**Title**: Genomics of local adaptation in blue tit

**Running title**: Population genomics in blue tit

**Authors**: Perrier C. 1, Rougemont Q. 2, Charmantier A. 3

**Addresses**: 1. Centre de Biologie pour la Gestion des Populations UMR CBGP, INRAE, CIRAD, IRD, Montpellier SupAgro, Univ Montpellier, Montpellier, France.

2. Département de Biologie, Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, Québec, Canada QC G1V OA6.

3. Centre d'Ecologie Fonctionnelle et Evolutive UMR 5175, Campus CNRS, 1919 route de Mende, 34293 Montpellier Cedex 5, France.

**Correspondence**: Perrier Charles charles5perier@icloud.com

**Abstract**

Understanding the genomic processes underlying local adaptation is a central aim of modern evolutionary biology. This task requires identifying footprints of local selection but also estimating spatio-temporal variation of populations' demography and variation in recombination rate and diversity along the genome. Here, we investigated these parameters in blue tit populations inhabiting neighbouring deciduous and evergreen forests and populations in an insular versus a continental context. Close populations from deciduous and evergreen habitats were weakly genetically differentiated ($F_{ST}$ = 0.004 on average), nevertheless with a significant effect of habitat type on the overall genetic structure. This low differentiation was consistent with the large effective population sizes (from 43,000 to 463,000) and the strong and long-lasting gene flow inferred by demographic modeling. In turn, insular and continental populations were moderately differentiated ($F_{ST}$ = 0.08 on average), which was consistent with the inference of moderate ancestral migrations followed by isolation since the end of the last glaciation. Weak and non-parallel footprints of divergent selection among deciduous and evergreen populations were consistent with their demography and the probable polygenic nature of local adaptation in these habitats. This contrasted with stronger outlier regions, more often in regions of low recombination, found between insular and continental populations. Lastly, we identified a genomic inversion on the continent, spanning 2.8Mb. These results provide insights into the demographic history and genetic architecture of local adaptation in blue tit populations at multiple geographic scales.

## Introduction

Local adaptation in heterogeneous environments imposing divergent selection on counterparts of a same species has fascinated scientists for decades (Kawecki & Ebert 2004; Blanquart *et al.* 2013). However, it is only with the recent advance in sequencing technologies that numerous recent empirical studies started discovering the genomic features and processes underlying local adaptation in heterogeneous environments (Savolainen *et al.* 2013; Tiffin & Ross-Ibarra 2014; Manel *et al.* 2016; Hoban *et al.* 2016). Especially, many studies used so-called $F_{ST}$ genome scans between populations (Excoffier *et al.* 2009; Narum & Hess 2011; de Villemereuil *et al.* 2014) to detect genes under selection and potentially implicated in local adaptations (Hohenlohe et al. 2010; Lamichhaney et al. 2016; Reid et al. 2016; Perrier et al. 2017). However, genome scans need improvements for example for studying polygenic variation (Rockman 2012; Wellenreuther & Hansson 2016). We also need to better look for structural variants such as inversions since they protect adaptive gene sets from recombination (Kirkpatrick 2006; 2010; Stapley *et al.* 2017; Wellenreuther *et al.* 2019) and hence enable their persistence and rapid redeployment in heterogeneous environments despite high gene flow (Lowry & Willis 2010; Sinclair-Waters et al. 2017; Todesco et al. 2019). In addition, scans for adaptive variants need to be interpreted in the light of variation of recombination along the genome (Stapley et al. 2017) since it shapes the local extent of diversity and of differentiation (Cutter & Payseur 2013; Tine et al. 2014; Burri et al. 2015; Rougemont et al. 2019). Lastly, genome scans results need to be interpreted in the light of the demographic history of populations. Especially,

difficult to discriminate genuine footprints of recent divergent selection from regions of reproductive isolation accumulated during allopatry showing increased differentiation in secondary contacts (Bierne et al. 2011; 2013). In true cases of divergence with gene flow it is also important to estimate spatio-temporal variation in gene flow and of effective population size since both parameters affect the possibility of local adaptations (Lenormand 2002).

Here, we aimed at investigating various demographic and genomic aspects of the adaptive divergence of a well-studied passerine bird, the Blue tit. Populations of small insectivorous passerines have long been used to study local adaptation, in both quantitative genetics and population genetics frameworks (Carbonell et al. 2003; Broggi et al. 2005; Laaksonen et al. 2015). In particular, several blue tit (*Cyanistes caeruleus*) populations breeding in a heterogeneous habitat in Southern France (Figure 1A) offer an ideal context to study local adaptation and have been subject to a long term project spanning more than 40 years (Blondel et al. 2006; Charmantier et al. 2016). These populations show marked quantitative phenotypic differences (Figure 1B&C), notably in morphological (e.g. tarsus length and body mass), life-history (lay date and clutch size) and behavioural traits (e.g. song characteristics and handling aggression) (Charmantier et al 2016). These phenotypic differences were found at two spatial scales. First, birds breeding in deciduous forest habitats are taller, more aggressive, and lay larger and earlier broods than birds in evergreen forests (see table 1 in Charmantier et al 2016). Strikingly, neighbouring populations in deciduous and evergreen habitats are weakly genetically differentiated (Porlier et al. 2012b; Szulkin et al. 2016; Dubuc-Messier et al.

2018), which questioned the mechanisms of persistence of the observed phenotypic differentiation against presumably large gene flow. Second, insular blue tits from Corsica, that might have been isolated since the sea level rise after the last glacial maximum (the sea level raised of 120 m from 17,000 to 5,000 years ago (Lambeck & Bard 2000; Jouet et al. 2006)), are smaller and more colourful than their mainland counterparts (again, see table 1 in Charmantier et al 2016) and are listed as two subspecies. Overall, traits displaying these strong phenotypic differences harboured significant genetic variation, with heritability ranging from 0.20-0.43 (eg for lay date, (Caro et al. 2009)) to 0.42-0.60 (eg for tarsus length, (Teplitsky et al. 2014; Delahaie et al. 2017; Perrier et al. 2018)), which were potentially adaptive (Lambrechts et al. 1997; Porlier et al. 2012a). In particular, the heritability of lay date and the breeding time gap between populations could result in reproductive isolation by breeding time, limiting gene flow and favouring local adaptation. The studied traits were typically quantitative (Charmantier et al. 2016), hence much probably controlled by a polygenic architecture (Perrier et al. 2018) composed of many loci with small individual effects, as found in similar traits in other passerine birds (Santure et al. 2013; Bosse et al. 2017; Hansson et al. 2018; Lundregan et al. 2018). Overall, given their phenotypic, demographic and genetic characteristics, these blue tit populations provide an ideal framework to study the genomic architecture of polygenic adaptation in heterogeneous environments.

We investigated genome wide patterns of genetic diversity and differentiation and the demographic history among several populations of blue tits from Southern France, in heterogeneous forest habitats (deciduous versus evergreen) and in insular

(Corsica island) and continental areas (Mainland France) in order to better understand the determinants of their local adaptation. The analysis was based on RADsequecing of birds sampled in four sites studied in the context of a long-term project (Blondel et al. 2006; Charmantier et al. 2016) together with three additional pairs of deciduous and evergreen forests in order to test for parallel evolution (see Figure 1A). First, we investigated variation in genetic diversity and differentiation in order to verify that habitat type and geographic distance explained a significant proportion of the genetic structure among populations (Szulkin et al. 2016). Second, we investigated the historical and contemporary demography of each deciduous and evergreen population pairs in order to better understand the potential cause of their differentiation, and the demography between the populations on mainland France and on the Corsica Island in order to verify their split time and their subsequent connectivity and effective population size. Third, we screened the genome for $F_{ST}$ outlier loci and inversions potentially implicated in adaptation in the two habitat types and between Corsica and Mainland France. Rather than identifying the genes under selection, we wanted to test whether the adaptation had a polygenic or oligogenic architecture and whether outlier loci would be more often found in regions with low recombination rate. We discuss our results in the light of the recent literature in population genomics, to decipher the role of genomic and demographic variations on the genetic and phenotypic divergence among blue tits in deciduous *versus* evergreen habitats and in continental *versus* insular contexts.

## Methods

### Sites and sampling

947 blue tits (*Cyanistes caeruleus*) were captured in ten locations (Figure 1, Table 1). Three locations are located in the south of mainland France (codes staring with "A": A1D, A2D, A2E) and seven in the Corsica Island (staring with "B": B3E, B4D, B4E, B5D, B5E, B6D, B6E). The numbers from 1 to 6 indicated the sampling area and hence the population pairs. Half of these sites were dominated by the deciduous downy oak *Quercus pubescens* (indicated by "D" in the code) and the other half are dominated by the evergreen holm oak *Q. ilex* (indicated by "E"). Four of these locations (A1D, B3E, B4D, B4E) are monitored as part of a long-term research program started in 1976 and have been described in previous studies (Blondel et al. 2006; Charmantier et al. 2016). The 6 other sites (A2D, A2E, B5D, B5E, B6D & B6E) were sampled in order to obtain further replicates of neighboring deciduous and evergreen populations. Capture and handling of the birds was conducted under permits provided by the *Centre de Recherches sur la Biologie des Populations d'Oiseaux* (CRBPO) and by the *Direction Départementale des Services Vétérinaires* (DDSV). Birds were captured during the reproductive period, from early April to late June, on their nesting territory. Birds were captured either in nest boxes during the feeding of their young (in the four sites studied on the long term), or using nets in the vicinity of their nest (in the other sites). Birds were banded with a unique metal ring provided by the CRBPO. Five to 20µl of blood were sampled from a neck or a wing vein from breeding blue tits. Blood was stored at 4°C in Queen's buffer (Seutin et al. 1991).

**Molecular biology and sequencing**

DNA extractions were achieved using Qiagen DNeasy Blood & Tissue kits and were randomized across sites. DNA was quantified using first a NanoDrop ND8000 spectrophotometer and then a Qubit 2.0 fluorometer with the DNA HS assay kit (Life Technologies). DNA quality was examined on agarose gels. Library preparation using RAD-seq (restriction-site-associated DNA sequencing; (Baird et al. 2008)) with the enzyme SbfI was done by Montpellier GenomiX (MGX) platform (CNRS, Montpellier). Each individual was identified using a unique 6 nucleotides tag, individuals were multiplexed in equimolar proportions by libraries of 36 individuals, and each library was sequenced on a lane of an Illumina HiSeq 2000. This design was used to obtain an average read depth of 50x. The DNA of three individuals was extracted twice and processed in different libraries to test for reliability of the genotyping process.

**Bioinformatics and data filtering**

Raw sequences were inspected with *FastQC* (Andrews 2010) for quality controls. Potential fragments of Illumina adapters were trimmed with *Cutadapt* (Martin 2011), allowing for a 10% mismatch in the adapter sequence. Reads were filtered for overall quality, demultiplexed and trimmed to 85bp using *process_radtags, from the Stacks software pipeline* 1.39 (Catchen et al. 2013), allowing for one mismatch in the barcode sequence. Sequencing RAD-tags resulted in a median value of 5,449,564 reads per individual. *BWA-MEM* 0.7.13 (Li & Durbin 2009) was used to map individual sequences against the reference genome of the great tit *Parus major* (Laine et al. 2016) and to produce *sam* files using default options. *Samtools* 0.1.19 (Li et al. 2009) was used to

build and sort *bam* files. Back in *Stacks*, we used *pstacks* to treat *bam* files, align the reads into matching stacks, infer loci, and detect SNPs at each locus. We used a minimum depth of coverage (m) of 5, the SNP model, and alpha = 0.05 (chi square significance level required to call a heterozygote or homozygote). *cstacks* was used to build the catalogue of loci, allowing 3 mismatches between sample loci when building the catalog. *sstacks* was used to match loci against the catalog. Lastly, *populations Stacks* program was used to genotype individuals. In this program, relatively permissive filters were applied, to retain SNP genotyped in at least 50% of individuals (all individuals from all sites grouped) and with a minimum read depth of 4. 350,941 SNP and 947 individuals were obtained in this "GeneralDataset".

We filtered this dataset using the programs *VCFtools* (Danecek et al. 2011) and *Plink* (Purcell et al. 2007). For subsequent analyses SNP were filtered for a minimum average read depth of 10 across all genotypes, and a maximum average read depth of 100 across all blue tit genotypes. SNP genotyped in less than 80% of the individuals were removed. We removed SNPs with heterozygosity ≥ 0.65 among individuals on Corsica or on mainland. Individuals below 85% genotyping rate were removed. Identity-By-State (IBS) was estimated for three replicated individuals (from the DNA extraction to SNP-calling) in order to investigate reliability of the entire genotyping process. The IBS measured between each replicated samples was high, ranging from 0.9989 to 0.9999, indicating very low genotyping error rate. These replicates were then removed from the dataset. Using the R packages *gdsfmt* and *SNPRelate* (Zheng et al. 2012), we measured the realized genomic relatedness (GRM) between individuals in order to remove highly

related individuals (*i.e.* full-sibs and parent-offspring). For each pair of individuals with relatedness ≥ 0.35, we removed one individual. This procedure was applied in order to limit biases due to highly related individuals. We then limited the number of individuals to 60, chosen at random, in populations (A1D, B3E, B4D & B4E) in which a large number of individuals were genotyped. This limitation was achieved in order to limit differences in analyses precision between populations due to sample size (All individuals will be used in another ongoing study of the genomic architecture of quantitative adaptive traits). We hence removed potential monomorphic SNPs and created the dataset "FilteredDataset" later used for all of the population genomics analyses. 144,773 SNPs and 454 individuals were kept in the final dataset (Table 1). The median missing rate across all SNPs for these individuals was 0.019. The median read depth across genotypes (SNPs x Individual) was 49.2. The number of SNPs per chromosome ranged from 143 (LGE22) to 15,738 (chromosome 2). Further filtering (e.g. MAF) was often operated, depending on the analyses, and thereafter mentioned if it was the case. Lastly, we produced a haplotype VCF file from these individuals with the population module of Stacks to perform demographic inference.

