Genomics of microgeographic adaptation in the Amazonian hyperdominant tree *Eperua falcata* Aubl. (Fabaceae)

Louise Brousseau[1,2,6], Paul V. A. Fine[3], Erwin Dreyer[4], Giovanni Giuseppe Vendramin[5], Ivan Scotti[6]

[1.] INRA, UMR EcoFoG Ecologie des Forêts de Guyane, Campus Agronomique BP709, 97387 Kourou cedex, French Guiana

[2.] IRD, UMR DIADE Diversité - Adaptation - Développement, 911 Avenue Agropolis, BP64501, 34394 Montpellier, France.

[3.] Department of Integrative Biology, University of California, Berkeley, 1005 Valley Life Science Building, Berkeley, California 94720-3140, USA.

[4.] INRA, Université de Lorraine, UMR EEF Ecologie et Ecophysiologie Forestière, 54280 Champenoux, France.

[5.] Institute of Biosciences and BioResources, National Research Council (IBBR-CNR), Division of Florence, Via Madonna del Piano 10, 50019 Sesto Fiorentino (FI), Italy.

[6.] INRA, URFM Ecologie des Forêts Méditerranéennes, Domaine Saint-Paul, Site Agroparc CS 40509, 84914 Avignon cedex 9, France.

Corresponding Author:

Louise Brousseau. IRD, UMR DIADE, 911 Avenue Agropolis, BP 64501, 34394 Montpellier, France. louise.brousseau@ird.fr

Abstract

Plant populations can undergo very localized adaptation, allowing widely distributed populations to adapt to divergent habitats in spite of recurrent gene flow. Neotropical trees - whose large and undisturbed populations span across a variety of environmental conditions and local habitat - are particularly good models to study this process. Here, we carried out a genome scan for selection through whole-genome sequencing of pools of populations, sampled according to a nested sampling design, to evaluate microgeographic adaptation in the hyperdominant Amazonian tree *Eperua falcata* Aubl. (Fabaceae). A high-coverage genomic resource of ~250 Mb was assembled *de novo* and annotated, leading to 32,789 predicted genes. 97,062 bi-allelic SNPs were detected over 25,803 contigs, and a custom Bayesian model was implemented to uncover candidate genomic targets of divergent selection. A set of 290 divergence outlier SNPs was detected at the regional scale (between study sites), while 185 SNPs located in the vicinity of 106 protein-coding genes were detected as replicated outliers between microhabitats within regions. These genes potentially underlie ecologically important phenotypes and indicate that adaptation to microgeographic habitat patchiness would affect genomic regions involved in a variety of physiological processes, among which plant response to stress (for e.g., oxidative stress, hypoxia and metal toxicity) and biotic interactions. Identification of genomic targets of microgeographic adaptation in the Neotropics is consistent with the hypothesis that local adaptation is a key driver of ecological diversification, operating across multiple spatial scales, from large- (i.e. regional) to microgeographic- (i.e. landscape) scales.

Key words: Amazonia, Neotropics, *Eperua falcata*, Pool-Seq, *de novo* assembly, genome scan, SNPs, G_{ST}, Bayesian modelling, adaptive divergence, local adaptation.

Introduction

Plant populations can undergo extremely localized adaptation, allowing populations of widely distributed species to adapt to divergent habitats in spite of recurrent gene flow. This phenomenon has been repeatedly observed in annual plants growing in strongly contrasting yet geographically adjacent habitats, particularly in cases with large differences in soil composition (Bradshaw 1960; Jain and Bradshaw 1966; Antonovics and Bradshaw 1970; Antonovics 2006; Gould *et al.* 2014); in some cases, molecular evidence also supports adaptation to habitat mosaics (Turner *et al.* 2010; Fustier *et al.* 2017). In a well-known study of *Howea* palm trees, Savolainen *et al.* (2006) provided strong evidence of a link between adaptation to habitat mosaics and sympatric speciation. The process underlying these instances of adaptation with gene flow has been dubbed 'microgeographic adaptation', whereby adaptive divergence occurs at geographical scales of the same order of magnitude as gene dispersal distance (Richardson *et al.* 2014). Adaptation to multiple optima within a continuous population has been identified as a major mechanism for the maintenance of genetic diversity (Delph and Kelly 2014). Moreover, conditions under which adaptive divergence can occur with gene flow have been explored theoretically with models including two environmental patches connected by varying levels of gene flow (Bulmer 1972; Hendry *et al.* 2001; Yeaman and Guillaume 2009; Yeaman and Otto 2011; Schmid and Guillaume 2017). Surprisingly, these simulations show that a wide range of migration-selection equilibrium states can lead to adaptation to each patch with maintenance of polymorphism (i.e. maintenance of patch specialists, as opposed to the emergence of a generalist genotype).

Data from perennial plants, and from trees in particular, suggest that this may be a general phenomenon, that can be readily observed at the phenotypic level (Yeaman and Jarvis 2006; Brousseau *et al.* 2013; Vizcaíno-Palomar *et al.* 2014; Lind *et al.* 2017) (but see Latreille & Pichot (2017) for a case of a complete lack of microgeographical adaptation). At the molecular level, association between alleles (or genotypes) and habitats has been detected in a variety of studies (see review by Savolainen *et al.* (2007)), suggesting an important role for intra-specific and intra-populational variation in the evolution of habitat specialization (Scotti *et al.* 2016). The advent of genome-wide approaches has increased detection power, making it possible to more precisely evaluate the genetic bases of microgeographic adaptation (Turner *et al.* 2010; Eckert *et al.* 2015; Izuno *et al.* 2017; Fustier *et al.* 2017; Lind *et al.* 2017). In general, those genome scan studies have found only a minority of loci (in the order of a few percent of the total) that exhibits microgeographic disruptive selection. Yet genome-wide,

sequencing-based approaches can go beyond a finer estimation of the number of loci involved in the process, because they permit formulating functional interpretations of the divergence process as sequence annotation can suggest putative function of the target loci (Turner *et al.* 2010; Eckert *et al.* 2010, 2015; Fustier *et al.* 2017).

Microgeographic adaptation involves the filtering of phenotypes and genotypes that favor adaptation to habitat patches, and therefore can be viewed as an initial step towards the diversification of populations, ecotypes and eventually species divergence. In particular, environmental filtering has frequently been invoked as a driver of niche partitioning and species diversification in the Neotropics (Fine *et al.* 2004, 2005, 2013; Baraloto *et al.* 2007; Kraft *et al.* 2008; Fine 2015); intra-populational diversification processes across environmental gradients and/or habitat patches thus may play a key role in the generation of functional diversity. Moreover, species' short-term adaptive potential (Harrisson *et al.* 2014; De Kort and Honnay 2017) largely depends on standing genetic variation. Microgeographic adaptation can thus play a role in the maintenance of adaptive genetic diversity. Measuring its extent in the wild is thus of major importance, as it could provide insights into the ability of populations to evolve towards new phenotypic and genetic optima under global climate change (GCC).

Trees are particularly good models to study microgeographic adaptation: many wild tree species have large populations with widespread distribution spanning a variety of microhabitats, harbor large amounts of genetic diversity, and display large dispersal distances (Hamrick and Godt 1990; Petit and Hampe 2006; Neale and Kremer 2011; Sork *et al.* 2013; Fetter *et al.* 2017). They therefore correspond quite closely to the conditions under which selection-migration equilibrium theory has been developed. Moreover, methods to detect locus-specific divergence (and infer selection) should work quite well with microgeographic processes in populations of the most common forest trees, because their relatively shallow (spatial) population structures and large effective population sizes should minimize the introduction of biases (e.g., because of drift) and the generation of false positives.

Lowland Amazonian rainforests, whose landscapes show fine patchiness of contrasting topographic and edaphic conditions varying over microgeographic scales (**Fig. 1a,c,d**), represent an ideal study system to investigate microgeographic adaptation processes. In Amazonian rainforests, the recent study by ter Steege *et al.* (2013) reported 227 hyperdominant tree species accounting for half of all stems (ter Steege *et al.* 2013). Some of these widespread species are also ecological generalists, whose preferences could be explained by individual phenotypic plasticity or by microgeographic adaptation to habitat

patchiness (Fortunel *et al.* 2016). We focus here on dense, almost monodominant stands of *Eperua falcata* Aubl. (Fabaceae) in French Guiana, that occur across environmentally heterogeneous areas, with the goal of identifying genomic loci undergoing divergence at the microgeographic scale between habitats. To do this, a genome scan was carried out through a nested sampling design combining regional and microgeographic spatial scales (i.e. replicates of microhabitats, ~300 m apart, in two study sites, ~300 km apart). We analyzed patterns of genomic divergence from large (i.e. regional) to microgeographic spatial scales within a hierarchical Bayesian framework (Brousseau *et al.* 2016) designed to identify locus-specific departures from genome-wide divergence in nested population samples.

Materials and Methods

Species model

Eperua falcata Aubl. (**Fig. S1**) is a reportedly non-nodulating (Villadas *et al.* 2007) tree species of Fabaceae (Caesalpinioideae). It is widespread in Northest Amazonia, from North Brazil to Venezuela through the Guiana shield (https://demo.gbif.org/species/2943522), and it is a 'hyperdominant' species, the 13th most abundant tree species of the Amazonian tree flora (ter Steege *et al.* 2013). It is an evergreen, late-successional canopy tree species which is often emergent. It is shade hemi-tolerant (Bonal, Barigah, *et al.* 2000) and can tolerate both drought (Bonal, Atger, *et al.* 2000; Bonal *et al.* 2001) and water-logging (Baraloto *et al.* 2007). Its distribution is very clumped and populations are commonly found at high densities. It is a generalist with respect to microgeographic edaphic conditions, although it is most abundant in water-logged forests of bottomlands (Baraloto *et al.* 2007). *E. falcata* flowers and fruits during the rainy season, and seeds reach maturity in April-May, synchronously among trees in all microhabitats. Its large pollen grains (>100µm) are dispersed by animals such as bats (Cowan 1975), while seed dispersal is autochorous: heavy seeds are dispersed around crowns of mother trees through explosive pod dehiscence.

