Population genomics of incipient allochronic divergence in the Pine Processionary Moth

Running Title: Genomics of allochronic divergence

Tanguy Muller[1,*], Mathieu Gautier[1], Éric Lombaert[2], Raphaël Leblois[1], Laure Sauné[1], Manuela Branco[3], Carole Kerdelhué[1], Charles Perrier[1,*]

[1] UMR CBGP, INRAE, CIRAD, IRD, Institut Agro, Univ. Montpellier, Montpellier, France

[2] UMR ISA, INRAE, Université Côte d'Azur, Sophia-Antipolis, France

[3] Forest Research Centre (CEF), School of Agriculture (ISA), University of Lisbon, Lisboa, Portugal

*Corresponding authors: tanguy.muller@inrae.fr and charles.perrier@inrae.fr

Abstract

Allochronic divergence is a key evolutionary mechanism that can frequently lead to incipient speciation. Although theoretical models suggest that such divergence is notably facilitated by small population size and genetic polymorphisms influencing reproductive timing, though constrained by genetic load, empirical validation remains limited. We investigated these predictions by re-analyzing a case of allochronic differentiation between two sympatric populations of pine processionary moth (*Thaumetopoea pityocampa*) in Portugal, using whole genome resequencing (IndSeq and PoolSeq) of those two populations and eight allopatric ones. We inferred the demographic history of those populations, assessed their genetic load, and searched for genomic regions associated with life cycle differences. Our analyses revealed a recent split between the sympatric allochronic populations, accompanied by a strong reduction in gene flow, bottlenecks, inbreeding, and accumulation of deleterious variants. Genome scans identified several loci associated with life cycle variation, including genes putatively involved in circadian rhythm regulation, predominantly located on the Z chromosome. We discuss how these empirical genomic findings support theoretical expectations that assortative mating driven by differences in reproductive timing, underpinned by polymorphisms in circadian genes, along with genetic drift and purge of genetic load at high-impact sites, can promote the onset and persistence of allochronic divergence.

Keywords

allochrony, demographic history, inbreeding, genetic load, circadian genes

Introduction

Differences in reproductive timing among individuals within a population can lead to the formation of reproductively isolated clusters that gradually accumulate genetic differences over time, a process called allochronic divergence, which may ultimately promote speciation (Coyne and Orr, 2004; Hendry and Day, 2005; Rundle and Nosil, 2005; Taylor and Friesen, 2017). Theoretical models developed by Devaux and Lande (2008) and Weis *et al.* (2014) suggest that, in sympatry, the formation of allochronic clusters of individuals may be promoted notably by genetic drift, the existence of genetic polymorphisms influencing reproductive time, positive assortative mating, and a short individual reproductive window within a longer reproductive season. Isolation by distance between allochronic clusters may also contribute to further reinforce allochronic divergence (Kunkel, 2023). In allopatry, in contrast, spatially divergent selection may be an important driver of allochronic divergence (Jarrett *et al.*, 2025). The long-term viability of newly formed allochronic populations could be facilitated by adaptations to the new life cycle (Jarrett *et al.*, 2025). In turn, it could be limited by inbreeding depression (Devaux and Lande, 2008), which can be exacerbated by the small population sizes often expected in such contexts. Although empirical investigations revealed the important role of genetic mutations at circadian genes underlying allochrony (Tessnow *et al.*, 2022), thorough empirical analyses are needed to evaluate the respective roles of the several evolutionary processes promoting the onset of allochronic divergence.

Divergence in the timing of reproductive activity can stem from variation in genes regulating biological rhythms (Tomioka and Matsumoto, 2015; Hutfilz, 2022). In insects, circadian genes such as *period*, *pdfr*, *CYC*, *cry*, or *timeless* play key roles in controlling diapause duration and termination, thereby ensuring synchronization of reproductive activity with seasonal cycles (Abruzzi *et al.*, 2017; Kozak *et al.*, 2019; Brady *et al.*, 2021). In parallel, genes involved in secondary sexual traits, pheromone production, olfaction, and assortative mating may further contribute to reinforcing reproductive isolation (Matessi *et al.*, 2002; Unbehend *et al.*, 2013; Kopp *et al.*, 2018). Several genome-wide scans comparing allochronic populations have highlighted the involvement of polymorphisms in circadian genes in driving phenological divergence. For instance, Pruisscher *et al.* (2021), using whole-genome resequencing, identified non-synonymous SNPs differentially fixed in the *period* gene between two populations of *Pieris napi* differing in their diapause behavior. However, although such genetic polymorphisms at circadian genes may constitute the raw material for phenological differences, assortative mating and allochronic divergence, it may be difficult to decipher the relative roles of genetic drift, selection resulting from assortative mating, and spatially varying selection, as evolutionary forces driving genetic differentiation at these loci.

Inferring demographic histories is required to determine whether allochronic divergence between two populations emerged in sympatry *via* local strong genetic drift together with non-selective assortative mating based on genetically controlled phenology (Devaux and Lande, 2008; Kunkel,

2023) or was initially produced in allopatry via spatially divergent selection on genetic polymorphisms controlling phenology (Safran *et al.*, 2016). Recent advances in statistical population genomics, such as those using simulation-based inference (e.g., Approximate Bayesian Computation or ABC, Raynal *et al.*, 2019) allow the inference of demographic events, including temporal fluctuations in population size and gene flow. In addition, methods have recently been developed to measure the fraction of the genome within regions of autozygosity, as well as to track the temporal dynamics of accumulation of these regions within individual genomes (Druet and Gautier, 2017). These complementary approaches may help to reveal the putative role of demography and assortative mating in the onset of allochronic divergence.

While the particular context of small population size may promote allochronic divergence, it is also predicted to decrease genetic diversity and increase genetic load. Indeed, in general, reduced effective population size associated with bottlenecks and inbreeding can lead to the accumulation of genetic load, as purifying selection becomes less efficient in removing deleterious variants (Charlesworth, 2009; Woolfit, 2009; Lanfear *et al.*, 2014). However, prolonged periods of small effective population size can also facilitate the purging of highly deleterious mutations, potentially mitigating some of the negative fitness consequences (Robinson *et al.*, 2018, 2023). Hence, one could hypothesize that successfully established allochronic populations may have gone through strong bottlenecks and significant purging of the genetic load. Empirically examining whether genetic load is purged during allochronic divergence is crucial to assess the long-term viability and evolutionary trajectories of allochronically divergent populations.

The pine processionary moth (PPM, *Thaumetopoea pityocampa*, Lepidoptera, Notodontidae) provides a unique system to study allochronic divergence. Indeed, in the Mata Nacional de Leiria, Portugal, a recently discovered population (Pimentel *et al.*, 2006; Santos *et al.*, 2007, 2011) exhibits a clear shift in reproductive timing compared to the locally sympatric winter population (Figure 1 A,B). This so-called "Leiria summer population" (LSP) reproduces in spring, with summer larval development, whereas the sympatric "Leiria winter population" (LWP) follows the typical phenology of the species, with reproduction occurring in summer and larval development occurring in winter. Both populations hence experience strong pre-mating reproductive isolation (Santos *et al.*, 2011; Burban *et al.*, 2016). Experimental rearing suggested a high heritability of the reproductive period (Branco *et al.*, 2017). Previous population genetic studies based on mitochondrial, microsatellite, and RAD-seq data (Santos *et al.*, 2007, 2011; Burban *et al.*, 2016; Leblois *et al.*, 2018) showed low mitochondrial differentiation but high nuclear differentiation between LSP and LWP, suggesting strong bottlenecks and a recent split between these two populations. However, these studies were limited in their ability to elucidate the genomic mechanisms underlying the phenological shift, notably due to the lack of a chromosome-level genome assembly and the sparsity of genetic markers used.

In this study, we leverage recently developed genomic resources for the PPM, including a chromosome-level reference genome (Gautier *et al.*, 2025) and whole genome resequencing data for both individuals (IndSeq) and pooled samples (PoolSeq) from different populations, to investigate the genetic basis and consequences of the PPM allochronic divergence. Specifically, our goals were (i) to reconstruct the recent demographic history of LSP and LWP, (ii) to investigate the genetic bases of phenological variation between these two populations, and (iii) to characterize and compare their respective genetic load. Following the previously aforementioned theoretical works (Devaux and Lande, 2008) and empirical studies on the PPM (Burban *et al.*, 2016; Leblois *et al.*, 2018), we hypothesized that recent bottlenecks and non-selective assortative mating relaying on genomic variations in circadian genes might have driven the recent onset of allochronic divergence of LSP and that genetic load might have been partly purged. By integrating demographic inferences, genome scans, and analyses of deleterious variation, we provide new empirical insights into the evolutionary mechanisms driving allochronic divergence and its genetic consequences.

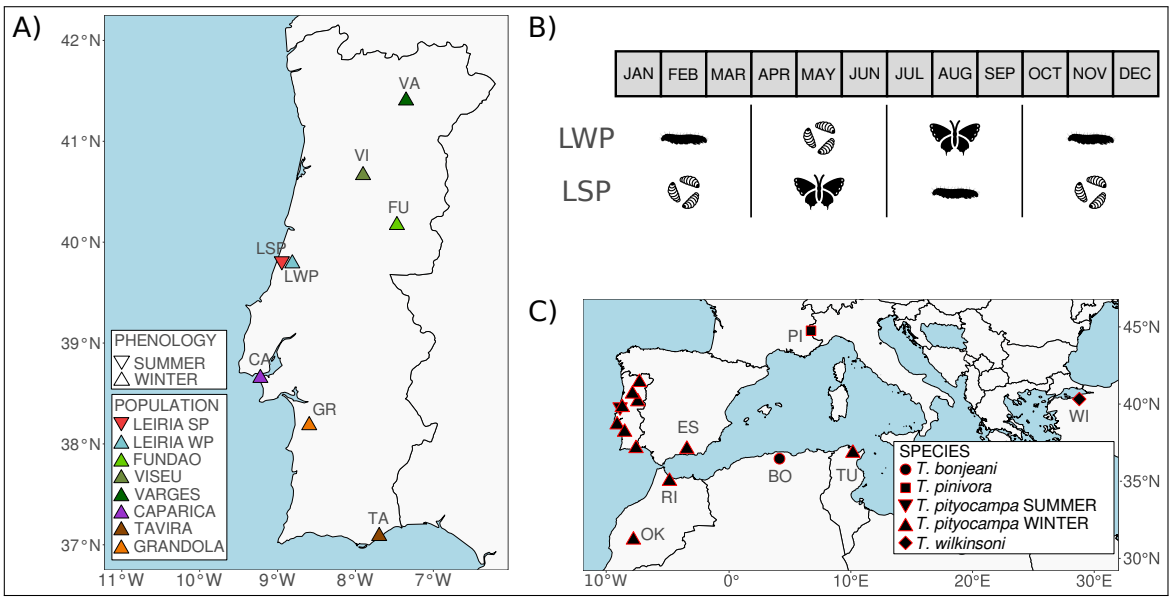

Figure 1: **Geographical locations of all sampled populations and life cycle of *T. pityocampa*. (A)** Map of the sampling locations for PPM populations over Portugal. Colors indicate the different Portuguese populations, and shapes their phenology. The population abbreviations stay the same throughout this study : LSP (Leiria Summer Population; n = 25), LWP (Leiria Winter Population; n = 18), FU (Fundão; n = 10), VI (Viseu; n = 4), VA (Varges; n = 4), CA (Caparica; n = 4), TA (Tavira; n = 4), GR (Grândola; n = 4). **(B)** Annual life cycle of the PPM, illustrating the LWP cycle (winter phenology) and the LSP cycle (summer phenology). Caterpillars represent the larval stage, pupae the pupal stage and moth the emergence period of adults. **(C)** Map of the sampling locations for all populations and species used in this study. Shapes indicate the sampled species. The population abbreviations stay the same throughout this study : ES (Sierra Nevada, Spain ; n = 2), RI (Rif, Morocco; n = 2), OK (Oukaimeden, Morocco; n = 2), TU (Tunis, Tunisia; n = 2), BO (*T. bonjeani*; n = 2), PI (*T. pinivora*; n = 2), WI (*T. wilkinsoni;* n = 2).

Methods

Sampling

A total of 369 individuals of PPM were collected between 2002 and 2012 in the context of previous studies (Burban *et al.*, 2016, 2020), in the summer (LSP) and winter (LWP) sympatric populations from the Leiria region (Figure 1A; Table 1; Table S1), as well as in more distant locations of Fundão (FU), Viseu (VI), Varges (VA), Caparica (CA), Tavira (TA), and Grândola (GR) in Portugal. For FU, CA, TA, and GR, adult individuals were captured using pheromone traps from mid-April to September. For VI and VA, larvae were collected from nests, with a single larva sampled per nest to avoid collecting siblings. For LSP and LWP, 12 and 6 larvae, and 13 and 12 adults, respectively, were collected following the same sampling strategy described above. All individuals were stored in 95% ethanol immediately after sampling and then kept at -18°C until DNA extraction. We also included 8 individuals referred to as "LateLSP", sampled alongside LWP individuals (i.e. with a winter phenology) but genetically (at microsatellite and SNP markers) clustering with LSP individuals, and hypothetically resulting from plastic variations of phenology in LSP (Burban *et al.*, 2020). We added 9 LSPxLWP hybrids ("F1") produced by Branco *et al.* (2017) in laboratory. Lastly, 14 individuals were included as outgroups (Figure 1C; Table 1, Table S1): 2 individuals of *T. pityocampa* from Cortijuela (ES, Spain), Bab Barred (RI, Rif region, Morocco), Oukaimeden (OK, Morocco) and Tunis (TU, Tunisia), 2 individuals of *T. bonjeani* from Tikjda (BO, Algeria), 2 individuals of *T. pinivora* from Alpes (PI, France) and 2 individuals of *T. wilkinsoni* from Mudanya (WI, Türkiye).

Table 1: **Study populations and resequencing strategy.** For each studied population is given a code, species name, country of origin, latitude, longitude, number of resequenced individuals (Ind-Seq) for both sexes (m for males and f for females), and number of individuals in resequenced pools. For pools, the ratio of the Z chromosome to the chromosome 1 median coverage was used to estimate sex ratio (see text for details).

Population code	Species	Country	Location	Latitude	Longitude	Ind-Seq n total (m;f)	Pool-Seq n total (m;f)
LSP	*T. pityocampa*	Portugal	Leiria	39.81	-8.96	25 (15;10)	50 (50;0)
LWP	*T. pityocampa*	Portugal	Leiria	39.78	-8.78	18 (12;6)	50 (50;0)
FU	*T. pityocampa*	Portugal	Fundão	40.15	-7.50	10 (10;0)	40 (40;0)
VI	*T. pityocampa*	Portugal	Viseu	40.66	-7.94	4 (1;3)	20 (13;7)
VA	*T. pityocampa*	Portugal	Varges	41.41	-7.37	4 (0;4)	26 (11;15)
CA	*T. pityocampa*	Portugal	Caparica	38.66	-9.20	4 (4;0)	35 (35;0)
TA	*T. pityocampa*	Portugal	Tavira	37.12	-7.65	4 (4;0)	37 (37;0)
GR	*T. pityocampa*	Portugal	Grândola	38.18	-8.58	4 (4;0)	35 (35;0)
F1	*T. pityocampa*	Portugal	Laboratory	NA	NA	9 (5;4)	17 (9;8)
LateLSP	*T. pityocampa*	Portugal	Leiria	39.81	-8.96	8 (7;1)	16 (16;0)
ES	*T. pityocampa*	Spain	Cortijuela	37.08	-3.47	2 (2;0)	-
RI	*T. pityocampa*	Morocco	Bab Barred	35.00	-4.90	2 (2;0)	-
OK	*T. pityocampa*	Morocco	Oukaimeden	31.20	-7.86	2 (2;0)	-
TU	*T. pityocampa*	Tunisia	Tunis	36.82	10.17	2 (1;1)	-
BO	*T. bonjeani*	Algeria	Tikjda	36.46	4.14	2 (2;0)	-
PI	*T. pinivora*	France	Alpes	44.75	6.73	2 (2;0)	-
WI	*T. wilkinsoni*	Türkiye	Mudanya; Istanbul	40.32	28.77	2 (1;1)	-

DNA extraction, library preparation and sequencing

For each Portuguese population, the DNA of 4 to 25 individuals was used for individual resequencing, and pools of 16 to 50 individuals' DNA were prepared for pool sequencing (Table 1). For each outgroup, the DNA of 2 individuals was used for individual resequencing (Table 1). All DNA samples, available from the aforementioned previous studies (Burban *et al.*, 2016, 2020) were quantified on a Qubit fluorometer (Thermo Fisher Scientific), and their quality assessed on agarose gels. Whole genome libraries were constructed using the Illumina Truseq kit following the manufacturer's recommendations. The quality of each library was inspected on a 2100 Bioanalyzer. The 114 libraries (10 Pool-Seq and 104 Ind-Seq) were paired-end (150 bp) sequenced at the MGX-Montpellier Genomix platform on an Illumina HiSeq2500 (4 libraries) or on a NovaSeq6000 (110 libraries).

