

Complex adaptive architecture of quantitative resistance erosion in a plant fungal pathogen

Thomas Dumartinet¹, Sébastien Ravel¹, Véronique Roussel¹, Luis Pérez Vicente², Jaime Aguayo³, Catherine Abadie¹, and Jean Carlier⁴

¹CIRAD Departement Systemes biologiques

²INISAV

³Anses Laboratoire de la sante des vegetaux Unite de Mycologie

⁴CIRAD BIOS

September 25, 2021

Abstract

Plant pathogens often adapt to plant genetic resistance so characterization of the architecture under-lying such an adaptation is required to understand the adaptive potential of pathogen populations. Erosion of banana quantitative resistance to a major leaf disease caused by polygenic adaptation of the causal agent, the fungus *Pseudocercospora fijiensis*, was recently identified in the northern Caribbean region. Genome scan and quantitative genetics approaches were combined to investigate the adaptive architecture underlying this adaptation. Thirty-two genomic regions showing host se-lection footprints were identified by pool sequencing of isolates collected from seven plantation pairs of two cultivars with different levels of quantitative resistance. Individual sequencing and phenotyping of isolates from one pair revealed significant and variable levels of correlation be-tween haplotypes in 17 of these regions with a quantitative trait of pathogenicity (the diseased leaf area). The multilocus pattern of haplotypes detected in the 17 regions was found to be highly varia-ble across all the population pairs studied. These results suggest complex adaptive architecture un-derlying plant pathogen adaptation to quantitative resistance with a polygenic basis, redundancy, and a low level of parallel evolution between pathogen populations. Candidate genes involved in quantitative pathogenicity and host adaptation of *P. fijiensis* were highlighted in genomic regions combining annotation analysis with available biological data.

Original article

Complex adaptive architecture of quantitative resistance erosion in a plant fungal pathogen

Running title: Adaptive architecture of resistance erosion

Thomas Dumartinet^{1,2}, Sébastien Ravel^{1,2}, Véronique Roussel^{1,2}, Luis Perez-Vicente³, Jaime Aguayo⁴, Catherine Abadie^{1,2}, Jean Carlier^{1,2}

¹CIRAD, UMR PHIM, F-34398 Montpellier, France, ²PHIM, Univ Montpellier, INRAe, CIRAD, Montpellier SupAgro, Montpellier, France, ³INISAV Instituto de Investigaciones de Sanidad Vegetal, Havana, Cuba, ⁴ANSES, Laboratoire de la Santé des Végétaux (LSV), Unité de Mycologie, Malzéville, France

Correspondance : Jean Carlier, jean.carlier@cirad.fr

Abstract

Plant pathogens often adapt to plant genetic resistance so characterization of the architecture underlying such an adaptation is required to understand the adaptive potential of pathogen populations. Erosion of banana quantitative resistance to a major leaf disease caused by polygenic adaptation of the causal agent,

the fungus *Pseudocercospora fijiensis*, was recently identified in the northern Caribbean region. Genome scan and quantitative genetics approaches were combined to investigate the adaptive architecture underlying this adaptation. Thirty-two genomic regions showing host selection footprints were identified by pool sequencing of isolates collected from seven plantation pairs of two cultivars with different levels of quantitative resistance. Individual sequencing and phenotyping of isolates from one pair revealed significant and variable levels of correlation between haplotypes in 17 of these regions with a quantitative trait of pathogenicity (the diseased leaf area). The multilocus pattern of haplotypes detected in the 17 regions was found to be highly variable across all the population pairs studied. These results suggest complex adaptive architecture underlying plant pathogen adaptation to quantitative resistance with a polygenic basis, redundancy, and a low level of parallel evolution between pathogen populations. Candidate genes involved in quantitative pathogenicity and host adaptation of *P. fijiensis* were highlighted in genomic regions combining annotation analysis with available biological data.

KEYWORDS

fungal plant pathogen, host adaptation, genome scan, quantitative genetics, Musa, *Pseudocercospora fijiensis*

1 | INTRODUCTION

The adaptation of a population to a new environment can involve traits controlled by only a few genes that have a major effect, but such oligogenic adaption is relatively rare (Bell, 2009; van't Hof et al., 2011; Bastide et al., 2016). Indeed, many adaptive traits are genetically complex and involve large numbers of loci, each of which contributes little to the phenotype (Pritchard et al., 2010; Sella & Barton, 2019). With the large amount of genomic data now available, many authors have been able to identify the genetic basis of complex adaptive traits in different organisms (Daborn, 2002; Cook et al., 2012; Linnen et al., 2013) but identifying the genetic basis of a polygenic trait is not sufficient to understand adaptive potential of a species. In addition, the effect size of the genes (i.e. their contribution to the genetic variance of a trait, (Park et al., 2010)), interactions between genes (i.e. additivity, dominance, epistasis and pleiotropy, (Hansen, 2006)) and redundancy (i.e. when several genotypes share the same phenotype by accumulating different combinations of mutations (Barghi et al., 2020)) need to be evaluated.

Identifying the genetic architecture of adaptive traits has been the main focus of two fields of evolutionary biology (Höllinger et al., 2019; Barghi et al., 2020). The first approach is based on molecular population genetics and assumes that adaptive traits result in the directional selection of a limited number of beneficial mutations that have major effects on the traits concerned. A hitchhiking effect on other linked loci leads to loss of diversity in the surrounding genomic regions; this footprint is called a “selective sweep” (Maynard-Smith & Haigh, 1974; Messer & Petrov, 2013). Genome scan methods have been developed to detect this footprint across the genome by measuring differentiation between populations, by detecting variations in the site frequency spectrum (SFS) and/or identifying haplotypes under strong linkage disequilibrium (reviewed by (Vitti et al., 2013, Vatsiou et al., 2016 and Pavlidis & Alachiotis, 2017)). The second approach is based on quantitative genetics and focuses on the phenotype to identify the genes responsible for phenotypic variation (Bazakos et al., 2017). Evolution of a polygenic trait is supposed to be the result of a collective effect of a large number of loci with infinitesimally small variations, leading to more subtle footprints called “shifts” (Barton et al., 2017; Boyle et al., 2017). Analyses of quantitative trait loci (QTL) or genome wide association studies (GWAS) are used to decipher the genetic architecture of a phenotypic trait by identifying correlations between loci and the phenotype (Barton & Keightley, 2002; Visscher et al., 2017). Molecular population genetics and quantitative genetics views are not incompatible. Pritchard and Di Rienzo in 2010 proposed a unifying view of polygenic adaptation as the result of sweeps and shifts acting simultaneously. Thus, combining the two approaches could be a good way to decipher the genetic architecture underlying polygenic adaptation (Gagnaire & Gaggiotti, 2016).

Genetic architecture of traits can be viewed as the genetic potential for phenotype variation through mutation. However, this concept is not sufficient to fully understand adaptation in natural populations, and Barghi et al. 2020 recently proposed the notion of adaptive architecture to better describe the adaptive

potential of species. This notion extends the genetic architecture concept by including other factors involved in population adaptation such as the frequency of contributing alleles, pleiotropy fitness constraints, and genetic forces other than mutation, including selection, drift, and recombination. All these factors play a role in shaping the relative contribution of genes to the adaptation of a population and also in the degree of parallelism when different populations are compared that evolve in the same environment, and could consequently be considered as replicates. Experimental evolution is one possible approach to investigate the genomic responses related to adaptation and to measure the degree of parallelism between populations faced with a controlled environmental constraint and has been successfully applied in *Drosophila* (Graves et al., 2017; Griffin et al., 2017) and *Escherichia coli* (Tenailon et al., 2012). Alternatively, in biological situations (like epidemics) that are difficult to reproduce in the laboratory, adaptive architecture can be investigated in natural systems comprising multiple populations that evolve independently in similar environments (Barghi et al., 2020).

The adaptive architecture concept proposed by Barghi et al. 2020 provides a unified framework to understand how pathogens adapt to plant genetic resistance which is more and more used in agriculture to control diseases as an alternative to applying chemicals. Two categories of resistance have been described in the literature: qualitative and quantitative resistance. Qualitative resistance is often based on ‘effector-triggered immunity’ (ETI), in which major genes confer almost complete protection after recognition of effectors produced by certain pathogen genotypes referred to as avirulent genotypes (Cowger & Brown, 2019). Qualitative resistance is usually not durable because the high specificity of the host-pathogen interactions exerts strong selective pressure on pathogen populations and can lead to rapid selection and fixation of a beneficial mutation (Parlevliet, 2002; Zhong et al., 2017), a process corresponding to the selective sweep concept described above. The genetic basis of quantitative resistance may rely on only a small number of QTLs but can be also polygenic, i.e. involve a large number of QTLs (Cowger & Brown, 2019). Diverse mechanisms can be involved and quantitative resistance is generally considered as the most durable (Pilet-Nayel et al., 2017). However, following changes in quantitative traits of pathogenicity (also referred to as aggressiveness (Lannou, 2012), many examples of erosion of quantitative resistance have recently been reported (reviewed in Pilet-Nayel et al. 2017, Cowger & Brown, 2019). In contrast to quantitative resistance of plants, only a few studies have provided information on the genetic basis of quantitative pathogenicity in pathogens. A complex genetic architecture of fungal quantitative pathogenicity was found in a comprehensive QTL mapping analysis of the wheat pathogen *Zymoseptoria tritici* supported by genome wide association studies (GWAS) of a global sample of isolates (Hartmann et al., 2017; Dutta et al., 2021). However, description of the adaptive architecture on one particular host requires comparison of several fungal populations which can have notable differences on standing genetic variation and population size (McDonald & Linde, 2002).

The ascomycete fungus *Pseudocercospora fijiensis*, which is responsible for black streak disease (BLS) of banana, is an interesting pathogen model to describe adaptive architecture to quantitative plant resistance. BLS is the most damaging foliar pathogens of banana worldwide (Guzmán et al., 2019). The BLS pandemic started around 1960 in South-East Asia/Oceania. In 1972, the disease was detected for the first time in Latin America, in Honduras, and spread rapidly throughout the region (Carlier et al., 2021a). The *Fundación Hondureña de Investigación Agrícola* (FHIA) produced several quantitatively resistant hybrids that were used in Cuba in the 1990s and 2000s and have been used in the Dominican Republic since 2005. However, after five to 10 years of cultivation, in both countries, erosion of resistance was reported in FHIA 18 and FHIA 21 cultivars in the field (Pérez Miranda et al., 2006; Guzmán et al., 2019). Local adaptation of *P. fijiensis* populations explaining the erosion of resistance of FHIA hybrids in the two countries was demonstrated in cross-inoculation experiments (Dumartin et al., 2019). An even more recent study based on pool sequencing (Pool-Seq) supported the existence of convergent adaptation in both resistant and susceptible cultivars in less than 10 genomic regions, suggesting oligogenic architecture underlies this adaptation (Carlier et al., 2021b). However, other genomic regions that did not converge were detected across the populations analyzed and neither redundancy nor phenotype-genotype relationship was tackled in that study.

