doi:10.1111/evo.14151

A region of the sex chromosome associated with population differences in diapause induction contains highly divergent alleles at clock genes

Peter Pruisscher,[1,2] Sören Nylin,[1] Christopher West Wheat,[1] and Karl Gotthard[1]

[1] *Department of Zoology, Stockholm University, Svante Arrheniusväg 18b, Stockholm S-106 91 Sweden*

[2] *E-mail: peter.pruisscher@gmail.com*

Received May 4, 2020

Accepted October 25, 2020

Developmental plasticity describes the capacity of individuals with the same genotype to induce permanent change in a phenotype depending on a specific external input. One well-studied example of adaptive developmental plasticity is the induction of facultative diapause in insects. Studies investigating the inheritance of diapause induction have suggested diverse genetic origins. However, only few studies have performed genome-wide scans to identify genes affecting the induction decision. Here we compare two populations of the butterfly *Pieris napi* that differ in the propensity to enter diapause, and despite showing a low genome-wide divergence, we identify a few genomic regions that show high divergence between populations. We then identified a single genomic region associated with diapause induction by genotyping diapausing and directly developing siblings from backcrosses of these populations. This region is located on the Z chromosome and contained three circadian clock genes, *cycle*, *clock*, and *period*. Additionally, *period* harbored the largest number of SNPs showing complete fixation between populations. We conclude that the heritable basis of between-population variation in the plasticity that determines diapause induction resides on the Z chromosome, with the *period* gene being the prime candidate for the genetic basis of adaptive plasticity.

KEY WORDS: **Crosses, diapause induction, genes, local adaptation, *Pieris napi*.**

Developmental plasticity describes the capacity of individuals with the same genotype to induce permanent change in a phenotype depending on a specific external input (Stearns 1989). As there is often genetic variation for the responsiveness to external stimuli, developmental plasticity can evolve by natural selection. This may produce adaptive plasticity, allowing organisms to express different phenotypes in different environments to optimize fitness in a context-dependent manner (Gotthard and Nylin 1995; Nettle and Bateson 2015). To unravel the evolutionary dynamics of this process, the underlying genetic variation in adaptive plasticity among natural populations needs to be identified (Lafuente et al. 2018).

One well-studied example of adaptive developmental plasticity is the induction of facultative diapause in insects (Danilevskii 1965; Tauber et al. 1986; Leimar et al. 2006; Gotthard and Berger 2010; Kivelä et al. 2017). This plasticity allows insects to express several generations during the favorable time of the year, while still having the capacity to enter diapause when the nonfavorable season approaches. Diapause is characterized by a suppression of development and reduced metabolic rate, which enables insects to survive until favorable conditions return (Tauber et al. 1986). Typically, environmental cues such as a decreasing photoperiod and temperature signal the approaching end of the growing season, inducing a switch from direct development to the diapause program (Lees 1955). As the change in photoperiod over the seasons is highly consistent between years, photoperiod often supersedes temperature as a reliable signal for initiating diapause (Lindestad et al. 2019). The photoperiodic signal that induces diapause is latitude-specific for many species,

[#] KG and CWW should be considered joint senior authors.

revealing an adaptive landscape of photoperiod-based diapause induction (Kimura 1988; Hahn and Denlinger 2010; Paolucci et al. 2013; Aalberg Haugen and Gotthard 2015; Lindestad et al. 2019).

Studies investigating the inheritance of diapause induction have suggested diverse genetic backgrounds, from single locus to polygenic inheritance (Kurahashi and Ohtaki 1977; Lumme and Keränen 1978; Doležel et al. 2005; Kawakami et al. 2010; Söderlind and Nylin 2011; Lehmann et al. 2016; Pruisscher et al. 2017). However, only few studies have performed genome-wide scans to identify genes affecting the induction decision (Pruisscher et al. 2018). As it appears unlikely that the genetic basis for variation in diapause induction is completely idiosyncratic to each species, it is important to explore potential commonalities across species to enable the study of shared physiological mechanisms and diapause evolution.

One influential hypothesis concerning photoperiodic induction of diapause is the so-called Bünning hypothesis, which posits that the circadian clock, a biochemical oscillator that cycles in a 24-hour period entrained by day light, is involved in the measurement of day length (Bünning 1936). Studies across a range of insects have implicated a link between circadian clock genes and diapause induction (Ikeno et al. 2010; Paolucci et al. 2016; Pruisscher et al. 2018), of which one notable example found alternative isoforms of the gene *timeless* that correlated with an adaptive cline in photoperiodic response (Tauber et al. 2007). However, there is also evidence suggesting the circadian clock and photoperiodic induction of diapause to be independent of each other (Emerson et al. 2009b; Bradshaw et al. 2012a). This would suggest that specific clock genes can act pleiotropically on diapause induction, or that the involvement of these genes is taxon-specific (Emerson et al. 2009a).

The present study investigates the genomic basis of population differences in diapause induction in the green-veined white butterfly *Pieris napi* (Lepidoptera, Pieridae). This species shows an adaptive cline across latitudes in the photoperiodic induction of diapause (Kivelä et al. 2015; Pruisscher et al. 2017), and also to a lesser extent in the effect of temperature on diapause induction (Kivelä et al. 2015). In a previous study *P. napi* populations from Barcelona (northern Spain) and Abisko (northern Sweden) were crossed to investigate the inheritance of the photoperiodic induction of diapause (Pruisscher et al. 2017). In the wild, the Barcelona populations may have up to four annual generations, whereas Abisko populations only have one generation per year. The growing season in Abisko is so short that these populations never express the direct development pathway. This adaptive difference persists in common garden conditions, indicating a genetic basis for the difference in the propensity to induce diapause (Pruisscher et al. 2017). Initial investigations into the genetic architecture of diapause induction using F1-hybrid crosses and backcrosses with the Abisko population identified a strong condition-dependent sex-linked component with an additional polygenic autosomal composition (Pruisscher et al. 2017).

The specific aim of this article was to identify high-quality candidate genes for mechanistic insights and future analysis on variation in diapause induction, and to place these results in the larger context of adaptive developmental plasticity. To do this, we first characterized overall genome-wide differences between the *P. napi* populations of Barcelona and Abisko using a pooled sequencing approach. Second, we sequenced population crosses of a previous study (Pruisscher et al. 2017), and performed a bulk segregant analysis (BSA) approach to identify the genomic regions associated with the diapause induction decision in five separate, between-population backcrosses. This replicated family design for our BSA allows us to gain general insights into the population level differences of induction, and to narrow down the chromosomal region of interest involved in this locally adapted reaction norm. We then investigated these genomic regions of interest to assess whether there was an enrichment of divergent regions in circadian clock genes.

