Genetic basis of allochronic differentiation in the fall armyworm

Sabine Hänniger1*, Pascaline Dumas2†, Gerhard Schöff3, Steffi Gebauer-Jung1, Heiko Vogel1, Melanie Unbehend1, David G. Heckel1 and Astrid T. Groot1,3

Abstract

Background: Very little is known on how changes in circadian rhythms evolve. The noctuid moth Spodoptera frugiperda (Lepidoptera: Noctuidae) consists of two strains that exhibit allochronic differentiation in their mating time, which acts as a premating isolation barrier between the strains. We investigated the genetic basis of the strain-specific timing differences to identify the molecular mechanisms of differentiation in circadian rhythms.

Results: Through QTL analyses we identified one major Quantitative trait chromosome (QTC) underlying differentiation in circadian timing of mating activity. Using RADtags, we identified this QTC to be homologous to Bombyx mori C27, on which the clock gene vrille is located, which thus became the major candidate gene. In S. frugiperda, vrille showed strain-specific polymorphisms. Also, vrille expression differed significantly between the strains, with the rice-strain showing higher expression levels than the corn-strain. In addition, RT-qPCR experiments with the other main clock genes showed that pdp1, antagonist of vrille in the modulatory feedback loop of the circadian clock, showed higher expression levels in the rice-strain than in the corn-strain.

Conclusions: Together, our results indicate that the allochronic differentiation in the two strains of S. frugiperda is associated with differential transcription of vrille or a cis-acting gene close to vrille, which contributes to the evolution of prezygotic isolation in S. frugiperda.

Background

Virtually all life on earth experiences a similar day-night cycle, yet some species have evolved to be day-active, while others are night-active. Even closely related species living in the same habitat can differ from each other in their daily activity rhythms [1]. The molecular mechanisms of circadian rhythms have been revealed in a broad range of eukaryotic species from algae [2] to mammals (reviewed in SM Reppert and DR Weaver [3]), and also in insects with the increasing number of sequenced insect genomes [4–7]. Surprisingly, despite this growing knowledge on the molecular basis of circadian rhythms, the evolution of differentiation in daily activity patterns is largely unexplored. In insects, differences in diurnal mating times have been found to prevent gene flow between populations [8–14]. By determining the genetic basis of allochronic differentiation between closely related species, or even between divergent populations within species, the initial steps causing differentiation in daily activity rhythms can be discovered, which is important for an understanding of the evolution of circadian rhythms on a micro-evolutionary time-scale.

An ideal model organism for the study of the evolution of circadian rhythms is the noctuid moth Spodoptera frugiperda (Lepidoptera: Noctuidae), as it consists of two naturally occurring morphologically identical strains that exhibit strain-specific timing of mating in the night [15, 16]. The so-called corn- and rice-strains seem to be in the process of ecological speciation in sympatry [17]. Although the hybridization rate is up to 16% in the field [18], the two strains do not merge into one panmictic population, which is probably prevented by a combination of different isolation barriers [17]. So far, three possible prezygotic mating barriers have been described in this species: a) differential host plant choice [19–23], b) strain-specific timing of mating in the night [15, 16], and c) female sex pheromone
differences [24–26]. Recent studies have shown that host preference in the field is not as specific as previously thought [27–29]. Therefore, habitat isolation seems to be a relatively weak prezygotic mating barrier. Differences in female sex pheromone composition are also likely to constitute a weak prezygotic mating barrier [26, 30]. As both strains consistently differ in their timing of reproductive activity at night [15, 16], allochronic divergence seems to be a major barrier separating the two S. frugiperda strains. The corn-strain calls, mates and oviposits approximately three hours earlier than the rice-strain, with only a small overlapping time-window between the strains [15, 16].

Allochronic speciation due to seasonal timing differences has been suggested for several insect species, e.g. crickets [31, 32], fruit flies [33, 34] and mosquitoes [13]. However, surprisingly little research has been conducted on the importance and exact genetic changes underlying temporal speciation (reviewed in AT Groot [1]). Recently, a study by Kaiser et al. [35] determined the genomic basis of circadian and circalunar timing adaptations in the midges Clunio marinus. Different naturally occurring strains of C. marinus emerge at different time points in the circadian as well as the circalunar rhythm and mate and oviposit shortly after. The abundance of different splice variants of the calcium/calmodulin-dependent kinase II.1 (CaMKII.1) is associated with these allochronic differences.

Also the two S. frugiperda strains differ in their diurnal mating patterns and we hypothesize that genetic and/or expression differences in one or more clock genes underlie their differences in timing of reproductive activity.

In general, biological clocks are a network of genes and gene products that enhance and suppress each other in a rhythmic manner, entrained by environmental factors such as light, temperature or tides [36, 37]. Within insects, the clock gene network is best described in the vinegar fly Drosophila melanogaster, where the network consists of two interlocked feedback loops [36, 38]; one involving the genes vri, PDP-1 (pdp1), clock (clk) and cycle (cyc); and the other incorporating period (per), timeless (tim), clk and cyc. In addition, kinases phosphorylate clock proteins (e.g. phosphorylation of PER by DOUBLETIME (DBT) and casein kinase IIα (CK2α)) and facilitate their accumulation [36], while cryptochrome 1 (cry1) functions as a circadian photoreceptor. Most of these genes are also present in Lepidoptera [5, 39–42], and are thus good candidate genes that may underlie the timing differences between the corn- and the rice-strain (see Fig. 1). Additionally, a second cryptochrome, cryptochrome 2 (cry2), is present in Lepidoptera and able to repress CLK:CYC mediated transcription [40, 41, Fig. 1]. Even though most of the clock genes and their organization in the circadian clock are well conserved across taxonomic groups [43, 44], differences among insects orders have been highlighted [41, 43, 45], especially in the daily oscillation patterns of gene expression [46–49].