The aim of the present work was thus to characterize the adaptive architecture underlying the quantitative resistance adaptation of *P. fijiensis*. To this end, we analyzed a large number of *P. fijiensis* samples from

susceptible and resistant cultivars in Cuba using a paired population sampling design. We first used a genome scan based on pool-sequencing data to detect host selection footprints in key genomic regions. Isolates from one location characterized for one trait of pathogenicity (the diseased leaf area) were individually sequenced to perform GWAS and to investigate correlations between the phenotype and the genotype in candidate genomic regions. We then combined all these data to compare adaptive architecture between populations.

2 | MATERIALS AND METHODS

2.1 | Sampling

Two sampling campaigns were conducted in Cuba in 2011 and in 2013 in three different locations (Villa Clara, Ciego de Avila and Matanzas) located between 20 and 300 km apart using the same paired population sampling design (Figure 1, Table 1). Infected banana leaves were collected in a banana plantation pair in each location except in Ciego de Avila, where two pairs were sampled in 2013. The same varieties were collected in all pairs; one plot was planted with a susceptible variety (Macho $\frac{3}{4}$ belonging to the banana AAB genomic group and the plantain subgroup) and the other with a resistant variety (the tetraploid FHIA 18 belonging to the banana AAAB genomic group). More than 1 000 strains were isolated representing 14 populations with a number of isolates per population ranging between 38 and 135. The six populations sampled in 2011 had already been analyzed to investigate local adaptation to banana quantitative resistance and the underlying genetic basis in *P. fijiensis* (Dumartinet et al., 2019; Carlier et al., 2021b). To further describe adaptive architecture, in this study, we added eight more populations sampled in 2013 in the same locations. Banana plantations are frequently replanted and different plantation pairs were collected in the two first locations in 2011 and in 2013. In Matanzas, the same plantations were sampled but they were replanted between 2011 and 2013. The data obtained from samples collected in the two years were then considered as spatial replicates but not as time series.

2.2 | *P. fijiensis* isolates, DNA extraction and sequencing

Mycelium cultures initiated by single ascospore isolated from necrotic lesions were identified as belonging to *P. fijiensis* and stored as described in Zapater et al. 2008. Genomic DNA was extracted from mycelium cultures as detailed in Halkett et al. 2010. Equimolar amounts of DNA from isolates of each population were then pooled (see details in Table 1) to reduce variation during pool sequencing (Pool-Seq), as suggested by Rode et al. 2018. The mean pool size of the six and eight populations collected in 2011 and 2013 was, respectively, 42.25 and 93.12 individuals per pool. The pools of isolates collected in 2011 were sequenced as described in Carlier et al. 2021b, at the Genome Quebec Innovation Centre at McGill University on a GAII platform for paired-end Illumina sequencing (read length of 100 pb, target depth of 80X). The pools of isolates collected in 2013 were sequenced for this study by Genewiz UK Ltd for paired-end HiSeq for paired-end Illumina sequencing (read length of 150 pb, target depth of 200X). Finally, the DNAs of 63 isolates sampled in the location Villa Clara in 2011 were sequenced for this study at the Genome Quebec Innovation Centre at McGill University for individual paired-end Illumina sequencing (read length of 100 pb, target depth of 30X).

2.3 | Mapping, variant calling and filtering

Genomic reads obtained from individuals and pools were mapped against the *P. fijiensis* reference genome (<https://genome.jgi.doe.gov/Mycfi2/Mycfi2.home.html>, Arango Isaza et al., 2016). Pool-sequencing data were treated using the same pipeline and filtering parameters as in Carlier et al. 2021b. Data available from the 2011 samples were rerun with the 2013 samples so the same versions of software were used for both. SNP calling was performed separately for the samples from the two years because some analyses were only possible using samples from 2011, for which some phenotypic data were available (see explanation below). After filtration (mapping quality > 30, minimum read count=3, minimum allelic frequency=0.03), 981 001 and 1 792 219 biallelic SNPs were detected in the six and eight populations collected in 2011 and 2013, respectively. For the sequencing of individuals, the genomic reads of 63 isolates were mapped separately using bwa v0.7.15 software (Li & Durbin, 2010) with bwa_men commands and default parameters. Duplicates were tagged and eliminated using Picard Toolkit v 2.7.0 (Picard Toolkit, 2019, Broad Institute, GitHubRe-

pository:<http://broadinstitute.github.io/picard/>) and `mark_duplicates` command. Genome Analysis Toolkit (GATK) v 4.1.4.0 (McKenna et al., 2010) was used for SNP calling with `Haplotypecaller` command and all individuals were merged in the same file in variant call format (VCF) with the `GenotypeGVCFs_merge`. The VCF file was filtered to keep only SNPs using GATK’s `SelectVariants` command and variants were filtered for quality with the `VariantFiltration` command with the same parameters as in Derbyshire et al. 2019. A second filter was then applied to each genotype from the VCF file using `vcftools` v.0.1.14 (Danecek et al., 2011) with the following parameters: `maf 0.01, minDP 4 maxDP 100, minGQ 20, max-missing 0.7`. After filtration, 758 407 SNPs were identified among the 63 isolates. The VCF file was converted using a custom script into FASTA files containing all individuals, the required format for some analyses below.

2.4 | Evaluation of quantitative pathogenicity

The quantitative pathogenicity of isolates sampled in 2011 (including the 63 isolates individually sequenced) was measured in Dumartinet et al. 2019, on the two cultivars studied (FHIA 18 and Macho 3) by performing *in-vitro* inoculations. Briefly, quantitative pathogenicity was estimated for 16 to 32 isolates per sample by measuring the diseased leaf area (DLA) on the susceptible (DLA-S) and the resistant cultivar (DLA-R) via *in-vitro* inoculation of detached leaf fragments. DLA combines two phenotypic traits, i.e., the within-host growth rate and infection efficiency, which are among the most influential epidemiological parameters in *P. fijiensis* (Landry et al., 2017) and appeared to be a good proxy of parasite fitness. Leaf fragments were collected on banana plants cultivated in the greenhouse for 5-7 months and inoculated with conidial suspensions of *P. fijiensis*. After inoculation, the fragments were incubated in a climate chamber and the DLA was measured at 60 dpi (days post inoculation). A mixed linear model accounting for all the experimental effects was developed to predict the least-squares means (LSMeans) of DLA-S or DLA-R at the population level and for each isolate.

2.5 | Genome scan analyses

Genomic patterns related to host selection were first investigated with Pool-Seq data using the procedure detailed in Carlier et al, 2021b. Briefly, two different genome scan approaches were used: the first, named BayPass (Gautier, 2015), is a genotype-environment association method, while the second, named PoolFreqDiff (Wiberg et al., 2017), is a differentiation-based method. The population pairs collected in 2011 or in 2013 were used in the same analysis using the two approaches to detect convergent signals between the replicates and to limit the detection of false positives. For BayPass, we used the standard covariate model with a qualitative covariate (called Cov-co) corresponding to the cultivars of origin (coded 1 for resistant and -1 for susceptible). Analysis with the samples collected in 2011 were also run with two other quantitative covariates (called Cov-dS and Cov-dR) corresponding to the least-squares means (LSMeans) of DLA-S or DLA-R estimated in Dumartinet et al. 2019. Three independent runs were conducted for all the BayPass analyses and produced very close results. For the poolFreqDiff method, we rescaled all the allele counts to the effective sample size (n_{eff}), as recommended by the authors (Wiberg et al., 2017). From the p-values estimated in the above analyses, we identified putative genomic regions under host selection using the local score approach (Fariello et al., 2017) which accounts for linkage disequilibrium from Pool-Seq data. Several values of the tuning parameter ξ (1, 2 or 3) were used as recommended in simulations (Bonhomme et al., 2019).

2.6 | Genome-Wide Association Studies (GWAS)

GWAS were performed on the two different phenotypic traits (DLA-R and DLA-S) estimated for each isolate in (Dumartinet et al., 2019) using two different statistical models: the multi-locus mixed linear model (MLMM, (Segura et al., 2012)) and the settlement of MLM under progressively exclusive relationship (SUPER) (Wen et al., 2018) both implemented in GAPIT software (Lipka et al., 2012). The first model accounts for the linkage disequilibrium between loci associated with traits while the second model is known to have higher statistical power than regular mixed linear models. All GWAS were corrected for population genetics by estimating a distance matrix between all isolates, and all single-marker p-values were corrected using the Benjamini–Hochberg FDR procedure (Benjamini & Hochberg, 1995). The above-mentioned local

score procedure was also applied to GWAS p-values to increase GWAS resolution to detect clusters of loci with low p-values, as in Bonhomme et al. 2019.

2.7 | Population genetics statistics

Genetic differentiation between all populations was estimated with SNP markers using pairwise F_{ST} (Weir & Cockerham, 1984). With pool-sequencing data, we ran the function `computePairwiseFSTmatrix` implemented in “Poolstat” (Hivert et al., 2018). F_{ST} were computed between each population pair, with all the SNP markers along the genome, or only with the SNPs found in all the candidate genomic regions identified with genome scan analyses, and neighbor joining dendrograms were computed using the R-function `hclust`. In the candidate regions, the differentiation between each population pair sampled in the same location and for each region was estimated and compared to an empirical null F_{ST} -distribution, as described in Carlier et al. 2021b.

Tajima’s D, nucleotide diversity (π), and Watterson’s Theta (ϑ) were estimated with Pool-Seq data on 1Kb non-overlapping windows distributed across the genome for each population sampled in 2011 and 2013 using PoPoolation software (Kofler et al., 2011). These three indexes were also estimated on 1Kb non-overlapping windows using the 63 isolates individually sequenced with the library Egglis v.3.0.0 (De Mita & Siol, 2012). For each population, we computed the median for the three statistics based on the distribution of each 1 kb window distributed across the genome as described in Carlier et al. 2021b. We then compared Tajima’s D distributions computed with SNPs found in the candidate genomic regions to the whole genome distribution using a Wilcoxon-Mann-Whitney test.

Linkage disequilibrium (LD) was estimated from the 63 isolates of *P. fijiensis* sequenced individually using the option `hap-r2` implemented in `vcftools` v.0.1.14 (Danecek et al., 2011). In the present study, LD was estimated for all pairs of loci, for each scaffold forming the core genome and for the two populations separately (CU1S2 and CU1R2). The pattern of LD decay was estimated by calculating the mean Pearson’s correlation coefficient (r^2) for intervals of 1 Kb grouping pairs of loci separated by the same distance.