Variant calling

Raw paired-end reads were filtered using *fastp* (v0.21.0; Chen *et al.*, 2018) with default parameters to remove contaminant adapter sequences and trim poor quality bases, with a phred-quality score < 15. Read pairs with either one read with a proportion of low-quality bases over 40% or containing more than 5 N bases were removed. The median percentage of reads filtered out per library was 1.52% (Figure S1A; Table S2). The median percentage of duplicated reads per library was 6.89% (Figure S1B; Table S2). Filtered reads were then mapped onto the *Tpit2.1* chromosome level reference genome (Gautier *et al.*, 2025) using the *bwa-mem2* program (v2.2.1 Vasimuddin *et al.*, 2019) with default options. Read alignments with a mapping quality Phred-score < 20 or PCR duplicates were removed using *SAMtools* software (v1.14; Danecek *et al.*, 2021) with the *view* and the *markdup* program. Finally, the mean depth, percentage of GC and insert size of each individual were checked with the option *stats* of *SAMtools*. The median percentage of reads mapped and properly paired per library was 97.65% for samples from Portugal and 93.95% for outgroups (Figure S2A,B; Table S2). The median read depth was 24.64X for individuals and 52.65X for the pools (Figure S2C; Table S2). We inferred the sex of individuals and sex ratios of pools by comparing the median read depth of the Z chromosome with that of chromosome 1, assuming that males were homogametic ZZ and that females were heterogametic ZW (Figure S3A,B; Table 1, Table S2).

We conducted 4 variant calling using *Freebayes* (v1.3.6; Garrison and Marth, 2012) based on: i) individuals and pools together (ALL_VCF), ii) individuals only (IND_VCF), iii) pools only (POOL_VCF), iv) individuals and pools together for LSP+LWP+FU populations only (DEMO_VCF), to be used in different subsequent analyses. For each calling, we used the following parameters, which allow to call variants for either individuals, pools or both (*-C 1, -F 0.01 , -G 1, -E -1 --max-complex-gap -1 --haplotype-length -1 -K --limit-coverage 500 -n 3 -m 30 -q 20*). For the specific case of the Z chromosome, the ploidy was set to 1 (*-p* option) for individual females to compute individual genotype likelihoods, while set to 2 for individual males and pools. Remaining multi-nucleotide

polymorphisms (MNP) were decomposed using the decompose function of *vt* with the option *-s* (Tan *et al.*, 2015). Further filters were applied using *bcftools* (v1.17; Danecek *et al.*, 2021), for SNP quality >= 30, absence of missing genotypes and missing individuals, reads on both strands and a minimum alternate haplotype count *(INFO/AO >= 4, and INFO/AO <=(INFO/DP-4)).* Finally, all indels were removed, and only biallelic SNPs were retained. Total site depth filter of *[200-5000]* was applied to ALL_VCF, while [200-4000] was applied to IND_VCF and POOL_VCF, and [100-3000] for the DEMO_VCF. This resulted in a total of 8,363,880 SNPs for the ALL_VCF, 8,622,533 SNPs for the IND_VCF, 9,573,120 SNPs for the POOL_VCF, and 7,537,291 SNPs for the DEMO_VCF.

Besides, two gVCF files were generated using *bcftools*, with options *-C 50 --max-depth 500 -q 20 -Q 20* : one file (OUT_gVCF) including the two highest-depth individuals from each of the LSP, LWP, and FU populations in Portugal, along with all individuals from the outgroup populations, and the second file containing all Portuguese individuals (IND_gVCF).

Characterization of population structure

All analyses were carried out separately on autosomal and Z-linked SNPs. The ALL_VCF, IND_VCF and POOL_VCF datasets were first filtered by discarding low frequency variants (i.e., MAF < 5% in at least one population) and pruned based on Linkage Disequilibrium (LD) assessed on all Ind-Seq data using the *--indep* option of *PLINK* (v1.9; Purcell *et al.*, 2007) considering a r^2 threshold of 0.1 and a 20 kb sliding window size (Figure S4). The remaining number of SNPs for autosomes (resp. Z chromosome) was 270,330 (resp. 13,655) for the ALL_VCF dataset; 271,712 (resp. 13,705) for the IND_VCF dataset; and 268,380 (resp. 13,608) for the POOL_VCF dataset. To provide a global view of the structuring of genetic diversity across all the data, we then performed a random allele PCA on the combined Pool-Seq and Ind-Seq data (filtered ALL_VCF dataset) using the *randomallele.pca* function implemented in the R package *poolfstats* (v3.0.0; Gautier *et al.*, 2022). Unsupervised hierarchical clustering was performed on individual genotypes (from the filtered IND_VCF dataset) using *Admixture* v1.3.0 (Alexander *et al.*, 2009) considering a number of clusters K ranging from 2 to 10 (the optimal K being assessed using cross-validation run with default settings). Pairwise-population F_{ST} were estimated from Pool-Seq data (filtered POOL_VCF dataset) using the unbiased estimators (Hivert *et al.*, 2018) implemented in the *compute.fstats* function of the R package *poolfstat.* To obtain unbiased estimates of the absolute nucleotide divergence D_{XY} for all pairwise population comparisons between LSP, LWP, FU and the outgroups, we relied on unfiltered whole-genome Ind-Seq data (OUT_gVCF) and used *Pixy* v1.2.7, (Korunes and Samuk, 2021). Finally, joint unfolded site frequency spectrum (SFS) graphs were plotted for SP versus WP and for WP versus FU, using allele frequencies among individuals in the POLARIZED_VCF.

Nucleotide diversity, inbreeding and effective population size

Nucleotide diversity (π) was estimated for each individual using *Pixy* v1.2.7, (Korunes and Samuk, 2021) on window sizes of 100kb and the IND_gVCF. A two-way ANOVA was used to test for π differences between the 10 populations and between autosomes and the Z chromosome. Post-hoc tests were conducted using estimated marginal means (i.e., least-squares means or adjusted means) with Bonferroni adjustment.

To characterize individual autozygosity, we relied on the hidden Markov model (HMM) based approach to detect homozygous-by-descent (HBD) segments (Druet and Gautier, 2017) implemented in the R package *RzooRoH* v0.3.2.1 (Bertrand *et al.*, 2019). Note that no LD pruning or MAF filtering was carried out (as recommended in e.g., Meyermans *et al.* (2020)) on the analyzed IND_VCF dataset. However, SNPs displaying high heterozygosity (Ho >= 70%) across all of the PPM individuals, or located within repeated elements identified within the reference genome, were discarded. In addition, since we provided population allele frequencies to *RZooRoH* that were estimated using Pool-Seq data, we applied an additional filter using the *poolfstats* package, requiring a minimum coverage of 10 reads per pool (*min.cov.per.pool = 10*). This resulted in a final dataset of 2,923,200 autosomal SNPs. In the *RZooRoH* models, autozygosity was partitioned into multiple Homozygous By Descent (HBD) classes defined by their rate parameter Rc, which determines their expected length and corresponds to groups of ancestors from different past generations. We here fitted the "layer" model (Druet and Gautier, 2022) specifying 13 HBD classes with rates Rc equal to 2; 4; 8; . . . ; 8,192 i.e. corresponding to (modeled) groups of ancestors living 1 to 4096 generations (of recombination) ago. In the absence of dense genetic maps for PPM and to facilitate such timing interpretation, SNP genetic positions were obtained from the physical maps assuming an average genomewide recombination rate of 4 cM per Mb as observed in several other species of *Lepidoptera* (Shipilina *et al.*, 2022; i Torres *et al.*, 2022; Palahí I Torres *et al.*, 2023). The underlying inference model has already been shown to be robust to variable local recombination rate (Druet and Gautier, 2017). As a matter of comparison, we also estimated the fraction of the genome presenting Runs of Homozygosity - RoH (FROH) using *Plink* v1.9 (v1.9; Purcell *et al.*, 2007) with the *--homozyg* function, based on the same dataset. Two categories of ROH were tested: ROH > 500 kb and ROH > 100 kb, to compare ancient and recent inbreeding. We used the following parameters *--homozyg-window-snp 50, --homozyg-snp 50, --homozyg-density 10, --homozyg-gap 100, --homozyg-window-het 2, and --homozyg-window-missing 5.* The individual FROH was calculated by dividing the total length of ROH across all autosomes by the total length of autosomes (579.27 Mb).

Historical effective population sizes (N_e) were inferred with the program *GONe* that implements the approach developed by Santiago *et al.* (2020), assuming a genome-wide average recombination rate of 4 cM/Mb (as justified above). To evaluate the potential impact of recent gene flow on the estimates, we ran analyses with parameter hc (recombination fraction threshold) set to 0.01

or 0.05 (Santiago *et al.*, 2020). For each population, we analyzed 100 independent subsets of 50,000 randomly selected SNPs. Near-contemporary N_e at each time point from generation 0 to generation 40 backward in time was then estimated as the harmonic mean of the 100 resulting estimates, following Feng *et al.* (2023). We also estimated the near-contemporary N_e for each population using *currentNe2* (Santiago *et al.*, 2025), after removing population-specific monomorphic sites. The analysis was performed with a constant recombination rate of 4 cM/Mb (-r 4) and accounting for population structure (-x).

Inference of the demographic history among LSP, LWP and FU

We relied on an Approximate Bayesian Computation (ABC) framework to compare different evolutionary scenarios among the sympatric LSP and LWP populations and the more distant winter population FU, and to estimate parameters for the best-supported scenario. Instead of using traditional ABC algorithms, we applied a Random Forest–based ABC approach (ABCRF) (Raynal *et al.*, 2019) implemented in the R package *abcrf* (v1.9; Pudlo *et al.*, 2016). This method represents a major improvement over classical ABC implementations in several respects. As highlighted by Pudlo et al. (2016), ABCRF (i) often achieves greater discriminative power among competing models, (ii) is more robust to the number and specific choice of summary statistics, (iii) substantially reduces computational cost (by at least an order of magnitude compared to standard rejection or regression-based ABC methods), and (iv) partially resolves the curse of dimensionality by reducing computational complexity in high-dimensional spaces. We compared three scenarios (Figure S5) corresponding to i) a simultaneous divergence of the three populations (concomitant_div); ii) a recent divergence of LSP from its sympatric population LWP (recent_div); iii) an old divergence of LSP from the two winter populations LWP and FU (old_div). Based on results previously obtained by Leblois *et al.* (2018), we introduced the possibility of bottleneck events in every population after the divergence event, and symmetric migration between the different populations up to their common ancestral one.

A total of 158 different summary statistics (see Table S3 for details) were then estimated on the Ind-Seq or the Pool-Seq data (136 and 22 statistics respectively) from the DEMO_VCF dataset described above. We then constructed a reference table by computing the same summary statistics on datasets simulated under each of the demographic scenarios using the coalescent simulator implemented in the program *msprime* (v1.2.0; Baumdicker *et al.*, 2022). Each simulated genome consisted of 1,000 segments of 10 kb with a fixed per-generation per-base recombination rate $r = 4.0 \times 10^{-8}$ and a fixed mutation rate $\mu = 2.9 \times 10^{-9}$, following Leblois *et al.* (2018). For each simulation, model parameters were sampled from the prior distributions detailed in Table S4. To further simulate Pool-Seq datasets, we relied on the *sim.readcounts* function of the R package *poolfstat* (v3.0.0; Gautier *et al.*, 2024a) based on the genetic dataset (transformed into sample allele count data) ob-

tained with *msprime* with parameters *lambda.cov* (mean of read coverages assumed to be Poisson distributed) set to 50 for LSP, LWP, and FU pool; *min.rc* (overall minimal read count threshold) set to 2; and *exp.eps* (experimental error rate introducing unequal contribution of individuals to the pool sequences) set to 1%; *seq.eps* (sequencing error rate) of 0.1%; and *genome.size* (to allow simulating spurious SNPs due to sequencing errors that may pass filtering criteria) set to 10 Mb (i.e., 1,000 segments of 10 kb).

To perform model selection, we simulated 50,000 datasets for each of the three demographic scenarios. We then used the *abcrf* function from the *abcrf* R package to build a random forest (RF) of 1,000 trees based on the resulting reference table to ensure stable estimation of the global error rate, as recommended by Collin *et al.* (2021). The best scenario was predicted 10 times using the *predict* function, along with the global and local error rates, across 10 replicate RF analyses based on training sets of 10,000 randomly chosen simulated datasets for each scenario. Finally, the parameter values for the best identified scenario were estimated using regression RF, as implemented in the function *regAbcrf* based on a new reference table constructed with 100,000 simulated datasets.

SNP annotation and polarization

We annotated and polarized SNPs in order to later inspect populations' genetic load and search for non-synonymous mutations potentially associated to allochrony.

To predict the potential effect of each SNP, we used *SnpEff* v5.2a (Cingolani *et al.*, 2012) and the gff annotation file of the reference genome applied to the complete VCF dataset (ALL_VCF). SNPs located within coding regions were categorized as synonymous (SYN"; low predicted impact), missense (MIS"; moderate impact), or loss-of-function (LOF"; high impact) variants. Variants outside coding regions were classified as either intergenic (IG"; located outside exons and splice sites) or "OTHER" (e.g., located within introns or untranslated regions). SnpEff predicted, on the autosomes and the Z respectively, 81,099 and 2,352 SYN, 53,136 and 1,772 MIS, and 764 and 22 LOF.

To determine the ancestral allelic state of each SNP site, we used bam files from a single individual for three outgroups: PPM from Oukaimeden, *T. wilkinsoni*, and *T. bonjeani* and compared their allelic states to those observed in the Portuguese populations under study. For each site, we first assessed whether each outgroup individual was polymorphic, monomorphic, or lacked sufficient coverage. Polymorphic sites were considered monomorphic if the minor allele frequency was within the lowest 10% of the frequency distribution, to avoid misclassifying spurious polymorphisms. Outgroup alleles were only considered informative if coverage > 5. Ancestral alleles were defined as the allele shared by at least two outgroups. As a final check, we retained only those sites for which the inferred ancestral allele matched either the reference or alternate allele in our VCF; sites failing this criterion were discarded.

Estimation of the genetic load in LSP, LWP and FU

To explore the genetic load of each population, we used SNPs called from Ind-Seq data that were successfully polarized and annotated as IG, SYN, MIS, or LOF. Analyses were conducted separately for autosomes and the Z chromosome, with only male individuals included for Z-linked analyses.

For LSP, LWP and FU, considering all the sites that were polymorphic among the Portuguese PPM populations, we estimated expected heterozygosity (He) for the different functional categories of variants (IG, SYN, MIS, LOF). We then calculated, for each individual, the ratio of heterozygosity at non-synonymous sites (including MIS and LOF variants) to heterozygosity at synonymous sites (SYN), expressed as HoN/HoS.

