0.954	1.152		-157.6	150.1		1.77E-03	1.63E-02	*	-1.46E-02	4.05E-03		-0.009	2.873	
		⊖ Slope	-14.960	-5.466	*	-5.105	-2.056	*	-3.105	-0.862	*	-433.3	-120.1	*	-2.42E-03	1.29E-02		-7.55E-03	1.23E-02		0.062	3.091	*
		⊖ Bottomland	-11.850	16.050	*	2.441	5.708	*	0.699	3.102	*	115.5	456.2	*	-2.19E-02	-6.03E-03	*	-7.38E-03	1.37E-02		-4.541	-1.39	*
	E. grandiflora	⊖ Hilltop	-22.290	-2.797	*	-6.738	-0.478	*	-4.740	-0.134	*	-582.6	73.5		-8.34E-03	2.25E-02		-9.55E-03	3.12E-02		-2.444	3.661	
		⊖ Slope	-18.470	-0.314	*	-5.457	0.372		-4.572	-0.284	*	-566.8	49.1		-1.56E-02	1.35E-02		-1.11E-02	2.68E-02		-2.411	3.355	
		⊖ Bottomland	10.250	34.460	*	2.399	10.170	*	2.105	7.824	*	121.5	934.1	*	-2.48E-02	1.36E-02		-4.30E-02	7.51E-03		-4.796	2.811	
Family effect	*E. falcata*	Bottomland	-38.330	17.510		-12.780	5.160		-9.605	3.590		-1511.0	386.7		-5.28E-02	3.55E-02		-4.83E-02	6.83E-02		-9.879	7.538	
		⊖ family1	-4.839	35.990		0.753	13.870		-1.472	8.176		-261.2	985.2		-4.17E-02	2.34E-02		-4.58E-02	3.93E-02		-9.065	3.808	
		⊖ family2	-7.429	22.880		-1.224	8.510		-1.705	5.456		-155.6	826.4		-2.08E-02	2.73E-02		-2.61E-02	3.72E-02		-3.847	5.516	
		⊖ family3	-34.690	9.084		-10.700	3.359		-7.777	2.566		-1006.0	393.9		-3.24E-02	3.63E-02		-7.23E-02	6.86E-02		-7.793	5.793	
		⊖ family4	-17.590	26.040		-7.576	6.436		-2.623	7.685		-142.8	1235.0		-9.12E-03	5.92E-02		-7.23E-02	1.83E-02		-0.8136	12.83	
		⊖ family5	6.826	45.020	*	2.779	15.040		-0.926	8.097		-199.3	1040.0		-5.30E-02	8.22E-03		-6.12E-02	1.85E-02		-11.26	1.031	
		⊖ family6	-35.540	5.248		-10.890	2.210		-8.635	1.003		-1196.0	56.4		-3.82E-02	2.67E-02		-2.28E-02	6.24E-02		-5.888	7.022	
		⊖ family7	-14.180	21.390		-5.647	5.775		-2.162	6.241		-292.8	875.6		1.53E-03	5.73E-02	*	-5.25E-02	2.18E-02		0.6784	11.74	*
		⊖ family8	-34.340	5.225		-11.640	1.070		-8.611	0.735		-1198.0	76.2		-5.76E-02	5.11E-03		-4.03E-02	4.23E-02		-12.38	0.04394	
		⊖ family9	-19.950	8.228		-7.338	1.710		-3.755	2.901		-452.1	472.5		-1.22E-02	3.26E-02		-2.36E-02	3.32E-02		-2.446	6.454	
	E. falcata	Slope	-16.310	19.640		-6.699	4.847		-2.425	6.070		-194.4	1013.0		2.89E-03	6.19E-02	*	-7.40E-02	1.06E-03		0.551	12.25	*
		⊖ family11	-5.335	32.090		-3.188	8.833		-0.472	8.372		47.3	1211.0	*	-1.31E-02	4.67E-02		-1.10E-01	-3.18E-02		-0.7922	11.01	
		⊖ family12	-41.700	-3.289		-12.480	-0.141		-9.959	-0.884	*	-1417.0	-205.2	*	-3.56E-02	2.48E-02		1.75E-02	9.77E-02	*	-6.986	4.978	
		⊖ family13	3.345	66.880	*	-2.122	18.280		0.237	15.250	*	-159.2	1728.0		-5.83E-02	4.19E-02		-1.59E-01	-2.60E-02	*	-12.26	7.578	
		⊖ family14	-27.140	8.606		-8.492	2.985		-6.137	2.307		-719.0	395.2		-2.38E-02	3.37E-02		-2.37E-02	5.09E-02		-4.246	7.133	
		⊖ family15	-39.240	13.410		-10.700	6.211		-9.296	3.144		-1226.0	354.0		-3.97E-02	4.25E-02		-1.47E-02	9.52E-02		-11.79	4.425	
		⊖ family16	-24.790	0.818		-8.094	0.130		-4.948	1.102		-637.0	217.7		1.60E-02	5.74E-02	*	-3.95E-02	1.40E-02		4.147	12.33	*
		⊖ family17	-15.400	22.720		-4.322	7.918		-3.780	5.224		-512.9	664.3		-4.42E-02	1.61E-02		-3.33E-02	4.63E-02		-7.536	4.404	
		⊖ family18	5.329	42.700	*	1.399	13.400		1.290	10.120	*	207.7	1369.0	*	-2.69E-02	3.23E-02		-6.02E-02	1.78E-02		-5.508	6.196	
		⊖ family19	-21.090	8.386		-5.667	3.799		-4.995	1.969		-772.9	180.4		-2.86E-02	1.83E-02		1.02E-02	7.17E-02	*	-8.155	1.136	
		⊖ family20	-15.320	28.520		-5.544	8.534		-4.155	6.203		-507.9	853.4		-5.95E-02	1.04E-02		-6.23E-02	2.92E-02		-9.528	4.307	
		⊖ family21	-10.210	24.650		-3.139	8.054		-3.665	4.569		-553.3	547.7		-5.95E-02	-4.97E-03	*	-1.65E-02	5.63E-02		-13.12	-2.224	
		⊖ family22	-19.200	10.970		-5.802	3.887		-4.297	2.831		-489.5	475.3		-2.14E-02	2.71E-02		-1.72E-02	4.58E-02		-3.819	5.81	
		⊖ family23	-40.430	-7.125		-11.050	-0.360		-10.590	-2.727	*	-1407.0	-337.5	*	-3.35E-02	1.92E-02		2.39E-02	9.34E-02	*	-5.145	5.312	
		⊖ family24	4.287	36.630	*	2.910	13.300		0.106	7.747	*	19.5	1141.0	*	-4.42E-02	8.64E-03		-1.93E-02	4.83E-02		-9.452	1.043	
	E. grandiflora	Hilltop	-28.390	12.240		-8.654	4.395		-7.567	2.033		-935.0	270.4		-3.33E-02	3.10E-02		-3.02E-02	5.47E-02		-3.532	9.192	
		⊖ family27	1.480	38.770	*	-0.312	11.660		-0.500	8.311		-115.9	1006.0		-3.50E-02	2.42E-02		-8.84E-02	-1.05E-02	*	-8.104	3.622	
		⊖ family28	-29.780	5.146		-10.470	0.751		-7.164	1.089		-927.5	138.5		-1.66E-02	4.01E-02		-6.10E-02	1.19E-02		-3.167	8.172	
		⊖ family29	-22.760	20.770		-6.822	7.156		-6.044	4.240		-847.0	535.5		-6.67E-02	1.94E-03		-2.51E-02	6.58E-02		-15.06	-1.498	*
		⊖ family30	-41.460	-5.305	*	-12.390	-0.782		-9.182	-0.640	*	-1284.0	-114.5	*	-1.11E-02	4.59E-02		-1.51E-02	6.04E-02		-3.711	7.53	
		⊖ family31	-10.910	15.630		-3.770	4.749		-2.541	3.726		-317.6	561.4		-2.72E-02	1.56E-02		-2.42E-02	3.12E-02		-5.362	3.103	
		⊖ family32	-12.830	19.470		-3.555	6.818		-3.053	4.579		-374.8	664.4		-2.78E-02	2.37E-02		-2.75E-02	3.99E-02		-5.686	4.517	
		⊖ family33	-7.745	43.840		-5.821	10.740		0.949	13.140	*	32.1	1622.0	*	-2.62E-02	5.64E-02		-8.18E-02	2.59E-02		-8.533	7.741	
		⊖ family34	-35.510	-2.513	*	-10.080	0.516	*	-8.386	-0.590	*	-1071.0	-9.3	*	-5.20E-03	4.76E-02		-1.09E-02	5.80E-02		4.548	15.06	*
	E. grandiflora	Bottom.	-7.715	25.830		-3.008	7.763		-1.107	6.817		-238.0	840.4		-2.52E-02	2.83E-02		-3.10E-02	3.90E-02		-4.164	6.417	
		⊖ family35	-25.830	7.715		-7.763	3.008		-6.817	1.107		-840.4	238.0		-2.83E-02	2.52E-02		-3.90E-02	3.10E-02		-6.417	4.164	
		⊖ family36	10.170	40.550	*	4.184	13.940		0.149	7.325	*	-71.6	990.3		-3.79E-02	9.80E-03		-4.19E-02	2.16E-02		-7.008	2.554	
		⊖ family37	-21.950	8.828	*	-8.057	1.826		-5.370	1.901		-692.5	318.1		-2.91E-02	1.97E-02		-5.01E-02	1.42E-02		-6.073	3.537	
		⊖ family38	6.805	32.490	*	-0.449	7.800		4.541	10.610	*	469.7	1336.0	*	1.47E-02	5.55E-02	*	-6.96E-02	-1.60E-02	*	0.6324	8.725	*
		⊖ family39	-39.310	-2.809	*	-11.070	0.653		-10.720	-2.101	*	-1360.0	-189.4	*	-5.61E-02	1.03E-03		-1.77E-03	7.44E-02	*	-8.974	2.333	
		⊖ family40	-32.050	-2.268	*	-9.121	0.443		-6.629	0.407		-891.2	101.3		-1.16E-02	3.72E-02		3.95E-03	6.61E-02	*	-2.349	7.309	
		⊖ family41	-22.560	6.963		-7.263	2.219		-5.159	1.816		-665.6	333.1		-2.35E-02	2.31E-02		-2.30E-02	3.86E-02		-4.552	4.648	
		Hilltop	-16.430	15.450		-5.713	4.525		-4.195	3.337		-509.8	535.6		-3.05E-02	1.94E-02		-5.05E-02	1.61E-02		-5.805	4.071	
		⊖ family43	-11.850	28.400		-3.351	9.575		-2.657	6.852		-488.6	800.5		-2.60E-02	3.74E-02		-3.26E-02	5.14E-02		-5.467	7.094	
		⊖ family44	4.151	6.956	*	1.444	2.345		0.806	1.469	*	107.8	204.1	*	-9.61E-03	-5.20E-03	*	-4.02E-03	1.84E-03		-2.322	-1.446	*
Seed mass	*E. falcata*	⊖ Seed mass	0.467	1.105	*	0.200	0.405		0.071	0.222	*	7.9	29.3	*	-9.11E-04	9.70E-05		-4.38E-04	8.94E-04		-0.1614	0.03814	
	E. grandiflora	⊖ Seed mass																					

Table S4: Parameters estimated by the Bayesian analysis of phenotypic value decomposition for seedlings height and growth rate.

			Height 6			Height 12			Height 18			Height 24			Growth rate 6-12			Growth rate 12-18			Growth rate 18-24	
			2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5		2.5	97.5	
Year effect		⊖2006	-3.942	0.095		-4.525	3.804		-10.540	2.833		-23.680	3.692		-0.260	0.754		-1.265	0.030		-1.099	1.093	
		⊖2007	5.103	8.100		3.015	9.476	*	1.934	12.320	*	3.304	23.490	*	-0.383	0.365		-0.307	0.695		-0.011	1.539	
		⊖2008	-5.758	-3.498	*	-8.405	-3.715	*	-7.301	0.256	*	-12.370	5.044		-0.536	0.040		0.059	0.777	*	-1.482	-0.093	*
Species effect		⊖ *E. grandiflora*	5.864	10.940	*	7.423	18.270	*	0.543	17.990	*	-6.199	26.960		-0.080	1.226		-1.648	0.189		-0.842	1.774	
		⊖ *E. falcata*	-10.940	-5.864	*	-18.270	-7.423	*	-17.990	-0.543	*	-26.910	6.264		-1.226	0.080		-0.189	1.648		-1.774	0.842	
Habitat-of-provenance effect		⊖Hilltop	-1.001	1.034		-2.905	1.528		-4.820	2.296		-6.542	5.060		-0.364	0.165		-0.434	0.255		-0.398	0.510	
	E. falcata	⊖ Slope	-2.690	-0.477	*	-5.922	-1.251	*	-9.481	-2.042	*	-13.460	-1.726	*	-0.615	-0.064	*	-0.682	-0.009	*	-0.679	0.238	
		⊖ Bottomland	0.450	2.669	*	1.760	6.486	*	3.100	10.540	*	2.523	14.620	*	0.133	0.729	*	0.058	0.785	*	-0.312	0.664	
	E. grandiflora	⊖ Hilltop	-3.770	1.283		-6.915	3.098		-15.050	1.081		-24.790	0.109		-0.588	0.694		-1.588	0.039		-1.315	0.569	
		⊖ Slope	-4.768	-0.666	*	-10.870	-2.058	*	-11.680	2.511		-18.230	4.591	*	-1.185	-0.111	*	-0.361	1.043		-0.829	0.946	
		⊖ Bottomland	1.055	6.746	*	1.957	14.470	*	1.243	21.400	*	4.966	34.660	*	-0.179	1.316		-0.513	1.430		-0.814	1.500	
Family effect	*E. falcata*	⊖Family1	-9.256	3.962		-20.040	9.548		-35.700	7.991		-60.250	4.126		-2.264	1.413		-3.221	1.101		-4.286	0.720	
		⊖ family2	-1.734	5.262		-8.108	6.836		-10.270	13.600		-12.660	36.330		-1.318	0.579		-0.838	1.570		-2.192	1.918	
		⊖ family3	-3.683	1.853		-9.931	1.856		-16.660	2.096		-13.340	23.830		-1.233	0.143		-1.535	0.361		-0.324	2.602	
		⊖ family4	-6.338	3.695		-12.350	9.544		-17.900	17.510		-34.180	18.460		-1.400	1.404		-1.626	1.727		-1.540	2.642	
		⊖ family5	-2.460	9.201		-8.551	16.910		-13.760	26.530		-15.100	40.200		-1.276	1.590		-1.628	2.022		-1.021	3.193	
		⊖ family6	-1.201	4.767		4.191	16.820	*	10.640	30.750	*	-0.157	47.490		0.734	2.224	*	0.691	2.676	*	-1.503	2.382	
		⊖ family7	-6.828	0.476		-14.030	1.405		-22.200	2.585		-44.100	6.636		-1.375	0.481		-1.833	0.617		-4.034	-0.046	*
		⊖ family8	-5.045	0.319		-13.040	-0.713	*	-15.210	4.520		-21.010	20.120		-1.383	0.051		-0.737	1.140		-1.414	2.011	
		⊖ family9	-0.950	8.843		-4.250	17.810		-15.640	19.970		-24.470	26.940		-0.850	1.764		-2.589	0.759		-1.407	2.395	
		⊖ family10	-2.896	1.813		-4.232	5.720		-4.981	11.300		-17.840	17.010		-0.379	0.903		-0.427	1.115		-1.602	1.154	
	Bottomland	⊖ family11	-2.543	6.638		-5.979	14.810		-8.277	25.060		-12.680	34.780		-0.849	1.603		-1.097	2.050		-1.044	2.534	
		⊖ family12	-0.446	9.190		-1.231	18.890		-1.130	31.720		-1.162	46.040		-0.473	1.975		-0.586	2.558		-1.360	2.330	
		⊖ family13	-6.907	1.647		-22.190	-4.572	*	-38.480	-9.816	*	-49.480	-2.737	*	-2.931	-0.687	*	-3.251	-0.516	*	-4.051	-0.272	*
		⊖ family14	-2.919	12.890		-1.944	32.010		-4.792	48.990		9.442	89.270		-0.400	3.785		-1.523	3.571		0.041	6.182	
		⊖ family15	-7.567	1.706		-18.190	1.337		-30.080	1.382		-39.230	5.802		-2.107	0.293		-2.707	0.191		-2.054	1.412	
		⊖ family16	-11.690	2.666		-20.950	9.796		-31.460	14.580		-51.940	14.890		-2.033	1.705		-1.387	3.267		-4.395	0.630	
		⊖ family17	-6.710	-1.675	*	-9.611	0.953		-16.420	0.530		-24.890	7.599		-0.599	0.711		-1.476	0.183		-1.809	0.715	
		⊖ family18	-3.098	3.199		-4.705	9.422		-2.057	20.670		-22.210	23.800		-0.439	1.165		-0.012	2.204		-1.940	1.875	
		⊖ family19	-2.880	3.507		-4.670	8.480		-9.068	12.110		-4.543	40.030		-0.543	1.080		-1.468	0.671		0.407	4.007	
		⊖ family20	-0.608	4.919		-7.752	4.212		-11.820	7.493		-32.010	5.196		-1.337	0.130		-1.126	0.804		-2.541	0.383	
		⊖ family21	-0.366	9.788		-6.019	16.270		-14.670	20.730		-18.590	41.010		-1.305	1.417		-2.093	1.371		-2.281	1.951	
		⊖ family22	-1.555	4.455		7.728	20.190	*	10.520	30.460	*	-17.990	24.060		0.954	2.543	*	0.172	2.081	*	-2.023	1.337	
		⊖ family23	-5.306	0.056		-13.520	-1.050	*	-16.440	3.446	*	-17.080	21.290		-1.478	-0.120	*	-0.874	0.929		-0.309	2.633	
		⊖ family24	-5.834	-0.427	*	-15.740	-3.640	*	-24.280	-4.837	*	-51.450	-12.120	*	-1.709	-0.302	*	-1.843	0.115		-2.691	0.488	
	Slope	⊖ family25	1.429	7.117	*	1.642	14.240	*	3.925	24.130	*	-1.935	41.500		-0.100	1.511		0.040	1.974	*	0.237	3.490	*
		⊖ family26	-6.126	3.802		-13.470	8.076		-19.020	15.620		-27.530	22.680		-1.445	1.126		-1.534	1.719		-2.705	1.472	
		⊖ family27	1.494	8.241	*	5.147	19.790	*	7.397	30.930	*	-3.424	42.850		0.421	2.210	*	-0.152	2.120		-2.470	1.255	
		⊖ family28	-5.262	1.537		-10.960	3.536		-16.220	7.037		-30.020	14.890		-1.123	0.560		-1.307	0.891		-3.026	0.612	
		⊖ family29	-3.143	7.457		-10.050	12.880		-20.690	15.120		-32.160	18.050		-1.492	1.239		-2.076	1.365		-2.504	1.512	
		⊖ family30	-9.722	-4.582	*	-21.420	-10.470	*	-30.770	-13.330	*	-53.750	-11.740	*	-2.278	-0.852	*	-1.872	-0.236	*	-3.169	0.279	
		⊖ family31	-4.197	0.935		-9.022	2.189		-15.130	2.695		-15.230	18.520		-0.980	0.320		-1.351	0.305		-0.617	2.015	
		⊖ family32	-3.137	3.530		-6.420	8.061		-13.010	10.100		-15.850	25.450		-0.748	0.980		-1.323	0.797		0.049	3.183	
		⊖ family33	-0.429	12.040		1.066	27.830	*	0.767	43.840	*	-10.970	53.810		-0.320	3.107		-0.685	3.294		-2.527	2.453	
		⊖ family34	-8.750	-1.922	*	-18.370	-4.484	*	-27.520	-5.289	*	-35.320	5.055		-1.878	-0.225	*	-1.852	0.238		-1.268	1.961	
	Bottom.	⊖ family35	-6.774	1.270		-8.744	9.266		-8.683	20.310		-20.390	23.200		-0.363	1.712		-0.534	2.265		-1.250	2.002	
		⊖ family36	-1.270	6.774		-9.266	8.744		-20.310	8.683		-23.060	20.770		-1.712	0.363		-2.265	0.534		-2.002	1.230	
E. grandiflora	Slope	⊖ family37	11.250	19.140	*	11.960	28.360	*	12.540	38.990	*	9.096	46.720	*	-0.080	1.934		-0.230	2.189		-1.155	1.764	
		⊖ family38	-3.346	2.369		-4.716	7.771		-10.670	9.360		-20.240	16.290	*	-0.457	0.966		-1.334	0.643		-1.456	1.452	
		⊖ family39	0.414	6.741	*	0.769	15.250	*	8.626	31.950	*	15.070	48.120	*	-0.095	1.618		0.953	3.215	*	0.690	3.213	*
		⊖ family40	-13.220	-4.036	*	-23.600	-3.786	*	-39.580	-7.714	*	-56.490	-10.540	*	-2.041	0.226	*	-3.172	-0.160	*	-3.673	-0.204	*
		⊖ family41	-13.040	-6.562	*	-22.460	-8.805	*	-32.160	-10.230	*	-42.010	-3.242	*	-1.955	-0.232	*	-2.078	0.189		-1.674	1.291	
	Hilltop	⊖ family42	1.242	7.865	*	-9.307	4.845		-12.600	10.200		-24.700	11.790		-2.109	-0.311	*	-0.912	1.257		-1.792	1.025	
		⊖ family43	1.060	9.097	*	0.219	17.390	*	-6.963	20.670		-19.130	20.530		-0.206	1.881		-1.550	1.032		-2.603	0.444	
		⊖ family44	-14.360	-4.830	*	-16.840	4.086		-21.990	11.520		-18.240	30.870		-0.959	1.628		-1.354	1.648		-0.446	3.425	
Seed mass	*E. falcata*	⊖ Seed mass	2.482	2.970	*	3.753	4.801	*	4.475	6.155	*	3.410	6.779	*	0.194	0.323	*	0.073	0.234	*	-0.099	0.172	
	E. grandiflora	⊖ Seed mass	0.232	0.350	*	0.314	0.567	*	0.349	0.757	*	0.060	0.875	*	0.014	0.045	*	-0.002	0.039		-0.051	0.010	

Table S5: Parameters estimated by the Bayesian analysis of phenotypic value decomposition for seedlings diameter and radial growth.

				Diameter 18			Diameter 24			Radial gr. with 18-24		
				2.5	97.5		2.5	97.5		2.5	97.5	
Year effect			⊖2006	0.248	1.030	*	-1.013	0.803		-0.145	-0.005	*
			⊖2007	-0.127	0.489		0.251	1.569	*	0.031	0.150	*
			⊖2008	-1.074	-0.420	*	-1.338	-0.232	*	-0.054	0.041	
Species effect			⊖ *E. grandiflorus*	1.017	2.031	*	0.428	2.440	*	-0.121	0.077	
			⊖ *E. fulcatus*	-2.031	-1.017	*	-2.440	-0.428	*	-0.077	0.121	
			⊖Hilltop	-0.284	0.134		-0.549	0.131		-0.049	0.013	
	E. fulcatus		⊖ Slope	-0.397	-0.133	*	-0.730	-0.043	*	-0.024	0.041	
Habitats-of-provenance effect			⊖ Bottomland	0.214	0.438	*	0.242	1.019	*	-0.023	0.043	
			⊖Hilltop	-0.522	0.434		-1.333	0.299		-0.123	0.021	
	E. grandiflorus		⊖ Slope	-0.887	-0.044	*	-0.842	0.597		0.014	0.143	*
			⊖ Bottomland	-0.102	1.092		-0.231	1.487		-0.109	0.041	
			⊖family1	-1.993	0.597		-4.334	-0.330	*	-0.404	-0.045	*
			⊖ family2	-0.134	1.272		-0.400	2.438		-0.141	0.143	
			⊖ family3	-0.415	0.302		-0.140	2.212		0.030	0.240	*
			⊖ family4	-1.376	0.730		-2.547	0.814		-0.124	0.144	
			⊖ family5	-0.903	1.488		-1.140	2.230		-0.073	0.232	
			⊖ family6	0.804	1.999	*	0.784	3.842	*	-0.033	0.229	
			⊖ family7	-1.403	0.079		-2.245	0.970		-0.192	0.098	
			⊖ family8	-1.033	0.119		-1.900	0.490		-0.214	0.013	
			⊖ family9	-0.845	1.230		-1.728	1.414		-0.148	0.129	
			⊖ family10	-0.844	0.099		-1.432	0.855		-0.054	0.143	
			⊖ family11	-1.170	0.804		-1.043	2.094		-0.020	0.239	
			⊖ family12	-0.027	1.925		-0.299	2.788		-0.125	0.137	
			⊖ family13	-2.303	-0.775	*	-3.370	-0.178	*	-0.281	0.010	
			⊖ family14	0.031	3.231	*	-0.241	4.438		-0.214	0.234	
			⊖ family15	-1.414	0.454		-2.031	0.944		-0.141	0.125	
			⊖ family16	-2.434	0.302		-2.864	1.349		-0.127	0.244	
			⊖ family17	-1.070	-0.054	*	-2.281	-0.031	*	-0.115	0.075	
			⊖ family18	0.180	1.343	*	-0.722	2.145		-0.130	0.118	
			⊖ family19	-0.093	1.143		0.214	3.193	*	0.022	0.279	*
			⊖ family20	-0.329	0.420		-2.019	0.312		-0.203	0.004	
			⊖ family21	-0.277	1.828		-0.941	2.462		-0.193	0.117	
Family effect			⊖ family22	0.437	1.420	*	-0.938	1.483		-0.111	0.120	
			⊖ family23	-1.320	-0.139	*	-1.441	0.784		-0.047	0.142	
			⊖ family24	-1.443	-0.488	*	-3.048	-0.291	*	-0.178	0.044	
			⊖ family25	0.185	1.382	*	-0.191	2.343		-0.033	0.220	
			⊖ family26	-1.483	0.579		-2.004	1.219	*	-0.102	0.194	
			⊖ family27	0.954	2.337	*	0.334	3.343	*	-0.144	0.129	
			⊖ family28	-1.330	0.054		-2.383	0.443		-0.139	0.137	
			⊖ family29	-0.740	1.384		-1.905	1.300		-0.238	0.044	
			⊖ family30	-1.904	-0.848	*	-3.123	-0.343	*	-0.183	0.042	
			⊖ family31	-0.444	0.618		-0.784	1.453		-0.081	0.114	
			⊖ family32	-0.374	0.798		-0.785	1.881		-0.034	0.202	
			⊖ family33	-0.575	2.184		-1.382	2.307		-0.234	0.111	
			⊖ family34	-2.037	-0.714	*	-2.944	-0.231	*	-0.154	0.091	
			⊖ family35	-1.040	0.440		-2.713	-0.013	*	-0.243	0.003	
			⊖ family36	-0.440	1.040		0.013	2.713	*	-0.003	0.243	
			⊖ family37	0.488	2.237	*	0.841	3.314	*	-0.022	0.195	
			⊖ family38	-0.541	0.433		-1.308	1.049		-0.129	0.089	
			⊖ family39	-0.910	0.473		-0.347	1.499		0.044	0.224	*
			⊖ family40	-1.591	0.499		-2.599	0.473		-0.244	0.009	
			⊖ family41	-1.459	-0.159	*	-2.792	-0.482	*	-0.181	0.024	
			⊖ family42	-1.244	0.107		-1.634	0.337		-0.044	0.134	
			⊖ family43	-0.201	1.439		-0.053	2.544		-0.028	0.201	
			⊖ family44	-1.030	0.948		-2.344	0.732		-0.244	-0.004	*
Seed mass	*E. fulcatus*		⊖ Seed mass	0.421	0.521	*	0.424	0.443	*	0.001	0.020	*
	E. grandiflorus		⊖ Seed mass	0.037	0.041	*	0.023	0.079	*	-0.001	0.003	

Table S6: Parameters estimated by the Bayesian analysis of phenotypic value decomposition for leaf traits.