![Fig. 1 Two feedback loops that define the circadian rhythm in Danaus plexippus. Adapted from S Zhan et al. [5]. In the first feedback loop (green proteins), the CLK:CYC dimer promotes the transcription of Per, Tim and Cry2. PER, TIM and CRY2 proteins enter the nucleus where a PER:Cry2 dimer inhibits the binding of the CLK:CYC dimer to E-Boxes and thus inhibits the expression of genes with E-Box promoters. In the second, modulatory feedback loop (yellow proteins), the CLK:CYC dimer promotes the transcription of Vri and/or Pdp1. VRI and PDP1 proteins enter the nucleus, where PDP1 promotes CLK transcription while VRI inhibits CLK transcription.](image-url)
the exact genetic changes and the mechanism(s) underlying allochronic differentiation in behavior will shed light on how temporal differentiation may evolve.

In this study, we determined the genetic basis of the major prezygotic isolation barrier, i.e. differentiation in the diurnal mating patterns of the two strains of *S. frugiperda*. To identify the genomic loci (chromosomes) that explain most of the variance in timing of mating activity in an unbiased way, we first conducted quantitative trait locus (QTL) analyses using AFLP markers and RADtags. To determine which of the candidate clock genes map to the major Quantitative trait chromosome (QTC), we indirectly mapped these genes onto our genetic map by homologizing the *S. frugiperda* linkage map to the *B. mori* chromosomes and identifying the location of the candidate genes on the homologous chromosomes. This approach is possible due to the high syntheny between *S. frugiperda* and *B. mori* [50]. Since *vrille* was located on the major QTC, we determined differences in expression levels, as well as sequence differences of this gene, between the two strains, and compared this to differences in the other main clock genes.

**Methods**

As *S. frugiperda* is a non-model organism, no assembled genome was available when we started investigating the genetic basis of the differences in mating time of the two strains. Hence, this manuscript presents results that were obtained with a variety of methods to overcome this lack of information. In the course of experiments, the genome of both the corn- and the rice-strain became available to the members of The Fall armyworm International Public Consortium, which facilitated genomic comparisons and expression analysis of all main clock genes. An overview of the chronology of the experiments is depicted in Additional file 1.

**Insects**

Individuals used for the QTL analysis descended from >200 rice-strain larvae and >100 corn-strain larvae, collected from different fields in Florida in 2003 and 2004, respectively (Additional file 2). These populations were reared for 10 (corn-strain) and 21 (rice-strain) generations in mass culture at the USDA-ARS in Gainesville, FL, before shipment to the Max Planck Institute for Chemical Ecology (MPICE) in 2007. These populations were also used by G Schöfl, DG Heckel and AT Groot [16]. We refer to these populations as CL1 and RL1 (Additional file 2). Unfortunately, these two populations died after six years of laboratory rearing. Therefore, we established new laboratory populations starting with ~300 larvae collected in Florida (rice-strain) and Puerto Rico (corn-strain) in 2010 (Additional file 2), which were shipped directly to MPICE, where all adults were screened for strain-specific COI markers [23], and separated accordingly into strain-specific colonies. We refer to these populations as CL2 and RL2 (Additional file 2).

All populations were reared in climate chambers with reversed light:dark (L:D) cycle (photophase started at 10 pm, scotophase started at noon) and 14:10 L:D photoperiod at 26 °C and 70% RH. Adults were fed with a 10% honey-water solution and random single-pair-matings were set up to maintain the populations and minimize inbreeding.

**Generation of backcrosses**

For the QTL analysis, we generated female-informative backcrosses (Additional file 3). Single pair matings between pure corn- and rice-strain individuals were performed to obtain F₁ hybrids (denoted CR from corn mothers and RC from rice mothers). Hybrid females were then backcrossed to pure rice-strain males to produce different backcross families (Additional file 3). Two backcross families (BCs) were used for the QTL analysis (BC_A: RCxR, BC_B: CRxR) (the first two letters of a backcross refer to the mother, the last letter to the father). The two rice-strain fathers used to generate both backcrosses were kin.

**Phenotyping backcross families**

To determine the phenotype for the QTL analysis, we observed the mating behavior of a) pure strain individuals in intra-strain (CxC, RxR) and inter-strain matings (CxR, RxC), b) hybrid females backcrossed to pure strain males (CRxC, CRxR, RCxC, RCxR), and c) female backcross offspring crossed to pure strain males (CR-RxC, CR-RxR, RC-RxC, RC-RxR). The observations of mating behavior were performed as described by [16] and summarized here. One to four day-old virgin females and males were set up in single pairs in clear plastic cups (16 oz.) and provided with 10% honey solution. All matings were set up simultaneously and placed in a walk-in climate chamber (26 °C, 70% RH, L:D 14:10) two hours before scotophase. In total, 320 to 400 couples were observed throughout the scotophase and one hour into photophase (in total 11 h), with a 30 min interval, i.e. each couple was observed once every 30 min. All pairs were observed for three consecutive nights starting at the first day of the mating. The onset time of the first mating was the phenotype used for the timing QTL analysis. After observation, all individuals were frozen at -80 °C for further genetic analysis.