**Analysis of genetic structure and effects of environmental variables**

We used PCAs with the function *snpgdsPCA* from *snprelate* to depict genetic structure between individuals. PCAs were run for the entire dataset, for continental and Corsican populations separately and for each pairs of deciduous and evergreen populations. We inferred admixture proportions for each individual using *Admixture* 1.23 (Alexander et al. 2009) with K-values ranging from 1 to 12 and 1,000 bootstraps.

The different clustering solutions were inferred by plotting cross-validations errors and by plotting individual admixture proportion. The effect of environmental variables, forest phenology (E versus D), geographic distance (latitude & longitude) and altitude, on genomic differentiation was measured using a redundancy analysis (RDA) (Legendre & Fortin 2010; Forester et al. 2015). We used the RDA implemented in the *Vegan* R package (Oksanen et al. 2007). We investigated the proportion of genetic variability explained by a constraining covariance matrix consisting of phenology, latitude, longitude, and altitude for each individual. We tested the global significance of the model using 1,000 permutations. We ran marginal effects permutations to address the significance of each variable. Ultimately, we focused on the effect of phenology alone, using partial RDA first taking into account the effect of latitude, longitude and altitude. Significance was tested running 1,000 permutations. For the PCA, the admixture analysis and the RDA, we selected from the dataset the SNPs with more than 95% genotyping rate, MAF > 0.05, retaining one SNP per locus and we remove SNP in linkage disequilibrium using the Plink command "indep 50 5 2". Genome wide differentiation between each sampling location was measured with Weir and Cockerham's $F_{ST}$ estimator (Weir & Cockerham 1984) implemented in *StAMPP* (Pembleton et al. 2013). $F_{ST}$ was estimated for all the SNPs, the ones on the autosomes and the ones on the sex chromosome Z separately. Significance was assessed using 1,000 bootstraps replicates.

**Analysis of demographic history**

Alternative models of divergence history including the effects of selection at linked sites affecting *Ne* and of differential introgression (*m*) were compared using an

ABC framework modified from (Roux et al. 2016). Six scenarios were compared for the 4 pairs of deciduous and evergreen populations in order to test whether the divergence between different habitats was not due to a divergence in different historical refugia but to contemporary ecological divergence. We included a model of panmixia (PAN), a model of equilibrium corresponding to the island model with two populations (EQ), a model of isolation with migration (IM), a model of secondary contact (SC), a model of divergence with migration during the first generations, *i.e.* ancestral migration (AM), and a model of strict isolation (SI). The prior and details of the simulation pipeline are fully described in supplementary text 1. The PAN model assumes that the two focal populations descent from a single panmictic population characterized by it effective size (Ne). The EQU model assumes that the two populations of size $N_{\mathrm{pop1}}$ and $N_{\mathrm{pop2}}$ are connected by continuous gene flow at a constant rate each generation. The SI, SC, and AM models all assume that an ancestral population of size $N_{\mathrm{ANC}}$ splits at $T_{split}$ into two daughter populations of size $N_{\mathrm{pop1}}$ and $N_{\mathrm{pop2}}$. Under SI no subsequent gene flow occurs. Under AM model, gene flow occurs from $T_{split}$ to $T_{am}$ and is followed by a period without gene flow. Under IM, gene flow is continuous after $T_{split}$. Under, the SC model, $T_{split}$ is followed by a period of strict isolation, after which a secondary contact starts $T_{\mathrm{sc}}$ generations. The EQU, IM, AM and SC models included migration as $M = 4\ N_0.m$, with $M_{1\leftarrow 2}$ being the number of migrants from population 2 to population 1 and $M_{2\leftarrow 1}$ being the reverse. The effect of linked selection and barriers to gene flow by modeling heterogeneous population size and heterogeneous migration were modeled using beta distributions as hyper-prior on each of these two parameters (Supplementary table 1).

These resulted in four alternative versions for models with gene flow EQ, AM, IM, SC (NhomoMhomo, NhomoMhetero, NheteroMhomo, NheteroMhetero) and two versions for PAN and SI (Nhomo and Nhetero). We used a modified ABC (Csilléry et al. 2010) pipeline from (Rougemont & Bernatchez 2018) to perform model selection, compute robustness and we estimated posterior probabilities of parameters.

We also investigated the historical demography of the populations from Corsica as compared to the ones from the continent. Gene flow has probably been impossible since the last de-glaciation and sea-level rise. Therefore, we compared models of ancient migration (AM) and Strict Isolation (SI). The pipeline described in the above section, integrating linked selection and barriers to gene flow, was run between the A1D samples (chosen on the continent because it had the largest sample size) and B4D (chosen in Corsica because it had both a large sample size and was from the same habitat as A1D). We used the same ABC pipeline for the model selection procedure and parameter estimation as described in the above section.

**Genomic diversity**

Genome wide genetic diversity was inferred for each sampling location in the dataset, by measuring expected heterozygosity (He) with *Vcftools*, proportion of polymorphic loci and MAF spectrums with R, and nucleotide diversity (π) with *Plink*. For each population and chromosome, and subsequently on average for the entire genome, Linkage Disequilibrium (LD) decay was measured with *Plink* and smoothed in R. To contrast with long term Ne estimates from coalescent simulation in our ABC modeling (see "Analysis of demographic history"), we also inferred recent Ne for each population

using *SNeP* V1.1 (Barbato et al. 2015), which uses LD data, with a MAF ≥ 0.05 filter per population. We investigated the nature of SNP variation, *i.e.* synonymous or non-synonymous, blasting on the reference genome and transcriptome of the great tit using *blastx* (McGinnis & Madden 2004). We kept hits with at least 90% similarity and more than 25 amino acid similarity between each read and the transcriptome. We kept only SNPs for which both the alternative and reference allele yields the same score.

**Identification of genomic footprints of selection**

We searched for outlier SNPs at two geographic levels: i) between the mainland and Corsican populations and ii) between local pairs of evergreen (E) and deciduous (D) populations, *i.e.* A2D vs A2E, B4D vs B4E, B5D vs B5E, B6D vs B6E. We identified SNPs potentially under divergent selection using *Bayescan* V2.1 (Foll & Gaggiotti 2008). We filtered each of the five datasets for a minimum MAF of 0.05. We used default parameters except for prior odds that we sat at 10,000 in order to limit false positives, and set the other parameters to default values. We investigated the parallelism across pairs of D-E environments. Then, we estimated $F_{ST}$ along windows along the genome, using either a 200kb average with *VCFtools*, or the function "snpgdsSlidingWindow" from the package SNPrelate to estimate $F_{ST}$ in blocks of 50 SNPs moving by 5 SNPs. The second window strategy was used in order to compensate for lower SNP density in regions of lower recombination that tend to exaggerate the contribution of individual SNPs in these regions and to dilute the individual SNP contribution in regions of higher recombination (see Perrier & Charmantier 2019 for a broader comment on this).

**Variation of genomic diversity and differentiation between and along chromosomes**

We investigated variation among chromosomes of $\pi$, $F_{ST}$, the proportion of non-synonymous SNPs and we tested for correlations between these parameters and chromosome length as well as local or chromosome averaged recombination rate. We also estimated whether $F_{ST}$ outliers were preferentially found more in regions of recombination. We estimated local recombination rates using a coalescent method implemented in *Ldhat* (McVean 2004) using linkage disequilibrium signal. Following these authors' recommendation, the dataset was split in blocks of 2,000 SNPs with 500 overlapping SNPs. A MAF of 0.05 and a maximum of 5% of missing data were allowed. The local recombination rate, $\rho = 4Ne\ r$, was then estimated in each blocks independently with the Bayesian reversible jump MCMC scheme implemented in interval. We use a block penalty of 5, with 30 millions MCMC iterations sampled every 5,000 iterations. The first 250,000 iterations were discarded. To speed up computations we used the pre-computed two locus likelihood table for n = 190 and assuming theta = 0.001. We estimated $\rho$ in a composite dataset with individuals from every population, habitats and sex. We also calculated $\rho$ separately for males and females.

**Detection of genomic inversions**

We searched for potential genomic inversions using a variety of descriptive statistics. First, we searched for genomic regions having a particularly low recombination rate and large long-distance linkage disequilibrium, nevertheless associated to a high density of SNPs and therefore unlikely to correspond to pericentromeric regions but rather to local suppression of recombination that may be due to inversions. Second, we implemented a PCA sliding window analysis in order to

identify portions of the genome with individuals carrying an inversion at the homozygous or heterozygous state or individuals exempt from the inversion (Ma & Amos 2012). We first used 10Mb windows sliding by 1Mb and then 1Mb windows sliding by 100kb, in order to use enough SNPs to perform PCAs. We then used *Lostruct* (Li & Ralph 2019), with k=2, sliding by 100 SNPs, to identify particular blocks of linked SNPs explaining an abnormally high proportion of variance between 2 groups of individuals (e.g. inverted and non-inverted). We verified, using admixture with K=2 for analyzing SNPs from an inversion, that putative heterozygous individuals had an admixture ratio close to 1:1 of both putative inverted and non-inverted homozygous clusters. We then inspected variations of $F_{ST}$ (per SNP) and $\pi$ (per 10kb window) along the genome among individuals that were inverted homozygous, non-inverted homozygous, and heterozygous for the inversion, looking for potentially reduced diversity and increased differentiation at the inversion. Finally, we tested whether the detected inversions were at Hardy-Weinberg equilibrium and whether the frequency of the inversions varied geographically and between evergreen and deciduous habitats.

## Results

### Genetic structure and effects of environmental variables

The admixture analysis suggested the existence of 2 main distinct genetic groups corresponding to continental and island populations (Figure 2A). Increasing K value contributed to delineate the populations, showing the existence of a weak structure between the 6 main areas. However, the coefficient of variation increased with K. There

was very little evidence for a genetic structure within each of the four pairs of D- and E-populations. PCAs also showed a clear structure between individuals from the continent and the island but little structure within each of these 2 groups (Supplementary figure 1). The full RDA model was significant ($p < 0.001$, Supplementary table 2), as well as the effects of each variable tested in the model: phenology (*i.e.* E versus D, $p = 0.006$), latitude ($p < 0.001$), longitude ($p < 0.001$) and altitude ($p < 0.001$). The first axis of the global RDA explained 4.54% of the variance, and was correlated mainly with latitude (-0.987) and longitude (+0.999) (Supplementary table 2, Figure 2B). The fourth axis of this global RDA displayed the strongest correlation with phenology (-0.856), and explained 0.27% of the variance. The partial RDA model conditioning for the effects of latitude, longitude and altitude, was globally significant ($p < 0.001$), as well as the effect of phenology ($p <0.001$). The axis of the partial RDA explained 0.29% of the variance and was correlated with phenology (0.922).

All pairwise $F_{ST}$ comparisons were significantly different from zero (Table 2). Average $F_{ST}$ calculated between island and continental populations was 0.081. It was on average 0.003 between close populations within the mainland and 0.005 within the island. $F_{ST}$ was low, on average 0.003, for each of the four pairs of D- and E- populations, ranging from 0.0006 between B6D and B6E to 0.0079 between B4D and B4E. The $F_{ST}$ estimated on the Z chromosome were on average 1.52 times higher than on the autosomes; 1.62 times higher when comparing populations from the continent versus the Corsica Island and 1.44 times higher when comparing neighboring populations in the island or in the continent. The low differentiation between populations was also

captured by low levels of sequence divergence, as measured by a Dxy of 0.41% between continent versus island populations and from 0.31% to 0.35% for pairs of D- and E-populations. Similarly, the net divergence (Da) value was 3e-04 among continent versus Island populations and below 3e-05 between close populations.