Sampling design and study sites

We sampled populations that replicated the environmental contrast 'bottomland *versus* terra-firme' (~300 m apart) at two regional locations (i.e. study sites) in French Guiana (~300

km apart), **Fig. 2a**. Laussat (western French Guiana) and Regina (eastern French Guiana) study sites are located on the coastal Guiana shield, about 300 km apart, and experience different rainfall regimes (~2,500 and 3,500 mm/year, respectively) (Brousseau *et al*. 2015). Each regional location harbors different microhabitats (hygromorphic sandy bottomlands and ferralitic *terra-firme* plateaus or hills) separated by short distances (~300 m). In Laussat, a plateau of low elevation is adjacent to a permanently water-logged bottomland, while Regina is composed of higher elevation hilltops bordered by abrupt slopes surrounding a seasonally flooded bottomland forest, **Fig. S2**. *E. falcata* trees are found through the study area, ensuring gene flow between tree populations occupying different microhabitats (Brousseau *et al.* 2015), **Fig. 2b**. *E. falcata* stem (> 20 cm diameter at breast height) density varies between 30 and 50 trees / ha at these locations (Brousseau *et al.* 2015). Twenty adult trees for each of the four *regional location* × *microhabitat* combination were sampled, totaling 80 individuals.

Molecular methods

Genomic DNA was extracted from fresh leaves according to a modified CTAB protocol (Doyle and Doyle 1987): CTAB buffer was supplemented in proteinase K in concentrations of 80 µg/ml, and the deproteinization step was realized twice (once with Dichloromethane:Chloroform:Isoamylalcool 24:24:1 and once with Chloroform:Isoamylalcool 24:1). Genomic DNA was eluted in Tris-EDTA 0.5X with RNAse A in concentration 10 µg/ml, and quantified twice using a Nanodrop 2000. Individuals belonging to the same regional location and microhabitat were pooled together in equimolarity, resulting in four libraries of pooled individuals with total DNA concentrations ranging between 180 and 240 ng/µl. The four libraries were sequenced (paired-end) on three lanes of an Illumina HiSeq 2500. Each library was distributed over the three lanes, and each lane received the different libraries in equal quantity.

Bioinformatics pipeline

Reads pre-processing. Paired-end reads (100 nt) were pre-processed (cleaned and filtered) before assembly. Individual bases of low quality were masked using the 'Fastq_masker' tool of the suite fastx_toolkit (http://hannonlab.cshl.edu/fastx_toolkit) under a quality threshold of 25. Reads-ends were trimmed using Fastq_trimming' (fastx_toolkit) by trimming the first and last ten bases of each reads (resulting in reads of 80 nt). Reads were further sorted and orphan

reads were excluded using Sickle (Joshi and Fas 2011) with the following parameters: minimum per base quality threshold = 25 and minimum read length = 70. The quality of the cleaned reads was assessed using 'fastqc'. Reads pre-processing resulted in 2,332,116,952 high-quality reads (**Table S1**). Cleaned reads have been deposited on the European Nucleotide Archive (ENA, EMBL), under project accession code PRJEB9879, sample accession codes ERS791957, ERS791958, ERS791959, ERS791960.

***De novo* assembly and mapping.** Because *E. falcata* is a non-model species with no reference genome, cleaned reads were assembled into a crude *de novo* genome assembly which constitutes a large new pan-genomic resource. Because classical assemblers do not easily support the numerous polymorphisms contained in libraries of pooled individuals (resulting in the segregation of reads from different populations pools into different assembled contigs and scaffolds), the sequenced libraries were *de novo* assembled using RayMeta (Boisvert *et al.* 2012), an assembler originally developed for meta-communities. The different libraries were assembled together in order to maximize the coverage of the *de novo* assembly, with a kmer size of 21. The assembled contigs were then aligned together in order to re-align redundant contigs using CD-HIT-EST (Li and Godzik 2006; Fu *et al.* 2012) with a sequence identity threshold of 95%. Non-redundant contigs were retained and used as reference to map *a posteriori* each library separately using bwa (Li and Durbin 2009): 'bwa aln' and 'bwa sampe' (only paired reads were mapped). The .sam files (one by library) were converted into .bam, sorted, indexed and merged into a unique .mpileup file using the samtools suite (Li *et al.* 2009): 'samtools view', 'samtools sort', 'samtools index' and 'samtools mpileup'.

Structural and functional annotation. The crude pangenomic reference was annotated with the suite implemented in Maker mpi v. 2.31.8 (Cantarel *et al.* 2008) and associated tools. In summary, repetitive elements were detected by repeatMasker v.4.0.3 (http://www.repeatmasker.org/) and *ab initio* gene prediction was realized by Augustus v.2.6.1 (Stanke *et al.* 2006) at the scaffold level using both the reference transcriptome of *E. falcata* (Brousseau *et al.* 2014) and a composite proteome reference composed of all published reference proteomes of Fabaceae (*Cajanus cajan*, *Cicer arietinum*, *Glycine max*, *Lupinus angustifolius*, *Medicago truncatula*, *Phaseolus angularis*, *Phaseolus vulgaris* and *Vigna radiata*) plus the reference proteomes of *Arabidopsis thaliana*. The predicted proteins were further BLASTed against the non-redundant protein database with NCBI BLASTp v.2.6.0. At last, gene ontology terms of each predicted gene were retrieved using Interproscan 5.21.60 (Jones *et al.* 2014).

SNP detection and Filtering. The absolute abundance of different nucleotides was counted for each assembled position using the PoPoolation2 mpileup2sync.jar script (Kofler *et al.* 2011). The .java.sync output file containing nucleotide counts for each position of the assembly was post-processed through custom perl scripts. (1) Monomorphic sites were discarded and rare SNPs rare were masked (replaced by 'N') under a minimum allele frequency (maf) threshold of 0.05, all libraries confounded (meaning that a minimum number of 8 alternative variants out of 160 sequenced gametes in the whole dataset was required for a polymorphism to be retained). (2) Among the remaining SNPs, only bi-allelic polymorphisms with a total maximum coverage of 200X over all libraries to avoid SNPs caused by potential misassembly of paralogs into the same contigs and scaffolds. (3) Bi-allelic SNPs with library-level coverage < 20X were discarded, resulting in a SNP dataset composed of SNPs with coverage > 20X within each of the four libraries. (4) Contiguous SNPs were further filtered out to avoid polymorphisms caused by copy number variation (CNV). (4 and 5) SNPs located within repeat elements (detected by repeatMasker) as well as SNPs located in contigs with SNP density > 0.02 were filtered out, once again to avoid SNPs located in paralogs.

Computation of basic diversity statistics

Nei's molecular diversity π (Nei 1987) was computed for each of the four sub-populations at the contig level with unphased SNP data, under the hypothesis of random mating, as the sum of unbiased estimates of heterozygosity at the nucleotide (SNP) level (Nei 1987):

$$\pi = \sum_{i}^{S} H_{e,i}$$

with S = number of polymorphic loci in a contig, and

$$H_{e,i} = H'_{e,i} \frac{n}{n-1}$$

where n = sample size and $H'_{e,i}$ = expected heterozygosity at the ith polymorphism in the contig. Because different contigs have different lengths, to compare across contigs we present $\pi_L = \pi/L$, where L is contig length, that is, molecular diversity per base.

Bayesian modelling of genomic divergence and outlier detection

A Bayesian modelling approach was used to identify outlier SNPs under selection. Brousseau *et al.* (2016) has described this Bayesian framework in detail and its power has been assessed through simulations, revealing low false discovery and false non-discovery rates. Briefly, allele frequencies were inferred from read counts and pairwise locus-specific G_{ST} was estimated in the model, thus integrating coverage variability across the genome and libraries.

$$n_{i,m} \sim Bin(p_{i,m}, N_{i,m})$$

$$q_{i,m} = 1 - p_{i,m}$$

where $n_{i,m}$ is the read count of the reference allele and $N_{i,m}$ the total number of read counts at marker m in population i. $p_{i,m}$ and $q_{i,m}$ are allele frequencies of the reference and the alternative alleles respectively.

$$GST_{(i,j),m} = (Ht_{(i,j),m} - Hs_{(i,j),m})/Ht_{(i,j),m}$$

$G_{ST(i,j),m}$ is the Nei's fixation index (i.e. genetic differentiation) between populations i and j at marker m. $H_{T(i,j),m}$ and $H_{S(i,j),m}$ are the expected heterozygosities in the total metapopulation and within subpopulations respectively:

$$Hs_{(i,j),m} = (p_{i,m}q_{i,m}) + (p_{j,m}q_{j,m})$$

$$Ht_{(i,j),m} = 2\bar{p}_m\bar{q}_m = 2(\frac{p_{i,m} + p_{j,m}}{2})(\frac{q_{i,m} + q_{j,m}}{2})$$

Pairwise G_{ST} was further partitioned into genome-wide and locus-specific components according to a hierarchical framework: neutral (genome-wide) divergence both between study sites and between microhabitats within each study sites (**Fig. 2a**), and adaptive (locus-specific) divergence both between sites and between microhabitats common to both sites.

$$GST_{(i,j),m} \sim N(\mu_{(i,j),m}, \sigma_R^2)$$

$$\mu_{(i,j),m} = kr_{(i,j)}(\mu r + \theta r_m) + kl_laussat_{(i,j)}(\mu l_laussat + \theta l_m) + kl_regina_{(i,j)}(\mu l_regina + \theta l_m)$$

$$kr_{(i,j)} = \begin{cases} 0 \\ 1 \end{cases}; kl_laussat_{(i,j)} = \begin{cases} 0 \\ 1 \end{cases}; kl_regina_{(i,j)} = \begin{cases} 0 \\ 1 \end{cases};$$

μr and μl are the global mean of differentiation (over all markers) at regional and local scales respectively: μr is the global mean of differentiation between sites, $\mu l_laussat$ is the global mean of differentiation between microhabitats within the study site of Laussat, and μl_regina is the global mean of differentiation between microhabitats within the study site of Régina.