			LMA 18		LMA 24		%N 18		%C 18		Amax 18		δ^{13}C 18	
			2.5	97.5	2.5	97.5	2.5	97.5	2.5	97.5	2.5	97.5	2.5	97.5
Year effect		⊖2006	3.682	6.592 *	2.955	8.190 *	-0.134	-0.018 *	-0.430	-0.122 *	-0.938	-0.320 *	-0.024	0.483
		⊖2007	0.066	2.205 *	-4.925	-1.292 *	0.181	0.264 *	-0.149	0.070	0.083	0.536 *	-0.121	0.205
		⊖2008	-7.123	-5.484 *	-4.181	-0.910 *	-0.182	-0.111 *	0.221	0.409 *	0.155	0.497 *	-0.372	-0.120 *
Species effect		⊖ *E. grandiflora*	8.887	12.610 *	8.169	14.350 *	-0.261	-0.117 *	-0.229	0.149	-0.415	0.338	-0.273	0.314
		⊖ *E. falcata*	-12.610	-8.887 *	-14.350	-8.169 *	0.117	0.261 *	-0.149	0.229	-0.338	0.415	-0.314	0.273
Habitat-of-provenance effect	*E. falcata*	⊖ Hilltop	-0.669	0.835	-1.332	0.790	-0.006	0.052	0.012	0.163 *	-0.221	0.087	-0.073	0.136
		⊖ Slope	-1.398	0.168	-2.316	-0.110 *	-0.005	0.052	-0.121	0.024	-0.267	0.068 *	-0.315	-0.107 *
		⊖ Bottomland	-0.286	1.335	0.363	2.672 *	-0.077	-0.017 *	-0.120	0.037	-0.003	0.334 *	0.076	0.291 *
	E. grandiflora	⊖ Hilltop	-0.435	3.155	-1.648	2.795	-0.036	0.109	-0.074	0.317	-0.995	-0.246 *	-0.454	0.106
		⊖ Slope	-3.226	-0.193 *	-2.781	1.359	-0.002	0.121	-0.147	0.173	-0.135	0.471 *	-0.394	0.051
		⊖ Bottomland	-1.867	2.414	-2.560	2.921	-0.181	-0.006 *	-0.355	0.098	0.012	0.872 *	0.060	0.692 *
Family effect	*E. falcata*	Bottomland	⊖family1	-5.785	4.044	-5.865	6.779	-0.066	0.292	-0.723	0.205	-1.176	0.795	-1.049	0.235
			⊖family2	-1.273	4.221	-5.293	4.195	-0.191	0.018	-0.536	-0.010	-0.199	0.853	-0.514	0.265
			⊖family3	-0.063	3.841	-3.450	3.429	-0.066	0.080	-0.176	0.205	-0.732	0.088	-0.607	-0.036 *
			⊖family4	-7.451	0.867	0.928	10.320 *	-0.237	0.051	-0.308	0.439	-0.947	0.615	-0.463	0.636
			⊖family5	-6.530	1.735	-9.190	0.763	-0.195	0.104	-0.030	0.739	-0.966	0.888	0.099	0.706 *
			⊖family6	-2.481	1.782	-1.144	7.803	0.018	0.176 *	0.137	0.552 *	-0.019	0.888	0.140	1.182 *
		Slope	⊖family7	-4.555	0.840	-6.981	2.388	-0.108	0.103	-0.047	0.501	-1.055	0.049	-0.370	0.387
			⊖family8	-2.000	2.079	-6.564	1.425	-0.104	0.060	-0.330	0.092	-0.609	0.196	-0.550	0.005
			⊖family9	0.977	8.571 *	-4.175	4.842	-0.197	0.091	-0.501	0.245 *	-0.569	0.902 *	-0.389	0.518
			⊖family10	-1.475	2.184	-3.904	2.583	0.017	0.159 *	-0.394	-0.022 *	0.178	0.887 *	-0.301	0.208
		Hilltop	⊖family11	-3.461	3.557	-7.571	0.857	-0.008	0.255	0.056	0.741 *	-0.805	0.570	0.009	0.931 *
			⊖family12	-8.027	-1.005 *	-10.110	-1.579 *	NA	NA	NA	NA	-0.723	0.727	NA	NA
			⊖family13	-2.873	3.777	-3.443	5.727	-0.222	0.034	-0.109	0.561	-1.780	-0.462 *	-0.774	0.179
			⊖family14	-11.400	0.654	-8.290	6.163	NA	NA	NA	NA	-0.014	2.374	NA	NA
			⊖family15	-1.983	4.952	-0.403	7.813	0.068	0.322 *	-0.442	0.219	-0.773	0.621	-0.652	0.319
			⊖family16	-4.419	5.727	1.028	12.790 *	-0.216	0.187	-0.347	0.685 *	-1.252	0.907	-0.668	0.751
			⊖family17	-2.467	1.270	-7.183	-1.255 *	-0.098	0.046	-0.455	-0.078 *	-0.359	0.406	-0.424	0.092
			⊖family18	-4.799	-0.220 *	-2.563	6.226	-0.131	0.063	-0.173	0.339	0.323	1.284 *	0.001	0.719 *
			⊖family19	0.609	5.326 *	-4.255	4.201	0.084	0.243 *	-0.242	0.169	-0.419	0.550	-0.320	0.306
			⊖family20	0.392	4.599 *	-0.300	6.583	-0.316	-0.134 *	-0.851	-0.368 *	-0.743	0.088	-0.444	0.177
			⊖family21	-3.304	4.766	-8.277	1.581	NA	NA	NA	NA	-0.787	0.733	NA	NA
			⊖family22	2.365	6.833 *	0.344	8.267 *	-0.131	0.029	-0.037	0.386	0.109	0.995 *	0.215	0.793 *
			⊖family23	-0.501	3.402	-4.229	2.765	-0.133	0.010	-0.328	0.042	-0.609	0.197 *	-0.361	0.180 *
			⊖family24	-3.780	0.277	-6.069	1.475	-0.049	0.093	-0.078	0.298	-0.965	-0.153 *	-0.731	-0.209
		Hilltop	⊖family25	-2.244	2.310	-2.142	5.531	-0.079	0.108	-0.097	0.394	0.171	1.027 *	-0.116	0.596
			⊖family26	-8.222	-0.902 *	-4.695	4.514	-0.369	-0.089 *	-0.474	0.276	-0.786	0.713	-0.678	0.352
			⊖family27	-1.716	3.444	-5.408	3.260	-0.005	0.195	-0.514	0.015	-0.023	0.998	0.197	0.968 *
			⊖family28	-4.321	0.556	-5.537	2.693	-0.029	0.170	-0.033	0.497	-0.409	0.613	-0.357	0.372
			⊖family29	-0.372	7.163	-4.799	5.103	-0.182	0.101	-0.308	0.518	-0.607	0.978	-0.517	0.539
			⊖family30	-3.181	0.917	-4.275	3.874	-0.131	0.019	-0.434	-0.044 *	-0.937	-0.161 *	-0.814	-0.237 *
			⊖family31	0.226	3.905 *	-2.940	3.268	-0.024	0.118	-0.113	0.259	-0.555	0.220	-0.561	-0.040 *
			⊖family32	0.586	5.464 *	-2.433	4.924	0.093	0.271 *	0.215	0.675 *	-0.216	0.777	-0.180	0.449 *
			⊖family33	-5.435	4.357	-6.077	5.709	-0.240	0.138	-0.607	0.299	-1.385	0.478	0.011	1.351 *
			⊖family34	-3.632	1.091	-3.988	3.595	-0.137	0.050	-0.534	-0.048 *	-0.941	0.100	-1.075	-0.418 *
	E. grandiflora	Bottom.	⊖family35	-4.501	1.430	-5.116	2.574	-0.227	0.028	-0.801	-0.139 *	-0.623	0.590	-0.466	0.437
			⊖family36	-1.430	4.501	-2.574	5.116	-0.028	0.227	0.139	0.801 *	-0.590	0.623	-0.437	0.466
		Slope	⊖family37	-3.346	2.433	-3.336	3.605	-0.163	0.064	-0.076	0.526	-0.557	0.622	-0.256	0.574
			⊖family38	-3.017	1.046	-2.093	4.944	-0.042	0.117	0.139	0.559 *	-0.649	0.206	0.188	0.735 *
			⊖family39	-4.710	0.202	-8.308	-2.434 *	-0.096	0.098	-0.220	0.305	0.145	1.096 *	-0.672	0.049
			⊖family40	-0.547	6.007	0.564	8.732 *	-0.217	0.043	-0.542	0.129	-1.288	0.094	-0.374	0.612
			⊖family41	-1.688	3.260	-4.110	2.924	0.009	0.194 *	-0.662	-0.154 *	-0.342	0.631	-0.764	-0.111 *
		Hilltop	⊖family42	-2.383	2.714	-3.304	3.414	-0.054	0.142	-0.094	0.421	-0.575	0.404	-0.047	0.648
			⊖family43	-3.406	2.306	-1.226	5.949	-0.096	0.141	0.060	0.668 *	-1.116	0.042	-0.406	0.336
			⊖family44	-3.356	3.963	-6.942	2.082	-0.203	0.057	-0.910	-0.188 *	-0.082	1.330	-0.718	0.255
Seed mass		⊖ Seed mass	0.289	0.661 *	0.217	0.830 *	-0.003	0.011	0.005	0.042 *	0.006	0.079 *	0.096	0.149 *
	E. grandiflora	⊖ Seed mass	0.054	0.142 *	0.000	0.145 *	-0.003	3.86E-04	-0.002	0.006 *	-0.009	0.009	0.006	0.019 *

Article n°5 - Local adaptation in tropical rainforest trees: response of *Eperua falcata* (Fabaceae) seedling populations from contrasted habitats to drought and to water-logging

Louise BROUSSEAU[1,2,3], Ivan SCOTTI[1], Erwin DREYER[2,3], Damien BONAL[1,2,3].

1. INRA, UMR Ecologie des Forêts de Guyane, Campus agronomique, BP 709, 97387 Kourou cedex, French Guiana

2. INRA, UMR 1137 « Ecologie et Ecophysiologie Forestières », 54280 Champenoux, France.

3. Université de Lorraine, UMR 1137 « Ecologie et Ecophysiologie Forestières », 54280 Champenoux, France.

Abstract

The impact of drought and flooding stress on functional traits and fitness was studied in populations of seedlings from contrasting habitats, which have previously been shown to display phenotypic differences in controlled conditions.

All provenances responded similarly to limiting conditions, with overall reduced growth and decreased specific leaf area under drought, and with changes in biomass allocation and decreased specific leaf area under flooding. Significant provenance x treatment interactions were observed, suggesting the possible existence of adaptive responses; however, these interactions were mostly due to the reduction of differences between provenances observed in limiting conditions. Trait correlation matrices also differed between provenance x treatment combinations, which may indicate an effect of differential active responses to stress. Finally, it was shown that known functional traits such as water use efficiency and specific leaf area have an effect on fitness in all environments, and although it could not be proven that provenances had the most adaptive trait values in the treatment that mimicked their environment of origin, observed trend suggest that the subpopulations may undergo local adaptation.

Introduction

The way populations and species vary in space and time in response to environmental cues has captivated the attention of biologists for a very long time (Darwin 1859). Mechanisms underlying the adaptation of plants to environmental variation both in space and time are a major focus in biology, ecology and environmental science, as plant populations and communities are the foundation of most ecosystems. Plants that survive and reproduce in a given environment necessarily cope with its peculiarities, and are therefore able to face challenges arising in, and exploit resources provided by, that particular set of biotic and abiotic conditions (Delph & Kelly 2013). When plant species occur in diverse habitats, a main *tenet* is that their populations may exhibit "local adaptation" (Endler 1977), which means that the variability of their traits allows for the (potential) maximisation of fitness in each environment. Local adaptation *sensu stricto* implies genetic divergence of populations under the effect of natural selection (Kawecki & Ebert 2004), but the role of phenotypic plasticity in the maintenance of populations in variable environments is also widely acknowledged (Miner *et al.* 2005).

Traits that are involved in adaptation to variable environmental conditions display G×E (genotype-by-environment) interactions, which indicate differences among the reaction norms of different genotypes (Kruuk *et al.* 2008). G×E interactions are suggestive of local adaptation when the trait is involved in the determination of fitness (Conner & Hartl 2004). The way individual traits affect survival and reproduction is summarised by their "selection gradients" (Lande & Arnold 1983), which describe the relationship between the value of a trait and (components of) fitness. It is common to observe that the trait values that maximise fitness vary as a function of the environment, and therefore populations adapted to different environments have different values of a trait. Relating trait values to local environment and to fitness is the basis for the stringent demonstration of local adaptation (Endler 1986, Linnen *et al.* 2009). The role a trait plays in determining individual fitness is ultimately at the core of the (evolutionary) definition of a "functional trait" (Violle *et al.* 2007); therefore, studying the distribution of traits in natural populations and their effect on fitness is a key to the understanding of functional relationships of populations with their habitat.

The investigation of how populations adapt to variable environments and consequently diverge phenotypically is all the more intriguing in tropical rainforest ecosystems, which harbour exceedingly large amounts of phenotypic diversity. At least part

of that diversity is thought to be driven by adaptive processes, according to the 'niche theory' of biodiversity (Hutchinson 1959), which has received at least indirect confirmation by the observation of association between species and environmental parameters (Sabatier *et al.* 1997, John *et al.* 2007, Vincent *et al.* 2011). Because species diversification must ultimately stem from initial genetic divergence between conspecific populations, studying mechanisms underlying genetic differentiation among populations has a direct impact on our understanding of biodiversity. A further aspect that makes tropical lowland rainforests appealing from the theoretical standpoint is the way habitat properties can interact with evolutionary processes. In these ecosystems, significant environmental variation occurs at the very local geographical scale, driven by topography and associated soil water content constraints. In bottomlands, forest trees face seasonal or permanent water-logging and associated flooding conditions (Ferry *et al.* 2010). Prolonged water-logging commonly results in soil hypoxia (Ponnamperuma 1972) that affects below-ground respiration (Epron *et al.* 2006), induces a decrease in available N (Luizao *et al.* 2004), and may severely constrain the survival of trees. Soil instability along the slopes increases the frequency of tree fall events that contribute to change the levels of available light (Ferry *et al.* 2010), and slope itself, plus reduced soil depth, induce water shortages (Sabatier *et al.* 1997). Moreover, topography variations affect the soil water regime through a decrease in soil water availability from bottomland to hilltops (Daws *et al.* 2002), but hilltops (often referred to as "*terra firme*") display deeper soils and therefore larger reservoirs of available soil water. This patchiness may have a significant effect on species and population structure in addition to the widely described successional effect controlled by light availability and related to gaps in the canopy (Ferry *et al.* 2010). Numerous sympatric tree species display a non-random spatial distribution related to soil variations (Clark *et al.* 1998, Clark *et al.* 1999, Palmiotto *et al.* 2004), in association with differences in tolerance to seasonal drought (Engelbrecht *et al.* 2007) or flooding (Lopez & Kursar 2003, Baraloto *et al.* 2007). Other species seem to be more generalist, and are able to colonise and develop in all three habitat types. *Eperua falcata* Aubl, widely present across the Guiana shield, is one of such species. Significant genetic structuring in relation with habitat-soil related types was described for several genes involved in water relations within *E. falcata* (Audigeos *et al.* 2013), and significant phenotypic divergence was detected under common conditions among seedlings collected from *E falcata* populations growing in the three habitats (Brousseau *et al.* 2013). These observations were made under conditions of close to optimal water supply, and the

sensitivity of the three sub-populations to either water-logging or soil water deficit was not addressed.

Selective pressure caused water-logging and drought is expected to act at very early life stages, when large numbers of seedlings die. The sensitivity of seedlings and saplings to water-logging and soil water deficit is therefore relevant to understand how local environment may result in selection.

In this study, we used a provenance test to investigate whether seedlings originating from habitats with highly contrasted soil water conditions display different growth, leaf traits, and biomass allocation patterns when submitted to contrasting soil water conditions. Seedling populations originated from three different habitats: bottomland, slope and hilltop (Brousseau *et al.* 2013). Eighteen months old seedlings from this test were submitted during 6 months to three contrasted levels of water availability (severe water deficit, optimal water supply and lasting water-logging). We hypothesized that seedlings from hilltop provenances would display a larger tolerance to soil water deficit and a smaller one to water-logging with respect to bottomland seedlings.

Differences among provenances in tolerance to water deficit or water-logging were assessed from the reduction in growth, biomass accumulation and from changes in relative biomass allocation to shoots and roots, as well as in leaf traits. We used a Bayesian approach to assess the relative effects of habitat among sub-populations and of treatments on seedling growth, and leaf traits. Using biomass accumulation traits as proxies for fitness, the extent of local adaptation was assessed.

Material and methods

Eperua falcata (Aubl.) (Fabaceae) is abundant in the coastal plains of French Guiana and the Guyana shield, and has a clear preference for seasonally water-logged bottomlands, but occurs also on seasonally dry slopes and well-drained hilltops (Baraloto *et al.* 2007). *E. falcata* is bat-pollinated (Cowan 1975) and disperses its heavy seeds by explosive dehiscence and gravity at short distances of a few meters (Forget 1989). Gene dispersal distance is probably about 140-500 m (Hardy, pers. comm.).

Plant material and growth conditions

The protocol to obtain the seedlings used in this experiment was already described in details by Brousseau *et al.* (2013). Basically, seeds allowing studying a total of 1363 seedlings in the present experiment were sampled as progenies from 34 mother trees in the 9-ha undisturbed plot 6 of the Paracou forest inventory site (5°18'N, 52°53'W) of French Guiana, South America (Gourlet-Fleury *et al.* 2004). To reach the required number of seedlings from each mother tree, seeds were collected over three successive fructification periods (February-March 2006, 2007 and 2008). The year of seed sampling was then taken into account in the different statistical analyses. This site is characterized by a rugged landscape formed by the alternation of 40 to 50 m-high hills, hill slopes, and bottomlands, varying in soil drainage type and water table depth. Seeds were assigned to one of the three habitats in agreement with soil properties and the topographic position of their mother trees: "Bottomlands" (B) with hydromorphic soils and a water table between 0 and 60 cm depth; "Slopes" (S) with a water-table always below 100 cm; "Hilltops" (H), with deep soils, deep vertical drainage and a water-table always below 150 cm. Seeds were laid down in germination boxes filled with a substrate made of river sand that was maintained humid using a sprinkler system. Two months after germination, they were transplanted into individual 12-l pots containing a mix of sand and A horizon soil (30/70 v/v). They were transferred into a glasshouse (ambient air temperature: 28.6 ± 2.2 °C; relative air humidity: 72.7 ± 8.6%; relative irradiance: 14.3 ± 2.3 %). More details may be found in Brousseau *et al.* (2013).

During 18 months, seedlings were maintained at a gravimetric soil water content close to field capacity, i.e. around 20g kg^{-1} dry soil by watering the pots every second or third morning. After this date, 1242 seedlings were assigned to three treatments:

A control treatment (C) with pots maintained close to field capacity by manually

watering every second or third day;

A water-logging treatment (F), in which each pot was inserted into a larger PVC container (diameter = 30 cm; height = 50 cm; volume = 35 l) allowing to maintain a permanent water table 3-cm above soil surface; the water level was manually adjusted every second or third day;

A water deficit treatment (D), in which the seedlings were left without irrigation until wilting occurred, i.e., when the orientation of the 3-4 most recent leaves changed from horizontal to close to vertical. Leaf angles were observed every two-three days. After the onset of leaf wilting, pots were immediately re-irrigated with 50, 100, 150 or 200 ml water depending on total leaf area and plant size. Pots were weighted 12-18 h after watering to ensure homogenous drainage with a balance (range 50 kg ± 20g) every second week over the experiment and when leaf wilting was first observed. This procedure allowed maintaining soil water content between wilting point and wilting point +100 g kg^{-1} soil water.

Progenies were uniformly assigned to the three treatments while within progenies, seedlings were randomly assigned to each treatment. Within each treatment, progenies were randomly distributed among blocks of 16 seedlings from 4 different progenies. A detailed account of the experimental design is provided in **Table 1**.

Table 1: Detailed account of the number of individuals in the different treatments.

	Control	Dry	Flooded	
Bottomland	180	121	124	425
Slope	274	128	136	538
Hilltop	202	116	82	400
	656	365	342	1363

A subsample of the F and C plants (n = 15 for both) was weighted similarly to test for the homogeneity within treatments and among progenies and habitats.

Gravimetric soil water content at the end of the experiment ($SWC_{g\text{-}24\ mo}$) was measured by sampling ~500 g fresh soil in each pot at the end of the experiment, drying the samples at 105°C for about 7 days until constant mass. Fresh (FW) and dry (DW) mass were recorded with a balance (model, accuracy), a weighted, in order to calculate $SWC_{g\text{-}24\ mo}$ (%) as: $SWC_{g24mo} = 100 * \frac{FW-DW}{DW}$. This value was used to derive actual SWC_{g} from the pot weight at all dates. The changes in whole pot weight were attributed to changes in soil water content, and in plant fresh weight. The latter were assessed from allometric relationships between plant height and plant fresh weight recorded at the onset of the experiment and at the end.

Plant and leaf traits

Plant height (H, cm) and stem diameter near collar (D, mm) were recorded at the onset of the experiment when the seedlings were 18 months old and at harvest (end of the experiment at t = 24 months) to derive mean stem and radial increment rates.

At harvest, six to eight mature and fully expanded leaflets were collected per plant, close to the top. Fresh leaf area of these leaves was measured with an area meter (Li-2100, Licor, Lincoln, Nebraska) and leaves were subsequently dried to constant weight at 60 °C for about three days in order to calculate leaf mass per area ratio (LMA, g m^{-2}) as the ratio of dry mass over fresh leaf area. Leaves, stems and roots were then carefully separated and the occurrence of adventitious roots above the collar was recorded (presence/absence). The total fresh leaf area of each plant was measured with an area meter (Li-2100, Licor, Lincoln, Nebraska). The leaf, root and stem components were then dried to constant weight at 60 °C for about three days and weighted to obtain leaf, root and stem masses (LM, RM and SM, g). Leaf mass ratio (LMR) and root mass ratio (RMR) were then calculated as the ratio of leaf or root dry mass over total plant dry mass.

Decompostion of the phenotypic value with a Bayesian modeling approach

Some mortality occurred during the experiment, and the dead plants were excluded from the analysis. Similarly, some small individuals that never reached the wilting point were excluded from the analysis.

Due to the complex experimental design (hierarchic effects, uncomplete and unbalanced dataset), we used a Bayesian model of phenotypic value decomposition to estimate the respective effects of habitat and soil water availability (water deficit, control, water-logging) on phenotypic characters. The model included the effects of habitat, treatment (estimated for each population of provenance) on seedling phenotypic values. Progeny effects were also included.

The model was corrected by the addition of a factor relative to years of seed sampling and seedling cultivation. Because each block did not receive seedlings from each habitat, and was associated with a unique condition, we excluded the 'block' effect that made the model unidentifiable.

$$Trait[i] \sim N(mean[i], \tau_{Res}) \quad \text{(eq.1)}$$

$$mean[i] = mu + \theta_{year[i]} + \theta_{provenance[i]} + \theta_{provenance\,[i],treatment[i]} + \theta_{progeny/provenance\,[i]}$$

In order to make the model identifiable, we constrained $\Sigma\theta_{year}=0$ (year effect), $\Sigma\theta_{provenance}=0$ (provenance effect), and $\theta_{provenance,Control} = 0$: specifying that "control" is null means that we defined the "control" treatment as a reference ; individual phenotypic values depend only on provenance and year. We used non-informative prior for all parameters (quasi-uniform normal law of large variance). The model was computed using WinBUGS® (20000 iterations, burning=200, thinning=50).

Phenotypic correlations among traits

Last, we analyzed Pearson's correlations among phenotypic traits in the different conditions. We estimated first the overall correlation (all provenances confounded) each treatment, and the correlation within each provenance and each treatment. Confidence intervals were obtained by bootstrap (2000 iterations).

Estimation of selection gradients

Because biomass accumulation in tropical tree seedlings is strongly associated to survival (Clark & Clark 1985, Howe 1990, Gerhardt 1996, Poorter 1999, Gilbert *et al.* 2001), we took two traits: H_{1824} (height change during the stress) and RG_{1824} (relative height growth during the stress), which summarise the capacity to continue accumulating biomass during the stress experiment, as proxies of the seedling survival component of fitness. Selection gradients were analysed using the multiple regression approach described in Lande & Arnold (1983) with three functional traits as explanatory variables: leaf mass per area at 24 months (LMA_{24}), carbon isotope discrimination at 18 months ($\delta^{13}C_{18}$), and seed mass (which is not a seedling trait *per se* but has an impact on biomass traits). These traits have been shown to affect fitness in plant populations from contrasting environments (Dudley 1996, Poorter & Bongers 2006). Both linear and quadratic regressions were tested for LMA_{24} and $\delta^{13}C_{18}$ for the three treatments and both fitness proxies. The hypothesis that we tested is that each provenance would have the trait value maximising fitness in the treatment corresponding to its habitat. The correspondence between treatments and habitats was designed based on known facts on the habitats and the treatments (Control = Hilltop; Drought = Slope; Flooding = Bottomland). All analyses were performed based on maternal family means of raw observed values.

