INTRODUCTION

Organisms in temperate regions often experience a radically different environment depending on the time of year, and face a major evolutionary challenge in adapting their life cycles to this seasonality. Strong seasonality will favour genotypes that execute particular life cycle events (e.g., mating; migration; entering or exiting dormancy) at the appropriate times of year. Because the seasonal cycle is not the same everywhere, but varies with local climate, optimal life cycles also vary geographically. Hence, seasonality provides an opportunity to study divergent adaptation across populations (Bradshaw et al., 2004; Conover & Schultz, 1995; Posledovich et al., 2015; Stinchcombe et al., 2004).

In insects, a central life history trait is voltinism, the number of generations produced per year. This includes high-latitude species in previously glaciated areas, meaning that divergent selection on life cycle traits has taken place during or shortly after recent colonization. Here, we use a population genomics approach to compare a set of nine Scandinavian populations of the butterfly Pararge aegeria that differ in life cycle traits (diapause thresholds and voltinism) along both north–south and east–west clines. Using a de novo-assembled genome, we reconstruct colonization histories and demographic relationships. Based on the inferred population structure, we then scan the genome for candidate loci showing signs of divergent selection potentially associated with population differences in life cycle traits. The identified candidate genes include a number of components of the insect circadian clock (timeless, timeless2, period, cryptochrome and clockwork orange). Most notably, the gene timeless, which has previously been experimentally linked to life cycle regulation in P. aegeria, is here found to contain a novel 97-amino acid deletion unique to, and fixed in, a single population. These results add to a growing body of research framing circadian gene variation as a potential mechanism for generating local adaptation of life cycles.

KEYWORDS

circadian clock, diapause, insect, population genomics, timeless, voltinism
generations per year is a nontrivial difference, as it is associated with drastic differences in generation time (Roff, 1980), and may necessitate or drive associated adaptations in terms of regulation of diapause induction (Bean et al., 2012; Lindestad et al., 2019), life history traits such as development rate and body size (Fischer & Fiedler, 2002; Masaki, 1972; Mousseau & Roff, 1989), and developmental pathway expression (Aalberg Haugen & Gotthard, 2015; Kivelä et al., 2013). Nonetheless, many insect species differ geographically in voltinism, typically in relation to the local length of the warm season (Braune et al., 2008; Faccoli & Bernardinelli, 2014; Hart et al., 1997; Hu et al., 2012; Levy et al., 2015).

Expression of a particular local voltinism is often tied to photoperiodism; that is, how the population responds to differences or changes in daylength. The most widely described regulatory mechanism underlying differences in voltinism is local adaptation of the critical daylength (CD) below which diapause is induced (Bean et al., 2012; Grevstad & Coop, 2015; Lindestad et al., 2019). CD tends to be locally adapted: in particular, a positive correlation with latitude has been recorded across a diverse set of taxa (Hut et al., 2013). Generally speaking, a lower CD means that diapause will be induced later in the year, extending the insect’s active season, which allows additional generations to take place in the same year. However, because voltinism expression depends on the interplay of several environmental and physiological factors (Grevstad & Coop, 2015), there is not a simple correspondence between a population’s CD and its voltinism.

Genetic studies have shown that interpopulation differences in diapause induction are generated by a combination of variation at loci of small and large effects (Ragland et al., 2019). In several cases (Dalla Benetta et al., 2019; Levy et al., 2015; Pruisscher et al., 2018; Tauber et al., 2007; Yamada & Yamamoto, 2011), the detected large-effect loci are components of the circadian clock, which exhibit a widely recognized, but controversial and poorly understood, link with photoperiodism across many insect species (Koštál, 2011; Meuti & Denlinger, 2013). To determine how widespread this association between life cycle adaptation and circadian gene variation is, there is a need for broad analyses that survey genome-wide variation potentially associated with these traits. Such analyses will also be aided by a better understanding of the historical and nonselective genetic backgrounds against which selection on life cycle traits has been acting.

The present study explores genetic variation in the speckled wood (Pararge aegeria), a woodland-associated satyrine butterfly that varies greatly in voltinism across its pan-European range. The northernmost P. aegeria populations are univoltine; southern populations have many overlapping generations per year and appear not to diapause at all (Nylin et al., 1995). Common-garden studies have demonstrated that this life cycle variation is to a large extent generated by local adaptation of photoperiodic plasticity (Lindestad et al., 2019; Nylin et al., 1995). In mainland Scandinavia, P. aegeria populations shift from univoltine in central Sweden to bivoltine in southern Sweden and Denmark (Figure 1a). The southern Swedish populations appear to be recently established, as no record of them exists before the 1930s, and they are still separated from the central Swedish populations by a distribution gap wherein P. aegeria is rare or absent (Eliasson et al., 2005; Nordström, 1955). Historical modelling from microsatellite data shows that both central Swedish and south Swedish populations probably immigrated from the south via Denmark (Tison et al., 2014), which would mean that south Swedish populations went extinct or nearly extinct after the initial northward colonization, and later rebounded either from local stock or through a second migration event via Denmark. In addition, bivoltine populations exist on the large Baltic islands of Öland and Gotland, on the same latitude (and hence experiencing the same photoperiod) as some of the univoltine mainland populations. The Öland population is genetically depauperate, while the Gotland population maintains quite high genetic diversity (Tison et al., 2014). The colonization histories of both island populations are unknown.

Previously, a study of genome-wide differentiation patterns between a univoltine population (Sundsvall) and a bivoltine population (North Skåne) at either end of the range of P. aegeria in Sweden (Pruisscher et al., 2018) identified 15 genomic regions showing signs of divergent selection. Single nucleotide polymorphism (SNP) genotypes in some of these regions—including nonsynonymous substitutions in two circadian loci, period and timeless—also correlated with diapause incidence in a laboratory test of interpopulation hybrids, suggesting that variation at these candidate loci contributes to maintaining voltinism variation across the studied geographical region by affecting photoperiodic plasticity. Furthermore, genotyping of three additional populations showed that thecline in voltinism from north to south is accompanied by a gradient of allele frequencies at the identified candidate loci.

However, these previous results cannot explain all of the life cycle variation seen in the studied populations. The clearest indication of this is the Gotland island population, which is bivoltine and exhibits a relatively low CD (Aalberg Haugen & Gotthard, 2015; Lindestad et al., 2019), despite closely resembling the univoltine, high-CD mainland populations in terms of variation at the 15 candidate loci (Pruisscher et al., 2018). In other words, if there is indeed a causal link between the identified candidate SNPs and diapause incidence in the studied populations, there must in addition be a significant amount of genetic variation modifying this trait that remains to be identified. Extending the genome-wide search to include more intermediate-latitude populations, especially populations with contrasting voltinism and different colonization histories, should help identify such variation.

Here we build on these earlier results in several ways. First, we assemble a high-quality genome for P. aegeria in order to improve our analyses. Second, we use whole-genome resequencing to reconstruct phylogeographical relationships between nine Scandinavian P. aegeria populations, and infer historical gene flow events. Third, we scan the populations for genomic regions showing signs of divergent selection potentially related to variation in life cycles, across both north–south and island–mainland voltinism clines. Our findings contribute to a more complete picture of the selective and historical processes underlying variation in insect life cycles, including
potential contribution from novel variation in circadian genes that may help resolve previous incongruities in this study system.