2.8 | Haplotype analysis in candidate genomic regions

The VCF files corresponding to each previously identified candidate region were extracted from the individual sequencing data using `vcftools` v.0.1.14 (Danecek et al., 2011) and converted into a FASTA file containing all individuals using a custom script. The software RAxML v.8.2.4 (Stamatakis, 2014) was used with a GTRGAMMA model (-m), rapid bootstrap method (-f), 630 seeds for the parsimony inference procedure (-p) to build maximum likelihood trees for each region and including the 63 individuals from location 1. Trees were visualized with the R package `ggtree` (Yu et al., 2017). The LD between haplotypes in pairs of candidate regions was tested with the R GenePop package (Rousset, 2008) and p-values were corrected for the number of tests by computing q-values. The R Poppr package (Kamvar et al., 2014) was used to estimate the index of association ($\bar{\pi}_D$) between combinations of haplotypes in multiple candidate regions. Redundancy analysis (RDA, (Rao, 1964)) was performed with the R `vegan` package (Oksanen et al., 2012). We used a RDA to identify correlations between the haplotypes defined in the candidate regions (explanatory variables) and the quantitative traits (response variables), i.e., the DLA-R and the DLA-S traits. The correlations between candidate regions and traits were statistically tested with a permutation test with 1 000 permutations. The multilocus adaptive pattern between populations was then further investigated using pool-sequencing data. The software `harp` (Kessner et al., 2013) was used to estimate the frequency of the haplotypes defined from individual sequencing in the 14 study pools.

2.9 | Gene content in the candidate genomic regions

The annotated reference genome of *P. fijiensis* was used to identify all the genes in all the genomic regions detected using the genome scan approaches and that were putatively involved in the host adaptation. The annotation (Go terms, KOG terms, and presence of a peptide signal) for these genes was retrieved from the GGF3 file of the JGI website (<https://genome.jgi.doe.gov/Mycfi2/Mycfi2.home.html>). We also tested whether these genes corresponded to *in silico* defined SSPs (Arango Isaza et al., 2016) or to proteins secreted

in vitro and *in planta* by comparing isolates with different pathogenicity levels (Escobar-Tovar et al., 2015) or genes expressed during infection in a transcriptome analysis (Noar & Daub, 2016). Protein sequences were subjected to a BLAST (Altschul et al., 1990) search in the pathogen-host interactions database (PHI-base, version 4.9), which currently contains around 6 000 genes proven to affect the outcome of host-pathogen interactions (Urban et al., 2015). For each BLAST search, we kept the gene in the PHI-base that had the lowest bitscore (and E value $< 1 \times 10^{-9}$) and an effect on quantitative pathogenicity (i.e. phenotype “loss of pathogenicity”, “reduced virulence”, “increased virulence”).

3 | RESULTS

3.1 | Population genetic structure

The F_{ST} values were relatively low (ranging from 0 to 0.055) indicating a weak genetic structure between the different populations (Table S1). A dendrogram constructed from the F_{ST} (Figure S1) shows that the populations initially clustered in pairs. However, sampling location is not a hierarchical level of the population structure since only the populations sampled in location 1 in 2011 and 2013 clustered together. The populations in location 3 were sampled in the same plantations in 2011 and 2013 but the plantations had been replanted between the two years and the differentiation observed could thus result from genetic drift after recolonization by *P. fijiensis*. Absence of clear hierarchical structure reinforced our decision to consider data from samples collected in 2011 and 2013 as spatial replicates but not as time series.

The nucleotide diversity (π), Watterson’s theta (Θ) and Tajima’s D estimated for each population were similar in populations sampled in the same year (in 2011 or in 2013) with Pool-Seq data. In populations sampled in 2011, the median values of π and Θ ranged from 0.0001 to 0.0019 and from 0.0003 to 0.0017, respectively. In populations sampled in 2013, the median values of π and Θ ranged from 0.0049 to 0.0072 and from 0.0040 to 0.0061, respectively. The greater genetic diversity detected in populations sampled in 2013 is certainly due to the fact that more individuals were sampled in 2013. However, Tajima’s D estimates were similar in the two sampling years, ranging from -0.429 to 0.055 and -0.316 to -0.005 in populations sampled in 2011 and 2013, respectively. For individual sequencing data, the median values of π and Θ were 0.0005 for both populations and both indices. Tajima’s D median was 0.0006 for population CU1S2 and 0.0005 for population CU1R2. The estimated values for all the indexes were lower with the individual sequences, probably due to the smaller number of SNPs detected. The data obtained by sequencing individuals were also used to estimate linkage disequilibrium (LD) decay over the distance separating two SNPs in each population. The distance corresponding to a LD decay of 50% was around 4 Kb when all the scaffolds constituting the core genome of *P. fijiensis* were taken into consideration (Figure S2). Overall, no difference in the LD decay pattern was observed between the two populations (CU1S2 and CU1R2). The association became random in sites located 500 kb apart whatever the population.

3.2 | Host selection footprints

After combining the results of all the genome scan analyses, a total of 32 genomic regions with sizes ranging from 0.59 to 57 kb were detected (Figure 2, Table 2). Among the 32 regions, 25 were only identified with BayPass, two only with PoolFreqDiff, and three with both methods (Table 2). Fifteen genomic regions were detected using the six populations sampled in 2011. Twelve of the regions had already been identified in our previous study (Carrier et al., 2021b), including the five major and convergent regions: S1R2-Cu, S2R1-Cu, S2R5-Cu, S12R1-Cu and S12R2-Cu. The detection of three new regions in the present study was certainly because a larger number of SNPs were found using the latest more efficient version of the software used for the mapping step. We ran the same analysis on the eight additional populations sampled in 2013 and a total of 19 genomic regions were detected including two (S1R2-Cu, S12R1-Cu) already detected in the 2011 pool and 17 newly detected.

The genetic differentiation estimated between populations in the 32 candidate genomic regions (from 4 974 SNPs) was higher than that of the whole genome (Figure S1, Table S1) with F_{ST} values ranging from 0.003 to 0.226. Although none of the regions was found in all seven population pairs analyzed, 29/32 were significantly differentiated in at least two pairs, including three regions (S1R2-Cu, S6R5-Cu, S12R1-Cu) in five pairs. A

complex population structure was observed with no apparent effect of the location, the sampling year, or the variety of origin (Figure S1). Tajima’s D computed from 1-kb non-overlapping window was significantly greater in the candidate genomic regions than in the whole genome in all the study populations and in either pool or individual sequencing in location 1 (Figure 3).

3.3 | Adaptive phenotypic variation in quantitative pathogenicity traits

Genome wide association studies (GWAS) were performed using the 63 samples collected in location 1 in 2011 that were both sequenced and phenotyped. No significantly associated SNP nor genomic regions were detected using two models either with a false discovery rate (FDR) of 5% or the local score procedure. Adaptive phenotypic variation was further investigated considering the 32 genomic regions showing host selection footprints. Phylogenetic trees were computed for each region from the genomic sequences of the 63 sequenced isolates (example of the region S1R2-Cu in Figure S3). The two regions with the lowest sizes (S3R0-Cu and S9R0-Cu) were first discarded because no SNP was detected from this data. In most of the other regions, the isolates clustered in two main groups of close DNA sequences, hereafter referred to as haplotype 1 and haplotype 2 (Table S2). However, two genomic regions (S9R1-Cu and S12R3-Cu) were discarded because clusters of haplotypes were difficult to identify due to missing data and possible recombinant haplotypes. In the remaining 28/30 regions, each isolate was assigned to either haplotype 1 or haplotype 2. Between one and six individuals with intermediate haplotypes that may have resulted from recombination between the two main haplotypes in four regions were discarded (Table S2). The distribution of the DLA-R and DLA-S traits was then compared between the two haplotypes using a non-parametric Wilcoxon test. If a significant difference was observed, the haplotype performing the best (which could be either haplotype 1 or haplotype 2) was considered as the advantageous haplotype in the variety concerned. The results suggest association of at least one haplotype with at least one trait in 17/28 regions. A significant difference between haplotypes was observed for both traits (DLA-R and DLA-S) in 3/28 regions, where haplotype 1 appeared advantageous in both cultivars in two regions (S2R1-Cu and S4R4-Cu) while in the latter (S9R3-Cu), haplotype 1 was advantageous in the resistant cultivar and haplotype 2 in the susceptible cultivar (Table 3 and S3). In 9/28 regions only haplotype 1 appeared to be more advantageous in the resistant cultivars. The opposite was observed in 5/28 in which only haplotype 2 appeared to be more advantageous in the susceptible cultivar.

To further investigate correlations with phenotype in the 28 candidate regions in which haplotypes could be defined, a redundancy analysis (RDA) was conducted. Differentiation between the two populations (CU1S2 and CU1R2) was first observed using this analysis according to the x axis (RDA1) but not according to the y axis (RDA2, Figure 4A). The angle between the two vectors corresponding to DLA-S and DLA-R approached 90° suggesting no correlation between the two traits. The sign of the RDA score indicates the direction of the correlation for a given trait, a positive or a negative score indicating that haplotype 1 or haplotype 2 is correlated with the trait, respectively. An unequal contribution to DLA-S and DLA-R of the 28 candidate genomic regions was observed when we plotted the RDA scores corresponding to all candidate regions (Figure 4B). RDA scores ranged from 0.419 to -0.232 for DLA-R and from 0.234 to -0.301 for DLA-S, suggesting heterogeneous contributions of the candidate regions for the two traits (Table S3). A significant correlation (permutation test with 10% threshold) was found using the RDA approach considering both traits in 13 genomic regions for which a significant difference in pathogenicity was previously observed between haplotype 1 and haplotype 2 using the Wilcoxon test (Table 3). The S2R1-Cu region was significantly correlated with both traits, with RDA scores of 0.347 and 0.234 for DLA-R and DLA-S, respectively, supporting the hypothesis that haplotype 1 is advantageous for both traits. Eight out of 13 regions were only correlated with DLA-R with positive RDA scores ranging from 0.419 to 0.285, suggesting that haplotype 1 is the advantageous haplotype for the resistant cultivars. By contrast, the RDA scores of the 4/13 regions showing a significant correlation with DLA-S ranged from -0.301 to -0.264 and this negative correlation supports the hypothesis that haplotype 2 is advantageous for the susceptible cultivar.