Results

Gravimetric soil water content at the end of the experiment was about 8.93% (sd= 2.90), 19.7% (sd=7.7) and 31.5% (sd=5.68) in the drought, control and water-logged treatment, respectively (**Fig. 1**).

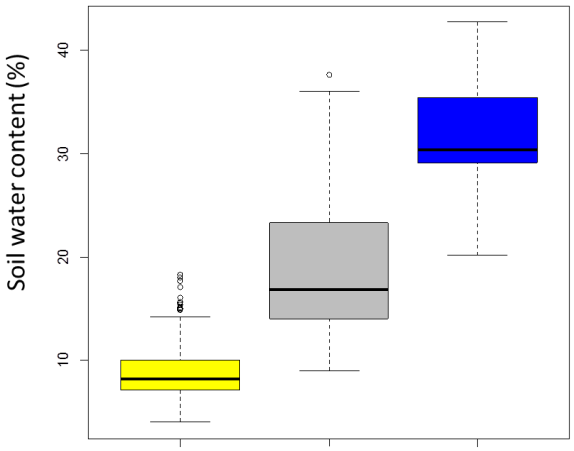


Figure 1: Boxplots of gravimetric soil water content at the end of the experiment in the three treatments: D, water deficit; C, control; F, water-logging.

A synthetic view of the statistical results (**Table 2**) shows that there are significant effects for treatments, provenance and family (progeny) for all traits.

Significant differences were found for all traits among treatments, but much less so among provenances. Direct treatment effects always overrode variations among provenances and variations due to interactions, suggesting that the phenotypic values were more influenced by the treatments than by the genetic structuring into sub-populations. Genotype × Environment interaction terms were nonetheless significant for all biomass traits and for LMA.

Table 2: Summary of the effects of the different factors (Treatment, Provenance, Family, Year) on the different recorded traits.

	$H_{18_{24}}$	$D_{18_{24}}$	TM	LM	RM	SM	LA	LMA_{24}	$LMA_{18_{24}}$	RMR	SMR	LMR	SWCtlp
Provenance	ns	ns	*	*	*	*	*	*	ns	ns	ns	*	ns
Treatment	*	*	*	*	*	*	*	*	*	*	*	*	NA
Interactions	ns	ns	*	*	*	*	ns	ns	ns	ns	ns	ns	NA
Year	ns	*	ns	*	ns	ns	*	*	*	*	ns	*	*
Family	*	*	*	*	*	*	*	*	*	*	*	*	*

Effects of water deficit and water-logging treatments.

In the water deficit treatments, seedlings reached the wilting point at a gravimetric soil water content of about 8.4%. All seedlings survived the treatment.

Water deficit affected growth: growth rate, radial growth, biomass accumulation, and leaf area were smaller than in the controls, **fig. 2 & 3**. Relative biomass allocation significantly differed under water deficit with respect to controls, with a larger shoot mass ratio (SMR) and a lower leaf mass ratio (LMR) in seedlings submitted to water deficit, and no change in root mass ratio (RMR), **fig. 4**. At leaf level, water deficit resulted in a visible increase of leaf mass to area ratio (LMA), **fig. 5**.

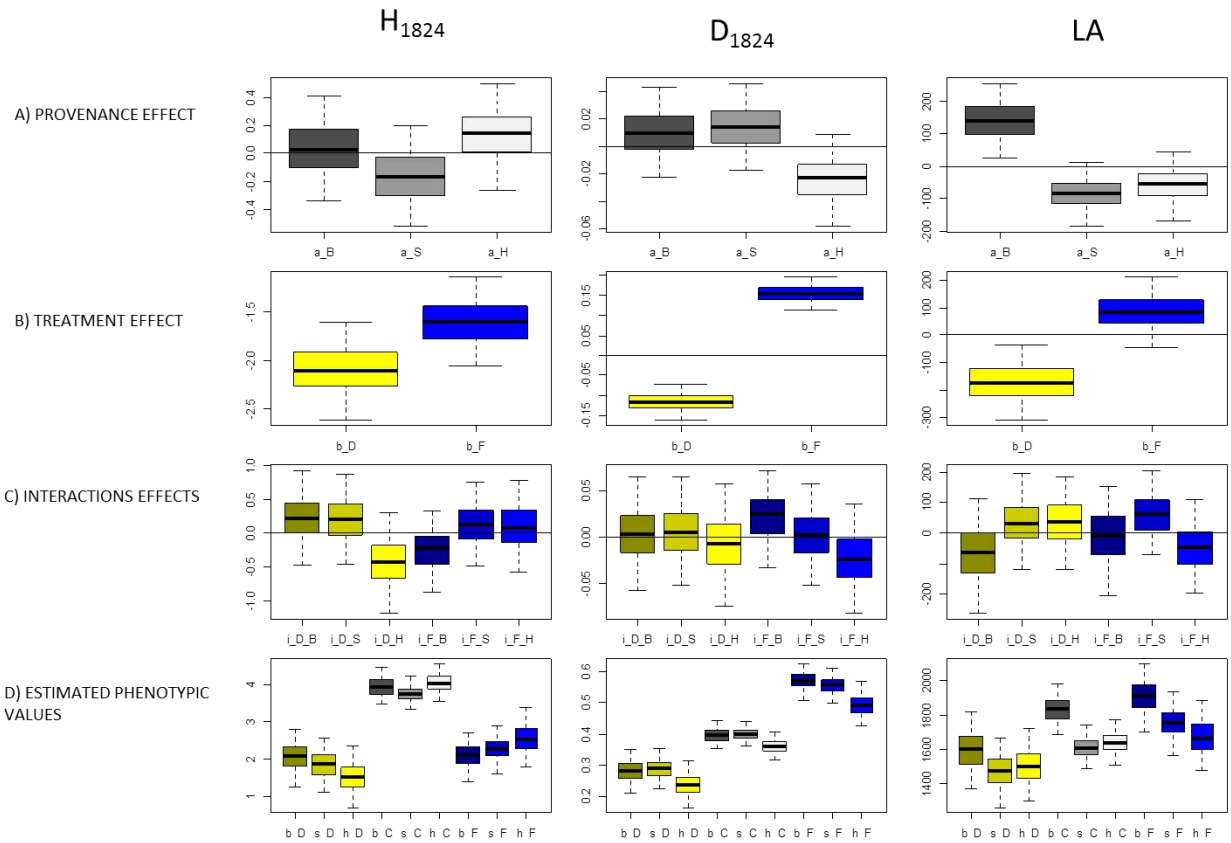


Figure 2: Boxplots of the Bayesian estimates (±CI, p= 0.05) of the height (H_{1824}) and of the diameter (D_{1824}) increment plus leaf area (LA) during the treatments. Units are: cm.month^{-1} for H_{1824}, mm.month^{-1} for D_{1824} and cm² for LA. The panels represent (A) the deviation of the three provenances (a_B: bottomland, a_S: slope, a_H: Hill tops); (B) the deviation from controls displayed in the drought (b_D) and water-logging (b_F) treatments; (C) the interaction effect presented as the deviance between controls and treated plants in the three provenances; and (D) the simulated phenotypic values in each provenance x treatment corresponding to sum sum of provenance + treatment + intecraction effects (from left to right: bottomland, slope and hilltop provenances in drought, bottomland, slope and hilltop in control, and bottomland, slope and hilltop in water-logging).

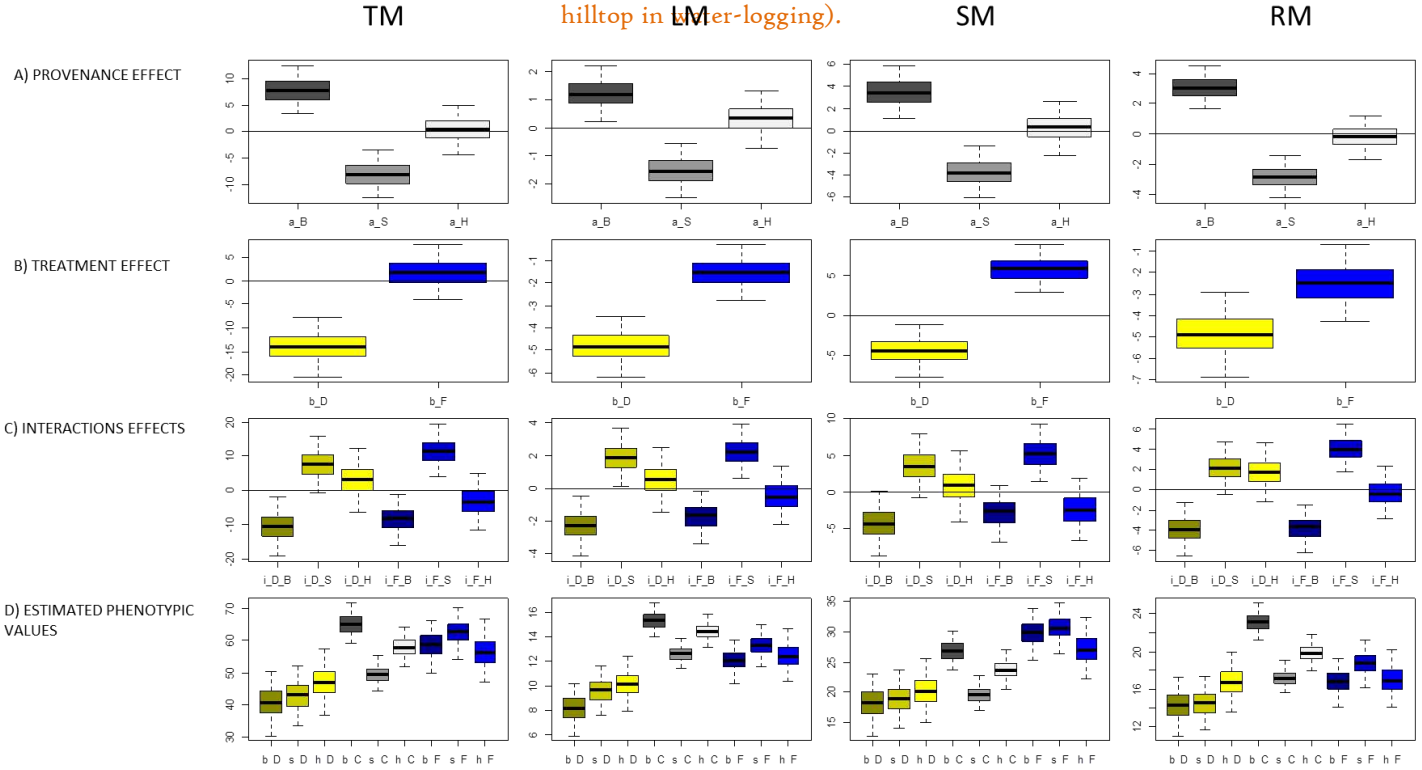


Figure 3: Boxplots of the Bayesian estimates (±CI, p= 0.05) of the total biomass (TM), leaf mass (LM), stem mass (SM), and root mass (RM) at the end of the experiment. Units are grams.

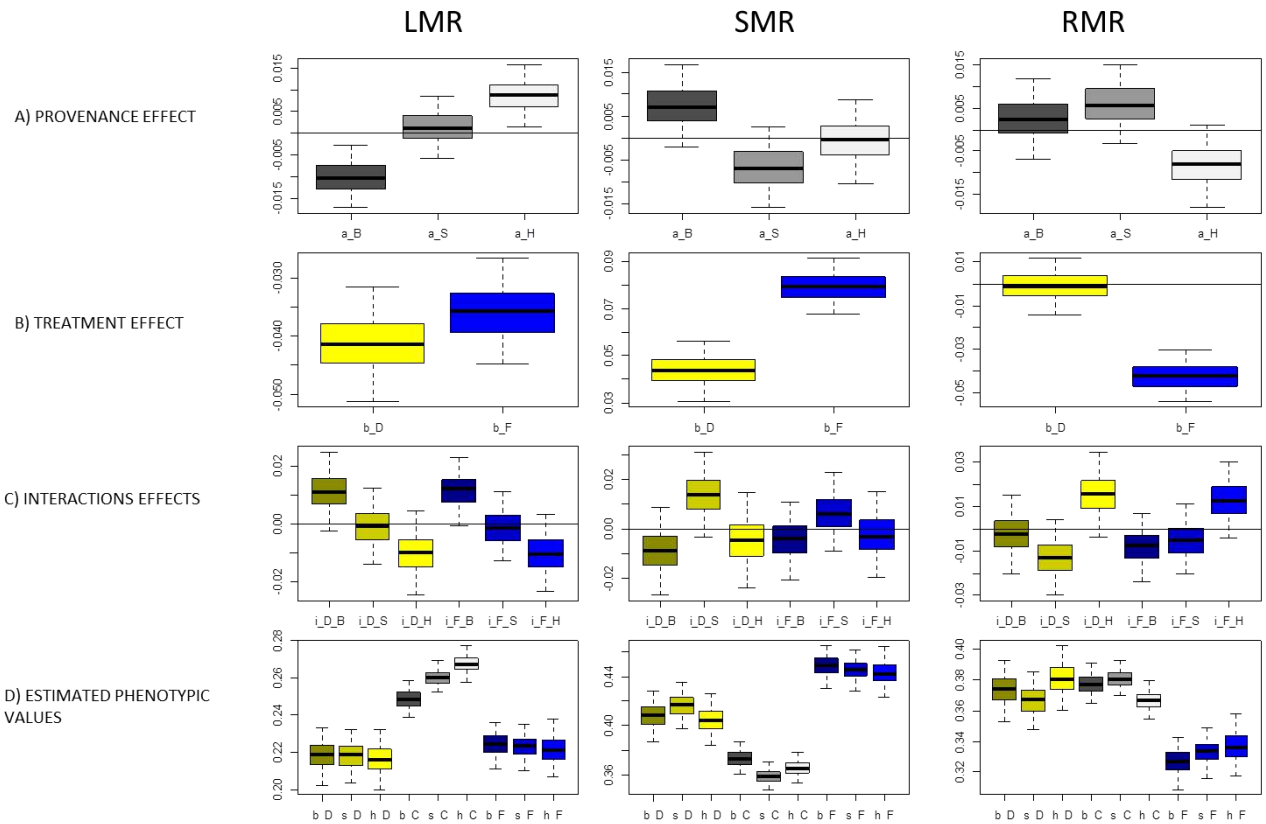


Figure 4: Boxplots of the Bayesian estimates (±CI, p= 0.05) of the leaf to total biomass ratio (LMR), the stem to total biomass ratio (SMR) and the root to total biomass ratio (RMR) at the end of the experiment

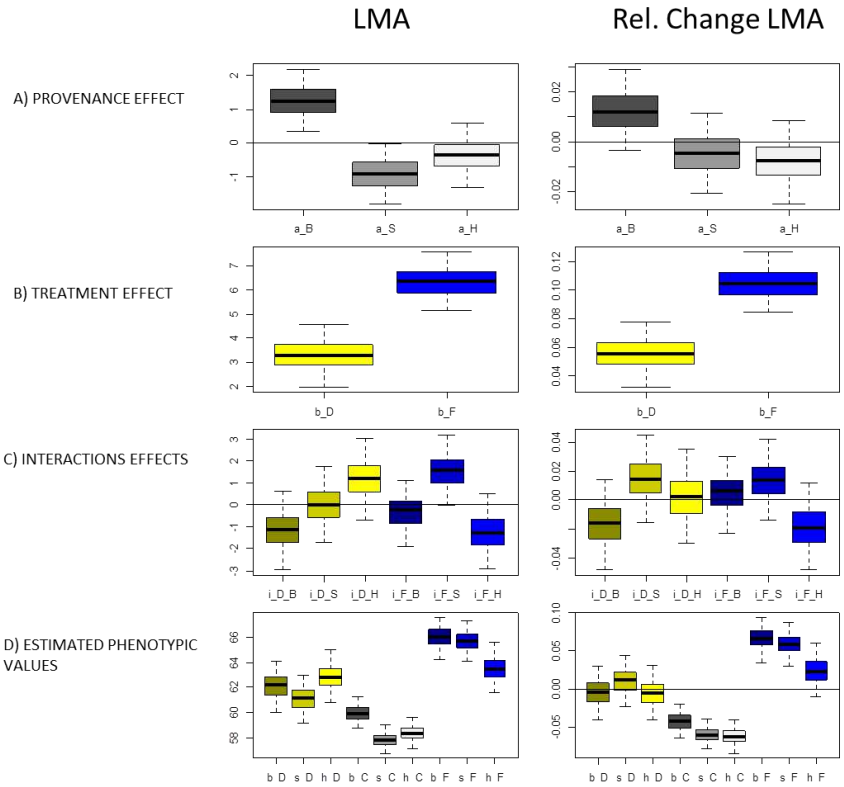


Figure 5: Boxplots of the Bayesian estimates (±CI, p= 0.05) of the leaf mass to area ratio (LMA) and the fraction 13C in leaf biomass ($\delta^{13}C$).

Surprisingly, despite its duration, water-logging had no effect on biomass accumulation. Nevertheless, it induced a significant decrease in stem elongation compensated by a large increase in radial growth, resulting in a significantly higher stem mass. The large increase in stem volume was accompanied by more limited decreases in leaf and root mass. Water-logging also induced significant changes in biomass allocation, with a higher LMR and RMR. Last, water-logging induced the production of adventitious roots in about 59% of water-logged seedlings, **fig. 6**. Water-logging also induced an increase of LMA.

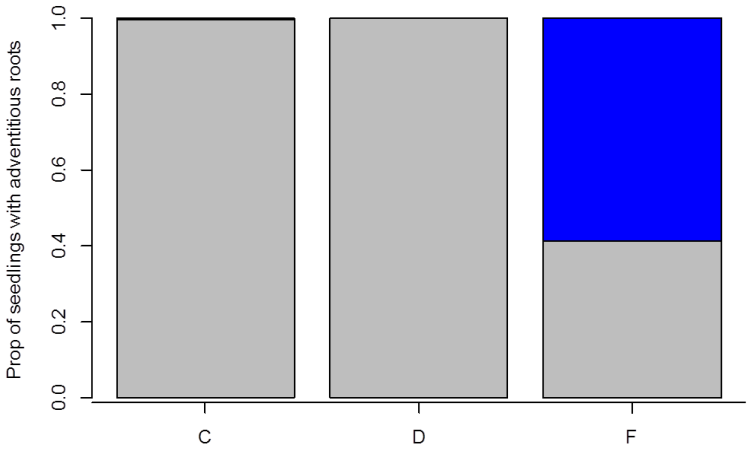


Figure 6: Frequency of the individuals with (blue) and without (grey) adventives roots in the three treatments (C= control, D=drought, F=flood).

Provenance effect

A significant effect of provenance was detected for biomass accumulation in leaves, stems, and roots: seedlings from bottomland displayed significantly higher biomass accumulation and leaf area than seedlings from the two other provenances. However, we did not find significant differences in stem growth rate between 18 and 24 months among provenances. Seedlings from hilltops had higher RMR than seedlings from the two other provenances, whereas seedlings from bottomland had slightly higher SMR. Last, a gradual structuring of LMR among provenance was observed: with seedlings from bottomlands displaying lowest LMR, and seedlings from hilltop the largest LMR.

Interactions effects

All provenances showed similar response to treatments (i.e. no provenance-by-treatment interactions) for the majority of traits.

For biomass accumulation -for which provenance where significantly different in control conditions- provenance-by-treatment interactions were significant but resulted in similar mean biomass accumulation of the different provenance in constraining (drought or water-logged) conditions. Posterior estimates of provenance contributions to phenotypic values revealed that such structuring between provenances was clear in control conditions but disappeared in constrained ones: provenance effects on biomass were comprised

between 59.1 to 71.6 g (95% CI with a median of 65.26g) for seedlings from bottomland, between 44.2 to 55.2 g for seedlings from slopes (median =49.3) and between 51.8 to 64.18g for hilltop provenance. Moreover, estimated leaf area was comprised between 1689 to 1984 cm^2 for seedlings from bottomlands (median=1837 cm^2), 1489 to 1740 for seedlings from slopes (median=1611 cm^2), and between 1506 to 1772 cm^2 for seedlings from hilltops (median=1641). In the drought and water-logging treatment, no significant differences were found between provenances in posterior estimates of phenotypic values (i.e. overlapping 95% credible intervals of simulated phenotypic values) in spite of significant overall provenance effect, see **fig. 2 to 5**.

Variations of phenotypic correlations among treatment and provenances

Trait-trait correlation patterns varied significantly among treatments and among provenances. In water-logged conditions, the correlation between longitudinal and radial growth rates was significantly higher for seedlings from hilltops that for seedlings from the two other provenances. $\delta^{13}C$ was positively correlated with longitudinal growth rate in drought and water-logging conditions only. Correlation between $\delta^{13}C$ and growth rates was significant for seedlings from hilltops in limiting conditions, but not in controls. RMR was negatively correlated with growth rates in all cases except for hilltop seedlings under drought. The correlation between SMR and longitudinal growth rate was not significant in hilltop seedlings under drought, whereas it was significant in all other groups. LMA was negatively correlated with longitudinal growth rate in constraining conditions only (but not for hilltop seedlings in water-logging), and was negatively correlated with radial growth in drought for all provenances and in water-logging for hilltop seedlings only.

Selection gradients

Out of twelve tests (two fitness estimators x two explanatory traits x three treatments), eight displayed significant or marginally significant multiple regression coefficients (**Table 3**). Coefficients were consistent, for a given combination of predicted and explanatory variables, across tests. Linear fitness gradients were consistently negative for LMA and the two quadratic relationships for $\delta^{13}C$ were concave. However, the distributions of the independent variables for the three provenances were largely overlapping in all treatments, thus preventing a test of association between independent and dependent variables as a function of the provenance. However, some trends can be

detected, although they are non-significant. **Fig. 7** shows the relationship between the H_{1824} fitness function and LMA for the three treatments. The Slope provenance has the lowest median for LMA, which corresponds to the highest fitness, in the Drought experiment, whereas the hilltop provenance has the lowest median in the Control experiment; it also has the lowest value (the highest fitness) in the Flooding experiment, contrary to the hypothesis. The plots for the remaining significant fitness gradients are provided as **Suppl. Figure S1**; **Table 3** reports a verbal evaluation of the concordance between trait patterns and the local adaptation hypothesis.

Table 3: Coefficients for the linear (α) and the quadratic (β) terms of the regressions. Only (marginally) significant values are displayed. Rpoxy: trait used as fitness proxy; D: drought; C: control; F: flooding. • : *P*-value between 0.1 and 0.05; *: *P*<0.05; **: *P*<0.01. Concordance with (trait) hypothesis: visual assessment on (non significant) relationships between traits, fitness and provenances; YES/NO indicate clear trends; (YES)/(NO) indicate inconclusive trends.

Proxy	Treatment	α(LMA_{24})	β(LMA_{18})	concordance with LMA hypothesis	α($\delta^{13}C_{18}$)	β($\delta^{13}C_{18}$)	concordance with $\delta^{13}C$ hypothesis
H_{1824}	D	-0.5558**	ns	YES	ns	ns	-
H_{1824}	C	-0.7304*	ns	(YES)	ns	ns	-
H_{1824}	F	-0.5851**	ns	NO	4.4108•	ns	YES
RG_{1824}	D	ns	ns	-	-2.6328•	-0.04785•	(YES)
RG_{1824}	C	-0.009592**	ns	(YES)	-5.4743*	-0.09952*	(NO)
RG_{1824}	F	-0.005797**	ns	NO	ns	ns	-

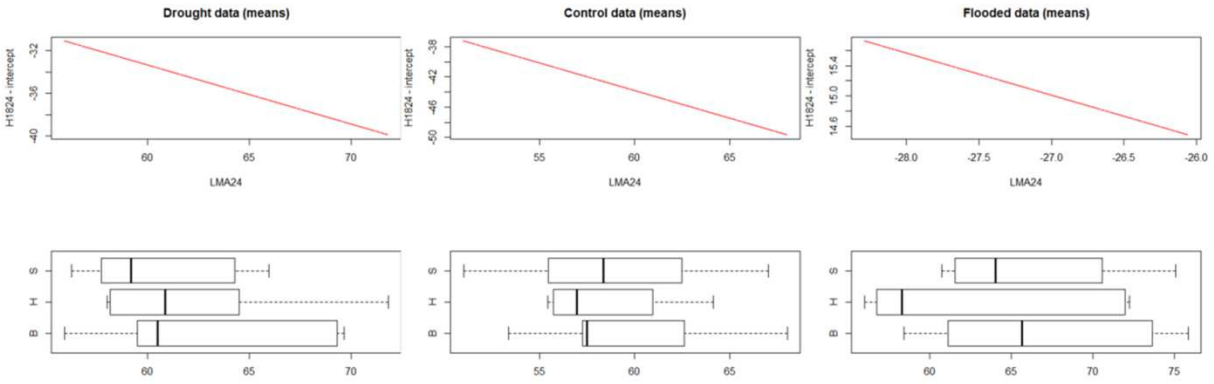


Figure 7: Upper panes: estimated selection gradients; lower panes: box plots of the distribution of predictor variables by provenance in each treatment.

Discussion

All provenances displayed large responses to both limiting conditions relative to controls, as expected. Drought conditions significantly reduced seedling growth, decreased LMR, which has been shown to reduce transpiration and consequently water loss (Chaves *et al.* 2003, Bréda *et al.* 2006), and increased LMA, which contributes to limit water requirements by reducing the leaf surface per unit of mass (SLA) (Wright 2002, Poorter *et al.* 2009). Water-logging induced a shift in stems development: higher SMR, lower RMR and lower LMR by comparison with plants in control conditions. Such variations in stems diametric growth may be explained by changes in bark thickness and/or xylem structure (size and amounts of vessels) (Kozlowski 1997). LMA also increased in flooded conditions, as observed in drought. These variations in leaf structure may be linked with loss of water uptake under water-logging (Poorter *et al.* 2009).