2 | MATERIALS AND METHODS

2.1 | Genome assembly and annotation

High-molecular-weight DNA was extracted from a single Pararge aegeria female from Skåne in southern Sweden, and used for constructing a 10× Genomics Chromium Genome library. Library preparation, sequencing and genome assembly were performed at the National Genomics Infrastructure, SciLifeLab, Stockholm, Sweden. Genome assembly was conducted using Supernova 1.2 with the phase option enabled; 12 separate assemblies were carried out, each using a different proportion of the 10× sequencing data (from 15% to 100%). Each generated assembly was assessed for two metrics: basic assembly stats, using Quast 4.0 (Gurevich et al., 2013), and genic content, using BUSCO version 3 (Waterhouse et al., 2018) with the database eukaryota_odb9. The genome assembly deemed optimal according to these metrics was then scaffolded using a 5-kb insert mate-pair library generated from the same Pararge aegeria individual, after quality filtering and mate-pair specific filtering using NextClip version 1.3 (Leggett et al., 2014), using the scaffolding software BEST version 2.0 (Sahlin et al., 2014). The genome was polished using genomic data from a single individual via the software Pilon version 1 (Walker et al., 2014). Next, the genome was collapsed to a haploid copy using Haplomerger2 version 20180603 (Huang et al., 2017), and again assessed for genic content using BUSCO, this time with the database insecta_odb9. Then, to improve gene model regions, the
assembly was super-scaffolded using the proteome from the Bicyclus anynana assembly by Nowell et al. (2017). This proteome (Bicyclus._anynana_nBa.0.1_-proteins.faa.gz) was downloaded from Lepbase (Challis et al., 2016). Superscaffolding was carried out using MESPA (Neethiraj et al., 2017), which is a wrapper for the high-performing protein-to-genome aligner SPALN2 (Iwata & Gotoh, 2012). Finally, the genome was annotated using the B. anynana proteome, as well as a pre-existing RNA sequencing (RNAseq) data set from P. aegeria, by running BRAKER2 (Hoff et al., 2016; Stanke et al., 2006, 2008) on an assembly that had been soft-masked for repetitive content using RED (Girgis, 2015). Functional annotation was generated using the EggNOGv5 online server with default settings (Huerta-Ceperas et al., 2019).

### 2.2 Pooled resequencing of populations

DNA for pool-sequencing was extracted from 201 P. aegeria individuals sampled from eight populations across Sweden and Denmark (Figure 1a; Table 1). Most individuals (n = 168) were wild adults caught in 2014–2016; of these, 12 female butterflies could not be used, but as each female was already mated when caught, one offspring per female was used instead. In order to improve sample size for two of the populations, Öland and Stockholm, 33 individuals were added that had been collected in 2010–2011 for a previous study, using similar extraction, pooling, and sequencing methods (Pruisscher et al., 2018). The Sundsvall data set was generated from 22 adult P. aegeria individuals caught in 2011 for a previous study, using similar extraction, pooling and sequencing methods (Pruisscher et al., 2018).

#### TABLE 1 DNA sample details. Populations considered as belonging to the northern or southern phylogeographical clusters are marked (N) or (S), respectively. Pool size is the total number of individuals used; of these, some (italicized numbers in parentheses) were caught for a previous study (Tison et al., 2014). For the Sundsvall pool, DNA extraction and sequencing had also been done previously (Pruisscher et al., 2018). For populations represented by two sampling locations, coordinates for the site contributing the largest part of the sample are listed first. Sex is the number of individuals of each sex used (this information was missing for a few individuals, marked “?”). DNA used refers to the total mass of DNA of the pooled sample

| Population       | Voltinism    | Coordinates            | Pool size | Sex (M/F/?) | Sampling years | DNA used (µg) |
|------------------|--------------|------------------------|-----------|-------------|----------------|---------------|
| Sundsvall (N)    | Univoltine   | 62.4°N, 17.5°E         | 22 (22)   | 14/8/0      | 2010, 2011     | 5.0           |
| Stockholm (N)    | Univoltine   | 59.6°N, 18.5°E         | 30 (16)   | 24/6/0      | 2010, 2015, 2016 | 4.0         |
| Småland Highlands (N) | Univoltine | 57.5°N, 14.1°E     | 22         | 15/7/0      | 2015, 2016     | 4.0           |
| Kalmar (N)       | Univoltine   | 56.9°N, 16.0°E         | 13         | 10/3/0      | 2016           | 4.0           |
| Gotland (island) | Bivoltine    | 57.4°N, 18.5°E         | 28         | 22/6/0      | 2014, 2016     | 4.0           |
| Öland (island)   | Bivoltine    | 56.6°N, 16.6°E         | 29 (17)   | 21/6/2      | 2010, 2011, 2015, 2016 | 4.0         |
| North Skåne (S)  | Bivoltine    | 56.3°N, 12.5°E         | 18         | 13/5/0      | 2015           | 4.0           |
| South Skåne (S)  | Bivoltine    | 55.6°N, 13.5°E         | 44         | 21/11/12    | 2014, 2016     | 4.5           |
| Copenhagen (S)   | Bivoltine    | 55.6°N, 12.6°E         | 17         | 15/2/0      | 2015, 2016     | 4.0           |

#### 2.3 Data set preparation

Raw sequencing files were filtered for polymerase chain reaction (PCR) duplicates using the clone_filter script of STACKS 1.21 (Catchen et al., 2013), and Illumina sequencing adapters were removed using BBduk2 (Bushnell, 2015). All nine PoolSeq read sets were mapped to the P. aegeria genome using NEXTGENMAP version 0.4.10 (Sedlazeck et al., 2013) at 90% identity. Alignments were filtered using SAMTOOLS version 1.6 (Li et al., 2009), keeping only properly paired reads with a map quality of 20 or higher.

Read depth (RD) exceeded pool size over large parts of the alignments, leading to a risk of unequal read sampling of sequenced individuals, and hence skewed estimates of nucleotide diversity (Hoban et al., 2016). Because the analyses relied on accurate diversity estimates, and for consistency across analyses, we applied a subsampling approach: reads were sampled without replacement up to a population-specific target RD. This meant that genomic positions where raw RD was lower than the target RD were discarded, and so, raw RD distributions were examined in order to pick an appropriate
target RD that allowed a majority of the data to be kept while still ensuring accurate allele frequency estimates. The chosen cutoff was the 5th percentile of raw RD for each respective population, or a minimum of 20.

Before subsampling, each alignment was converted into a population-wise pileup file using samtools. To avoid spurious SNP calling, indels were filtered out (with a 5-bp margin on each side) using scripts from the popoolation version 1.2.2 package (Kofler, Orozco-terWengel, et al., 2011). Next, also using popoolation, reads were subsampled from each pileup up to the target RD. Pileup positions in the 99th percentile of per-population RD were also excluded. The resulting nine subsampled population-wise pileups (details in Appendix S1: Table S1) served as the input for all downstream analyses. A PHRED quality score cutoff of 20 and a minimum count of 2 for SNP alleles were used in all subsequent analyses.

### 2.4 Population history and connectivity analyses

To estimate overall population differentiation, the average genome-wide fixation index \( F_{ST} \) was calculated for all possible population pairs using popoolation2 version 1.201 (Kofler et al., 2011). All nine input pileups were joined together as separate columns in a single mpileup file, which was then converted into the sync format. SNPs were called on this sync file, yielding about 7.8 million SNPs (the precise number of SNPs called in each analysis is shown in Appendix S1: Table S2). \( F_{ST} \) was calculated on the called SNP variation in nonoverlapping windows across the whole genome. The window size used was 5000 bp, as this was found to be large enough to prevent noisy results, but small enough to still capture population differences. The mean \( F_{ST} \) across all windows for each pairwise population comparison was then calculated. Here and in all other analyses, \( F_{ST} \) was calculated from allele frequencies as per Hartl and Clark (2007), as implemented by default in popoolation. Additionally, overall genetic diversity in each population was quantified by measuring nucleotide diversity (\( \pi \)) and Tajima's \( D \) across each single-population pileup, again in 5,000-bp windows, using popoolation. The \( D \) statistic measures the relationship between nucleotide diversity and the number of variable sites, and can provide information on demographic history and selection dynamics (Tajima, 1989a, 1989b). For both the \( F_{ST} \), \( \pi \) and Tajima’s \( D \) analyses, only genomic windows where at least 50% of positions met the RD requirements were analysed.