$kr_{i,j}$ and $kl_{i,j}$ are binary matrices describing whether the populations i and j belong to the same study site (regional scale) and to the same microhabitat within each study site (local

scale): $kr_{i,j}$ = 1 if the two populations *i* and *j* belong to the different study sites, $kr_{i,j}$ = 0 otherwise; $kl_laussat_{i,j}$ = 1 if the two populations *i* and *j* belong to different microhabitats within the study site of Laussat, $kl_laussat_{i,j}$ = 0 otherwise; $kl_regina_{i,j}$ = 1 if the two populations *i* and *j* belong to different microhabitats within the study site of Régina, $kl_regina_{i,j}$ = 0 otherwise.

θr_m and θl_m are locus-specific parameters describing the departure from the global means of differentiation at regional (i.e. between sites) and local (i.e. between microhabitats) scales. Posteriors of θr_m and θl_m are normally distributed allowing thus to attribute an empirical Bayesian p-value (eBP-value) to each SNP. SNPs with an eBP < 1% were considered as outliers under divergent (upper tail) or homogenizing (lower tail) selection between sites (θr_m) or microhabitat within sites (θl_m).

The Bayesian model was encoded in stan (Carpenter *et al.* 2016) and compiled into C++ using Rstan. Because Bayesian inference is computationally expensive, the model was applied to subsets of 10,000 SNPs on a Dell Power Edge CG220 Intel Xeon (CPU E5-2670 v2 @2.50Ghz) cluster node with parallelization on 5 CPU (with parameters: virtual memory vmem = 25G, virtual memory limit h_vmem=30G and parallel_fill = 5) and explicit chains parallelization in R with the package 'parallel'. Both dataset hashing and explicit parallelization optimized computation times and the model ran in ~ 4 hours / subset. Outputs were post-processed with R, including chain convergence checking and eBP values inference from the posterior distribution of locus-specific parameters. Due to the large number of SNPs screened, a correction for multiple-testing was applied on eBP values with the package 'fdrtool' with a FDR threshold of 10%. The proportions of outliers under selection was compared between geographic scales with chi-squared tests ('prop.test' function). Enrichment tests were further conducted through X^2 tests in order to test (i) the enrichment of outliers located within exons, within a predicted gene and in the neighborhood (< 5 kb) of a predicted gene compared to the whole SNP dataset, as well as (ii) the enrichment of GO terms between predicted genes neighboring outlier SNPs compared to all predicted genes of the entire reference.

Results

The crude reference genome produced here, although incomplete and fragmented, constitutes a new, valuable pan-genomic resource containing key information about population-level genetic diversity (i.e. tens of thousands of loci scattered throughout the

genome, a first for a non-model tree species of Amazonia). The four population pools were assembled *de novo* into a high-coverage pan-genomic reference: *de novo* assembly led to 325,249 contigs in 272,865 scaffolds for a total assembly length of about 250 Mb (252,364,908 bases assembled *de novo* of which 249,308,901 were successfully mapped *a posteriori* by the different libraries, **Table S1)** and a mapping coverage of 57.81X on average, **Fig. S3**. The structural annotation predicted 32,789 genes and 117,278 exons, and detected 124,287 repeated elements. The functional annotation of the predicted genes resulted in 32,075 successfully BLASTed genes, and only 732 predicted genes did not return any BLAST hit. Additionally, gene-ontology terms (GO terms) were retrieved for 18,577 predicted genes, while the remaining 14,212 genes were not associated to any GO term.

A total of 97,062 bi-allelic SNPs scattered across 25,803 contigs and 23,276 scaffolds were identified, according to a stringent procedure, with a final SNP density of 0.003/base (*sd*=0.004) with a Ti/Tv=1.80, **Table S2** and **Fig. S4**. Normalized molecular diversity π_L varied between 9×10^{-6} and 8×10^{-3} (**Table S3**), with little variation among populations. 41,050 SNPs (42.3%) were located in scaffolds with at least one predicted gene while 56,012 SNPs (57.7%) were located in scaffolds without any predicted gene. 26,435 SNPs (27.2%) were located within a predicted gene, of which 10,863 (11.2%) within exons, while 14,615 SNPs were in the neighborhood of a predicted gene with a mean distance to a predicted gene of 737 bp (with min = 1bp and max = 27.2 kb).

The Bayesian model revealed low genome-wide differentiation between populations and significant departures from neutral expectations for a small subset of SNPs. Genome-wide genetic differentiation (average G_{ST} across loci) ranged between 0.026 and 0.032 at both geographical scales, **Fig. S5**. Accordingly, genome-wide parameters of divergence estimated through G_{ST} partitioning were not higher between regions than between microhabitats within regions, as confirmed by overlapping credible intervals, **Table S4**. Locus-specific departures from genome-wide divergence (i.e. outlier SNPs) were identified from the distribution of locus-specific parameters (θr_m and θl_m). Between 375 and 2,045 outlier SNPs were detected to be under selection with an eBP threshold of 1%: 375 SNPs under homogenizing selection at regional scale, 2,045 SNPs under divergent selection at regional scale, 510 SNPs under homogenizing selection at local scale and 1,770 SNPs under divergent selection at local scale. After multiple-testing correction, the number of significant SNPs fell to 475 outlier SNPs (0.49% of all SNPs screened): 290 SNPs (in 268 contigs) under divergent selection at the regional scale and 185 SNPs (in 168 contigs) under divergent selection at the local scale, with no outlier shared between the regional and the local scales. Outlier SNPs under divergent

selection were characterized by clines in allele frequencies between regional locations and/or local microhabitats. Clines in allele frequency across sites and microhabitats are illustrated in **Fig. 4** for different kinds of neutral and outlier SNPs. A total of 140 outlier SNPs (29.5% out of 475 outlier SNPs) were located within predicted genes, among which 39 (8.2%) were within exons, and 86 are in the neighborhood (< 5kb) of one or several predicted genes (**Supplementary files S1 and S2**). However, outlier SNPs were not more likely to be located within exons (X^2 = 3.9368, df = 1, p-value = 0.05), within predicted genes (X^2 = 1.0846, df = 1, p-value = 0.30), or in the neighborhood of predicted genes (X^2 = 3.1964, df = 1, p-value = 0.07), compared to any SNP in the entire dataset. Nevertheless, 87 GO-terms were significantly enriched in the subset of genes surrounding outlier SNPs (84 genes) compared to all predicted genes of the entire reference at the regional scale, and 100 GO-terms were significantly enriched in the subset of genes surrounding outlier SNPs (106 genes) at the local scale, **Table S5.**

Discussion

This study analyzes patterns of genomic divergence in wild and undisturbed forest stands of the Amazonian hyperdominant tree *Eperua falcata* in French Guiana. For this species, former studies have already reported high heritability in quantitative growth and leaf traits in a common garden experiment (Brousseau et al., 2013), suggesting high levels of standing genetic variation for functional phenotypic traits. In addition, seedlings native to different microhabitats displayed genetically-driven differences in phenotypic traits probably due to a combination of neutral processes (i.e. short-distance seed flow and local inbreeding) and local adaptation (Audigeos *et al.* 2013; Brousseau *et al.* 2015). However, past studies were not able to conclusively investigate the evolutionary drivers behind microgeographic genetic divergence because they were based on datasets restricted to few (e.g. candidate gene approach, Audigeos *et al.* (2013)) or anonymous markers (AFLPs, Brousseau *et al.* (2015)), and the study of microgeographic adaptation lacked a comprehensive survey at the genomic scale. Here, we provide a new, genomic-scale, evidence of microgeographic adaptation associated to (topographic and edaphic) habitat patchiness in this species.

Patterns of genome-wide divergence.

Genome-wide (supposedly neutral) differentiation was low at both the local and regional geographic scales, ranging from 2 to 3% (**Fig. S5**), similarly to what previously found by Brousseau *et al.* (2015), who reported F_{ST} values between 2 and 3% between microhabitats within sites, and about 3% between these study sites (based on a screening of about 1,100 AFLPs). In two additional study sites in French Guiana (Paracou and Nouragues), another study (Audigeos *et al.* 2013) provided comparable estimates: $F_{ST} \cong 1\%$ between study sites and F_{ST} from 1 to 2% between microhabitats within sites. The SNP density that we estimated here (a mean of 3 SNP every 1,000 bp, **Table S2**) was quite low: Neale and Kremer (2011) for example reported that SNP frequency in trees is generally in the range 1 SNP per 100 bp. However, in our study the stringent SNP filtering probably filtered out many actual polymorphisms, leading to an under-estimation of genome-wide SNP density.