3.4 | Relationship between multilocus genotype and traits related to quantitative pathogenicity

Analysis of multilocus genotypes was conducted on the 63 isolates individually sequenced across the 28 regions in which haplotypes could be defined. No significant multilocus disequilibrium was detected in the two

populations considering the two haplotypes previously defined across the 28 regions ($0.008 > \bar{r}_D > 0.009$), the 12 regions only associated with DLA-R ($0 > \bar{r}_D > 0.007$), or the eight regions only with DLA-S ($0 > \bar{r}_D > 0.009$). Furthermore, based on the same haplotypes, no significant pairwise linkage disequilibrium (LD) was detected between the 28 regions in the two populations (CU1S2 and CU1R2). However, the number of advantageous haplotypes in the resistant cultivar accumulated per isolate ranged from 2 to 10 (out of 12 possible advantageous haplotypes) and the correlation of this number with the level of aggressiveness in the resistant cultivar (measured with DLA-R) was positive (Spearman’s correlation coefficient $\rho=0.39$, $p=2.21e^{-16}$). On the other hand, the number of advantageous haplotypes in the susceptible cultivar accumulated per isolate ranged from 0 to 7 (out of 8 possible advantageous haplotypes) and the correlation of this number with the level of aggressiveness in the susceptible cultivar (measured with DLA-S) was also positive (Spearman’s correlation coefficient $\rho=0.38$ and $p=4.093 e^{-9}$). These results suggested that the most aggressive isolates in both cultivars tend to accumulate more advantageous haplotypes in the corresponding cultivars across the candidate regions but in different combinations.

The frequencies of haplotype 1 or haplotype 2 defined from individual sequencing in the 17 genomic regions associated with DLA-R and/or DLA-S (Table 3) were estimated in the 14 study pools with the software harp (71, Table S4). This program uses an expectation-maximization (EM) algorithm to infer the maximum-likelihood estimated frequencies of a known haplotype in a pool of individuals. To validate this approach, we first compared the frequencies of the two haplotypes between estimates resulting from individual and pool sequencing in samples from location 1 for which both kinds of data were available (Chi2 test). Haplotype frequencies differed significantly in pool and individual sequencing in 4/17 regions (S1R3-Cu, S2R3-Cu, S4R2-Cu, S4R4-Cu). Visual examination of the alignments from the sequencing of individuals revealed a large number of recombinant haplotypes, missing data, and/or point mutations in these regions that could have biased our estimation of haplotype frequencies in pools. These four regions were consequently not used for further analysis. A significant difference in haplotype frequencies (Fisher’s exact test) was detected between population pairs in some of the 13 remaining regions and in some of the locations, indicating high heterogeneity between populations but a tendency was nevertheless observed (Table 4). For the seven regions correlated with DLA-R, haplotype 1, which was advantageous in the resistant cultivars, was always significantly more frequent in the populations of this cultivar sampled, with one exception (region S8R1-Cu). In the four regions correlated with DLA-S, the advantageous haplotype 2 in the susceptible cultivar was always more frequently found in the populations of this cultivar sampled. In the S2R1-Cu region correlated with both DLA-R and DLA-S, haplotype 1 was more aggressive considering both traits but was found to be more frequent only in the population of the resistant variety sampled in location 3 in 2013. For the region S9R3-Cu haplotypes 1 and 2 were considered to be advantageous in the resistant and susceptible cultivar, respectively. In this region in location 1 in 2011 and 2013, haplotypes 1 and 2 were significantly more frequent in populations of resistant cultivars and susceptible cultivars, respectively, suggesting diversifying selection. Overall, our results concerning multilocus haplotypes suggested a low degree of convergence between the different locations, and the regions involved in the two traits are not necessarily the same from one location to another. However, some populations accumulated several haplotypes across regions that are advantageous in their cultivar of origin.

3.5 | Gene content in the candidate genomic regions

A total of 118 genes distributed in 25/32 candidate regions were identified using the annotated reference genome of *P. fijiensis* (Arango Isaza et al., 2016). However, due to linkage disequilibrium and the hitchhiking effects, it is unlikely that all these genes are involved in pathogenicity and adaptation. The number of genes per region ranged from 1 to 12. All the information on these genes available to date is provided in supplementary (Table S5). None of the candidate genes corresponded to the putative small secreted proteins (SSPs) identified *in silico* in Arango-Isaza et al. 2016, which are putative effectors involved in plant-pathogen interactions. However, secreted molecules other than SSPs can also be effectors (Rovenich et al., 2014). We identified 10 genes distributed across nine regions with a peptide signal that could correspond to effectors (Table 5). Only one gene among these genes was identified in the *in-vitro* secretome of *P. fijiensis* published by Escobar-Tovar et al. 2015. Four other genes found in different regions were associated with

increased expression in leaf tissues infected by *P. fijiensis* in the transcriptome study published by Noar & Daub 2016. Finally, we found 17 genes distributed in 10 genomic regions corresponding to homologous genes indexed in the PHI-base (a database containing more than 6 000 genes involved in host-pathogen interactions, (Urban et al., 2015)) that are known to lead to a total loss, reduction, or gain in pathogenicity in other fungal species. A total of 12 genes located in the 17 regions significantly correlated with DLA-R and/or DLA-S were also identified in at least one of the other previous analyses. Based on all the data available at present, we considered several genes as candidates involved in quantitative pathogenicity and host adaptation (highlighted in Table 5 and details in Table S5): three major candidate genes (highlighted in red in Table 5 and Table S5) were located in regions significantly correlated with DLA-R or DLA-S and previously identified in at least two of the databases consulted, other good candidates (highlighted in grey) were located in regions correlated to DLA-S or DLA-R and identified in one of the database (9 genes) or found in at least two of the databases consulted (4 genes).

4 | DISCUSSION

The aim of this study was to investigate the adaptive architecture involved in the response of the fungal pathogen *Pseudocercospora fijiensis* to selection exerted by banana quantitative resistance. We combined genome scan and quantitative genetic approaches to compare paired population samples of *P. fijiensis* collected on banana cultivars with different levels of quantitative resistance in Cuba. The results provide first insights into the adaptive architecture behind the response to quantitative resistance in a fungal plant pathogen which was revealed to be complex.

Thirty-two putatively selected genomic regions were detected using the genome scan approach, suggesting a polygenic basis of host adaptation in *P. fijiensis* (Table 2, Figure 2). Genome sequencing of pools of isolates (Pool-Seq), a cost-effective method that enables estimation of allele frequency using large samples (Schlötterer et al., 2014), was used in this first approach. The main problem using genome scan methods is the detection of false positives (Vatsiou et al., 2016). Considering signals that are convergent between several replicates is one way to limit the number of false positives, which is what we did in this study (Lotterhos & Whitlock, 2015; Hoban et al., 2016). As also suggested in Dalongeville et al. 2018 to limit the number of false positives, we combined different genome scan methods and used the local score method that considers linkage disequilibrium between SNPs showing a significant signal to delimit genomic regions (Bonhomme et al., 2019; Fariello et al., 2017). It is worth noting that the sizes of the genomic regions delimited in the present study using the local score method were in line with the results of linkage disequilibrium analysis. The average size of the candidate regions was around 14 kb, a distance for which there still was some linkage disequilibrium between two adjacent SNP markers (Figure S2). We used the same paired design as in Carlier et al. 2021b, but we more than doubled the number of population pairs analyzed (Table 1, Figure 1). The 8/14 new populations were sampled two years later in the same locations and in the same cultivars. Based on this larger number of samples, a convergent selection footprint was detected in at least two locations in 90% (27/30) of the candidate genomic regions. As suggested in Carlier et al. 2021b, increasing the number of locations analyzed allowed us to detect more candidate regions. The genetic basis of host adaptation in *P. fijiensis* consequently appears to be more polygenic than we previously thought. The genome scan based on Pool-Seq was an efficient and low-cost way to obtain insights into the genetic basis of adaptation of *P. fijiensis* to quantitative resistance. However, to further characterize the genetic architecture of this adaptation, other approaches were required to identify adaptive variation and estimate the contribution of the candidate genomic regions in adaptive traits.

The genetic architecture of adaptation of *P. fijiensis* to quantitative resistance was further investigated by searching for correlations between a trait (the diseased leaf area, DLA) involved in quantitative pathogenicity and haplotypes identified from individual sequencing of isolates originating from one location (Table 3). This trait, previously used to detect local adaptation in the populations sampled in 2011 (Dumartinet et al., 2019), appeared to be a good proxy of parasite fitness (see Material and Method). A correlation was found between 61% (17/28) of the candidate regions and DLA measured in both cultivars. These results first confirmed that more than half the candidate regions detected by genome scan were associated with a phenotypic variation

related to quantitative pathogenicity in at least one cultivar. In 14 regions out of 28, we were able to detect a haplotype that would confer a fitness advantage only on the resistant cultivar (9 regions) or the susceptible one (5 regions). These results support the hypothesis that adaptation to quantitative resistance can involve specific host-pathogen (or genotype x genotype) interactions that may result in a local adaptation pattern already described in the same cultivars (Dumartinet et al., 2019). However, the advantageous haplotypes identified in one cultivar did not result in a disadvantage in the other cultivar, thus supporting the absence of a fitness-cost, as previously observed in Dumartinet et al. 2019. Host specificity in genes involved in quantitative pathogenicity was also suggested in the fungus *Z. tritici* (which has similar biology to *P. fijiensis*; Ohm et al., 2012) by comparing two wheat cultivars (Hartmann et al., 2017). Three regions (S2R1-Cu, S4R4-Cu and S9R3-Cu) were found to be significantly correlated with the trait measured in the two cultivars. In the two first regions, the same haplotype conferred an advantage on both cultivars, meaning that the same genes in these regions could be selected in both cultivars or alternatively, different linked genes, since the regions contained several genes. In the third region, two different haplotypes were correlated with increased pathogenicity in the two cultivars, again suggesting the existence of specific host-pathogen interactions. No correlation was detected between the DLA and 39% (11/28) of the candidate genomic regions. This could be due to insufficient statistical power or, alternatively, the genes in these regions may play a role in other traits related to the pathogen's life cycle not measured in this study, such as the latent period, spore production rate, or the latent infectious period (Guzmán et al., 2019). The level of disease in the field depends on the value taken by all these quantitative traits (Lannou, 2012). Currently, no method is available to measure all these traits in *P. fijiensis* in laboratory conditions but methods that test associations at the population level like BayPass may make it possible to study some of them directly in the field.