**Demographic history**

When deciphering the historical demography of deciduous and evergreen population pairs, the hierarchical model choice procedure strongly rejected models of strict isolation (SI), ancient migration and of panmixia (PAN), which were associated with a posterior probability of 0 (Supplementary table 3). Instead, models with gene flow were highly supported. In three of the four pairs, the equilibrium model (EQU) received the highest posterior probability with P(EQU) = 0.99 and 0.95, 0,94 in the B6D vs B6E, B5D vs B5E and B4D vs B4E comparisons, respectively. In the A2D vs A2E comparisons, the best supported model was the isolation with migration model, P(IM) = 0,74. Across all models, comparisons with heterogeneous gene flow and heterogeneous effective population size were not supported, indicating that, if genetic barriers or linked selection were at play they could not be detected. Demographic parameters were estimated for each pair of populations under the best model (Supplementary table 3). Posterior distributions were well differentiated from their prior indicating that estimated parameters were confidently estimated. Effective population size ($N_e$) where higher in Deciduous than in Evergreen habitat with value ranging from 43,000 to 463,000, and of 139,250 on average (*Ne*1 in Table 2, Figure 3A, Supplementary table 3). $N_e$ tended to be higher on the continent than on the islands. Finally, under the EQ

model, we found that in 3 out of 4 instances, migration rates (m) were highly asymmetric, stronger from the Deciduous to the Evergreen habitat (average *m*~2e-4) than from the Evergreen to the deciduous habitat (average *m ~ 6e-5;* Figure 3A, Supplementary table 3). This, combined with higher effective population sizes of the deciduous populations resulted into far higher gene flow (Nm) from the deciduous to the Evergreen habitats (average = 39 individuals/generation) than from the Evergreen to the deciduous ones (average = 3 individuals by generation)

Deciphering the historical demography between Corsica and the continent, the model choice unambiguously showed that the ancient migration models (AM) strongly outperformed the strict isolation modelss (SI), with p(AM) = 0.988 (Supplementary table 4). Among the AM model, the simplest model without linked selection or barriers to gene flow was the best supported, with p = 0.743. Parameters estimates revealed posterior generally well differentiated from the prior providing increased confidence (Supplementary table 4). Effective population size of the continent was 397,685 [CI = 244,375 -- 600,105], around 7.5 times greater than the one of the island, 52,990 [CI = 33,065 -- 78,475] (Figure 3AB, Ne2 in Table 2, Supplementary table 4). Our analysis also indicated a strong population size change during the process of population isolation, since the ancestral Ne was estimated at only 240 [CI = 65 -- 685]. Split time was estimated at 1,417,480 [CI = 740,000 -- 2,394,000] generations ago, hence around 3.2M years ago (assuming a 2.3 years generation time (Charmantier et al. 2004)) and gene flow subsequently stopped 4540 [CI = 2500 -- 6360] generations ago, hence around 10,000 years ago. The ancestral migration rate between Corsica and the continent, from

the split in two populations to the end of the latest gene flow between these populations, was inferred to have been asymmetric, with a 5 times larger migration rate from the island to the continent (*m ~ 1.9e-4)* than the other way around *(m ~ 3.8e-5*), however without generating strong difference in gene flow between the two groups given the lower effective size of the population on the island compared to the one on the continent, (i.e. respective number of migrants of 10 and 15 from Corsica to the Continent and in the reverse direction, with overlapping credible intervals, see details in Supplementary table 3).

**Genetic diversity**

The continental populations displayed significantly higher patterns of expected heterozygosity (on average 0.169, table 2) than Island populations (on average 0.147), all t-tests' p-values < 0.001. On the contrary there was no significant difference of expected heterozygosity at smaller spatial scale among populations within the island (either same or different tree phenology, E- and D-) or within the mainland (all t-tests' p-value > 0.05). Similarly, nucleotide diversity, $\pi$, was slightly higher for mainland populations than island populations (on average 0.0037 versus 0.0033, respectively; Table 2) but was not different between populations within the island or within the mainland. More SNPs were polymorphic in continental populations (77,306 to 80,616 SNPs for a MAF > 0.05) than in island populations (62,728 to 65,876 SNPs for a MAF > 0.05). The MAF spectrum showed enrichment of variants with smaller frequencies in continental populations versus island populations (Supplementary figure 2A). LD decayed rapidly in the first 5kb and was lower in populations from the continent

(especially A1D) compared to island populations (Supplementary figure 2B). The pattern of LD decay was similar among chromosomes (eg for chromosomes 1, 2 and Z, Supplementary figure 3). Contemporary $N_e$ inferred from LD varied from 142 (in A2E) to 355 (in A1D) (Ne2 in Table 2, Supplementary figure 2C). $N_e$ values were rather similar within pairs of E- and D- populations and varied more markedly between pairs of populations, irrespective of their location on the continent or the island. Among the entire dataset, 2599 SNPs were identified as non-synonymous while 6751 SNPs were identified as synonymous. As, expected, we observed lower allele frequencies for non-synonymous than for synonymous variants at the different spatial scales considered, either at the population level, at the level of mainland populations or at the Corsican level (all t.tests' p-values < 0.000001). However, we did not find significant enrichment for non-synonymous variants in populations from either deciduous *vs* evergreen environments or from Corsica *vs* mainland populations.

**Footprints of selection**

Manhattan plots of $F_{ST}$ per SNP revealed no clear evidence for high peaks of $F_{ST}$ for the five comparisons considered (Figure 4 & 5). Throughout the five Bayescan tests (Corsica vs mainland populations and the 4 D-4 population pairs), we identified 61 SNPs with a log10(BF) > 0 among which 18 SNPs had a log10(BF) > 1 (Supplementary table 5, Supplementary figure 4). None of these SNPs were detected twice among the different tests. Among these 61 SNPs, 27 were found for the Corsica vs mainland test, 2 were found for A2D vs A2E, 9 were found for B4D vs B4E, 0 were found for B5 vs B5E and 2 were found for B6D vs B6E (Supplementary table 5). Among these 61 SNPs, 19 genes

were identified. Sliding windows of $F_{ST}$ along the genome helped revealing a few outlier windows, of relatively modest intensities in the case of deciduous and evergreen pairwise tests (Figure 4, Supplementary table 6), with almost no parallelism between the tests. More noticeable outlier windows were found for the Corsica *vs* continent comparison (Figure 5, Supplementary table 7).

**Variation of genomic diversity and differentiation between and along chromosomes**

Local recombination rate differed between chromosomes and sexes (ANOVA $p < 2e\text{-}16$ & $p < 2e\text{-}16$). Average recombination rate, rho, per chromosome was negatively correlated to chromosome length, with (linear model $p = 1.6e\text{-}4$) or without ($p = 3.4e\text{-}4$, Supplementary figure 5A) the chromosome Z. The Z chromosome had the smallest recombination rate, 3.28 time smaller than the average, while it was 2.39 time longer than the average. Recombination rates in females and males were highly correlated ($p = 2.0e\text{-}13$, Supplementary figure 5B) and on average 1.23 times higher in females than in males. To test how recombination rate constrained genetic diversity and differentiation, we performed a series of linear regressions and Anovas.

Nucleotide diversity, $\pi$, differed between chromosomes and with recombination rate (ANOVA $p < 2e\text{-}16$ & $p < 2e\text{-}16$). Average $\pi$ per chromosome was negatively correlated to average recombination rate per chromosome ($p = 1.3e\text{-}3$, Supplementary figure 5D). The Z chromosome showed the highest deviations from this correlation, given its large size and small average recombination rate, with a 22% reduced $\pi$ compared to the average. From a genomic perspective, all chromosomes taken

together, $\pi$ was positively correlated to local recombination rate (Supplementary figure 5E; $p < 2e-16$).

$F_{ST}$ between continental and Corsican populations differed between chromosomes and with recombination rate (ANOVA $p < 2e-16$ & $p < 2e-16$). Average $F_{ST}$ per chromosome between continental and Corsican populations was not correlated to chromosome length (Supplementary figure 5G). Average $F_{ST}$ per chromosome between continental and Corsican populations was negatively correlated to average recombination rate ($p = 1.3e-2$). This correlation was not significant when the chromosomes Z and 27 were excluded ($p = 0.21$, Supplementary figure 5G). $F_{ST}$ between continental and Corsican populations was negatively correlated to local recombination rate, all chromosomes taken together (Figure 6A; $p < 2e-16$). $F_{ST}$ in each pairs of deciduous and evergreen populations was not correlated to local recombination rate, all chromosomes taken together (Supplementary figure 5H). $F_{ST}$ SNP outliers between Corsica and the mainland populations were more often found in regions of low recombination (ki2 p-value < 0.01, Figure 6B) and average recombination rate was on average twice smaller for outlier SNPs compared to the rest of the dataset (t-test p-value = 3.1e-6). This enrichment in regions of low recombination was not observed for $F_{ST}$ SNP outliers between deciduous and evergreen populations (Figure 6B).

The proportion of non-synonymous mutations, versus synonymous, differed between chromosomes and with recombination rate (ANOVA $p = 4.5e-10$ & $p = 0.03$, respectively). The Z chromosome had the highest proportion of non-synonymous mutations (45%), 1.83 times higher than the average (of 24%) of other chromosomes.

This proportion of non-synonymous mutations was marginally negatively correlated to average recombination rate per chromosome (p = 6.5e-2, Supplementary figure 5i). Considering the entire genome, the proportion of non-synonymous mutations was not correlated to local recombination rate (p = 0.37, Supplementary figure 5J).

**Genomic inversions**

We detected one putative inversion on chromosome 3, spanning 2.8Mb, from position 11,838,789 to 14,661,550, and containing 390 SNPs. The estimated recombination rate was on average 9 times lower (t-test: p<2.2e-16) in this region of the chromosome compared to the rest of the chromosome (Figure 7A). The SNP density did not appear reduced at the location of this putative inversion (Figure 7A). The sliding PCA detected either 2 groups (individuals from Corsica and from mainland) outside of the putative inversion, and 4 groups at the putative inversion location, separating individuals from Corsica, and 3 groups for mainland individuals (Figure 7B). We assumed that the inverted segment was the one with the lower frequency in mainland and that was absent from Corsica. The Lostruct analysis confirmed for this same region the presence of four successive blocks of 100 SNPs highly discriminating individuals (Figure 7C). The average $F_{ST}$ among the three groups of mainland individuals was much higher (t-test: p<2.2e-16, Figure 7D) for SNPs in the putative inversion ($F_{ST}$ = 0.26) than for SNPs outside of this region ($F_{ST}$ = 5.8e-4). $\pi$ was lower (t-test: p=1.9e-5, Figure 7E) for SNPs located in the putative inversion for inverted homozygous individuals ($\pi$ = 2.5e-5) compared to non-inverted homozygous individuals ($\pi$ = 5.3e-5). Both inverted and non-inverted homozygous individuals had lower $\pi$ at this inversion compared to

heterozygous individuals at this region ($\pi$ = 7.6e-5) and compared to the three types for the rest of the chromosome ($\pi$ ranging from 7.4e-5 to 7.6e-5). The PCA (Figure 7F) and the admixture analysis (Figure 7G) using the SNPs from the inversion clearly delineated inverted, non-inverted and heterozygous individuals, with heterozygous individuals falling at intermediate eigenvalue between the two categories of homozygous individuals. LD was higher (Figure 7H) for the region of the putative inversion (Little LD was found when only homozygous non inverted individuals were kept in the analysis). We did not found any evidence for clear variations of read depth in or in the close vicinity of this inversion. This region contains 15 genes (Supplementary table 8). Out of 113 individuals from the mainland, we found 4 inverted homozygous, 23 heterozygous, and 86 non-inverted homozygous, hence we observed no deviation from the Hardy-Weinberg equilibrium (ki2 p-value > 0.1). The percentage of each genotype was similar in each mainland location and habitat type (the inverted segment was at 11% in A1D, 16% in A2D, 17% in A2E, and 14% across these 3 locations).

## Discussion

In this study, we investigated demographic history and genome wide patterns of genetic diversity and differentiation among several populations of blue tits presenting strong phenotypic differences in deciduous and evergreen forest habitats and in insular and continental contexts. Demographic modelling revealed that each pair of deciduous and evergreen populations most likely derived from locally unique ancestral populations. This modeling also showed large gene flow and effective population sizes

(Ne) among deciduous and evergreen habitats. Moreover, gene flow was higher from deciduous to evergreen populations than the reverse and Ne was larger in the former than the later, concordant with a higher productivity of the deciduous forest patches than the evergreen ones. We identified slight and mostly unrepeated footprints of divergent selection across these population pairs, which is consistent with their demography and the likely polygenic nature of several traits implicated in their local adaptation. Between insular and continental populations, we found large effective population sizes and a break up of gene flow since the end of the last glaciation, explaining moderate genome-wide differentiation. We identified several divergent genomic regions, enriched in regions of low recombination, between these groups of populations. We also identified a putative genomic inversion spanning 2.8Mb, polymorphic in continental populations only. We discuss these results in the context of the vast scientific knowledge acquired on these blue tit populations during the last four decades and more generally in the context of genomics of adaptation with gene flow.