Low genome-wide differentiation suggests that high levels of gene flow in *E. falcata* regularly homogenize genome-wide genetic diversity both between regions and between microhabitats within regions. The consequences of recurring inter-habitat gene flow on population adaptability could be paradoxical. On one hand, extensive gene flow can allow the spread of advantageous alleles between populations, thus granting high adaptive potential for populations established in heterogeneous environments. On the other hand, gene flow may also counteract the effect of natural selection, with migration load halting adaptive divergence of populations into locally adapted subpopulations (Lenormand 2002; Le Corre and Kremer 2012; De Kort *et al.* 2015; Pujol *et al.* 2018). The deleterious effect of gene flow on adaptive divergence could explain the low proportion of outliers detected (see below). Nevertheless, high gene flow (and resulting low genome-wide differentiation) did not completely swamp the signal of adaptive divergence between populations.

Microgeographic adaptation.

Locus-specific (putatively adaptive) departures from genome-wide divergence were detected for 0.49% of the 97,062 SNPs analyzed. We detected significantly more outliers at the regional than at the microgeographic scale: 290 *versus* 185 outlier SNPs ($X^2 = 22.826$, df = 1, $p = 1.77 \times 10^{-6}$). This result can be explained as the logical consequence of an order of magnitude difference in the geographical distances between populations: 300 meters between microhabitats within sites versus 300 km between study sites. Indeed, ecological conditions are probably more variable - and divergent selection stronger - at the regional scale than divergent selective pressures between microhabitats separated by hundreds of meters, and the

effect of divergence would be reinforced by more limited gene flow. The proportion of outliers detected over microgeographic scales, although lower, remains nevertheless non trivial. Because the Bayesian model focuses on replicated footprints of adaptation only, it is extremely unlikely that the detected outliers could have been structured between microhabitats in two different regions purely by chance (Savolainen *et al.* 2013). Moreover, no outlier was detected jointly at both spatial scales, as shown by the heart-shaped joint distribution of regional- and local-scale locus-specific parameters (**Fig. 3**). Instead, outliers were always specific to only one spatial scale, suggesting that different portions of the genome respond to different types of environmental constraints; in future studies, local adaptation in wild populations should be investigated by integrating different spatial scales, including microgeographic habitat patchiness, into the sampling design. It is however important to stress that our Bayesian approach may miss footprints of polygenic adaptation, if adaptation acts in different ways in the two study sites (through different selection strengths and/or different loci affected) which is known to be common in wild trees (Berg and Coop 2014; Lind *et al.* 2017; Harrisson *et al.* 2017). Indeed, F_{ST}-based methods are more prone to detect adaptation caused by loci of large effects than by loci with small effects, especially since adaptation occurs with ongoing gene flow (Hedrick 2006; Yeaman and Whitlock 2011; Le Corre and Kremer 2012; Savolainen *et al.* 2013). In other words, the set of outliers detected is probably enriched with loci of large effects. Even if the Bayesian model employed here proved its robustness to detect footprints of polygenic adaptation caused by weak selection strengths (s=0.05) in simulations (Brousseau *et al.* 2016), it may still miss footprints of polygenic adaptation if the strength of selection is variable among loci (Sork *et al.* 2013).

By providing evidence of adaptation to microgeographic habitat patchiness, our results agree with previous studies showing soil variation as a driver of adaptation in trees. Indeed, previous studies have reported examples of adaptation to soil nutrient composition such as potassium deficiency in *Cornus florida* (Cornaceae) (Pais *et al.* 2017), to aluminum and phosphate ion concentrations in *Pinus contorta* (Pinaceae) (Eckert *et al.* 2012), to substrate age in *Metrosideros polymorpha* (Myrtaceae) (Izuno *et al.* 2017), to soil types in *Abies religiosa* (Pinaceae) (Méndez-González *et al.* 2017), or to soil water availability in *Pinus albicaulis* (Pinaceae) (Lind *et al.* 2017). Preliminary evidence of microgeographic adaptation has been recently reported in Neotropical tree species, including *E. falcata* (Audigeos *et al.* 2013; Brousseau *et al.* 2013, 2015; Torroba-Balmori *et al.* 2017). At the community level, the study by Fortunel *et al.* (2016) suggests a mechanism through which microgeographic adaptation may take place: seedlings mortality of *terra firme* species was high when planted

in seasonally flooded habitats. Topography-linked microgeographic patchiness is likely driven by a complex interaction of abiotic and biotic factors (including soil types (Misiewicz and Fine 2014), water-logging and flooding events, water availability (Clark *et al.* 1999), seasonal soil drought (Markewitz *et al.* 2010), soil fertility (John *et al.* 2007), light availability (Ferry *et al.* 2010), competition and predation (Fine *et al.* 2004)), hence many different classes of loci may respond to selection for microgeographic adaptation. It is, however, difficult to identify the exact genomic targets of natural selection, particularly in non-model species, because of potential hitchhiking between causal mutations and the analyzed SNPs, and because of the lack of chromosome-level genome assembly as well as no reliable genome-wide estimates of linkage disequilibrium in this species. Despite these limitations, the identity of the genes neighboring outlier SNPs may provide useful information, as many 'divergence' outlier SNPs detected here were located within or near (< 5 kb) protein-coding genes involved in a variety of biological functions (**Supplementary files S1 and S2**), with 106 genes neighboring a genomic footprint of microgeographic divergence (*i.e.* one or several outlier SNPs) and 100 Gene Ontology terms (corresponding to 49 genes) significantly enriched in the subset of genes neighboring outlier SNPs, **Table S5**. Some of them are particularly meaningful regarding their biological significance in terms of survival, growth, defense and reproduction (**Supplementary files S1 and S2**).

We detected several genes involved in plant response to stress near outlier SNPs. For example, we detected a glutaredoxin-C11 (predgene_010110) involved in managing the level of reactive oxygen species (ROS) in plant cells (Mittler *et al.* 2004), a H_2O_2-responsive gene encoding a metallopeptidase M24 family protein which is up-regulated under stress in *Arabidopsis thaliana* (Ordoñez *et al.* 2014), and a gene encoding a plant cysteine oxidase 4-like (predgene 003921) involved in response to flooding and soil hypoxia (Weits *et al.* 2014; White *et al.* 2017). Oxidative stress is very common in bottomlands, where water-logging and seasonal flooding events cause soil hypoxia and oxygen deprivation in root cells, leading to the production of ROS (Blokhina *et al.* 2003). This observation is concordant with a previous candidate-gene approach in which a catalase (i.e. a gene involved in ROS detoxification) has been reported under selection in this species (Audigeos *et al.* 2013). We also detected five genes with possible implications in lipid biosynthesis (predgenes 006075, 019500, 002577, 001066, 000689), as for example a palmitoyl-acyl carrier protein thioesterase which is known to play an essential role for plant growth and seed development (Dörmann *et al.* 2000; Bonaventure *et al.* 2003). Lipid metabolism is of high importance for plant development, from early germination to reproduction. Indeed, lipids are the primary material for cutin, and

they are largely involved in stress signaling and resistance (Okazaki and Saito 2014). We also detected two genes involved in proline amino acid biosynthesis (predgenes 007023 and 007024), which is accumulated in response to drought stress in plants (Fu *et al.* 2018). A calmodulin-like protein (probable CML48, predgene 022002) was also detected in the neighborhood of a microgeographic outlier. Calmodulin and calmodulin-like plant calcium sensors have evolved unique functions of high importance in the regulation of plant development and plant response to stress (Perochon *et al.* 2011). We also detected a probable cellulose synthase A (predgene 000745) whose mutations can alter morphology or induce resistance to cellulose biosynthesis inhibitors as reported in *A. thaliana* (Richmond 2000), a gene for plant metal tolerance (predgene 014347), and a mRNA-decapping enzyme-like (predgene 011170) which is of importance for early seedling development (Goeres *et al.* 2007).

We also detected two genes mediating biotic interactions - through plant defense against pathogens and recognition of symbiotic bacteria - as potential targets of microgeographic adaptation. In particular, plant chemical defense against herbivores and pathogens is likely to play a key role in microgeographic adaptation; previous studies have experimentally demonstrated that a trade-off between growth and antiherbivore defenses strengthens habitat specialization between congeneric species within large Neotropical plant families (e.g. Annonaceae, Burseraceae, Euphorbiaceae, Fabaceae, Malvaceae) (Fine *et al.* 2004, 2006, 2013). Here, we detected a chitinase (predgene 000254), a family of enzymes known for their role in resistance to pathogens (Punja and Zhang 1993; Grover 2012). We also discovered a gene involved in recognition of symbiotic bacteria (LysM receptor-like kinase, predgene 002974) which enables the model legume *Lotus japonicus* to recognize its bacterial symbiont (Madsen *et al.* 2003; Radutoiu *et al.* 2003; Ryals *et al.* 2008). Symbiotic nitrogen fixation is a specific feature of legumes, and nutrient availability largely varies across microhabitats (Ferry *et al.* 2010). The presence of genes involved in the perception of symbiotic bacteria near genomic footprints of microgeographic adaptation in this supposedly non-nodulating legume (Villadas *et al.* 2007) is particularly interesting. However, accurate functional characterizations of the proteins surrounding outlier SNPs are required to draw robust conclusions about the physiological processes involved in microgeographic adaptation.

Microevolution in the (Neo)tropics.