In order to investigate population history and potential gene flow, a phylogeographical analysis was conducted using treemix version 1.13 (Pickrell & Pritchard, 2012). treemix takes genome-wide allele frequency distributions as input, reconstructs a bifurcating phylogeny and identifies population pairs that share more allele frequency variation than the reconstructed tree can account for. The software then allows these incongruences to be resolved by sequentially adding inferred migration events between points on the tree, where each migration event is modelled as a unidirectional donation of a specific percentage of a population’s allelic variation. Input allele frequencies were extracted from the same nine-population sync file as for the \( F_{ST} \) analysis. treemix was run five times, each time allowing for one additional inferred migration event, from zero migration events to four. Because previous results indicate that all Swedish regions were settled from the south (Tison et al., 2014), Copenhagen was set as the root population in all runs.

### 2.5 Scanning for footprints of selection \( (F_{ST} \text{ and } \pi) \)

As a first step towards identifying genomic regions potentially involved in voltinism regulation (including, but not necessarily limited to, regulation of CD), we carried out a set of pairwise population comparisons. The aim was to find genomic regions showing (i) high differentiation between two populations of differing voltinism, and (ii) relatively low genetic diversity in at least one of the two populations, which together may indicate the past occurrence of a divergent selective sweep (Carneiro et al., 2014; Maynard Smith & Haigh, 1974; Reid et al., 2016).

Because the phylogeographical analyses (Figure 1) indicated that the four univoltine populations (Sundsvall, Stockholm, Kalmar and the Småland Highlands) were most closely related to one another, they were here treated as a single phylogeographical unit, hereafter referred to as the “northern cluster.” In order to analyse this cluster as a single unit, we merged the input pileups using a custom awk-based bash script that concatenated the base calls from the constituent populations at each genomic position, generating a single base call column representing all reads for this cluster. Likewise, Copenhagen and the two Skåne populations (which were also closely related) were treated as a single bivoltine population (hereafter “southern cluster”), and their pileups were combined in the same way. The three comparisons, then, were Öland versus the northern cluster, Gotland vs. the northern cluster, and the southern cluster versus the northern cluster, each intended to represent an at least semi-independent bivoltine/univoltine comparison on a different overall genetic background.

For each comparison, patterns of genetic variation (nucleotide diversity and \( F_{ST} \)) were calculated across the genome, again in nonoverlapping 5-kb windows. Nucleotide diversity (\( \pi \)) was calculated using popoolation version 1.2.2 on individual pileups for each geographical region in the pair, while \( F_{ST} \) was calculated using popoolation2 version 1.201 on a single mpileup of all four regions (where the northern and southern clusters had been merged into a single base call column each, as described above). Finally, in addition to the sliding-window analyses, \( F_{ST} \) was calculated in pairwise mpileups on a nucleotide-by-nucleotide basis (i.e., for each individual SNP in the genome). Biallelic SNPs with an \( F_{ST} \) of 0.9 or higher were considered strongly differentiated (when two populations are completely fixed for different alleles of an SNP, \( F_{ST} = 1 \)).

Outlier windows were selected based on four criteria: (i) an overall window \( F_{ST} \) in the 99th percentile of windows for that comparison, (ii) nucleotide diversity in the 1st percentile of windows for either population in that comparison, (iii) containing at least one strongly
differentiated SNP and (iv) being located in a genetic region. The last criterion was assessed using bedtools version 2.21.0 (Quinlan & Hall, 2010), by cross-referencing the list of potential outlier regions with the genetic regions identified in the genome annotation. A genetic region was defined as all introns and exons of a putative gene, plus the 5 kb on either side (i.e., immediately upstream and downstream), to account for potential selection on regulatory mutations. Each hit region was inspected visually for placement of differentiated SNPs etc. using the Integrative Genomics Viewer (Robinson et al., 2011), and the identity of the gene(s) overlapping with each outlier region was checked by running the putative amino acid sequences against the NCBI database using blastp (Camacho et al., 2009), specifying Lepidoptera (moths and butterflies) as the focal taxon.

2.6 | Scanning for loci associated with CD

To further identify genomic regions potentially associated with voltinism variation, while also accounting for the underlying demographic relationships among the studied populations, we applied a genotype–phenotype association approach using BAYPASS (Gautier, 2015). For this analysis we focused specifically on CD for pupal diapause induction, one of the main component traits of voltinism, as it is a continuous metric that allows for higher resolution than pairwise population comparisons.

Population values of CD were collated from published laboratory estimates (Aalberg Haugen & Gotthard, 2015; Lindestad et al., 2019), as well as two recent unpublished data sets for Kalmar and Copenhagen, respectively. Most of the studied populations have had their CD estimated more than once, and different rearing temperatures (16–20°C) have been used. Higher temperature decreases diapause propensity in P. aegeria, hence lowering the CD (Lindestad et al., 2019). On the assumption that CD is a linear function of temperature within the studied range, population means were calculated from a regression model with each estimate as one data point, population as a factor and temperature as a covariate (Appendix S1: Figure S1). The fitted per-population CDs at 18°C (Figure 1a) were then used as the phenotypic trait values for the BAYPASS analyses.

BAYPASS uses a Bayesian approach to identify loci whose allele frequencies correlate with a phenotypic trait across populations, while controlling for the underlying population covariance structure. The input used was the same allele frequency table as for the TREEMIX analysis. (Although the TREEMIX and BAYPASS analyses were run independently, the population structure inferred by BAYPASS and used as the basis for its analysis was highly similar to the one produced by TREEMIX.) CD phenotype values were scaled using the -scalecov command, and BAYPASS was run using the standard (default) model.

The analysis was run in triplicate, and for each analysed SNP, the median BF (Bayes factor; used as a phenotype association score) across the three runs was taken as the output, thereby filtering out spurious results due to stochastic behaviour of the Bayesian computation process. Because preliminary analyses suggested that rare alleles were strongly biasing the results, SNPs with a minor allele count lower than 10 (across all nine populations in the analysed set) were discarded.

As before, bedtools was used to cross-reference the BAYPASS output with our genome annotation to focus on outlier SNPs located in or near genes, yielding a set of genes potentially correlated with CD. As this outlier set was quite large, gene-set enrichment analysis (GSEA) was carried out using topgo version 2.34.0 (Alexa & Rahnenführer, 2016), in order to test for overrepresentation of particular functional gene categories among the outliers. Two GSEAs were performed: one for all genes containing SNPs with a BF score above 15, and one for all genes containing SNPs with a BF score above 20. This increased the chance of identifying larger patterns, and allowed us to gauge the extent to which the GSEA approach was sensitive to the significance cutoff used.

2.7 | Testing for sex-linked loci

Because sex-linked variation has been connected to diapause traits in other studies in Lepidoptera (Pruisscher et al., 2020), the combined set of outlier genes from the F2+/r and BAYPASS (BF >20) analyses (n = 379) was tested for Z-linkage. Using blastp (Camacho et al., 2009), the amino acid sequences for the outlier gene models was cross-referenced against a database of all Z-chromosome gene sequences from a recent, chromosome-level assembly of the P. aegeria genome (Lohse et al., 2021 [version 1; awaiting peer review]). blast hits with 70% identity or higher were considered matching genes.

2.8 | Scanning for population-specific indels

Because one of the identified outlier genes contained a large, population-specific deletion, a follow-up analysis was conducted in order to assess how common this type of structural variation is in the P. aegeria genome. For this analysis, the read depth-filtered data set was not used, so as not to unnecessarily lose data. Instead, a single mpileup file was generated for all nine populations. Positions where read depth equalled zero for all populations were filtered out, and an indel scan was then run in 100-bp windows using a custom awk-based bash script. Each 100-bp window where a single population had zero reads, while all other populations showed an average read depth of 20 or more across the window, was scored as an indel. Note that using this method, potential indels varying within a population, or shared between more than one population, were not counted.