**Genome wide divergence between populations in deciduous and evergreen habitats**

Although we found a significant effect of habitat on genetic structure, the genetic differentiation between neighbouring deciduous and evergreen populations was low ($F_{ST}$ ranging from 0.0006 to 0.0079). This result was in line with the primary observations realized earlier on a restricted set of populations (Porlier et al. 2012b; Szulkin et al. 2016). Accordingly, we found high and asymmetric gene flow (on average 39 individuals/generation from deciduous to evergreen populations). Yet, this quantification of gene flow and the low genetic structure contrasted with the

demographic knowledge collected on blue tits. Indeed, demographic studies suggested very restricted dispersal between these populations, with 4 dispersal events observed between B4D and B4E (5.6 km apart) and none between the B3E and either B4D or B4E (24.1 km), among a total of 2788 males, 2672 females and 25,158 nestling ringed in the three main Corsican sites between 1976 and 2018 with 47% of parents captured more than once and a mean recruitment rate of 6% (Charmantier, com pers). Our interpretation of this result is triple. First, dispersal estimation on the field using capture-mark-recapture is a very challenging and demanding task. Since natal dispersal in blue tits classically ranges between 330 m and 4 km (see (Tufto et al. 2005; Ortego et al. 2011)), the long-term monitoring sites in Corsica equipped with nest-boxes (black circled dots in figure 1C) are not ideally spaced to identify the origin of immigrants and the destination of emigrants, and only a small fraction of the landscape favourable to blue tit breeding is covered. Second, only few migrants are sufficient to decrease the genetic distance between two populations, measured through the $F_{ST}$ (Marko & Hart 2011; Cayuela et al. 2018), and in that regard, our results may be compatible with the few dispersal events recorded throughout the years. Lastly, the number of migrants estimated here should not be interpreted at the restricted temporal scale of the study. Rather, they should be considered at a larger temporal scale during which dispersal may have varied between years or even decades. We found large Ne (ranging from 43,000 to 463,000), on average larger in deciduous than in evergreen populations. This could be explained by the higher productivity of deciduous forests, resulting in larger clutches and more fledging in this habitat, compared to evergreen forests. However, the

coalescent Ne estimated here is Kingman's coalescent (Kingman 1982) which corresponds to the time for all lineages to coalesce into a common ancestor and hence shouldn't be interpreted as the number of breeding individuals. In turn, the estimations of Ne using the LD signal (Ne ranging from 142 to 355) could be much more comparable to the numbers of reproductive individuals found in the forest patches. However, given that this Ne estimation method is biased when applied in opened populations connected by strong gene flow (Waples & Do 2008), as in the case of these blue tit populations, these numbers should be interpreted cautiously.

Demographic modeling indicated that secondary gene flow between previously isolated populations, which could have eroded most genome-wide differentiation outside of regions implicated in local adaptation and/or acting as reproductive barriers, in particular in low recombining areas (Bierne et al. 2013), was very unlikely. Indeed, secondary contacts were strongly rejected in favor of models supporting ongoing gene-flow without divergence (EQ model) in 3 out of 4 population pairs, the last one supporting an isolation with migration model (IM). These results suggest that these population pairs have been continuously connected by relatively high gene flow. Hence genetic divergence between populations in each pair did not originate or build up on pre-existing genetic divergence accumulated during an allopatric phase but rather is supported as a case of genuine recent ecological divergence (Pinho & Hey 2010; Wang & Bradburd 2014). However, as often, our demographic estimates should be interpreted cautiously in the absence of a well-documented mutation rate in blue tit. Considering

credible intervals around parameter estimates nevertheless informs on the uncertainty of our parameters and their biological relevance.

We did not find strong footprints of divergent selection in any of the four pairs of populations (Figure 4). This result is in line with the high migration rates observed, which should limit the potential for local adaptation (since the accumulation of allelic differentiation is very unlikely) (Lenormand 2002). The observed outliers were not found repeatedly across the 4 deciduous-evergreen comparisons but rather each outlier was found only once, and they were not enriched in regions of low recombination. First, this pattern is consistent with the best demographic model being the equilibrium model and not a secondary contact. Indeed, the later would more often create repeated and strong outliers located in regions implicated in reproductive isolation between the ancient populations (eg (Rougemont et al. 2016)). Second, this pattern suggests that local adaptation in these populations could involve alleles of small effects, probably transient (Yeaman & Whitlock 2011; Yeaman 2015). Third, our result is also consistent with the fact that most documented traits that are involved in the adaptation of blue tits to habitat type (such as clutch size or laying date (Lambrechts et al. 1997; Blondel et al. 1998)) are quantitative and most likely rely on a polygenic architecture with alleles of small effects (see e.g. (Santure et al. 2013)). However, some statistical issues may limit the interpretation of these results. First, currently available genome-scan methods to detect alleles of small effects are statistically limited, especially when applied to relatively small datasets (Rockman 2012; Hoban *et al.* 2016). Second, a potential lack of power in detecting outliers could arise from the combination of the use of

RADsequencing and a rapid LD decay along the genome, that may result in too low resolution especially in regions with larger recombination rates (Lowry et al. 2017). Further analyses of the potential outliers found here, as well as a new analysis with higher marker density and more individuals, are needed to better document and discuss the potential genes and biological processes implicated. Particularly, quantitative genomics will be useful to establish links between the level of divergent selection on these putative targets and their effect size on phenotypic trait variation (Stinchcombe & Hoekstra 2007; Gagnaire & Gaggiotti 2016). This would notably reveal whether genes under divergent selection are also responsible for the observed phenotypic variations and assessing of how much of phenotypic variation is adaptive.

Lastly, we searched for inversions potentially associated to phenotypic variation and/or segregating in both habitats. We combined multiple evidences (Figure 6) for a putative inversion on chromosome 3 in continental populations. However, the proportion of inversions in the genomes of deciduous- *versus* evergreen-breeding birds did not differ, which most probably precludes a putative role in local adaptation in these habitats. Besides, this polymorphism followed the Hardy-Weinberg equilibrium and we found 4 homozygous inverted individuals, suggesting this inversion does not involve an accumulation of lethal recessive deleterious alleles (Jay et al. 2019). The functional consequences of this inversion on variation in life history traits in blue tit would need further investigations, as achieved in other songbirds (Tuttle et al. 2016; Knief et al. 2017; Kim et al. 2017). In order to genotype this inversion in more birds and to link it to putative phenotypic variation, we could use a PCR-RFLP approach (da Silva et al. 2019).

**Genome wide divergence between populations in continental and insular contexts**

Our demographic modelling approach of Corsican and continental populations revealed that a model with ancient migration was the most probable, which is coherent with the history previously reconstructed for the blue tit complex. Split time of the ancestral populations in two populations connected by gene flow (~3M years) was relatively coherent with the diversification time found in the literature for the entire blue tit complex (5M (Illera et al. 2011)). Gene flow between the island and the continent likely stopped around 10,000 years, which is compatible with a stop of gene flow due to the rise of the sea level and hence an increase of the flooded area between Corsica and the continent at the end of the last glaciation, from 17,000 to 5,000 years ago (Lambeck & Bard 2000; Jouet et al. 2006). The population expansion during the diversification period and the gene flow 5 times larger from the island to the continent than from the continent to the island were coherent with a suspected recolonisation of the continent from the island, as proposed by Illera et al. (2011) based on analyses of nuclear and mitochondrial DNA sequences. The larger effective population size on the continent than on the island may be due to multiple sources of colonisation of the continent (Taberlet et al. 1992) and to the much larger continental area compared to the island. It would be interesting to analyse one population from the South-East of Europe, one from Sardinia and one from the Iberian Peninsula, in order to determine whether the large size estimated for the continental population studied here might be explained by a mixed recolonisation from these refugium (Kvist et al. 1999; 2004) and whether insular populations consistently have reduced effective population size. It

would also be interesting to explore a model allowing for multiple cycles of isolations and contacts that could have resulted from successive glacial cycles (Hewitt 2004).

Given the demographic parameters inferred, it is not surprising that the genetic differentiation between populations from the continent and the island was moderate (FST = 0.08). However, this differentiation value contrasted with the much larger divergence at mitochondrial DNA observed for these populations (OST = 0.67 between A1D and B4D (Kvist et al. 2004)). This discrepancy could first be explained by the typically smaller Ne and hence larger drift and divergence in mitochondrial DNA. Second, accumulation of cytonuclear incompatibilities could limit its introgression (Burton & Barreto 2012). Alternatively, gene flow may have been greater for males than females during the colonisation process from Corsica to the continent, at the end of the last glacial period. However, this hypothesis is not supported by the general observation that females disperse on average further away than males and that rare long distance dispersals are also achieved by females (Tufto et al. 2005). It could be interesting to inspect mitochondrial DNA variations for the birds studied here and verify whether we do obtain the same pattern as in (Kvist et al. 2004). In any case, one should bear in mind that our results stem from integrating coalescent patterns observed across thousands of loci, therefore providing increased resolution to investigate the determinants of demographic divergence compared to approaches based on fewer mitochondrial and nuclear loci.

The contemporary demography inferred among the populations from Corsica and mainland France, especially the absence of gene flow since approximately 4500

generations and the large effective population size, likely established a suitable context for local adaptation to build up given. Moreover, the level of divergence ($F_{ST}$ = 0.08) was probably low enough to detect divergent outlier loci potentially implicated in local adaptation. We identified outlier SNPs as well as outlier 200 kb windows showing elevated differentiation, primarily in regions of lower recombination. $F_{ST}$ was also in average higher in these regions. These patterns of increased divergence in regions with low recombination are commonly observed (Nachman & Payseur 2011; Cruickshank & Hahn 2014; Gagnaire et al. 2018), including in other bird species (Burri et al. 2015; Spurgin et al. 2019). It is well documented that such increased differentiation in regions with low recombination is not necessarily due to positive selection, or at least not alone, and that it is largely influenced by the effect of recombination in interaction with background selection (Charlesworth et al. 1993; Burri et al. 2015; Perrier & Charmantier 2019). Moreover, a uniform $F_{ST}$ sliding window, sized in Kb, is expected to dilute signatures of selection in regions of the genome where LD is decaying fast while it is expected to over-represent regions where LD decays much more slowly, hence increasing even more the potential false positives in regions with low recombination (Beissinger *et al.* 2015; Perrier & Charmantier 2019). Although perfectible, our sliding window approach using a SNP unit attempted fixing this issue and successfully captured several new outlier regions outside of deserts of recombination. This second window approach should however be improved for example by estimating the local neutral enveloped, whose computation should integrate local variations in LD, recombination, diversity, but also the demography of populations. As mentioned for the study of

adaptive divergence between deciduous and evergreen populations, complementary analyses integrating both genome scans and quantitative genomics would improve our comprehension of the genomic and phenotypic divergence observed between Corsican and continentla blue tits (Stinchcombe & Hoekstra 2007; Gagnaire & Gaggiotti 2016).

Lastly, the genomic inversion detected on chromosome 3 on the continent was not present in Corsica, which could suggest that it was younger than the last period of gene flow between both populations. The age of this inversion would need to be inferred (Lohse *et al.* 2015). However, the observed pattern could also suggest that i) this inversion had a selective disadvantage in Corsica or ii) that is was lost by drift or iii) that it arrived in South of France via gene flow from another refugium, after the end of gene flow between Corsica and the continent. To investigate this, this inversion could be genotyped in several other areas, including islands and several refugium.

**Conclusion and perspectives**

Our study demonstrated the usefulness of demographic modeling and the analysis of the variation of genomic diversity and recombination along the genome for better investigating the genetic determinants of local adaptation in blue tit. Especially, demographic modeling rejected the hypothesis of a secondary contact between deciduous and evergreen populations and favoured a situation with continuous gene flow (Bierne *et al.* 2011; 2013). This also helped contextualising how large gene flow much probably constrained local adaptation (Lenormand 2002) and favoured its architecture based on alleles of small effects (Yeaman 2015). Besides, modelling also refined our knowledge about the divergence between insular and continental

populations, between which gene flow likely stopped since ten thousands years. We also verified the relationship between local recombination rate and differentiation, that should probably be integrated in genome scans looking for footprints of selection (Beissinger *et al.* 2015; Burri *et al.* 2015; Berner & Roesti 2017; Perrier & Charmantier 2019). However, it will probably be important to increase both sample size and marker density (Lotterhos & Whitlock 2015; Hoban *et al.* 2016) in order to better detect loci with small effects that most likely contributed to quantitative phenotypic variation and local adaptation in blue tits. Lastly, the putative inversion found here would necessitate further analyses since this type of structural variation is often implicated in phenotypic variation (Kirkpatrick 2006; 2010; Stapley *et al.* 2017; Wellenreuther *et al.* 2019).