Methods

SAMPLING DESIGN

For the full rearing design refer to Pruisscher et al. (2017). Briefly, in 2014 a population of Abisko (Sweden, 68.36°N, 18.79°E), that had already spent one generation in diapause in the laboratory, was crossed with a newly caught population sample from Barcelona (Spain, 42.23°N, 3.10°E) to generate F1 hybrids. The Abisko population was also crossed within itself to continue the Abisko line (independent families to avoid inbreeding). In 2015, the F1 hybrids and pure population line of Abisko were used to generate backcrosses. In the present study we used offspring from crosses between an Abisko female and F1-hybrid males of both reciprocal crosses (i.e., either an Abisko female and a Barcelona male, or an Abisko male and a Barcelona female). In particular, three families from an Abisko × ♀Abisko/♂Barcelona cross, and two families of an Abisko × ♀Barcelona/♂Abisko cross were used (Table 1). All offspring from these crosses were reared under a light:dark cycle of 23 hours:1 hour, at 20°C. After pupating, each offspring was put into individual cups and monitored daily for eclosion. Pupae that had not eclosed within three weeks, were considered to be in diapause. Diapausing individuals were put into constant darkness at 2°C for five months, after which they were brought back into the rearing conditions to eclose. All eclosed individuals were sampled at day 2 of adult life by putting them into individual storage containers into a freezer at −80°C.

Table 1. Five backcrosses and the number of direct developing and diapausing individuals for each sex under 23 hours of light per day at 20°C.

Family	F × M	Males Direct	Diapause	Diapause (%)	Females Direct	Diapause	Diapause (%)
106	A × AB	18	43	70.5	30	39	56.5
110	A × AB	14	82	85.4	27	51	65.4
115	A × AB	9	80	89.9	34	66	66
128	A × BA	31	54	63.5	35	46	56.8
135	A × BA	11	35	76.1	27	26	49.1

SEQUENCING

The initial results of Pruisscher et al. (2017) indicated an inheritance of diapause induction linked to the Z chromosome, a sex-linked effect. As females are the heterogametic sex in Lepidoptera (they have one Z and one W chromosome), the female offspring of these five crosses were sequenced to investigate candidate genes for variation in diapause induction.

DNA was extracted from each individual using a DNeasy blood and tissue kit (Qiagen) with an extra RNase A treatment to remove potential RNA contamination. DNA quality was checked on 2% agarose gels stained with GelRed, to ensure minimal fragmentation, and UV-Vis spectrometer (NanoDrop 8000; Thermo Scientific) to assess purity. All samples showed minimal fragmentation on a gel, and high purity with an absorbance 260/280 >1.7 and <2.0. For each family and pathway, samples were combined at equal concentration, resulting in 10 pools of 5 μg RNA-free gDNA. SciLifeLab (Uppsala, Sweden) performed the library preparation (Illumina TruSeq DNA PCR-free library) and sequencing (Illumina HiSeq2000, 100-bp paired-end reads, 450 bp insert size).

READ FILTERING

PCR duplicates were removed from the raw sequencing files using the clone_filter script of Stacks-1.21 (Catchen et al. 2013), after which Illumina sequencing adaptors were removed and reads were quality trimmed to a minimum Phred base quality of 20 as well as discarding broken read pairs, using BBDUK2 (BBMap v34.86, Bushnell, http://sourceforge.net/projects/bbmap/).

MAPPING

All Pool-Seq read data were mapped to the *P. napi* v1.1 genome assembly (Hill et al. 2019) using NEXTGENMAP v0.5.0 at default settings (Sedlazeck et al. 2013). Alignments were filtered using samtools v1.6 (Li 2009), only keeping reads that mapped correctly as pairs (Table S1).

FST DIVERGENCE

Differentiation between the pure populations, as well as for each within-family pair of direct development versus diapause, was quantified by combining the filtered pairs of read sets into an mpileup file using samtools v1.6 (Li 2009), in which indels and 2 bp on either side of the indel were masked using popoolation v1.2.2 (Kofler et al. 2011a). Next, FST was calculated in 50 kb nonoverlapping windows using popoolation2 v1201 (Kofler et al. 2011b), keeping windows where >50% of a window had a read coverage between 10 and 500.

SYNTENY WITH *Bombyx mori* AND *Zerene cesonia*

To investigate the chromosomal location of *P. napi* modScaffolds that were not associated to a chromosome, we aligned these modScaffolds and the Z chromosome of *P. napi* against the chromosome assembly of *Zerene cesonia*, which is in the same family as *P. napi* (Pieridae) (Rodrigues-Caro et al Rodriguez-Caro et al. 2020), at the DNA level using Whole Genome VISTA webserver (Mayor et al. 2000; Frazer et al. 2004), which is a wrapper for Shuffle-LAGAN, a global alignment algorithm (Brudno et al. 2003). We also used the protein sequences of these scaffolds to search against the protein sequences of the silk moth *Bombyx mori* using DIAMOND v0.9.10 (Buchfink et al. 2015) on default settings, extracting the top hits for each protein sequence. Both analyses gave similar results, thus only the alignment with *Z. cesonia* is shown.

GENE SET ENRICHMENT ANALYSIS

Gene set enrichment analysis was used to investigate the candidate region for enrichment of gene sets, using topGO considering Parent-Child relationships (Alexa and Rahnenfuhrer 2018), comparing the set of genes present in the candidate region against the rest of the annotated genes in the genome.

To test specifically for an enrichment of circadian clock genes, the GO term GO:0032922, circadian regulation of gene expression, was added to the annotation of the genome for a custom set of 33 genes (Table S2). These genes represent the orthologs of 35 components of the circadian clock as identified in the monarch butterfly *Danaus plexippus* (Merlin et al. 2009). These genes were identified in *P. napi* using DIAMOND v0.9.10 (Buchfink et al. 2015) on default settings, comparing the *P. napi* protein sequences to the *Danaus plexippus* protein sequences of