3 | RESULTS

3.1 | Pararge aegeria genome assembly

A total of 148.12 million reads (85% >Q30) were generated during sequencing, and using different proportions of this raw read data for assembly greatly affected assembly statistics (Appendix S1: Figure 1b).
For the 12 initial linked read assemblies generated, N50 values ranged from 1.8 to 16.9 kbp, while total assembly length ranged from 198.1 to 466.7 Mbp. The 100% and 35% assemblies showed the highest N50 values (16.9 and 14.2 kbp, respectively), as well as the largest number of scaffolds at least 50 kbp in length (n = 778 and 661, respectively). The 35% assembly, however, additionally showed superior gene content and quality, with 74% of BUSCO entries being full length and single copy, and the lowest number of missing orthologues (7.3%). Given these results, the 35% assembly was used as the backbone for subsequent scaffolding, polishing, haplomerging, protein scaffolding and annotation.

The final genome was 479 Mbp in length, and consisted of 26,567 scaffolds (of which 95 scaffolds were >1 Mbp), with a maximum scaffold length of 3,041,493 bp, an N50 of 561 kbp and a total of 10.9% non-ATGC characters. BUSCO analysis of the final genome was much improved, with 88% entries complete and single-copy, 5.7% fragmented and 5.7% missing. A total of 23,567 high-quality gene models were annotated, of which 15,993 had functional annotations, and 33% of the genome was found to contain repetitive content.

### 3.2 Population history and connectivity analyses

Phylogeographical analysis suggested a deep split between southern versus northern + island *P. aegeria* populations, with the two Skåne populations as the sister group to all other Swedish populations, and all four northern univoltine populations clustering together (Figure 1b). At least two admixture events could be inferred from the residual allelic variation. First, Gotland appears to derive a large part of its genetic variation (~21%) from a population related to the Stockholm population. Second, Southern Skåne is inferred to have received a smaller amount of genetic variation (~5%) that is shared by all northern and island populations. Subsequent migration edges were not considered geographically likely to reflect true gene flow events, and did not notably improve tree fit, so were not considered further (Appendix S1: Figure S3).

Average genome-wide *F*$_{ST}$ varied between population comparisons (Figure 1d). Mainland populations showed a weak isolation-by-distance pattern, with low differentiation between the three closely grouped southern (Copenhagen and Skåne) populations, and generally higher differentiation between the northernmost population (Sundsvall) and other populations. However, most comparisons across the north/south distribution gap gave an *F*$_{ST}$ around 0.07–0.08, regardless of geographical distance. As for the two island populations, Öland was relatively strongly differentiated (*F*$_{ST}$ = 0.14) from all other populations, while Gotland was mildly differentiated (*F*$_{ST}$ = 0.08) from all mainland populations. In terms of population-wide patterns of genetic diversity, the three southern populations were relatively diverse, with a tendency toward negative values of Tajima’s *D*, while the northern mainland populations showed intermediate diversity and less negative values of *D*. Again, the two island populations yielded quite different results: Öland had low nucleotide diversity and near-zero average Tajima’s *D*, while Gotland showed similar patterns of diversity to the northern mainland populations (Figure 1c).

### 3.3 Regions showing possible selection footprints

All three univoltine/bivoltine comparisons yielded genic regions showing low π, high overall *F*$_{ST}$ and at least one strongly differentiated...
SNP (Figure 2a–c). All outlier SNPs detected were unique to a single comparison (Appendix S1: Table S3), and the discovery rate corresponded to the overall differentiation between the two geographical regions in the pair. When comparing the southern cluster to the northern cluster (Figure 2c), overall genome-wide \( F_{ST} \) was low (mean = 0.04), and only one region met all of the search criteria: a putative glucose dehydrogenase gene, which contained a non-synonymous exonic substitution. Several of the 15 outlier loci detected in Pruisscher et al.’s (2018) scan of two of the populations represented in this comparison (Sundsval and North Skåne) also showed a high \( F_{ST} \)/low \( \pi \) signal here, including the circadian gene \textit{timeless}. However, the previously described candidate SNPs in these genes did not register as strongly differentiated here, as one of the populations in the northern cluster, Kalmar, harboured the "southern" alleles at intermediate frequencies. By far the largest number of outlier SNPs was found when comparing Öland to the northern cluster: 13 windows met all search criteria (Figure 2b). Two outlier loci for this comparison contained non-synonymous substitutions: kinesin-associated protein 3 and a calcium-binding protein. Meanwhile, the Gotland/northern cluster comparison (Figure 2a) yielded outlier windows in three genes. The first was a programmed cell death protein; the latter two were \textit{timeless} and \textit{period}, both of which contained SNPs unique to Gotland, including a non-synonymous (Pro→Ser) substitution in \textit{period}.

### 3.4 Regions showing association with critical daylength

Using \textsc{baypass} to test for associations between allele frequency and CD across all nine populations, while controlling for their underlying demographic relationships, yielded 1084 genes containing at least one SNP with a BF score above 15. Of these, 346 genes further contained at least one SNP with a BF score above 20 (Figure 2d). Gene-set enrichment analysis revealed few robust patterns among the outlier genes. For genes with a BF above 15, the most significantly enriched GO terms included connections to Toll signalling, limb formation, and responses to fungi and toxins (Table 2). Narrowing the set to genes with a BF above 20 instead yielded a rather different pattern of enriched GO terms, relating to amino acid metabolism and subcellular localization/transport of proteins (Table 3).

One notable result that was reflected in the GSEA at both sensitivity levels was the presence of outlier genes relating to rhythmic processes and circadian function. Five circadian clock genes were found to contain SNPs correlated with CD across the studied populations: \textit{clockwork orange} (BF = 41), \textit{timeless} (BF = 32), \textit{period} (BF = 24), \textit{timeless}2 (BF = 16) and \textit{cryptochrome} (BF = 15). This also partly accounts for the emphasis on subcellular protein localization/transport among the enriched GO terms (Table 3), as such transport is part of circadian clock function. Circadian-related GO-term enrichment was further boosted by the presence of a 5-HT (serotonin) receptor among the outliers (BF = 17).

One of the outlier regions contained two genes previously described as candidate loci for diapause regulation in \textit{P. aegeria}: \textit{kinesin-associated protein 3} and \textit{carotinoid o-acyltransferase}. Previous work indicates that these two genes are in linkage with \textit{timeless} (Pruisscher et al., 2018). The highest-scoring SNPs in the \textsc{baypass} output (BF > 50) were found clustered in a 30-kb region containing several genes tied to basic cellular functions: \textit{importin (subunit) \( \alpha \)}, the beta subunit of mitochondrial ATP synthase, the ribosomal protein L24e and the ribosomal export protein NMD3. Overall, the \textsc{baypass} outlier set was dominated by genes showing acline in allele frequency from north to south, reflecting the correlation between CD and latitude in the studied populations. However, the detailed clinal pattern varied greatly between different outlier genes, with some alleles gently shifting frequencies over latitudes, and others showing a more abrupt sharp north/south cutoff (Figure 3). A list of \textsc{baypass} outlier genes (BF > 20) is provided in Appendix S2.