Although largely exploratory, our results open a new window into the evolutionary processes underlying the generation of (Neo)tropical biological diversity, and support the hypothesis of niche evolution as a driver of habitat specialization and diversification (Leibold 2008). Microgeographic adaptation driven by a complex interaction of abiotic and biotic factors may be an initial step towards habitat specialization and ecological (sympatric) speciation by trees (Savolainen *et al.* 2006; Feder *et al.* 2012). In this study we provide support for this process at the intra-populational level in an Amazonian tree species. However, the efficacy of microgeographic adaptive divergence at a small but non negligible fraction of loci in driving ecological speciation (as opposed to the effect of genome-wide neutral divergence, caused by strong restrictions in gene flow, in assisting allopatric speciation) warrants further investigation. It is also possible that microgeographic adaptation contributes to the generalist and hyperdominant distribution of *E. falcata* (ter Steege *et al.* 2013). In this view, local adaptation to habitat patches would maintain multiple adaptation optima (i.e. multiple genetic combinations of alleles at different loci), thus pre-adapting the populations to a multitude of environmental situations. Indeed, two evolutionary processes can theoretically allow populations to cope with environmental changes, such as GCC, and avoid extinction (Aitken *et al.* 2008; Harrisson *et al.* 2014): populations can migrate to new areas to track ecological niches, causing shifts in the species' distribution, and/or adapt locally to the new conditions through the action of natural selection (Hoffmann and Sgro 2011). These processes are likely to occur simultaneously and populations' response to GCC can be viewed as a 'race where populations are tracking the moving optima both in space and time' (Kremer *et al.* 2012). From the standpoint of the entire metapopulation of *E. falcata* in the region, local adaptation to multiple optima would be tantamount to population-level balancing selection, which is considered to be major component of adaptive potential (Barrett and Schluter 2008; Delph and Kelly 2014). Such a reservoir of adaptively useful genetic diversity may help the species deal with environmental challenges.

We provided here a new, genomic-scale, evidence of divergent selection at both regional and microgeographic scales in the Amazonian hyperdominant *Eperua falcata*, suggesting that microgeographic environmental heterogeneity caused by variations in topography and edaphic factors creates the conditions for local adaptation in lowland Neotropical rainforests. We notably detected signals of adaptive divergence at many SNPs, whose surrounding genes may ultimately constitute the genomic basis for local adaptation. Accurate functional annotation further gave some interesting clues as to the biological processes targeted by divergent selection, putting the emphasis on the potential roles of

abiotic stresses related to water-logging and drought, and of biotic interactions, in driving local adaptation. We firmly believe that the democratization of both next-generation sequencing technologies and Bayesian methods will contribute to fill the gaps in our understanding of trees evolution in the Neotropics, and we strongly recommend to integrate different spatial scales (including microgeographic habitat patchiness) to future studies addressing the processes of adaptation and diversification in lowland rainforests of Amazonia.

Acknowledgements

This study has been funded by the ANR '*BIOADAPT FLAG'* (ref. ANR-12-ADAPT-0007-01) and benefited from an '*Investissement d'Avenir*' grant (CEBA, ref. ANR-10-LABX-25-01), both managed by the French *Agence Nationale de la Recherche*. Louise Brousseau has been funded by a young investigator grant '*Contrat Jeune Scientifique*' (CJS) of the *French National Institute for Agronomical Research* (INRA). We are grateful to the *Genotoul bioinformatics platform Toulouse Midi-Pyrenees* (Bioinfo Genotoul) for providing help and computing and storage resources. We thank Saint-Omer Cazal and Julien Engel for technical assistance and for botanical identifications. We also thank Dr Anna Stavrinides for helpful discussions about the functional interpretation of the results.

Author contribution

LB, ED and IS designed the experiment. LB sampled leaf material, extracted and conditioned DNA, analyzed the data including the development of the Bayesian model under the supervision of IS and GVV. All authors wrote the article.

References

Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S. **2008**. Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* **1**: 95–111.

Antonovics J. **2006**. Evolution in closely adjacent plant populations X: long-term persistence of prereproductive isolation at a mine boundary. *Heredity* **97**: 33–37.

Antonovics J, Bradshaw AD. **1970**. EVOLUTION IN CLOSELY ADJACENT PLANT POPULATIONS. VIII. Clinal patterns at a mine boundary. *Heredity* **25**: 349–362.

Audigeos D, Brousseau L, Traissac S, Scotti-Saintagne C, Scotti I. **2013**. Molecular

divergence in tropical tree populations occupying environmental mosaics. *Journal of Evolutionary Biology* **26**: 529–544.

Baraloto C, Morneau F, Bonal D, Blanc L, Ferry B. **2007**. Seasonal water stress tolerance and habitat associations within four neotropical tree genera. *Ecology* **88**: 478–489.

Barrett RDH, Schluter D. **2008**. Adaptation from standing genetic variation. *Trends in Ecology & Evolution* **23**: 38–44.

Berg JJ, Coop G. **2014**. A Population Genetic Signal of Polygenic Adaptation (M W. Feldman, Ed.). *PLoS Genetics* **10**: e1004412.

Blokhina O, Virolainen E, Fagerstedt K V. 2003. Antioxidants, oxidative damage and oxygen deprivation stress: A review. *Annals of Botany* **91**: 179–194.

Boisvert S, Raymond F, Godzaridis E, Laviolette F, Corbeil J. **2012**. Ray Meta: scalable *de novo* metagenome assembly and profiling. *Genome Biology* **13**: R122.

Bonal D, Atger C, Barigah TS, Ferhi A, Guehl JM, Ferry B. **2000**. Water acquisition 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, Barigah TS, Granier A, Guel JM. **2000**. Late-stage canopy tree species with extremely low d13C and high stomatal sensitivity to seasonal soil drought in the tropical rainforest of French Guiana. *Plant, Cell and Environment* **23**: 445–449.

Bonal D, Guehl JM, Champenoux F. **2001**. Contrasting patterns of leaf water potential and gas exchange responses to drought in seedlings of tropical rainforest species. *Functional Ecology* **15**: 490–496.

Bonaventure G, Salas JJ, Pollard MR, Ohlrogge JB. **2003**. Disruption of the FATB gene in Arabidopsis demonstrates an essential role of saturated fatty acids in plant growth. *The Plant Cell* **15**: 1020–33.

Bradshaw A. **1960**. Population differentiation in *Agrostis tenuis* Sibth. III. Populations in varied environments. *New Phytol* **59**: 92–103.

Brousseau L, Bonal D, Cigna J, Scotti I. **2013**. Highly local environmental variability promotes intrapopulation divergence of quantitative traits: An example from tropical rain forest trees. *Annals of Botany* **112**: 1169–1179.

Brousseau L, Foll M, Scotti-Saintagne C, Scotti I. **2015**. Neutral and adaptive drivers of microgeographic genetic divergence within continuous populations: The case of the neotropical tree *Eperua falcata* (aubl.). *PLoS ONE* **10**: e0121394.

Brousseau L, Postolache D, Lascoux M, *et al.* 2016. Local adaptation in European firs assessed through extensive sampling across altitudinal gradients in southern Europe. *PLoS*

ONE **11**: e0158216.

Brousseau L, Tinaut A, Duret C, Lang T, Garnier-Gere P, Scotti I. **2014**. High-throughput transcriptome sequencing and preliminary functional analysis in four Neotropical tree species. *BMC Genomics* **15**: 238.

Bulmer MG. **1972**. Multiple Niche Polymorphism. *American Naturalist* **106**: 254–257.

Cantarel BL, Korf I, Robb SMC, *et al.* 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research* **18**: 188–96.

Carpenter B, Gelman A, Hoffman M, *et al.* 2016. Stan: A probabilistic programming language. *Journal of statistical Software* **in press**.

Clark DB, Palmer MW, Clark DA. **1999**. Edaphic fators and the landscape-scale distribution of tropical rain forest trees. *Ecology* **80**: 2662–2675.

Le Corre V, Kremer A. **2012**. The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* **21**: 1548–1566.

Cowan RS. **1975**. *A monograph of the genus Eperua (Leguminosae-Caesalpinioideae)*. Smithsonian Institution Press.

Delph LF, Kelly JK. **2014**. On the importance of balancing selection in plants. *New Phytologist* **201**: 45–56.

Dörmann P, Voelker TA, Ohlrogge JB. **2000**. Accumulation of palmitate in Arabidopsis mediated by the acyl-acyl carrier protein thioesterase FATB1. *Plant Physiology* **123**: 637–44.

Doyle JJ, Doyle JL. **1987**. A rapid DNA isolation procedure from small quantities of fresh leaf tissues. *Phytochem Bull* **19**: 11–15.

Eckert AJ, van Heerwaarden J, Wegrzyn JL, *et al.* 2010. Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* **185**: 969–82.

Eckert AJ, Maloney PE, Vogler DR, Jensen CE, Mix AD, Neale DB. **2015**. Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics & Genomes* **11**: 42.

Eckert AJ, Shahi H, Datwyler SL, Neale DB. **2012**. Spatially variable natural selection and the divergence between parapatric subspecies of lodgepole pine (*Pinus contorta*, Pinaceae). *American Journal of Botany* **99**: 1323–1334.

Feder JL, Egan SP, Nosil P. **2012**. The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**: 342–350.

Ferry B, Morneau F, Bontemps J-DD, Blanc L, Freycon V. **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**: 106–116.

Fetter KC, Gugger PF, Keller SR. **2017**. Landscape Genomics of Angiosperm Trees: From Historic Roots to Discovering New Branches of Adaptive Evolution In: New York, NY: Springer New York, 1–31.

Fine PVA. **2015**. Ecological and Evolutionary Drivers of Geographic Variation in Species Diversity. *Annual Review of Ecology, Evolution, and Systematics* **46**: 369–392.