Seedlings from the different provenances displayed both significant differences in their mean phenotypic value for growth and leaf traits (biomass accumulation, leaf area, LMR and LMA), and contrasted growth strategies in particular conditions as revealed by the variations in the strength in the correlations between traits among provenances in several conditions. Moreover, the different provenances were not equally affected by the treatments, as revealed by the significant provenance-by-treatment interactions for several traits. This suggests that evolutionary processes result in heterogeneous distribution of genetic diversity among habitat types (Schemske 1984). Nevertheless, differences among the Bayesian posterior distributions of phenotypic values for the three provenances were clear in control conditions but disappeared in limiting (stressful) conditions, suggesting that significant interactions are due to the absence of phenotypic differentiation among provenances in constraining conditions in spite of significant structuring in controlled conditions. This result is not surprising because non-limiting growth conditions commonly allow the full expression of genetically driven phenotypic divergences, whereas these differences are often hidden in constraining condition due to large environmental effects (Wilson et al. 2006, Sogaard *et al.* 2008, Sogaard *et al.* 2009). For example, seedlings from bottomlands, that produced more biomass than other provenances in control conditions, did not accumulate more biomass during treatments than seedlings from other provenances. This is probably because in water-logging conditions, seedlings with larger biomass would have taken oxygen up from the substrate faster than smaller seedlings and probably suffered from long-term hypoxia.

All provenances had similar growth and leaf properties in constraining conditions (drought or water-logging), suggesting absence of local adaptation to specific stresses, as already described in numerous species (Hereford 2009). Yet, observed trait differences in control conditions cannot be easily attributed to pure drift (Brousseau *et al.* 2013). Differences in trait-trait correlations between provenances, as a function of treatments, are also possibly linked to adaptive responses to environmental stresses (Robinson *et al.* 2009). The analysis of selection gradients showed that traits such as water use efficiency and leaf mass per area have an effect on fitness in all environments, indicating that selective processes operate on seedlings. The distribution of phenotypic values of different provenances largely overlapped, precluding the possibility to identify provenances displaying optimum trait values in each treatment. Nevertheless, trends in the two characters (**Table 3**, **Fig. 7**, **Suppl. Fig. S1**) hint that at least some patterns are compatible with local adaptation. Ongoing reciprocal transplant experiments in natural conditions will help clarify such trends.

Local adaptation mechanisms contribute significantly to the maintenance of genetic and phenotypic diversity (Delph & Kelly 2013), which are the fuel of persistence of plant populations in a changing environment. Global warming models predict large increases in atmospheric CO_2 concentration and temperature by 2100 in Amazonia. No notable change is expected in annual precipitation but changes in their seasonality, leading to increased risks of seasonal flooding or drought in this region (Betts *et al.* 2004, Neelin *et al.* 2006, Galbraith *et al.* 2010). Recent studies, based on both controlled and *in situ* experiments, revealed that large scale drought effects have already occurred across Amazonia, and that the tropical rainforest tree already were severely affected by such drought episodes with important effects on growth, wood production and the resulting carbon sink (Phillips *et al.* 2009, Lola da Costa *et al.* 2010); some of these events have already lead to local tree mortality (Meir and Woodward 2010); the ability of tropical rainforest tree population to adapt to current, rapid climate change will depend on their genetic diversity in climate-related functional traits (Jump 2005, Savolainen 2007, Aiken 2008), that is, in turn, negatively affected by strong selective pressures exerted by these changes. Integrating the role of adaptive and plastic responses as those identified here to community-level models of forest dynamics will lead to better predictions of the impact of global change on forest ecosystems.

Article n°5 – Supplementary figures

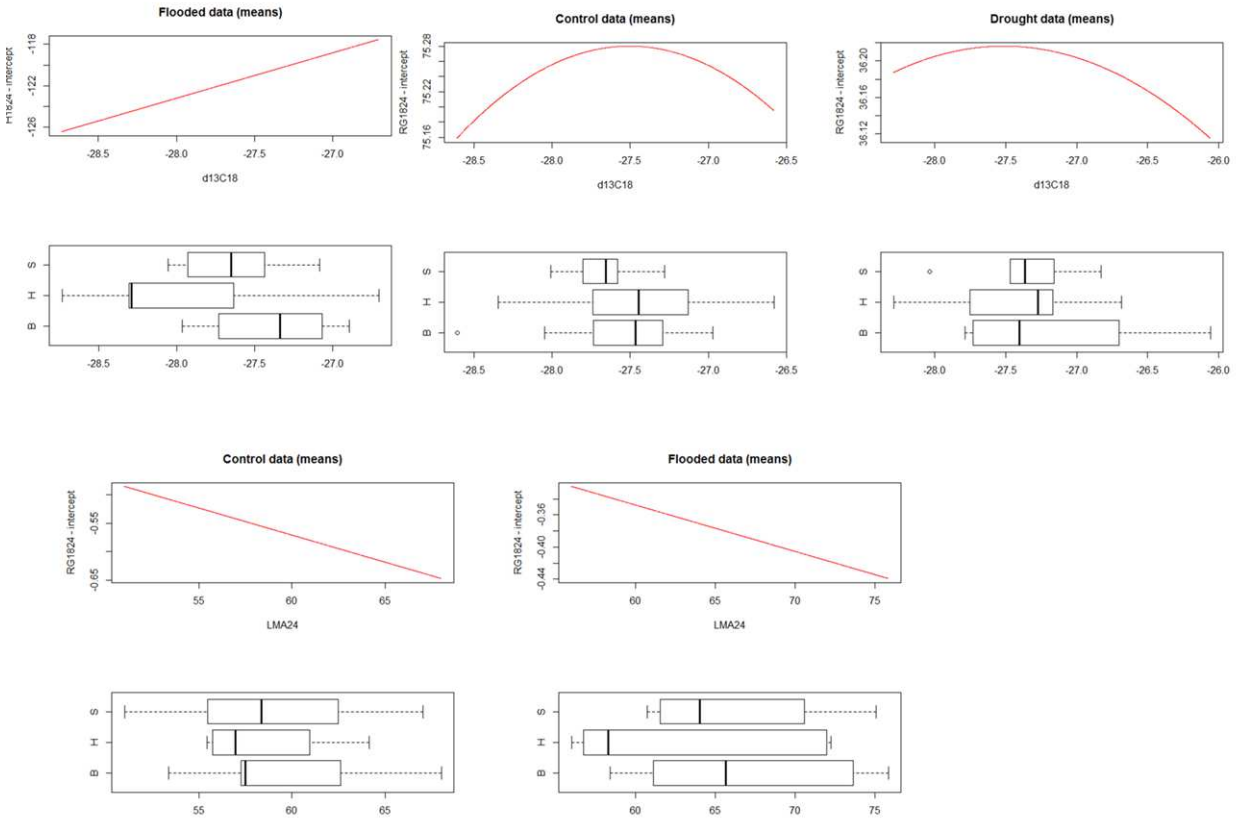


Figure S1: Additional selection gradient plots

Preliminary results - Reciprocal transplants

Reciprocal transplants are probably one of the most suitable experimental strategies to test the hypothesis of local adaptation based on phenotypic traits. This experiment was set-up during the first year of my PhD (2010), and was coupled with molecular analysis (AFLPs), already presented in **article n°2**. The goal of this experiment was to dissociate genetic and environmental sources of phenotypic variations in natural conditions.

This section aims at briefly describing the experiment, and presents the preliminary results. However, the experiment will be continued until 2015 at the earliest.

Material and Methods

Study sites and seed sampling

Seeds were sampled from the two populations of *Eperua falcata* established in the sites of Laussat (x=214508.277 ; y=606318.383 UTM WGS1984) and Regina (x=362876.245 ; y=476934.114 UTM WGS1984) in March 2010.

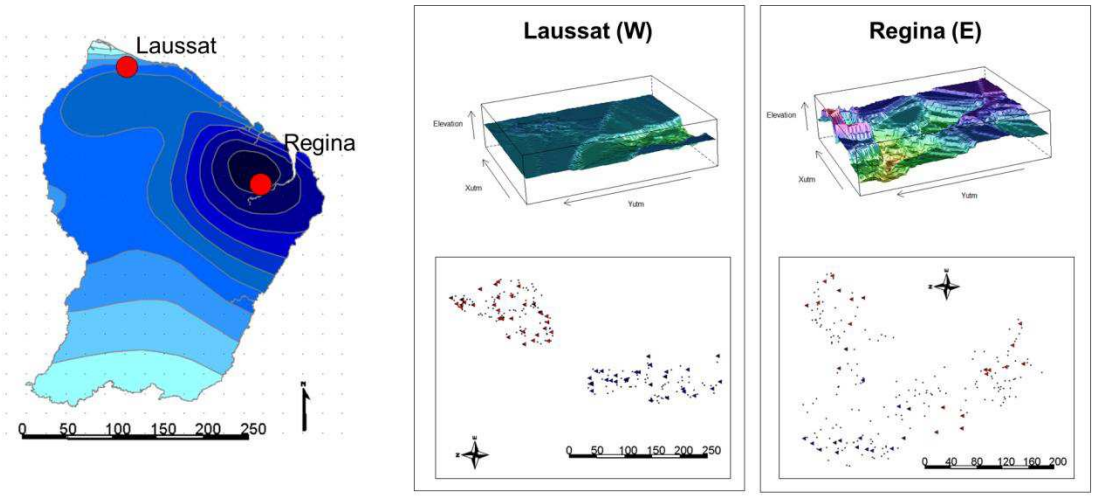


Figure 1: Sampling sites, sampled trees (circles) and seedling (triangles), and topography (interpolated from heterogeneous GPS elevation data).

The two study sites have similar soil properties, but differ in rainfall, with a mean annual precipitation of 2500 mm and 3500 mm in years 2010 and 2011 leading to longer dry periods in Laussat (**Fig. 2**).

The two sites covered different habitat types but had slight topographic differences (**Fig. 1**). Laussat is composed by a permanently water-logged bottomland (elevation= ~38.17 meters) and a plateau of low elevation (~54.65m). Regina is composed by a seasonally water-logged bottomland with flooding events in rainy season (~56.3m), hilltops of high elevation (~84.6m) and important slopes (~69.2m).

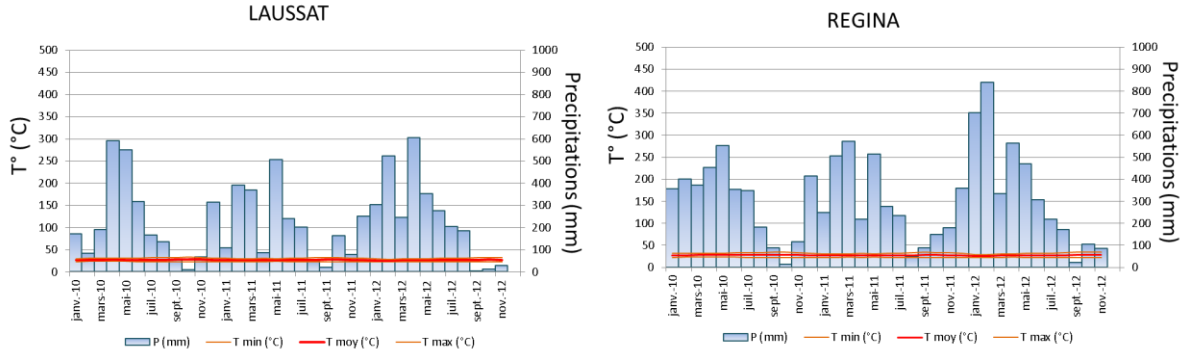


Figure 2: Climate diagrams for the two study sites.

In both sites, hygromorphic soils of bottomlands are characterized by a large accumulation of organic matter until 1 meter traducing that soil formation was dominated by an excess of water. On the contrary, soils from terra-firme are characteristic of well-drained ferralitic soils rich in iron oxides with a sand-clay texture allowing free vertical drainages (**Fig. 3**).

Figure 3: Example of soil toposequences from the bottomland (left) and the plateau (right) of Laussat

All *Eperua falcate* trees of dbh>20cm were mapped within a continuous area of 6.7 ha in Regina (**Fig. 1**). Due to a higher population density in Laussat, sampling was restricted to two areas of 2.5ha and 1.8ha, one in the bottomland and one in the plateau. Population densities ranged from 29.9 adult trees/ha to 48.11 trees/ha in Regina and Laussat respectively.

Seeds were sampled around fructifying mother trees according to a grid layout; the identity of surrounding reproductive trees was recorded for each seed. Seeds were sown in individual pots containing forest soils in a shade-house daily watered until cotyledons emerged.

Seedling growth and transplants

About one month after seeds sowing, young and vigorous seedlings were transplanted into the experimental gardens in the undisturbed forests of Laussat and Regina. Seedlings were randomly distributed in 12 transplant gardens of dimension 10m x 10m (3 gardens within each site and local habitat) with a randomization between blocks within gardens. A total of 813 seedlings were transplanted.

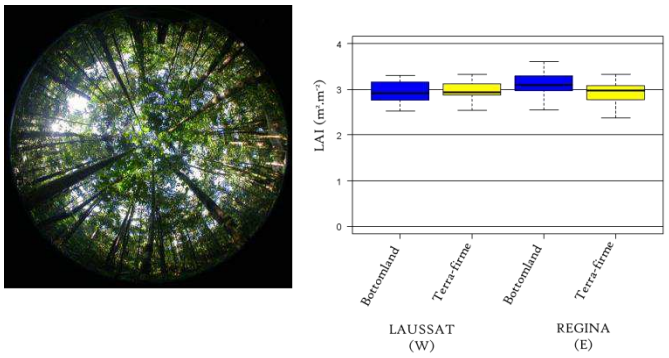


Figure 5: Leaf area index (LAI) estimated in the transplant sites (averaged among gardens, right) through hemispherical photographs (left).

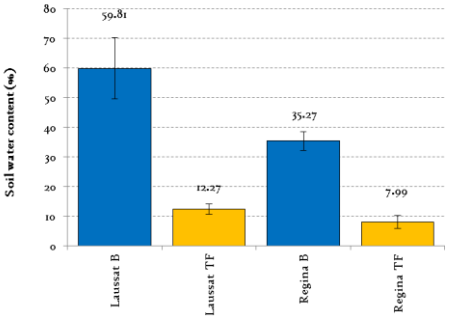


Figure 4: Soil water content in each study site and local habitat (averaged over blocks and over 3 measurements: 09/2012, 04/2012 and 09/2013)

Figure 4 shows soil water content within each local habitat and transplant site. Finally, we checked that light irradiance was similar among the transplant gardens by assessing canopy opening through hemispherical photographs (**Fig. 5**).

Phenotypic traits measurements

Seedlings survival and growth were followed every six months during the two first years after transplants, and will be followed once a year until the end of the experiment. Growth was assessed by measuring height, diameter (and growth rates), the number of growth units, and the number leaves and leaflets. Leaflet area and total leaf area were estimated through the allometric relation between leaflet dimensions and their area (**Fig. 6**). The relation was calibrated by measuring leaf area with a planimeter for a variety of leaflets from un-transplanted seedlings. (Relative) stem elongation rate, diameter growth rate, as well as variations in leaf area between two dates were calculated as follow:

$$\Delta H = \frac{(Ht2 - Ht1)/Ht1}{t2 - t1}$$

Variations of discrete variables between two dates (variations in growth unit numbers, number of leaf, and leaflet production) were assessed by calculating the absolute difference between the two dates.

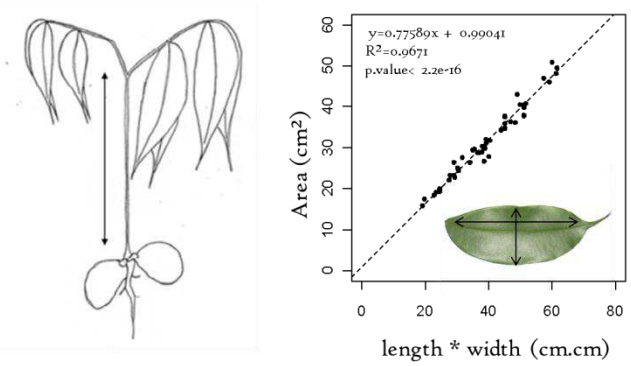


Figure 6: Allometric relation bewteen leaflets dimensions and area.

Several leaf traits will be analyzed. In April 2013, one leaflet per seedling was sampled and leaflet chlorophyll content was assessed using a SPAD. Several disks of leaf tissue (154 mm²) were sampled and dried for further measurements of leaf mass per area (LMA) and leaf nitrogen content. The remaining of each leaflet was frozen and store at -80°C for further genetic investigations. At the end of the experiment, seedlings will be destroyed to measure their dry biomass, and biomass allocation to leaves, stems, and roots.

Statistical analysis

I used a linear model of phenotypic value decomposition similar to those used in the two experiments in glasshouse (**articles n°4 & n°5**) but extended to include the numerous factors tested here. Seedlings phenotypic value was thus decomposed into genetic ('provenance') and environmental effects apprehended at both regional and local scales, with all their interactions. The model was calibrated by a Bayesian method using OPENBUGS, leading to the formalization:

$$\boldsymbol{P \sim N(mean_i, \tau_R)}$$

$$\boldsymbol{mean_i = \mu + a_j + b_k + c_l + d_m + g_{lm} + I(GxG)_{j,k} + I(ExE)_{l,m} + I_{regio}(GxE)_{j,l} + I_{local}(GxE)_{k,m}}$$

$mean_i$ is the individual phenotypic value

τ_R is the residual precision $(1/\sigma_R^2)$

a_j is the effect of the j regional provenance (site)

b_k is the effect of the k local provenance (local habitat)

c_l is the effect of the l regional transplant environment (site)

d_m is the effect of the m local transplant environment (local habitat)

g_{lm} is the effect of the gardens within each l regional transplant and m local habitat

$I(GxG)_{j,k}$ is the "gene-by-gene" interaction between regional and local provenance

$I(ExE)_{l,m}$ is the "environment-by-environment" interaction between regional and local transplant environments

$I_{regio}(GxE)_{j,l}$ is the regional "gene-by-environment" interaction between regional provenance and regional transplant environment

$I_{local}(GxE)_{k,m}$ is the local "gene-by-environment" interaction between local provenance and local transplant environment

Thus, each parameter captures the phenotypic difference from the overall global mean induced by a given level of a given factor. Each parameter was sorted a priori in a non-informative quasi-uniform distribution (a normal distribution centered on o with large variance, i.e. a normal distribution of very small precision). As in classical ANOVA, we used the classical constraint "$\sum \alpha_i = 0$" to make the model identifiable.

For main effects, t consists in defining the effect of a given level as the sum of the other level effects, and in sorting the effects of all other levels in a non-informative normal distribution. For example,

$\boldsymbol{a_{j=1} = -a_{j=2}}$

$\boldsymbol{a_{j=2} \sim N(0, 0.0001)}$

For interactions terms, it consists in fixing interactions at o for one factor (the factor is set as a reference), and setting "$\sum \alpha_i = 0$" for the other factors. For example, for the interaction $I_{local}(GxE)_{k,m}$

Provenance bottomland transplanted into bottomland:

$I_{local}(GxE)_{k=1,m=1} = 0$

Provenance bottomland transplanted into hilltop:

$I_{local}(GxE)_{k=1,m=2} = 0$

Provenance hilltop transplanted into bottomland:

$I_{local}(GxE)_{k=2,m=1} = -I_{local}(GxE)_{k=2,m=2}$

Provenance hilltop transplanted into hilltop:

$I_{local}(GxE)_{k=2,m=2} \sim N(0, 0.00001)$

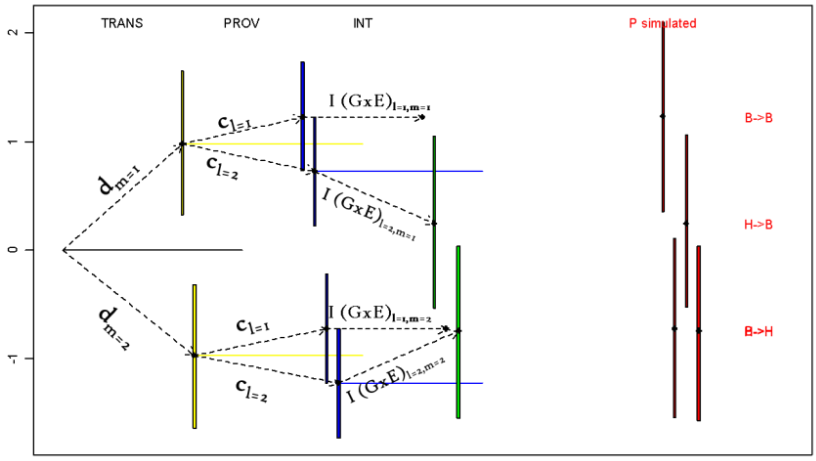


Figure 7: Phenotypic value decomposition into 'provenance', 'transplant environment' and their gene-by-environment interactions, focusing on local scale.

Because, the measured characters include different kinds of data (continuous such as height and diameter, discrete such as the number of leaves, or binary such as the survival), we used different statistical distributions to draw prior distributions of phenotypic values: a normal distribution for continuous characters, a Poisson

distribution for discrete characters, and a Bernoulli distribution for survival. The difference of number of leaves between two dates (differences between two Poisson distributions) was drawn in a Skemall distribution. Complete BUGS codes are available in supplementary materials.

This approach is very powerful for resolving complex linear models with numerous factors involved, by finely dissecting phenotypic value into interesting effects free from others sources of variations. F**igure 7** shows how phenotypic values were partitioned, focusing on local effects.

Preliminary results and brief discussion

Provenance and transplant main effects were significant for numerous traits. **Table 1** summarizes the significance of each factor for each phenotypic trait. Detailed figures of parameters estimated for main effects are provided in supplementary figures S1 to S3.

A significant effect of the regional site of provenance was found for survival, height, stem elongation rate, diameter growth rate, leaflet production and leaflet. Until 17 months after transplant, seedlings from Laussat had a greater survival probability than seedlings from Regina, but this effect became non-significant at 17 months, due to higher death rates in seedlings from Laussat between 12 and 17 months. From 5 months, seedlings from Laussat were smaller than seedlings from Regina. These differences appeared before at early-developmental stages, before the first measurement and were maintained until 17 months. Seedlings from Laussat had, however, a higher growth rate between 5 and 12 months; this effect disappeared between 12 and 17 months, and was unsufficient to inhibit differences in stems elongations installed at early-stages. Total diameter growth rate estimated from 5 to 17 months were slightly higher in seedlings from Regina, but this effect was insufficient to lead to significant differences in diameter at 17 months between the two regional provenances. The number of leaflets produced was equal whatever the provenance. Leaflet production between 5 and 12 months was slightly higher in seedlings from Laussat but this effect became non-significant between 12 and 17 months, and was not detected on leaf production between 5 and 17 months. Seedlings from Regina had slightly higher leaflet area at 5 and 12 months, but any differences in total leaf area were detected at any date.

Regional transplant site had significant effects on survival and leaf production. From 5 months, seedlings transplanted in Laussat had a greater survival probability than seedlings transplanted in Regina. Seedlings transplanted into Laussat produced more leaves between 12 and 17 months, but this did not lead to more leaves at 17 months and this effect was not detected by analyzing total leaf production between 5 and 17 months.

Local native habitat was significant for survival, total number of leaflets and leaflet production. Seedlings from bottomland had a lower survival probability than seedlings from hilltops, and produced more leaflets between 5 and 12 months, leading to more leaflets at 12 and 17 months. Without differences in the total number of leaves, in total leaf area, and in the average leaflet area, it is probable that a combination of a slightly more leaflets per leaves and a slightly lower leaflet area in seedlings from bottomland would had led to similar total leaf area among local provenances.

Local transplant environment had significant effects on several growth traits: height, stem elongation rate, diameter, growth unit development (GU and ΔGU), leaf production (NL and ΔNL), leaflet area and leaflet production (la and Δla). At 5 months, seedlings transplanted into bottomland were taller than seedlings transplanted onto hilltops. However, stem elongation rates became lower in bottomland between 12 and 17 months, leading to similar seedlings height whatever the local habitat at 12 and 17 months.

From 5 months, seedlings transplanted into bottomlands had a higher diameter than seedlings from hilltops. No difference in diameter growth was detected between 5 and 17 months, leading to the maintenance of differences between provenances. Seedlings transplanted onto hilltops produced more growth units between 12 and 17 months than seedlings transplanted into bottomland. This resulted in seedlings with more growth units (but similar height) in hilltops by comparison with bottomlands. Seedlings transplanted into bottomland had a higher leaf area without significant differences, neither in the number of leaflets, nor in leaflet areas. Seedlings from bottomland produced fewer leaves and leaflets between 12 and 17 months, leading to fewer leaves and leaflets at 17 months. From 12 months, leaf area became similar between local provenances. Because *Eperua falcata* is widespread in water-logged habitats, we expected that survival probability would be higher in bottomland. Surprisingly, survival probability was equal in both habitat types. Furthermore, bottomland habitat had a negative effect on the majority of growth traits (except for diameter), suggesting that seedlings grew better on hilltops than in bottomlands.

No 'gene-by-environment' (I_{GxE}), 'gene-by-gene' (I_{GxG}) or 'environment-by-environment' (I_{ExE}) interaction was significant. This suggests that the effect of local native habitat was similar for the two regional provenances (I_{GxG}), the effect of transplant habitat was similar in the two transplant sites, and that (regional or local) transplant environments had the same effect whatever the (regional or local) provenance (I_{GxE}).

Moreover, several factors seem to have increasing effects over time, whereas other factors' effects tend to disappear. For example, the effect of regional provenance became non-significant at 17 months for survival, and from 12 months for stem elongation rate, leaf production, and leaflet area. For these traits, differences among groups are equalized over time. For provenance effects, the disappearance of significant effects over time may reveal the existence of apparent genetic effects, probably due to maternal effects (such as seed quality). On the contrary, regional provenance effect became significant for total diametric increment from 5 to 17 months. In the same way, the effect of regional transplant site on

survival became significant at 12 months. Differences in stem elongation rate, leaves and leaflet productions (NL, ΔNL, Nl, and ΔNL) between local transplant site became significant from 12 months, and differences in NGU became significant at 17 months that was associated to a significant variation in total ΔNGU between 5 and 17 months.