Five genetic regions (containing a total of seven genes) appeared as outliers in both the \( F_{ST} \) and \textsc{baypass} analyses; these showed a high degree of overlap with the candidate regions reported by Pruisscher et al. (2018). Results for these regions are summarized in Table 4. In

| Term ID      | Term name                                      | Total | Outliers | \( p \)   |
|--------------|-----------------------------------------------|-------|----------|----------|
| GO:0008592   | Regulation of Toll signalling pathway         | 12    | 5        | .0004    |
| GO:0061013   | Regulation of mRNA catabolic process          | 5     | 3        | .0012    |
| GO:0009649   | Entrainment of circadian clock                | 11    | 4        | .0015    |
| GO:0074449   | Proximal/distal pattern formation, imaginal disc | 17    | 5        | .0017    |
| GO:009954    | Proximal/distal pattern formation              | 22    | 5        | .0024    |
| GO:002783    | Antifungal peptide biosynthetic process       | 9     | 4        | .0027    |
| GO:002810    | Regulation of antifungal peptide biosynthetic process | 9     | 4        | .0027    |
| GO:0045752   | Positive regulation of Toll signalling pathway | 6     | 3        | .0032    |
| GO:009636    | Response to toxic substance                   | 106   | 11       | .0038    |
| GO:0035223   | leg disc pattern formation                    | 12    | 4        | .0048    |

**Table 2** The 10 most significantly enriched "biological process" GO terms for genes with BF > 15 in the \textsc{baypass} analysis. Total = number of genes assigned a given GO term in the \textit{Pararge aegeria} genome annotation; Outliers = number of genes assigned the same GO term in the outlier set; \( p \) = parent/child-adjusted Fisher \( p \)-value.
total, only 12 outlier loci appeared to be located on the Z chromosome (Appendix S1: Table S4). These included period, which is previously known to be Z-linked in Lepidoptera (Pruisscher et al., 2020), but did not include any of the other loci showing a high $F_{ST}/\pi$ signal. Hence, the results of the analyses did not appear to be primarily driven by sex-linked variation.

### Table 3

The 10 most significantly enriched “biological process” GO terms for genes with BF > 20 in the BAYPASS analysis. Total = number of genes assigned a given GO term in the Pararge aegeria genome annotation; Outliers = number of genes assigned the same GO term in the outlier set; $p$ = parent/child-adjusted Fisher $p$-value.

| Term ID      | Term name                                           | Total | Outliers | $p$   |
|--------------|-----------------------------------------------------|-------|----------|-------|
| GO:0006575   | Cellular modified amino acid metabolic process      | 39    | 4        | .0019 |
| GO:0045475   | Locomotor rhythm                                    | 41    | 4        | .0032 |
| GO:0046835   | Carbohydrate phosphorylation                        | 7     | 2        | .0047 |
| GO:0006749   | Glutathione metabolic process                       | 18    | 3        | .0053 |
| GO:0051169   | Nuclear transport                                    | 105   | 7        | .0056 |
| GO:0007166   | Cell surface receptor signalling pathway            | 662   | 15       | .0082 |
| GO:0046685   | Response to arsenic-containing substance            | 12    | 2        | .0084 |
| GO:0051234   | Establishment of localization                       | 1339  | 29       | .0117 |
| GO:0007010   | Cytoskeleton organization                           | 587   | 11       | .0138 |
| GO:0048511   | Rhythmic process                                    | 92    | 5        | .0163 |

### Table 4

Genic regions that ranked as outliers in both the BAYPASS (BF > 20) and $F_{ST}/\pi$ analyses. “Comparison” indicates which population contrast returned a given region as having high $F_{ST}$ and low $\pi$. “BF” shows the Bayes Factor for the highest-scored SNP in each gene. Genes that were among the 15 candidate loci reported by Pruisscher et al. (2018) are marked with an asterisk (*). Note that the outlier region in contig m232 was large (45 kbp) and contained three genes with high BF scores.

| Contig | Comparison          | BF  | Genes in region          | Notes                  |
|--------|---------------------|-----|--------------------------|------------------------|
| m431   | South vs. north     | 31  | glucose dehydrogenase-like | Nonsyn. SNP            |
| m442   | Gotland vs. north   | 32  | timeless*                | Nonsyn. SNP; large deletion |
| c2968  | Gotland vs. north   | 24  | period*                  | Nonsyn. SNP            |
| m594   | Öland vs. north     | 18  | MORN4 homologue          |                        |
| m232   | Öland vs. north     | 32  | RAB14                    |                        |
|        |                     |     | choline/carnitine        |                        |
|        |                     |     | 0-acyltransferase*       |                        |
|        |                     |     | kinesin-associated protein 3* | Nonsyn. SNP |
timeless as well as most of exon 14, shortening the predicted TIM protein sequence by 97 amino acids, or 8% of its length. The deletion was judged unlikely to be a bioinformatic artefact, as reads from all other eight populations mapped to this region (Figure 4). In addition, PCR amplification using a pair of primers that spanned this region yielded a long DNA fragment in two Stockholm individuals, but a short fragment in two Gotland individuals, consistent with a 1.5-kbp deletion in the Gotland samples (Appendix S1: Figure S4).

A genome-wide scan of population-specific indel variation revealed putative deletions at an additional 325 locations in the genome, some of a similar size, but most only one or two hundred bp in size (Appendix S1: Figure S5). In 96 cases, the missing sequence data lay within an exon. The majority of putative deletions were in Sundsvall, although that population may have a somewhat higher risk of false positives, as its read depth was generally lower across the genome. Öland showed deletions at a few loci, and Gotland at one additional locus. However, the deletion in timeless was the largest population-specific indel detected in any population.

4 | DISCUSSION

Scandinavian Pararge aegeria constitute a set of relatively recently (i.e., postglacially) immigrated populations showing a certain degree of local adaptation to a climate gradient. Consistent with this, the results presented here suggest a genomic landscape of mild interpopulation differentiation, interspersed with spots of stronger divergence with potential adaptive importance. The outlier searches were conducted with unspecific criteria, the aim being to identify any genic regions potentially associated with selection on voltinism variation. Given this, it is notable that the output of both the selection-footprint scans and phenotype association analyses so prominently featured circadian genes, a group of genes that have previously been implicated in generating variation in life cycle regulation in P. aegeria (Pruisocher et al., 2018) as well as other species (see below). The exonic deletion in timeless was found essentially by luck, as our methods included a visual examination of each detected outlier region, revealing the local drop in read depth (Figure 4). Judging by follow-up scans, such large, population-specific deletions are not particularly common in P. aegeria (Appendix S1: Figure S5). The chance nature of its discovery highlights the risk, inherent to typical SNP-based genomic methods, of overlooking structural variation that may be important to adaptation (Hoban et al., 2016); however, bioinformatic tools do exist that may facilitate the discovery of such variation (Chen et al., 2009; Rausch et al., 2012).

4.1 | Circadian gene variation

Of the circadian genes emerging as outliers in the present analyses, timeless, period and cryptochrome are the best characterized. The first two are negative oscillators in the core circadian loop across most insects, including butterflies (Sandrelli et al., 2008; Zhu et al., 2008), while cryptochrome (cry1) codes for the main light receptor that allows the circadian clock to synchronize with the daylight cycle (Yuan et al., 2007; Zhu et al., 2008). Allelic variation in all three of these genes has previously been tied to geographical differences in diapause traits: timeless in pitcher plant mosquitoes (Mathias et al., 2007) and two different Drosophila species (Tauber et al., 2007; Yamada & Yamamoto, 2011); period in corn borer moths (Kozak et al., 2019) and parasitic wasps (Paolucci et al., 2016); and cryptochrome in Drosophila triauraria (Yamada & Yamamoto, 2011).

One of the strongest BAYPASS outliers, containing several variable SNPs (Figure 3b), was clockwork orange (cwo). In Drosophila, the CWO protein acts synergistically with the main period/timeless loop by binding to and repressing several circadian oscillator genes (including tim, per and cwo itself) in the late night, thus forming an essential subloop that helps keep circadian rhythmicity at high amplitude (Kadener et al., 2007; Tomioka & Matsumoto, 2015). While this function has also been inferred in lepidopterans, it has not yet been experimentally confirmed (Brady et al., 2021). Less is known about the function of timeless2 (also known as "mammalian timeless" or timout), although it appears to help entrain the insect circadian clock to light (Benna et al., 2010), and has been tied to population variation in circadian activity rhythms in rice borer moths (Zhu et al., 2019).