In the case of polygenic adaptation, variants can have different effect sizes on a given phenotypic trait and measuring these effects provides insight into the genes that contribute the most to adaptation (Park et al., 2010; Shabana et al., 2018). The effect sizes can be estimated using GWAS (Korte & Farlow, 2013). However, in the present study, no association between loci and phenotypic traits was detected using this analysis. GWAS is not always appropriate to study the genetic basis of highly variable traits and/or traits involving a large number of loci with minor effects, because associations can only be detected with using a large number of individuals (Barton & Keightley, 2002; Korte & Farlow, 2013; Visscher et al., 2017). Concerning quantitative pathogenicity in plant pathogenic fungi, SNPs associated with a trait related to reproduction in two cultivars were detected using the GWAS approach in *Z. tritici* (Hartman et al. 2017). However, in a recent GWAS analysis with the same pathogen, SNPs associated with a trait related to reproduction and to leaf disease area (like the trait used in the present study) were detected in only 3/12 and 2/12 inoculated wheat cultivars, respectively, although the authors used a not too conservative statistical threshold (false discovery rate, FDR=10%, (Dutta et al., 2021)). The size effect of the candidate regions detected in the present study was rather tackled by estimating their contribution to the two traits using redundancy analysis. In the present study, the redundancy analysis (RDA, Figure 4) indicated unequal contribution of the regions associated with both study cultivars, thus suggesting that these regions may contain variation in genes with different effect sizes. Moreover, different multilocus genotypes across the candidate regions leading to an increase in quantitative pathogenicity were found among the study populations, suggesting genetic redundancy among the loci involved in adaptation of a fungus like *P. fijiensis* to its host (Table 4). Genetic redundancy has been proven to create heterogeneous signatures of adaptation and therefore to influence the adaptive architecture of polygenic traits, as already observed in plants or animals facing environmental variations (Yeaman et al., 2016; Barghi et al., 2019).

Although convergent adaptation was revealed between some *P. fijiensis* populations in most candidate regions (Table 2), overall, a low level of convergence was found across all the populations and the 32 candidate regions analyzed. In some evolve and resequence experiments, parallel evolution was found to be relatively rare (Graves et al., 2017; Griffin et al., 2017). Barghi et al. 2020 suggested that non-parallel evolution should be considered as the most likely scenario because the majority of adaptive traits are complex and polygenic. Non-parallelism is expected for 1) polygenic traits controlled by multiple genes with small effects, 2) when there is some redundancy between the genes involved in the adaptive trait, and 3) when populations are

differentially affected by evolutionary forces (Barghi et al., 2020). The results obtained in this study suggest a polygenic basis for host adaptation with genetic redundancy. In addition, the 14 natural populations studied could have been affected by demographic events. As already discussed in Carlier et al. 2021b, the change in the allele frequency spectrum observed in all the 14 populations, with an increase in intermediate-frequency alleles (Figure 3), may reflect a concomitant effect in candidate genomic regions of population contraction and selection on *P. fijiensis* Cuban populations. Constraints such as crop management, pesticide applications or introduced resistance create bottlenecks and genetic drift will randomly maintain some mutations that are putatively beneficial but not others (McDonald & Linde, 2002). Thus, the different history of the populations studied could play a role in the non-parallelism we observed. Furthermore, the *P. fijiensis* populations are panmictic (Carlier et al., 2021a) and recombination should also favor non-parallelism (Barghi et al., 2020). Finally, other factors such as pleiotropy and epistasis not taken into consideration in the present study can influence parallelism (Bailey et al., 2017) and further investigations are needed to better understand the relative role of all the potential factors shaping adaptive architecture in plant pathogen populations.

The results of this study suggest a polygenic basis for adaptation to quantitative resistance and specific host-pathogen interactions in *P. fijiensis*. Specific interactions are not always detected in erosion of quantitative resistance in other plant pathogens (Cowger & Brown, 2019) and general adaptation to quantitative resistance could emerge in such situations even through selective sweeps in a few genes or through polygenic adaptation. General adaptation could lead to an impasse in the use of quantitative resistance since greater pathogen aggressiveness could be selected (Zhan et al., 2015). On the other hand, specific interactions in different cultivars, as observed in *P. fijiensis* and other plant pathogens (Montarry et al., 2012; Delmas et al., 2016; Frézal et al., 2018), can lead to local adaptation patterns (Dumartinet et al., 2019). Specific interactions in different cultivars associated with fitness cost could lead to antagonist selection pressures on the pathogen populations. Assuming that cultivars are the main habitat of pathogen populations, such a situation would resemble the so-called antagonistic pleiotropy process. In this process, alleles have opposite effects on fitness in different habitats, and this is the most important form of genotype x environment interaction involved in local adaptation (Kawecki & Ebert, 2004; Mitchell-Olds et al., 2007; Anderson et al., 2013). Antagonistic adaptation to different quantitative-resistant cultivars could thus be exploited to define durable resistance constraining the evolution of pathogen populations. To this end, fitness cost and adaptive architecture of pathogen populations need to be first analyzed in a wide range of resistant cultivars using similar approaches to the ones applied in this study. Finally, from the without a priori approach used in this study, we were able to highlight major candidate genes which accumulated several characteristics and could now, using functional analysis, be further investigated to better understand the mechanisms involved in the quantitative pathogenicity of fungi such as *P. fijiensis*.

ACKNOWLEDGMENTS

We would like to thank Mathieu Gautier, Florian Charriat, Stéphane De Mita, Pierre Alexandre Gagnaire for helpful discussions. This work is a part of Thomas Dumartinet's PhD financed by a CIRAD/ANSES grant. Experiments and sampling were supported by the Interreg IV Caribbean Program (grant number 31409-Cabaré), European Regional Development Fund and the *Conseil Régional de Guadeloupe* through the MALIN project, and *Agropolis Fondation* (grant number 1504-004- E-Space).

AUTHOR CONTRIBUTIONS

JC designed the project. JC, CA and JA acquired the funding and supervised this work. JC, CA, LPV collected the samples and VR made the collection of isolates. TD and VR acquired sequencing data. Bioinformatics analyses were performed by TD, JC and SR. TD and JC interpreted the results and wrote the manuscript. CA and JA reviewed the manuscript.

DATA AVAILABILITY STATEMENT

Data and Code are deposited at DRYAD (<https://doi.org/10.5061/dryad.rxdwbrv8r>). The repository contains raw DNA sequences, variant calling files (VCF, Sync) from individual and pool sequencing, and custom scripts used in the present study (temporary

link:<https://datadryad.org/stash/share/LTML4RVZ0Sr0y7KNqyNImIGd87N3jRdRF5cxo-rX3tc>).

ORCID

Jean Carlier <https://orcid.org/0000-0002-6967-1852>

Sébastien Ravel <https://orcid.org/0000-0001-6663-782X>

REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). *Basic Local Alignment Search Tool* . 8.
- Anderson, J. T., Lee, C.-R., Rushworth, C. A., Colautti, R. I., & Mitchell-Olds, T. (2013). Genetic trade-offs and conditional neutrality contribute to local adaptation: GENETIC BASIS OF LOCAL ADAPTATION. *Molecular Ecology* , 22 (3), 699–708. <https://doi.org/10.1111/j.1365-294X.2012.05522.x>
- Arango Isaza, R. E., Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C. F., Jong, T. V. de, Vries, I. de, Dietrich, R., Farmer, A. D., Ferreira, C. F., Garcia, S., Guzman, M., Hamelin, R. C., Lindquist, E. A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., ... Kema, G. H. J. (2016). Combating a Global Threat to a Clonal Crop: Banana Black Sigatoka Pathogen *Pseudocercospora fijiensis* (Synonym *Mycosphaerella fijiensis*) Genomes Reveal Clues for Disease Control. *PLOS Genetics* , 12 (8), e1005876. <https://doi.org/10.1371/journal.pgen.1005876>
- Bailey, S. F., Blanquart, F., Bataillon, T., & Kassen, R. (2017). What drives parallel evolution?: How population size and mutational variation contribute to repeated evolution. *BioEssays* , 39 (1), e201600176. <https://doi.org/10.1002/bies.201600176>
- Barghi, N., Hermisson, J., & Schlötterer, C. (2020). Polygenic adaptation: A unifying framework to understand positive selection. *Nature Reviews Genetics* , 21 (12), 769–781. <https://doi.org/10.1038/s41576-020-0250-z>
- Barghi, N., Tobler, R., Nolte, V., Jakšić, A. M., Mallard, F., Otte, K. A., Dolezal, M., Taus, T., Kofler, R., & Schlötterer, C. (2019). Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biology* , 17 (2), e3000128. <https://doi.org/10.1371/journal.pbio.3000128>
- Barton, N. H., Etheridge, A. M., & Véber, A. (2017). The infinitesimal model: Definition, derivation, and implications. *Theoretical Population Biology* , 118 , 50–73. <https://doi.org/10.1016/j.tpb.2017.06.001>
- Barton, N. H., & Keightley, P. D. (2002). Understanding quantitative genetic variation. *Nature Reviews Genetics* , 3 (1), 11–21. <https://doi.org/10.1038/nrg700>
- Bastide, H., Lange, J. D., Lack, J. B., Amir, Y., & Pool, J. E. (2016). *Oligogenic Adaptation, Soft Sweeps, and Parallel Melanic Evolution in Drosophila melanogaster* [Preprint]. *Genetics*. <https://doi.org/10.1101/058008>
- Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., & Loudet, O. (2017). New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype. *Annual Review of Plant Biology* , 68 (1), 435–455. <https://doi.org/10.1146/annurev-arplant-042916-040820>
- Bell, G. (2009). The Oligogenic View of Adaptation. *Cold Spring Harbor Symposia on Quantitative Biology* , 74 (0), 139–144. <https://doi.org/10.1101/sqb.2009.74.003>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* , 289–300.
- Bonhomme, M., Fariello, M. I., Navier, H., Hajri, A., Badis, Y., Miteul, H., Samac, D. A., Dumas, B., Baranger, A., Jacquet, C., & Pilet-Nayel, M.-L. (2019). A local score approach improves GWAS resolution and detects minor QTL: Application to *Medicago truncatula* quantitative disease resistance to multiple *Aphanomyces euteiches* isolates. *Heredity* , 123 (4), 517–531. <https://doi.org/10.1038/s41437-019-0235-x>

- Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* , 169 (7), 1177–1186. <https://doi.org/10.1016/j.cell.2017.05.038>
- Carlier, J., Robert, S., Roussel, V., Chilin-Charles, Y., Lubin-Adjanoh, N., Gilabert, A., & Abadie, C. (2021a). Central American and Caribbean population history of the *Pseudocercospora fijiensis* fungus responsible for the latest worldwide pandemics on banana. *Fungal Genetics and Biology* , 148 , 103528. <https://doi.org/10.1016/j.fgb.2021.103528>
- Carlier, J., Bonnot, F., Roussel, V., Ravel, S., Martinez, R. T., Perez-Vicente, L., Abadie, C., & Wright, S. (2021b). Convergent Adaptation to Quantitative Host Resistance in a Major Plant Pathogen. *MBio* , 12 (1). <https://doi.org/10.1128/mBio.03129-20>
- Cook, L. M., Grant, B. S., Saccheri, I. J., & Mallet, J. (2012). Selective bird predation on the peppered moth: The last experiment of Michael Majerus. *Biology Letters* , 8 (4), 609–612. <https://doi.org/10.1098/rsbl.2011.1136>
- Cowger, C., & Brown, J. K. M. (2019). Durability of Quantitative Resistance in Crops: Greater Than We Know? *Annual Review of Phytopathology* , 57 (1), 253–277. <https://doi.org/10.1146/annurev-phyto-082718-100016>
- Daborn, P. J. (2002). A Single P450 Allele Associated with Insecticide Resistance in *Drosophila*. *Science* , 297 (5590), 2253–2256. <https://doi.org/10.1126/science.1074170>
- Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S., & Manel, S. (2018). Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics* , 19 (1). <https://doi.org/10.1186/s12864-018-4579-z>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics* , 27 (15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- De Mita, S., & Siol, M. (2012). EggLib: Processing, analysis and simulation tools for population genetics and genomics. *BMC Genetics* , 13 (1), 27. <https://doi.org/10.1186/1471-2156-13-27>
- Delmas, C. E. L., Fabre, F., Jolivet, J., Mazet, I. D., Richart Cervera, S., Delière, L., & Delmotte, F. (2016). Adaptation of a plant pathogen to partial host resistance: Selection for greater aggressiveness in grapevine downy mildew. *Evolutionary Applications* , 9 (5), 709–725. <https://doi.org/10.1111/eva.12368>
- Derbyshire, M. C., Denton-Giles, M., Hane, J. K., Chang, S., Mousavi-Derazmahalleh, M., Raffaele, S., Buchwaldt, L., & Kamphuis, L. G. (2019). A whole genome scan of SNP data suggests a lack of abundant hard selective sweeps in the genome of the broad host range plant pathogenic fungus *Sclerotinia sclerotiorum*. *PLOS ONE* , 14 (3), e0214201. <https://doi.org/10.1371/journal.pone.0214201>
- Dumartinet, T., Abadie, C., Bonnot, F., Carreel, F., Roussel, V., Habas, R., Martinez, R. T., Perez-Vicente, L., & Carlier, J. (2019). Pattern of local adaptation to quantitative host resistance in a major pathogen of a perennial crop. *Evolutionary Applications* . <https://doi.org/10.1111/eva.12904>
- Dutta, A., Hartmann, F. E., Francisco, C. S., McDonald, B. A., & Croll, D. (2021). Mapping the adaptive landscape of a major agricultural pathogen reveals evolutionary constraints across heterogeneous environments. *The ISME Journal* . <https://doi.org/10.1038/s41396-020-00859-w>
- Eoche-Bosy, D., Gautier, M., Esquibet, M., Legeai, F., Bretaudeau, A., Bouchez, O., Fournet, S., Grenier, E., & Montarry, J. (2017). Genome scans on experimentally evolved populations reveal candidate regions for adaptation to plant resistance in the potato cyst nematode *Globodera pallida* . *Molecular Ecology* , 26 (18), 4700–4711. <https://doi.org/10.1111/mec.14240>

- Escobar-Tovar, L., Guzman-Quesada, M., Sandoval-Fernandez, J. A., & Gomez-Lim, M. A. (2015). Comparative analysis of the in vitro and in planta secretomes from *Mycosphaerella fijiensis* isolates. *Fungal Biology* , 119 (6), 447–470. <https://doi.org/10.1016/j.funbio.2015.01.002>
- Fariello, M. I., Boitard, S., Mercier, S., Robelin, D., Faraut, T., Arnould, C., Recoquillay, J., Bouchez, O., Salin, G., Dehais, P., Gourichon, D., Leroux, S., Pitel, F., Leterrier, C., & SanCristobal, M. (2017). Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: The local score approach. *Molecular Ecology* , 26 (14), 3700–3714. <https://doi.org/10.1111/mec.14141>
- Frezal, L., Jacqua, G., & Neema, C. (2018). Adaptation of a Fungal Pathogen to Host Quantitative Resistance. *Frontiers in Plant Science* , 9 . <https://doi.org/10.3389/fpls.2018.01554>
- Gagnaire, P.-A., & Gaggiotti, O. E. (2016). Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology* , 62 (6), 603–616. <https://doi.org/10.1093/cz/zow088>
- Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics* , 201 (4), 1555–1579. <https://doi.org/10.1534/genetics.115.181453>
- Graves, J. L., Hertweck, K. L., Phillips, M. A., Han, M. V., Cabral, L. G., Barter, T. T., Greer, L. F., Burke, M. K., Mueller, L. D., & Rose, M. R. (2017). Genomics of Parallel Experimental Evolution in *Drosophila* . *Molecular Biology and Evolution* , msw282. <https://doi.org/10.1093/molbev/msw282>
- Griffin, P. C., Hangartner, S. B., Fournier-Level, A., & Hoffmann, A. A. (2017). Genomic Trajectories to Desiccation Resistance: Convergence and Divergence Among Replicate Selected *Drosophila* Lines. *Genetics* , 205 (2), 871–890. <https://doi.org/10.1534/genetics.116.187104>
- Guzman, M., Perez-Vicente, L., Carlier, J., Abadie, C., De Lapeyre de Bellaire, L., Carreel, F., Marin, D. H., Romero, R. A., Gauhl, F., Pasberg-Gauhl, C., & et al. (2019). Black leaf streak. In *Handbook of Diseases of Banana, Abaca and Enset* (Jones DR edition, pp. 41–115). CAB International.
- Halkett, F., Coste, D., Rivas Platero, G. G., Zapater, M. F., Abadie, C., & Carlier, J. (2010). Genetic discontinuities and disequilibria in recently established populations of the plant pathogenic fungus *Mycosphaerella fijiensis*: GENETIC STRUCTURE AND COLONIZATION. *Molecular Ecology* , 19 (18), 3909–3923. <https://doi.org/10.1111/j.1365-294X.2010.04774.x>
- Hansen, T. F. (2006). The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics* , 37 (1), 123–157. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110224>
- Hartmann, F. E., Sanchez-Vallet, A., McDonald, B. A., & Croll, D. (2017). A fungal wheat pathogen evolved host specialization by extensive chromosomal rearrangements. *The ISME Journal* , 11 (5), 1189–1204. <https://doi.org/10.1038/ismej.2016.196>
- Hivert, V., Leblois, R., Petit, E., Gautier, M., & Vitalis, R. (2018). Measuring genetic differentiation from Pool-seq data. *BioRxiv* . <https://doi.org/10.1101/282400>
- Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *The American Naturalist* , 188 (4), 379–397. <https://doi.org/10.1086/688018>
- Hollinger, I., Pennings, P. S., & Hermisson, J. (2019). Polygenic adaptation: From sweeps to subtle frequency shifts. *PLOS Genetics* , 15 (3), e1008035. <https://doi.org/10.1371/journal.pgen.1008035>
- Kamvar, Z. N., Tabima, J. F., & Grunwald, N. J. (2014). *Poppr* : An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* , 2 , e281. <https://doi.org/10.7717/peerj.281>

- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters* , 7 (12), 1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>
- Kessner, D., Turner, T. L., & Novembre, J. (2013). Maximum Likelihood Estimation of Frequencies of Known Haplotypes from Pooled Sequence Data. *Molecular Biology and Evolution* , 30 (5), 1145–1158. <https://doi.org/10.1093/molbev/mst016>
- Kofler, R., Orozco-terWengel, P., De Maio, N., Pandey, R. V., Nolte, V., Futschik, A., Kosiol, C., & Schlotterer, C. (2011). PoPoolation: A Toolbox for Population Genetic Analysis of Next Generation Sequencing Data from Pooled Individuals. *PLoS ONE* , 6 (1), e15925. <https://doi.org/10.1371/journal.pone.0015925>
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* , 9 (1), 29.
- Landry, C., Bonnot, F., Ravigne, V., Carlier, J., Rengifo, D., Vaillant, J., & Abadie, C. (2017). A foliar disease simulation model to assist the design of new control methods against black leaf streak disease of banana. *Ecological Modelling* , 359 , 383–397. <https://doi.org/10.1016/j.ecolmodel.2017.05.009>
- Lannou, C. (2012). Variation and Selection of Quantitative Traits in Plant Pathogens. *Annual Review of Phytopathology* , 50 (1), 319–338. <https://doi.org/10.1146/annurev-phyto-081211-173031>
- Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics* , 26 (5), 589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Linnen, C. R., Poh, Y.-P., Peterson, B. K., Barrett, R. D. H., Larson, J. G., Jensen, J. D., & Hoekstra, H. E. (2013). Adaptive Evolution of Multiple Traits Through Multiple Mutations at a Single Gene. *Science* , 339 (6125), 1312–1316. <https://doi.org/10.1126/science.1233213>
- Lipka, A. E., Tian, F., Wang, Q., Peiffer, J., Li, M., Bradbury, P. J., Gore, M. A., Buckler, E. S., & Zhang, Z. (2012). GAPIT: Genome association and prediction integrated tool. *Bioinformatics* , 28 (18), 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* , 24 (5), 1031–1046. <https://doi.org/10.1111/mec.13100>
- Maynard-Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* , 23 (1), 23–35. <https://doi.org/10.1017/S0016672300014634>
- McDonald, B. A., & Linde, C. (2002). Pathogen population genetics, evolutionary potential, and durable resistance. *Annual Review of Phytopathology* , 40 (1), 349–379. <https://doi.org/10.1146/annurev.phyto.40.120501.101443>
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* , 20 (9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- Messer, P. W., & Petrov, D. A. (2013). Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* , 28 (11), 659–669. <https://doi.org/10.1016/j.tree.2013.08.003>
- Mitchell-Olds, T., Willis, J. H., & Goldstein, D. B. (2007). Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics* , 8 (11), 845–856. <https://doi.org/10.1038/nrg2207>
- Montarry, J., Cartier, E., Jacquemond, M., Palloix, A., & Moury, B. (2012). Virus adaptation to quantitative plant resistance: Erosion or breakdown? *Journal of Evolutionary Biology* , 25 (11), 2242–2252. <https://doi.org/10.1111/j.1420-9101.2012.02600.x>