Even if significant, main factors effects were often small (i.e. they induced small deviance from the overall global mean, μ), thus suggesting that both provenance and transplants effects little affect seedlings growth. Moreover, phenotypic differences among local provenances were less clear that those observed in non-limiting conditions (**Article n°4**). Several causes may be advanced here:

(1) Seedlings grew necessarily more slowly in natural conditions than in non-limiting ones: at 5 months, seedlings transplanted into natural conditions measured about 16.94 cm high, against 33.05 cm at 6 months in non-limiting conditions. At 12 months, seedlings transplanted in field measured 22.07 cm against 55.13 cm in non-limiting conditions. At 17 months, seedlings that grew in the field measured only 26.19 cm against 77.02 cm for seedlings that grew in the shade-house. Thus, at the same age, seedlings that grew in field and in shadehouse are not at the same ontogenetic stage. The relative effects of maternal, genetic and environmental sources of phenotypic variations are known to vary across ontogenetic stages. Maternal effects are expected to decrease over time, whereas both truly genetically- and environmentally-driven phenotypic differences among seedling groups would appear after a sufficient time allowing sufficient differences. Thus, the (small) seedlings analyzed here are probably too young to detect large effects, and traits that were not significant in the first measurements and that became significant will probably become more significant. Moreover, seedlings growth at early-ontogenetic stages is probably confused with a part of maternal effects. This expectation is supported by several traits for which a significant 'provenance' effect became not significant through time. In a preliminary analysis, I included seed fresh mass as a co-factor in the model. This effect was removed because not significant, thus suggesting that variation in seed mass did not significantly affect seedlings growth. However, the model did not investigate how seed mass may be structured among regional and local provenances (a part of the provenance effects may hide the effect of seed mass differences among provenances). Moreover, seed mass describes only

part of maternally-induced seed differences: seed quality is also an important point (i.e. mother trees inhabiting a nutrient-rich environment would produce seeds of better quality than mother trees inhabiting poor environments) that cannot be taken into account, because analyzing seeds quality is a destructive method that prevents seed sowing). For these reasons, it is necessary to follow seedlings growth over a longer time period to draw conclusions about this experiment.

(2) Common garden experiments in non-limiting conditions experiments allow the full expression of genetically-driven phenotypic divergences among groups by minimizing environmental sources of variations. In natural conditions, where the factors influencing seedlings growth are multiple and complex, large environmental variability may completely hide inherent phenotypic divergences among provenances. In particular, environmental heterogeneity associated with regional and local habitat transplant environment, coupled with a large variability among gardens and probably numerous other factors not taken in account (included into residual variability) may induce large phenotypic variations and mask the inherent structuring of phenotypic traits observed in common garden.

(3) In the common garden experiment, the largest phenotypic divergences among provenances were detected for seedlings biomass, which is an integrator of all above-ground and below-ground compartments. Here, we lack information about the most informative trait. Up to now, above-ground biomass may not be estimated through the measured traits (stem dimensions and leaf area without knowledge of leaf mass per area). Moreover, we have no idea about the below-ground biomass of seedlings. Both below-and above-ground biomass, as well as biomass allocation to leaves, stems and roots will be assessed at the end of the experiment.

Table 1: Measured phenotypic traits (with abbreviations and units) and synthetic results of phenotypic value decomposition.
"*" indicates that the factor is significant, i.e. the parameters estimated for the different levels are not overlapping, based on 95% credible intervals.
L: Laussat, R: Regina, B: bottomland, H: hilltop.

				Regional provenance	Regional transplant environment	GxE interactions at regional scale	Local provenance	Local transplant environment	GxE interactions at local scale	GxG interactions	ExE interactions	Gardens
Trait	Abbrev	Unit	Model	a_j	c_l	$I_{regio}(GxE)_{jl}$	b_k	d_m	$I_{local}(GxE)_{km}$	$I(GXG)_{jk}$	$I(EXE)_{lm}$	g_{lm}
Survival	S_t		Bernoulli	* (L>R)	ns	ns	* (B<TF)	ns	ns	ns	ns	ns
	S_5		Bernoulli	* (L>R)	* (L>R)	ns	* (B<TF)	ns	ns	ns	ns	*
	S_{11}		Bernoulli	* (L>R)	* (L>R)	ns	* (B<TF)	ns	ns	ns	ns	*
	S_{17}		Bernoulli	ns	* (L>R)	ns	* (B<TF)	ns	ns	ns	ns	*
Height	H_5	cm	gaussian	* (L<R)	ns	ns	ns	* (B>TF)	ns	ns	ns	*
	H_{11}		gaussian	* (L<R)	ns	ns	ns	ns	ns	ns	ns	*
	H_{17}		gaussian	* (L<R)	ns	ns	ns	ns	ns	ns	ns	*
Stem elongation rate	$\Delta H_{5:11}$	$cm.cm^{-1}.month^{-1}$	gaussian	* (L>R)	ns	ns	ns	ns	ns	ns	ns	*
	$\Delta H_{11:17}$		gaussian	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	ns
	$\Delta H_{5:17}$		gaussian	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	*
Diameter	D_5	mm	gaussian	ns	ns	ns	ns	* (B>TF)	ns	ns	ns	*
	D_{11}		gaussian	ns	ns	ns	ns	* (B>TF)	ns	ns	ns	*
	D_{17}		gaussian	ns	ns	ns	ns	* (B>TF)	ns	ns	ns	*
Diameter growth rate	$\Delta D_{5:11}$	$mm.mm^{-1}.month^{-1}$	gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*
	$\Delta D_{11:17}$		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*
	$\Delta D_{5:17}$		gaussian	* (L<R)	ns	ns	ns	ns	ns	ns	ns	*
Number of growth units	NGU_5		Poisson	ns	ns	ns	ns	ns	ns	ns	ns	*
	NGU_{11}		Poisson	ns	ns	ns	ns	ns	ns	ns	ns	*
	NGU_{17}		Poisson	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	*
Variation of growth unit number	$\Delta NGU_{5:11}$		Skemmall	ns	ns	ns	ns	ns	ns	ns	ns	ns
	$\Delta NGU_{11:17}$		Skemmall	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	*
	$\Delta NGU_{5:17}$		Skemmall	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	*
Number of leaves	NL_5		Poisson	ns	ns	ns	ns	ns	ns	ns	ns	ns
	NL_{11}		Poisson	ns	ns	ns	ns	ns	ns	ns	ns	ns
	NL_{17}		Poisson	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	ns
Leaf production	$\Delta NL_{5:11}$		Skemmall	ns	* (L>R)	ns	ns	ns	ns	ns	ns	*
	$\Delta NL_{11:17}$		Skemmall	ns		ns	ns	* (B<TF)	ns	ns	ns	*
	$\Delta NL_{5:17}$		Skemmall	ns		ns	ns	* (B<TF)	ns	ns	ns	*
Leaf area	LA_5	cm^2	gaussian	ns	ns	ns	ns	* (B>TF)	ns	ns	ns	*
	LA_{11}		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	ns
	LA_{17}		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	ns
Leaf area variations	$\Delta LA_{5:11}$	$cm^2.cm^{-1}.month^{-1}$	gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*
	$\Delta LA_{11:17}$		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*
	$\Delta LA_{5:17}$		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*
Number of leaflets	NI_5		Poisson	ns	ns	ns	ns	ns	ns	ns	ns	*
	NI_{11}		Poisson	ns	ns	ns	* (B<TF)	ns	ns	ns	ns	*
	NI_{17}		Poisson	ns	ns	ns	* (B<TF)	* (B<TF)	ns	ns	ns	*
Leaflet production	$\Delta NI_{5:11}$		Skemmall	* (L>R)	ns	ns	* (B<TF)	ns	ns	ns	ns	*
	$\Delta NI_{11:17}$		Skemmall	ns	ns	ns	ns	* (B<TF)	ns	ns	ns	*
	$\Delta NI_{5:17}$		Skemmall	ns	ns	ns	ns	ns	ns	ns	ns	*
Leaf area per leaflet	LaI_5	cm^2	gaussian	* (L<R)	ns	ns	ns	ns	ns	ns	ns	*
	LaI_{11}		gaussian	* (L<R)	ns	ns	ns	ns	ns	ns	ns	*
	LaI_{17}		gaussian	ns	ns	ns	ns	ns	ns	ns	ns	*

Perspectives

Phenotypic traits measurements

Until the end of the experiment, seedling growth will be followed every year. We intend to extend the span of the recorded traits by including several leaf traits (leaf carbon, nitrogen, and chlorophyll content, LMA) and by including seedlings biomass and biomass allocation ratios.

Environmental characterization

We also envisage improving environmental characterization, and particularly soil water content properties. Because soil water content varies widely between seasons and days, it is not a good estimator of soil water constraints when measured instantaneously. Due to technical constraints, we could not set up permanent sensors to automatically measure soil water content at regular time intervals. Instead, we propose to finely analyze soil granulometry, and to use soil-climate models that correlate rainfall and soil water content depending on soil structure and texture.

Genetic analyses:

The phenotypic approach will be complemented by fine genetic analyses on both adult tree populations and transplanted seedlings. We intend to develop a genome scan approach through high-throughput genotyping or re-sequencing of a large variety of expressed sequenced (chosen among the unigenes described from 454 analysis, **Article n°3**), including both candidate genes and randomly chosen loci. Genetic data will be used to carry out both population genetics and association studies. Matching genetic data from adult trees and from seedlings will allow us to assess seedling relatedness. Association genetics will allow identifying loci linked with quantitative traits (i.e. loci for which allelic state is correlated with quantitative traits). Finally, high-throughput re-sequencing or genotyping of characterized unigenes (expressed genes with known function, including several candidate genes) will allow extending the genome scan approached initially developed on anonymous (AFLPs) markers (**Article n°2**) and will lead to more precise conclusions about the identity of genes targeted by divergent selection across habitat types.

Comparing results with other tree species:

In Parallel to this PhD, the reciprocal transplant experiment involves a variety of biological tree species with different degrees of genetic divergence, and particularly two *Carapa* species with complete reproductive isolation, and two *Symphonia* species organized in a species complex with partial reproductive isolation. With the inclusion of *Eperua falcata*

sub-populations, this experiment will allow to dissect environmental and genetic sources of phenotypic variations across different levels of genetic differentiation (i.e. different stages of speciation process): from intra-specific differentiation to complete isolation between close species.

Preliminary results – Supplementary materials

Linear model of phenotypic value decomposition

```
### Likelyhood for continuous traits

for (i in 1:Ndata){

Trait[i]~dnorm(mean[i],tau_R)

mean[i]<-mu+a[Prov_Geo[i]]+b[Trans_Geo[i]]+c[Prov_Topo[i],j]+d[Trans_Topo[i],j]+
i_GxG[Prov_Geo[i],Prov_Topo[i],j]+i_ExE[Trans_Geo[i],Trans_Topo[i]]+
i_GxE_geo[Prov_Geo[i],Trans_Geo[i]]+i_GxE_topo[Prov_Topo[i],Trans_Topo[i]]+
g[Garden[i]]

}

### Likelyhood for discrete traits

for( i in 1:Ndata ) {

Trait[i,j]~dpois(mean[i,j])

log(mean[i,j])<-mu+
a[Prov_Geo[i]]+b[Trans_Geo[i]]+c[Prov_Topo[i],j]+d[Trans_Topo[i],j]+
i_GxG[Prov_Geo[i],Prov_Topo[i],j]+i_ExE[Trans_Geo[i],Trans_Topo[i]]+
i_GxE_geo[Prov_Geo[i],Trans_Geo[i]]+i_GxE_topo[Prov_Topo[i],Trans_Topo[i]]+
g[Garden[i]]

}

### Likelyhood for binary traits

for( i in 1:Ndata ) {

Trait[i]~dbern(p[i])
logit(p[i])<-F[i]
F[i]~dnorm(mean[i],1)
mean[i]<-
a[Prov_Geo[i]]+b[Trans_Geo[i]]+c[Prov_Topo[i],j]+d[Trans_Topo[i],j]+
i_GxG[Prov_Geo[i],Prov_Topo[i],j]+i_ExE[Trans_Geo[i],Trans_Topo[i]]+
i_GxE_geo[Prov_Geo[i],Trans_Geo[i]]+i_GxE_topo[Prov_Topo[i],Trans_Topo[i]]+
g[Garden[i]]

}

### Likelyhood for traits based on Skemall distribution (differences between two
Poisson)

Cst <- 10000 # this just has to be large enough to ensure all p[i]'s < 1
UNIFLIM <- 100

for (i in 1:Ndata){
Delta_Trait[i]<-Trait_t2[i]-Trait_t1[i]
zeros[i] <- 0
zeros[i] ~ dpois(zeros.mean[i])
zeros.mean[i] <- -loglike[i] + Cst


loglike[i] <-
```

```
-lambda1[i] + Theta1[i]*log(lambda1[i]) - logfact(Theta1[i])
-lambda2[i] + Theta2[i]*log(lambda2[i]) - logfact(Theta2[i])

log(lambda1[i])<-
mu+a[Prov_Geo[i]]+b[Trans_Geo[i]]+c[Prov_Topo[i]]+d[Trans_Topo[i]]+
i_GxG[Prov_Geo[i],Prov_Topo[i]]+i_ExE[Trans_Geo[i],Trans_Topo[i]]+
i_GxE_geo[Prov_Geo[i],Trans_Geo[i]]+i_GxE_topo[Prov_Topo[i],Trans_Topo[i]]+
g[Garden[i]]

log(lambda2[i])<-
mu+a[Prov_Geo[i]]+b[Trans_Geo[i]]+c[Prov_Topo[i]]+d[Trans_Topo[i]]+
i_GxG[Prov_Geo[i],Prov_Topo[i]]+i_ExE[Trans_Geo[i],Trans_Topo[i]]+
i_GxE_geo[Prov_Geo[i],Trans_Geo[i]]+i_GxE_topo[Prov_Topo[i],Trans_Topo[i]]+
g[Garden[i]]

Theta1[i]<- Theta3[i] * (Delta_Trait[i]+Theta4[i]) + (1-Theta3[i]) * Theta4[i]
Theta2[i]<- Theta3[i] * Theta4[i]+ (1-Theta3[i]) * (-Delta_Trait[i]+Theta4[i])
Theta3[i]<-step(Delta_Trait[i])
Theta5[i]~dunif(0, UNIFLIM)
Theta4[i]<-trunc(Theta5[i])

            }

### Priors definition

# global mean and residuals
mu[j]~dnorm(0,0.0001)     # not for discrete traits
tau_R[j]~dgamma(0.01,0.01) # residual precision (only for continuous traits)

# main effects
a[1]<-(-a[2]) ; a[2]~dnorm(0,0.0001)
b[1]<-(-b[2]) ; b[2]~dnorm(0,0.0001)
c[1]<-(-c[2]) ; c[2]~dnorm(0,0.0001)
d[1]<-(-d[2]) ; d[2]~dnorm(0,0.0001)

# Gene-by-gene interactions
i_GxG[1,1]<-0 ; i_GxG[1,2]<-0
i_GxG[2,1]<-(-(i_ GxG [2,2])) ; i_GxG[2,2]~dnorm(0,0.0001)

# Environment-by-environment interactions
i_ExE[1,1]<-0 ; i_ ExE[1,2]<-0
i_ ExE[2,1]<-(-(i_ExE[2,2])) ; i_ ExE[2,j]~dnorm(0,0.0001)

# Gene-by-environment interactions (regional scale)
i_GxE_geo[1,1]<-0 ; i_GxE_geo[1,2]<-0
i_GxE_geo[2,1]<-(-(i_GxE_geo[2,2])) ; i_GxE_geo[2,2]~dnorm(0,0.0001)

# Gene-by-environment interactions (local scale)
i_GxE_topo[1,1]<-0 ; i_GxE_topo[1,2]<-0
i_GxE_topo[2,1]<-(-(i_GxE_topo[2,2])) ; i_GxE_topo[2,2]~dnorm(0,0.0001)

# Gardens effects (with 3 gardens within each regional site and local habitat)
g[1]<-(-(g[2]+g[3])) ; g[2]~dnorm(0,0.0001) ; g[3]~dnorm(0,0.0001)
g[4,j]<-(-(g[5]+g[6])) ; g[5]~dnorm(0,0.0001) ; g[6]~dnorm(0,0.0001)
g[7,j]<-(-(g[8]+g[9])) ; g[8]~dnorm(0,0.0001) ; g[9]~dnorm(0,0.0001)
g[10,j]<-(-(g[11]+g[12])) ; g[11]~dnorm(0,0.0001) ; g[12]~dnorm(0,0.0001)
```

Preliminary results – Supplementary figures

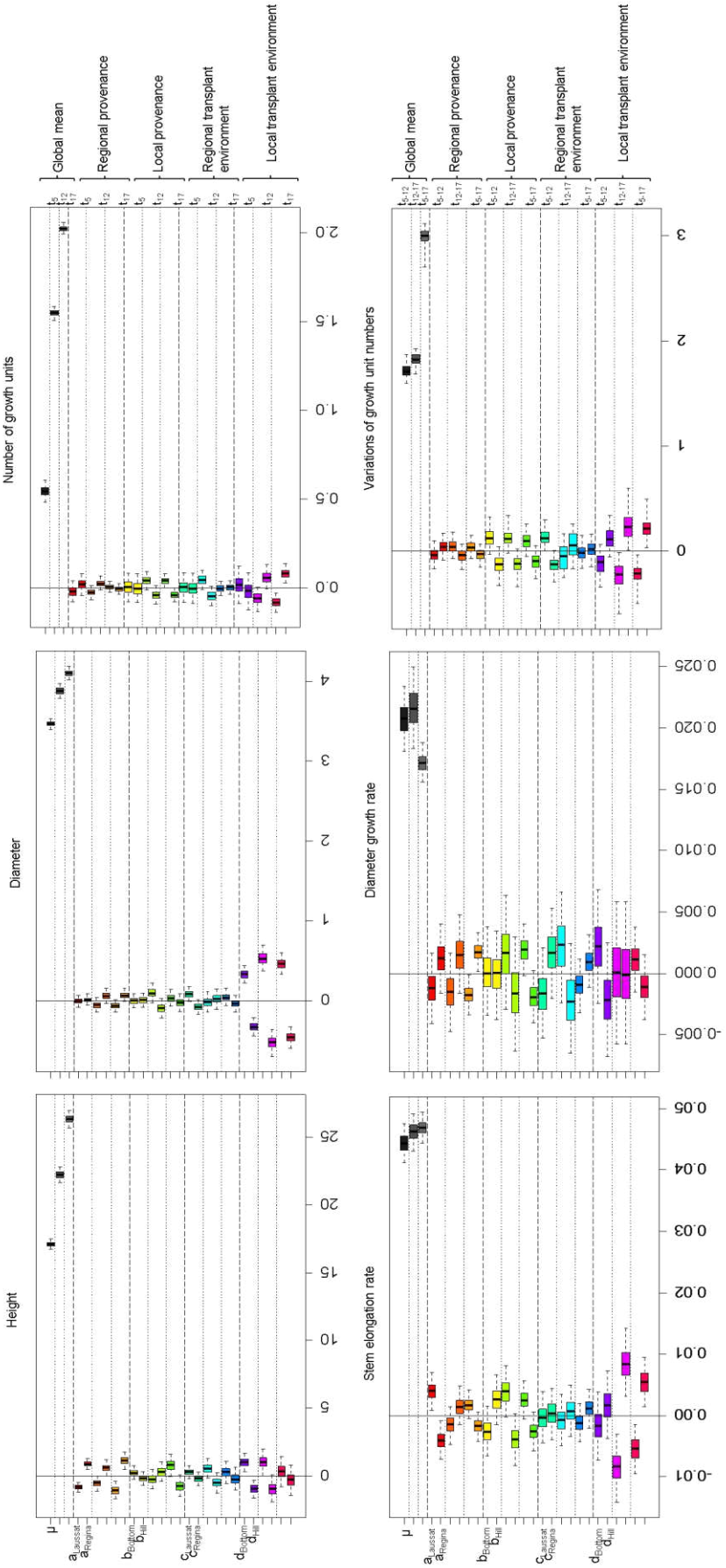


Figure S1 : Global mean and main effects estimations for characters relative to stem growth. *μ*: global mean, a: Regional provenance effects, b: Local provenance effects, c: Regional transplant site effect, d: regional transplant environment effects).

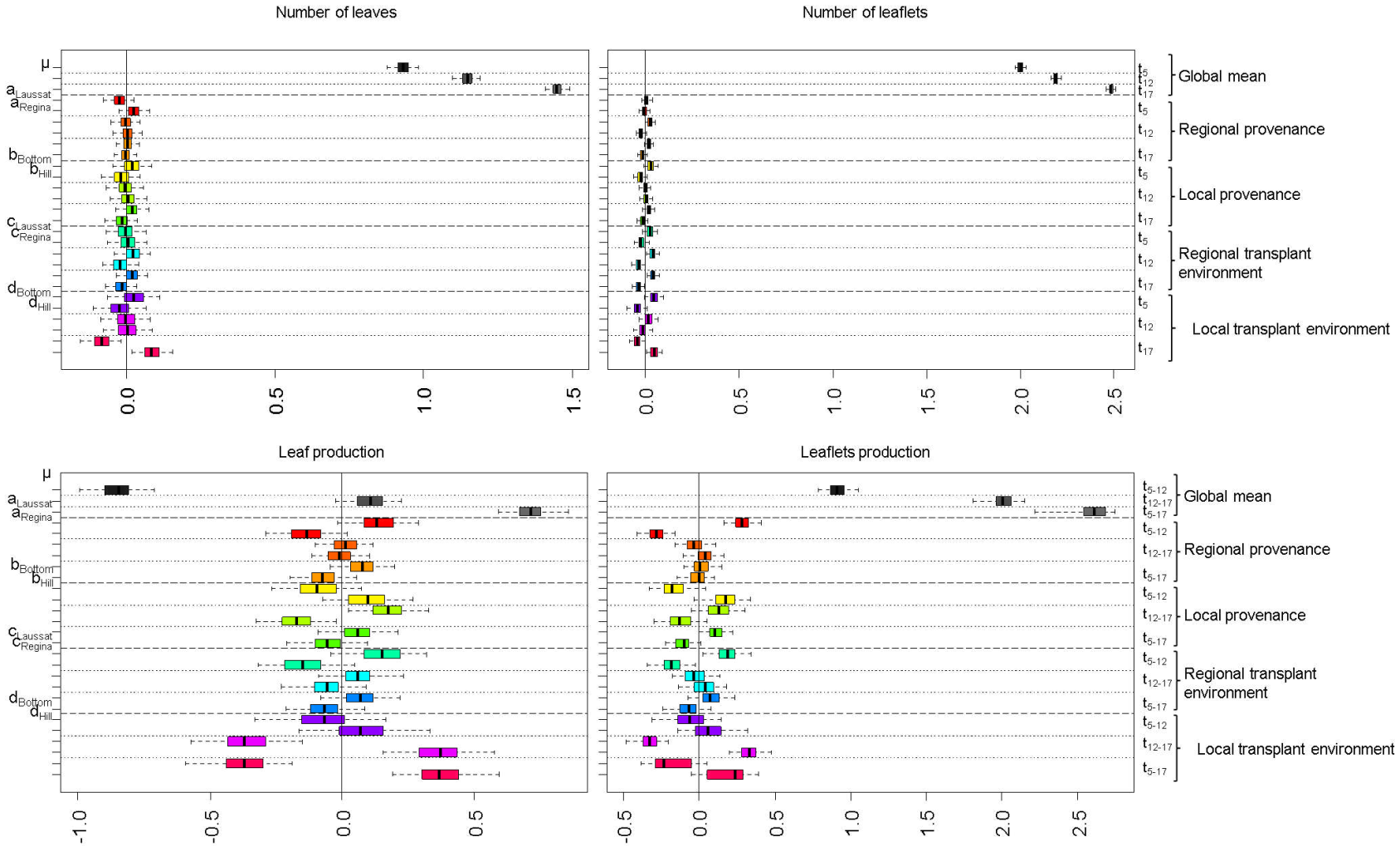


Figure S2: Global mean and main effects estimations for characters relative to leaf and leaflet production μ: global mean, a: Regional provenance effects, b: Local provenance effects, c: Regional transplant site effect, d: regional transplant environment effects).

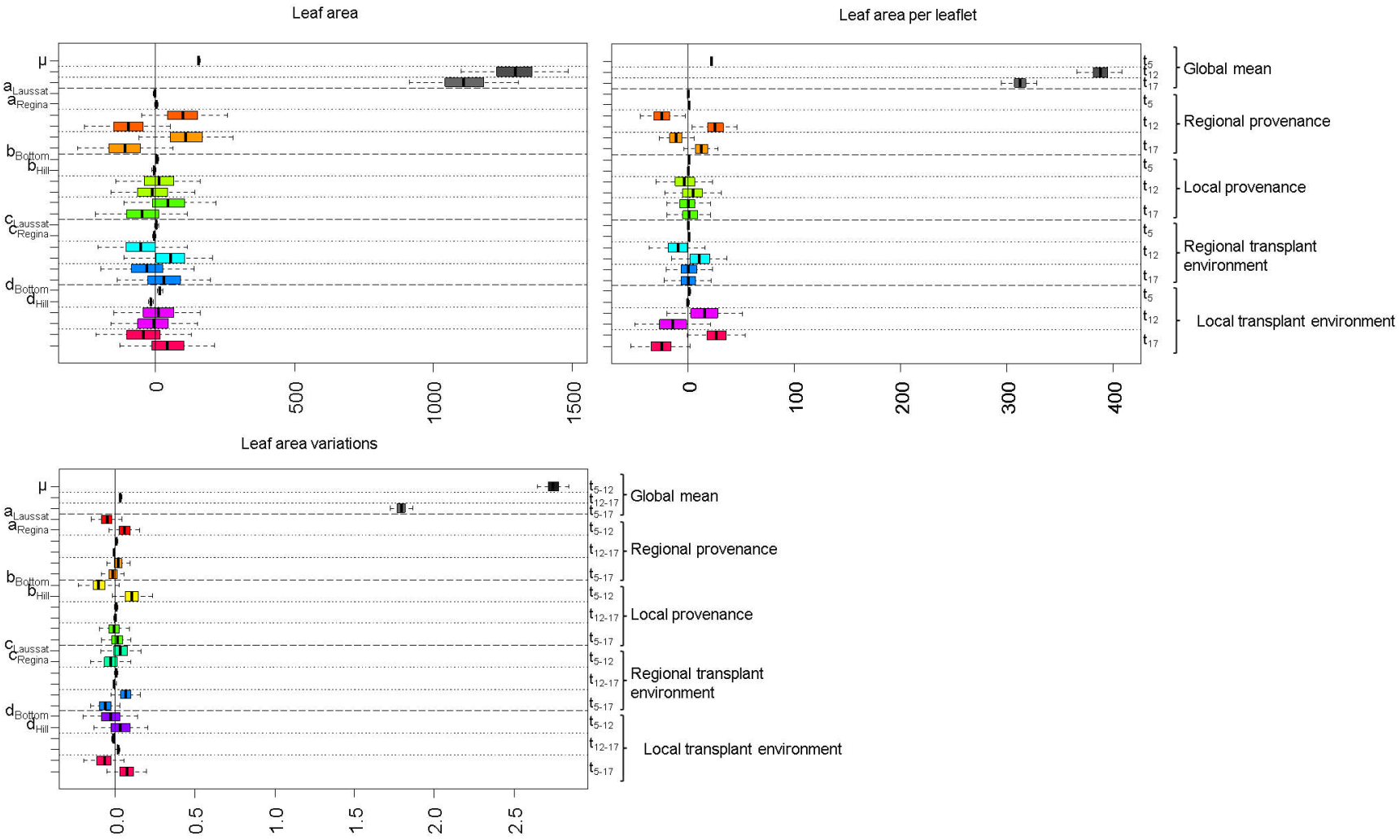


Figure S3: Global mean and main effects estimations for characters relative to leaf area expansion
μ: global mean, a: Regional provenance effects, b: Local provenance effects, c: Regional transplant site effect, d: regional transplant environment effects).