Using the subset of polarized, polymorphic sites shared among LSP, LWP, and FU, we then analyzed the site frequency spectrum (SFS) and the proportions of segregating and fixed ancestral or derived variants across the functional categories. Finally, we used these same data to calculate the R_{XY} statistic (Do *et al.*, 2015; Grossen *et al.*, 2020; Dussex *et al.*, 2021) which quantifies the relative excess or deficit of derived alleles in population X compared to population Y. The statistic was computed for the SYN, MIS, and LOF categories, using IG variants as a proxy for neutral expectations. Confidence intervals for R_{XY} were obtained via a block jackknife approach following Xue *et al.* (2015), by dividing the genome into 100 blocks with equal numbers of consecutive SNPs. To further investigate the R_{XY} patterns, we generated joint SFS plots comparing LSP, LWP, and FU across the different functional categories.

Search for genomic bases associated with allochrony

To pinpoint genomic regions potentially involved in the phenological shift in LSP, genome scans were performed using the C2 based contrast analysis proposed by Olazcuaga *et al.* (2020) as implemented in the latest version of the *BayPass* (v3.0 Gautier, 2015), which allows joint analyses of Ind-Seq and Pool-Seq data (Camus *et al.*, 2024). More precisely, we ran the *BayPass* core model with default parameters on the ALL_VCF dataset, with Ind-Seq data provided in the form of genotype likelihoods (GL) and Pool-Seq data in the form of read counts (see the *BayPass* manual for details), analyzing autosomes and the Z chromosome separately. The C2 statistic was defined to contrast samples representative of summer populations (i.e., LSP and LateLSP) and the winter populations (LWP, FU, TA, GR, VA, CA, VI). To identify and delineate significant windows associated with allochrony, i.e. with an excess of high C2 values, we relied on the local-score approach described in Fariello *et al.* (2017) and (re)implemented in the R function *compute.local.scores* available in the *BayPass* (v3.0 Gautier, 2015; Camus *et al.*, 2024), using default options but $\xi = 2$.

To go further into the description of the significant windows for association with allochrony identified using *BayPass*, we estimated F_{ST} between LSP, LWP, FU, and LateLSP pools over 10 SNPs windows with the *compute.fstats* function from *poolfstat 3.0.0* (using a MAF cutoff of 0.05).

We also relied on Ind-Seq data to estimate D_{XY} between LSP, LWP, FU, and outgroups, as well as average π for LSP and LWP, using 2 Kb window sizes (corresponding on average on the Z chromosome to 10-SNP windows) with *Pixy*, and Tajima's D with *VCFtools 0.1.16* (Danecek *et al.*, 2011). Joint unfolded SFS graphs were plotted for SP versus WP and WP versus FU based on polarized ind-seq data. *LDBlockShow* v1.40 (Dong *et al.*, 2021) was used to assess LD patterns. Genotype_plots were used to depict genotypic structure between populations (Whiting, 2022). The read depth of LSP and LWP Ind-Seq libraries were calculated from BAM files using *Mosdepth* (Pedersen and Quinlan, 2018) over 2 kb windows to identify potential candidate CNVs. *Smoove* (https://github.com/brentp/smoove) was used to search for short INDELs and inversions based on Ind-Seq bam files. Lastly, we used *Bedtools* 2.31.1 (Quinlan and Hall, 2010) to list candidate genes located under each window showing high C2 values and we looked for the potential presence of non-synonymous mutations identified using SnpEff.

Results

Population structure and differentiation

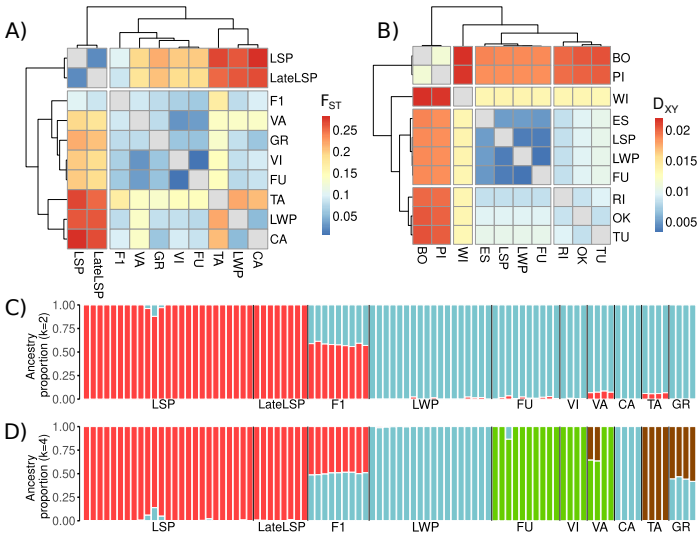

Figure 2: **Genetic structure among populations from Portugal. (A)** Pairwise F_{ST} estimates between Portuguese populations. **(B)** pairwise D_{XY} estimates between three Portuguese populations (LSP, LWP, and FU) and outgroups. Unsupervised ancestry assignments based on the program Admixture **(C)** at k = 2 and **(D)** at k = 4. All analyses were performed on autosomes. Population codes are defined in Table 1.

F_{ST}, PCA, and Admixture analyses on autosomes revealed strong genetic structure among populations and especially between the summer and winter populations. The pairwise F_{ST} values were high between LSP and winter populations, e.g. 0.259 with LWP, 0.197 with FU, and 0.278 with CA (Figure 2A; Table S5), but was only (1.80×10^{-3}) with LateLSP. In contrast, pairwise F_{ST} between winter populations was lower, ranging from (3.00×10^{-3}) between FU and VI to 0.216 between LWP

and TA. In turn, D_{XY} between LSP and LWP was relatively low and comparable to D_{XY} between LWP and FU (3.50×10^{-3} and 3.10×10^{-3}, respectively, Figure 2B; Table S6). D_{XY} values between both LSP and LWP versus outgroups were high (1.2×10^{-2} on average in both cases). The PCA first axis separated the summer populations (LSP and LateLSP), F1 and the other populations (Figure S6). The Admixture analysis yielded a minimal cross-validation error for K = 2, separating summer and winter populations into two distinct genetic clusters (Figure 2C). Increasing the K parameter value further delineated the winter populations (Figure 2D). Finally, while the joint SFS for LSP and LWP showed strong differentiation, with LSP displaying slightly more fixed variants (Figure S7A), the joint SFS for LWP and FU exhibited low differentiation, with LWP showing more fixed variants (Figure S7B).

Demographic history inferred using ABCRF

The highest RF classification mean votes (ranging from 391 to 456) and estimated posterior probabilities (ranging from 0.55 to 0.63) provided the strongest support for the recent_div compared to the other two scenarios (Table S7, Figure S8). However, although the old_div scenario was always the least supported, the concomitant_div scenario occasionally received the highest posterior probability in some RF analysis replicates. This is consistent with the relatively recent divergence between LWP and FU, given that the concomitant_div scenario represents a limiting case of old_div where no differentiation is assumed between the population ancestral to both LWP/LSP and FU.

Table 2: **Results for estimation of parameters of interest for the scenario recent_div.** Mean and median estimates for each parameter were calculated using an ABCRF run conducted with a reference table containing 158 summary statistics from 100,000 simulations. Quantiles at 5% (lower bound) and 95% (upper bound) were used to obtain the 90% confidence interval (CI). Mean and median global and local NMAE (normalized mean absolute error, which is the average absolute difference between the point estimate and the true simulated value divided by the true simulated value) were also estimated.

Parameter	**Posterior point estimates of**		**90% CI**		**Global (prior) NMAE computed from**		**Local (posterior) NMAE computed**	
Parameter	**Mean**	**Median**	**Lower bound**	**Upper bound**	**Mean**	**Median**	**Mean**	**Median**
m_ancPP	3.24×10^{-4}	3.09×10^{-4}	7.76×10^{-6}	2.19×10^{-2}	2.12	2.07	0.45	0.45
m_anc_LSPFU	1.74×10^{-3}	2.29×10^{-3}	2.09×10^{-5}	1.29×10^{-1}	2.34	2.29	0.68	0.67
m_rec_LSPFU	9.55×10^{-4}	1.38×10^{-3}	4.37×10^{-5}	8.71×10^{-3}	0.85	0.76	0.17	0.16
m_anc_LSPLWP	5.75×10^{-3}	5.62×10^{-3}	1.41×10^{-4}	2.24×10^{-1}	2.49	2.46	1.10	1.11
m_rec_LSPLWP	6.03×10^{-4}	6.17×10^{-4}	7.94×10^{-5}	4.07×10^{-3}	0.99	0.91	0.13	0.13
m_anc_LWPFU	2.19×10^{-3}	2.09×10^{-3}	4.07×10^{-5}	1.26×10^{-1}	1.57	1.53	0.72	0.71
m_rec_LWPFU	1.15×10^{-3}	1.78×10^{-3}	2.63×10^{-5}	1.58×10^{-2}	0.84	0.76	0.20	0.18
N_e_anc_PP	5.13×10^{5}	6.31×10^{5}	2.14×10^{5}	8.91×10^{5}	0.01	0.01	0.00	0.00
N_e_bot_LSP	5.49×10^{0}	3.24×10^{0}	1.15×10^{0}	1.74×10^{2}	4.56	3.90	4.64	3.10
N_e_bot_FU	1.10×10^{2}	9.12×10^{1}	2.14×10^{0}	1.38×10^{4}	2.79	2.34	3.25	3.04
N_e_bot_LWP	1.78×10^{1}	1.26×10^{1}	1.51×10^{0}	6.92×10^{2}	4.17	3.58	1.73	1.47
N_e_LSP	1.07×10^{3}	6.46×10^{2}	1.15×10^{2}	5.13×10^{4}	0.19	0.18	0.23	0.21
N_e_FU	1.12×10^{4}	1.32×10^{4}	5.75×10^{2}	1.15×10^{5}	0.14	0.14	0.14	0.14
N_e_LWP	1.38×10^{3}	1.70×10^{3}	1.51×10^{2}	8.51×10^{3}	0.14	0.14	0.16	0.16
N_e_Old_FU	1.95×10^{4}	3.02×10^{4}	5.50×10^{2}	1.66×10^{5}	0.19	0.18	0.16	0.16
N_e_Old_LWP	2.34×10^{3}	1.86×10^{3}	1.26×10^{2}	6.46×10^{4}	0.21	0.20	0.23	0.22
N_e_ancfound_FU	1.23×10^{4}	1.82×10^{4}	2.14×10^{2}	3.09×10^{5}	0.24	0.24	0.23	0.23
N_e_ancfound_LWP	6.03×10^{3}	5.62×10^{3}	1.32×10^{2}	2.69×10^{5}	0.25	0.25	0.28	0.28
N_e_found_LWP	4.79×10^{1}	3.55×10^{1}	1.55×10^{0}	4.90×10^{3}	48.49	46.37	4.29	3.91
N_e_found_LSP	4.90×10^{1}	3.63×10^{1}	1.35×10^{0}	6.46×10^{3}	5.46	4.92	7.88	7.50
$Tsplit_PP$	2.24×10^{3}	1.95×10^{3}	5.75×10^{2}	1.29×10^{4}	0.08	0.07	0.09	0.09
$Tsplit_LWP$	8.71×10^{1}	7.41×10^{1}	1.29×10^{1}	8.71×10^{2}	0.24	0.23	0.25	0.25
$Tbot$	2.51×10^{1}	2.04×10^{1}	1.05×10^{1}	1.29×10^{2}	0.22	0.20	0.19	0.17

Figure 3A and Table 2 give the estimated values for the parameters of the recent_div scenario (see Figure S9, S10 and S11 and Table S8 for the concomitant_div scenario), and revealed a recent divergence between LSP and LWP, occurring approximately a hundred generations ago (74, 90% CI: 13–871). A bottleneck likely occurred around 20 generations ago (CI: 10–129) in all three populations, with estimated effective sizes of few individuals only for LSP (3, CI: 1-174), 13 individuals in LWP (CI: 1-692) and 91 in FU (CI: 2-13,490; Figure 3A; Table 2). Contemporary gene flow among the three populations was significantly reduced following this bottleneck, with the most pronounced reduction observed between LSP and LWP (6.17×10^{-4}; 90% CI: $7.94 \times 10^{-5} - 3.98 \times 10^{-3}$). Finally, LSP showed the lowest current N_e (645 individuals; 90% CI: 115–51,286), while FU showed the highest current N_e (13,182 individuals; 90% CI: 575–114,815). Parameters specifying past events, such as ancient migration or ancient N_e, and founder events were not accurately estimated with large Normalized Mean Absolute Error and flat posterior distribution (Figure S12 and S13; Table 2).

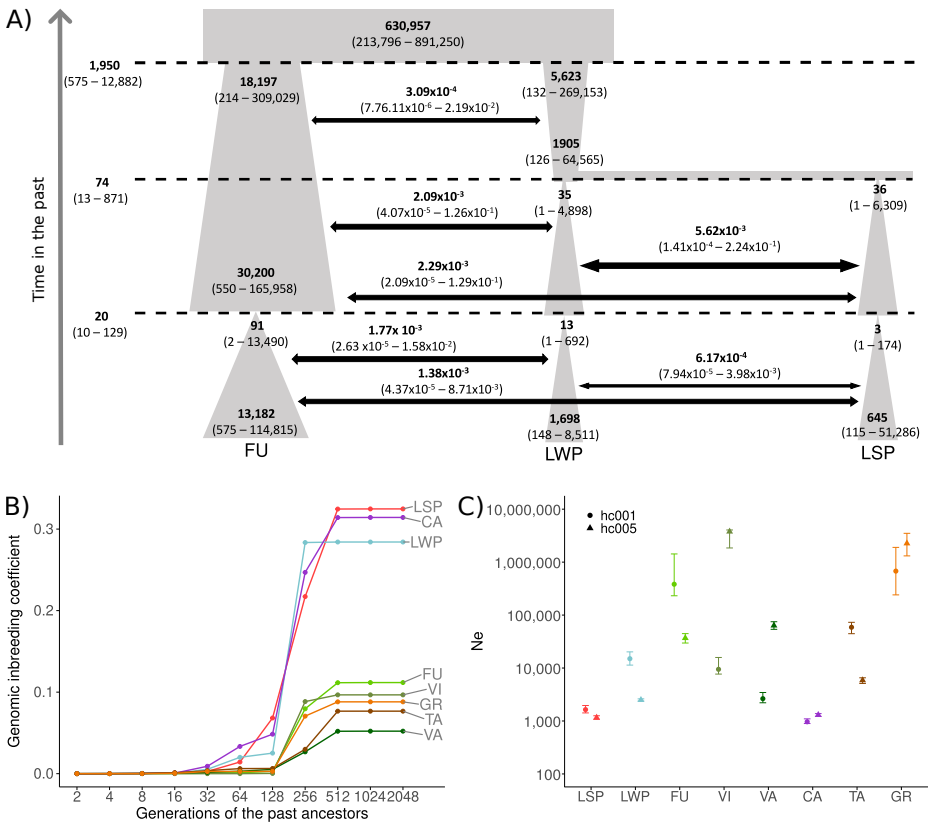

Figure 3: **Inference of demographic history, inbreeding and effective population sizes. (A)** Graphical representation of the best demographic scenario, recent_div, of divergence between LSP, LWP and FU populations. Split time and bottleneck time (horizontal dashed lines), populations size and migration rates (horizontal arrows) were estimated from 100,000 simulations. Bold values correspond to the median estimates of the parameters, and numbers in parentheses indicate 95% confidence intervals. **(B)** Genomic inbreeding inferred by RZooRoH on autosomes. Each line represents the mean value of the individual genomic inbreeding coefficient for a population, cumulated through time. The X-axis represents the generations of ancestors contributing to each HBD class. Only HBD classes ranging from 2 to 2048 generations are shown. **(C)** Near-contemporary effective population size (0 to 50 generations) estimated with GONE on autosomes. For each population, the mean and range of N_e estimates obtained using a recombination fraction (hc) of 0.01 and 0.05 are shown.