**Genetic map construction**

DNA of 90 randomly chosen backcross females (44x RC-R, 46x CR-R) as well as of their parents and grandparents were used to generate AFLP markers. Female backcross individuals were chosen to construct the
markers. Seventeen markers in each BC family did not differentiate primer combinations up to a maximum of 26, some consisted of at least two AFLP markers from autosomes (Additional file 4). Each chromosomative AFLP markers were used to identify the 30 of 465 (in BC family A) and 514 (in BC family B) individuals with two female-informative backcross families. A total of QTL analysis present in both backcrosses (Additional file 4) were used to construct linkage maps are not necessarily homologous. Markers in each linkage map, so that the same numbers in the different linkage groups were utilized to identify RAD markers segregating in the same pattern with RAD tools [53]. The AFLP markers showed a specific presence (1) /absence (0) pattern in backcross individuals for each chromosome, thus RAD markers showing the same 1/0 pattern in the backcross individuals were identified as belonging to the same specific linkage group. All sequences matching an AFLP segregation pattern were pooled across the two linkage maps.

To identify candidate QTC, each chromosome was tested for a significant difference in the phenotype (i.e. onset time of first mating) between the homo- and heterozygote backcross females. The two backcrosses were also combined for this analysis, to increase the sample size and thus the possibility to detect QTC. Because of the absence of crossing-over in female Lepidoptera [52], each identified QTC corresponds to an individual chromosome, on average 1/31 of the genome. Statistical analysis was performed with R 2.5.0 (R-Development-Core-Team, 2007) and SAS® software (SAS institute, Cary, NC, USA, 2002-2008). To assess how much of the variance is explained by the different QTC (R² value) we conducted a two-sided t-test and a GLM (The GLM procedure of the SAS software, with onset time of the first copulation as dependent variable). Chromosomes with a significant correlation (P<0.05) were considered a QTC.

Homologizing linkage maps to Bombyx mori chromosomes, using RAD markers

To identify candidate genes on the QTC, the linkage map was homologized to the reference genome of B. mori, using restriction site associated DNA (RAD) analysis (see Baxter et al [53] and Groot et al. [54]). DNA of parents, female grandparents and 11 backcross individuals per backcross family was digested with the SbfI restriction enzyme (New England Biolabs, Ipswich, MA, USA), barcoded, pooled, sheared and amplified, following the procedure described in Groot et al [54]. The pool was paired-end sequenced (50 bp fw, 50 bp rev) by FASTERIS (Geneva, Switzerland) with a HiSeq Illumina sequencer, resulting in 76 million reads. The reads were separated by barcodes into pools per individual and filtered for quality (q10 = 99%). On average, there were 5-10 different paired-end reads per forward read (Additional file 5). The segregation patterns that were obtained with the AFLP markers for different linkage groups were utilized to identify RAD markers segregating in the same pattern with RAD tools [53]. The AFLP markers showed a specific presence (1) /absence (0) pattern in backcross individuals for each chromosome, thus RAD markers showing the same 1/0 pattern in the backcross individuals were identified as belonging to the same specific linkage group. All sequences matching an AFLP segregation pattern were pooled across the individuals, after which the paired-end sequences were retrieved, resulting in 30 FASTA files (one file per chromosome). Each group was assembled into RAD contigs using CLC Genomics Workbench (CLC bio
traction, cDNA synthesis and RT-qPCR reaction were conducted, as described in AT Groot et al. [54] and summarized here. Pools of five heads were ground with mortar and pestle in liquid nitrogen, RNA was isolated using Direct-zol™ RNA MiniPrep (Zymo Research corp.), and DNase was digested by adding 10 µl 10x Turbo DNase buffer and 1 µl Turbo DNase (Ambion, LIFE TECHNOLOGIES, Darmstadt, Germany). cDNA was synthesized from 1000 ng RNA using Verso cDNA synthesis kit (Thermo Fisher Scientific, Schwerte, Germany). RT-qPCR experiments were conducted with 5 ng cDNA per reaction, 2 technical replicates on each plate, using ABSolute Blue QPCR SYBR Green Low Rox Mix (Thermo Fisher Scientific, Schwerte, Germany) and Applied Biosystems 7500 Real-Time PCR (ThermoFisher Scientific) (see Additional file 8 for details). The Elongation Initiation Factor 1α (eIF1α) was used as the reference gene and amplified for all samples. eIF1α was chosen as the reference gene as it showed the most stable expression over the different time points in a pre-experiment. Relative expression levels were calculated as copy numbers per 1000 copies eIF1α.

To verify our first RT-qPCR results, we conducted a second qPCR experiment where we focused on the 12 most relevant time points in which the two strains showed differences in clock expression levels, i.e. from 5 h before (-5 h) until 6 h into scotophase (+6 h). For this second experiment, we again observed mating couples for one night, as described above, and extracted RNA extraction in the second night, for which we chose corn-strain females that showed reproductive behavior (female calling, copulation) early in the night and rice-strain females that showed reproductive behavior later in the night. Every hour, 6 - 10 females of both strains were transferred from the rearing cups to a 10 ml Falcon tube, immediately frozen in liquid nitrogen and kept at -80 °C. This time, RNA was isolated individually from six heads, providing six biological replicates per strain per time point. RNA extraction, cDNA synthesis and RT-qPCR reaction were conducted with 5 ng cDNA per reaction, this time 3 technical replicates on each plate, using 5X Hot firepol EvaGreen® qPCR mix plus (ROX) (Thermo Fisher Scientific, Schwerte, Germany) and a Bio-Rad CFX machine (Biotum) (see Additional file 8 for details). eIF1α was again used as the reference gene and amplified for all samples. Relative expression levels were calculated as above.

Statistical differences in strain-specific expression levels of the clock genes were tested in R (R Development Core Team 2010). For each time point, we tested
the normality of the data with a Shapiro-Wilk test (shapiro.test function under R). To ensure that the variances were equal in both rice and corn dataset, we ran a Fisher test (var.test function under R). Finally, we used a Student test to compare the expression levels of two samples for each gene (t.test function in R) with a Bonferroni correction for multiple comparisons.