## Tables

**Table 1. Study populations' names, abbreviations, habitat types, locations, and sample size.**

| Name | Abbreviation | Region | Site | Type of oaks | Latitude | Longitude | Altitude | N-GeneralDataset | N-FilteredDataset |
|---|---|---|---|---|---|---|---|---|---|
| D-Rouviere | A1D | Mainland | Rouviere | Deciduous | 43.664000 | 3.668000 | 248 | 203 | 60 |
| D-Quissac | A2D | Mainland | Quissac | Deciduous | 43.891036 | 3.996931 | 104 | 35 | 27 |
| E-Quissac | A2E | Mainland | Quissac | Evergreen | 43.926673 | 3.989759 | 116 | 31 | 24 |
| E-Pirio | B3E | Corsica | Pirio | Evergreen | 42.376000 | 8.750000 | 145 | 180 | 60 |
| D-Muro | B4D | Corsica | Muro | Deciduous | 42.551000 | 8.923000 | 283 | 215 | 60 |
| E-Muro | B4E | Corsica | Muro | Evergreen | 42.589000 | 8.967000 | 99 | 116 | 60 |
| D-Zilia | B5D | Corsica | Zilia | Deciduous | 42.523685 | 8.873689 | 102 | 42 | 38 |
| E-Zilia | B5E | Corsica | Zilia | Evergreen | 42.520633 | 8.884057 | 267 | 38 | 38 |
| D-OlmiCappella | B6D | Corsica | OlmiCappella | Deciduous | 42.540078 | 8.993915 | 1027 | 44 | 44 |
| E-OlmiCappella | B6E | Corsica | OlmiCappella | Evergreen | 42.529414 | 9.021652 | 881 | 43 | 43 |
| Total/average | | | | | | | | 947 | 454 |

**Table 2. Diversity and differentiation across populations.** $F_{ST}$, effective population size (Ne) estimated with ABC (Ne1), using models for Deciduous-Evergreen population pairs or for Corsica-Continent, Ne estimated with a LD method (Ne2), expected heterozygosity (He) and nucleotide diversity ($\pi$). Color gradient indicates differentiation intensity (Darker red for higher differentiation and darker blue for lowest differentiation). Underlined values indicate deciduous and evergreen pairs.

| | $F_{ST}$ | | | | | | | | | | Ne1 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A1D | A2D | A2E | B3E | B4D | B4E | B5D | B5E | B6D | B6E | Deciduous-Evergreen | Corsica-Continent | Ne2 | He | $\pi$ |
| A1D | | 0.0045 | 0.0043 | 0.0812 | 0.0830 | 0.0853 | 0.0796 | 0.0804 | 0.0785 | 0.0784 | Na | 397,000 | 355 | 0.1712 | 0.0042 |
| A2D | 0.0045 | | 0.0029 | 0.0832 | 0.0849 | 0.0872 | 0.0799 | 0.0810 | 0.0793 | 0.0792 | 463,000 | na | 161 | 0.1673 | 0.0034 |
| A2E | 0.0043 | 0.0029 | | 0.0828 | 0.0846 | 0.0871 | 0.0796 | 0.0807 | 0.0790 | 0.0787 | 113,000 | na | 142 | 0.168 | 0.0034 |
| B3E | 0.0812 | 0.0832 | 0.0828 | | 0.0058 | 0.0089 | 0.0067 | 0.0071 | 0.0056 | 0.0058 | Na | Na | 334 | 0.1488 | 0.0033 |
| B4D | 0.0830 | 0.0849 | 0.0846 | 0.0058 | | 0.0079 | 0.0063 | 0.0067 | 0.0061 | 0.0062 | 59,000 | 53,000 | 310 | 0.1489 | 0.0031 |
| B4E | 0.0853 | 0.0872 | 0.0871 | 0.0089 | 0.0079 | | 0.0093 | 0.0097 | 0.0083 | 0.0085 | 58,000 | na | 271 | 0.148 | 0.0031 |
| B5D | 0.0796 | 0.0799 | 0.0796 | 0.0067 | 0.0063 | 0.0093 | | 0.0025 | 0.0035 | 0.0032 | 119,000 | na | 210 | 0.1454 | 0.0036 |
| B5E | 0.0804 | 0.0810 | 0.0807 | 0.0071 | 0.0067 | 0.0097 | 0.0025 | | 0.0036 | 0.0033 | 79,000 | na | 216 | 0.1458 | 0.0036 |
| B6D | 0.0785 | 0.0793 | 0.0790 | 0.0056 | 0.0061 | 0.0083 | 0.0035 | 0.0035 | | 0.0006 | 180,000 | na | 268 | 0.1463 | 0.0035 |
| B6E | 0.0784 | 0.0792 | 0.0787 | 0.0058 | 0.0062 | 0.0085 | 0.0032 | 0.0033 | 0.0006 | | 43,000 | na | 260 | 0.1463 | 0.0034 |

## Figures captions

**Figure 1. A)** Map of the sampling locations of blue tits on the mainland and in Corsica. In red habitats dominated by deciduous oaks and in green habitats dominated by evergreen oaks. Sites with long-term monitorings are circled in black (A1D is D-Rouvière, B3E is D-Pirio, D4D is D-Muro and B4E is E-Muro). On the lower panel, illustrative phenotypic differences (with standard deviations) among the four main populations with long term monitorings in **B)** life history (laying date and clutch size) and **C)** male tarsus length (data from Table 1 in Charmantier et al 2016 Evol Appl, see this reference for detailed information on differences in many other traits, on data sampling and on sample sizes).

**Figure 2. Population structure among populations.** A) Admixture for K = 2, 6 and 10. RDA showing the influence of latitude, longitude, altitude and phenology on genetic structure.

**Figure 3. Demographic parameters for each of the four pairs of blue tits populations in deciduous and evergreen habitats, estimated using equilibrium models (A) and for the historical demography of continent and Corsica, estimated using an ancestral migration model (B).** In panel A, circle size is proportional to effective population size (Ne) and arrow width is proportional to migration rate. In panel B, rectangle width is proportional to log10 (effective population size), arrow width is proportional to migration rate, and split time and time of last migration are indicated in number of generations.

**Figure 4. $F_{ST}$ Manhattan plots between the 4 population pairs in deciduous and evergreen forests.** $F_{ST}$ are given either by SNP, by 200kb windows or by 50 SNPs windows. For graphic simplicity, the 3 pairs of populations in Corsica were grouped for the sliding window $F_{ST}$ Manhattan plots (BD *vs* BE).

**Figure 5. $F_{ST}$ Manhattan plots between Corsica and the mainland.** $F_{ST}$ are given either by SNP, by 200kb windows or by 50 SNPs windows.

**Figure 6. Correlations between local recombination rate across the genome and A) and genetic differentiation and B) and outlier frequency.**

**Figure 7. Putative genetic inversion on chromosome 3.** A) Recombination rate variation and SNP density; B) individual values for PCA eignevector 1; C) MDS 1 results from Lostruct; D) $F_{ST}$ between the three groups of individuals detected in mainland populations relatively to the putative inversion, Nucleotide diversity for the three groups of individuals detected in mainland populations relatively to the putative inversion; F) PCA results for the putative inversion; G) Individual admixture proportions for the putative inversion; H) Linkage disequilibrium for SNPs found in the putative inversion (plus 1 Mb each side of the inversion), for all of the individuals from the mainland populations.

**Acknowledgments**

We whish to thank the many people who helped on the blue tit long-term project. We also wish to thank Philippe Perret, Samuel Perret, Christophe De Franceschi, Jacques Blondel, Marta Szulkin, Monica Arias, Patricia Sourouille, Annick Lucas & Boris Delahaie. This project was funded by the European Research Council (Starting grant ERC-2013-StG-337365-SHE to AC) and the OSU-OREME.

**Authors' contributions:**

CP conceived the study, performed field-work, lab analyses, bioinformatics analyses, statistical analyses, interpreted the data and wrote the manuscript. QR performed bioinformatics analyses, statistical analyses, interpreted the data and edited the manuscript. AC got the grant, conceived the study, performed field-work, interpreted the data and edited the manuscript.

## References


Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Research, 19, 1655–1664.

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Available: http://www.bioinformatics.babraham.ac.uk.

Baird NA, Etter PD, Atwood TS et al. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers (JC Fay, Ed,). PlosOne 3, e3376–7.

Barbato M, Orozco-terWengel P, Tapio M, Bruford MW (2015) SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. Frontiers in Genetics, 6, 1294–7.

Beissinger TM, Rosa GJM, Kaeppler SM, Gianola D, De Leon N (2015) Defining window-boundaries for genomic analyses using smoothing spline techniques. Genetics Selection Evolution, 47, 30–9.

Berner D, Roesti M (2017) Genomics of adaptive divergence with chromosome-scale heterogeneity in crossover rate. Molecular Ecology, 114, 7061–19.

Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it…? why are FST outliers sometimes so frequent? Molecular Ecology, 22, 2061–2064.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20, 2044–2072.

Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation (D Ebert, Ed,). Ecology Letters, 16, 1195–1205.

Blondel J, Maistre M, Perret P, Hurtrez-Boussès S, Lambrechts M (1998) Is the small clutch size of a Corsican blue tit population optimal? Oecologia, 117, 80–89.

Blondel J, Thomas DW, Charmantier A et al. (2006) A Thirty-Year Study of Phenotypic and Genetic Variation of Blue Tits in Mediterranean Habitat Mosaics. Bioscience, 56, 661–673.

Bosse M, Spurgin LG, Laine VN et al. (2017) Recent natural selection causes adaptive evolution of an avian polygenic trait. Science (New York, N.Y.), 358, 365–368.

Broggi J, Hohtola E, Orell M, Nilsson J-Å (2005) Local adaptation to winter conditions in a passerine spreading north: a common‐garden approach. Evolution, 59, 1600–1603.

Burri R, Nater A, Kawakami T et al. (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. Genome Research, 25, 1656–1665.

Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky-Muller incompatibilities? Molecular Ecology, 21, 4942–4957.

Carbonell R, Perez-Tris J, Telleria JL (2003) Effects of habitat heterogeneity and local adaptation on the body condition of a forest passerine at the edge of its distributional range. Biological journal of the linnean society, 479–488.

Caro SP, Charmantier A, Lambrechts M et al. (2009) Local adaptation of timing of reproduction: females are in the driver's seat. Functional Ecology, 23, 172–179.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. Molecular Ecology, 22, 3124–3140.

Cayuela H, Rougemont Q, Prunier JG et al. (2018) Demographic and genetic approaches to study dispersal in wild animal populations: A methodological review. Molecular Ecology, 27, 3976–4010.

Charlesworth B, Morgan MT, Charlesworth D (1993) The Effect of Deleterious Mutations on Neutral Molecular Variation. Genetics, 134, 1289–1303.

Charmantier A, Doutrelant C, Dubuc-Messier G, Fargevieille A, SZULKIN M (2016) Mediterranean blue tits as a case study of local adaptation. Evolutionary Applications 9, 135–152.

Charmantier A, Kruuk LEB, Blondel J, Lambrechts MM (2004) Testing for microevolution in body size in three blue tit populations. Journal of Evolutionary Biology, 17, 732–743.

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23, 3133–3157.

Csilléry K, Blum MGB, Gaggiotti OE, François O (2010) Approximate Bayesian Computation (ABC) in practice. Trends in Ecology & Evolution, 25, 410–418.

Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. Nature Reviews Genetics, 14, 262–274.

da Silva VH, Laine VN, Bosse M et al. (2019) The Genomic Complexity of a Large Inversion in Great Tits. (J Mank, Ed,). Genome Biology and Evolution, 11, 1870–1881.

Danecek P, Auton A, Abecasis G et al. (2011) The variant call format and VCFtools. Bioinformatics, 27, 2156–2158.

de Villemereuil P, Frichot É, Bazin É, François O, Gaggiotti OE (2014) Genome scan methods against more complex models: when and how much should we trust them? Molecular Ecology, 23, 2006–2019.

Delahaie B, Charmantier A, Chantepie S et al. (2017) Conserved G-matrices of morphological and life-history traits among continental and island blue tit populations. Heredity ,119, 76-87.

Dubuc-Messier G, Caro SP, Perrier C et al. (2018) Gene flow does not prevent personality and morphological differentiation between two blue tit populations. Journal of Evolutionary Biology, 23, 960–11.

Excoffier L, Hofer T, Foll M (2009) Detecting loci under selection in a hierarchically structured population. Heredity, 103, 285–298.

Foll M, Gaggiotti O (2008) A Genome-Scan Method to Identify Selected Loci Appropriate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics, 180, 977–993.

Forester BR, Jones MR, Joost S, Landguth EL, Lasky JR (2015) Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular Ecology, 25, 104–120.

Gagnaire P-A, Gaggiotti OE (2016) Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. Current Zoology, 62, 603–616.

Gagnaire P-A, Lamy J-B, Cornette F et al. (2018) Analysis of Genome-Wide Differentiation between Native and Introduced Populations of the Cupped Oysters Crassostrea gigas and Crassostrea angulata. Genome Biology and Evolution, 10, 2518–2534.

Hansson B, Sigeman H, Stervander M et al. (2018) Contrasting results from GWAS and QTL mapping on wing length in great reed warblers. Molecular Ecology Resources, 18, 867–876.

Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary (KJ Willis, KD Bennett, D Walker, Eds,). Philosophical Transactions of the Royal Society B: Biological Sciences, 359, 183–195.

Hoban S, Kelley JL, Lotterhos KE et al. (2016) Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. The American Naturalist, 188, 379–397.

Hohenlohe PA, Bassham S, Etter PD et al. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. Plos Genetics, 6, e1000862–e1000862.

Illera J-C, KOIVULA K, Broggi J et al. (2011) A multi-gene approach reveals a complex evolutionary history in the Cyanistes species group. Molecular Ecology, 20, 4123–4139.

Jay P, Chouteau M, Whibley A et al. (2019) Mutation accumulation in chromosomal inversions maintains wing pattern polymorphism in a butterfly. bioRxiv, 10, 1.

Jouet G, Berné S, Rabineau M et al. (2006) Shoreface migrations at the shelf edge and sea-level changes around the Last Glacial Maximum (Gulf of Lions, NW Mediterranean). Marine Geology, 234, 21–42.

Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. Ecology Letters, 7, 1225–1241.

Kim K-W, Bennison C, Hemmings N et al. (2017) A sex-linked supergene controls sperm morphology and swimming speed in a songbird. Nature Ecology & Evolution, 1, 1168–1176.

Kingman JFC (1982) The coalescent. Stochastic Processes and their Applications, 13, 235–248.

Kirkpatrick M (2006) Chromosome Inversions, Local Adaptation and Speciation. Genetics, 173, 419–434.

Kirkpatrick M (2010) How and Why Chromosome Inversions Evolve. PLoS Biology, 8, e1000501–5.

Knief U, Forstmeier W, Pei Y et al. (2017) A sex-chromosome inversion causes strong overdominance for sperm traits that affect siring success. Nature Ecology & Evolution, 1–8.

Kvist L, Ruokonen M, Lumme J, Orell M (1999) The colonization history and present-day population structure of the european great tit (Parus major major). Heredity, 82 (Pt 5), 495–502.