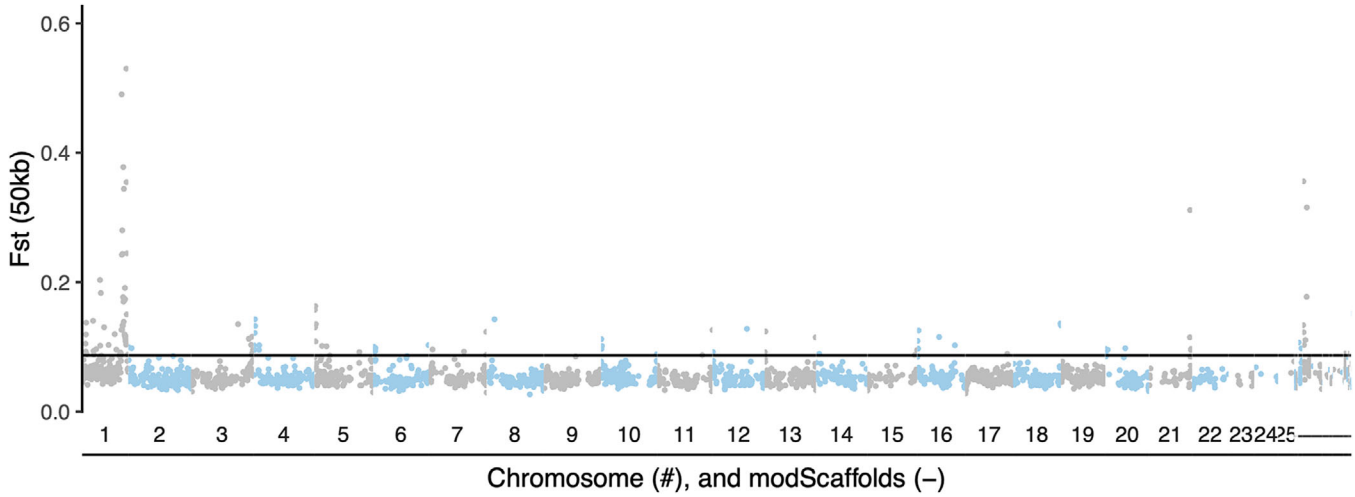

Figure 1. Genome-wide divergence measured as FST in 50 kb nonoverlapping windows between Barcelona and Abisko. The solid horizontal line indicates the 95th percentile at FST = 0.087. The 25 chromosomes are indicated by number, and modScaffolds are indicated by a dash on the *X*-axis. Chromosome 1 is the Z chromosome.

these genes (Table S2), and selecting genes with a minimum identity of 50%, and an *E*-value $\leq 1.0 \times 10^{-10}$.

Results

DIVERGENCE BETWEEN BARCELONA AND ABISKO

The *P. napi* genome assembly consists of 25 chromosomes, 1 mitochondrial sequence, and 2943 unplaced "modScaffolds" for which the chromosomal location is unknown (Hill et al. 2019). A total of 91.7% of the reads from the Barcelona population, and 92% of the reads from the Abisko populations mapped to this assembly (Table S1). Differentiation between the populations measured as FST in 50 kb nonoverlapping windows revealed an overall low population divergence between Barcelona and Abisko (FST mean = 0.056, SD = 0.026). When using the 95th percentile (FST > 0.087) as criterion for being an outlier, there were windows present on most chromosomes that fulfilled this, but there was a notable concentration of outlier windows on chromosome 1, which is the Z chromosome in the assembly (Fig. 1).

DIFFERENTIATION IN BACKCROSSES

The diapause induction decision was investigated using the offspring of five crosses between Abisko females and F1-hybrid males (Table 1), from a larger previous study that investigated the inheritance of diapause induction in this species (Pruisscher et al. 2017). Female butterflies inherit their single Z chromosome from their father who produce Z chromosomes that have gone through at least one recombination event. The F1-hybrid males were the result of a cross between Abisko and Barcelona, resulting in their daughters being hemizygous at any given locus on the Z for either the Abisko or Barcelona allele. The expectation in this study was that diapausing females would carry an Abisko allele at the locus of interest, and the directly developing females to have a Barcelona allele. Chi-square tests on the propensity to enter diapause revealed that in the offspring of these backcrosses, the males showed *a* ≈77% diapause incidence with a difference between crosses ($X^2 = 22.87$, $P < 0.001$), whereas in females the propensity of ≈59% was not significantly different between crosses ($X^2 = 5.77$, $P = 0.217$). In the females, this pattern resembled a sex chromosome linked inheritance of diapause induction (Pruisscher et al. 2017), which predicts the expression of diapause in half of the individuals because that is the proportion carrying Z-linked diapause induction genes from Abisko.

To identify which of the genomic outliers that were detected between the populations could be involved in diapause adaptation, the females of the five backcrosses were sequenced in pools; five pools for direct and five pools for diapause. More than 91% of the reads mapped to the assembly for each of the 10 pools (Table S1). Differentiation between these direct development and diapause pools as measured using FST in 50 kb nonoverlapping windows revealed three highly divergent regions among a uniform background of no differentiation, consistent over all the five crosses (Fig. 2A).

Outlier regions for each cross were defined as having a window-based FST above the 95th percentile, corresponding to FST > 0.095 in family 106, FST > 0.14 in family 110, FST > 0.256 in family 115, FST > 0.141 in family 128, and FST > 0.25 in family 135. More than 71% of all unique outlier regions were shared between all five crosses (Fig. 2B), and these outliers were located in only three genomic regions: these were Chromosome 1, and two anonymous smaller contigs called modScaffold_17_1, and modScaffold_95_1 (Table 2).

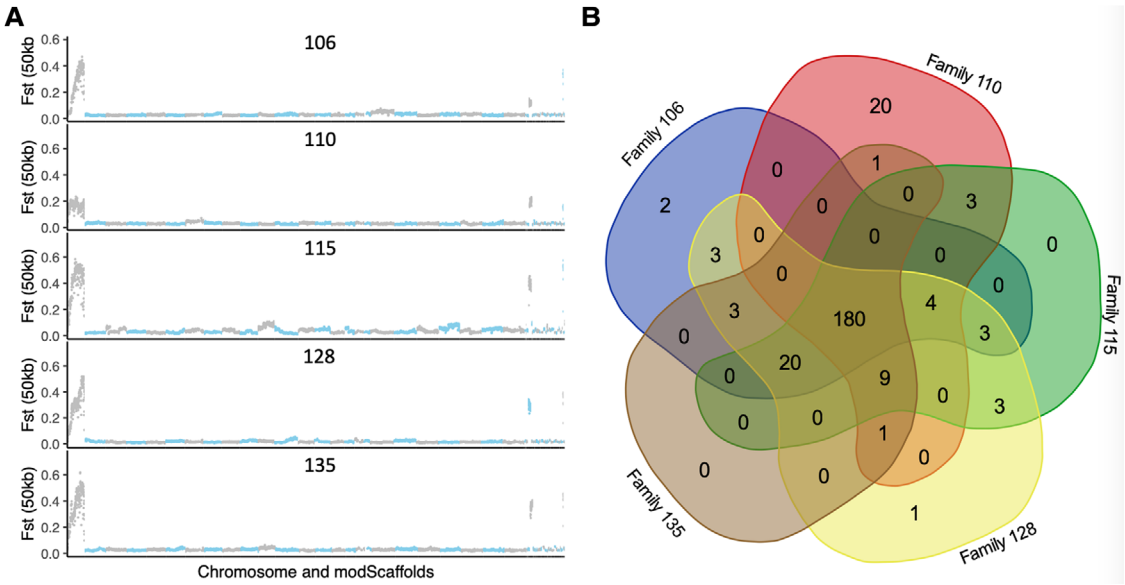

Figure 2. (A) Genome-wide divergence measured as FST in 50 kb nonoverlapping windows between direct development and diapause pools of five backcrosses. (B) Venn diagram of the overlap of the most divergent regions between crosses. Of the total of 253 unique regions, 180 (71%) were shared between all five families, and 213 (84%) were common to at least four families.