We also tested for significant differences between the whole expression pattern of the corn-strain and the rice-strain for individual clock genes. For this, we first used a Shapiro-Wilk normality test for each dataset (dataset = expression data 1 gene for one strain for each timepoint) followed by an F test to compare the variances of both strains per clock gene. Finally, we used a Two Sample T-test to test for a significant difference between the corn-strain expression and the rice-strain expression for each clock gene.

All graphs were made in R with the ggplot2 package [57].

**Sequence differences of the main candidate gene, vrille**

To assess strain-specific sequence differences in vril, the sequence of the gene was established stepwise. First degenerate primers based on insect ESTs and genomic sequences (gb|AY526608.1, gb|AY576272.1, gb|AADK01019845.1) were used to obtain partial sequences. After obtaining the sequences, primers were designed to sequence further. The DNA Walking SpeedUp™ Kit II (SEEGENE, Eschborn, Germany) was used to obtain the sequence upstream of the coding sequence (see Additional file 9 for all primers used). To determine exon/intron structure, the coding region was sequenced from cDNA. Subsequently, parts of the gene were sequenced in 88 different samples (including backcross individuals and corn- and rice-strain individuals from different regions; Additional file 2), using Sanger-sequencing and Sequencher 4.10.1 for analysis. All obtained sequences are available in GenBank (accession numbers KM675483 - KM675658). Subsequently, the S. frugiperda genome for both strains (http://ww6.inra.fr/lepidodb/SfruDB) and an in-house RNAseq database of larval guts became available [55]. With the full length mRNA acquired from the RNAseq database and BLASTed against the genome, the full sequence of vril was obtained, including a large intron in the 5’UTR. The corn-strain genome was not complete in this region, thus two BAC clones (AUA00AAA25YL06FM1, AUA0AAA20YH15RM1) spanning the region were obtained from the Centre National de Ressources Génomiques Végétales (CNRGV, Toulouse, France) and shotgun sequenced using Sanger sequencing and Sequencher for analysis. Based on the alignment of the rice-strain genome from SfruDB and the BAC clone sequences, additional parts of the intron were sequenced in 12 corn-strain and 12 rice-strain individuals from the CL_1 and RL_1 populations, as well as the parental and F_1 generations of the backcross families.

When the genome of S. frugiperda became available, the other main clock genes were annotated in both strains as part of the WGS project [55]. The coding sequences of each gene of both strains were aligned to identify polymorphisms between the strains. The protein domains were identified on http://prosite.expasy.org/ and using Protein BLAST on http://blast.ncbi.nlm.nih.gov/.

**Results**

**QTL analysis**

QTCs were identified by testing each linkage group for a significant association with the phenotypic trait, i.e. onset time of first mating. Resulting P-values and R^2-values (see Additional file 6) refers to the combined analysis of both BCs to use the biggest sample size possible. One QTC (Sf_C25, P < 0.0001, R^2 = 0.19) had a major effect on the variance in the strain-specific timing of mating. This QTC is homologous to Bombyx mori chromosome 27 (Bm_C27) and explained 19% of the variance of the onset time of first mating (Fig. 2). Bm_C27 is 14.5 Mb in size (52.8 cm) and represents 3.3% of the total B. mori genome [56, 58]. The difference in the onset time of mating between heterozygous and homozygous individuals (carrying only rice-strain copies of the chromosome) for Sf_C27 did not differ between the combined or individual analysis of backcrosses (Additional file 10). The LOD scores of all linkage groups are shown in Additional File 11.

Even though the major QTC Sf_C25 explained 19% of the variance between the strains, it only achieved a power of 0.55, i.e. only 55% of QTCs of this magnitude could be detected with our setup on average (Additional file 12). This is due to the small sample size and implies a chance of missing or underestimating minor QTCs. However, the finding of the same major QTC in two independent backcrosses strengthens our finding. We did detect three minor QTCs in the combined analysis that were not consistently present in both families when analyzed individually: Sf_C28 (Bm_C2, P = 0.014, R^2 = 0.08), Sf_C30 (Bm_C6, P = 0.0104, R^2 = 0.08) and Sf_C20 (Bm_C12, P = 0.023, R^2 = 0.07).

**Mapping the candidate genes**

The main circadian rhythm genes are located on the following chromosomes (Additional file 6): per, clk, cyc and PdP1 on the sex chromosome (Bm_C01), jetlag on Bm_C3 (Sf_C10), tim on Bm_C4 (Sf_C31), CK2α on Bm_C5 (Sf_C09), cry2 on Bm_C15 (Sf_C23), CK2β and cry1 on Bm_C15 (Sf_C23), dbt on Bm_C17 (Sf_C17), shaggy on Bm_C18 (Sf_C13), clockwork orange on
Bm_C22 (Sf_C6), slimb on Bm_C24 (Sf_C12, 32), vri on Bm_C27 (Sf_C25) and CK1α on Bm_scaf256 (which has not been mapped to a B. mori chromosome, and thus cannot be homologized). Thus, of all candidate genes, only vri mapped to the one major QTC Bm_C27 (Sf_C25). Since the reciprocal F1 hybrids (CR and RC) did not differ in their onset time of mating (Additional file 13; [16]), the involvement of the sex chromosome in the timing differentiation between the two strains can be excluded, which thus excludes per, clk, cyc and PdP1, which are located on the sex chromosome.

Expression differences of the main candidate gene vri
The main candidate gene vriille showed a clear daily oscillation pattern in its expression in both strains and the overall expression of vriille differed between the strains (Fig. 3). We found significant differences in expression levels between the two strains in the first 24 h experiment (experiment 1, \( P = 0.00065 \)), as well as in the second 12 h experiment (experiment 2, \( P < 0.0001 \)) (Fig. 3). These differences were due to overall significantly higher expression levels of vri in the rice-strain compared to the corn-strain.