Two other outlier loci merit mention for their connection to this subject. While not a core circadian gene, the BAYPASS outlier casein kinase 1α has been tentatively linked to circadian clock regulation in silk moths (Trang et al., 2006). The highest-scored BAYPASS outlier region (BF = 55) contained the gene importin subunit α. Importins control the transport of proteins into the nucleus; this has been shown to be essential to both fly and mammal clock function for importin β, albeit not for importin α (Lee et al., 2015).

Although a central role for circadian genes in the regulation of diapause in particular and photoperiod-controlled traits in general
has been well established, it remains controversial to what extent the circadian clock as such forms a part of the daylength-sensing machinery, as opposed to some circadian genes merely having a pleiotropic role in photoperiodism (Bradshaw & Holzapfel, 2010; Meuti & Denlinger, 2013; Saunders et al., 2004). In either case, the connection between the two time-keeping systems complicates adaptive interpretations for two reasons. First, because different latitudes experience different day/night regimes that may need to be compensated for, circadian genes could potentially vary between populations as a result of direct selection on clock function, with no regard to photoperiodic traits (Bradshaw & Holzapfel, 2010). Second, clock traits such as circadian period length have themselves been shown to correlate with latitude in various species (Hut et al., 2013; Pivarciova et al., 2016), and in some cases this has been linked to allometric variation in clock genes (Dalla Benetta et al., 2019; Kozak et al., 2019). It is unclear if this type of variation is a side-effect of selection on photoperiodism, or if clock traits themselves are locally adapted to different selective environments (Salmela & Weinig, 2019). In general, the close correlation between latitude and CD across the populations studied here means that, despite controlling for demographic background, the BayPASS analysis cannot necessarily disentangle the effects of selection on diapause traits from selection on circadian rhythmicity or anything else that may covary with latitude. Functional assays of diapause regulation and/or circadian rhythmicity will be required to understand the adaptive role of the genetic variation found across these populations.

The most drastic novel variation was found in timeless, where Gotland P. aegeria lack 8% of the gene’s coding sequence. It is difficult to say what the effect the deletion may have. The lepidopteran circadian clock is relatively well described at the level of whole protein interactions (Brady et al., 2021; Zhu et al., 2008), and experimental data exist on the function of specific parts of a few circadian proteins (Zhang et al., 2017), but to our knowledge the lepidopteran TIM protein has not yet been studied at this level of mechanistic detail. Results from silk moths show that, at least on an amino acid sequence level, both TIM and PER contain many of the functional domains described in Drosophila (Iwai et al., 2006). These include the cytoplasmic localization domain (CLD) characterized in TIM by Saéz and Young (1996), which overlaps partially with the Gotland-specific deletion (Appendix S1: Figure S6). In D. melanogaster, this domain prevents TIM from entering the nucleus; its effect is overridden by dimerization of TIM with PER, which allows the core transcriptional feedback loop of the circadian clock to be closed.

Predicting the functional consequences of circadian gene variation across insect orders is problematic, not least because the insect circadian clock machinery is variable both in which components are present and what roles they play (Yuan et al., 2007). In the monarch (Danaus plexippus), the foremost model species for butterfly circadian function, the core feedback loop is not closed by a TIM/PER dimer, as in Drosophila. Instead, this role is filled by CRY2, a light-insensitive parologue of CRY, which forms a trimer with TIM and PER (Zhu et al., 2008). Nonetheless, if TIM–PER interaction sites are conserved, the overlap of the Gotland deletion with the putative silk moth CLD is an indication that this mutation may have some effect on the functioning of the TIM protein. Finally, we note an interesting parallel with a previously known case of a length polymorphism in timeless, whose allele frequency cline also correlates with diapause variation (Sandrelli et al., 2007; Tauber et al., 2007). However, in that case the protein length variation results from alternative start codons rather than a deletion, and hence affects the N-terminus, the opposite end of the TIM polypeptide from where the Gotland deletion is located.

### 4.2 Population history and adaptive divergence

Overall genetic differentiation among the studied populations was fairly low, with most comparisons yielding an average genome-wide $F_{ST}$ of 0.06–0.08 (Figure 1d). The exception was Öland, which showed relatively strong differentiation from all other populations. This probably stems from the low allometric diversity of this population (Figure 1c), which appears to have gone through a demographic bottleneck during its history, as also indicated by its long branch in Figure 1b. On the mainland, both Treemix results and genome-wide $F_{ST}$ values indicate two well-separated population clusters with differing voltinism. Gene flow across the mainland voltinism divide is probably limited, given that the two southernmost univoltine populations, Kalmar and the Småländ Highlands, are just as differentiated from the bivoltine populations as Stockholm is, at twice the geographical distance (Figure 1d). In general, regions of elevated $F_{ST}$ between the compared populations showed no particular tendency toward lowered nucleotide diversity in either population, indicating that the largest part of genome-wide differentiation between Scandinavian $P. aegeria$ populations has occurred through neutral genetic processes over the course of postglacial migration, rather than through selection. Overall nucleotide diversity was higher for southern populations, which confirms earlier microsatellite results (Tison et al., 2014) and is consistent with gentle clinal loss of diversity during post-Pleistocene northward expansion from glacial refugia. Interestingly, low average $\pi$ values correlated with high values of $D$ (Figure 1c). This may reflect a relatively short time having passed since northern colonization, with rare variants having been lost to drift and not yet replaced by novel mutations.

The Treemix analysis suggests that South Skåne (but not North Skåne) has received a small amount of novel genetic variation shared by all northern and island populations. This could be interpreted as gene flow along the Baltic coast. However, a more likely explanation may be secondary contact, where the recent (i.e., 1930s) migration wave from Denmark encountered a small number of $P. aegeria$ still present in southeastern Sweden from the first migration wave that produced the northern/island populations. Our analyses also suggest a possible explanation for the surprisingly high genetic diversity of the geographically isolated island of Gotland (Figure 1c), namely that this population may have originated through admixture of different founder populations. The results of these phylogeographical analyses are fully
compatible with, and contribute additional detail to, the picture of genetic diversity in Northern European *P. aegeria* suggested by earlier results (Tison et al., 2014).

Given the differentiation in both voltinism regulation (Lindestad et al., 2019; Nylin et al., 1995) and associated life history traits (Aalberg Haugen & Gotthard, 2015) displayed by Scandinavian *P. aegeria* populations, one of the aims of the present study was to investigate the extent to which selective (local adaptation) versus nonselective processes (genetic history) have contributed to the expressed life cycle patterns. A particularly good indicator of selection is functional differentiation between populations despite a certain degree of gene flow, as it suggests that phenotypic differentiation is continually being maintained by local adaptation (Kawecki & Ebert, 2004). Given that univoltine and bivoltine populations clustered separately in the phylogeographical analyses, with little evidence of recent or ongoing genetic exchange between northern and southern mainland populations, such conclusions cannot as easily be drawn here. However, it should be noted that voltinism is a labile trait on ecological timescales (Alttermatt, 2010), generated through a combination of genetic and environmental differences between populations (Lindestad et al., 2019), and has been observed to evolve quickly through modification of photoperiodic plasticity (Bean et al., 2012; Yamanaka et al., 2008). For these reasons, a classical phylogenetic–comparative interpretation with a single transition of voltinism states along the population tree may be misleading. Assuming a postglacial colonization scenario where a univoltine range margin (as today represented by Sundsvall) tracks the upper limit of the viable climate envelope for the species, it is likely that currently bivoltine *P. aegeria* populations in this region were univoltine when first established.