Table 2. Number of outlier windows (>95th percentile) shared in the backcross families for the three regions they occur in.

Present in *n* families	Region Chr_1	modScaf_17_1	modScaf_95_1
Five	136	37	7
Four	30	3	1
Three	1	1	0
Two	9	0	1
One	23	0	0

To place the two anonymous scaffolds in a chromosomal context, a synteny analysis was performed using the *B. mori* proteome and the protein sequences of the genes present on these two scaffolds, In total 66 of the 77 genes found on modScaffold_17_1 were found on *B. mori* chromosome 1, as well as 26 of the 28 genes found on modScaffold_95_1, providing evidence that these scaffolds are also part of chromosome 1 in *P. napi*. Synteny of these scaffolds was also explored using nucleotide alignment with Pierid butterfly *Z. cesonia*, confirming the protein synteny with *B. mori* (Figs. 3D and S1).

CANDIDATE GENES FOR DIAPAUSE INDUCTION

The 180 windows that showed consistent divergence between phenotypes in all five backcrosses were intersected with the outliers of the population comparison (FST > 95th percentile). This yielded 46 windows of overlap, containing a total of 100 annotated genes. 89 of these genes were located on the terminal end of chromosome 1, together with 11 on modScaffold_17_1, and none on modScaffold_95_1 (Fig. 3).

Within these 100 genes, a total of 25 genes contained 219 SNPs with an FST > 0.52 (95th percentile) within their exon boundaries (Table 3). When only examining fixed differences (FST = 1.0) a total of three genes contained 25 SNPs. The gene annotated as *period circadian protein* contained the majority of SNPs at both FST cutoffs (Table 3). Of the in total 25 fixed SNPs between populations, 21 SNPs showed nonsynonymous changes: one in the gene *Protein zer-1 homolog*, three in the gene *Dynein heavy chain 6, axonemal*, and 17 in *period circadian protein* (Table S3).

GENE SET ENRICHMENT ANALYSIS

To conduct an enrichment analysis for certain gene sets, the 100 genes as well as the subset of 25 genes were compared against the rest of the genome. This revealed an enrichment (P <0.01) for circadian regulation of gene expression, as well as signal transduction processes and signal transducer activity (Table 4). The most significant GOterm in both comparisons was linked to three genes, annotated as *cycle, clock*, and *period circadian protein*.

Discussion

Here we demonstrate that the genetic basis for population variation in diapause induction in the butterfly *P. napi* is predominantly located on the Z chromosome. Using a pooled sequencing approach to identify divergent regions between our two focal populations, and a bulk-segregant analysis on

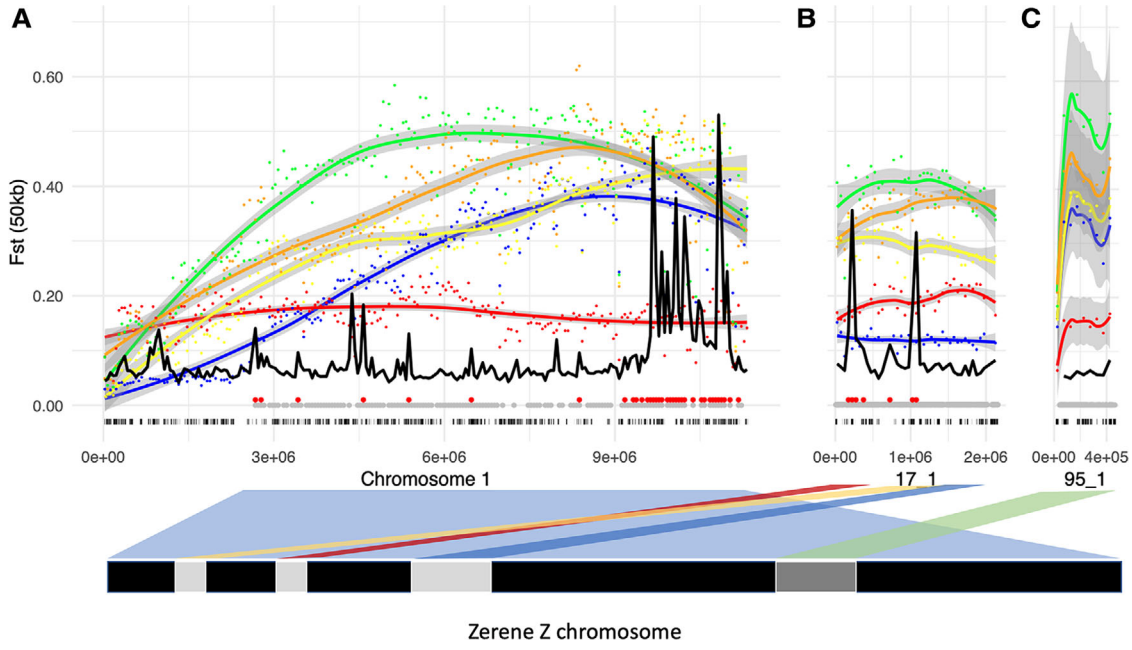

Figure 3. QTL results and genome-wide divergence in 50 kb windows for (A) Chromosome 1, (B) modScaffold_17_1, and (C) modScaffold_95_1. The black line indicates population differentiation (FST) as measured in 50 kb windows. The colored dots represent the FST values of the comparisons between direct and diapausing individuals in the five crosses: family 106 is blue, 110 is red, 115 is green, 128 is yellow, and 135 is orange, and the corresponding lines are smoothed means and their standard deviations. The gray dots on the *X*-axis represent the 180 outlier windows shared between the five crosses, and red dots represent the 46 outlier windows shared between comparisons of phenotypes within the crosses and the population comparisons. Black vertical bars below the *X*-axis represent gene models present on the scaffolds. (D) Synteny plot showing orthology between *P. napi* Chromosome 1, modScaffolds 17_1 and 95_1, and *Zerene cesonia* Z chromosome.

backcrosses between these populations to directly interrogate the genome for regions involved in diapause induction, we were able to associate genes from this region with diapause induction differences between Abisko and Barcelona populations. These findings corroborate the sex-linked effect found in the results of earlier work on the inheritance of diapause in this species (Pruisscher et al. 2017). Here, our genomic approach allowed us to identify high-quality candidate genes on the Z chromosome associated with diapause induction, revealing an enrichment for the circadian clock, represented by *cycle, clock*, and *period*, and revealing fixed nonsynonymous substitutions in the genes: *protein zer-1*, *dynein heavy chain 6*, and *period*.