Population diversity, inbreeding, and effective population size

Nucleotide diversity (π) estimated over all autosomes varied greatly among populations (2-ways ANOVA, p-value < 0.001), in line with very different demographic histories. The genetic diversity was the lowest in the LSP and LateLSP populations ($\pi = 2.1 \times 10^{-3}$); intermediate in the LWP and CA populations ($\pi = 2.5 \times 10^{-3}$); and ranged from $\pi = 3.0 \times 10^{-3}$ (for TA) to $\pi = 3.4 \times 10^{-3}$ (for VA) in the other Portuguese populations (Figure S14).

RZooRoH analyses showed much higher levels of genomic inbreeding coefficients in LSP, LWP and CA (from 0.284 for LWP to 0.325 for LSP) compared to other populations (from 0.052 for VA to 0.112 for FU). Importantly, *RZooRoH* inferred that most autozygosity concentrated in HBD segments was associated with ancestors living 64 to 256 generations ago (corresponding to classes 128 to 512; Figure 3B), while little genomic inbreeding originated from more recent (< 64 generations) or more ancient (> 512 generations) ancestors. Moreover, there was little inter-individual variance in the level of inbreeding and the age of large HBD segments within each population (Figure S15). In line with these analyses, LSP, LWP and CA exhibited high FROH values ranging from 0.280 to 0.336, approximately 3 times higher than those of other Portuguese populations, which ranged from 0.059 (for VA) to 0.106 (for FU, Figure S16). The same pattern was observed for the fraction of the genome covered by ROH > 500 Kb (Figure S17A). Similarly, the number of both short and long ROH displayed comparable patterns (Figure S17B,C,D).

Finally, the near contemporary N_e estimated with *GONe* indicated a relatively small recent N_e (< 1000) for the LSP summer population and its two LWP and CA neighbors $N_e < 5000$ and $N_e < 1000$, respectively, compared to more distant populations (Figure 3C, Figure S18A,B). Similar results were found using *currentNe2*, with N_e of 999 in LSP (90% CI:540-1849), 652 in LWP (90% CI=330-1289), and 113 in CA (90% CI:33-391). However, *currentNe2* was unable to converge for FU, VI, VA, TA, and GR, which was likely due to the large N_e of these populations.

Genetic load

Expected heterozygosity was lower in LSP compared to LWP and in LWP compared to FU, for each category of sites (Figure 4A). For each population, expected heterozygosity was lower at LOF sites compared to MIS sites, and at MIS sites compared to SYN and IG sites (Figure 4A). HoN/HoS ratio was higher in LSP and LWP compared to FU (Post-hoc Dunn's tests with Bonferroni correction; p-values $= 1.7 \times 10^{-3}$ and 2.0×10^{-5}, respectively) and was not different between LSP and LWP (p-value $= 3.5 \times 10^{-1}$; Figure 4B). There were lower proportions of homozygous derived genotypes at LOF sites, compared to MIS sites, and at MIS sites compared to SYN and IG sites, in each population (Figure 4C), suggesting an effect of purifying selection at MIS sites and more strikingly at LOF sites. However, proportions of homozygous derived genotypes were higher in LSP than in LWP and in LWP than in FU, for every category of variant (Figure 4C), suggesting accumulation of

genetic load in LSP compared to LWP and in LWP compared to FU, likely resulting from reduced purifying selection and increased drift. The SFS (Figure S19) indicated similar trends. R_{XY} values > 1 were observed when comparing LSP to LWP, LSP to FU, and LWP to FU, for the LOF sites (Figure 4D), indicating a relative accumulation of genetic load at LOF sites in LSP compared to LWP and FU, and in LWP compared to FU. R_{XY} values for the SYN and MIS classes were close to 1, suggesting little differences in purge-accumulation of genetic load signals at these sites between populations. Yet, R_{XY} values were not significantly different to 1 when using z-score two-tailed test. Joint SFS between populations for each category of sites illustrated the accumulation of derived alleles at LOF variants in LWP and more importantly in LSP, compared to FU (Figure S20).

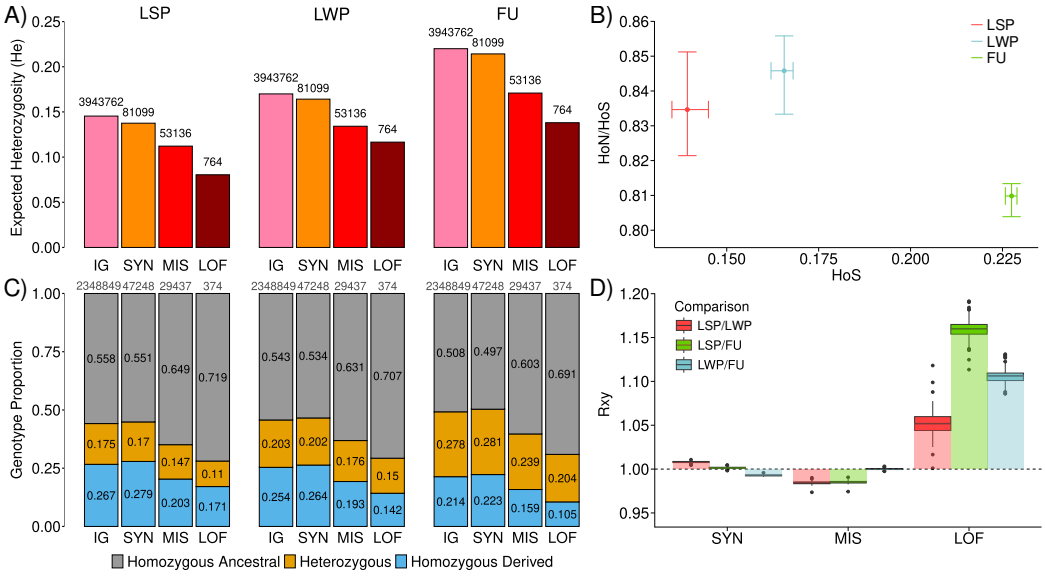

Figure 4: **Estimates of genetic load in LSP, LWP and FU on autosomes. (A)** Expected heterozygosities at IG, SYN, MIS, and LOF sites, with the count of variants indicated above each category. **(B)** Relationship between the ratio of individual heterozygosity at MIS (HoN) to SYN (HoS) sites and individual heterozygosity at SYN (HoS) sites. Each point represents the population median, with confidence intervals based on individual-level variation. **(C)** Proportions of genotypes that were either homozygous ancestral, homozygous derived or heterozygous, at IG, SYN, MIS, and LOF sites, with the count of variants indicated above each category. **(D)** R_{XY} statistics at SYN, MIS, and LOF sites.

Genome scans between summer and winter populations

The C2 contrast analyses comparing summer and winter population samples allowed identifying 135 significant windows (102 genes) on autosomes and 20 windows (25 genes) on the Z chromosome (Figure 5A, 5B; Table S9). Among these significant windows, some —including those showing the highest C2 values and their surrounding regions (+/- 10 kb)— contain several genes known to influence biological rhythms: *NPAS2* (neuronal PAS domain-containing protein 2-like), $GABA_B$ (gamma-aminobutyric acid receptor subunit beta-like), *Coq8* (atypical kinase COQ8B, mitochondiral), *NEURAL4 / NHR* (neuralized-like protein 4), *NEP-2* (neprilysin-2) and *SP2353* (basement membrane-specific heparan sulfate proteoglycan core protein-like). Although the *period* gene was

not in a C2 local score window, it lied - 100 kb upstream of the second highest C2 local score window. F_{ST} between either LSP and LWP and between LateLSP and LWP were on average high across the entire Z chromosome, while F_{ST} between LSP and LateLSP were consistently close to zero (Figure 5C). The region around *NEURAL4 / NHR* and *NEP-2* showed particularly high F_{ST} values between either LSP or LateLSP and LWP (Figure 5D). This genetic differentiation pattern was similar across the several comparisons between LSP and winter populations (Figure S21).

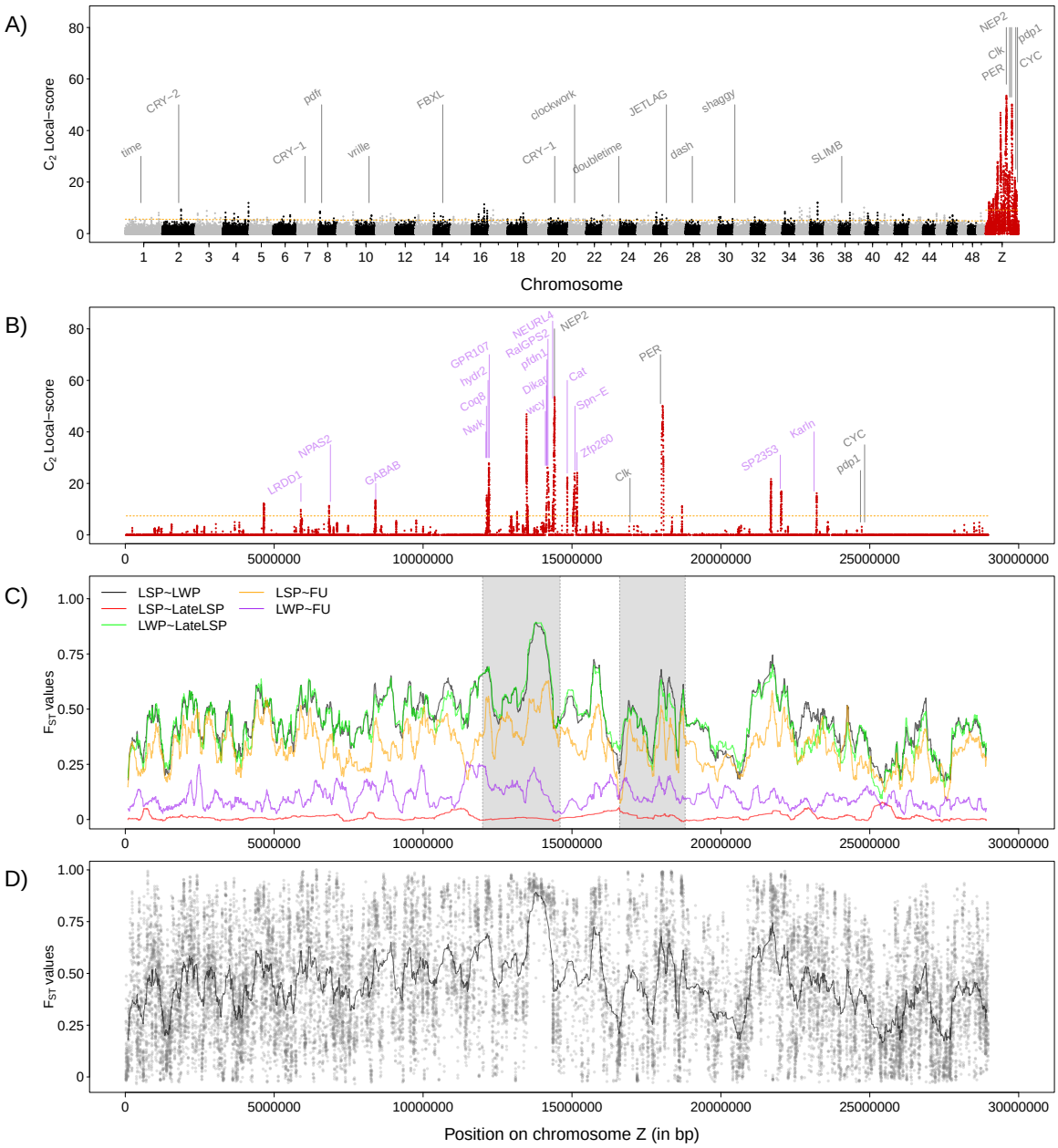

Figure 5: **Genome scan looking for loci underlying phenological variations.** Manhattan plot of the Local-score based on C2 p-values for the entire genome **(A)** and for the Z chromosome only **(B)**. The horizontal dashed lines indicate the chromosome-specific 1% significance threshold for the local score. Genes potentially linked to circadian rhythm and diapause are written in grey. Z-linked genes related to other functions and identified in high C2 local scores are written in purple. **(C)** F_{ST} values along the Z chromosome for five population pairs. The light grey rectangles correspond to the regions zoomed in on Figure 6. **(D)** F_{ST} values along the Z chromosome between LSP and LWP. Each point represents the estimated F_{ST} value between LSP and LWP pools for 10-SNPs windows.

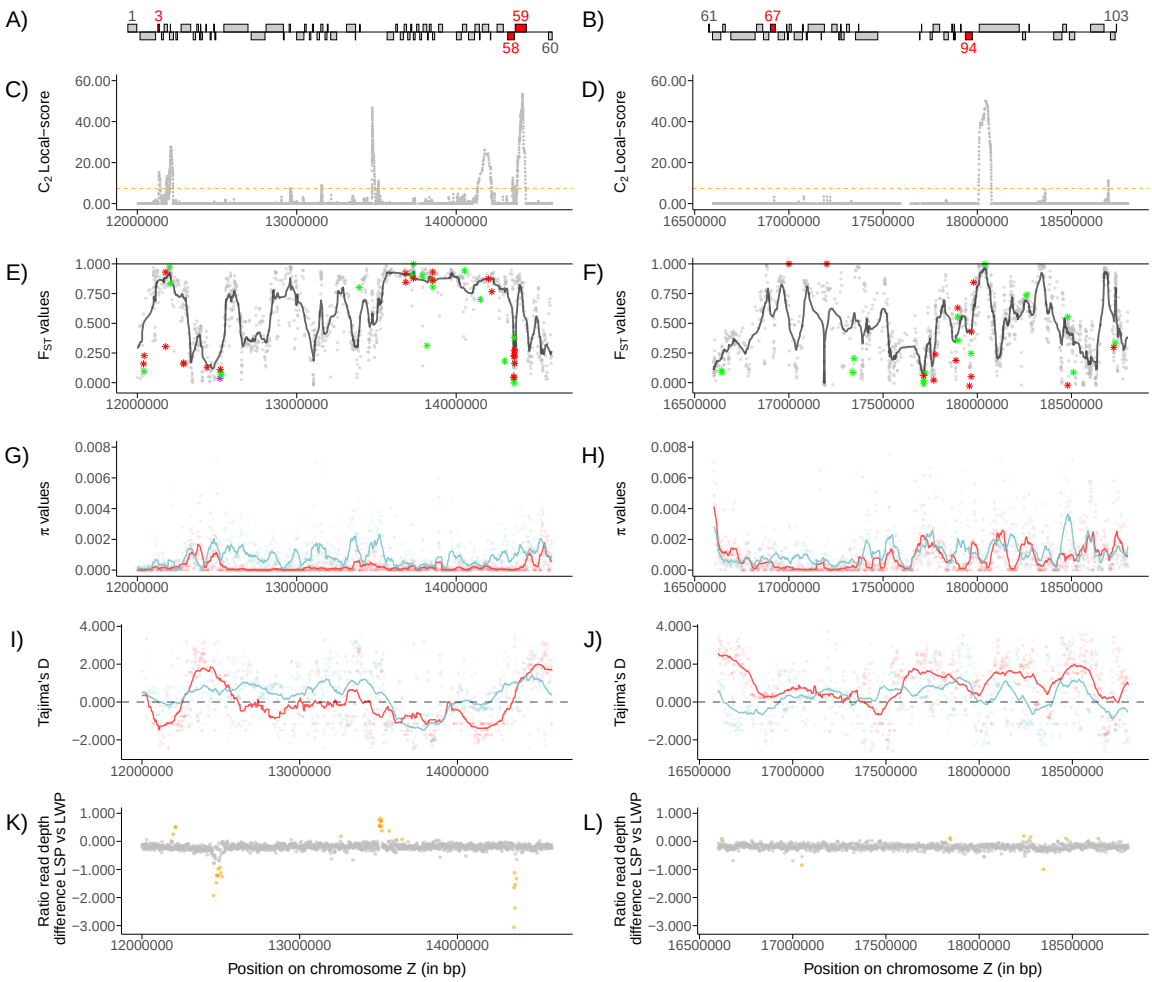

Figure 6: **Genetic characteristics of two genomic regions, from 12.00 Mb to 14.60 Mb and from 16.60 Mb to 18.80 Mb, showing high genetic contrasts between LSP and LWP. (A & B)**: Annotated genes (see also Table S10) present within both genomic regions. Red rectangles indicate genes associated with functions related to circadian rhythm and diapause in insects (3: *Coq8*; 58: *NEURL4*; 59: *NEP-2*; 67: *Clk*; 94: *period*) and grey rectangles represent other genes. **C & D**: Manhattan plots of the Local-score based on C2 p-values, zoomed in on the Z chromosome. The black dashed line represents the chromosome-specific 1% significance threshold for the local score. **E & F**: Pairwise F_{ST} between LSP and LWP. Each point represents an estimated value calculated over a 10-SNP sliding window. Green, red and purple stars respectively indicate tolerant, deleterious and loss of function mutations, with their SNP F_{ST} values. **G & H**: Pi value of LSP (in red) and LWP (in blue), with each point representing the average Pi over a 2,000 bp window. **I & J**: Tajima's D in LSP (in red) and LWP (in blue), where each point represents the average Tajima's D over a 2,000 bp window. **K & L**: Read depth difference between LSP and LWP (see methods), with each point corresponding to average depth over a 2,000 bp window. Negative values indicate an excess of read depth in LWP relative to LSP, and positive values indicate the opposite. Orange dots indicate SNPs exceeding the 99th percentile threshold of the distribution.