- Noar, R. D., & Daub, M. E. (2016). Transcriptome sequencing of *Mycosphaerella fijiensis* during association with *Musa acuminata* reveals candidate pathogenicity genes. *BMC Genomics* , 17 (1). <https://doi.org/10.1186/s12864-016-3031-5>
- Ohm, R. A., Feau, N., Henrissat, B., Schoch, C. L., Horwitz, B. A., Barry, K. W., Condon, B. J., Copeland, A. C., Dhillon, B., Glaser, F., Hesse, C. N., Kosti, I., LaButti, K., Lindquist, E. A., Lucas, S., Salamov, A. A., Bradshaw, R. E., Ciuffetti, L., Hamelin, R. C., . . . Grigoriev, I. V. (2012). Diverse Lifestyles and Strategies of Plant Pathogenesis Encoded in the Genomes of Eighteen Dothideomycetes Fungi. *PLoS Pathogens* , 8 (12), e1003037. <https://doi.org/10.1371/journal.ppat.1003037>
- Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., Minchin, R. M., O'Hara, R. B., Gavin, L. S., Solymos, P., Stevens, M. H., Szoecs, E., & Wagner, H. (2012). *vegan: Community Ecology Package* (R package version 2.5-6.) [Computer software]. <https://CRAN.R-project.org/package=vegan>
- Park, J.-H., Wacholder, S., Gail, M. H., Peters, U., Jacobs, K. B., Chanock, S. J., & Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature Genetics* , 42 (7), 570–575. <https://doi.org/10.1038/ng.610>
- Parlevliet, J. E. (2002). Durability of resistance against fungal, bacterial and viral pathogens; present situation. *Euphytica* , 124 (2), 147–156. <https://doi.org/10.1023/A:1015601731446>
- Pavlidis, P., & Alachiotis, N. (2017). A survey of methods and tools to detect recent and strong positive selection. *Journal of Biological Research-Thessaloniki* , 24 (1). <https://doi.org/10.1186/s40709-017-0064-0>
- Perez Miranda, M. P., Vicente, L. P., & Trujillo, R. (2006). *Variabilidad de mycosphaerella fijiensis morelet. Estabilidad de la resistencia a sigatoka negra de los clones hibridos de la FHIA* . 11.
- Pilet-Nayel, M.-L., Moury, B., Caffier, V., Montarry, J., Kerlan, M.-C., Fournet, S., Durel, C.-E., & Delourme, R. (2017). Quantitative Resistance to Plant Pathogens in Pyramiding Strategies for Durable Crop Protection. *Frontiers in Plant Science* , 8 . <https://doi.org/10.3389/fpls.2017.01838>
- Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation – not by sweeps alone. *Nature Reviews Genetics* , 11 (10), 665–667. <https://doi.org/10.1038/nrg2880>
- Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology* , 20 (4), R208–R215. <https://doi.org/10.1016/j.cub.2009.11.055>
- Rao, C. R. (1964). The Use and Interpretation of Principal Component Analysis in Applied Research. *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)* , 26 (4), 329–358.
- Rode, N. O., Holtz, Y., Loridon, K., Santoni, S., Ronfort, J., & Gay, L. (2018). How to optimize the precision of allele and haplotype frequency estimates using pooled-sequencing data. *Molecular Ecology Resources* , 18 (2), 194–203. <https://doi.org/10.1111/1755-0998.12723>
- Rousset, F. (2008). genepop'007: A complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources* , 8 (1), 103–106. <https://doi.org/10.1111/j.1471-8286.2007.01931.x>
- Rovenich, H., Boshoven, J. C., & Thomma, B. P. (2014). Filamentous pathogen effector functions: Of pathogens, hosts and microbiomes. *Current Opinion in Plant Biology* , 20 , 96–103. <https://doi.org/10.1016/j.pbi.2014.05.001>
- Schlötterer, C., Tobler, R., Kofler, R., & Nolte, V. (2014). Sequencing pools of individuals—Mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics* , 15 (11), 749–763. <https://doi.org/10.1038/nrg3803>
- Segura, V., Vilhjálmsson, B. J., Platt, A., Korte, A., Seren, Ü., Long, Q., & Nordborg, M. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* , 44 (7), 825–830. <https://doi.org/10.1038/ng.2314>

- Sella, G., & Barton, N. H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics* , 20 (1), 461–493. <https://doi.org/10.1146/annurev-genom-083115-022316>
- Shabana, Shahid, S. U., & Hasnain, S. (2018). Use of a gene score of multiple low-modest effect size variants can predict the risk of obesity better than the individual SNPs. *Lipids in Health and Disease* , 17 (1). <https://doi.org/10.1186/s12944-018-0806-5>
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* , 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R. L., McDonald, P., Bennett, A. F., Long, A. D., & Gaut, B. S. (2012). The Molecular Diversity of Adaptive Convergence. *Science* , 335 (6067), 457–461. <https://doi.org/10.1126/science.1212986>
- Urban, M., Pant, R., Raghunath, A., Irvine, A. G., Pedro, H., & Hammond-Kosack, K. E. (2015). The Pathogen-Host Interactions database (PHI-base): Additions and future developments. *Nucleic Acids Research* , 43 (D1), D645–D655. <https://doi.org/10.1093/nar/gku1165>
- van't Hof, A. E., Edmonds, N., Dalikova, M., Marec, F., & Saccheri, I. J. (2011). Industrial Melanism in British Peppered Moths Has a Singular and Recent Mutational Origin. *Science* , 332 (6032), 958–960. <https://doi.org/10.1126/science.1203043>
- Vatsiou, A. I., Bazin, E., & Gaggiotti, O. E. (2016). Detection of selective sweeps in structured populations: A comparison of recent methods. *Molecular Ecology* , 25 (1), 89–103. <https://doi.org/10.1111/mec.13360>
- Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics* , 101 (1), 5–22. <https://doi.org/10.1016/j.ajhg.2017.06.005>
- Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting Natural Selection in Genomic Data. *Annual Review of Genetics* , 47 (1), 97–120. <https://doi.org/10.1146/annurev-genet-111212-133526>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution* , 38 (6), 1358. <https://doi.org/10.2307/2408641>
- Wen, Y.-J., Zhang, H., Ni, Y.-L., Huang, B., Zhang, J., Feng, J.-Y., Wang, S.-B., Dunwell, J. M., Zhang, Y.-M., & Wu, R. (2018). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics* , 19 (4), 700–712. <https://doi.org/10.1093/bib/bbw145>
- Wiberg, R. A. W., Gaggiotti, O. E., Morrissey, M. B., & Ritchie, M. G. (2017). Identifying consistent allele frequency differences in studies of stratified populations. *Methods in Ecology and Evolution* , 8 (12), 1899–1909. <https://doi.org/10.1111/2041-210X.12810>
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., Nurkowski, K. A., Smets, P., Wang, T., Gray, L. K., Liepe, K. J., Hamann, A., Holliday, J. A., Whitlock, M. C., Rieseberg, L. H., & Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science* , 353 (6306), 1431–1433. <https://doi.org/10.1126/science.aaf7812>
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* , 8 (1), 28–36. <https://doi.org/10.1111/2041-210X.12628>
- Zapater, M.-F., Abadie, C., Pignolet, L., Carlier, J., & Mourichon, X. (2008). Diagnosis of *Mycosphaerella* spp., responsible for *Mycosphaerella* leaf spot diseases of bananas and plantains, through morphotaxonomic observations. *Fruits* , 63 (6), 389–393. <https://doi.org/10.1051/fruits:2008039>
- Zhan, J., Thrall, P. H., Papaix, J., Xie, L., & Burdon, J. J. (2015). Playing on a Pathogen's Weakness: Using Evolution to Guide Sustainable Plant Disease Control Strategies. *Annual Review of Phytopathology* , 53 (1),

19–43. <https://doi.org/10.1146/annurev-phyto-080614-120040>

Zhong, Z., Marcel, T. C., Hartmann, F. E., Ma, X., Plissonneau, C., Zala, M., Ducasse, A., Confais, J., Compain, J., Lapalu, N., Amselem, J., McDonald, B. A., Croll, D., & Palma-Guerrero, J. (2017). A small secreted protein in *Zymoseptoria tritici* is responsible for avirulence on wheat cultivars carrying the *Stb6* resistance gene. *New Phytologist*, 214 (2), 619–631. <https://doi.org/10.1111/nph.14434>

TABLES

Table 1 Sampling information on the *P. fijiensis* populations collected in 2011 and 2013 in three locations (Provinces) in Cuba on the susceptible (Macho 3) and the resistant banana cultivar (FHIA18). The phenotype and genetic group corresponding to the two cultivars of origin and the number of isolates used for each pool or for individual sequencing are indicated for each population.

Sampling year	Location / Province (code)	Cultivar of origin			Population code	Number of isolates	
		Name (code)	Phenotype	Group		Pool Sequencing	Individual Sequencing
2011	Villa Clara (1)	Macho 3/4 (S2)	Susceptible	AAB	CU1S2_2011	40	32
		Fhia18 (R2)	Resistant	AAAB	CU1R2_2011	48	31
	Ciego de Avila (2)	Macho 3/4 (S2)	Susceptible	AAB	CU2S2_2011	49	0
		Fhia18 (R2)	Resistant	AAAB	CU2R2_2011	58	0
	Matanzas (3)	Macho 3/4 (S2)	Susceptible	AAB	CU3S2_2011	38	0
		Fhia18 (R2)	Resistant	AAAB	CU3R2_2011	38	0
2013	Villa Clara (1)	Macho 3/4 (S2)	Susceptible	AAB	CU1S2_2013	95	0
		Fhia18 (R2)	Resistant	AAAB	CU1R2_2013	58	0
	Ciego de Avila (2)	Macho 3/4 (S2)	Susceptible	AAB	CU2S2_2013	135	0
		Fhia18 (R2)	Resistant	AAAB	CU2R2_2013	46	0
	Ciego de Avila (2b)	Macho 3/4 (S2)	Susceptible	AAB	CU2bS2_2013	106	0
		Fhia18 (R2)	Resistant	AAAB	CU2bR2_2013	106	0
	Matanzas (3)	Macho 3/4 (S2)	Susceptible	AAB	CU3S2_2013	103	0
		Fhia18 (R2)	Resistant	AAAB	CU3R2_2013	96	0

Hosted file

image2.wmf available at <https://authorea.com/users/437202/articles/538997-complex-adaptive-architecture-of-quantitative-resistance-erosion-in-a-plant-fungal-pathogen>

Table 2 Candidate genomic regions involved in host adaptation detected among Pool-Seq data using genome scan analysis.

Note: The genome scan results and genetic differentiation (F_{ST}) obtained with the pools sampled in 2011 and in 2013 are presented in this table. Information concerning every region detected with genome scan analyses, i.e., the name of the region, the corresponding scaffold, the position on the scaffold (start/end) and the size of the region are provided. For each genomic region, tick symbols highlighted in orange indicate which method (BayPass with the covariate Cov-co for 2011 and 2013, BayPass with the covariates Cov-dR or Cov-dS for 2011 or poolFreqDiff for 2011 and 2013) and which local score tuning parameter value ($\xi=1, 2, \text{ or } 3$) was used to detect the region. Significant F_{ST} values ($p < 0.05$) between each pair of populations sampled in the same location are indicated in green.