Kvist L, Viiri K, Dias PC, Rytkönen S, Orell M (2004) Glacial history and colonization of Europe by the blue tit Parus caeruleus. Journal of Avian Biology, 35, 352–359.

Laaksonen T, Sirkiä PM, Calhim S et al. (2015) Sympatric divergence and clinal variation in multiple coloration traits of Ficedulaflycatchers. Journal of Evolutionary Biology, 28, 779–790.

Laine VN, Gossmann TI, Schachtschneider KM et al. (2016) Evolutionary signals of selection on cognition from the great tit genome and methylome. Nature Communications, 7, 1–9.

Lambeck K, Bard E (2000) Sea-level change along the French Mediterranean coast for the past 30000 years. Earth and Planetary Science Letters, 175, 203–222.

Lambrechts M, Blondel J, Maistre M, Perret P (1997) A single response mechanism is responsible for evolutionary adaptive variation in a bird's laying date. Proceedings of the National Academy of Sciences of the United States of America, 94, 5153–5155.

Lamichhaney S, Han F, Berglund J, Wang C (2016) A beak size locus in Darwin's finches facilitated character displacement during a drought. Science, 352, 470-474.

Legendre P, Fortin M-J (2010) Comparison of the Mantel test and alternative

approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. Molecular Ecology Resources, 10, 831–844.

Lenormand T (2002) Gene flow and the limits to natural selection. Trends in Ecology & Evolution, 17, 183–189.

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

Li H, Ralph P (2019) Local PCA Shows How the Effect of Population Structure Differs Along the Genome. Genetics, 211, 289–304.

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

Lohse K, Clarke M, Ritchie MG, Etges WJ (2015) Genome-wide tests for introgression between cactophilic Drosophila implicate a role of inversions during speciation. Evolution, 69, 1178–1190.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology, 24, 1031–1046.

Lowry DB, Willis JH (2010) A Widespread Chromosomal Inversion Polymorphism Contributes to a Major Life-History Transition, Local Adaptation, and Reproductive Isolation (NH Barton, Ed,). PLoS Biology, 8, e1000500–14.

Lowry DB, Hoban S, Kelley JL et al. (2017) Responsible RAD: Striving for best practices in population genomic studies of adaptation. Molecular Ecology Resources, 17, 366–369.

Lundregan SL, Hagen IJ, Gohli J et al. (2018) Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. Molecular Ecology, 27, 3498–3514.

Ma J, Amos CI (2012) Investigation of Inversion Polymorphisms in the Human Genome Using Principal Components Analysis (F Emmert-Streib, Ed,). PLOS ONE, 7, e40224–12.

Manel S, Perrier C, Pratlong M et al. (2016) Genomic resources and their influence on

the detection of the signal of positive selection in genome scans. Molecular Ecology, 25, 170–184.

Marko PB, Hart MW (2011) The complex analytical landscape of gene flow inference. Trends in Ecology & Evolution, 26, 448–456.

McGinnis S, Madden TL (2004) BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Research, 32, W20–W25.

McVean GAT (2004) The Fine-Scale Structure of Recombination Rate Variation in the Human Genome. Science, 304, 581–584.

Nachman MW, Payseur BA (2011) Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. Philosophical Transactions of the Royal Society B: Biological Sciences, 367, 409–421.

Narum SR, Hess JE (2011) Comparison of FST outlier tests for SNP loci under selection. Molecular Ecology Resources, 11, 184–194.

Oksanen J, Kindt R, Legendre P et al. (2007) The vegan package. Community ecology package, 10, 631–637.

Ortego J, García-Navas V, Ferrer ES, Sanz JJ (2011) Genetic structure reflects natal dispersal movements at different spatial scales in the blue tit, Cyanistes caeruleus. Animal Behaviour, 82, 131–137.

Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Molecular Ecology Resources, 13, 946–952.

Perrier C, Charmantier A (2019) On the importance of time scales when studying adaptive evolution. Evolution Letters 3, 240-247.

Perrier C, Delahaie B, Charmantier A (2018) Heritability estimates from genomewide relatedness matrices in wild populations: Application to a passerine, using a small sample size. Molecular Ecology Resources, 3, e3376–16.

Perrier C, Ferchaud A-L, Sirois P, Thibault I, Bernatchez L (2017) Do genetic drift and accumulation of deleterious mutations preclude adaptation? Empirical investigation using RADseq in a northern lacustrine fish. Molecular Ecology, 26, 6317–6335.

Pinho C, Hey J (2010) Divergence with Gene Flow: Models and Data. Annual Review of Ecology Evolution and Systematics, 41, 215–230.

Porlier M, Charmantier A, Bourgault P et al. (2012a) Variation in phenotypic plasticity and selection patterns in blue tit breeding time: between- and within-population comparisons. Journal of Animal Ecology, 81, 1041–1051.

Porlier M, Garant D, Perret P, Charmantier A (2012b) Habitat-Linked Population Genetic Differentiation in the Blue Tit Cyanistes caeruleus. Journal of Heredity, 103, 781–791.

Purcell S, Neale B, Todd-Brown K et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics, 81, 559–575.

Reid NM, Proestou DA, Clark BW et al. (2016) The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science, 354, 1305–1308.

Rockman MV (2012) The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution, 66, 1–17.

Rougemont Q, Bernatchez L (2018) The demographic history of Atlantic Salmon ( Salmo salar) across its distribution range reconstructed from Approximate Bayesian Computations. Evolution, 1–3.

Rougemont Q, Gagnaire P-A, Perrier C et al. (2016) Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. Molecular Ecology, 26, 142–162.

Rougemont Q, Moore J-S, Leroy T et al. (2019) Demographic history, linked selection, and recombination shape the genomic landscape of a broadly distributed Pacific salmon. 123, 887–37.

Roux C, Fraïsse C, Romiguier J et al. (2016) Shedding Light on the Grey Zone of Speciation along a Continuum of Genomic Divergence (C Moritz, Ed,). PLoS Biology, 14, e2000234–23.

Santure AW, Cauwer I, Robinson MR et al. (2013) Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. Molecular

Ecology, 22, 3949–3962.

Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. Nature Publishing Group, 14, 807–820.

Seutin G, White BN, Boag PT (1991) Preservation of avian blood and tissue samples for DNA analyses. Canadian Journal of Zoology-Revue Canadienne De Zoologie, 69, 82–90.

Sinclair-Waters M, Bradbury IR, Morris CJ et al. (2017) Ancient chromosomal rearrangement associated with local adaptation of a post-glacially colonized population of Atlantic Cod in the northwest Atlantic. Molecular Ecology, 27, 339-351.

Spurgin LG, Bosse M, Adriaensen F et al. (2019) The great tit HapMap project: a continental-scale analysis of genomic variation in a songbird. bioRxiv, 1–24.

Stapley J, Feulner PGD, Johnston SE, Santure AW, Smadja CM (2017) Variation in recombination frequency and distribution across eukaryotes: patterns and processes. Philosophical Transactions of the Royal Society B: Biological Sciences, 372, 20160455–18.

Stinchcombe JR, Hoekstra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity, 100, 158–170.

Szulkin M, Gagnaire PA, Bierne N, Charmantier A (2016) Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. Molecular Ecology, 25, 542–558.

Taberlet P, Meyer A, Bouvet J (1992) Unusual mitochondrial DNA polymorphism in two local populations of blue tit Parus caeruleus. Molecular Ecology, 1, 27–36.

Teplitsky C, Tarka M, Møller AP et al. (2014) Assessing Multivariate Constraints to Evolution across Ten Long-Term Avian Studies. 9, e90444–15.

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution, 29, 1–8.

Tine M, Kuhl H, Gagnaire P-A et al. (2014) European sea bass genome and its variation

provide insights into adaptation to euryhalinity and speciation. Nature Communications, 5, 1–10.

Todesco M, Owens GL, Bercovich N et al. (2019) Massive haplotypes underlie ecotypic differentiation in sunflowers. bioRxiv, 19, 337–66.

Tufto J, Ringsby TH, Dhondt AA, Adriaensen F, Matthysen E (2005) A parametric model for estimation of dispersal patterns applied to five passerine spatially structured populations. American Naturalist, 165, E13–26.

Tuttle EM, Bergland AO, Korody ML et al. (2016) Divergence and Functional Degradation of a Sex Chromosome-like Supergene. Current Biology, 1–8.

Wang IJ, Bradburd GS (2014) Isolation by environment. Molecular Ecology, 23, 5649–5662.

Waples RS, Do C (2008) ldne: a program for estimating effective population size from data on linkage disequilibrium. Molecular Ecology Resources, 8, 753–756.

Weir BS, Cockerham CC (1984) Estimating F-Statistics for the Analysis of Population Structure. Evolution, 38, 1–14.

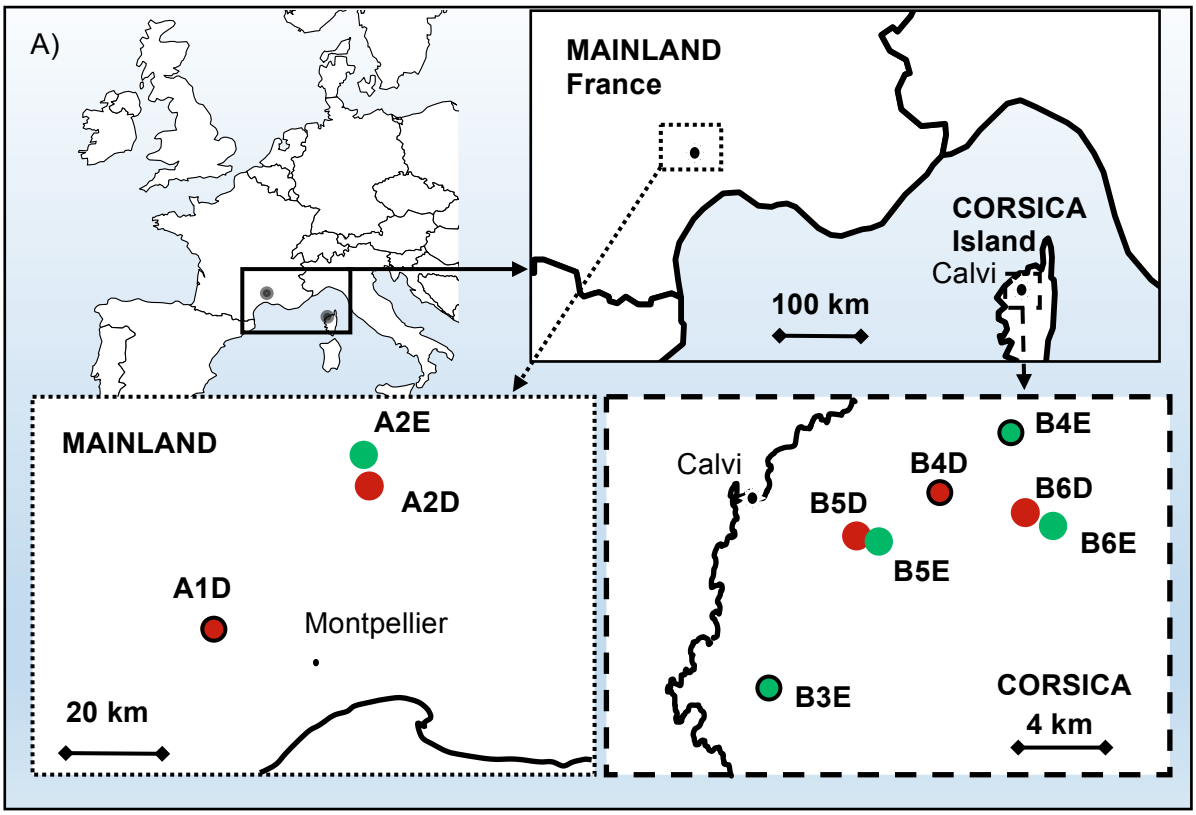
Wellenreuther M, Hansson B (2016) Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. Trends in Genetics, 32, 155–164.

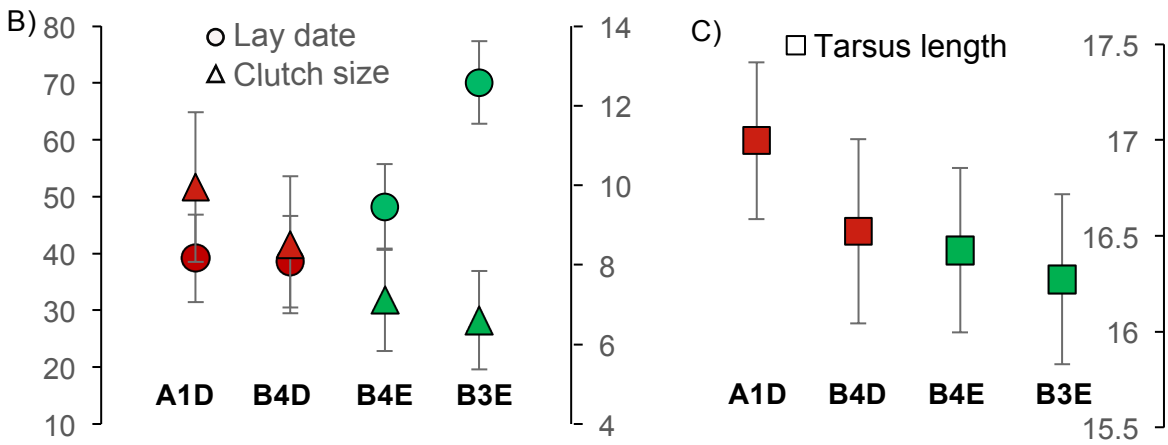
Wellenreuther M, Merot C, Berdan E, Bernatchez L (2019) Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Molecular Ecology, 28, 1203–1209.