Interestingly, the bivoltine Gotland population, which in terms of the candidate SNPs for diapause regulation characterized by Pruisscher et al. (2018) has a distinctly “northern” profile, is here shown to harbour additional, unique variants at two of these same candidate loci (timeless and period). In other words, instead of simply carrying a “northern,” high-CD allele of these two genes, this population in fact displays a unique, derived allele of both genes. If these local variants have a phenotypic effect, this could be interpreted as a result of selection for bivoltinism through a lowered CD following colonization by univoltine genetic stock. It is also worth noting that the Gotland population appears to have an intrinsically longer diapause than neighbouring populations, possibly as an adaptation to partial bivoltinism (Lindestad et al., 2020). Differences in the rate of diapause termination are associated with variation at period in the moth *Ostrinia nubilalis* (Kozak et al., 2019; Levy et al., 2015), marking the novel circadian gene variation in Gotland as a possible parallel.

## 5 CONCLUSIONS

European populations of *Pararge aegeria* are a useful model system for studying postglacial migration and climate adaptation (Aalberg Haugen & Gotthard, 2015; Hill et al., 1999; Pateman et al., 2016; Tison et al., 2014). The present results expand on earlier analyses in this species by providing a more fine-scale picture of both neutral and putatively adaptive variation. In a wider perspective, it is especially notable that circadian loci, which appear to play a key role in adaptation to geographical variation in seasonality across a range of insect species, again emerge as potentially under selection. While any hypothetical effects of the novel mutations described here (either on circadian functioning or on photoperiodic diapause induction) are unknown, these loci stand out as promising targets for further investigation in vivo.

### ACKNOWLEDGEMENTS

This work was financed by the Knut and Alice Wallenberg Foundation (KAW 2012.0058), the Bolin Centre for Climate Research at Stockholm University (to K. Gotthard), and the Swedish Research Council grants VR-2012-3715 (to S. Nylin), VR-2012-4001 (to C.W.W.) and VR-2017-04500 (to K. Gotthard). Genome assembly received 50% cofunding from SciLifeLab National Programmes (2014/R2-77). High Performance Computing resources were provided by the Swedish National Infrastructure for Computing (SNIC), partially funded by the Swedish Research Council through grant agreement nos. 2020-05284 and 2020-05609. Many thanks to Maria de la Paz Celorio Mancera for help with DNA extraction and sample preparation, and to Peter Pruisscher for input on bioinformatic methods.

### OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at https://doi.org/10.5281/zenodo.5780591.

### DATA AVAILABILITY STATEMENT

Bioinformatic scripts have been made available through Zenodo (Lindestad et al., 2021a). Sequence data and genome assembly data have been deposited through EMBL-EBI (Lindestad et al., 2021b).

### ORCID

Olle Lindestad https://orcid.org/0000-0002-8226-3167
Sören Nylin https://orcid.org/0000-0003-4195-8920
Christopher W. Wheat https://orcid.org/0000-0003-1863-2340
Karl Gotthard https://orcid.org/0000-0002-4560-6271

### REFERENCES

Aalberg Haugen, I. M., & Gotthard, K. (2015). Diapause induction and relaxed selection on alternative developmental pathways in a butterfly. *Journal of Animal Ecology*, 84, 464–472. https://doi.org/10.1111/1365-2656.12291
Alexa, A., & Rahnenführer, J. (2016). topGO: Enrichment Analysis for Gene Ontology. R package version 2.34.0. https://bioconductor.org/packages/release/bioc/html/topGO.html
Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaption. Ecology Letters, 7, 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kivelä, S. M., Välimäki, P., & Gotthard, K. (2013). Seasonality maintains alternative life-history phenotypes. Evolution, 67, 3145–3160. https://doi.org/10.1111/evo.12181

Koštál, V., Orozco-terWengel, P., de Maio, N., Pandey, R. V., Rolte, V., Futschik, A., Kosil, C., & Schlötterer, C. (2011). Population: A toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One, 6, e15925. https://doi.org/10.1371/journal.pone.0015925

Koštál, V., Pandey, R. V., & Schlötterer, C. (2011). PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). Bioinformatics, 27, 3435–3436. https://doi.org/10.1093/bioinformatics/btr589

Koštál, V. (2006). Eco-physiological phases of insect diapause. Journal of Insect Physiology, 52, 113–127. https://doi.org/10.1016/j.jinsphys.2005.09.008

Koštál, V. (2011). Insect photoperiodic calendar and circadian clock function: Independence, cooperation, or unity? Journal of Insect Physiology, 57, 538–556. https://doi.org/10.1016/j.jinsphys.2010.10.006

Kozak, G. M., Wadsworth, C. B., Kahne, S. C., Bogdanowicz, S. M., Harrison, R. G., Coates, B. S., & Dopman, E. B. (2019). Genomic basis of circannual rhythm in the European Corn Borer Moth. Current Biology, 29, 3501–3509. https://doi.org/10.1016/j.cub.2019.08.053

Lee, Y., Jang, A. R., Francey, L. J., Sehgal, A., & Hogenesch, J. B. (2015). Quantitative trait loci associated with photoperiodic response and stage of diapause in the pitcher-plant mosquito, Wyeomyia smithii. Genetics, 176, 391–402.

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. Genetical Research, 23, 23–35. https://doi.org/10.1017/S0016672300014634

Meuti, M. E., & Denlinger, D. L. (2013). Evolutionary links between circadian clocks and photoperiodic diapause in insects. Integrative and Comparative Biology, 53, 131–143. https://doi.org/10.1093/icb/ict023

Mousseau, T. A., & Roff, D. A. (1989). Adaptation to seasonality in a cricket: Patterns of phenotypic and genotypic variation in body size and diapause expression along a cline in season length. Evolution, 43, 1483–1496. https://doi.org/10.2307/15586466.1989.tb02598.x

Neethiraj, R., Hornett, E. A. E., Hill, J. A., & Wheat, C. W. (2017). Investigating the genomic basis of discrete phenotypes using a Pool-Seq-only approach: New insights into the genetics underlying colour variation in diverse taxa. Molecular Ecology, 26, 4990–5002. https://doi.org/10.1111/mec.14205

Nordström, F. (1955). De fennoskandiska dagfjärilarnas utbredning. Lepidoptera, Diurna (Rhopalocera & Hesperioida). Kungliga Fysiografiska Sällskapets Handlingar (NF), 66, 1–157.

Nowell, R. W., Elsworth, B., Oostra, V., Zwaan, B. J., Wheat, C. W., Saastamoinen, M., Saccheri, I. J., van’t Hof, A., Wes, B. K., Connahs, H., Aslam, M. L., Kumar, S., Challis, R. J., Monteiro, A., Brakefield, P. M., & Blaxter, M. (2017). A high-coverage draft genome of the mycalesine butterfly Bicyclus anynana. GigaScience, 6, gix035. https://doi.org/10.1093/gigascience/gix035

Nylin, S., Wickman, P.-O., & Wiklund, C. (1995). Life-cycle regulation and life history plasticity in the speckled wood butterfly: Are reaction norms predictable? Biological Journal of the Linnean Society, 55, 143–157. https://doi.org/10.1111/j.1095-8312.1995.tb01056.x

Paolucci, S., Salis, L., Vermeulen, C. J., Beukeboom, L. W., & van de Zande, L. (2016). QTL analysis of the photoperiodic response and clinal distribution of period alleles in Nasonia vitripennis. Molecular Ecology, 25, 4805–4817.

Pateman, R. M., Thomas, C. D., Hayward, S. A. L., & Hill, J. K. (2016). Macro- and microclimatic interactions can drive variation in species’ habitat associations. Global Change Biology, 22, 556–566. https://doi.org/10.1111/gcb.13056

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genetics, 8, e1002967.