Table 3 Relationship between haplotypes in candidate genomic regions involved in host adaptation and traits related to quantitative pathogenicity

Hosted file

image3.wmf available at <https://authorea.com/users/437202/articles/538997-complex-adaptive-architecture-of-quantitative-resistance-erosion-in-a-plant-fungal-pathogen>

Note: Summary of the results obtained with the Wilcoxon test or the redundancy analysis (RDA) for all the genomic regions detected using genome scan analyses. Information (name, corresponding scaffold and size) for each region is given. For the two phenotypic traits investigated (DLA-R or DLA-S), the haplotype (identified with individual sequencing data) significantly correlated with the corresponding trait (i.e. the advantageous haplotype) is indicated by a colored circle (blue = Haplotype 1, green = Haplotype 2). Green ticks or red crosses indicate which method was able to define the advantageous haplotype. The regions in

which haplotypes were identified as being significantly associated with a higher diseased leaf area measured on the resistant cultivars are written in blue, on the susceptible cultivar in green, and on both cultivars in blue and green. DLA-R and DLA-S=Diseased Leaf Area measured on the resistant (R) or the susceptible (S) cultivars.

Table 4 Architecture of host adaptation in *P. fijiensis* populations

Hosted file

image4.wmf available at <https://authorea.com/users/437202/articles/538997-complex-adaptive-architecture-of-quantitative-resistance-erosion-in-a-plant-fungal-pathogen>

Note: The predominant haplotype in the 14 Cuban populations and 13 candidate regions was identified by comparing haplotype frequencies between each population pair (susceptible versus resistant). The regions in which haplotypes were previously identified as being significantly associated with a higher diseased leaf area (DLA) measured on the resistant cultivars are written in blue, on the susceptible cultivar in green, and on both cultivars in blue and green. The predominant haplotype in each population is identified by a blue circle when haplotype 1 is predominant and by a green circle when haplotype 2 is predominant. Darker shades of each color indicate that the Fisher's exact used to test for difference in haplotype frequencies was significant at a 5% threshold, and paler colors at a 10% threshold.

Hosted file

image5.wmf available at <https://authorea.com/users/437202/articles/538997-complex-adaptive-architecture-of-quantitative-resistance-erosion-in-a-plant-fungal-pathogen>

Table 5 Summary of the gene annotation results in candidate genomic regions involved in host adaptation

Note: For each candidate genomic region, the total number of genes identified and the number of genes detected in the different annotation databases consulted (i.e. presence of a peptide signal, in the *in-vitro* secretome of *P. fijiensis*, in the *in-planta* transcriptome of *P. fijiensis* and of homologs in the Pathogen-Host Interactions database (PHI-Base)) are indicated. Major gene candidates involved in quantitative pathogenicity and host adaptation (highlighted in red) were located in regions significantly correlated to DLA-R or DLA-S and had several hits in the databases consulted. Other good candidates (highlighted in grey) were located in regions significantly correlated to DLA-R or DLA-S but had only one hit in the databases consulted or were not in correlated regions but have several hits in the databases (details and references are given in Table S5). DLA-R and DLA-S=Diseased Leaf Area measured on the resistant (R) or the susceptible (S) cultivars.

FIGURES

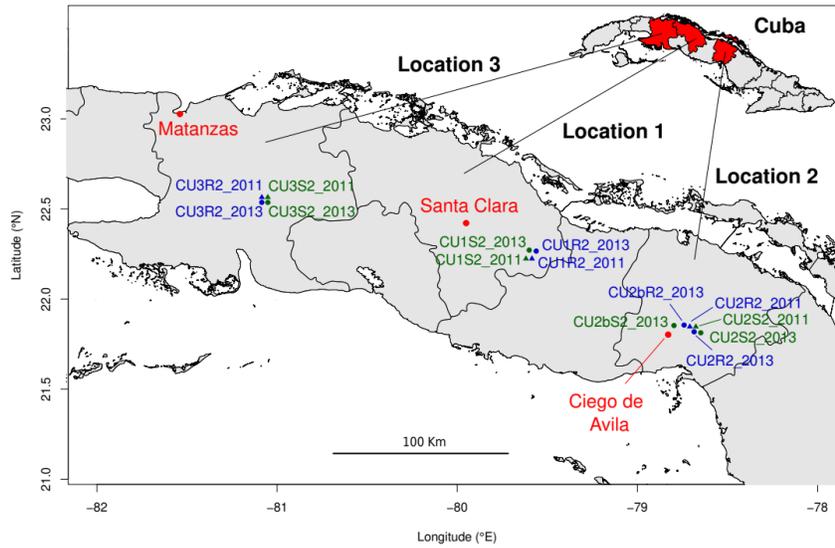


Figure 1 Map detailing the paired population design for samples of *P. fijiensis* collected in Cuba. The three Cuban provinces sampled are shown in red. The 14 populations studied were collected from either a susceptible (green) or a resistant cultivar (blue) and the locations of the sampled plots are indicated by triangles for populations collected in 2011 and by dots for populations collected in 2013.

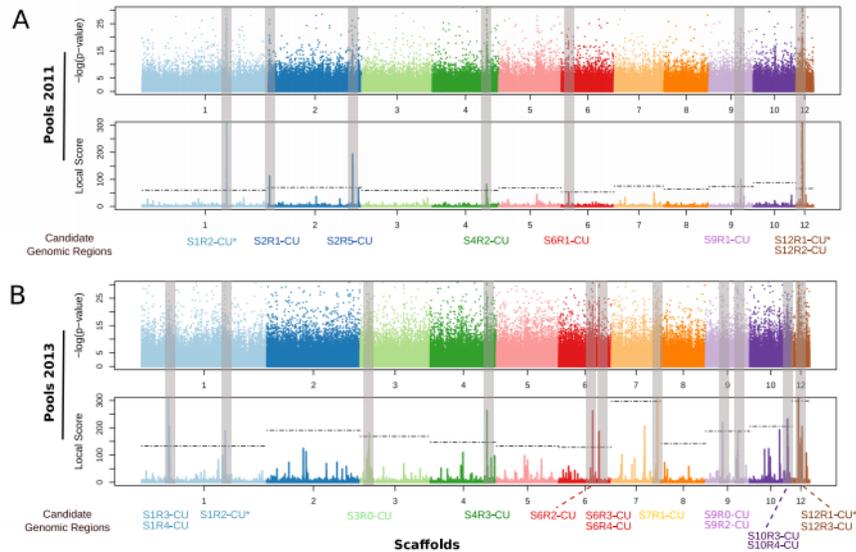


Figure 2 Manhattan plots showing selection footprints detected between samples collected from susceptible or resistant cultivars in Cuba in 2011 (A) and in 2013 (B). The x axis shows the distribution of each SNP along the core genome of *P. fijiensis* (scaffolds are represented by different colors) and the y axis indicates the p-value of each SNP obtained with the association test using BayPass and the covariate Cov-Co or the local score obtained with $\xi=2$. The horizontal dashed lines correspond to the chromosome-wide threshold α

= 1% calculated for each scaffold. For both sampling years, all genomic regions containing SNPs with local scores above the significance threshold are indicated by gray vertical bars. Regions identified in populations sampled in 2011 and 2013 are indicated with a * symbol.

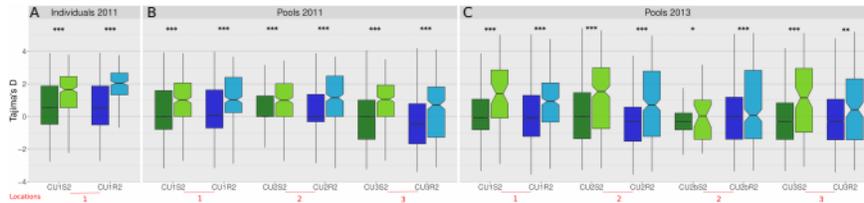


Figure 3 Boxplots representing Tajima’s D distribution for the whole genome (dark colors) or for the candidate genomic regions (pale colors) in all the populations sampled on the susceptible variety (green) and the resistant variety (blue) for the 63 sequenced isolates collected from location 1 in 2011 (A), or for the pool sampled in 2011 (B) and for the pool sampled in 2013 (C). For each population, the two distributions (whole genome versus candidate genomic regions) were compared using a Wilcoxon-Mann-Whitney test and the significance of the p-values is indicated with symbols (***: p-values < 0.001, **: < 0.01, *: < 0.05).

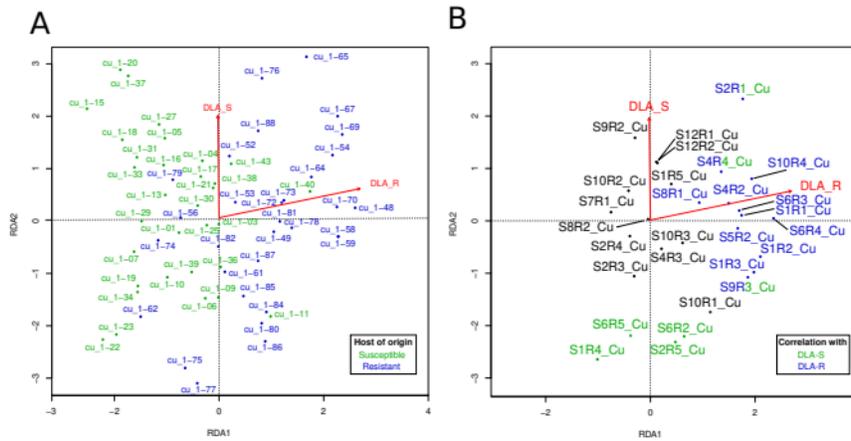


Figure 4 Redundancy analysis (RDA) computed with DLA-R and DLA-S traits and in candidate genomic regions. This figure shows the projection of all isolates sampled from FHIA 18 (blue) or Macho $\frac{3}{4}$ (green) and the two variables (DLA-R and DLA-S) in the RDA space (A) and the projection showing the contribution to the two variables of the 28 genomic regions (B), with regions significantly correlated with DLA-R and DLA-S indicated in blue and green, respectively. Regions in black were correlated to none of both traits. DLA-R and DLA-S=Diseased Leaf Area measured on the resistant (R) or the susceptible (S) cultivars.

SUPPLEMENTARY MATERIAL

Supplementary_Information.pdf

Supplementary_Tables.xlsx