Among the Z chromosome candidate windows, we focused on two regions of interest, showing high C2 and F_{ST} and including *Coq8*, *NEURL4*, *NEP-2*, *Clk* (circadian locomoter output cycles protein kaput) and *period* genes.

First, in the surroundings of the *Coq8* gene, from 12.10 to 12.25 Mb (Figure 6A; Table S10), we observed high C2 local score windows and F_{ST} value between LSP and LWP pools (Figure 6C, 6E). π in this region was low in both LSP and LWP (Figure 6G). Tajima's D was negative in LSP

while it was close to 0 for LWP (Figure 6I). One potentially deleterious amino acid change from Val to Ala, differentially fixed between LWP and LSP, was found in the *Coq8* gene (Table S11).

Then, the region between 13.60 and 14.30 Mb, upstream of the *NEURL4* and *NEP-2* genes (figure 6A; Table S10), showed several windows with a high C2 local score (Figure 6C) and an almost constantly high F_{ST} value between LSP and LWP pools (Figure 6E). The joint SFS between LSP and LWP individuals in this region showed a large proportion of SNPs with opposite allele frequencies (Figure S7C), while LWP and FU exhibited more similar allele frequencies (Figure S7D). This was also clearly depicted by genotype plots (Figure S22, S23). LD was particularly high in this highly differentiated region (Figure S24). π in this region was lower in both LSP (mean: $\pi = 2.0 \times 10^{-4}$) and LWP (mean: $\pi = 8.0 \times 10^{-4}$) relative to the Z chromosome's π average (mean LSP: $\pi = 7.0 \times 10^{-3}$; mean LWP: $\pi = 1.3 \times 10^{-3}$), and π was here lower in LSP than in LWP (Figure 6G). In addition, Tajima's D was negative over large regions both in LSP and in LWP, from 13.33 Mb to 14.33 Mb and 13.57 Mb to 14.10 Mb, and more negative in LSP than in LWP from 13.90 Mb to 14.33 bp (Figure 6I). The D_{XY} scan between LSP and LWP or LSP and FU in this region revealed no significantly high values (Figure S25) and the D_{XY} tree's topologies remained consistent between the entire Z chromosome and this highly differentiated region (Figure S26). Although a few small structural variants were detected by the *Smoove* analysis, we did not identify large INDELs or inversions in this region (not shown). *Mosdepth* nevertheless identified from 14,358,000 to 14,376,000 bp, within *NEURL4* and upstream *NEP-2* (Figure 6K), a lower read depth in LSP (18.13 X on average) than in LWP (42.70 X on average), which may reflect a short duplication or CNV in LWP and/or deletion in LSP. In this specific region, we found an excess of heterozygous individuals in LWP (Ho = 0.83; 140 SNPs with Ho > 0.5), but not in LSP (Ho = 0.25; 0 SNP with Ho > 0.5), which may reflect the presence of a polymorphic duplication in LWP which was absent (i.e. deleted) in LSP (Figure S27). Evidence of this putative duplication was also found in FU and other winter populations, but not in LateLSP. The excess of heterozygosity in LWP in this region between 14,358,000 to 14,376,000 bp was responsible for the observed narrow drop in F_{ST} values at this region (Figure S27). No deleterious mutation with a strong differentiation between LSP and LWP was found in *NEP-2* or *NEURL4*. However, there were several potentially deleterious amino acid changes in the *NEURL4* region showing an excess of read depth and of heterozygosity in LWP, that were at intermediate frequencies in LWP (Table S11) and fixed or almost fixed in LSP and that should hence be considered carefully.

In the surroundings of the *Clk* gene, from 16.93 to 16.96 Mb (Figure 6B; Table S10), we observed high F_{ST} windows (Figure 6D, 6F). π in this region was low in both LSP and LWP (Figure 6H). Tajima's D was slightly negative in LWP and positive or close to 0 in LSP, from 16.50 Mb to 17.40 Mb (Figure 6J). No deleterious mutation with strong differentiation between LSP and LWP was found in *Clk*.

Immediately downstream of the *period* gene (Figure 6B) from 18.00 Mb to 18.05 Mb, we observed a window with a high C2 local score and a cluster of windows with high F_{ST} values (Figure 6D,F; Table S10). There was no strong decrease in π (Figure 6H) nor of Tajima's D (Figure 6J). One potentially deleterious amino acid change from Val to Ala showing a frequency of 0 in LWP and 0.8 in LSP was found in the *period* gene (Table S11).

Genetic differentiation, diversity and genetic load on the Z chromosome

Pairwise F_{ST} values were higher for the Z chromosome than for autosomes for most of the population pairs (Figure 7A, Figure S28; Table S12). However, this higher differentiation at the Z chromosome was much more pronounced for population pairs including either LSP or LateLSP. In turn, the D_{XY} values for the Z chromosome between LSP or LateLSP were comparable to those of the winter Portuguese populations, although D_{XY} on autosomes was globally slightly higher than on the Z chromosome (Figure S29,S30; Table S13). π values varied significantly between autosomes and the Z chromosome across populations (ANOVA, p-value ¡ 0.001, Figure S31), with a particularly reduced π on the Z chromosome in LSP and LateLSP (median values of 6.0×10^{-4}), compared to other winter populations (median values ranging from $\pi = 1.3 \times 10^{-3}$ to $\pi = 1.9 \times 10^{-3}$). Interestingly, while LSP and LWP showed relatively low autosomal π compared to other populations, the average ratio of π on the Z chromosome over π on autosomes was significantly lower for LSP (0.244; Figure 7B) than for LWP (0.523) or the other Portuguese populations (ranging from 0.507 to 0.590), suggesting a striking reduction of genetic diversity on the Z chromosome in LSP.

The PCA on the Z chromosome showed a clear population genetic structure (Figure S32), with the F1 females grouping either with LSP or LWP individuals (as expected given that they inherited only one Z copy from either a LSP father or a LWP father). Accordingly, the admixture analysis on the Z chromosome revealed the two main distinct genetic groups corresponding to the summer and winter populations (Figure S33A,B), with F1 males mixed between both groups and females grouping with either LSP or LWP individuals. The minimum cross-validation error was observed at K = 2, delineating summer and winter populations. Finally, the joint SFS for LSP and LWP showed strong differentiation on the Z chromosome, with many differentially fixed variants, especially between LSP and LWP (Figure S7E). The joint SFS for LWP and FU on the Z chromosome exhibited less differentiation (Figure S7F).

Similarly as for autosomes, He at the Z chromosome was lower at LOF sites compared to MIS sites, and at MIS sites compared to SYN and IG sites, across all three populations (Figure S34A). Additionally, expected heterozygosity was consistently lower in LSP compared to LWP, and in LWP compared to FU, across all categories of variants. HoN/HoS ratio was higher in LSP and LWP compared to FU (p-values = 2.2×10^{-2} and 3.5×10^{-2}, respectively) and was not different between

LSP and LWP (p-value = 1; Figure S34B). In contrast to autosomes, the proportion of homozygous derived genotypes at polarized LOF sites was lower in LSP compared to LWP and FU (Figure S34C), although there were few LOF sites on the Z (n = 12). Very low R_{XY} values were observed when comparing LSP to LWP and LSP to FU at LOF sites (Figure S34D), suggesting stronger purging of LOF derived alleles in LSP compared to LWP and FU. The site frequency spectrum (SFS, Figure S35) supported these trends. Joint SFS analyses further illustrated an excess of derived alleles at SYN and MIS sites in LSP compared to LWP and LSP, but the opposite patterns for the few LOF sites (Figure S36).

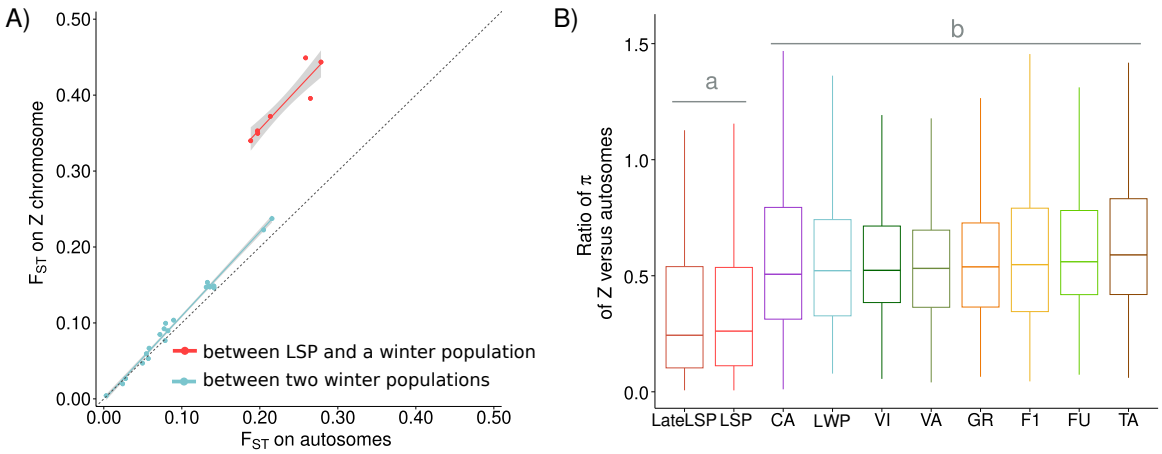

Figure 7: **Over-differentiation and reduced nucleotide diversity on the Z chromosome in LSP. (A)** Pairwise F_{ST} values on the Z chromosome compared to pairwise F_{ST} values on autosomes. The red dots represent F_{ST} values involving comparisons between LSP and winter populations, and the blue dots indicate F_{ST} values for comparisons between winter populations. **(B)** Ratio of the median nucleotide diversity (π) calculated for the Z chromosome over the median π calculated for autosomes, for male individuals. Statistically significant differences were indicated letter *a* and *b*.

Discussion

In this study, we investigated the genetic causes and consequences of the allochronic divergence between pine processionary moth populations in Portugal. We empirically explored theoretical predictions (Devaux and Lande, 2008; Weis *et al.*, 2014) proposing that small N_e, non-selective assortative mating based on the reproduction date, and the existence of polymorphisms at genes controlling for the phenology of reproduction would promote the onset of allochronic divergence. We found a strong differentiation between LSP and LWP sympatric allochronic populations, together with evidence for recent gene flow reduction (but not to the point of total isolation), as previously shown for this system (Santos *et al.*, 2011; Burban *et al.*, 2016, 2020; Leblois *et al.*, 2018). In addition, we found statistical support for recent strong inbreeding and reduction of N_e not only in LSP but also in LWP, suggesting that a common environmental force might have driven a reduction in size of both populations. Rather than being purged, the genetic load in LSP appears to have increased, likely as a consequence of strong drift and reduced efficacy of selection. Our most salient

result is the identification of a marked reduction in genetic diversity and enhanced differentiation in several Z-linked regions associated with allochrony, including genes known to be implicated in the regulation of biological rhythms in insects. Moreover, the entire Z chromosome in LSP showed an over-differentiation and a strong reduction in nucleotide diversity, compared to other populations, likely indicating both a strong bottleneck and the onset of a barrier to gene flow on the LSP Z chromosome. We discuss how these empirical genomic findings support theoretical expectations that assortative mating driven by reproductive timing differences underpinned by polymorphisms in circadian genes, along with enhanced genetic drift and purge of genetic load, can promote the onset and persistence of allochronic divergence.

Demographic history of the allochronic summer population

The most likely scenario of divergence of LSP inferred by ABCRF suggested a split between LSP and LWP approximately 74 generations before sampling, concomitant with strong bottlenecks in both populations. In line with this result, RzooRoh analyzes detected inbreeding generated 64-256 generations ago in LSP, LWP and CA. Similar patterns were reported, for example, by Gautier *et al.* (2024b), who observed a high proportion of short HBD segments in a cattle population that experienced a strong recent founder effect followed by rapid expansion. Importantly, these results of a recent split between LSP and LWP and recent inbreeding in both populations support a scenario of recent sympatric allochronic divergence caused by increased genetic drift, in line with theoretical expectations from Devaux and Lande (2008). Conversely, the results do not support alternative divergence scenarios, such as an ancestral divergence of LSP (potentially in parapatry) and/or a major role of spatially varying selection acting on phenology.

Among the potential environmental drivers of such recent bottlenecks and inbreeding, the management of the pine forest in Portugal may have played an important role. Indeed, large-scale intensive afforestation in Mata National de Leiria started in 1892 and continued throughout the 20th century (Pinho, 2012) with a particular intensification in the 1930s, primarily using maritime pine *Pinus pinaster*. These efforts expanded further in the 1960s, incorporating clear-cut practices and the development of 120 km of forest roads. This widespread plantation of maritime pine may have facilitated the establishment of new populations in which bottlenecks may have occurred via founder effects. In addition, important forest fires have been documented in Portugal, especially in pine tree forests (Mendes, 2007; Aguiar *et al.*, 2021), which potentially triggered bottlenecks in pine processionary moth populations.

The fact that our estimated split time between LSP and LWP was slightly more recent than the one inferred by Leblois *et al.* (2018) (560 generations) could be due to the fact that we included as a third population in our ABCRF analysis a more distant and more diverse winter population (FU) compared to Leblois *et al.* (2018) (CA). Besides, the accuracy of demographic parameter estimation

may have been improved by the use of a much larger SNP dataset and a more precise SNP calling thanks to the use of a new chromosome-level reference genome (Fuentes-Pardo and Ruzzante, 2017; Martchenko and Shafer, 2023). In addition, using *msprime* enabled the simulation of numerous coalescent scenarios (Kelleher *et al.*, 2019; Liu *et al.*, 2020) and using ABCRF allowed further improvement of the accuracy of parameter inference by incorporating a diverse set of numerous non-independent summary statistics from PoolSeq and IndSeq for parameter estimation (Pudlo *et al.*, 2016; Raynal *et al.*, 2019).