BIBLIOGRAPHY

Achere, V., J. M. Favre, et al. (2005). "Genomic organization of molecular differentiation in Norway spruce (*Picea abies*)." Mol Ecol **14**: 3191-3201.

Aitken, S. N., S. Yeaman, et al. (2008). "Adaptation, migration or extirpation: climate change outcomes for tree populations." Evolutionary Applications **1**(1): 95-111.

Albert, C. H., W. Thuiller, et al. (2010). "Intraspecific functional variability: extent, structure and sources of variation." Journal of Ecology **98**: 604-613.

Aldrich, P. R. and J. L. Hamrick (1998). "Reproductive Dominance of Pasture Trees in a Fragmented Tropical Forest Mosaic." Science **281**: 103-105.

Allendorf, F. W., P. A. Hohenlohe, et al. (2010). "Genomics and the future of conservation genetics." Nature reviews genetics **11**: 697-709.

Angert, A. L. and D. W. Schemske (2005). "The evolution of species' distributions: reciprocal transplants accross the elevation ranges of *Mimulus cardinalis* & *M. lewisii*." Evolution Int J Org Evolution **59**(8): 1671-1684.

Anhuf, D., M. P. Ledru, et al. (2006). "Paleo-environmental change in Amazonian and African rainforest during the LGM." Paleogeography, Paleoclimatology, Paleoecology **239**: 510-527.

Antonovics, J. (1971). "Effects of a heterogeneous environment on the genetics of natural populations." Am Sci **59**: 539-599.

Arias, P. A., R. Fu, et al. (2011). "Changes in cloudiness over the Amazon rainforests during the last two decades: diagnostic and potential causes." Climate Dynamics **37**(5-6): 1151-1164.

Arita, H. T. and E. Vazquez-Dominguez (2008). "The tropics: cradle, museum or casino? A dynamic null model for latitudinal gradients of species diversity." Ecol Lett **11**(7): 653-663.

Audigeos, D. (2010). Thèse: Analyse de la variabilité génétique quantitative et moléculaire dans le complexe d'espèces d'arbres tropicaux *Eperua* en Guyane française.

Audigeos, D., A. Buonamici, et al. (2010). "Aquaporins in the wild: natural genetic diversity and selective pressure in the PIP gene family in five Neotropical tree species." BCM Evolutionary Biology **10**: 202.

Audigeos, D., L. Brousseau, et al. (2013). "Molecular divergence in tropical tree populations occupying environmental mosaics." J Evol Biol **26**(3): 529-544.

Babik, W., R. K. Bultin, et al. (2009). "How sympatric is speciation in the Howea palms of Lord Howe Island?" Mol Ecol **18**: 3629-3638.

Balaguer, L., E. Martinez-Ferri, et al. (2001). "Population divergence in the plasticity of the response of *Quercus coccifera* to the light environment." Funtional Ecology **15**(1): 124-135.

Balding, D. J. and R. A. Nichols (1995). "A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity." Genetica **96**(1-2): 3-12.

Balding, D. J., M. Greenhalgh, et al. (1996). "Population genetics of STR loci in Caucasians." Int J Legal Med **108**: 300-305.

Balloux, F. (2001). "EASYPOP (Version 1.7): A Computer Program for Population Genetics Simulations." Journal of Heredity **92**: 301-302.

Banta, J. A., J. Dole, et al. (2007). "Evidence of local adaptation to coarse-grained environmental variation in *Arabidopsis thaliana*." Evolution Int J Org Evolution **61**(10): 2419-2432.

Baraloto, C., D. E. Goldberg, et al. (2005). "Performance Trade-offs among tropical tree seedlings in contrasting microhabitats." Ecology **86**: 2461-2472.

Baraloto, C., D. Bonal, et al. (2006). "Differential seedling growth response to soil resource availability among nine neotropical tree species." J Trop Ecol **22**: 487-497.

Baraloto, C., F. Morneau, et al. (2007). "Seasonal water stress tolerance and habitat associations within four neotropical tree genera." Ecology **88**: 478-489.

Baraloto, C. and P. Couteron (2010). "Fine-scale Microhabitat Heterogeneity in a French Guianan Forest: Tropical Forest Microhabitat Heterogeneity." Biotropica **42**: 420-428.

Barbazuk, W. B., S. J. Emrich, et al. (2007). "SNP discovery via 454 transcriptome sequencing." The Plant Journal **51**(5): 910-918.

Bariteau, M. (1992). "Régénération naturelle de la forêt tropicale humide de Guyane : étude de la répartition spatiale de Qualea rosea Aublet, Eperua falcata Aublet et Symphonia globulifera Linnaeus f." Ann For Sci **49**(4): 359-382.

Barthe, S. (2012). Thèse: Signature de transition démographique pour les arbres de la Forêt Tropicale Humide du plateau des Guyanes.

Barthes, B. (1991). "Influence des caractères pédologiques sur la répartition spatiale de deux espèces du genre Eperua (Caesalpiniaceae) en forêt guyanaise = Influence of soil conditions on the spatial distribution of two Eperua species (Caesalpiniaceae) in the rain forests of French Guyana." Revue d'écologie **46**: 303-320.

Beaumont, M. A. and R. A. Nichols (1996). "Evaluating Loci for Use in the Genetic Analysis of Population Structure." Proceedings of The Royal Society B: Biological Sciences **263**: 1619-1626.

Beaumont, M. A. and D. J. Balding (2004). "Identifying adaptive genetic divergence among populations from genome scans." Mol Ecol **13**(4): 969-980.

Beaumont, M. A. (2005). "Adaptation and speciation: what can Fst tell us?" Trends in Ecology & Evolution **20**(8): 435-440.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the false discovery rates: a practical and powerful approach to multiple testing." Journal of the Royal Statistical Society B **57**: 289-300.

Bergmann, F. and L. t. u. Forstgenefik (1978). "The allelic distribution at an acid phosphatase locus in Norway spruce (*Picea abies*) along similar climatic gradients." Theoretical and Applied Genetics **52**: 57-64.

Betts, R. A., P. M. Cox, et al. (2004). "The role of ecosystem-atmosphere interactions in simuklated Amazonian precipitation decrease and forest dieback under glocal climate warming." Theor. Appl. Climatol. **78**: 157-175.

Bierne, N., J. Welch, et al. (2011). "The coupling hypothesis: why genome scans may fail to map local adaptation genes." Mol Ecol **20**(10): 2044-2072.

Bierne, N., D. Roze, et al. (2013). "Pervasive selection or is it? Why are FST outliers sometimes so frequent?" Mol Ecol **22**(8): 2061-2064.

Blanca, J., J. Canizares, et al. (2011). "Transcriptome characterization and high throughput SSRs and SNPs discovery in *Cucurbita pepo* (Cucurbitaceae)." BMC Genomics **12**(1): 104.

Blokhina, O., E. Virolainen, et al. (2003). "Antioxidants, Oxidative Damage and Oxygen Deprivation Stress: a Review." Ann Bot (Lond) **91**(2): 179-194.

Bolnick, D. I. and P. Nosil (2007). "Natural selection in populations subject to a migration load." Evolution Int J Org Evolution **61**(9): 2229-2243.

Bonal, D., D. Sabatier, et al. (2000). "Interspecific variability of d13Cn amon trees in rainforest of French Guiana: functional groups and canopy integration." Oecologia **124**: 454-468.

Bonal, D., C. Atger, et al. (2000). "Water acquiqition patterns of two wet tropical canopy tree species of French Guiana as inferred from H218O extraction profiles." Ann For Sci **57**: 717-724.

Bonal, D. and J. M. Guehl (2001). "Contrasting patterns of leaf water potential and gas exchange responses to drought in seedlings of tropical rainforest species." Functional Ecology **15**: 490-496.

Bonal, D., C. Born, et al. (2007). "The successional status of tropical rainforest tree species is associated with differences in leaf carbon isotope discrimination and functional traits." Ann For Sci **64**: 169-176.

Bonin, A., P. Taberlet, et al. (2006). "Explorative Genome Scan to Detect Candidate Loci for Adaptation Along a Gradient of Alm6titude in the Common Frog (*Rana temporaria*)." Mol Biol Evol **23**: 773-783.

Bonin, A., D. Ehrich, et al. (2007). "Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists." Mol Ecol **16**(18): 3737-3758.

Born, C., O. J. Hardy, et al. (2008). "Small-scale spatial genetic structure in the Central African rainforest tree species *Aucoumea klaineana*: a stepwise approach to infer the impact of limited gene dispersal, population history and habitat fragmentation." Mol Ecol **17**(8): 2041-2050.

Bossdorf, O., C. L. Richards, et al. (2008). "Epigenetics for ecologists." Ecol Lett **11**: 106-115.

Bouck, A. and T. Vision (2006). "The molecular ecologist's guide to expressed sequence tags: ESTs." Mol Ecol **16**: 907-924.

Bréda, N., R. Huc, et al. (2006). "Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences." Ann For Sci **63**: 625-644.

Brendel, O., D. Pot, et al. (2002). "Genetic parameters and QTL analysis of d13C and ring width in maritime pine." Plant, Cell and Environment **25**: 945-953.

Broccoli, A. J. (2000). "Tropical Cooling at the Last Glacial Maximum: An Atmosphere Mixed Layer Ocean Model Simulation." J Clim **13**: 951-976.

Brousseau, L., D. Bonal, et al. (2013). "Highly local environmental variability promotes intra-population divergence of quantitative traits: an example from tropical rainforest trees." Ann Bot (Lond) **112**: 1169-1179.

Brown, G. R., G. P. Gill, et al. (2004). "Nucleotide diversity and linkage disequilibrium in loblolly pine." Proceedings of The National Academy of Sciences **101**: 15255-15260.

Browne, W. J. and D. Draper (2006). "A comparison of Bayesian and likelihood-based methods for fitting multilevel models." Bayesian Analysis **1**: 473-514.

Burgarella, C., M. Navascuas, et al. (2012). "Recent population decline and selection shape diversity of taxol-related genes." Mol Ecol **21**(12): 3006-3021.

Burke, E. J., S. J. Brown, et al. (2006). "Modeling the Recent Evolution of Global Drought and Projections for the Twenty-First Century with the Hadley Centre Climate Model." J. Hydrometeor **7**(5): 1113-1125.

Burt, A. and V. Koufopanou (2004). "Homing endonuclease genes: the rise and fall and rise again of a selfish element." Curr Opin Genet Dev **14**(6): 609-615.

Bush, M. B. and P. E. De Oliveira (2006). "The rise and fall of the Refugial Hypothesis of Amazonian speciation: a paleoecological perspective." Biota Neotropica **6**.

Bush, M. B., M. R. Silman, et al. (2004). "48,000 Years of Climate and Forest Change in a Biodiversity Hot Spot." Science **303**: 827-829.

Byars, S. G., W. Papst, et al. (2007). "Local adaptation and cogradient selection in the alpine plant, *Poa hiemata*, along a narrow altitudinal gradient." Evolution Int J Org Evolution **61**(12): 2925-2941.

Caisse, M. and J. Antonovics (1978). "Evolution in closely adjacent plant populations." Heredity **40**(3): 371-384.

Callister, A. and S. Collins (2008). "Genetic parameter estimates in a clonally replicated progeny test of teak (*Tectona grandis* Linn. f.)." Tree Genet Genomes **4**: 237-245.

Campbell, D. and L. Bernatchez (2004). "Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes." Mol Biol Evol **21**(5): 945-956.

Carlson, J. E., K. E. Holsinger, et al. (2011). "Plant responses to climate in the cape floristic region of South Africa: Evidence for adaptive differentiation in the Proteaceae." Evolution Int J Org Evolution **65**(1): 108-124.

Carnegie, A., I. Johnson, et al. (2004). "Variation among provenances and families of blackbutt (*Eucalyptus pilularis*) in early growth and susceptibility to damage from leaf spot fungi." Canadian Journal of Forest Research **34**: 2314-2326.

Cavers, S., B. Degen, et al. (2005). "Optimal sampling strategy for estimation of spatial genetic structure in tree populations." Heredity **95**: 281-289.

Cavers, S. and C. W. Dick (2013). "Phylogeography of Neotropical trees." Journal of Biogeography **40**(4): 615-617.

Chaves, M. M., J. P. Maroco, et al. (2003). "Understanding plant responses to drought - from genes to the whole plant." Functional Plant Biology **30**: 239-264.

Chen, J., T. Källman, et al. (2012). "Disentangling the Roles of History and Local Selection in Shaping Clinal Variation of Allele Frequencies and Gene Expression in Norway Spruce (*Picea abies*)." Genetics **191**(3): 865-881.

Chevin, L.-M., R. Lande, et al. (2010). "Adaptation, Plasticity, and Extinction in a Changing Environment: Towards a Predictive Theory." PLoS Biol **8**(4): e1000357.

Chevreux, B., T. Pfisterer, et al. (2004). "Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs." Genome Res **14**(6): 1147-1159.

Cho, Y., Y.-L. Qiu, et al. (1998). "Explosive invasion of Plant Mitochondria by a Group I Intron." Evolution Int J Org Evolution **95**: 14244-14249.

Chown, S. L. and K. J. Gaston (2000). "Areas, cradles and museums: the latitudinal gradient in species richness." Trends Ecol Evol **15**(8): 311-315.

Chybicki, I. J., A. Oleksa, et al. (2011). "Increased inbreeding and strong kinship structure in *Taxus baccata* estimated from both AFLP and SSR data." Heredity **107**(6): 589-600.

Cincotta, R. P., J. Wisnewski, et al. (2000). "Human population in the biodiversity hotspots." Nature **404**: 990-992.

Clark, D. B. and D. A. Clark (1985). "Seedling dynamcs of a tropical tree: impacts of herbivory and meristem damage." Ecology **66**: 1884-1892.

Clark, J. S., E. Macklin, et al. (1998). "Stages and Spatial Scales of Recruitment Limitation in Southern Appalachian Forests." Ecol Monogr **68**: 213-235.

Clark, D. B., M. W. Palmer, et al. (1999). "Edaphic fators and the landscape-scale distribution of tropical rain forest trees." Ecology **80**: 2662-2675.

Cloutier, D., M. Kanashiro, et al. (2006). "Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl." Mol Ecol **16**: 1-13.

Cloutier, D., O. J. Hardy, et al. (2007). "Low Inbreeding and High Pollen Dispersal Distances in Populations of Two Amazonian Forest Tree Species." Biotropica **39**(3): 406-415.

Colinvaux, P. A., P. E. De Oliveira, et al. (2000). "Amazonian and neotropical plant communities on glacial time-scales: The failure of the aridity and refuge hypotheses." Quaternary Science Reviews **19**: 141-169.

Collevatti, R. G., D. Grattapaglia, et al. (2001). "High resolution microsatellite based analysis of the mating system allows the detection of significant biparental inbreeding in *Caryocar brasiliense*, an endangered tropical tree species." Heredity **86**: 60-67.

Collevatti, R. G., J. S. Lima, et al. (2010). "Spatial Genetic Structure and Life History Traits in Cerrado Tree Species: Inferences for Conservation." Natureza & Conservacao **08**: 54-59.

Collevatti, R. G. and J. D. Hay (2011). "Kin structure and genotype-dependent mortality: a study using the neotropical tree *Caryocar brasiliense*." Journal of Ecology **99**(3): 757-763.

Collinet, F. (1997). Thèse: Essai de regroupements des principales espèce structurantes d'une forêt dense humide d'après l'analyse de leur répartition spatiale (forêt Paracou-Guyane).

Colpaert, N., S. cavers, et al. (2005). "Sampling tissue for DNA analysis of trees: trunk cambium as an alternative to canopy leaves." Silvae Genet **54**: 265-269.

Condit, R., S. P. Hubbell, et al. (1996). "Species-area and species-individual relationships for tropical trees: a comparison of three 50-ha plots." Journal of Ecology **84**: 549-562.

Condit, R., S. Aguilar, et al. (2004). "Tropical forest dynamics across a rainfall gradient and the impact of an El Niño dry season." J Trop Ecol **20**: 51-72.

Conesa, A. and S. Götz (2008). "Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics." International Journal of Plant Genomics **2008**: 1-12.

Conner, J. K. and D. L. Hartl (2004). A Primer of Ecological Genetics.

Coop, G., D. Witonsky, et al. (2010). "Using Environmental Correlations to Identify Loci Underlying Local Adaptation." Genetics **185**: 1411-1423.

Cornelissen, J. H. C., S. Lavorel, et al. (2003). "A handbook of protocols for standardised and easy measurement of plant functional traits worldwide." Australian Journal of Botany **51**(4): 335-380.

Cornelius, J. (1994). "Heritabilities and additive genetic coefficients of variation in forest trees." Canadian Journal of Forest Research-revue Canadienne De Recherche Forestiere **24**: 372-379.

Costa e Silva, J., G. Dutkowski, et al. (2005). "Across-site heterogeneity of genetic and environmental variancesin the genetic evaluation of *Eucalyptus globulus* trials for height growth." Ann For Sci **62**: 183-191.

Coste, S., J. C. Roggy, et al. (2005). "Leaf photosynthetic traits of 14 tropical rain forest species in relation to leaf nitrogen concentration and shade tolerance." Tree Physiol **25**: 1127-1137.

Cotterill, P. P. (1987). "Short note: On estimating heritability according to practical applications." Silvae Genet **36**: 46-48.

Coutand, C., M. Chevolot, et al. (2010). "Mechanosensing of stem bending and its interspecific variability in five neotropical rainforest species." Ann Bot (Lond) **105**: 341-347.

Cowan, R. S. (1975). A monograph of the genus Eperua (Leguminosae-Caesalpinioideae), Smithsonian Institution Press.

Cowling, S. A., M. A. Maslin, et al. (2001). "Paleovegetation Simulations of Lowland Amazonia and Implications for Neotropical Allopatry and Speciation." Quaternary Research **55**: 140-149.

Craft, J. A., J. A. Gilbert, et al. (2010). "Pyrosequencing of *Mytilus galloprovincialis* cDNAs: tissue-specific expression patterns." PLoS ONE **5**.

Cushman, J. C. and H. J. Bohnert (2000). "Genomic approaches to plant stress tolerance." Curr Opin Plant Biol **3**(2): 117-124.

Cutler, S., M. Ghassemian, et al. (1996). "A Protein Farnesyl Transferase Involved in Abscisic Acid Signal Transduction in *Arabidopsis*." Science **273**(5279): 1239-1241.

Da Silveira Lobo Sternberg, L. (2001). "Savanna-forest hysteresis in the tropics." Global Ecology and Biogeography **10**: 369-378.

Darwin, C. (1859). On the origin of speciesby means of natural selection, or the preservation of favoured races in the struggle for life, Murray, London.

Daws, M. I., C. E. Mullins, et al. (2002). "Topographic position affects the water regime in a semideciduous tropical forest in Panama." Plant Soil **238**: 79-90.

Dawson, I. K., R. Waugh, et al. (1997). "Simple sequence repeats provide a direct estimate of pollen-mediated gene dispersal in the tropical tree *Gliricidia sepium*." Mol Ecol **6**(2): 179-183.

De Carvalho, D., P. K. Ingvarsson, et al. (2010). "Admixture facilitates adaptation from standing variation in the European aspen (*Populus tremula L.*), a widespread forest tree." Mol Ecol **19**(8): 1638-1650.

De Pristo, M. A., E. Banks, et al. (2011). "A framework for variation discovery and genotyping using next-generation DNA sequencing data." Nat Genet **43**(5): 491-498.

De Souza Gonçalves, P., M. L. T. De Moraes, et al. (2005). "Genetic variation in growth traits and yield of Rubber trees (*Hevea brasiliensis*) growing in the Brazilian state of Sao Paulo." Genet Mol Biol **28**: 765-772.

Debarre, F. and S. Gandon (2010). "Evolution of specialization in a spatially continuous environment." J Evol Biol **23**: 1090-1099.

Degen, B., E. Bandou, et al. (2004). "Limited pollen dispersal and biparental inbreeding in *Symphonia globulifera* in French Guiana." Heredity **93**(6): 585-591.

Degen, B. and D. W. Roubik (2004). "Effects of Animal Pollination on Pollen Dispersal, Selfing, and Effective Population Size of Tropical Trees: A Simulation Study." Biotropica **36**(2): 165-179.

Delph, L. F. and J. K. Kelly (2013). "On the importance of balancing selection in plants." New Phytologist.

Denslow, J. S. (1987). "Tropical Rainforest Gaps and Tree Species Diversity." Annu Rev Ecol Syst **18**(1): 431-451.

Dick, C. W., G. Etchelecu, et al. (2003). "Pollen dispersal of tropical trees (*Dinizia excelsa*: Fabaceae) by native insects and African honeybees in pristine and fragmented Amazonian rainforest." Mol Ecol **12**(3): 753-764.

Dick, C. W., E. Bermingham, et al. (2007). "Extreme long-distance dispersal of the lowland tropical rainforest tree *Ceiba pentandra L.* (Malvaceae) in Africa and the Neotropics." Mol Ecol **16**(14): 3039-3049.

Dick, C., O. Hardy, et al. (2008). "Spatial Scales of Pollen and Seed-Mediated Gene Flow in Tropical Rain Forest Trees." Tropical Plant Biology **1**(1): 20-33.

Dirzo, R. and P. H. Raven (2003). "Global state of biodiversity and loss." Annual Review of Environment and Resources **28**(1): 137-167.

Doligez, A. and H. I. Joly (1997). "Genetic diversity and spatial structure within a natural stand of a tropical forest tree species, *Carapa procera* (Meliaceae), in French Guiana." Heredity **79**(1): 72-82.

Donovan, L. A., H. Maherali, et al. (2011). "The evolution of the worldwide leaf economics spectrum." Trends in Ecology & Evolution **26**(2): 88-95.

Doyle, J. J. and J. L. Doyle (1987). "A rapid DNA isolation procedure from small quantities of fresh leaf tissues." Phytochem. Bull. **19**: 11-15.

Dudley, S. (1996). "Differing selection on plant physiological traits in response to environmental water availability: a test of adaptive hypothesis." Evolution Int J Org Evolution **50**: 92-102.

Duminil, J., H. Caron, et al. (2006). "Blind population genetics survey of tropical rainforest trees." Mol Ecol **15**(12): 3505-3513.

Dutech, C., J. Seiter, et al. (2002). "Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana." Mol Ecol **11**(4): 725-738.

Eckert, A., J. L. Wegrzyn, et al. (2009). "Multilocus Patterns of Nucleotide Diversity and Divergence Reveal Positive Selection at Candidate Genes Related to Cold-hardiness in Coastal Douglas-fir (*Pseudotsuga menziesii* var. *menziesii*)." Genetics **183**: 289-298.

Eckert, A. J., A. D. Bower, et al. (2009). "Association Genetics of Coastal Douglas Fir (*Pseudotsuga menziesii* var. *menziesii*, Pinaceae). I. Cold-Hardiness Related Traits." Genetics **182**: 1289-1302.

Eckert, A. J., A. D. Bower, et al. (2010). "Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae)." Mol Ecol **19**: 3789-3805.

Eckert, A. J., J. V. Heerwaarden, et al. (2010). "Patterns of population structure and environmental associations to aridity across the range of Loblolly Pine (*Pinus taeda L.*, Pinaceae)." Genetics **185**: 969-982.

Egan, S. P., P. Nosil, et al. (2008). "Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of Neochlamisus ebbianae leaf beetles." Evolution Int J Org Evolution **62**(5): 1162-1181.

Ehleringer, J. R. and T. A. Cooper (1988). "Correlations between carbon isotope ratio and microhabitat in desert plants." Oecologia **76**: 562-566.

Ehrlich, P. R. and P. H. Raven (1996). "Differentiation of populations." Ecological Applications **6**: 1036-1046.

Ekblom, R. and J. Galindo (2011). "Applications of next generation sequencing in molecular ecology of non-model organisms." Heredity **107**: 1-15.

Ellegren, H. (2008). "Sequencing goes 454 and takes large-scale genomics into the wild." Mol Ecol **17**: 1629-1631.

Ellison, A. M. (1996). "An Introduction to Bayesian Inference for Ecological Research and Environmental Decision-Making." Ecological Applications **6**: 1036-1046.

Emrich, S. J., W. B. Barbazuk, et al. (2007). "Gene discovery and annotation using LCM-454 transcriptome sequencing." Genome Res **17**(1): 69-73.

Endler, J. (1977). Geographic variation, speciation and clines, Princeton, NJ: Princetn University Press.

Endler, J. (1986). Natural selection in the wild. Monographs in population biology.

Engelbrecht, B. M. J. and T. A. Kursar (2003). "Comparative drought-resistance of seedlings of 28 species of co-occurring tropical woody plants." Oecologia **136**: 383-393.

Engelbrecht, B. M. J., L. S. Comita, et al. (2007). "Drought sensitivity shapes species distribution patterns in tropical forests." Nature **447**: 80-83.

Epperson, B. K. (2003). Geographical Genetics, Princeton University Press.

Epron, D., A. Bosc, et al. (2006). "Spatial variation of soil respiration across a topographic gradient in a tropical rain forest in French Guiana." J Trop Ecol **22**: 565-574.

Evanno, G., S. Regnaut, et al. (2005). "Detecting the number of clusters of individuals using the software structure: a simulation study." Mol Ecol **14**(8): 2611-2620.

Eveno, E., C. Collada, et al. (2008). "Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate gees as revealed by genetic differenciation analyses." Mol Biol Evol **25**: 417-437.

Excoffier, L., P. E. Smouse, et al. (1992). "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human Mitochondrial DNA restriction data." Genetics **131**: 479-491.

Excoffier, L., T. Hofer, et al. (2009). "Detecting loci under selection in a hierarchically structured population." Heredity **103**(4): 285-298.

Excoffier, L. and H. E. L. Lischer (2010). "Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows." Molecular Ecology Resources **10**(3): 564-567.

Falush, D., M. Stephens, et al. (2003). "Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies." Genetics **164**(4): 1567-1587.