In comparison, the other clock genes showed rhythmic expression levels as well, except for clk and cyc which did not...
not show a clear daily oscillation pattern (Fig. 4a-c). Three clock genes, *per*, *tim* and *cry2* did not differ between the strains in either experiment (Fig. 4a1-6a3). While in the first 24 h experiment three additional clock genes did show significant differences between the two strains (*pdp1* $P = 0.00543$, *clk* $P = 0.00222$, and *cyc* $P = 0.00002$; Fig. 4b, c), only *pdp1* also showed significant differences in the second 12 h experiment ($P = 0.0038$). All differences between the two strains were due to significantly higher expression levels in the rice-strain than in the corn-strain (Figs. 3 and 4).

**Sequence differences of the candidate gene *vrille***

*Vrille* is a short gene without introns in the protein coding region, coding for a 367 aa protein, followed by a 1234 bp 3' UTR. The 5' UTR is divided into a 45 bp segment and a 375 bp segment by an intron that contains regulatory elements, namely 11 Ebox elements (Ebox A – Ebox K) with the core sequence CACGTG (Fig. 3). Near Eboxes E, F and I, five strain-specific polymorphisms were identified in the investigated corn-strain and rice-strain populations and the maternal grandmothers of BC_A and BC_B (Table 1).

The availability of a corn-strain and a rice-strain variant of the *S. frugiperda* genome assembly enabled us to compare the sequences of the main clock genes in these assemblies (see Additional file 14). Sequence differences of the main clock genes were assessed by identifying non-synonymous as well as synonymous SNPs in the protein coding regions. We mostly found synonymous (syn) SNPs between the corn-strain and the rice-strain genome.

![Fig. 4](image-url)
assembly, with the exception of clk, cyc and tim, which showed two, two and one non-synonymous (non-syn) SNPs, respectively (Table 2). All non-synonymous SNPs were located in non-conserved domains of these genes (Additional file 14).

Discussion

In this study, we aimed to identify the genetic basis of the main prezygotic isolation barrier between the two strains of *S. frugiperda*, i.e. allochronic differentiation. We found one consistent Quantitative trait chromosome (QTC) that significantly accounted for the difference in the onset time of mating in the two strains, *Sf_C25*, which is homologous to *Bm_C27*. Detecting only one major QTC is of note because the timing of behavior is a complex trait, depending on the complex network of the circadian clock and its interlocked feedback loops of transcription and translation (see Fig. 1). Thus, one might anticipate that the difference in mating time between the corn and rice strains would represent a polygenic trait affected by multiple loci of modest to small effect. Yet, differences in one or more genes on *Sf_C25* explain 19% of the variance between the strains, while three minor QTCs in the combined analysis were not consistently present in both families when analyzed individually. In addition, no known genes involved in the circadian rhythm are located on the *Bombyx* homologs of these minor QTCs. Possibly, allochronic differentiation between the two strains is affected by an interaction between different factors involved in the circadian rhythm regulation.

A limitation of our indirect mapping approach is the different number of autosomes in *B. mori* (27) and *S. frugiperda* (30). When homologizing *Helicoverpa armigera* and *B. mori*, Sahara et al. [63] found *Bm* chromosomes 11, 23 and 24 to be merged from two chromosomes each. For *Bm* chromosome 23, we identified two homologous chromosomes in *S. frugiperda*: *Sf_C12* and *Sf_C29*. Whether *Bm_C11* and *Bm_C24* are also represented by two chromosomes in *S. frugiperda* remains to be determined. The incomplete homology did not affect our result, because a)

---

**Table 1** Variation in the regulatory intron in the 5’UTR of *vri*

| Position | Close to | Type       | Population | Individual |
|----------|----------|------------|------------|------------|
|          |          |            | CL_1 Corn  | RL_1 Rice  | mgmA Rice | mgmB Corn |
| 4717     | EboxE    | SNP        | T          | A          | A          | T          |
| 4727     | EboxE    | SNP        | T          | A          | A          | T          |
| 4816     | EboxE    | IN/DEL     | –          | –          | TTCGAA     | –          |
| 4870     | EboxF    | SNP        | A          | C          | C          | A          |
| 6690     | EboxI    | SNP        | T          | A          | A          | n.a.       |

Single nucleotide polymorphisms (SNPs) and insertions/deletion (IN/DEL) between 12 individuals from a corn-strain population (CL_1) and 12 individuals from a rice-strain population (RL_1) as well as in the maternal grandmothers (mgm) of BC_A and B (originating from these populations). Mgm = maternal grandmother; Sample name followed by Corn (= corn-strain) or Rice (= rice-strain); n.a. not available due to sequencing restrictions.

**Table 2** Number of SNPs between the corn-strain and rice-strain variant of the *S. frugiperda* genome assembly

| Gene name | Amino acids | Syn SNPs total | Syn SNPs per aa | Non-syn SNPs total | Non-syn SNPs per aa |
|-----------|-------------|----------------|-----------------|--------------------|---------------------|
| *pdp1*    | 263         | 3              | 0.0114          | 0                  | 0.0000              |
| *dbt*     | 346         | 1              | 0.0029          | 0                  | 0.0000              |
| *vri*     | 367         | 17             | 0.0463          | 0                  | 0.0000              |
| *cry1*    | 528         | 50             | 0.0947          | 0                  | 0.0000              |
| *cry2*    | 793         | 23             | 0.0290          | 0                  | 0.0000              |
| *per*     | 1222        | 34             | 0.0278          | 0                  | 0.0000              |
| *clk*     | 614         | 29             | 0.0472          | 2                  | 0.0033              |
| *cyc*     | 681         | 55             | 0.0808          | 2                  | 0.0029              |
| *tim*     | 1279        | 28             | 0.0219          | 1                  | 0.0008              |
we confirmed the position of vrille on our major Sf QTC, b) all minor QTC have a confident homologue in B. mori (see Fig. 2, Additional file 6), none of which contain known clock genes, and c) we homologized all chromosomes with known clock genes. The high synteny level between B. mori and S. frugiperda [50] also supports our conclusion that vrille is the only clock gene located on a QTC in S. frugiperda.