Despite evidence of recent strong bottlenecks in LSP and LWP, intermediate (i.e. relatively high) values of contemporary N_e were estimated (645 using ABCRF and 1,000 using *GONe*). Such N_e in LSP is much lower than N_e estimates sometimes exceeding 10,000 in several other Lepidoptera species (Bortoluzzi *et al.*, 2023; García-Berro *et al.*, 2023), but is also considerably higher than in endangered or compromised populations of other Lepidoptera species where N_e often falls below 100 or even 50 individuals (Andersen *et al.*, 2014; Marí-Mena *et al.*, 2019). Hence, the contemporary N_e values estimated for LSP appear high relative to the fact that this population experienced a relatively recent severe bottleneck and is reproductively isolated due to its allochrony. Such relatively high contemporary N_e in LSP may be first explained by the persistence of a certain amount of gene flow from winter populations into LSP, at least at autosomal regions of the genome. In addition, such a large N_e could have been caused by a rapid growth of the LSP population, which is in line with the observed geographical expansion of LSP along the coast of Portugal (Godefroid *et al.*, 2016). Such growth could have been facilitated by a relaxation of both intraspecific competition and predation or parasitism by natural enemies (Santos *et al.*, 2013). LSP population growth could have also been favored by local climatic conditions compatible (although not optimal) with a summer life cycle (Santos *et al.*, 2011; Godefroid *et al.*, 2016).

Our three-time lower estimation of m between LSP and LWP than between LWP and FU is consistent with the results from Leblois *et al.* (2018) who also reported reduced gene flow between LSP and winter populations. Although the restricted gene flow between LSP and LWP is consistent with the LSP population allochronic life cycle leading to reproductive isolation, it is higher than expected under a strict allochrony. This may notably be explained by phenological plasticity in the summer population. Some individuals - referred to as LateLSP - notably emerge with a winter phenology while they have a summer (LSP) genetic background. Moreover, hybrids between LSP and LWP genetic backgrounds have already been detected in nature (Burban *et al.*, 2016) and may then reproduce with either LSP or LWP individuals (see next discussion section). However, such gene flow between LSP and LWP challenges the possibility of maintaining these two genetically differentiated allochronic populations. One explanation for the maintenance of LSP despite gene flow may be the existence of a heterogeneous gene flow along the genome, with reduced gene flow at regions controlling phenology and assortative mating.

Z-linked genes associated with allochronic divergence

The identification of marked allelic differences between summer and winter populations at the Z chromosome, especially at genes associated with neurotransmission processes related to circadian rhythm regulation, is overall in line with empirical work in insects reporting Z-linked circadian genes associated with phenological variations (Kozak *et al.*, 2019; Pruisscher *et al.*, 2021; Zhang *et al.*, 2022; Tessnow *et al.*, 2022, 2025). In particular, the *NEP-2* gene, found in our highest-scoring BayPass window showing a large high F_{ST} window immediately after a putative structural variant, belongs to the family of synaptic membrane endopeptidases *neprilysin*, which are known to influence the activity of pigment dispersing factor - *PDF* (Isaac *et al.*, 2007). Indeed, Isaac et al. (2007) showed in adults *Drosophila melanogaster* that neprilysin inactivates *PDF* through cleavage at the Serine7-Leucine8 peptide bond, resulting in greatly reduced signaling at the *PDF*-receptor encoded by *pdfr* (Lear *et al.*, 2005). *PDF* is known to be a key neurotransmitter regulating rhythmic circadian locomotor behaviors by stabilizing the *period*/*timeless* heterodimer (Seluzicki *et al.*, 2014) and by stimulating *Clk*/*CYC*-dependent transcription in a time-dependent manner (Sabado *et al.*, 2017). In addition, *PDF* synchronizes organismal activity with the annual cycle of seasons via *pdfr*, with potential epistatic control of diapause termination occurring through the release of ecdysone, as discovered in the silkmoth *Bombyx mori* (Iga *et al.*, 2014; Kozak *et al.*, 2019). Interestingly, *PDF* abundance in small ventral lateral neurons (s-LNvs) of *Drosophila melanogaster* is lower on cold and short days compared to warm and long days, indicating an effect of both photoperiod and temperature on *PDF*, with the effect primarily attributed to temperature (Hidalgo *et al.*, 2023; Hamanaka *et al.*, 2024). Hence, our result may suggest a genetic polymorphism between LSP and LWP regulating the expression of *NEP-2* as a putative modifier of diapause timing, a hypothesis that requires further investigation. The potential role of *NEURL4* may be more indirect given that *NEURL4* has been shown to be involved in cellular metabolism regulation (Liu and Boulianne, 2017). Since it is located immediately upstream *NEP-2*, both genes could be considered as a functional module, with *NEURL4* delivering cellular energy delivered for *NEP-2* neprilysin production. Hence, a genetic polymorphism modifying the expression of *NEURL4*, or more likely of both *NEURL4* and *NEP-2*, in LSP could alter diapause timing. Alternatively, a genetic polymorphism within the *NEURL4* gene in LSP, modifying cellular energy regulation, could alter *NEP-2* neprilysin production and hence diapause timing. Other genes - *NPAS2*, *GABAR*, *Coq8* and *SP2353* - potentially involved in circadian or biological rhythm were also detected in significant, although lower, C2 local score windows. *NPAS2* is a paralog of the *Clk* gene, can functionally replace *Clk* to regulate circadian rhythms in mammals (DeBruyne *et al.*, 2007; Peng *et al.*, 2021). Furthermore, *NPAS2* pairs with *BMAL1* to form a heterodimer, and together they target the transcription of *period* and *CRY* genes, both crucial for regulating the circadian clock in mammals (Vitaterna *et al.*, 1994).*GABA* receptors have been shown to affect diapause hormone levels in the silkworm *Bombyx mori* (Cui *et al.*, 2021), and disrupting

the *period* gene in silkworms influences feedback regulation of the *GABA* pathway by modifying the circadian clock pathway's signal output. The *Coq8* gene was found to affect neuronal phenotypes in *Drosophila*, especially the development and maintenance of photoreceptors (Hura *et al.*, 2022). The *SP2353* gene was found to cause abnormal photoreceptor axon pathfinding and/or differentiation phenotypes in *Drosophila* (Marrone *et al.*, 2011). Lastly, the *period* gene, immediately upstream of a high C2 local score window in our study, is well known to influence seasonal timing in Lepidoptera (Sandrelli *et al.*, 2007), notably affecting egg hatching (Ikeda *et al.*, 2019, 2021; Nartey *et al.*, 2021) and being associated with the induction of photoperiodic diapause as well as its termination (Kozak *et al.*, 2019; Pruisscher *et al.*, 2018, 2021; Nelson *et al.*, 2022).

From an evolutionary perspective, the location on the Z chromosome of genetic polymorphisms putatively controlling biological rhythms may be particularly important. Indeed, given the reduced N_e of sex chromosomes compared to autosomes (driven by heterogametic ZW females), a faster evolution of these chromosomes is predicted (*Fast-Z evolution* - Vicoso and Charlesworth (2006)). Faster-Z likely explains increased genetic differentiation and reduced diversity at the Z chromosome compared to autosomes, in all the populations of our study, as well as in other Lepidoptera (Meisel and Connallon, 2013; Mongue *et al.*, 2022). However, the over-differentiation of the Z chromosome in the summer population was much higher relative to genome-wide differentiation than in the winter populations, indicating more differentiation in LSP than solely expected *via* the Faster-Z effect. The fact that D_{XY} values were low between summer and winter populations along the Z chromosome suggests that the over-differentiation Z of LSP arose recently from a common local gene pool. Importantly, the Z of LSP showed a nearly four-fold reduction of nucleotide diversity compared to autosomes, which was twice the nearly two-fold reduction observed in other populations, suggesting a dramatic bottleneck on the Z of LSP, causing its over-differentiation. Overall, the most parsimonious explanation of such reduced diversity and increased differentiation of the Z in LSP may be a strong founder effect caused by non-selective assortative mating driven by high effect Z-linked mutations, for example linked to *NEP-2*, underlying the phenological shift of LSP, together with heterogeneous gene flow along the genome (low gene flow at autosomes and almost null gene flow at the Z chromosome). In contrast, a role of divergent selection *a fortiori* as a trigger of allochrony seems unlikely given the small N_e in LSP and in particular at the time of LSP split, especially so in sympatry. Nevertheless, after the onset of the allochronic divergence, during population growth after bottleneck, positive selection might have acted in LSP on several genes implicated in the several adaptations theoretically needed to fit its new phenology (Jarrett *et al.*, 2025). This expectation might be in line with our finding of numerous genes associated to phenology, distributed across the Z chromosome and with large variations in c2 local scores, resembling a polygenic signature of response to positive selection in LSP. Lastly, a biased sex ratio towards females in LSP during its founding event could also have contributed to the reduction of genetic diversity on the Z

of this population. This hypothesis is however difficult to test using genetic data given the strong genomic footprints of selection on the Z chromosome in LSP which likely drove an extreme reduction of genetic diversity at the entire Z.

Yet, the molecular nature, age, and contribution of each of the putative genetic variants associated with the allochronic shift in LSP require further investigation. In addition, analyzing INDELs, CNVs, and inversions for both LSP and LWP based on long read sequencing data (Wadsworth *et al.*, 2015; Bastide *et al.*, 2022) would help to characterize putative structural variants between and better infer point mutations. RNAseq or RNAi studies of the aforementioned genes throughout different life stages and seasons would also validate their potential roles. Simulations in SLIM (Haller and Messer, 2023) could also help to better understand the dynamics of the evolution of these Z-linked genetic variants associated with allochrony. Lastly, laboratory crosses between LSP and LWP and genotyping of recombinant families, although very challenging, would probably allow us to pinpoint more precisely loci underlying the phenological variations between both populations (Branco *et al.*, 2017).

Extent of inbreeding and genetic load in the allochronic summer population

Genome-wide inbreeding levels in LSP, LWP and CA were high compared to other populations sampled in Portugal, likely resulting from the aforementioned bottlenecks of higher intensity in the planted forests between Lisbon and the Mata Nacional de Leiria. These high inbreeding levels were comparable to those observed in critically endangered vertebrate species (Robinson *et al.*, 2016, 2018; Grossen *et al.*, 2020; Dussex *et al.*, 2023; Kardos, 2023) or invasive insect species (Taylor *et al.*, 2024a). The relatively short ROH segments found (mostly between 100 kb and 1 Mb, with very few exceeding 1 Mb) correspond to a size range commonly reported in Lepidoptera species with low N_e or limited dispersal (Bortoluzzi *et al.*, 2023). The difference in inbreeding levels between, on one hand, LSP, LWP, and CA and, on the other hand, the other populations, is comparable to what was found between populations of the Lepidoptera *Satyrium semiluna* (MacDonald *et al.*, 2025), with different levels of spatial isolation. Overall, these results confirm a peculiar history of PPM coastal populations (represented here by LSP, LWP, and CA), which experienced stronger bottlenecks and stronger inbreeding compared to more inland Portuguese populations.

In each of the three populations for which we studied genetic load - LSP, LWP, FU - we found signals consistent with past selection acting on both the realized and masked load on autosomes, as reflected by lower proportions of derived homozygote and heterozygous genotypes at LOF sites compared to MIS sites, and to MIS sites compared to SYN sites. The fact that the masked load - measured through the proportion of heterozygous non-synonymous sites - decreased from FU to LWP to LSP is likely attributable to genetic drift, which reduces overall genetic diversity under strong

bottlenecks. An opposite pattern was found for the realized load - estimated via the proportion of derived homozygote non-synonymous sites - which decreased from LSP to LWP to FU. This likely reflects that drift not only reduces diversity but also promotes fixation of deleterious alleles, especially in small or inbred populations. On the other hand, Rxy allowed us to assess the relative excess of deleterious alleles compared to a neutral expectation. The high Rxy values for LOF sites in LSP (compared to LWP and FU) and in LWP (compared to FU) suggest that drift alone is insufficient to explain these patterns. This overall signal of relative accumulation of deleterious mutations in LSP and LWP compared to FU, and in LSP compared to LWP, likely reflects that selection was less efficient in purging the genetic load in LSP and LWP compared to FU and in LSP compared to LWP. This pattern may reflect that effective population sizes in LSP and LWP became too small to sustain efficient purifying selection, leading to a reduced ability to eliminate deleterious variants. The greater accumulation of genetic load in LSP might be explained by even smaller N_e and lower m compared to LWP. These results align with theoretical predictions that slightly deleterious variants (MIS) and masked genetic load tend to accumulate under relaxed purifying selection in small or bottlenecked populations (Ohta, 1992; Keller, 2002; Charlesworth and Eyre-Walker, 2007; Charlesworth, 2009). However, our results contrast with some empirical studies reporting a deficit of strongly deleterious mutations in small or inbred populations, particularly in endangered species or populations that have experienced severe bottlenecks (Grossen *et al.*, 2020; Kleinman-Ruiz *et al.*, 2022; Bourgeois *et al.*, 2024; Taylor *et al.*, 2024b; Lombaert *et al.*, 2025), but align with other empirical studies showing no purging or accumulation of strongly deleterious mutations (Femerling *et al.*, 2023; Zeitler *et al.*, 2023). One may wonder whether the high proportion of homozygous derived genotypes at strongly deleterious sites could have reduced the fitness of LSP, potentially threatening the long-term persistence of this allochronic population. Answering this question may require rearing LSP and LWP in a common garden and inspecting proxies of individual fitness. In turn, the signal of genetic purge detected at strongly deleterious mutations on the Z chromosome in LSP only may be due to the aforementioned particularly strong bottleneck that might have occurred at this chromosome in LSP. This result is also in line with the high efficiency of purifying selection on strongly deleterious recessive mutations on sex chromosomes since they are hemizygous in ZW genotypes, as shown in (Rousselle *et al.*, 2016).

Conclusion

Several of our results are in line with the theoretical expectation that small N_e, assortative mating between individuals sharing similar reproductive time, and polymorphisms at genes controlling for phenology may promote the onset of allochronic divergence (Devaux and Lande, 2008; Kunkel, 2023). In particular, we found evidence for recent strong bottlenecks in LSP and in LWP and in another nearby population, suggesting that a regional environmental driver was causative of the genetic

drift and inbreeding that may then have allowed allochronic divergence in LSP. Strong genetic differentiation between LSP and LWP in the vicinity of circadian genes on the Z chromosome was probably underlying phenological differences and associated with strong assortative mating. The localization of these genetic polymorphisms on the Z chromosome has likely been an important factor facilitating the divergence between LSP and LWP, further limiting gene swamping from the winter populations in the summer population. Further work is required to characterize in depth the molecular nature, age, and role of mutations underlying allochrony in LSP.

Data availability statement

All raw data (fastq files for Pool-Seq and Ind-Seq data) will be available in the SRA repository BioProject: PRJNA1269716 (publicly available upon paper publication). Scripts will be available at https://github.com/TanguyMuller/Genomic_of_Allochrony.

Acknowledgments and funding

The authors wish to thank the ANR for a collaborative grant (LOADEXP; ANR-22-CE02-0009) to C.P. which covered half of sequencing costs and half PhD scholarship to T.M.; funding from INRAE to C.P. covering half of the sequencing costs and half of a PhD scholarship; the MGX-GENOMICS platform for sequencing; the Genotoul Bio-informatic platform. We are grateful to A. Ducasse, J. Moulin, R. Vitalis, V. Danneville, J.M. Gigleux, P. Doinjashvili, C. Robin, for helping with administrative tasks; P. Nouhaud, M. De Navascues, A. Fraimout, and M. Uhl for discussions regarding bioinformatic and statistical analyses; K. Ipekdal (Hacettepe University, Ankara, Türkiye), N. Rahim (Higher National School of Biotechnology Taoufik Khaznadar, Constantine, Algeria) and F. Goussard (URZF, Orléans, France) for sharing samples of *T. wilkinsoni*, *T. bonjeani* and *T. pinivora*, respectively, M.L. Ben Jamâa (Institut National Agronomique de Tunis, Tunisia) for samples from Tunisia, and J. Hodar (University of Granada, Spain) for samples from Spain.