Given the large geographical distance between Barcelona and Abisko (≈4000 km), the overall genetic differentiation between the populations of *P. napi* was low (FST = 0.056). Other studies of butterflies in Europe have found higher genetic divergence at substantially shorter distances: *Pararge aegeria*, populations showed an FST of 0.09 at ≈900 km distance in Sweden (Pruisscher et al. 2018), and populations of *Melanargia galathea* around the Carpathian basin showed an FST of 0.07 at across ≈800–1200 km (Schmitt et al. 2006). The relatively low level of genetic differentiation between our distant *P. napi* populations indicates a reasonably high level of gene flow across latitudes. Against this background of low FST, the outlier regions of divergence identified here are striking and highly suggestive of local adaptation due to selection against foreign alleles at these loci (Fig. 1).

There was a strong overlap of the most divergent regions between populations and the outlier regions between phenotypes (diapause and direct development) in the backcrosses between F1 hybrids and the northern populations. This allowed us to associate the diapause induction decision with candidate genes that show significant divergence among the populations. The region of divergence is very compact, and for these genes clumped together showing divergence, it is not unlikely that most of these genes show divergence because they are in linkage with the loci under selection, although further testing would be required to identify the causal genes. We identify three genes with nonsynonymous changes: *zer1* is a protein involved in ubiquitin ligase (Vasudevan et al. 2007), *dynein heavy chain 6* is a protein involved in generating force of cilia, and shows differences in protein levels between resting and active embryos in the invertebrate *Brachionus plicatilis* (Ziv et al. 2017), while *period* is part of the circadian clock. One caveat that must be stated is that although

Table 3. Genes with SNPs above 95th percentile (FST > 0.52) and fixed (FST = 1) in exons.

					SNPs in exons	
Region	Start	Gene model	Gene ID	Gene description	FST > 0.52	FST = 1
Chr_1	2681865	PIENAPG00000001687	cycle	Cycle	4	
Chr_1	2694150	PIENAPG00000004714	asrij	OCIA domain-containing protein 1	2	
Chr_1	3442616	PIENAPG00000001823		Hypothetical protein	3	
Chr_1	4563642	PIENAPG00000008013	clock	Circadian locomoter output Cycles protein kaput	5	
Chr_1	6463591	PIENAPG00000009908	lana	Laminin subunit alpha	3	
Chr_1	9609468	PIENAPG00000008902	med23	Mediator of RNA polymerase II transcription subunit 23	3	
Chr_1	9678631	PIENAPG00000012168	socs7	Isoform 2 of Suppressor of cytokine signaling 7	11	
Chr_1	9784988	PIENAPG00000007342	pgrp-lc	Isoform x of Peptidoglycan-recognition protein LC	1	
Chr_1	9804877	PIENAPG00000012797	nach	Sodium channel protein Nach	3	
Chr_1	9829607	PIENAPG00000001733	slc24a4	Isoform 3 of sodium/potassium/calcium exchanger 4	6	
Chr_1	10068617	PIENAPG00000004734	cg12084	Protein zer-1 homolog	11	2
Chr_1	10090109	PIENAPG00000008181	gs2_2	Glutamine synthetase 2 cytoplasmic	3	
Chr_1	10147954	PIENAPG00000000462		Hypothetical protein	1	
Chr_1	10182292	PIENAPG00000001970	ttc39b	Tetratricopeptide repeat protein 39B	2	
Chr_1	10743887	PIENAPG00000001929		Hypothetical protein	1	
Chr_1	10820560	PIENAPG00000002997	dnah6	Dynein heavy chain 6, axonemal	40	5
Chr_1	10837977	PIENAPG00000004980	dnah6_2	Dynein heavy chain 6, axonemal	3	
Chr_1	10856669	PIENAPG00000012635	dnah6_3	Dynein heavy chain 6, axonemal	21	
Chr_1	10874137	PIENAPG00000010877	wac	WW domain-containing adapter protein w. coiled-coil	16	
Chr_1	10886099	PIENAPG00000001980	pfdn1	Prefoldin subunit 1	1	
Chr_1	10899791	PIENAPG00000004775	ralgps1	Ras-specific guanine nucleotide-releasing factor RalGPS	3	
Chr_1	10999784	PIENAPG00000001119	slc5a12	Sodium-coupled monocarboxylate transporter 2	1	
Chr_1	11009850	PIENAPG00000012945	slc24a4_	Isoform 3 of Sodium/potassium/calcium exchanger 4	3	
mS_17	209488	PIENAPG00000004337	period	Period circadian protein	68	18
mS_17	996594	PIENAPG00000002439	unc-89	Muscle M-line assembly protein unc-89	4	

Shown is the position and the number of SNPs in exons.

this genetic variation is suggestive, differences might also arise due to expression-level changes or expression of alternative isoforms from variation present outside of exons.

It is striking that this type of genome-wide scan for genetic variation associated with variation in diapause induction strongly implies a chromosomal region containing three important clock genes. However, because of the close chromosomal proximity of all the divergent genes, it is difficult to directly link selection for differences in diapause induction to differences in the clock, particular clock genes or other genes in the outlier chromosomal region. Nevertheless, it is suggestive that the gene *period* showed the highest SNP variation of any of the identified genes. Moreover, that SNP variation in clock genes correlate with photoperiodic diapause induction agrees closely with a previous

Table 4. Significant GO terms with a *P*-value <0.01 for the 100 genes within the 46 outlier windows or in the 25 genes with 95th percentile SNPs in exons.