Falush, D., M. Stephens, et al. (2007). "Inference of population structure using multilocus genotype data: dominant markers and null alleles." Mol Ecol Notes **7**: 574-578.

Farquhar, G. D., M. H. O'Leary, et al. (1982). "On the Relationship Between Carbon Isotope Discrimination and the Intercellular Carbon Dioxide Concentration in Leaves." Aust J Plant Physiol **9**: 121-137.

Fernandez, G. C. J. and J. C. Miller (1985). "Estimation of heritability by parent-offspring regression." Theoretical and Applied Genetics **70**: 650-654.

Ferry, B., F. Morneau, et al. (2010). "Higher treefall rates on slopes and waterlogged soils result in lower stand biomass and productivity in a tropical rain forest." Journal of Ecology **98**(1): 106-116.

Field, C. B., M. J. Behrenfeld, et al. (1998). "Primary Production of the Biosphere: Integrating Terrestrial and Oceanic Components." Science **281**(5374): 237-240.

Fine, P. V. A., D. C. Daly, et al. (2005). "The contribution of edaphic heterogeneity to the evolution and biodiversity of Burseraceae trees in the western Amazon." Evolution Int J Org Evolution **59**: 1464-1478.

Fine, P. V. A. and S. W. Kembel (2010). "Phylogenetic community structure and phylogenetic turnover across space and edaphic gradients in western Amazonian tree communities." Ecography **34**(4): 552-565.

Flint-Garcia, S. A., J. M. Thornsberry, et al. (2003). "Structure of linkage disequilibrium in plants." Annu Rev Plant Biol **54**(1): 357-374.

Foll, M. and O. Gaggiotti (2008). "A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective." Genetics **180**(2): 977-993.

Ford, E. B. (1964). Ecological Genetics, Chapman & Hall.

Forget, J. M. (1989). "La Regeneration Naturelle D'une Espece Autochore de La Foret Guyanaise: *Eperua falcata* Aublet (Caesalpiniaceae)." Biotropica **21**: 115-125.

Fournier-Level, A., A. Korte, et al. (2011). "A Map of Local Adaptation in *Arabidopsis thaliana*." Science **334**(6052): 86-89.

Freeland, J. R., P. Biss, et al. (2010). "Selection pressures have caused genome-wide population differentiation of *Anthoxanthum odoratum* despite the potential for high gene flow." J Evol Biol **23**: 776-782.

Frias-Lopez, J., Y. Shi, et al. (2008). "From the Cover: Microbial community gene expression in ocean surface waters." Proceedings of The National Academy of Sciences **105**: 3805-3810.

Futschik, A. and C. Schlõtterer (2010). "Massively Parallel Sequencing of Pooled DNA Samples: The Next Generation of Molecular Markers." Genetics **186**: 207-218.

Galbraith, D., P. E. Levy, et al. (2010). "Multiple mechanisms of Amazonian forest biomass losses in three dynamic global vegetation models under climate change." New Phytologist **187**: 647-665.

Ganal, W. M., T. R. Altmann, et al. (2009). "SNP identification in crop plants." Curr Opin Plant Biol **12**(2): 7-145.

Gerhardt, K. (1996). "Effects of root competition and canopy openess on survival and growth of tree seedlings in a tropical seasonal dry forest." For Ecol Manage **82**: 33-48.

Gilbert, G. S., K. E. Harms, et al. (2001). "Effects of seedling size, El Nino drought, seedling density and distance to nearest conspecific adult on 6-year survival of Ocotea whitei seedlings in Panama." Oecologia **127**: 509-516.

Givnish, T. J. (1999). "On the causes of gradients in tropical tree diversity." Ecology **87**: 193-210.

Gonzales-Rodriguez, V., I. C. Barrio, et al. (2012). "Within-populm6ation variability influences early seedling establishment in four Mediterranean oaks." Acta Oecologica **41**: 82-89.

Gonzalez-Martinez, S. C., E. Ersoz, et al. (2006). "DNA sequence variation and selection of Tag single-nucleotide polymorphisms at candidate genes for drought stress response in *Pinus taeda L.*" Genetics **172**: 1915-1926.

Gonzalez-Martinez, S. C., K. V. Krutovsky, et al. (2006). "Forest-tree population genomics and adaptive evolution." New Phytologist **170**: 227-238.

Gonzalez-Rodriguez, V., R. Villar, et al. (2011). "Maternal influences on seed mass effect and initial seedling growth in four *Quercus* species." Acta Oecologica **37**: 1-9.

Gonzalo-Turpin, H. and L. Hazard (2009). "Local adaptation occurs along altitudinal gradient despite the existence of gene flow in the alpine plant species *Festuca eskia*." Journal of Ecology **97**(4): 742-751.

Goudet, J., S. Nuenschwander, et al. (2009). Factors influencing the Fst/Qst contrast under neutrality and selection. ESEB 12th Congress, Turin.

Gourlet-Fleury, S., V. Favrichon, et al. (2004). Ecology & Management of a Neotropical Rainforest. Lessons Drawn from Paracou, a Long-Term Experimental Research Site in French Guiana Consequences of silvicultural treatments on stand dynamics at Paracou, Elsevier, Paris.

Granier, A., R. Huc, et al. (1996). "Transpiration of natural rain forest and its dependence on climatic factors." Agricultural and Forest Meteorology **78**(1-2): 19-29.

Gravel, D., C. D. Canham, et al. (2006). "Reconciling niche and neutrality: the continuum hypothesis." Ecol Lett **9**(4): 399-409.

Guo, S. W. and E. A. Thompson (1992). "Performing the exact test of Hardy-Weinberg proportion for multiple alleles." Biometrics **48**: 361-372.

Guo, X.-Y., Z.-Y. Huang, et al. (2011). "A comparison of physiological, morphological and growth responses of 13 hybrid poplar clones to flooding." Forestry **84**(1): 1-12.

Haapanen, M., P. Veiling, et al. (1997). "Progeny Trial Estimates of Genetic Parameters for Growth and Quality Traits in Scots Pine." Silva Fennica **31**: 3-12.

Haffer, J. (1969). "Speciation in Amazonian Forest Birds." Science **165**: 131-137.

Hamilton, M. B. (2009). Population Genetics, Wiley-Blackwell.

Hampe, A., L. E. Masri, et al. (2010). "Origin of spatial genetic structure in an expanding oak population." Mol Ecol **19**: 459-471.

Hancock, A. M., B. Brachi, et al. (2011). "Adaptation to Climate Across the *Arabidopsis thaliana* Genome." Science **334**(6052): 83-86.

Hannrup, B., L. Wilhelmsson, et al. (1998). "Time Trends for Genetic Parameters of Wood Density and Growth Traits in *Pinus sylvestris L.*" Silvae Genet **47**: 214-219.

Hardesty, B. D., S. P. Hubbell, et al. (2006). "Genetic evidence of frequent long-distance recruitment in a vertebrate-dispersed tree." Ecol Lett **9**(5): 516-525.

Hardy, O. J. and X. Vekemans (1999). "Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models." Heredity **83**(2): 145-154.

Hardy, O. J. and X. Vekemans (2002). "spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels." Mol Ecol Notes **2**: 618-620.

Hardy, O. J., L. Maggia, et al. (2006). "Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species." Mol Ecol **15**(2): 559-571.

Harms, K. E., R. Condit, et al. (2001). "Habitat associations of trees and shrubs in a 50-ha neotropical forest plot." Journal of Ecology **89**(6): 947-959.

Hartl, D. L. and A. G. Clark (2007). Principles of population genetics, Sinauer Associates Inc. Publishers.

Hawkins, B. A., M. A. Rodroguez, et al. (2011). "Global angiosperm family richness revisited: linking ecology and evolution to climate." Journal of Biogeography **38**(7): 1253-1266.

Hayes, B., J. K. Laerdahl, et al. (2007). "An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences." **265**: 82-90.

Hedrick, P. W. (2006). "Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics." Annual Review of Ecology, Evolution, and Systematics **37**(1): 67-93.

Helyar, S. J., J. Hemmer-Hansen, et al. (2011). "Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges." Molecular Ecology Resources **11**: 123-136.

Hereford, J. (2009). "A Quantitative Survey of Local Adaptation and Fitness Trade-Offs." Am Nat **173**(5): 10.

Hermisson, J. (2009). "Who believes in whole-genome scans for selection?" Heredity **103**(4): 283-284.

Herrera, C. M. and P. Bazaga (2008). "Population-genomic approach reveals adaptive floral divergence in discrete populations of a hawk moth-pollinated violet." Mol Ecol **17**(24): 5378-5390.

Heschel, M. S. and C. Riginos (2005). "Mechanisms of selection for drought stress tolerance and avoidance in *Impatiens capensis* (Balsaminaceae)." Am J Bot **92**(1): 37-44.

Heuertz, M., X. Vekemans, et al. (2003). "Estimating seed vs. pollen dispersal from spatial genetic structure in the common ash." Mol Ecol **12**: 2483-2495.

Heuertz, M., E. De Paoli, et al. (2006). "Multilocus Patterns of Nucleotide Diversity, Linkage Disequilibrium and Demographic History of Norway Spruce [*Picea abies* (L.) Karst]." Genetics **174**(4): 2095-2105.

Heywood, J. S. (1991). "Spatial Analysis Of Genetic Variation In Plant Populations." Annu Rev Ecol Syst **22**: 335-355.

Hill, W. G. and A. Robertson (1968). "Linkage disequilibrium in finite populations." Theoretical and Applied Genetics **38**: 226-231.

Hill, W. and B. Weir (1988). "Variances and covariances of squared linkage disequilibria in finite populations." Theor Popul Biol **33**: 54-78.

Hodge, G. R. and T. L. White (1992). "Genetic parameters estimates for growth traits at different ages in Splash Pine and some implications for breeding." Silvae Genet **41**: 252-261.

Hodge, G., W. Dvorak, et al. (2002). "Growth, provenance effects and genetic variation of *Bombacopsis quinata* in field tests in Venezuela and Colombia." For Ecol Manage **158**: 273-289.

Hoffmann, A. A. and C. M. Sgro (2011). "Climate change and evolutionary adaptation." Nature **470**(7335): 479-485.

Hoorn, C., F. P. Wesselingh, et al. (2010). "Amazonia through times: Andean Uplift, Climate change, Landscape Evolution and biodiversity." Science **330**: 927-931.

Hovenden, M. J. and J. K. Vander Schoor (2004). "Nature vs nurture in the leaf morphology of Southern beech, Nothofagus cunninghamii (Nothofagaceae)." New Phytologist **161**: 585-594.

Howe, H. F. (1990). "Survival and growth of juvenile *Virola surinamensis* in Panama: effects of herbivory and canopy closure." J Trop Ecol **6**: 259-280.

Howe, G. T. and A. M. Brunner (2005). "An evolving approach to understanding plant adaptation." New Phytologist **167**(1): 1-5.

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography, Princeton University Press.

Hubbell, S. P., F. He, et al. (2008). "How many tree species are there in the Amazon and how many of them will go extinct?" Proceedings of The National Academy of Sciences **105**: 11498-11504.

Hufford, K. M. and S. J. Mazer (2003). "Plant ecotypes: genetic differentiation in the age of ecological restoration." Trends in Ecology & Evolution **18**(3): 147-155.

Huse, S., J. Huber, et al. (2007). "Accuracy and quality of massively parallel DNA pyrosequencing." Genome Biol **8**(7): R143.

Hutchinson, G. (1959). "Homage to Santa Rosalia or why are there so many kinds of animals?" Am Nat **93**: 145-159.

Ingvarsson, P. K. (2005). "Nucleotide Polymorphism and Linkage Disequilibrium Within and Among Natural Populations of European Aspen (*Populus tremula L.*, Salicaceae)." Genetics **169**: 945-953.

Ingvarsson, P. K. (2008). "Multilocus Patterns of Nucleotide Polymorphism and the Demographic History of *Populus tremula*." Gmenetics **180**: 329-340.

Jabot, F., R. S. Etienne, et al. (2008). "Reconciling neutral community models and environmental filtering: theory and an empirical test." Oikos **117**(9): 1308-1320.

Janzen, D. H. (1970). "Herbivores and the Number of Tree Species in Tropical Forests." American Naturalist **104**: 501-528.

John, R., J. W. Dalling, et al. (2007). "Soil nutrients influence spatial distributions of tropical tree species." Proceedings of The National Academy of Sciences **104**(3): 864-869.

Johns, T. C., J. M. Gregory, et al. (2003). "Anthropogenic climate change for 1860 to 2100 simulated with the HadCM3 model under updated emissions scenarios." Climate Dynamics **20**(6): 583-612.

Johnson, G. and B. Gartner (2006). "Genetic variation in basic density and modulus of elasticity of coastal Douglas-fir." Tree Genet Genomes **3**: 25-33.

Joost, S., A. Bonin, et al. (2007). "A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation." Mol Ecol **16**(18): 3955-3969.

Jump, A. S. and J. Penuelas (2005). "Running to stand still: adaptation and the response of plants to rapid climate change." Ecol Lett **8**: 1010-1020.

Jump, A. S., J. M. Hunt, et al. (2006). "Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*." Mol Ecol **15**(11): 3469-3480.

Jump, A. S. and J. Penuelas (2006). "Extensive spatial genetic structure revealed by AFLP but not SSR molecular markers in the wind-pollinated tree, *Fagus sylvatica*." Mol Ecol **16**: 925-936.

Jump, A. S., R. Marchant, et al. (2008). "Environmental change and the option value of genetic diversity." Trends Plant Sci **14**: 1360-1385.

Jump, A. S., L. Rico, et al. (2012). "Wide variation in spatial genetic structure between natural populations of the European beech (*Fagus sylvatica*) and its implications for SGS comparability." Heredity **108**(6): 633-639.

Kahn, F. (1987). "The distribution of palms as a function of local topography in Amazonian terra-firme forests." Experientia **43**: 251-259.

Kanagaraj, R., T. Wiegand, et al. (2011). "Tropical tree species assemblages in topographical habitats change in time and with life stage." Journal of Ecology **99**: 1441-1452.

Kaufman, S. R., P. E. Smouse, et al. (1998). "Pollen-mediated gene flow and differential male reproductive success in a tropical pioneer tree, *Cecropia obtusifolia* Bertol. (Moraceae): a paternity analysis." Heredity **81**(2): 164-173.

Kawecki, T. J. and D. Ebert (2004). "Conceptual issues in local adaptation." Ecol Lett **7**: 1225-1241.

Kimura, M. (1985). The neutral theory of molecular evolution, Cambridge University Press.

Klein, E. K. and S. Oddou-Muratorio (2011). "Pollen and seed dispersal inferred from seedling genotypes: the Bayesian revolution has passed here too." Mol Ecol **20**(6): 1077-1079.

Klironomos, F. D., J. Berg, et al. (2013). "How epigenetic mutations can affect genetic evolution: model and mechanism." Bioessays **35**: 571-578.

Konuma, A., Y. Tsumura, et al. (2000). "Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis." Mol Ecol **9**(11): 1843-1852.

Koonin, E. and Y. Wolf (2009). "Is evolution Darwinian or/and Lamarckian?" Biol Direct **4**(1): 42.

Koufopanou, V., M. R. Goddard, et al. (2002). "Adaptation for Horizontal Transfer in a Homing Endonuclease." Mol Biol Evol **19**(3): 239-246.

Kozlowski, T. T. (1997). "Responses of woody plants to flooding and salinity." Tree Physiology Monograph **17**(7): 1-29.

Kraft, N. J., R. Valencia, et al. (2008). "Functional Traits and Niche-Based Tree Community Assembly in an Amazonian Forest." Science **322**: 580-582.

Kremer, A. and V. Le Corre (2012). "Decoupling of differentiation between traits and their underlying genes in response to divergent selection." Heredity **108**: 375-385.

Kriuki, M., M. rolfe, et al. (2006). "Diameter growth performance varies with species functional-group and habitat characteristics in subtropical rainforests." For Ecol Manage **225**: 1-14.

Kruuk, L. E. B., J. Slate, et al. (2008). "New answers for old questions: The evolutionary quantitative genetics of wild animal populations." Annual review of Ecology, Evolution and Systematics **39**: 525-548.

Kumar, S. and M. Blaxter (2010). "Comparing de novo assemblers for 454 transcriptome data." BMC Genomics **11**(1): 571.

Lande, R. and J. S. Arnold (1983). "The measurement of selection on correlated characters." Evolution Int J Org Evolution **37**(6): 1210-1226.

Latimer, A. M., J. A. Silander, et al. (2005). "Neutral Ecological Theory Reveals Isolation and Rapid Speciation in a Biodiversity Hot Spot." Science **309**(5741): 1722-1725.

Latouche-Halé, C., A. Ramboer, et al. (2004). "Long-distance pollen flow and tolerance to selfing in a neotropical tree species." Mol Ecol **13**(5): 1055-1064.

Le Corre, V. and A. Kremer (2012). "The genetic differentiation at quantitative trait loci under local adaptation." Mol Ecol **21**(7): 1548-1566.

Le Provost, G., J. Paiva, et al. (2003). "Seasonal variation in transcript accumulation in wood-forming tissues of maritime pine (*Pinus pinaster* Ait.) with emphasis on a cell wall glycine-rich protein." Planta **217**: 820-830.

Leimu, R. and M. Fischer (2008). "A meta-analysis of Local adaptation in plants." PLoS ONE **3**: 1-8.

Lenormand, T. (2002). "Gene flow and the limits to natural selection." Trends in Ecology & Evolution **17**(4): 183-189.

Leonardi, S. and P. Menozzi (1996). "Spatial structure of genetic variability in natural stands of *Fagus sylvatica L.* (beech) in Italy." Heredity **77**: 359-368.

Lexer, C., J. A. Joseph, et al. (2010). "Genomic Admixture Analysis in European Populus spp. Reveals Unexpected Patterns of Reproductive Isolation and Mating." Genetics **186**(2): 699-712.

Li, H. and W. Stephan (2006). "Inferring the Demographic History and Rate of Adaptive Substitution in *Drosophila*." PLoS Genet **2**(10): e166.

Li, W., R. Fu, et al. (2008). "Observed change of the standardized precipitation index, its potential cause and implications to future climate change in the Amazon region." Philosophical Transactions of The Royal Society B: Biological Sciences **363**(1498): 1767-1772.

Librado, P. and J. Rozas (2009). "DnaSP v5: a software for comprehensive analysis of DNA polymorphism data." Bioinformatics Applications in The Biosciences **25**: 1451-1452.

Linhart, Y. B. and M. C. Grant (1996). "Evolutionry significance of local genetic differentiation in plants." Annu Rev Ecol Syst **27**: 237-277.

Linnen, C. R., E. P. Kingsley, et al. (2009). "On the origin and spread of an adaptive allele in deer mice." Science (80-) **325**: 1095-1098.

Lister, R., B. D. Gregory, et al. (2009). "Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond." Curr Opin Plant Biol **12**(2): 107-118.

Loiselle, B. A., V. L. Sork, et al. (1995). "Spatial Genetic Structure of a Tropical Understory Shrub, *Psychotria officinalis* (Rubiaceae)." Am J Bot **82**: 1420-1425.

Lola da Costa, A., D. Galbraith, et al. (2010). "Effects of 7 yr experimental drought on vegetation dynamics and biomass storage of an eastern Amzonian rainforest." New Phytologist **187**: 579-591.

Lopez, O. R. and T. A. Kursar (2003). "Does flood tolerance explain tree species distribution in tropical seasonally flooded habitats?" Oecologia **136**: 193-204.

Lowe, A. J., B. Jourde, et al. (2003). "Fine-scale genetic structure and gene flow within Costa Rican populations of mahogany (*Swietenia macrophylla*)." Heredity **90**(3): 268-275.

Luikart, G., P. R. England, et al. (2003). "The power and promise of population genomics: from genotyping to genome typing." Nat Rev Genet **4**(12): 981-994.

Luizao, R. C., F. J. Luizao, et al. (2004). "Variation of carbon and nitrogen cycling processes along a topographic gradient in a central Amazonian forest." Global Change Biology **10**: 592-600.

Lunn, D. J., A. Thomas, et al. (2000). "WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility." Statistics and Computing **10**: 325-337.

Makinen, H. S., J. M. Cano, et al. (2008). "Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (*Gasterosteus aculeatus*) populations." Mol Ecol **17**(15): 3565-3582.

Malhi, Y. and J. Wright (2004). "Spatial patterns and recent trends in the climate of tropical rainforest regions." Philosophical Transactions of The Royal Society B: Biological Sciences **359**: 311-329.

Manel, S., M. K. Schwartz, et al. (2003). Landscape genetics: combining landscape ecology and population genetics, Elsevier Science Publishers. **18:** 189-197.

Manel, S., B. N. Poncet, et al. (2010). "Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*." Mol Ecol **19**(17): 3824-3m4835.

Margam, V. M., B. S. Coates, et al. (2011). "Transcriptome Sequencing, and Rapid Development and Application of SNP Markers for the Legume Pod Borer *Maruca vitrata* (Lepidoptera: Crambidae)." PLoS ONE **6**(7): e21388.

Marthews, T. R., D. F. R. P. Burslem, et al. (2008). "Modelling Direct Radiation and Canopy Gap Regimes in Tropical Forests." Biotropica **40**(6): 676-685.

Mayle, F. E. and M. J. Power (2008). "Impact of a drier Early Mid-Holocene climate upon Amazonian forests." Philosophical Transactions of The Royal Society B: Biological Sciences **363**(1498): 1829-1838.

Medri, C., E. A. Ruas, et al. (2011). "Genetic diversity and flooding survival in *Aegiphila sellowiana* (Lamiaceae), a typical tree species from upland riparian forests." Genetic and Molecular Research **10**: 1084-1091.

Meier, C. I. and C. Leuschner (2008). "Genotypic variation and phenotypic plasticity in the drought response of fine roots of European beech." Tree Physiol **28**(2): 13.

Meier, K., M. M. Hansen, et al. (2011). "An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta L.*): footprints of sm4election at microsatellite DNA loci." Heredity **106**(3): 488-499.

Meudt, H. M. and A. C. Clarke (2007). "Almost Forgotten or Latest Practice? AFLP applications, analyses and advances." Trends Plant Sci **12**: 106-117.

Meyer, C. L., R. Vitalis, et al. (2009). "Genomic pattern of adaptive divergence in *Arabidopsis halleri*, a model species for tolerance to heavy metal." Mol Ecol **18**(9): 2050-2062.

Miller, N. J., M. Ciosi, et al. (2007). "Genome scan of *Diabrotica virgifera* for genetic variation associated with crop rotation tolerance." Journal of Applied Entomology **131**(6): 378-385.

Mimura, M. and S. N. Aitken (2010). "Local adaptation at the range peripheries of *Sitka spruce*." J Evol Biol **23**(2): 249-258.

Miner, G. B., E. S. Sultan, et al. (2005). "Ecological consequences of phenotypic plasticity." Trends in Ecology and Evolution **20**(12): 685-692.

Miranda-Moreno, L. F., A. Labbe, et al. (2007). "Bayesian multiple testing procedures for hotspot identification." Accident Analysis and Prevention **39**: 1192-1201.

Mittelbach, G. G., D. W. Schemske, et al. (2007). "Evolution and the latitudinal diversity gradient: speciation, extinction and biogeography." Ecol Lett **10**(4): 315-331.

Mittler, R. (2002). "Oxidative stress, antioxidants and stress tolerance." Trends Plant Sci **7**(9): 405-410.

Molino, J.-F. and D. Sabatier (2001). "Tree Diversity in Tropical Rain Forests: A Validation of the Intermediate Disturbance Hypothesis." Science **294**(5547): 1702-1704.

Montesinos-Navarro, A., J. Wig, et al. (2011). "*Arabidopsis thaliana* populations show clinal variation in a climatic gradient associated with altitude." New Phytologist **189**(1): 282-294.

Morozova, O. and A. M. Marra (2008). "Applications of next-generation sequencing technologies in functional genomics." Genomics **92**(5): 10.

Morozova, O., M. Hirst, et al. (2009). "Applications of New Sequencing Technologies for Transcriptome Analysis." Annu Rev Genomics Hum Genet **10**(1): 135-151.

Mousseau, T. A. and C. W. Fox (1998). "The adaptive significance of maternal effects." Trends in Ecology & Evolution **13**(10): 403-407.

Murray, M. C. and M. P. Hare (2006). "A genomic scan for divergent selection in a secondary contact zone between Atlantic and Gulf of Mexico oysters, *Crassostrea virginica*." Mol Ecol **15**(13): 4229-4242.

Namroud, M. C., J. Beaulieu, et al. (2008). "Scanning the genome for gene single nucleotide polymorphisms involved in adaptative population diferenciation in white spruce." Mol Ecol **17**: 3599-3613.

Narum, S. R. and J. E. Hess (2011). "Comparison of FST outlier tests for SNP loci under selection." Molecular Ecology Resources **11**: 184-194.

Nason, J. D., E. A. Herre, et al. (1998). "The breeding structure of a tropical keystone plant resource." Nature **391**(6668): 685-687.

Navarro, C., S. Ward, et al. (2002). "The tree *Cedrela odorata* (Meliaceae): A morphologically subdivided species in Costa Rica." Rev Biol Trop **50**: 21-29.

Navarro, C., F. Montagnigni, et al. (2004). "Genetic variability of *Cedrela odorata* Linnaeus: results of early performance of provenances and families from Mesoamerica grown in association with coffee." Forest Ecology & Management **192**: 217-227.

Neale, D. B. and O. Savolaineen (2004). "Association genetics of complex traits in conifers." Trends Plant Sci **9**: 325-330.

Neelin, J. D., M. Münnich, et al. (2006). "Tropical drying trends in global warming models and observations." PNAS **103**: 6110-6115.

Nei, M. (1987). Molecular Evolutionary Genetic, Columbia University Pres, New York.

Nguyen-Queyrens, A. and F. Bouchet-Lannat (2003). "Osmotic adjustment in three-year-old seedlings of five provenances of maritime pine (*Pinus pinaster*) in response to drought." Tree Physiol **23**: 397-404.

Nielsen, R., S. Williamson, et al. (2005). "Genomic scans for selective sweeps using SNP data." Genome Res **15**(11): 1566-1575.

Nielsen, R., M. J. Hubisz, et al. (2009). "Darwinian and demographic forces affecting human protein coding genes." Genome Res **19**(5): 838-849.

Nijhout, H. F. (2003). "Development and evolution of adaptive polyphenisms." Evol Dev **5**: 9-18.

Nosil, P., S. P. Egan, et al. (2008). "Heterogeneous genomic differentiation between walking-stick ecotypes: Isolation by adaptation and multiple roles for divergent selection." Evolution Int J Org Evolution **62**(2): 316-336.