**Candidate gene vrille**

Within the network of the circadian clock genes in insects, vrille is a powerful player (e.g. in fire ants [64], pea aphids [65] and bean bugs [66]) and best described in Drosophila [36, 38, 67, 68]. VRI inhibits clk transcription, and since a dimer of CLK and CYC promotes many E-Box promoted genes, clk inhibition represses transcription of the core clock genes. Consequently, vrille mutants have altered behavioral rhythms [68]. Hence, in S. frugiperda a strain-specific difference in vrille structure or expression may cause a strain-specific expression difference in other clock genes, leading to a timing difference in behavior.

Our qPCR results show that vrille is consistently higher expressed in the rice-strain compared to the corn-strain (Fig. 3). Since VRI inhibits the transcription of clk, which is needed for the expression of most circadian rhythm genes, differences in vrille expression may impact the expression of the downstream players of the interlocked feedback loops. Interestingly, we only found strain-specific expression differences in the genes that constitute the second, modulatory feedback loop: vrille, pdp1, clk and cyc with the rice-strain showing an overall higher expression of these genes in at least one of the experiments. Pdp1, clk and cyc are all located on the sex chromosome, which is not involved in the timing differentiation between the strains (see Additional file 13). Thus, the observed expression pattern differences reflect downstream temporal regulation dependent upon the initial upstream genetic difference in the expression of vrille. The fact that only one feedback loop shows expression differences between the strains is surprising, as the CLK-CYC dimer connects and promotes the transcription of the downstream genes in both feedback loops (Fig. 1). Yet, mRNA expression differences do not directly translate to differences at the protein level. For the generally low turnover of the CLK and CYC proteins, the total effect of mRNA expression differences on protein abundance can be subtle. Additionally, there is always competition for the CLK-CYC dimers between the E-Box promoter genes, and basic helix-loop-helix transcription factors like CLK and CYC exhibit differences in their binding specificities to E-Box binding sites [69]. Since per and tim showed higher expression peaks compared to pdp1 and vrille, it is possible that per- and tim E-Boxes attract CLK-CYC dimers more easily and are thus less affected by a limited CLK-CYC dimer abundance.

Differences in expression levels of clock genes have been shown to be linked to different activity patterns. For example, the migratory songbird Emberiza melanocephala shows significant differences in clock gene expression levels, rather than phase shifts, between life history states that differ in their daily activity patterns (e.g. night activity in the migratory life state and day activity in the pre-migratory state) [70]. In insects, the cricket Gryllus bimaculatus shows lower expression of tim and per in the nocturnal nymphs compared to the diurnal adults [71]. Whether and how the differences in transcription levels of clock genes in S. frugiperda strains may affect behavioral variation in timing of sexual activities remains to be determined.

In our search for sequence differences in vrille that might account for the timing difference, we found five strain-specific polymorphisms surrounding E-boxes between the corn-strain and a rice-strain population from Florida and in the parents of the backcross families used for the QTL analysis (Fig. 5). Since the binding specificity of basic helix-loop-helix transcription factors, such as CLK and CYC, is influenced by the genomic region surrounding the E-box binding site [69], a less efficient binding of a transcription factor to the active vrille E-Box element(s) in e.g. the corn-strain could change the expression of vrille. Alternatively, a cis-regulatory element regulating this gene could be situated on the same chromosome at a more distant region that we did not yet sequence. Mutations in cis-regulatory elements generally cause expression differences [72, 73] and are hypothesized to be key elements of evolutionary changes [74].

When aligning the protein coding regions of the other annotated clock genes in the corn-strain and rice-strain variant of the S. frugiperda genome, we identified a number of synonymous SNPs in every clock gene, and five non-synonymous SNP in per, cyc and tim. However, none of these non-synonymous SNPs affected conserved protein domains of these genes. Also, these SNPs are based solely on the genome alignments and were not confirmed by testing multiple corn- and rice-strain individuals and may thus not be strain-specific. Tim is located on Sf_C31 (Bm_C04), which is not a QTC for the timing differences. Also, tim did not show any strain-specific expression differences (Fig. 4a). Clk and cyc are situated on the sex-chromosome, which does not underlie the strain-specific timing differences, based on the observation of reciprocal backcrosses. The expression differences of these genes can be explained as a downstream-effect of differences in the expression of vrille and pdp1.
Comparisons of clock gene expression patterns between species

To compare the expression patterns of the main clock genes in the night-active Lepidoptera *S. frugiperda* to expression data in other species is difficult, as expression studies addressing the clock genes are often conducted under Zeitgeber time (ZT) conditions (L:D 12:12), while we used a longer day (L:D 14:10). With this difference in mind, some expression levels can be compared.

In the day-active migratory Lepidoptera *Danaus plexippus*, and in the day-active *Drosophila melanogaster*, vri expression is highest early in the scotophase [38, 68]. In *S. frugiperda*, vri peaks earlier, namely 4 h before the scotophase in the rice-strain and 3–1 h before the scotophase in the corn-strain (24 h experiment). In *S. frugiperda*, pdp1 shows an expression peak 8 h after vri in the rice-strain and 4-7 h after vri in the corn-strain. In *Drosophila* expression levels of pdp1 peaked 6 h after vri [38] and thus more comparable to the rice-strain. These different times of peak expression levels are expected, as VRI and PDP1 proteins both bind to the promoter region of *clock* at different time points to inhibit or facilitate clock expression, respectively.