			100 Genes			25 Genes		
GO.ID	Term	Annotated	Obs.	Exp.	*P*-value	Obs.	Exp.	*P*-value
GO:0007623	Circadian rhythm	33	4	0.34	0.00031	3	0.14	0.00029
GO:0032922	Circadian regulation of gene expression	33	4	0.34	0.00031	3	0.14	0.00029
GO:0048511	Rhythmic process	33	4	0.34	0.00031	3	0.14	0.00029
GO:0007165	Signal transduction	350	10	3.59	0.0021	6	1.45	0.0019
GO:0007154	Cell communication	355	10	3.64	0.0023	6	1.48	0.0020
GO:0023052	Signaling	355	10	3.64	0.0023	6	1.48	0.0020
GO:0044700	Single-organism signaling	355	10	3.64	0.0023	6	1.48	0.0020
GO:0051716	Cellular response to stimulus	393	10	4.03	0.0049	6	1.63	0.0034
GO:0050896	Response to stimulus	426	11	4.37	0.0027	6	1.77	0.0051
GO:0044763	Single-organism cellular process	822	14	8.43	0.0274	8	3.42	0.0096

Given is the GOterm ID, how many genes are annotated within the genome with this term, and for each analysis the expected number of genes, the significant number of genes, and the *P*-value for the gene set enrichment analysis. The list is sorted by significance in the 25 genes (see Tables S4 and S5 for the full GO lists).

genome-wide association study in another temperate butterfly, the Speckled wood (Pruisscher et al. 2018). In that study variation in the *period* gene also showed an association with diapause induction, although the strongest effect was due to variation in the autosomal clock gene *timeless*. In combination these results strongly suggests that the evolution of local differences in photoperiodic induction of diapause is partly due to evolutionary dynamics of clock genes. That partly different circadian clock genes are implicated in different species is interesting (Tauber et al. 2007; Ikeno et al. 2010; Paolucci et al. 2016; Pruisscher et al. 2018; Kozak et al. 2019) and consistent with the hypothesis that these effects on photoperiodism are to some degree effectuated by differences in the circadian clock itself, rather than being pleiotropic effects of the clock genes on other processes. If the effects of clock genes on photoperiodism would be entirely due to pleiotropic effects of some of those genes, it seems likely that the identity of genes associated with photoperiodism should show little variation among species. In contrast, the circadian clock is dependent on many different proteins and their interactions, and it seems plausible that several alternative mutations in circadian clock genes may have similar effects on the clock. If changes in the clock itself underlies variation in photoperiodism, we expect more variation among species in exactly which clock genes associate with variation in photoperiodism. In line with this argument a recent study showed that specific genetic variation at clock genes, period (per) and pigment-dispersing factor receptor, simultaneously affected both photoperiodic-dependent diapause termination and circadian behavior of adult moths (Kozak et al. 2019).

Photoperiodic induction of diapause in insects typically shows relatively continuous variation across latitudes and this is also true for *P. napi* (Kivelä et al. 2015; Pruisscher et al. 2017). Here we have compared populations from different ends of the latitudinal distribution of *P. napi* and identified a region of the genome that has large effects and is fixed among populations. This genetic variation alone cannot explain the continuous adaptation of diapause induction across latitudes. Instead, it is very likely that much of the additional adaptive variation in photoperiodism that occur at latitudes intermediate to these two populations is due to still unknown genetic variation. This may be genetic variation at many other loci of small effects and/or to the presence of many yet undescribed alleles for loci located in the divergent region of the Z chromosome discovered here. Indeed, the variation among our families in diapause induction may be reflecting variation at these other loci. Additionally, several other studies have identified allelic variation at different clock genes across latitudinal distributions of other insects (Mathias et al. 2007; Tauber et al. 2007; Yamada and Yamamoto 2011; Paolucci et al. 2016; Pruisscher et al. 2018; Kozak et al. 2019). In any case, our results highlight a distinct chromosomal region for exploring genetic variation underlying adaptive variation in the photoperiodic induction of diapause in *P. napi* and potentially other insect species.

One notable feature is that the strongest differentiation between the populations is aggregated on the sex chromosome. Diapause induction appears to have a sex chromosome linked inheritance in several insect species (Hagen and Scriber 1989; Ikten et al. 2011; Chen et al. 2014; Fu et al. 2015; Pruisscher et al.

2018) and in the Lepidoptera a disproportional number of polymorphic traits appear to be sex-linked, even though the sex chromosomes only represent a fragment of the entire genome (Sperling 1994; Janz 1998). As sex chromosomes and autosomes are inherited differently, the relative rates of evolution are often expected to differ between sex-linked and autosomal genes. The so-called faster X-effect (Charlesworth et al. 1987) has emphasized that genes on sex chromosomes should adapt more quickly than genes on autosomes when beneficial mutations show some level of recessiveness (Charlesworth et al. 1987; Meisel and Connallon 2013). However, recent developments of the theory focusing in particular on local adaptation within species suggests that patterns of sex-specific migration may be more important that dominance relationships for explaining faster divergence of sex chromosomes under local adaptation (Lasne et al. 2017; Lasne et al. 2019). In fact, these models indicate that, under a wide set of assumptions, local adaptation is predicted to be more pronounced on the sex chromosomes compared to autosomes, when migration is biased toward the heterogametic sex. Unfortunately, there is no information on the potential sex-dependent dispersal of *P. napi* and further studies are needed to elucidate whether these differences in diapause induction arose through a faster X effect, or if the genes needed to create variation in this trait happen to be located on the sex chromosome.

It is still unknown whether candidate loci for adaptive plasticity affect multiple traits, or if they are specific to a given plastic response such as diapause induction, and if so, at what stage they do act. It is possible that the genes identified here are involved in the mechanisms that interpret environmental conditions at different stages of the induction process, from the perception of the environmental signal to the interpretation and transmission of the signal to downstream processes. If the genetic changes observed here are involved in the perception of the environmental cue, they could theoretically have effects on multiple traits at the same time as it would be more upstream in the induction process compared to being part of more downstream signaling pathways.

The evolution of plasticity depends on the strength of selection and the predictability of environmental fluctuations (Leimar et al. 2006). To better understand the evolution of adaptive plasticity it is necessary to characterize its genetic background, and this has only been performed in very few studies (Czypionka et al. 2018). One open point of discussion on adaptive plasticity considers that either specific loci are responsible for determining a plastic response, or alternatively that loci are co-opted by a number of traits that are adapted to a certain environment (Via et al. 1995; Sgrò et al. 2016). In the case of diapause induction, this study suggests that natural variation in the plastic response to photoperiod is related to genetic variation at a genomic region that includes several circadian clock genes. To what degree the circadian clock itself has a causal effect on diapause induction is still not clear (Bradshaw and Holzapfel 2017), but it is tempting to speculate that the clock, and its underlying genetic background, may be a molecular mechanism for many examples of plasticity in life-cycle timing.

AUTHOR CONTRIBUTIONS

PP, SN, KG, and CWW designed research; PP performed research and analyzed data; PP, KG, and CWW wrote the manuscript.

ACKNOWLEDGMENTS

The authors would like to acknowledge the support from Science for Life Laboratory, the National Genomics Infrastructure, NGI and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. Funding was provided by the Swedish Research Council (VR grant numbers 2012-3715, 2010- 5341, 621-2012-4001), Academy of Finland (grant number 131155), the Bolin Centre for Climate Research at Stockholm University and the Knut and Alice Wallenberg Foundation (grant number 2012.0058). Sequencing data are available at NCBI Bioproject PRJNA684967.