Fine PVA, Daly DC, Cameron KM. **2005**. The contribution of edaphic heterogeneity to the evolution and diversity of Burceraceae trees in the Western Amazon. *Evolution* **59**: 1464–1478.

Fine PVA, Mesones I, Coley PD. **2004**. Herbivores promote habitat specialization by trees in Amazonian forests. *Science* **305**: 663–665.

Fine PVA, Metz MR, Lokvam J, *et al.* 2013. Insect herbivores, chemical innovation, and the evolution of habitat specialization in Amazonian trees. *Ecology* **94**: 1764–1775.

Fine PVA, Miller ZJ, Mesones I, *et al.* 2006. The growth-defense trade-off and habitat specialization by plants in Amazonian forests. *Ecology* **87**: S150–S162.

Fortunel C, Paine CETT, Fine PVAA, *et al.* 2016. There's no place like home: seedling mortality contributes to the habitat specialisation of tree species across Amazonia. *Ecology Letters* **19**: 1256–1266.

Fu Y, Ma H, Chen S, Gu T, Gong J. **2018**. Control of proline accumulation under drought via a novel pathway comprising the histone methylase CAU1 and the transcription factor ANAC055. *Journal of Experimental Botany* **69**: 579–588.

Fu L, Niu B, Zhu Z, Wu S, Li W. **2012**. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150–3152.

Fustier M-A, Brandenburg J-T, Boitard S, *et al.* 2017. Signatures of local adaptation in lowland and highland teosintes from whole-genome sequencing of pooled samples. *Molecular Ecology* **26**: 2738–2756.

Goeres DC, Van Norman JM, Zhang W, Fauver NA, Spencer M Lou, Sieburth LE. **2007**. Components of the Arabidopsis mRNA decapping complex are required for early seedling development. *The Plant Cell* **19**: 1549–64.

Gould B, McCouch S, Geber M. **2014**. Variation in soil aluminium tolerance genes is associated with local adaptation to soils at the Park Grass Experiment. *Molecular Ecology* **23**: 6058–6072.

Grover A. **2012**. Plant Chitinases: Genetic Diversity and Physiological Roles. *Critical Reviews in Plant Sciences* **31**: 57–73.

Hamrick JL, Godt MJW. **1990**. Allozyme diversity in plant species. *Plant population genetics, breeding, and genetic resources.*: 43–63.

Harrisson KA, Amish SJ, Pavlova A, *et al.* 2017. Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular Ecology* **26**: 6253–6269.

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P. **2014**. Using genomics to characterize evolutionary potential for conservation of wild populations. *Evolutionary Applications* **7**: 1008–1025.

Hedrick PW. **2006**. Genetic Polymorphism in Heterogeneous Environments: The Age of Genomics. *Annual Review of Ecology, Evolution, and Systematics* **37**: 67–93.

Hendry AP, Day T, Taylor EB. **2001**. Population mixing and the adaptive divergence of quantitative traits in discrete populations: a theoretical framework for empirical tests. *Evolution; international journal of organic evolution* **55**: 459–66.

Hoffmann AA, Sgro CM. **2011**. Climate change and evolutionary adaptation. *Nature* **470**: 479–485.

Izuno A, Kitayama K, Onoda Y, *et al.* 2017. The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha* (Myrtaceae). *Molecular Ecology* **26**: 1515–1532.

Jain SK, Bradshaw AD. **1966**. Evolutionary divergence among adjacent plant populations. I. The evidence and its theoretical analysis. *Heredity* **21**: 407–441.

John R, Dalling JW, Harms KE, *et al.* 2007. Soil nutrients influence spatial distributions of tropical tree species. *Proceedings of the National Academy of Sciences of USA* **104**: 864–869.

Jones P, Binns D, Chang H-Y, *et al.* 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics (Oxford, England)* **30**: 1236–40.

Joshi NA, Fas JN. **2011**. Sickle: A sliding-window, adaptive, qualiy-based trimming tool for FastQ files (Version 1.33). Available at https://github.com/najoshi/sickle.

Kofler R, Pandey RV, Schlötterer C. **2011**. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* **27**: 3435–3436.

De Kort H, Honnay O. **2017**. Assessing Evolutionary Potential in Tree Species Through Ecology-Informed Genome Screening In: Pontarotti P, ed. *Evolutionary Biology: Self/Nonself Evolution, Species and Complex Traits Evolution, Methods and Concepts*. Cham: Springer International Publishing, 313–327.

De Kort H, Vandepitte K, Mergeay J, Mijnsbrugge K V., Honnay O. **2015**. The

population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales. *Heredity* **115**: 415–425.

Kraft NJ, Valencia R, Ackerly DD. **2008**. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science* **322**: 580–582.

Kremer A, Ronce O, Robledo-Arnuncio JJ, *et al.* 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* **15**: 378–392.

Latreille AC, Pichot C. **2017**. Local-scale diversity and adaptation along elevational gradients assessed by reciprocal transplant experiments: lack of local adaptation in silver fir populations. *Annals of Forest Science* **74**: 77.

Leibold MA. **2008**. Ecology: Return of the niche. *Nature* **454**: 39–41.

Lenormand T. **2002**. Gene flow and the limits to natural selection. *Trends in Ecology and Evolution* **17**: 183–189.

Li H, Durbin R. **2009**. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**: 1754–1760.

Li W, Godzik A. **2006**. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659.

Li H, Handsaker B, Wysoker A, *et al.* 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079.

Lind BM, Friedline CJ, Wegrzyn JL, *et al.* 2017. Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology* **26**: 3168–3185.

Madsen EB, Madsen LH, Radutoiu S, *et al.* 2003. A receptor kinase gene of the LysM type is involved in legumeperception of rhizobial signals. *Nature* **425**: 637–640.

Markewitz D, Devine S, Davidson EA, Brando P, Nepstad DC. **2010**. Soil moisture depletion under simulated drought in the Amazon: Impacts on deep root uptake. *New Phytologist* **187**: 592–607.

Méndez-González ID, Jardón-Barbolla L, Jaramillo-Correa JP. **2017**. Differential landscape effects on the fine-scale genetic structure of populations of a montane conifer from central Mexico. *Tree Genetics & Genomes* **13**: 30.

Misiewicz TM, Fine PVA. **2014**. Evidence for ecological divergence across a mosaic of soil types in an Amazonian tropical tree: *Protium subserratum* (Burseraceae). *Molecular Ecology* **23**: 2543–2558.

Mittler R, Vanderauwera S, Gollery M, Van Breusegem F. **2004**. Reactive oxygen gene network of plants. *Trends in Plant Science* **9**: 490–498.

Neale DB, Kremer A. **2011**. Forest tree genomics: growing resources and applications. *Nature Review Genetics* **12**: 111–122.

Nei M. **1987**. *Molecular Evolutionary Genetic*. Columbia University Pres, New York.

Okazaki Y, Saito K. **2014**. Roles of lipids as signaling molecules and mitigators during stress response in plants. *The Plant Journal* **79**: 584–596.

Ordoñez NM, Marondedze C, Thomas L, *et al.* 2014. Cyclic mononucleotides modulate potassium and calcium flux responses to H_2O_2 in Arabidopsis roots. *FEBS Letters* **588**: 1008–1015.

Pais AL, Whetten RW, Xiang Q. **2017**. Ecological genomics of local adaptation in *Cornus florida* L. by genotyping by sequencing. *Ecology and Evolution* **7**: 441–465.

Perochon A, Aldon D, Galaud J-P, Ranty B. **2011**. Calmodulin and calmodulin-like proteins in plant calcium signaling. *Biochimie* **93**: 2048–2053.

Petit RJ, Hampe A. **2006**. Some evolutionary consequences of being a tree. *Annu Rev Ecol Evol Syst* **37**: 187–214.

Pujol B, Blanchet S, Charmantier A, *et al.* 2018. The Missing Response to Selection in the Wild. *Trends in Ecology & Evolution*.

Punja ZK, Zhang YY. **1993**. Plant chitinases and their roles in resistance to fungal diseases. *Journal of Nematology* **25**: 526–40.

Radutoiu S, Madsen LH, Madsen EB, *et al.* 2003. Plant recognition of symbiotic bacteria requires two LysM receptor-like kinases. *Nature* **425**: 585–592.

Richardson JL, Urban MC, Bolnick DI, Skelly DK. **2014**. Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution* **29**: 165–176.

Richmond T. **2000**. Higher plant cellulose synthases. *Genome Biology* **1**: reviews3001.1.

Ryals JA, Neuenschwander UH, Willits MG, *et al.* 2008. A LysM Receptor-Like Kinase Plays a Critical Role in Chitin Signaling and Fungal Resistance in Arabidopsis. *The Plant Cell Online* **20**: 471–481.

Savolainen V, Anstett M-CC, Lexer C, *et al.* 2006. Sympatric speciation in palms on an oceanic island. *Nature* **441**: 210–213.

Savolainen O, Lascoux M, Merila J. **2013**. Ecological genomics of local adaptation. *Nature Review Genetics* **14**: 807–820.

Savolainen O, Pyhäjärvi T, Knürr T. **2007**. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* **38**: 595–619.

Schmid M, Guillaume F. **2017**. The role of phenotypic plasticity on population differentiation. *Heredity* **119**: 214–225.

Scotti I, González-Martínez SC, Budde KB, Lalagüe H. **2016**. Fifty years of genetic studies: what to make of the large amounts of variation found within populations? *Annals of Forest Science* **73**: 69–75.

Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB. **2013**. Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes* **9**: 901–911.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. **2006**. *AUGUSTUS*: ab initio prediction of alternative transcripts. *Nucleic Acids Research* **34**: W435-9.

ter Steege H, Pitman NCA, Sabatier D, *et al.* 2013. Hyperdominance in the Amazonian tree flora. *Science* **342**.

Torroba-Balmori P, Budde KB, Heer K, *et al.* 2017. Altitudinal gradients, biogeographic history and microhabitat adaptation affect fine-scale spatial genetic structure in African and Neotropical populations of an ancient tropical tree species. *PLoS ONE* **12**: e0182515.

Turner TL, Bourne EC, Von Wettberg EJ, Hu TT, Nuzhdin S V. **2010**. Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. *Nature Genetics* **42**: 260–263.

Villadas PJ, Fernández-López M, Ramírez-Saad H, Toro N. **2007**. Rhizosphere-bacterial community in *Eperua falcata* (Caesalpiniaceae) a putative nitrogen-fixing tree from French Guiana rainforest. *Microbial Ecology* **53**: 317–327.

Vizcaíno-Palomar N, Revuelta-Eugercios B, Zavala MA, Alía R, González-Martínez SC. **2014**. The Role of Population Origin and Microenvironment in Seedling Emergence and Early Survival in Mediterranean Maritime Pine (*Pinus pinaster* Aiton). *PLoS ONE* **9**: e109132.

Weits DA, Giuntoli B, Kosmacz M, *et al.* 2014. Plant cysteine oxidases control the oxygen-dependent branch of the N-end-rule pathway. *Nature Communications* **5**: 3425.

White MD, Klecker M, Hopkinson RJ, *et al.* 2017. Plant cysteine oxidases are dioxygenases that directly enable arginyl transferase-catalysed arginylation of N-end rule targets. *Nature Communications* **8**: 14690.

Yeaman S, Guillaume F. **2009**. Predicting adaptation under migration load: the role of genetic skew. *Evolution* **63**: 2926–2938.

Yeaman S, Jarvis A. **2006**. Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proceedings Biological Sciences* **273**: 1587–93.

Yeaman S, Otto SP. **2011**. Establishment and maintenance of adaptive genetic divergence

under migration, selection, and drift. *Evolution* **65**: 2123–2129.

Yeaman S, Whitlock MC. **2011**. The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**: 1897–1911.

Author information, data availability

Sequencing data have been deposited on the European Nucleotide Archive (ENA, EMBL), under the project accession code PRJEB9879. It contains the raw reads (accessions ERS791957, ERS791958, ERS791959, ERS791960) and the draft *de novo* genome (accession ERS791961). SNPs data are available on Figshare under the DOI 10.6084/m9.figshare.4212126. Data are under embargo until formal publication of this manuscript. In the meanwhile, correspondence and requests should be addressed to Louise Brousseau (louise.brousseau@ird.fr).

Online supporting information

Supplementary figures S1 to S5.

Supplementary tables S1 to S5.

Supplementary file S1. Interactive gene network to easily navigate through the functional annotation of genes neighboring 'divergence' outlier SNPs (< 5 kb) at the microgeographic scale. Green bubbles indicate predicted genes ; blue and pink bubbles indicate Gene ontology terms, with size depending on the number of predicted genes neighboring an outlier SNP flagged with each GO term ; pink bubbles indicate GO terms significantly enriched between the subset of predicted genes neighboring outlier SNPs compared to the all predicted genes.

Supplementary file S2. BLASTp and functional annotation of genes neighboring (< 5 kb) 'divergence' outlier SNPs at the microgeographic scale.

Figures legend

Fig. 1. (a) Relief map of forest landscapes in the region of Régina, French Guiana. The map was created through the French 'Géoportail' (https://www.geoportail.gouv.fr/): official geospatial data were provided by the French Institute of Geographic and Forest Information ©IGN 2017. (b) *Eperua falcata* sapling. (c) and (d) Forest communities associated with topography and related soil types: *terra-firme* forests on ferralitic soils at the top of hills (c), water-logged forests on hydromorphic soils in bottomlands (d).

Fig. 2. (a) Sampling design scheme and correspondence with genome-wide parameters of divergence estimated through the hierarchical Bayesian model at regional (μr) and local scales ($\mu l_{Laussat}$ and $\mu l_{Régina}$ respectively). (b) Distribution of geographic distances between trees in the study sites of Laussat (Western French Guiana) and Régina (Eastern French Guiana). Vertical dotted bars indicate the range of habitat turnover in each study site.

Fig. 3. (a) Locus-specific parameters (θr_m and θl_m) and normal distribution inferred by the Bayesian model. Vertical bars indicate the location of outliers under homogenizing (lower tail) and divergent (upper tail) selection. Left: regional scale (θr_m), right: local scale (θl_m). (b) Proportion of outlier SNPs under selection at regional (i.e. between study sites) and at local (i.e. between microhabitats within sites) scales. (c) Joint distribution of regional and local-scale locus-specific parameters inferred through the Bayesian model. Blue dots indicate outliers under divergent selection at the regional scale, green dots indicate outliers under divergent selection at the local scale.

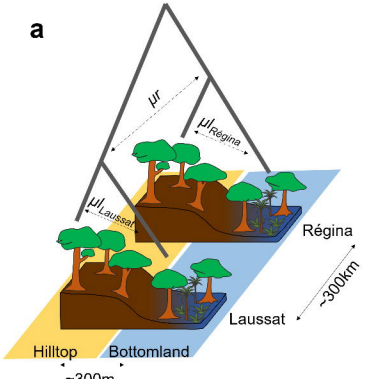
Fig. 4. Clines in allele frequency between regional locations (i.e. study sites) and microhabitats for the different kinds of (randomly sampled) neutral and outlier SNPs at different geographical scales.