Nosil, P. and C. P. Sandoval (2008). "Ecological Niche Dimensionality and the Evolutionary Diversification of Stick Insects." PLoS ONE **3**(4): e1907.

Nosil, D. P. Funk, et al. (2009). "Divergent selection and heterogeneous gnomic divergence." Mol Ecol **18**: 375-402.

Novaes, E., D. Drost, et al. (2008). "High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome." BMC Genomics **9**(1): 312.

Nussey, D. H., A. J. Wilson, et al. (2007). "The evolutionary ecology of individual phenotypic plasticity in wild populations." J Evol Biol **20**(3): 831-844.

Nystedt, B., N. R. Street, et al. (2013). "The Norway spruce genome sequence and conifer genome evolution." Nature **advance online publication**: 579-584.

O'Brien, E. K., R. A. Mazanec, et al. (2007). "Provenance variation of ecologically important traits of forest trees: implications for restoration." Journal of Applied Ecology **44**: 583-593.

Oddou-Muratorio, S., A. Bontemps, et al. (2010). "Comparison of direct and indirect genetic methods for estimating seed and pollen dispersal in *Fagus sylvatica* and *Fagus crenata*." For Ecol Manage **259**: 2151-2159.

Oden, N. D. and R. R. Sokal (1986). "Directional autocorrelation: an extension of spatial correlograms to two dimensions." Syst Zool **35**: 608-617.

Oleksyn, J., J. Modrzynski, et al. (1998). "Growth and physiology of Picea abies populations from elevational transects: common garden evidence for altitudinal ecotypes and cold adaptation." Functional Ecology **12**(4): 573-590.

Olson, D. M., E. Dinerstein, et al. (2001). "Terrestrial Ecopregions of the World: A New Map of Life on Earth." Bioscience **51**: 933-938.

Palmiotto, P. A., S. J. Davies, et al. (2004). "Soil-related habitat specialization in dipterocarp rain forest tree species in Borneo." Journal of Ecology **92**: 609-623.

Parchman, T., K. Geist, et al. (2010). "Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery." BMC Genomics **11**(1): 180.

Parelle, J., E. Dreyer, et al. (2010). Genetic variability and Determinism of adaptation of plants to soil waterlogging. Waterlogging signalling and tolerance in plants, Springer-Verlag, Heidelberg (DEU)**:** 241-265.

Pariset, L., S. Joost, et al. (2009). "Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean." BMC Genet **10**(1): 7.

Perata, P., W. Armstrong, et al. (2011). "Plants and flooding stress." New Phytologist **190**(2): 269-273.

Petit, R. J. and A. Hampe (2006). "Some Evolutionary Consequences of Being a Tree." Annual Review of Ecology, Evolution, and Systematics **37**(1): 187-214.

Phillips, P. C. (2005). "Testing hypotheses regarding the genetics of adaptation." Genetica **123**: 15-24.

Phillips and et al. (2009). "Drought sensitivity of the Amazon rainforest." Science **323**: 1344-1347.

Picoult-Newberg, L., T. E. Ideker, et al. (1999). "Mining SNPs From EST Databases." Genome Res **9**(2): 167-174.

Pigliucci, M. (2001). Phenotypic plasticity : beyond nature and nurture, The Johns Hopkins University Press.

Pigliucci, M., J. Murren, et al. (2006). "Phenotypic plasticity and evolution by genetic assimilation." J Exp Biol **209**: 2362-2367.

Plotkin, J. B., M. D. Potts, et al. (2000). "Species-area Curves, Spatial Aggregation, and Habitat Specialization in Tropical Forests." J Theor Biol **207**(1): 81-99.

Poncet, B. N., D. Hermann, et al. (2010). "Tracking genes of ecological relevance using a genome scan in two independent regional population samples of *Arabis alpina*." Mol Ecol **19**(14): 2896-2907.

Ponnamperuma, F. N. (1972). "The chemistry of submerged soils." Advances in Agronomy **24**: 29-96.

Pons, O. and R. J. Petit (1996). "Measuring and testing genetic differentiation with ordered versus unordered alleles." Genetics **144**: 1237-1245.

Poorter, L. (1999). "Growth responses of 15 rain-forest tree species to a light gradient: the relative importance of morphological and physiological traits." Functonal ecology **13**: 396-410.

Poorter, L. and F. Bongers (2006). "Leaf traits are good predictors of plant performance across 53 rain forest species." Ecology **87**: 1733-1743.

Poorter, L. and L. Markesteijn (2008). "Seedling Traits Determine Drought Tolerance of Tropical Tree Species." Biotropica **40**(3): 321-331.

Poorter, L., S. J. Wright, et al. (2008). "Are functional traits good predictors of demographic rates? Evidence from five neotropical forests." Ecology **89**(7): 1908-1920.

Poorter, H., U. l. Niinemets, et al. (2009). "Causes and consequences of variation in leaf mass per area (LMA): a meta-analysis." New Phytologist **182**(3): 565-588.

Pop, M. and S. L. Salzberg (2008). "Bioinformatics challenges of new sequencing technology." Trends in Genetics **24**: 142-149.

Pritchard, J. K., M. Stephens, et al. (2000). "Inference of Population Structure Using Multilocus Genotype Data." Genetics **155**(2): 945-959.

Prunier, J., J. Laroche, et al. (2011). "Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce." Mol Ecol **20**: 1702-1716.

Qian, S. S. and Z. Shen (2007). "Ecological Applications of Multilevel Analysis of Variance." Ecological applications of multilevel analysis of variance **10**: 2489-2495.

Quesada, C. A., J. Lloyd, et al. (2009). "Regional and large-scale patterns in Amazon forest structure and function are mediated by variations in soil physical and chemical properties." Biogeosciences Discussions **6**(2): 3993-4057.

Rafalski, A. (2002). "Applications of single nucleotide polymorphisms in crop genetics." Curr Opin Plant Biol **5**(2): 94-100.

Rajakaruna, N., G. E. Bradfield, et al. (2003). "Adaptive differentiation in response to water stress by edaphic races of *Lasthenia californica* (Asteraceae)." Int J Plant Sci **164**: 371-376.

Randall Hugues, A., B. D. Inouye, et al. (2008). "Ecological consequences of genetic diversity." Ecol Lett **11**: 609-623.

Reich, P. B., M. B. Walters, et al. (1994). "Photosynthesis-nitrogen relations in Amazonian tree species." Oecologia **97**: 62-72.

Remington, D. L., J. M. Thornsberry, et al. (2001). "Structure of linkage disequilibrium and phenotypic associations in the maize genome." Proceedings of The National Academy of Sciences **981**: 11479-11484.

Rice, K. J., D. R. Gordon, et al. (1993). "Phenotypic variation in seedlings of a 'keystone' tree species (Quercus douglasii): the interactive effects of acorn source and competitive environment." Oecologia **96**: 537-547.

Richardson, B., G. Rehfeldt, et al. (2009). "Congruennt climate-related genecological responses from molecular markers and quantitative traits for western white pine (*Pinus monticola*)." Int J Plant Sci **170**: 1120-1131.

Ricklefs, R. E. (1977). "Environmental heterogeneity and plant species diversity: a hypothesis." American Naturalist **111**: 376-381.

Ricklefs, R. E. (2006). "The unified neutral theory of biodiversity: Do the numbers add up?" Ecology **87**: 1424-1431.

Ridley, M. (2003). Evolution, Wiley-Blackwell.

Riebler, A., L. Held, et al. (2008). "Bayesian Variable Selection for Detecting Adaptive Genomic Differences Among Populations." Genetics **178**: 1817-1829.

Rix, K., A. Gracie, et al. (2012). "Paternal and maternal effects on the response of seed germination to high temperatures in *Eucalyptus globulus*." Ann For Sci **69**: 673-679.

Robert, A. (2003). "Simulation of the effect of topography and tree falls on stand dynamics and stand structure of tropical forests." Ecol Modell **167**(3): 287-303.

Robinson, M. R., A. J. Wilson, et al. (2009). "The impact of environmental heterogeneity on genetoc architecture in a wild popultion of Soay Sheep." Genetics **181**: 1639-1648.

Rochon, C., H. A. Margolis, et al. (2007). "Genetic variation in growth of *Guazuma crinita* (Mart.) trees at an early age in the Peruvian Amazon." For Ecol Manage **243**(2-3): 291-298.

Roulin, A. and C. Dijkstra (2003). "Genetic and environmental components of variation in eumelanin and phaeomelanin sex-traits in the barn owl." Heredity **90**: 359-364.

Rundle, H. D. and P. Nosil (2005). "Ecological speciation." Ecol Lett **8**: 336-352.

Russo, S. E., S. J. Davies, et al. (2005). "Soil-related performance variation and distributions of tree species in a Bornean rain forest." Journal of Ecology **93**: 879-889.

Sabatier, D., M. Grimaldi, et al. (1997). "The influence of soil cover organization on the floristic and structural heterogeneity of a Guianan rain forest." Plant Ecology **131**: 81-108.

Salvaudon, L., T. Giraud, et al. (2008). "Genetic diversity in natural populations: a fundamental component of plant–microbe interactions." Curr Opin Plant Biol **11**: 135-143.

Sanchez-Gomez, D., F. Valladares, et al. (2006). "Performance of seedlings of Mediterranean woody species under experimental gradients of irradiance and water availability: trade-offs and evidence for niche differentiation." New Phytologist **170**(4): 795-806.

Sarkar, S. and T. Fuller (2003). "Generalized norms of reaction for ecological developmental biology." Evol Dev **5**(1): 106-115.

Savolainen, O., F. Bokma, et al. (2004). "Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of *Pinus sylvestris* to climatic changes." For Ecol Manage **197**: 79-89.

Savolainen, O., T. Pyhäjärvi, et al. (2007). "Gene Flow and Local Adaptation in Trees." Annu. Rev. Ecol. Evol. Syst. **38**: 595-619.

Savolainen, V., M.-C. Anstett, et al. (2006). "Sympatric speciation in palms on an oceanic island." Nature **441**(7090): 210-213.

Savolainen, O. (2011). "The Genomic Basis of Local Climatic Adaptation." Science **334**(6052): 49-50.

Scascitelli, M., K. D. Whitney, et al. (2010). "Genome scan of hybridizing sunflowers from Texas (Helianthus annuus and H. debilis) reveals asymmetric patterns of introgression and small islands of genomic differentiation." Mol Ecol **19**(3): 521-541.

Scheffer, S. J. and D. J. Hawthorne (2007). "Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): apparent discordance among marker systems." Mol Ecol **16**(13): 2627-2637.

Schemske, D. W. (1984). "Population Structure and Local Selection in *Impatiens pallida* (Balsaminaceae), A Selfing Annual." Evolution Int J Org Evolution **38**: 817-832.

Schlichting, C. D. and H. Smith (2002). "Phenotypic plasticity: linking molecular mechanisms with evolutionary outcomes." Evol Ecol **16**: 189-211.

Schlõtterer, C. (2002). "Towards a molecular characterization of adaptation in local populations." Current Opinion in Genetics & Development **12**(6): 683-687.

Schluter, D. (2001). "Ecology and the origin of species." Trends in Ecology and Evolution **16**: 372-380.

Schmitt, J. and S. Gamble (1990). "The effect of distance from the parental site on offspring performance and inbreeding depression in *Impatiens capensis*: a test of the local adaptation hypothesis." Evolution Int J Org Evolution **44**: 2022-2030.

Schnitzer, S. A. and W. P. Carson (2001). "Treefall Gaps and the Maintenance of Species Diversity in a Tropical Forest." Ecology **82**(4): 913-919.

Schroeder, J. I., J. M. Kwak, et al. (2001). "Guard cell abscisic acid signalling and engineering drought hardiness in plants." Nature **410**(6826): 327-330.

Scotti, I. (2010). "Adaptative potential in forest tree populations: what is it, and how can measure it?" Ann For Sci **67**: 801.

Scotti, I., L. Calvo-Vialettes, et al. (2010). "Genetic variation for growth, morphological, and physiological traits in a wild population of the Neotropical shade-tolérant rainforest tree *Sextonia rubra* (Mez) van der Werff (Lauraceae)." Tree Genetics and Genomes **6**: 319-329.

Scotti-Saintagne, C., S. Mariette, et al. (2004). "Genome Scanning for Interspecific Differentiation Between Two Closely Related Oak Species [Quercus robur L. and Q. petraea (Matt.) Liebl.]." Genetics **168**(3): 1615-1626.

Scotti-Saintagne, C., C. W. Dick, et al. (2012). "Phylogeography of a species complex of lowland Neotropical rain forest trees (*Carapa*, Meliaceae)." Journal of Biogeography **40**(4): 676-692.

Scotti-Saintagne, C., C. W. Dick, et al. (2012). "Amazon diversification and cross-Andean dispersal of the widespread Neotropical tree species *Jacaranda copaia* (Bignoniaceae)." Journal of Biogeography **40**(4): 707-719.

Seeb, J. E., G. Carvalho, et al. (2011). "Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms." Molecular Ecology Resources **11**: 1-8.

Shao, H.-B., Q.-J. Guo, et al. (2007). "Understanding molecular mechanism of higher plant plasticity under abiotic stress." Colloids and Surfaces B: Biointerfaces **54**(1): 37-45.

Silva, D. C. G., M. C. C. G. Carvalho, et al. (2010). "Evidence of ecotypic differentiation between populations of the tree species *Parapiptadenia rigida* due to flooding." Genetics and Molecular Research **2**: 797-810.

Siol, M., S. I. Wright, et al. (2010). "The population genomics of plant adaptation." New Phytologiist **188**: 313-332.

Slatkin, M. and L. Voelm (1991). "FST in a hierarchical island model." Genetics **127**(3): 627-629.

Slatkin, M. and L. Excoffier (1996). "Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm." Heredity **76**: 377-383.

Sloan, D. B., S. R. Keller, et al. (2012). "De novo transcriptome assembly and polymorphism detection in the flowering plant *Silene vulgaris* (Caryophyllaceae)." Molecular Ecology Resources **12**: 333-343.

Sogaard, G., A. Granhus, et al. (2009). "Effect of frost nights and day and night temperatures during dormancy induction on frost hardiness, tolerance to cold storage of Norway spruce." Trees **23**: 1295-1307.

Sogaard, G., O. Johnsen, et al. (2008). "Climatic control of bud burst in young seedlings of nine provenances of Norway spruce." Tree Physiol **28**: 311-320.

Sork, V. L., K. A. Stowe, et al. (1993). "Evidence for local adaptation in closely adjacent sub-populations of northern Red Oak (*Quercu rubra* L.) exressed as a resistance to leaf herbivores." Am Nat **142**: 928-936.

Sotelo Montes, C., J. Beaulieu, et al. (2007). "Genetic variation in wood shrinkage and its correlations with tree growth and wood density of *Calycophyllum spruceanum* at an early age in the Peruvian Amazon." Canadian Journal of Forest Research **37**(5): 966-976.

Sotelo-Montes, C., J. Beaulieu, et al. (2007). "Genetic variation in wood shrinkage and its correlations with tree growth and wood density of *Calycophyllum spruceanum* at an early age in the Peruvian Amazon." Canadian Journal of Forest Research **37**: 966-976.

Soto-Cerda, B. J. and S. Cloutier (2013). "Outlier Loci and Selection Signatures of Simple Sequence Repeats (SSRs) in Flax (*Linum usitatissimum* L.)." Plant Molecular Biology Reporter: 1-13.

Stacy, E. A., J. L. Hamrick, et al. (1996). "Pollen dispersal in low-density populations of three neotropical tree species." American Naturalist **148**(2): 275-298.

Stapley, J., J. Reger, et al. (2010). "Adaptation genomics: the next generation." Trends in Ecology and Evolution **25**: 705-711.

Stephens, M. and P. Donnelly (2003). "A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data." Am J Hum Genet **73**: 1162-1169.

Stephens, M., N. J. Smith, et al. (2001). "A New Statistical Method for Haplotype Reconstruction from Population Data." Am J Hum Genet **68**: 978-989.

Storz, J. F. and J. M. Dubach (2004). "Natural selection drives altitudinal divergence at the albumin locus in deer mice, *Peromyscus maniculatus*." Evolution Int J Org Evolution **58**(6): 1342-1352.

Storz, J. F. (2005). "Population genetics: Nonrandom dispersal and local adaptation." Heredity **95**: 3-4.

Storz, J. F. (2005). "Using genome scans of DNA polymorphism to infer adaptive population divergence." Mol Ecol **14**: 671-688.

Storz, J. F. and J. K. Kelly (2008). "Effects of Spatially Varying Selection on Nucleotide Diversity and Linkage Disequilibrium: Insights From Deer Mouse Globin Genes." Genetics **180**: 367-379.

Strasburg, J. L., C. Scotti-Saintagne, et al. (2009). "Genomic Patterns of Adaptive Divergence between Chromosomally Differentiated Sunflower Species." Mol Biol Evol **26**(6): 1341-1355.

Strasburg, J. L., N. A. Sherman, et al. (2011). "What can patterns of differentiation across plant genomes tell us about adaptation and speciation?" Phil. Trans. R. Soc. B **367**: 364-373.

Street, N. R., O. Skogström, et al. (2006). "The genetics and genomics of the drought response in *Populus*." The Plant Journal **48**: 32-341.

Streiff, R., T. Labbe, et al. (1998). "Within-population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isozymes and microsatellites." Mol Ecol **7**: 317-328.

Stute, M., M. Forster, et al. (1995). "Cooling of Tropical Brazil (5°C) during the Last Glacial Maximum." Science **269**: 379-383.

ter Steege, H., V. G. Jetten, et al. (1993). "Tropical rain forest types and soil factors in a watershed area in Guyana." Journal of Vegetation Science **4**(5): 705-716.

ter Steege, H. and D. S. Hammond (2001). "Character convergence, diversity and disturbance in tropical rainforest in Guyana." Ecology **82**: 3197-3212.

ter Steege, H., N. C. A. Pitman, et al. (2006). "Continental-scale patterns of canopy tree composition and function across Amazonia." Nature **443**(7110): 444-447.

Terborgh, J., N. Pitman, et al. (2002). Seed dispersal and frigivory: Ecology, Evolution and Conservation Maintenance of Tree Dievsrity in Troical Forests, CABI Publishing.

Torres, T. T., M. Metta, et al. (2008). "Gene expression profiling by massively parallel sequencing." Genome Res **18**: 172-177.

Townsend, A. R., C. C. Cleveland, et al. (2007). "Controls over foliar N:P ratios in tropical rain forests." Ecology **88**(5): 107-118.

Tsumura, Y., T. Kado, et al. (2007). "Genome Scan to Detect Genetic Structure and Adaptive Genes of Natural Populations of *Cryptomeria japonica*." Genetics **176**(4): 2393-2403.

Tuomisto, H. and K. Ruokolainen (1997). "The role of ecological knowledge in explaining biogeography and biodiversity in Amazonia." Biodiversity and Conservation **6**: 347-357.

Tuomisto, H. (2006). "Edaphic niche differentiation among Polybotrya ferns in western Amazonia: implications for coexistence and speciation." Ecography **29**(3): 273-284.

Turner, T. L., E. C. Bourne, et al. (2010). "Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils." Nat Genet **42**(3): 260-263.

Valencia, R., R. B. Foster, et al. (2004). "Tree species distributions and local habitat variation in the Amazon: large forest plot in eastern ecuador." Journal of Ecology **92**: 214-229.

Van Kleunen, M. and M. Fischer (2005). "Constraints on the evolution of adaptive phenotypic plasticity in plants." New Phytologist **166**(1): 12.

Van Tassell, C. P., T. P. L. Smith, et al. (2008). "SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries." Nat Meth **5**(3): 247-252.

Vasemagi, A. and C. R. Primmer (2005). "Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies." Mol Ecol **14**(12): 3623-3642.

Vasquez, J. and W. S. Dvorak (1996). "Trends in variances and heritabilities with stand development of tropical pines." Canadian Journal of Forest Research **26**: 1473-1480.

Vekemans, X. and O. J. Hardy (2004). "New insights from fine-scale spatial genetic structure analyses in plant populations." Mol Ecol **13**: 921-935.

Vera, J. C., C. W. Wheat, et al. (2008). "Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing." Mol Ecol **17**: 1636-1647.

Veron, V., H. Caron, et al. (2005). "Gene flow and mating system of the tropical tree *Sextonia rubra*." Silvae Genet **54**: 275-280.

Via, S. and J. West (2008). "The genetic mosaic suggests a new role for hitchhiking in ecological speciation." Mol Ecol **17**: 4334-4345.

Vilas, A., F. A. Perez, et al. (2012). "A simulation study on the performance of differentiation-based methods to detect selected loci using linked neutral markers." J Evol Biol **25**(7): 1364-1376.

Vincent, G., J.-F. Molino, et al. (2011). "The relative importance of dispersal limitation and habitat preference in shaping spatial distribution of saplings in a tropical moist forest: a case study along a combination of hydromorphic and canopy disturbance gradients." Ann For Sci **68**: 357-370.

Violle, C., M.-L. Navas, et al. (2007). "Let the concept of trait be functional!" Oikos **116**(5): 882-892.

Visscher, P. M., W. G. Hill, et al. (2008). "Heritability in the genomics era: concepts and misconceptions." Nat Rev Genet **9**(4): 255-266.

Vormisto, J., H. Tuomisto, et al. (2004). "Palm distribution patterns in Amazonian rainforests: What is the role of topographic variation?" Journal of Vegetation Science **15**: 485-494.

Vornam, B., N. Decarli, et al. (2004). "Spatial Distribution of Genetic Variation in a Natural Beech Stand (*Fagus sylvatica* L.) Based on Microsatellite Markers." Conservation Genetics **5**: 561-570.

Vos, P., R. Hogers, et al. (1995). "AFLP: a new technique for DNA fingerprinting." Nucleic Acids Res **23**: 4407-4414.

Wagner, F. (2011). Thèse: La réponse des forêts tropicales humides aux variations climatiques : évolution de la structure et de la dynamique des peuplements forestiers guyanais.

Ward, M., C. W. Dick, et al. (2005). "To self, or not to self: A review of outcrossing and pollen-mediated gene flow in neotropical trees." Heredity **95**(4): 246-254.

Ward, S. E., K. E. Wightman, et al. (2008). "Early results from genetic trials on the growth of Spanish cedar and its susceptibility to the shoot borer moth in the Yucatan Peninsula, Mexico." For Ecol Manage **255**(2): 356-364.

Watterson, G. A. (1975). "On the number of segregating sites in genetical models without recombination." Theor Popul Biol **7**: 256-276.

Webb, C. O. and D. R. Peart (2000). "Habitat associations of trees and seedlings in a Bornean rain forest." Journal of Ecology **88**(3): 464-478.

Weber, A. P. M., K. L. Weber, et al. (2007). "Sampling the *Arabidopsis* Transcriptome with Massively Parallel Pyrosequencing." Plant Physiol **144**: 32-42.

Wei, X. and N. M. G. Borralho (1997). "Genetic Control of Wood Basic Density and Bark Thickness and Their Relationships with Growth Traits of *Eucalyptus urophylla* in South East China." Silvae Genet **46**: 245-250.

Weir, B. S. and C. C. Cockerham (1984). "Estimating F-statistics for the analysis of population structure." Evolution Int J Org Evolution **38**: 1358-1370.

Westoby, M., D. S. Falster, et al. (2002). "Plant Ecological Strategies: Some Leading Dimensions of Variation Between Species." Annu Rev Ecol Syst **33**(1): 125-159.

White, G. M., D. H. Boshier, et al. (2002). "Increased pollen flow counteracts fragmentation in a tropical dry forest: An example from *Swietenia humilis Zuccarini*." Proceedings of The National Academy of Sciences **99**(4): 2038-2042.

Wicker, T., E. Schlagenhauf, et al. (2006). "454 sequencing put to the test using the complex genome of barley." BMC Genomics **7**: 275.

Wiggins, D. A. (1989). "Heritability of body size in cross-fostered Tree Swallow broods." Evolution Int J Org Evolution **43**: 1808-1811.

Wightman, K. E., S. E. Ward, et al. (2008). "Performance and genetic variation of big-leaf mahogany (*Swietenia macrophylla* King) in provenance and progeny trials in the Yucatan Peninsula of Mexico." For Ecol Manage **255**(2): 346-355.

Wilding, C. S., R. K. Butlin, et al. (2001). "Differential gene exchange between parapatric morphs of Littorina saxatilis detected using AFLP markers." J Evol Biol **14**(4): 611-619.

Willekens, H., S. Chamnongpol, et al. (1997). "Catalase is a sink for H_2O_2 and is indispensable for stress defence in C3 plants." EMBO J **16**(16): 4806-4816.

Wilson, A. J., J. M. Pemberton, et al. (2006). "Environmental coupling of selection and heritability limits evolution." PLoS Biol **4**: e216.

Wolf, J. B. and M. J. Wade (2009). "What are maternal effects (and what are they not)?" Philosophical Transactions of The Royal Society B: Biological Sciences **364**(1520): 1107-1115.

Wright, S. (1992). "Seasonal drought, soil fertility and the species density of tropical forest plant communities." Trends in Ecology and Evolution **7**: 260-263.

Wright, S. J. (2002). "Plant diversity in tropical forests: a review of mechanisms of species coexistence." Oecologia **130**: 1-14.

Wright, I. J., M. Westboy, et al. (2002). "Convergence toward higher leaf mass per area in dry and nutrient-poor habitats has different consequences for leaf life span." Journal of Ecology **90**: 534-543.

Wright, I. J., P. B. Reich, et al. (2004). "The worldwide leaf economics spectrum." Nature **428**(6985): 821-827.

Wright, S. I. and B. S. Gaut (2005). "Molecular Population Genetics and the Search for Adaptive Evolution in Plants." Mol Biol Evol **22**: 506-519.

Yahara, K., M. Fukuyo, et al. (2009). "Evolutionary maintenance of selfish homing endonuclease genes in the absence of horizontal transfer." Proceedings of The National Academy of Sciences **106**: 18861-18866.

Yatabe, Y., N. C. Kane, et al. (2007). "Rampant Gene Exchange Across a Strong Reproductive Barrier Between the Annual Sunflowers, *Helianthus annuus* and *H. petiolaris*." Genetics **175**(4): 1883-1893.

You, F., N. Huo, et al. (2011). "Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence." BMC Genomics **12**(1): 59.

Zamudio, F., P. Rozenberg, et al. (2005). "Genetic variation of wood density components in a radiata pine progeny test located in the south of Chile." Ann For Sci **62**(2): 105-114.