DATA ARCHIVING

The archival location is available upon acceptance.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

LITERATURE CITED

Aalberg Haugen, I. M., and K. Gotthard. 2015. Diapause induction and relaxed selection on alternative developmental pathways in a butterfly. J. Anim. Ecol. 84:464–472.

Alexa, A., and J Rahnenfuhrer. (2018). topGO: Enrichment Analysis for Gene Ontology R package.

Bradshaw, W. E., and C. M. Holzapfel. 2017. Natural variation and genetics of photoperiodism in *Wyeomyia smithii*. Adv. Genet. 99:39–71.

Bradshaw, W. E., K. J. Emerson, and C. M. Holzapfel. 2012a. Genetic correlations and the evolution of photoperiodic time measurement within a local population of the pitcher-plant mosquito, *Wyeomyia smithii*. Heredity 108:473–479.

Brudno, M., S. Malde, A. Poliakov, C. B. Do, O. Couronne, I. Dubchak, and S. Batzoglou. 2003. Glocal alignment: finding rearrangements during alignment. Bioinformatics 19:i54–i62.

Bünning, E. 1936. Die endonome Tagesrhythmik als Grundlage der photoperiodischen Reaktion. Berichte der Deutschen Botanischen Gesellschaft 54:590–607.

Buchfink, B., C. Xie, and D. H. Huson. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12:59–60. https://doi.org/10.1038/nmeth.3176. Epub 2014 Nov 17. PMID: 25402007

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

Charlesworth, B., J. A. Coyne, and N. H. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. 130:113–146.

Chen, C., L. Xiao, H. M. He, J. Xu, and F. S. Xue. 2014. A genetic analysis of diapause in crosses of a southern and a northern strain of the cabbage beetle *Colaphellus bowringi* (Coleoptera: chrysomelidae). Bull. Entomol. Res. 104:586–591.

Czypionka, T., P. D. Fields, J. Routtu, E. van den Berg, D. Ebert, and L. De Meester. 2018. The genetic architecture underlying diapause termination in a planktonic crustacean. Mol. Ecol. 28:998–1008.

Danilevskii, A. S. 1965. Photoperiodism and seasonal development of insects. Oliver and Boyd, Edinburgh.

Doležel, D., H. Vaněčková, I. Šauman, and M. Hodkova. 2005. Is period gene causally involved in the photoperiodic regulation of reproductive diapause in the linden bug, *Pyrrhocoris apterus*? J. Insect Physiol. 51:655–659.

Emerson, K. J., W. E. Bradshaw, and C. M. Holzapfel. 2009a. Complications of complexity: integrating environmental, genetic and hormonal control of insect diapause. Trends Genet. 25:217–225.

Emerson, K. J., S. J. Dake, W. E. Bradshaw, and C. M. Holzapfel. 2009b. Evolution of photoperiodic time measurement is independent of the circadian clock in the pitcher-plant mosquito, *Wyeomyia smithii*. J. Comp. Physiol. A 195:385–391.

Frazer, K. A., L. Pachter, A. Poliakov, E. M. Rubin, and I. Dubchak. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32:W273–W279.

Fu, S., C. Chen, L. Xiao, H. He, and F. Xue. 2015. Inheritance of diapause in crosses between the northernmost and the southernmost strains of the asian corn borer *Ostrinia furnacalis*. PLoS One 10:e0118186.

Gotthard, K., and D. Berger. 2010. The diapause decision as a cascade switch for adaptive developmental plasticity in body mass in a butterfly. J. Evol. Biol. 23:1129–1137.

Gotthard, K., and S. Nylin. 1995. Adaptive plasticity and plasticity as an adaptation: a selective review of plasticity in animal morphology and life history. Oikos 74:3.

Hagen, R. H., and J. M. Scriber. 1989. Sex-linked diapause, color, and allozyme loci in *Papilio glaucus*: linkage analysis and significance in a hybrid zone. J. Hered. 80:179–185.

Hahn, D. A., and D. L. Denlinger. 2010. Energetics of Insect Diapause. Ann. Rev. Entomol. 56:103–121.

Hill, J., P. Rastas, E. A. Hornett, R. Neethiraj, N. Clark, N. Morehouse, M. de la Paz Celorio-Mancera, J. C. Cols, H. Dircksen, C. Meslin et al. 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci. Adv. 5:eaau3648.

Ikeno, T., S. I. Tanaka, H. Numata, and S. G. Goto. 2010. Photoperiodic diapause under the control of circadian clock genes in an insect. BMC Biol. 8:1.

Ikten, C., S. R. Skoda, T. E. Hunt, J. Molina-Ochoa, and J. E. Foster. 2011. Genetic variation and inheritance of diapause induction in two distinct voltine ecotypes of *Ostrinia nubilalis* (Lepidoptera: Crambidae). Ann. Entomol. Soc. Am. 104:567–575.

Janz, N. 1998. Sex-linked inheritance of host-plant specialization in a polyphagous butterfly. Proc. Proc. R. Soc. Biol. Sci. 265:1675–1678.

Kawakami, Y., H. Numata, K. Ito, and S. G. Goto. 2010. Dominant and recessive inheritance patterns of diapause in the two-spotted spider mite *Tetranychus urticae*. J. Hered. 101:20–25.

Kimura, M. T. 1988. Interspecific and geographic variation of diapause intensity and seasonal adaptation in the *Drosophila auraria* species complex (Diptera: Drosophilidae). Funct. Ecol. 2:177.

Kivelä, S. M., B. Svensson, A. Tiwe, and K. Gotthard. 2015. Thermal plasticity of growth and development varies adaptively among alternative developmental pathways. Evolution 69:2399–2413.

Kivelä, S. M., M. Friberg, C. Wiklund, and K. Gotthard. 2017. Adaptive developmental plasticity in a butterfly: mechanisms for size and time at pupation differ between diapause and direct development. Biol. J. Linn. Soc. 122:46–57.

Kofler, R., P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol, and C. Schlötterer. 2011a. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6:e15925.

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

Kozak, G. M., C. B. Wadsworth, S. C. Kahne, S. M. Bogdanowicz, R. G. Harrison, B. S. Coates, and E. B. Dopman. 2019. Genomic basis of circannual rhythm in the European corn borer moth. Curr. Biol. 29:3501–3509.

Kurahashi, H., and T. Ohtaki. 1977. Crossing between nondiapausing and diapausing races of *Sarcophaga peregrina*. Experientia 33:186–187.