As for other clock genes, in both *D. plexippus* and *D. melanogaster*, tim expression peaks 2 h into scotophase (ZT 14) [75], while we found the highest expression levels of tim at 4 h into scotophase. In addition, in *D. plexippus*, per expression peaks 2 h into scotophase (ZT 14) [75], while we found the highest expression levels of per at 6-7 h (corn-strain) to 10 h into scotophase (rice-strain; 24 h experiment). Since TIM, PER and CRY2 proteins form a trimer to enter the nucleus, similar expression patterns are expected in these genes, so that the peak differences between per and tim in *S. frugiperda* is surprising.

Expression differences in clock genes are especially interesting in the context of diurnal and nocturnal activity patterns. Martin-Fairey et al. [76] investigated PER protein expression levels in diurnal grass rats *Arvicanthis niloticus*. When comparing individuals that exhibited the usual day-active behavior to individuals that had adopted a night-active behavior, they found PER protein expression levels comparable to nocturnal and diurnal rodents [76]: The expression of PER protein peaks in the early morning in diurnal grass rats [76, 77] while it peaks in the late night in nocturnal rodent species [78, 79]. To be able to make such comparisons in Lepidoptera, clock gene expression studies across species should be conducted under comparable conditions.

**Conclusion**

In summary, we identified one major QTC for the timing difference in mating between the two *S. frugiperda* strains. The clock gene vri (vri) is located on this QTC and thus the major candidate for the strain-specific timing differences. Strain-specific expression differences, as well as strain-specific polymorphisms in the regulatory region of vri, support the hypothesis that vri plays a key role in the timing differentiation of these two strains. As allochronic, diurnal differentiation is likely the major isolation barrier driving divergence between *S. frugiperda* populations, the mechanism by which vri and the other circadian clock genes influence this differentiation should be elucidated, which is possible through fine-scale mapping and functional analyses. This will advance our understanding of the molecular basis of incipient speciation in sympathy.
## Additional files

| Additional file | Description |
|----------------|-------------|
| Additional file 1 | Chronological outline of experiments. (PDF 86 kb) |
| Additional file 2 | Spodoptera frugiperda populations. (PDF 68 kb) |
| Additional file 3 | Generation of female-informative backcross families for QTL analysis. (PDF 61 kb) |
| Additional file 4 | Numbers of scored AFLP markers. (PDF 56 kb) |
| Additional file 5 | Coverage of RAD sequences. (PDF 224 kb) |
| Additional file 6 | Homologous chromosomes of *S. frugiperda* and *B. mori*. (PDF 20 kb) |
| Additional file 7 | Details of BLAST hits of SfBC contigs to scaffolds of Sf genome and of BLAST hits of Sf genome scaffolds to Bm genome. (PDF 57 kb) |
| Additional file 8 | PCR conditions used in the mentioned experiments. (PDF 47 kb) |
| Additional file 9 | Primer combinations and annealing temperatures (Tm). (PDF 71 kb) |
| Additional file 10 | Mean onset time of mating for individuals that are homozygous or heterozygous for the QTC SF_C25. (PDF 107 kb) |
| Additional file 11 | LCO scores for all linkage groups. (PDF 36 kb) |
| Additional file 12 | Power analysis for backcross families. (PDF 317 kb) |
| Additional file 13 | Mating time in *S. frugiperda* hybrids. (PDF 29 kb) |
| Additional file 14 | Position of SNPs between the corn-strain and rice-strain of *S. frugiperda* in the coding regions of the major clock genes. (PDF 76 kb) |

## Acknowledgements

We thank Domenica Schnablauch, Susanne Donnerhacke and Antje Schmalz for their support in molecular analyses; Steffen Reifarth and Johannes Fleischmann for their support with insect observations, Simon Baxter for donating RAD P1 adapters and for help with RAD sequencing, Rob Meagher, Carlos Blanco and Laura Juarez for insects and DNA samples.

## Funding

This research was funded by the Deutsche Forschungsgemeinschaft (P.S.GR362721), the National Science Foundation (awards IOS-1052238 and IOS-1456973), and the Max-Planck-Gesellschaft.

## Availability of data and material

The Illumina RAD sequences are available from the Edmond repository under http://dx.doi.org/10.17617/3k [80]. Sequence data is accessible via GenBank (KM675483 - KM675658).

## Authors' contributions

ATG, GS, SH, PD and DGH designed the experiments. SH, PD, ATG, GS conducted the experiments. SGL, HV, DGH and SH conducted the bioinformatics analysis of the NGS data. PD, SH, DNG and ATG analyzed the data. SH, ATG, MU and PD drafted the manuscript. All authors read and approved the final manuscript.