Bibliography

Abruzzi, K. C., Zadina, A., Luo, W., Wiyanto, E., Rahman, R., Guo, F., Shafer, O., and Rosbash, M. 2017. RNA-seq analysis of *Drosophila* clock and non-clock neurons reveals neuron-specific cycling and novel candidate neuropeptides. *PLOS Genetics*, 13(2): e1006613.

Aguiar, F. C., Rodrigues, C., Pina, J. P., and Soares, P. 2021. Regeneration of Riparian and Maritime Pine Forests after a Large Wildfire on the Largest Public Forest of Portugal. *Forests*, 12(4): 477.

Alexander, D. H., Novembre, J., and Lange, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9): 1655–1664.

Andersen, A., Simcox, D. J., Thomas, J. A., and Nash, D. R. 2014. Assessing reintroduction schemes by comparing genetic diversity of reintroduced and source populations: A case study of the globally threatened large blue butterfly (*Maculinea arion*). *Biological Conservation*, 175: 34–41.

Bastide, H., López-Villavicencio, M., Ogereau, D., Lledo, J., Dutrillaux, A.-M., Debat, V., and Llaurens, V. 2022. Genome assembly of 3 Amazonian *Morpho* butterfly species reveals Z-chromosome rearrangements between closely related species living in sympatry. *GigaScience*, 12: giad033.

Baumdicker, F., Bisschop, G., Goldstein, D., Gower, G., Ragsdale, A. P., Tsambos, G., Zhu, S., Eldon, B., Ellerman, E. C., Galloway, J. G., Gladstein, A. L., Gorjanc, G., Guo, B., Jeffery, B., Kretzschumar, W. W., Lohse, K., Matschiner, M., Nelson, D., Pope, N. S., Quinto-Cortés, C. D., Rodrigues, M. F., Saunack, K., Sellinger, T., Thornton, K., Van Kemenade, H., Wohns, A. W., Wong, Y., Gravel, S., Kern, A. D., Koskela, J., Ralph, P. L., and Kelleher, J. 2022. Efficient ancestry and mutation simulation with msprime 1.0. *Genetics*, 220(3): iyab229.

Bertrand, A. R., Kadri, N. K., Flori, L., Gautier, M., and Druet, T. 2019. RZooRoH: An R package to characterize individual genomic autozygosity and identify homozygous-by-descent segments. *Methods in Ecology and Evolution*, 10(6): 860–866.

Bortoluzzi, C., Wright, C. J., Lee, S., Cousins, T., Genez, T. A., Thybert, D., Martin, F. J., Haggerty, L., of Life Project Consortium, D. T., Blaxter, M., *et al.* 2023. Lepidoptera genomics based on 88 chromosomal reference sequences informs population genetic parameters for conservation. *bioRxiv*, pages 2023–04.

Bourgeois, Y., Warren, B. H., and Augiron, S. 2024. The burden of anthropogenic changes and mutation load in a critically endangered harrier from the Reunion biodiversity hotspot, *Circus maillardi*. *Molecular Ecology*, 33(6): e17300.

Brady, D., Saviane, A., Cappellozza, S., and Sandrelli, F. 2021. The Circadian Clock in Lepidoptera. *Frontiers in Physiology*, 12: 776826.

Branco, M., Paiva, M., Santos, H. M., Burban, C., and Kerdelhué, C. 2017. Experimental evidence for heritable reproductive time in 2 allochronic populations of pine processionary moth. *Insect Science*, 24(2): 325–335.

Burban, C., Gautier, M., Leblois, R., Landes, J., Santos, H., Paiva, M.-R., Branco, M., and Kerdelhue, C. 2016. Evidence for low-level hybridization between two allochronic populations of the pine processionary moth, *Thaumetopoea pityocampa* (Lepidoptera: Notodontidae). *Biological Journal of the Linnean Society*, 119(2): 311–328.

Burban, C., Rocha, S., Leblois, R., Rossi, J.-P., Sauné, L., Branco, M., and Kerdelhué, C. 2020. From sympatry to parapatry: a rapid change in the spatial context of incipient allochronic speciation. *Evolutionary Ecology*, 34(1): 101–121.

Camus, L., Rode, N., Serga, S., Loiseau, A., Chen, X., Iampietro, C., Kenis, M., Marande, W., Mensch, J., Parinello, H., *et al.* 2024. New insights into past and future invasion of *Drosophila suzukii* using novel genomic resources and statistical methods to combine individual and pool sequencing data. *bioRxiv*, pages 2024–10.

Charlesworth, B. 2009. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3): 195–205.

Charlesworth, J. and Eyre-Walker, A. 2007. The other side of the nearly neutral theory, evidence of slightly advantageous back-mutations. *Proceedings of the National Academy of Sciences*, 104(43): 16992–16997.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17): i884–i890.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., and Ruden, D. M. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w ^{1118} ; iso-2; iso-3. *Fly*, 6(2): 80–92.

Collin, F., Durif, G., Raynal, L., Lombaert, E., Gautier, M., Vitalis, R., Marin, J., and Estoup, A. 2021. Extending approximate Bayesian computation with supervised machine learning to infer demographic history from genetic polymorphisms using DIYABC Random Forest. *Molecular Ecology Resources*, 21(8): 2598–2613.

Coyne, J. A. and Orr, H. A. 2004. *Speciation*, pages 545–545. Sinauer Associates, Inc. Publ, Sunderland, Massachusetts.

Cui, W.-Z., Qiu, J.-F., Dai, T.-M., Chen, Z., Li, J.-L., Liu, K., Wang, Y.-J., Sima, Y.-H., and Xu, S.-Q. 2021. Circadian Clock Gene Period Contributes to Diapause via GABAeric-Diapause Hormone Pathway in *Bombyx mori*. *Biology*, 10(9): 842.

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., and 1000 Genomes Project Analysis Group 2011. The variant call format and VCFtools. *Bioinformatics*, 27(15): 2156–2158.

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., and Li, H. 2021. Twelve years of SAMtools and BCFtools. *GigaScience*, 10(2): giab008.

DeBruyne, J. P., Weaver, D. R., and Reppert, S. M. 2007. CLOCK and NPAS2 have overlapping roles in the suprachiasmatic circadian clock. *Nature Neuroscience*, 10(5): 543–545.

Devaux, C. and Lande, R. 2008. Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proceedings of the Royal Society B: Biological Sciences*, 275(1652): 2723–2732.

Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S., and Reich, D. 2015. No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nature Genetics*, 47(2): 126–131.

Dong, S.-S., He, W.-M., Ji, J.-J., Zhang, C., Guo, Y., and Yang, T.-L. 2021. LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Briefings in Bioinformatics*, 22(4): bbaa227.

Druet, T. and Gautier, M. 2017. A model-based approach to characterize individual inbreeding at both global and local genomic scales. *Molecular Ecology*, 26(20): 5820–5841.

Druet, T. and Gautier, M. 2022. A hidden Markov model to estimate homozygous-by-descent probabilities associated with nested layers of ancestors. *Theoretical Population Biology*, 145: 38–51.

Dussex, N., Van Der Valk, T., Morales, H. E., Wheat, C. W., Díez-del Molino, D., Von Seth, J., Foster, Y., Kutschera, V. E., Guschanski, K., Rhie, A., Phillippy, A. M., Korlach, J., Howe, K., Chow, W., Pelan, S., Mendes Damas, J. D., Lewin, H. A., Hastie, A. R., Formenti, G., Fedrigo, O., Guhlin, J., Harrop, T. W., Le Lec, M. F., Dearden, P. K., Haggerty, L., Martin, F. J., Kodali, V., Thibaud-Nissen, F., Iorns, D., Knapp, M., Gemmell, N. J., Robertson, F., Moorhouse, R., Digby, A., Eason, D., Vercoe, D., Howard, J., Jarvis, E. D., Robertson, B. C., and Dalén, L. 2021. Population genomics of the critically endangered kākāpō. *Cell Genomics*, 1(1): 100002.

Dussex, N., Tørresen, O. K., Van Der Valk, T., Le Moullec, M., Veiberg, V., Tooming-Klunderud, A., Skage, M., Garmann-Aarhus, B., Wood, J., Rasmussen, J. A., Pedersen, , Martin, S. L., Røed, K. H., Jakobsen, K. S., Dalén, L., Hansen, B. B., and Martin, M. D. 2023. Adaptation to the High-Arctic island environment despite long-term reduced genetic variation in Svalbard reindeer. *iScience*, 26(10): 107811.

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., and San-Cristobal, M. 2017. Accounting for linkage disequilibrium in genome scans for selection without individual genotypes: The local score approach. *Molecular Ecology*, 26(14): 3700–3714.

Femerling, G., Van Oosterhout, C., Feng, S., Bristol, R. M., Zhang, G., Groombridge, J., P. Gilbert, M. T., and Morales, H. E. 2023. Genetic Load and Adaptive Potential of a Recovered Avian Species that Narrowly Avoided Extinction. *Molecular Biology and Evolution*, 40(12): msad256.

Feng, X., Löytynoja, A., and Merilä, J. 2023. Estimating recent and historical effective population size of marine and freshwater sticklebacks. preprint, Genomics.

Fuentes-Pardo, A. P. and Ruzzante, D. E. 2017. Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations. *Molecular Ecology*, 26(20): 5369–5406.

García-Berro, A., Talla, V., Vila, R., Wai, H. K., Shipilina, D., Chan, K. G., Pierce, N. E., Backström, N., and Talavera, G. 2023. Migratory behaviour is positively associated with genetic diversity in butterflies. *Molecular Ecology*, 32(3): 560–574.

Garrison, E. and Marth, G. 2012. Haplotype-based variant detection from short-read sequencing. Number: arXiv:1207.3907 arXiv:1207.3907 [q-bio].

Gautier, M. 2015. Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, 201(4): 1555–1579.

Gautier, M., Vitalis, R., Flori, L., and Estoup, A. 2022. f -Statistics estimation and admixture graph construction with Pool-Seq or allele count data using the R package *poolfstat*. *Molecular Ecology Resources*, 22(4): 1394–1416.

Gautier, M., Coronado-Zamora, M., and Vitalis, R. 2024a. Estimating hierarchical f–statistics from pool–seq data. *bioRxiv*, pages 2024–11.

Gautier, M., Micol, T., Camus, L., Moazami-Goudarzi, K., Naves, M., Guéret, E., Engelen, S., Lemainque, A., Colas, F., Flori, L., and Druet, T. 2024b. Genomic Reconstruction of the Successful Establishment of a Feralized Bovine Population on the Subantarctic Island of Amsterdam. *Molecular Biology and Evolution*, 41(7): msae121.

Gautier, M., Nouhaud, P., Lagnel, J., Sauné, L., Legeai, F., Parinello, H., Gschloessl, B., Marande, W., Lopez-Roques, C., Dorkeld, F., Branco, M., Perrier, C., and Kerdelhué, C. 2025. A chromosome-level assembly of the pine processionary moth (*Thaumetopoea pityocampa*) genome. *bioRxiv*.

Godefroid, M., Rocha, S., Santos, H., Paiva, M.-R., Burban, C., Kerdelhué, C., Branco, M., Rasplus, J.-Y., and Rossi, J.-P. 2016. Climate constrains range expansion of an allochronic population of the pine processionary moth. *Diversity and Distributions*, 22(12): 1288–1300.

Grossen, C., Guillaume, F., Keller, L. F., and Croll, D. 2020. Purging of highly deleterious mutations through severe bottlenecks in Alpine ibex. *Nature Communications*, 11(1): 1001.

Haller, B. C. and Messer, P. W. 2023. SLiM 4: Multispecies Eco-Evolutionary Modeling. *The American Naturalist*, 201(5): E127–E139.

Hamanaka, Y., Hasebe, M., and Shiga, S. 2024. Neural mechanism of circadian clock-based photoperiodism in insects and snails. *Journal of Comparative Physiology A*, 210(4): 601–625.

Hendry, A. P. and Day, T. 2005. Population structure attributable to reproductive time: isolation by time and adaptation by time: ISOLATION BY TIME AND ADAPTATION BY TIME. *Molecular Ecology*, 14(4): 901–916.

Hidalgo, S., Anguiano, M., Tabuloc, C. A., and Chiu, J. C. 2023. Seasonal cues act through the circadian clock and pigment-dispersing factor to control EYES ABSENT and downstream physiological changes. *Current Biology*, 33(4): 675–687.e5.

Hivert, V., Leblois, R., Petit, E. J., Gautier, M., and Vitalis, R. 2018. Measuring Genetic Differentiation from Pool-seq Data. *Genetics*, 210(1): 315–330.

Hura, A. J., Hawley, H. R., Tan, W. J., Penny, R. J., Jacobsen, J. C., and Fitzsimons, H. L. 2022. Loss of *Drosophila* Coq8 results in impaired survival, locomotor deficits and photoreceptor degeneration. *Molecular Brain*, 15(1): 15.

Hutfilz, C. 2022. Endocrine Regulation of Lifespan in Insect Diapause. *Frontiers in Physiology*, 13: 825057.

i Torres, A. P., Höök, L., Näsvall, K., Shipilina, D., Wiklund, C., Vila, R., Pruisscher, P., and Backström, N. 2022. The fine-scale recombination rate variation and associations with genomic features in a butterfly. preprint, Genomics.

Iga, M., Nakaoka, T., Suzuki, Y., and Kataoka, H. 2014. Pigment Dispersing Factor Regulates Ecdysone Biosynthesis via *Bombyx* Neuropeptide G Protein Coupled Receptor-B2 in the Prothoracic Glands of *Bombyx mori*. *PLoS ONE*, 9(7): e103239.

Ikeda, K., Daimon, T., Sezutsu, H., Udaka, H., and Numata, H. 2019. Involvement of the Clock Gene *period* in the Circadian Rhythm of the Silkmoth *Bombyx mori*. *Journal of Biological Rhythms*, 34(3): 283–292.

Ikeda, K., Daimon, T., Shiomi, K., Udaka, H., and Numata, H. 2021. Involvement of the Clock Gene Period in the Photoperiodism of the Silkmoth *Bombyx mori*. *Zoological Science*, 38(6).

Isaac, R. E., Johnson, E. C., Audsley, N., and Shirras, A. D. 2007. Metabolic inactivation of the circadian transmitter, pigment dispersing factor (PDF), by neprilysin-like peptidases in *Drosophila*. *Journal of Experimental Biology*, 210(24): 4465–4470.

Jarrett, B. J. M., Downing, P. A., and Svensson, E. I. 2025. Meta-analysis reveals that phenotypic plasticity and divergent selection promote reproductive isolation during incipient speciation. *Nature Ecology & Evolution*, 9(5): 833–844.

Kardos, M. 2023. Genomes of an endangered rattlesnake show that neutral genetic variation predicts adaptive genetic variation and genetic load. *Proceedings of the National Academy of Sciences*, 120(49): e2316880120.

Kelleher, J., Wong, Y., Wohns, A. W., Fadil, C., Albers, P. K., and McVean, G. 2019. Inferring whole-genome histories in large population datasets. *Nature Genetics*, 51(9): 1330–1338.

Keller, L. 2002. Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, 17(5): 230–241.

Kleinman-Ruiz, D., Lucena-Perez, M., Villanueva, B., Fernandez, J., Saveljev, A. P., Ratkiewicz, M., Schmidt, K., Galtier, N., Garcia-Dorado, A., and Godoy, J. A. 2022. Purging of deleterious burden in the endangered Iberian lynx. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 119(11).

Kopp, M., Servedio, M. R., Mendelson, T. C., Safran, R. J., Rodríguez, R. L., Hauber, M. E., Scordato, E. C., Symes, L. B., Balakrishnan, C. N., Zonana, D. M., and Van Doorn, G. S. 2018. Mechanisms of Assortative Mating in Speciation with Gene Flow: Connecting Theory and Empirical Research. *The American Naturalist*, 191(1): 1–20.