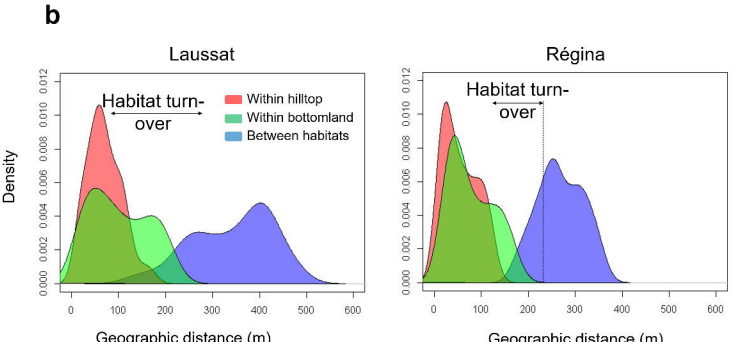
a

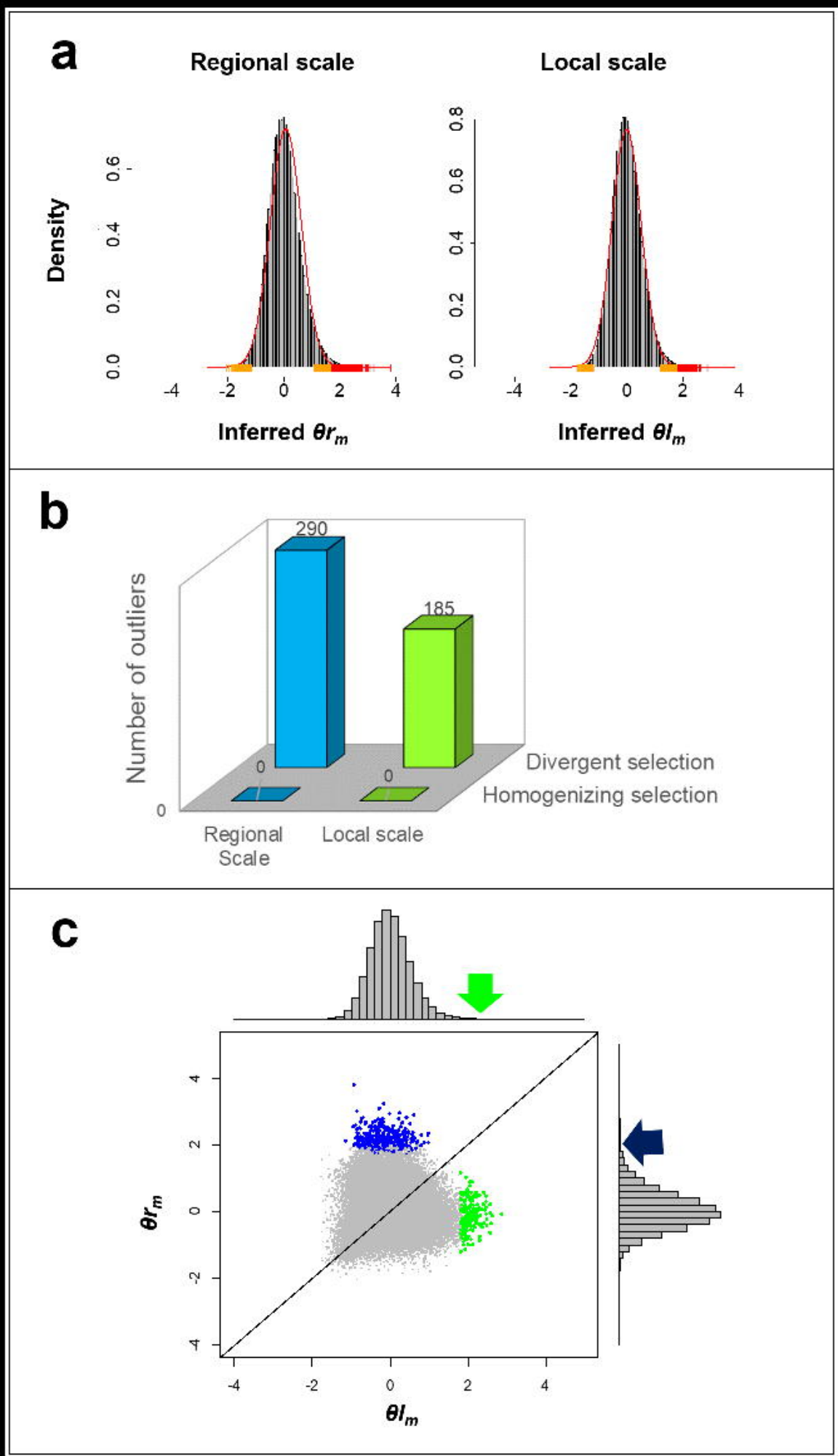
b

c

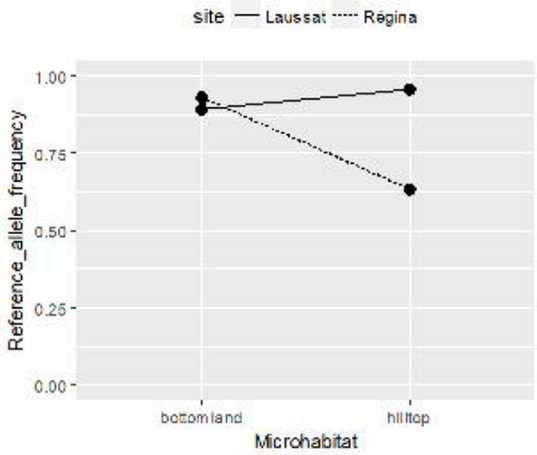
d


a
μr
μl Régina
μl Laussat
Régina
Laussat
~300km
Hilltop
Bottomland
~300m

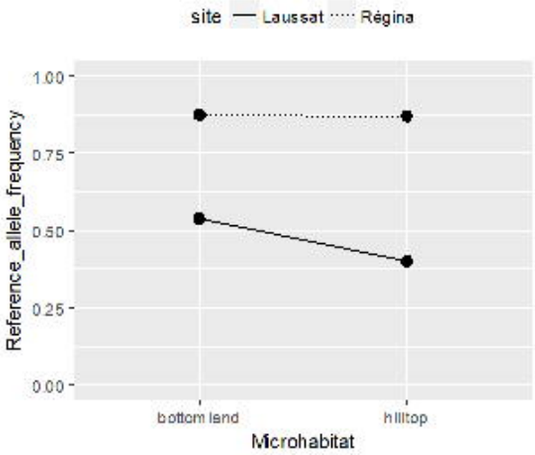

b
Laussat
Régina
Habitat turn-over
Within hilltop
Within bottomland
Between habitats
Density
0.000
0.002
0.004
0.006
0.008
0.010
0.012
0
100
200
300
400
500
600
Geographic distance (m)


a
Regional scale
Local scale
Density
0.0
0.2
0.4
0.6
0.8
-4
-2
0
2
4
Inferred θr_m
Inferred θl_m
b
Number of outliers
290
185
0
0
0
Divergent selection
Homogenizing selection
Regional Scale
Local scale
c
θr_m
θl_m
-4
-2
0
2
4

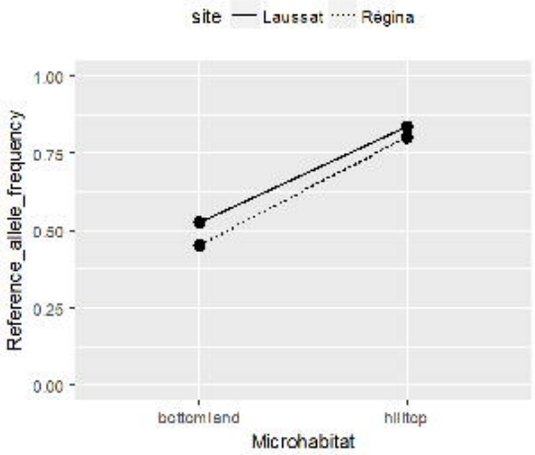
Regional scale	Neutral	Divergent selection	Neutral
Local scale	Neutral	Neutral	Divergent selection



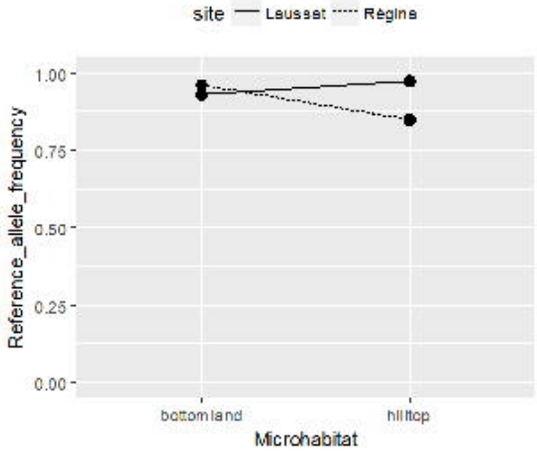
contig-2712000008 pos.550



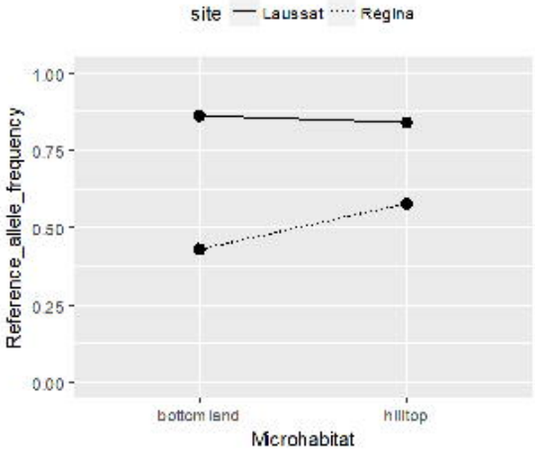
contig-6243000001 pos.1017



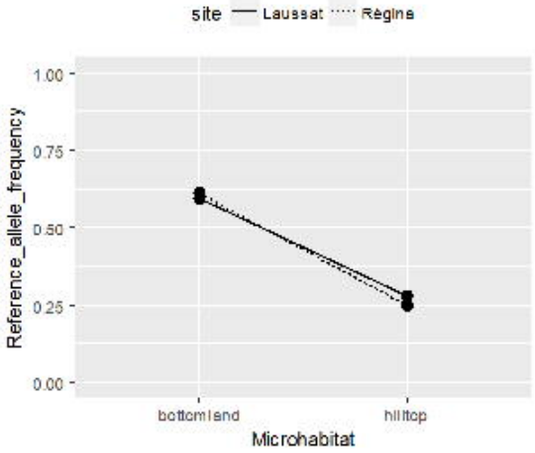
contig-9559000008 pos.801



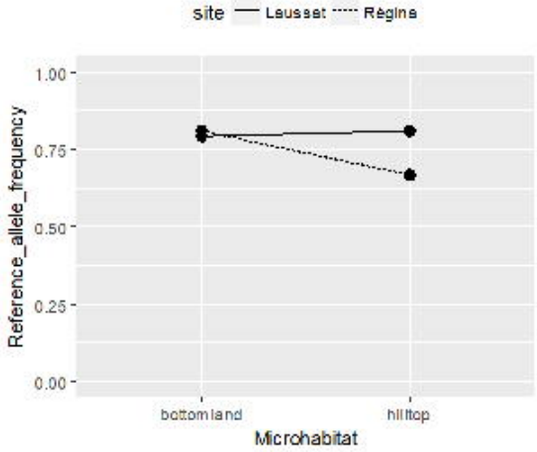
contig-18833000002 pos.62



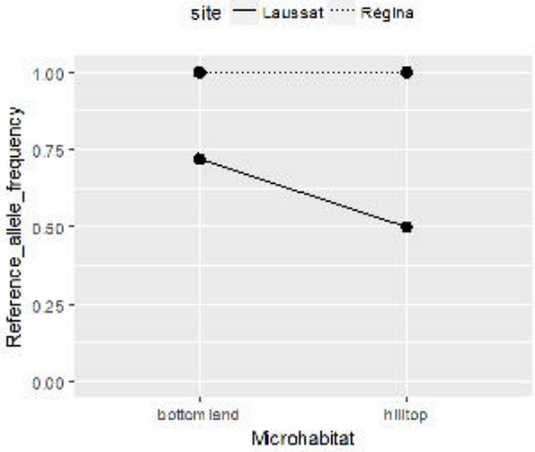
contig-7033000004 pos.701



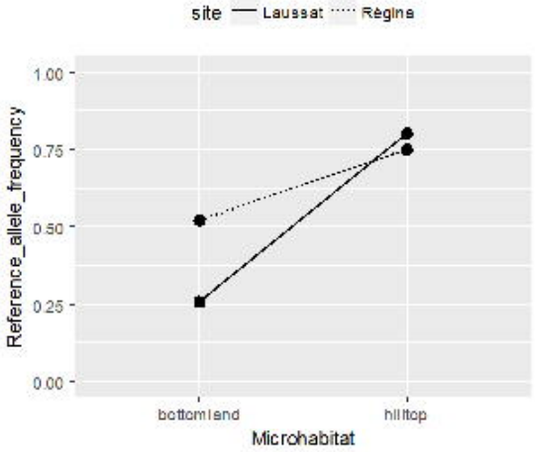
contig-1873000006 pos.763



contig-12808000003 pos.566



contig-19312000001 pos.821



contig-17997000007 pos.579