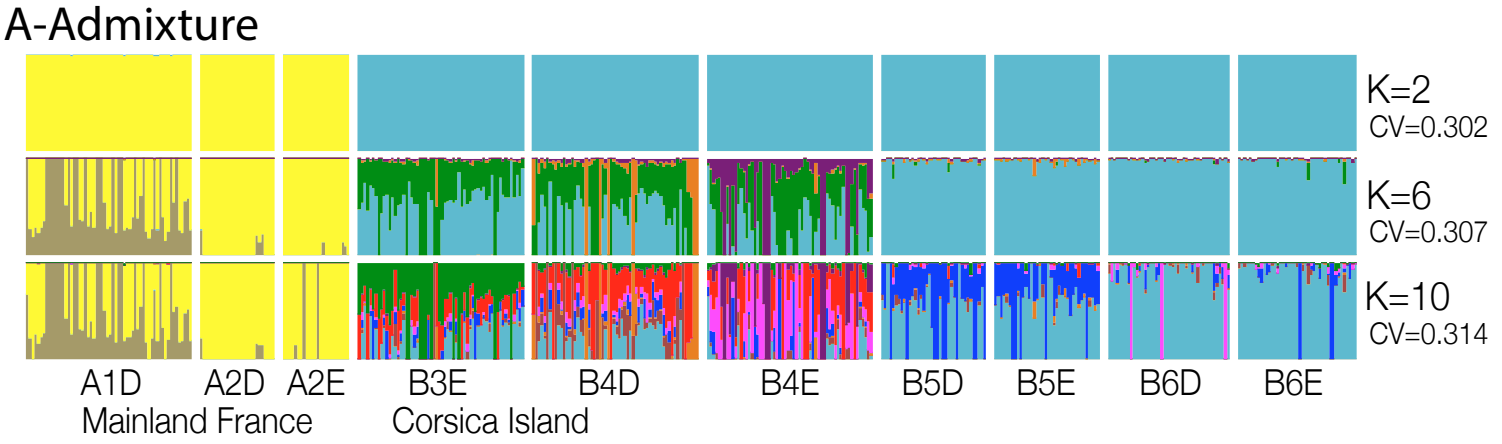
Yeaman S (2015) Local Adaptation by Alleles of Small Effect. American Naturalist, 186 Suppl 1, S74–89.

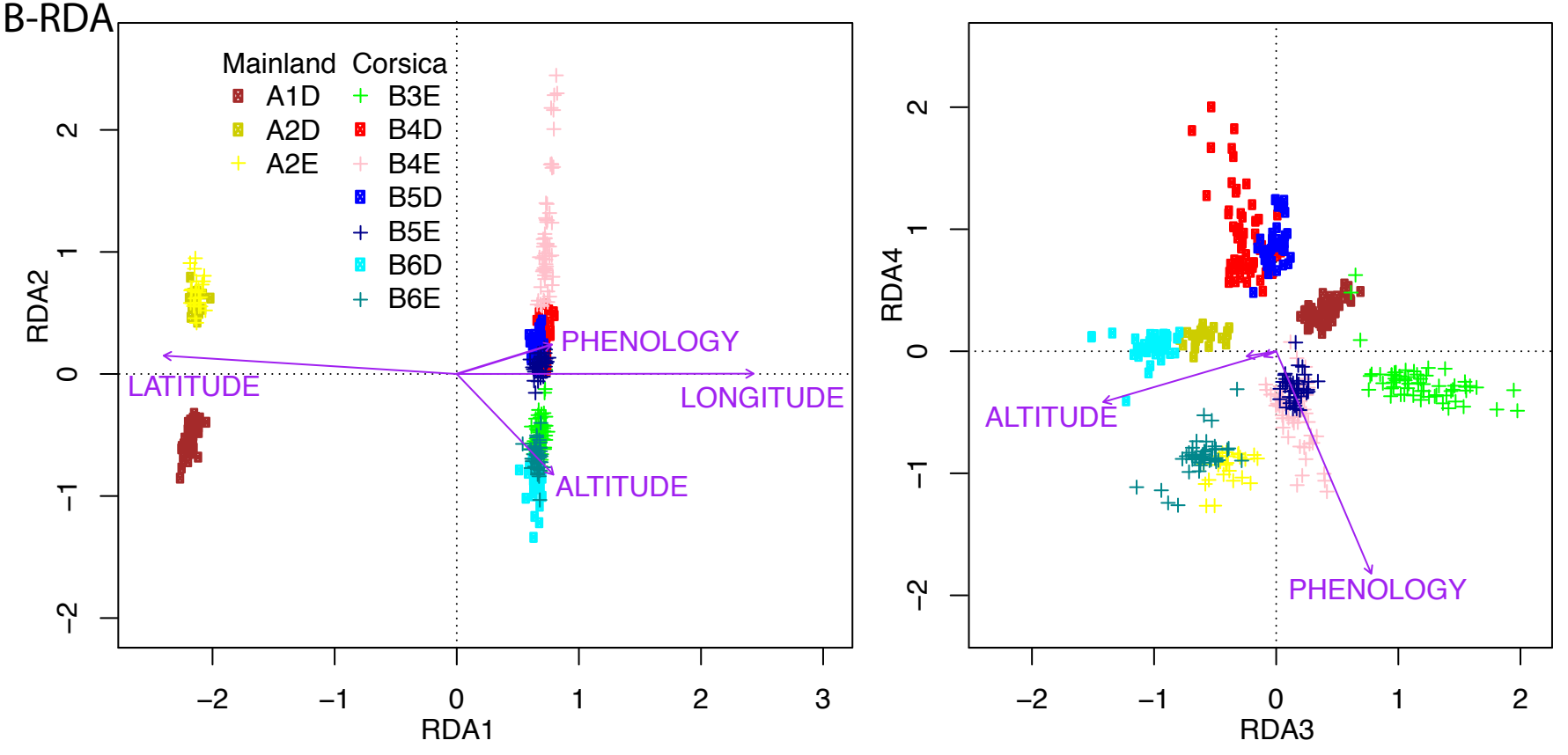
Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. Evolution, 65, 1897–1911.

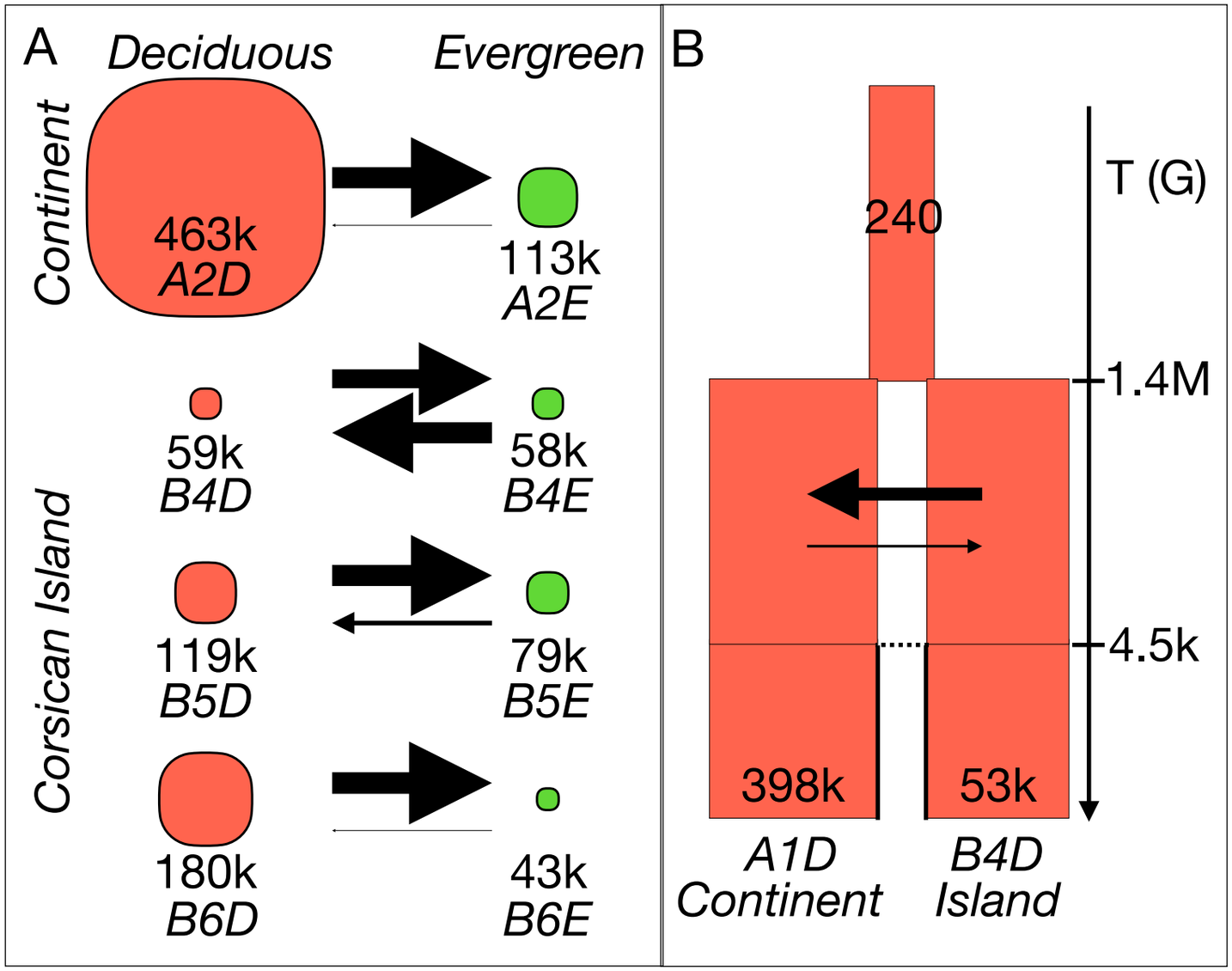
Zheng X, Levine D, Shen J et al. (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics, 28, 3326–3328.

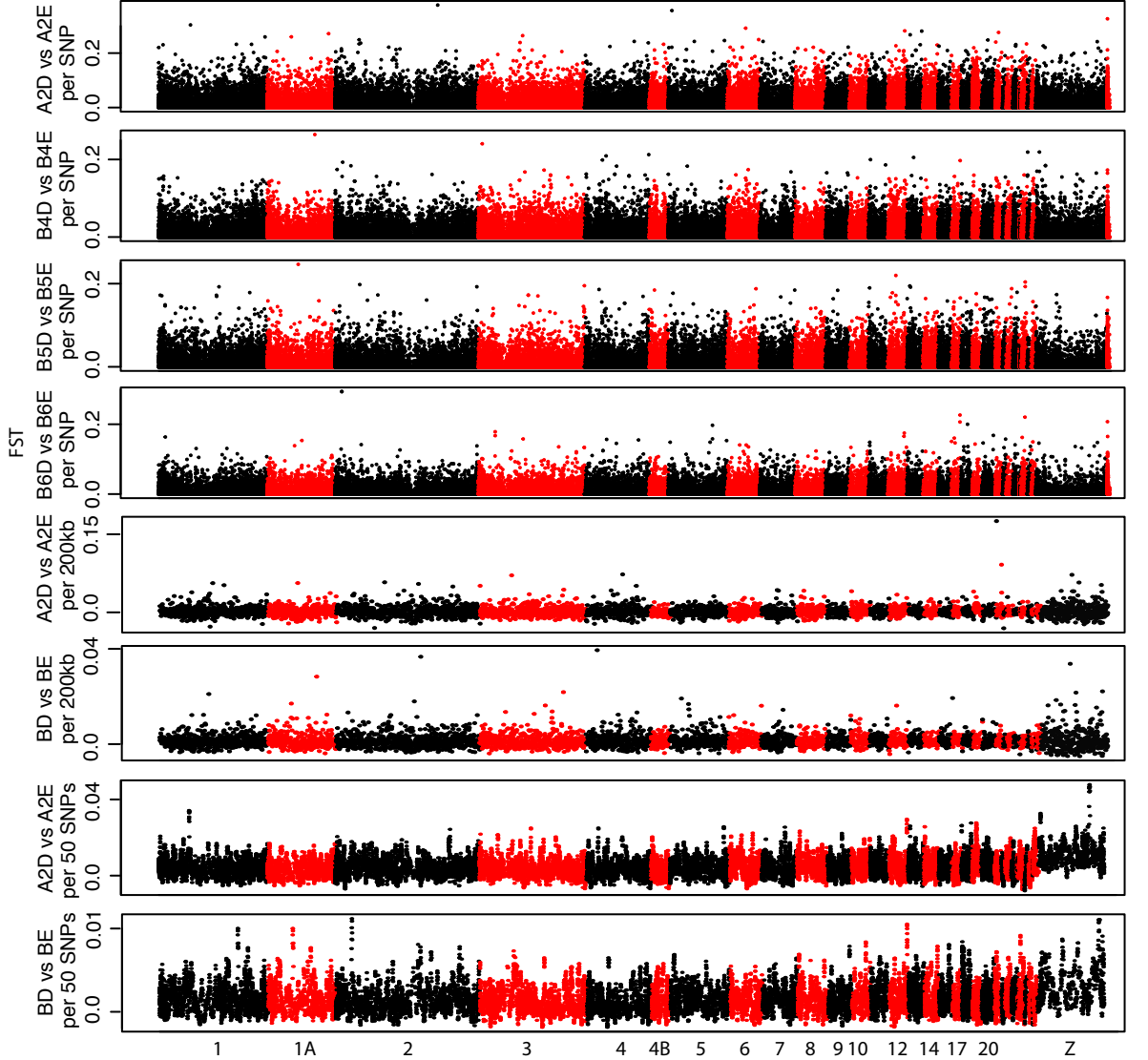

A)
MAINLAND
France
CORSICA
Island
Calvi
100 km
MAINLAND
A2E
A2D
A1D
Montpellier
20 km
Calvi
B4E
B4D
B6D
B5D
B6E
B5E
B3E
CORSICA
4 km