Lafuente, E., D. Duneau, and P. Beldade. 2018. Genetic basis of thermal plasticity variation in *Drosophila melanogaster* body size. PLoS Genet. 14:e1007686.

Lasne, C., C. M. Sgrò, and T. Connallon. 2017. The relative contributions of the X chromosome and autosomes to local adaptation. Genetics 205:1285.

Lasne, C., B. Van Heerwaarden, C. M. Sgrò, and T. Connallon. 2019. Quantifying the relative contributions of the X chromosome, autosomes, and mitochondrial genome to local adaptation. Evolution 73:262–277.

Lees, A. D. 1955. The physiology of diapause in arthropods. Cambridge Univ. Press, Cambridge.

Lehmann, P., A. Margus, and L. Lindström. 2016. Inheritance patterns of photoperiodic diapause induction in *Leptinotarsa decemlineata*. Physiol. Entomol. 41:218–223.

Leimar, O., P. Hammerstein, and T. J. M. Van Dooren. 2006. A new perspective on developmental plasticity and the principles of adaptive morph determination. Am. Nat. 167:367–376.

Li, H. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

Lindestad, O., C. W. Wheat, S. Nylin, and K. Gotthard. 2019. Local adaptation of photoperiodic plasticity maintains life cycle variation within latitudes in a butterfly. Ecology 100:e02550.

Lumme, J., and L. Keränen. 1978. Photoperiodic diapause in *Drosophila lummei* Hackman is controlled by an X-chromosomal factor. Hereditas 89:261–262.

Mathias, D., L. Jacky, W. E. Bradshaw, and C. M. Holzapfel. 2007. Quantitative trait loci associated with photoperiodic response and stage of diapause in the pitcher-plant mosquito, *Wyeomyia smithii*. Genetics 176:391–402.

Mayor, C., M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin, K. A. Frazer, L. S. Pachter, and I. Dubchak. 2000. VISTA: visualizing global DNA sequence alignments of arbitrary length. Bioinformatics 16:1046–1047.

Meisel, R. P., and T. Connallon. 2013. The faster-X effect: integrating theory and data. Trends Genet. 29:537–544.

Merlin, C., R. J. Gegear, and S. M. Reppert. 2009. Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies. Science 325:1700–1704.

Nettle, D., and M. Bateson. 2015. Adaptive developmental plasticity: what is it, how can we recognize it and when can it evolve? Proc. R. Soc. B Biol. Sci. 282:20151005.

Paolucci, S., L. van de Zande, and L. W. Beukeboom. 2013. Adaptive latitudinal cline of photoperiodic diapause induction in the parasitoid *Nasonia vitripennis* in Europe. J. Evol. Biol. 26:705–718.

Paolucci, S., L. Salis, C. J. Vermeulen, L. W. Beukeboom, and L. van de Zande. 2016. QTL analysis of the photoperiodic response and clinal

distribution of period alleles in *Nasonia vitripennis*. Mol. Ecol. 25:4805–4817.

Pruisscher, P., H. Larsdotter-Mellström, C. Stefanescu, S. Nylin, C. W. Wheat, and K. Gotthard. 2017. Sex-linked inheritance of diapause induction in the butterfly *Pieris napi*. Physiol. Entomol. 84:257–265.

Pruisscher, P., S. Nylin, K. Gotthard, and C. W. Wheat. 2018. Genetic variation underlying local adaptation of diapause induction along a cline in a butterfly. Mol. Ecol. 27:3613–3626.

Rodriguez-Caro, L., J. Fenner, C. Benson, S. M. Van Belleghem, and B. A. Counterman. 2020. Genome assembly of the dogface butterfly *Zerene cesonia*. Genome Biol. Evol. 12:3580–3585.

Schmitt, T., J. C. Habel, M. Zimmermann, and P. Müller. 2006. Genetic differentiation of the marbled white butterfly, *Melanargia galathea*, accounts for glacial distribution patterns and postglacial range expansion in southeastern Europe. Mol. Ecol. 15:1889–1901.

Sedlazeck, F. J., P. Rescheneder, and A. Haeseler von. 2013. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29:2790–2791.

Sgrò, C. M., J. S. Terblanche, and A. A. Hoffmann. 2016. What can plasticity contribute to insect responses to climate change? Ann. Rev. Entomol. 61:433–451.

Söderlind, L., and S. Nylin. 2011. Genetics of diapause in the comma butterfly Polygonia c-album. Physiol. Entomol. 36:8–13.

Sperling, F. A. H. 1994. Sex-linked genes and species differences in lepidoptera. Can. Entomol. 126:807–818.

Stearns, S. C. 1989. The evolutionary significance of phenotypic plasticity. BioScience 39:436–445.

Tauber, M. J., C. A. Tauber, and S. Masaki. 1986. Seasonal adaptations of insects. Oxford Univ., New York.

Tauber, E., M. Zordan, F. Sandrelli, M. Pegoraro, N. Osterwalder, C. Breda, A. Daga, A. Selmin, K. Monger, C. Benna et al. 2007. Natural selection favors a newly derived timeless allele in *Drosophila melanogaster*. Science 316:1895–1898.

Vasudevan, S., N. G. Starostina, and E. T. Kipreos. 2007. The *Caenorhabditis elegans* cell-cycle regulator ZYG-11 defines a conserved family of CUL-2 complex components. EMBO Rep. 8:279–286.

Via, S., R. Gomulkiewicz, G. De Jong, S. M. Scheiner, C. D. Schlichting, and P. H. Van Tienderen. 1995. Adaptive phenotypic plasticity: consensus and controversy. Trends Ecol. Evol. 10:212–217.

Yamada, H., and M. T. Yamamoto. 2011. Association between circadian clock genes and diapause incidence in *Drosophila triauraria*. PLoS One 6:e27493.

Ziv, T., V. Chalifa-Caspi, N. Denekamp, I. Plaschkes, S. Kierszniowska, I. Blais, A. Admon, and E. Lubzens. 2017. Dormancy in embryos: insight from hydrated encysted embryos of an aquatic invertebrate. Mol. Cell. Proteom. 16:1746–1769.

Associate Editor: G. D. Hurst
Handling Editor: T. Chapman

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table S1. Number of reads mapping to the genome for each data set, and how many mapped as proper pairs in the right orientation.
Table S2. The circadian clock genes in *Danaus plexippus* and their orthologs in *Pieris napi*.
Table S3. Fixed SNPs and their corresponding codon changes.
Table S4. Gene set enrichment of the 100 outlier genes.
Table S5. Gene set enrichment for the 25 outlier genes.
Figure S1. Local alignment between *Zerene cesonia* and *Pieris napi*.