Korunes, K. L. and Samuk, K. 2021. pixy : Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, 21(4): 1359–1368.

Kozak, G. M., Wadsworth, C. B., Kahne, S. C., Bogdanowicz, S. M., Harrison, R. G., Coates, B. S., and Dopman, E. B. 2019. Genomic Basis of Circannual Rhythm in the European Corn Borer Moth. *Current Biology*, 29(20): 3501–3509.e5.

Kunkel, J. 2023. *Phenological Assortative Mating And The Evolution of Flowering Time*. Master's thesis, University of Toronto (Canada).

Lanfear, R., Kokko, H., and Eyre-Walker, A. 2014. Population size and the rate of evolution. *Trends in Ecology & Evolution*, 29(1): 33–41.

Lear, B. C., Merrill, C. E., Lin, J.-M., Schroeder, A., Zhang, L., and Allada, R. 2005. A G Protein-Coupled Receptor, groom-of-PDF, Is Required for PDF Neuron Action in Circadian Behavior. *Neuron*, 48(2): 221–227.

Leblois, R., Gautier, M., Rohfritsch, A., Foucaud, J., Burban, C., Galan, M., Loiseau, A., Sauné, L., Branco, M., Gharbi, K., Vitalis, R., and Kerdelhué, C. 2018. Deciphering the demographic

history of allochronic differentiation in the pine processionary moth *Thaumetopoea pityocampa*. *Molecular Ecology*, 27(1): 264–278.

Liu, C.-C., Shringarpure, S., Lange, K., and Novembre, J. 2020. Exploring Population Structure with Admixture Models and Principal Component Analysis. In J. Y. Dutheil, editor, *Statistical Population Genomics*, volume 2090, pages 67–86. Springer US, New York, NY. Series Title: Methods in Molecular Biology.

Liu, S. and Boulianne, G. L. 2017. The NHR domains of Neuralized and related proteins: Beyond Notch signalling. *Cellular Signalling*, 29: 62–68.

Lombaert, E., Blin, A., Porro, B., Guillemaud, T., Bernal, J. S., Chang, G., Kirichenko, N., Sappington, T. W., Toepfer, S., and Deleury, E. 2025. Unraveling genetic load dynamics during biological invasion: insights from two invasive insect species. *Peer Community Journal*, 5.

MacDonald, Z. G., Dupuis, J. R., Glasier, J. R. N., Sissons, R., Moehrenschlager, A., Shaffer, H. B., and Sperling, F. A. H. 2025. Whole-Genome Evaluation of Genetic Rescue: The Case of a Curiously Isolated and Endangered Butterfly. *Molecular Ecology*, 34(4): e17657.

Marrone, A. K., Kucherenko, M. M., Rishko, V. M., and Shcherbata, H. R. 2011. New Dystrophin/Dystroglycan interactors control neuron behavior in *Drosophila* eye. *BMC Neuroscience*, 12(1): 93.

Martchenko, D. and Shafer, A. B. A. 2023. Contrasting whole-genome and reduced representation sequencing for population demographic and adaptive inference: an alpine mammal case study. *Heredity*, 131(4): 273–281.

Marí-Mena, N., Naveira, H., Lopez-Vaamonde, C., and Vila, M. 2019. Census and contemporary effective population size of two populations of the protected Spanish Moon Moth (*Graellsia isabellae*). *Insect Conservation and Diversity*, 12(2): 147–160.

Matessi, C., Gimelfarb, A., and Gavrilets, S. 2002. Long-term Buildup of Reproductive Isolation Promoted by Disruptive Selection: How Far Does it Go? *Selection*, 2(1-2): 41–64.

Meisel, R. P. and Connallon, T. 2013. The faster-X effect: integrating theory and data. *Trends in Genetics*, 29(9): 537–544.

Mendes, A. 2007. The portuguese forests. Universidade Católica Portuguesa: Porto.

Meyermans, R., Gorssen, W., Buys, N., and Janssens, S. 2020. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC Genomics*, 21(1): 94.

Mongue, A. J., Hansen, M. E., and Walters, J. R. 2022. Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages*. *Evolution*, 76(2): 332–345.

Nartey, M. A., Sun, X., Qin, S., Hou, C., and Li, M. 2021. CRISPR/Cas9-based knockout reveals that the clock gene *timeless* is indispensable for regulating circadian behavioral rhythms in *Bombyx mori*. *Insect Science*, 28(5): 1414–1425.

Nelson, T. D., MacDonald, Z. G., and Sperling, F. A. H. 2022. Moths passing in the night: Phenological and genomic divergences within a forest pest complex. *Evolutionary Applications*, 15(1): 166–180.

Ohta, T. 1992. The Nearly Neutral Theory of Molecular Evolution. *Annual Review of Ecology and Systematics*, 23(1): 263–286.

Olazcuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., Diepenbrock, L. M., Kenis, M., Zhang, J., Chen, X., Borowiec, N., Facon, B., Vogt, H., Price, D. K., Vogel, H., Prud'homme, B., Estoup, A., and Gautier, M. 2020. A Whole-Genome Scan for Association with Invasion Success in the Fruit Fly *Drosophila suzukii* Using Contrasts of Allele Frequencies Corrected for Population Structure. *Molecular Biology and Evolution*, 37(8): 2369–2385.

Palahí I Torres, A., Höök, L., Näsvall, K., Shipilina, D., Wiklund, C., Vila, R., Pruisscher, P., and Backström, N. 2023. The fine-scale recombination rate variation and associations with genomic features in a butterfly. *Genome Research*, 33(5): 810–823.

Pedersen, B. S. and Quinlan, A. R. 2018. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics*, 34(5): 867–868.

Peng, L. U., Bai, G., and Pang, Y. 2021. Roles of NPAS2 in circadian rhythm and disease. *Acta Biochimica et Biophysica Sinica*, 53(10): 1257–1265.

Pimentel, C., Calvão, T., Santos, M., Ferreira, C., Neves, M., and Nilsson, J.- 2006. Establishment and expansion of a *Thaumetopoea pityocampa* (Den. & Schiff.) (Lep. Notodontidae) population with a shifted life cycle in a production pine forest, Central-Coastal Portugal. *Forest Ecology and Management*, 233(1): 108–115.

Pinho, J. R. 2012. As ciências da vegetação ea intervenção dos serviços florestais. *Gestão e conservação da flora e da vegetação de Portugal e da África Lusófona." In Honorium" do Professor Catedrático Emérito Ilídio Rosário dos Santos Moreira. ISA Press, Lisboa*, pages 217–235.

Pruisscher, P., Nylin, S., Gotthard, K., and Wheat, C. W. 2018. Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. *Molecular Ecology*, 27(18): 3613–3626.

Pruisscher, P., Nylin, S., Wheat, C. W., and Gotthard, K. 2021. A region of the sex chromosome associated with population differences in diapause induction contains highly divergent alleles at clock genes. *Evolution*, 75(2): 490–500.

Pudlo, P., Marin, J.-M., Estoup, A., Cornuet, J.-M., Gautier, M., and Robert, C. P. 2016. Reliable ABC model choice via random forests. *Bioinformatics*, 32(6): 859–866.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J., and Sham, P. C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, 81(3): 559–575.

Quinlan, A. R. and Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6): 841–842.

Raynal, L., Marin, J.-M., Pudlo, P., Ribatet, M., Robert, C. P., and Estoup, A. 2019. ABC random forests for Bayesian parameter inference. *Bioinformatics*, 35(10): 1720–1728.

Robinson, J., Ortega-Del Vecchyo, D., Fan, Z., Kim, B., vonHoldt, B., Marsden, C., Lohmueller, K., and Wayne, R. 2016. Genomic Flatlining in the Endangered Island Fox. *Current Biology*, 26(9): 1183–1189.

Robinson, J., Kyriazis, C. C., Yuan, S. C., and Lohmueller, K. E. 2023. Deleterious Variation in Natural Populations and Implications for Conservation Genetics. *Annual Review of Animal Biosciences*, 11(1): annurev–animal–080522–093311.

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., and Wayne, R. K. 2018. Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, 28(21): 3487–3494.e4.

Rousselle, M., Faivre, N., Ballenghien, M., Galtier, N., and Nabholz, B. 2016. Hemizygosity enhances purifying selection: lack of fast-z evolution in two satyrine butterflies. *Genome biology and evolution*, 8(10): 3108–3119.

Rundle, H. D. and Nosil, P. 2005. Ecological speciation. *Ecology Letters*, 8(3): 336–352.

Sabado, V., Vienne, L., Nunes, J. M., Rosbash, M., and Nagoshi, E. 2017. Fluorescence circadian imaging reveals a PDF-dependent transcriptional regulation of the *Drosophila* molecular clock. *Scientific Reports*, 7(1): 41560.

Safran, R. J., Scordato, E. S. C., Wilkins, M. R., Hubbard, J. K., Jenkins, B. R., Albrecht, T., Flaxman, S. M., Karaardıç, H., Vortman, Y., Lotem, A., Nosil, P., Pap, P., Shen, S., Chan, S., Parchman, T., and Kane, N. C. 2016. Genome-wide differentiation in closely related populations: the roles of selection and geographic isolation. *Molecular Ecology*, 25(16): 3865–3883.

Sandrelli, F., Cappellozza, S., Benna, C., Saviane, A., Mastella, A., Mazzotta, G. M., Moreau, S., Pegoraro, M., Piccin, A., Zordan, M. A., Cappellozza, L., Kyriacou, C. P., and Costa, R. 2007. Phenotypic effects induced by knock-down of the *period* clock gene in *Bombyx mori*. *Genetical Research*, 89(2): 73–84.

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., and Caballero, A. 2020. Recent Demographic History Inferred by High-Resolution Analysis of Linkage Disequilibrium. *Molecular Biology and Evolution*, 37(12): 3642–3653.

Santiago, E., Köpke, C., and Caballero, A. 2025. Accounting for population structure and data quality in demographic inference with linkage disequilibrium methods. *Nature Communications*, 16(1): 6054.

Santos, H., Rousselet, J., Magnoux, E., Paiva, M.-R., Branco, M., and Kerdelhué, C. 2007. Genetic isolation through time: allochronic differentiation of a phenologically atypical population of the pine processionary moth. *Proceedings of the Royal Society B: Biological Sciences*, 274(1612): 935–941.

Santos, H., Burban, C., Rousselet, J., Rossi, J.-P., Branco, M., and Kerdelhué, C. 2011. Incipient allochronic speciation in the pine processionary moth (*Thaumetopoea pityocampa*, Lepidoptera, Notodontidae): Allochronic speciation in the PPM. *Journal of Evolutionary Biology*, 24(1): 146–158.

Santos, H. M., Paiva, M., Rocha, S., Kerdelhué, C., and Branco, M. 2013. Phenotypic divergence in reproductive traits of a moth population experiencing a phenological shift. *Ecology and Evolution*, 3(15): 5098–5108.

Seluzicki, A., Flourakis, M., Kula-Eversole, E., Zhang, L., Kilman, V., and Allada, R. 2014. Dual PDF Signaling Pathways Reset Clocks Via TIMELESS and Acutely Excite Target Neurons to Control Circadian Behavior. *PLoS Biology*, 12(3): e1001810.

Shipilina, D., Näsvall, K., Höök, L., Vila, R., Talavera, G., and Backström, N. 2022. Linkage mapping and genome annotation give novel insights into gene family expansions and regional recombination rate variation in the painted lady (*Vanessa cardui*) butterfly. *Genomics*, 114(6): 110481.

Tan, A., Abecasis, G. R., and Kang, H. M. 2015. Unified representation of genetic variants. *Bioinformatics*, 31(13): 2202–2204.

Taylor, B. A., Tembrock, L. R., Sankovitz, M., Wilson, T. M., Looney, C., Takahashi, J., Gilligan, T. M., Smith-Pardo, A. H., and Harpur, B. A. 2024a. Population genomics of the invasive Northern

Giant Hornet Vespa mandarinia in North America and across its native range. *Scientific Reports*, 14(1): 10803.

Taylor, R. S. and Friesen, V. L. 2017. The role of allochrony in speciation. *Molecular Ecology*, 26(13): 3330–3342.

Taylor, R. S., Manseau, M., Keobouasone, S., Liu, P., Mastromonaco, G., Solmundson, K., Kelly, A., Larter, N. C., Gamberg, M., Schwantje, H., Thacker, C., Polfus, J., Andrew, L., Hervieux, D., Simmons, D., and Wilson, P. J. 2024b. High genetic load without purging in caribou, a diverse species at risk. *Current Biology*, page S0960982224001441.

Tessnow, A. E., Raszick, T. J., Porter, P., and Sword, G. A. 2022. Patterns of genomic and allochronic strain divergence in the fall armyworm, *Spodoptera frugiperda* (J.E. Smith). *Ecology and Evolution*, 12(3): e8706.

Tessnow, A. E., Nagoshi, R. N., Meagher, R. L., Gilligan, T. M., Sadd, B. M., Carrière, Y., Davis, H. N., Fleischer, S. J., Richers, K., Palumbo, J. C., Porter, P., Verle Rodrigues, J. C., and Sword, G. A. 2025. Genomic patterns of strain-specific genetic structure, linkage, and selection across fall armyworm populations. *BMC Genomics*, 26(1): 116.

Tomioka, K. and Matsumoto, A. 2015. Circadian molecular clockworks in non-model insects. *Current Opinion in Insect Science*, 7: 58–64.

Unbehend, M., Hänniger, S., Meagher, R. L., Heckel, D. G., and Groot, A. T. 2013. Pheromonal Divergence Between Two Strains of *Spodoptera frugiperda*. *Journal of Chemical Ecology*, 39(3): 364–376.

Vasimuddin, M., Misra, S., Li, H., and Aluru, S. 2019. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 314–324, Rio de Janeiro, Brazil. IEEE.

Vicoso, B. and Charlesworth, B. 2006. Evolution on the X chromosome: unusual patterns and processes. *Nature Reviews Genetics*, 7(8): 645–653.

Vitaterna, M. H., King, D. P., Chang, A.-M., Kornhauser, J. M., Lowrey, P. L., McDonald, J. D., Dove, W. F., Pinto, L. H., Turek, F. W., and Takahashi, J. S. 1994. Mutagenesis and Mapping of a Mouse Gene, *Clock* , Essential for Circadian Behavior. *Science*, 264(5159): 719–725.

Wadsworth, C. B., Li, X., and Dopman, E. B. 2015. A recombination suppressor contributes to ecological speciation in OSTRINIA moths. *Heredity*, 114(6): 593–600.

Weis, A. E., Nardone, E., and Fox, G. A. 2014. The strength of assortative mating for flowering date and its basis in individual variation in flowering schedule. *Journal of Evolutionary Biology*, 27(10): 2138–2151.

Whiting, J. R. 2022. Jimwhiting91/genotype_plot: Genotype plot.

Woolfit, M. 2009. Effective population size and the rate and pattern of nucleotide substitutions. *Biology Letters*, 5(3): 417–420.

Xue, Y., Prado-Martinez, J., Sudmant, P. H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., Chen, Y., Yngvadottir, B., Cooper, D. N., de Manuel, M., Hernandez-Rodriguez, J., Lobon, I., Siegismund, H. R., Pagani, L., Quail, M. A., Hvilsom, C., Mudakikwa, A., Eichler, E. E., Cranfield, M. R., Marques-Bonet, T., Tyler-Smith, C., and Scally, A. 2015. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science*, 348(6231): 242–245.

Zeitler, L., Parisod, C., and Gilbert, K. J. 2023. Purging due to self-fertilization does not prevent accumulation of expansion load. *PLOS GENETICS*, 19(9).

Zhang, J., Zhang, F., Tay, W. T., Robin, C., Shi, Y., Guan, F., Yang, Y., and Wu, Y. 2022. Population genomics provides insights into lineage divergence and local adaptation within the cotton bollworm. *Molecular Ecology Resources*, 22(5): 1875–1891.