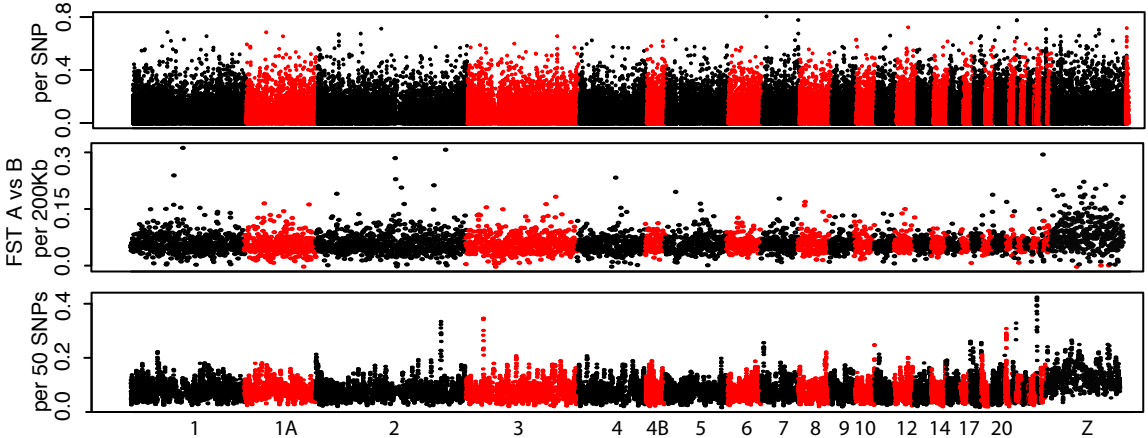

B)
Lay date
Clutch size
80
70
60
50
40
30
20
10
14
12
10
8
6
4
A1D
B4D
B4E
B3E
C)
Tarsus length
17.5
17
16.5
16
15.5
A1D
B4D
B4E
B3E

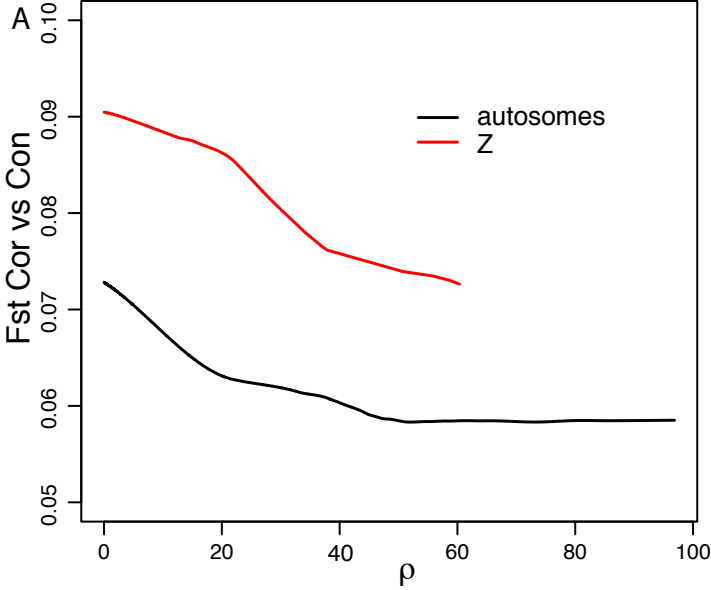

A-Admixture
K=2
CV=0.302
K=6
CV=0.307
K=10
CV=0.314
A1D
A2D
A2E
B3E
B4D
B4E
B5D
B5E
B6D
B6E
Mainland France
Corsica Island

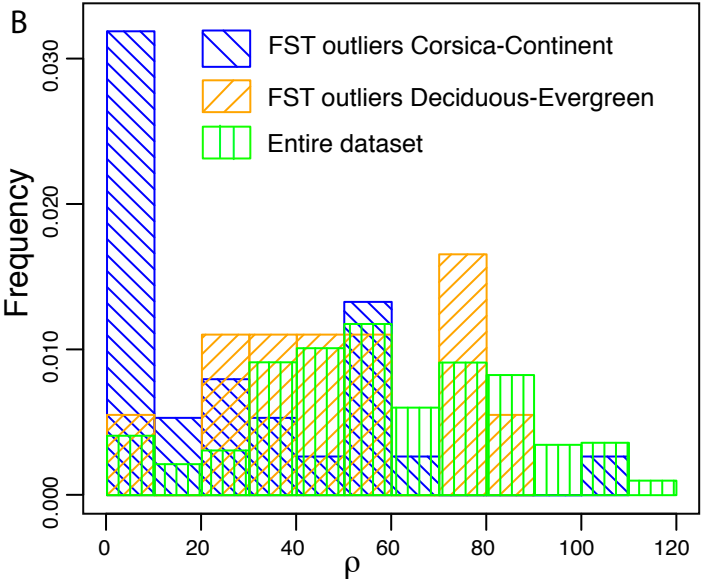

B-RDA
Mainland
A1D
A2D
A2E
Corsica
B3E
B4D
B4E
B5D
B5E
B6D
B6E
PHENOLOGY
LATITUDE
LONGITUDE
ALTITUDE
RDA1
RDA2
RDA3
RDA4
ALTITUDE
PHENOLOGY

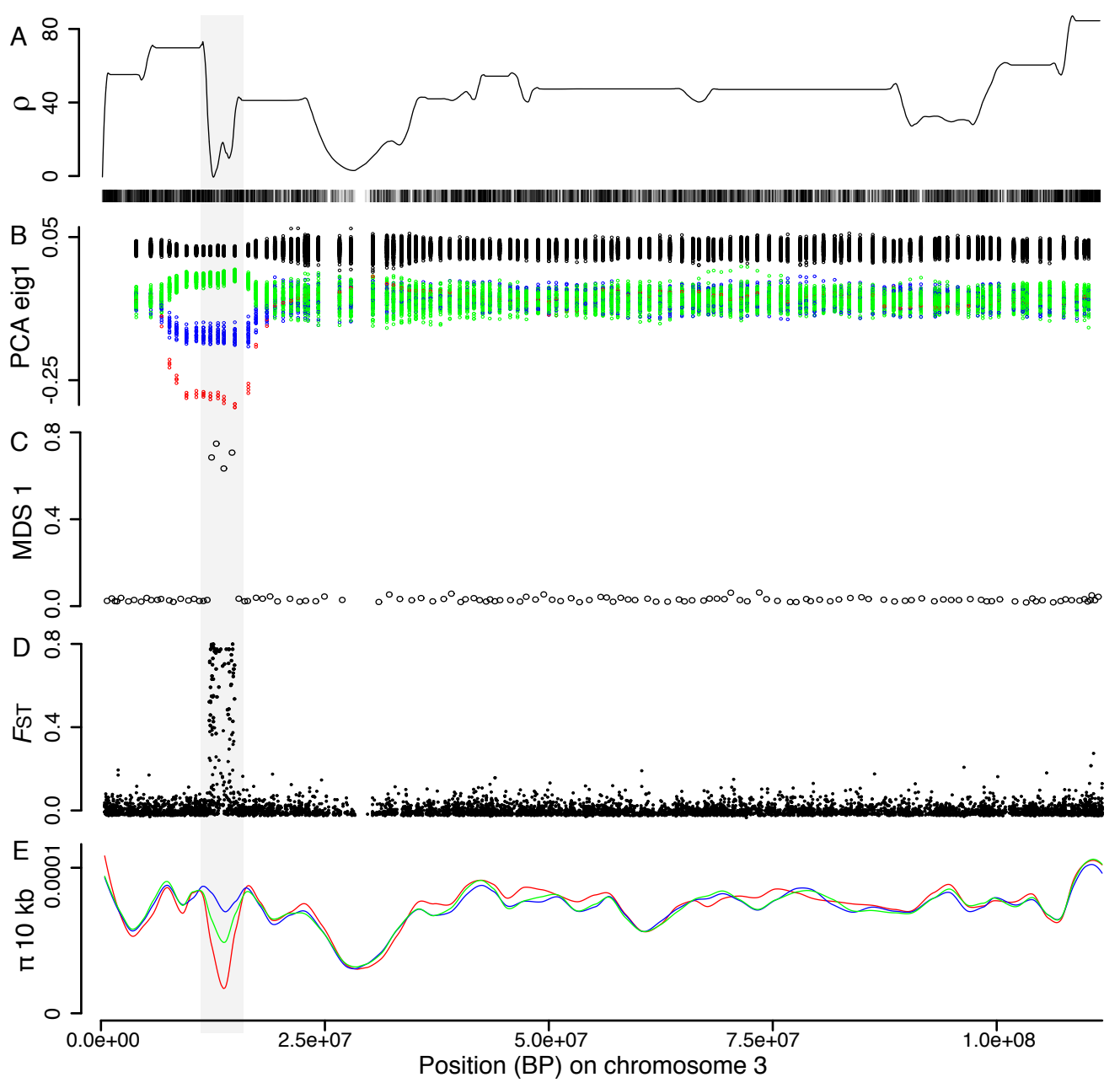

A
Deciduous
Evergreen
Continent
463k
A2D
113k
A2E
59k
B4D
58k
B4E
Corsican Island
119k
B5D
79k
B5E
180k
B6D
43k
B6E
B
T (G)
240
1.4M
4.5k
398k
53k
A1D
Continent
B4D
Island

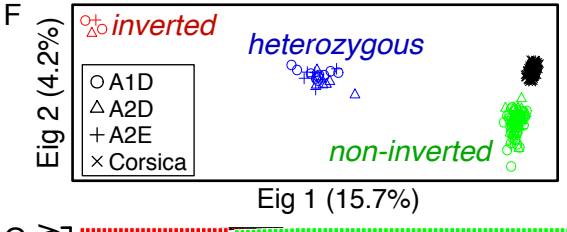

FST
A2D vs A2E per SNP
B4D vs B4E per SNP
B5D vs B5E per SNP
B6D vs B6E per SNP
A2D vs A2E per 200kb
BD vs BE per 200kb
A2D vs A2E per 50 SNPs
BD vs BE per 50 SNPs
0.0
0.2
0.15
0.04
0.01
1
1A
2
3
4
4B
5
6
7
8
9
10
12
14
17
20
Z

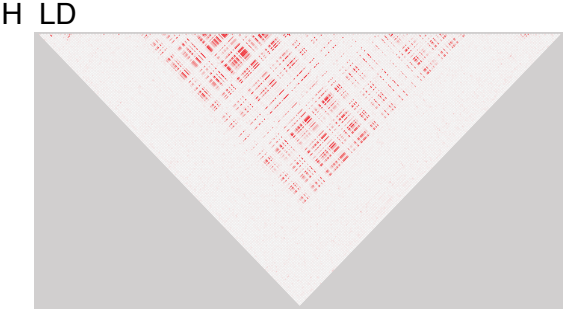

FST A vs B
per SNP
0.0
0.4
0.8
per 200Kb
0.0
0.15
0.3
per 50 SNPs
0.0
0.2
0.4
1
1A
2
3
4
4B
5
6
7
8
9
10
12
14
17
20
Z

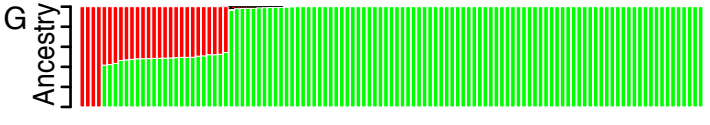

A
Fst Cor vs Con
0.05
0.06
0.07
0.08
0.09
0.10
autosomes
Z
0
20
40
60
80
100
ρ


B
Frequency
0.000
0.010
0.020
0.030
FST outliers Corsica-Continent
FST outliers Deciduous-Evergreen
Entire dataset
0
20
40
60
80
100
120
ρ


A
ρ
80
40
0
B
PCA eig1
0.05
-0.25
C
MDS 1
0.8
0.4
0.0
D
FST
0.8
0.4
0.0
E
π 10 kb
0.0001
0
0.0e+00
2.5e+07
5.0e+07
7.5e+07
1.0e+08
Position (BP) on chromosome 3


F
inverted
heterozygous
non-inverted
A1D
A2D
A2E
Corsica
Eig 2 (4.2%)
Eig 1 (15.7%)


H LD


G
Ancestry