Pivarciova, L., Vanecová, H., Provazník, J., Wu, B.-C.-H., Pivarcí, M., Pecková, O., Bazalová, O., Cada, Š., Kment, P., Kotwica-Rolinska, J., & Dolezel, D. (2016). Unexpected geographic variability of the free running period in the linen bug Pyrrhocoris apterus. Journal of Biological Rhythms, 31, 568–576.

Posledovich, D., Toftegård, T., Wiklund, C., Ehrlén, J., & Gotthard, K. (2015). Latitudinal variation in diapause duration and post-winter development in two pierid butterflies in relation to phenological specialization. Oecologia, 177, 181–190. https://doi.org/10.1007/s00442-014-3125-1

Pruisscher, P., Nylin, S., Gotthard, K., & Wheat, C. W. (2018). Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. Molecular Ecology, 27, 3613–3626. https://doi.org/10.1111/mec.14829

Pruisscher, P., Nylin, S., Wheat, C. W., & Gotthard, K. (2020). A region of the sex chromosome associated with population differences in diapause induction contains highly divergent alleles at clock genes. Evolution, 75, 490–500. https://doi.org/10.1111/evo.14151

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. Bioinformatics, 26, 841–842. https://doi.org/10.1093/bioinformatics/btp033

Ragland, G. J., Armbruster, P. A., & Meuti, M. E. (2019). Evolutionary and functional genetics of insect diapause: A call for greater integration. Current Opinion in Insect Science, 36, 74–81. https://doi.org/10.1016/j.cois.2019.08.003
Rausch, T., Zichner, T., Schlattl, A., Stütz, A. M., Benes, V., & Korbel, J. O. (2012). DELLY: Structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics, 28, i333–i339. https://doi.org/10.1093/bioinformatics/bts378

Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., Karchner, S. I., Hahn, M. E., Nacci, D., Oleksiak, M. F., Crawford, D. L., & Whitehead, A. (2016). The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science, 354, 1305–1308. https://doi.org/10.1126/science.aa4993

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. Nature Biotechnology, 29, 24–26. https://doi.org/10.1038/nbt.1754

Roff, D. (1980). Optimizing development time in a seasonal environment: The “ups and downs” of clinal variation. Oecologia, 45, 202–208. https://doi.org/10.1007/BF00346461

Saez, L., & Young, M. W. (1996). Regulation of nuclear entry of the Drosophila clock proteins period and timeless. Neuron, 17, 911–920. https://doi.org/10.1016/S0896-6273(00)80222-6

Sahlin, K., Veszi, F., Nystedt, B., Lundeberg, J., & Arvestad, L. (2014). BESSST - Efficient scaffolding of large fragmented assemblies. BMC Bioinformatics, 15, 281. https://doi.org/10.1186/1471-2105-15-281

Salmela, M. J., & Weinig, C. (2019). The fitness benefits of genetic variation in circadian clock regulation. Current Opinion in Plant Biology, 39, 354–348. https://doi.org/10.1016/j.molbev.2019.06.003

Sandrelli, F., Costa, R., Kyriacou, C. P., & Rosato, E. (2008). Comparative analysis of circadian clock genes in insects. Insect Molecular Biology, 17, 447–463. https://doi.org/10.1111/j.1365-2583.2008.00832.x

Sandrelli, F., Tauber, E., Pegoraro, M., Mazzotta, G., Cisotto, P., Landskron, J., Stanewsky, R., Piccin, A., Rosato, E., Zordan, M., Costa, R., & Kyriacou, C. P. (2007). A molecular basis for natural selection at the timeless locus in Drosophila melanogaster. Science, 316, 1898–1900.

Saunders, D. S., Lewis, R. D., & Warman, G. R. (2004). Photoperiodic induction of diapause: Opening the black box. Physiological Entomology, 29, 1–15. https://doi.org/10.1111/j.1365-3303.2004.00369.x

Sedlacek, F. J., Rescheneder, P., & Von Haeseler, A. (2013). NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics, 29, 2790–2791. https://doi.org/10.1093/bioinformatics/btt468

Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics, 24, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke, M., Schönfeld, O., Morgenstern, B., & Waack, S. (2006a). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics, 7, 62. https://doi.org/10.1186/1471-2105-7-62

Stinchcombe, J. R., Weinig, C., Ungerer, M., Olsen, K. M., Mays, C., Breda, C., Daga, A., Selmin, A., Monger, K., Benna, C., Rosato, E., Kyriacou, C. P., & Costa, R. (2007). Natural selection favors a newly derived timeless allele in Drosophila melanogaster. Science, 316, 1895–1898.

Tauber, M. J., Tauber, C. A., & Masaki, S. (1986). Seasonal adaptations of insects. Oxford University Press.

Tison, J. L., Edmark, V. N., Sandoval-Castellanos, E., Van Dyck, H., Tammaru, T., Välimäki, P., Dalén, L., & Gotthard, K. (2014). Signature of post-glacial expansion and genetic structure at the northern range limit of the speckled wood butterfly. Biological Journal of the Linnean Society, 113, 136–148. https://doi.org/10.1111/bij.12327

Tomioka, K., & Matsumoto, A. (2015). Circadian molecular clockworks in non-model insects. Current Opinion in Insect Science, 7, 58–64. https://doi.org/10.1016/j.cois.2014.12.006

Trang, L. T. D., Sehadowa, H., Ichihara, N., Iwai, S., Mita, K., & Takeda, M. (2006). Casein kinases I of the silkworm, Bombyx mori: Their possible roles in circadian timing and developmental determination. Journal of Biological Rhythms, 21, 335–349. https://doi.org/10.1177/074830406291734

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One, 9, e112963. https://doi.org/10.1371/journal.pone.0112963

Waterhouse, R. M., Seppey, M., Simão, F. A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E. V., & Zdobnov, E. M. (2018). BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular Biology and Evolution, 35, 543–548. https://doi.org/10.1093/molbev/msx319

Yamada, H., & Yamamoto, M. T. (2011). Association between circadian clock genes and diapause incidence in Drosophila triauraria. PLoS One, 6, e27493. https://doi.org/10.1371/journal.pone.0027493

Yamanaka, T., Tatsuki, S., & Shimada, M. (2008). Adaptation to the new land or effect of global warming? An age-structured model for rapid volitism change in an alien lepidopteran pest. Journal of Animal Ecology, 77, 585–596. https://doi.org/10.1111/j.1365-2656.2008.01367.x

Yuan, Q., Metterville, D., Briscoe, A. D., & Reppert, S. M. (2007). Insect cryptochrome: Gene duplication and loss define diverse ways to construct insect circadian clocks. Molecular Biology and Evolution, 24, 948–955. https://doi.org/10.1093/molbev/msm011

Zhang, Y., Markert, M. J., Groves, S. C., Hardin, P. E., & Merlin, C. (2017). Vertebrate-like CRYPTOCHROME 2 from monarch regulates circadian transcription via independent repression of CLOCK and BMAL1 activity. Proceedings of the National Academy of Sciences of the United States of America, 114, E7516–E7525. https://doi.org/10.1073/pnas.170214114

Zhu, H., Sauman, I., Yuan, Q., Cesselin, A., Emery-Le, M., Emery, P., & Reppert, S. M. (2008). Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. PLoS Biology, 6, e4. https://doi.org/10.1371/journal.pbio.0060004

Zhu, L., Feng, S., Gao, Q., Liu, W., Ma, W.-H., & Wang, X.-P. (2019). Host population related variations in circadian clock gene sequences and expression patterns in Chilo suppressalis. Chronobiology International, 36, 969–978.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher’s website.

How to cite this article: Lindestad, O., Nylin, S., Wheat, C. W., & Gotthard, K. (2022). Local adaptation of life cycles in a butterfly is associated with variation in several circadian clock genes. Molecular Ecology, 31, 1461–1475. https://doi.org/10.1111/mec.16331