## Competing interests

The authors have no financial or non-financial competing interests.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Author details

1 Max Planck Institute for Chemical Ecology, Entomology, Hans-Knöll-Str. 8, 07745 Jena, Germany. 2 Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands. 3 DKMS Life Science Labs, Fiedlerstr, 34, 01307 Dresden, Germany.
24. Groot AT, Marr M, Schöff G, Lorenz S, Svatos A, Heckel DG. Host strain specific sex pheromone variation in Spodoptera frugiperda. Front Zool. 2008;5:21.
25. Lima ER, McNeil JN. Female sex pheromones in the host races and hybrids of the fall armyworm, Spodoptera frugiperda (Lepidoptera: Noctuidae). Chemoecology. 2009;19(1):29–36.
26. Unbehend M, Hänniger S, Meagher RL, Heckel DG, Groot AT. Pheromonal divergence between two strains of Spodoptera frugiperda. J Chem Ecol. 2013;39(3):364–76.
27. Juárez ML, Murua MG, Garcia MG, Ontivero M, Vera MT, Vilardi JC, Groot AT, Castagnaro AP, Gastamiriz G, Willink E. Host association of Spodoptera frugiperda (Lepidoptera: Noctuidae) corn and rice strains in Argentina, Brazil, and Paraguay. J Econ Entomol. 2012;105(2):573–82.
28. Juárez ML, Schöff G, Vera M, Vilardi J, Murua M, Willink E, Hänniger S, Heckel DG, Groot AT. Population structure of Spodoptera frugiperda maize and rice host forms in South America: are they host strains? Entomol Exp Appl 2014.
29. Groot AT, Unbehend M, Hänniger S, Juárez ML, Kost S, Heckel DG, et al. Evolution of reproductive isolation of Spodoptera frugiperda. In: Pheromone Communication in Moths: Evolution, Behavior and Application. Oakland, CA: UC Press; 2006:129.
30. Unbehend M, Hänniger S, Vásquez GM, Juárez ML, Reissig D, McNeil JN, Meagher RL, Jenkins DA, Heckel DG, Groot AT. 2013. Geographic variation in sexual attraction of S. podoptera frugiperda corn- and rice-strain males to pheromone lures. PLoS One. 9:89255.
31. Danley PD, Decarvalho TN, Fergus DJ, Shaw KL. Reproductive asynchrony and the divergence of Hawaiian cricket. Ethology. 2007;113(2):1125–32.
32. Fergus DJ, Shaw KL. Circadian rhythms and period expression in the Hawaiian cricket genus Laupala. Behav Genet. 2014;43(3):241–53.
33. Prabhakaran PM, Sheeba V. Sy mpatric Drosophilid species interactions and circadian clock genes in insects. Insect Mol Biol. 2008;17(5):447.
34. Kaiser TS, Poehn B, Szkiba D, Preussner M, Sedlazeck FJ, Zrim A, Neumann T, Hänniger S. Linking proximate and ultimate control of the circadian clock in monarch butterfly migration. Science. 2003;300(5623):1303–8.
35. Kaiser TS, Poehn B, Szkiba D, Preussner M, Sedlazeck FJ, Zrim A, Neumann T, Hänniger S. Linking proximate and ultimate control of the circadian clock in monarch butterfly migration. Science. 2003;300(5623):1303–8.
36. Groot AT, Estock ML, Horovitz JL, Hamilton J, Santangelo RG, Schal C, Gould F. QTL analysis of sex pheromone blend differences between two closely related moths: Insights into divergence in biosynthetic pathways. Insect Biochem Mol Biol. 2009;39(8):568–77.
37. Heikel DG. Comparative genetic linkage mapping in insects. Annu Rev Entomol. 1993;38:381–408.
38. Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. PLoS One. 2011;6(4):11.
39. Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Mol Biol Evol. 2007;24(4):948–58.
40. Groot AT, Staudacher H, Barthel A, Inglis O, Schöff G, Santangelo RG, Gebauer-Jung S, Vogel H, Emerson J, Schal C, et al. One quantitative trait locus for intra- and interspecific variation in a sex pheromone. Mol Ecol. 2013;22(4):1065–80.
41. Guoin A, Bretaudeau A, Nam K, Gimenez S, Avery J-M, Duvic B, Hilliou F, Durand N, Montagné N, et al. Extreme gene family evolution underlies adaptation for polyophagous Spodoptera frugiperda. 2017; submitted.
42. Shiomura M, Minami H, Suetusu Y, Ohyanagi H, Satoh C, Antonio B, Nagamura Y, Kodono-Okauda K, Kajiwara H, Sezutsu H, et al. KAIKObase: An integrated silkworm genome database and data mining tool. BMC Genom. 2009;10:18.
43. Vidal-Khakham H. ggplot2: Elegant Graphics for Data Analysis. New York: Springer; 2009.
44. Xie QY, Wang J, Zhou ZY, Li RQ, Fan W, Cheng DJ, Cheng TC, Qin JJ, Duan J, Xu HF, et al. The genome of a lepidopteran model insect, the silkworm Bombyx mori. Insect Biochem Mol Biol. 2008;38(12):1036–45.
45. Gleason JM, Ritchie MG. Do quantitative trait loci for a courtship song difference between Drosophila simulans and D. sechelia coincide with candidate genes and interspecific QTL? Genetics. 2004;166(3):933–11.
46. Moehring AJ, Mackay TFC. The quantitative genetic basis of male mating behavior in Drosophila melanogaster. Genetics. 2004;167(3):1249–63.
47. Shaw KL, Parsons YM, Lesnitz SC. QTL analysis of a rapidly evolving speciation phenotype in the Hawaiian cricket Laupala. Mol Ecol. 2007;16(14): 2879–92.
48. Gleason JM, James RA, Wicker-Thomas C, Ritchie MG. Identification of quantitative trait loci function through analysis of multiple cuticular hydrocarbons differing between Drosophila simulans and Drosophila sechellia females. Heredity. 2009;103(5):416–42.
49. Sahara K, Yoshido A, Shibata F, Fujikawa-Kojima N, Okabe T, Tanaka-Okuyama Y, Yasukochi Y. FH51 identification of Helicoverpa armigera and Mamestra brassicae chromosomes by BAC and fosmid probes. Insect Biochem Mol Biol. 2009;39(8):568–74.
50. d’Alencon E, Sezutsu H, Legeai F, Permal E, Bernard-Samain S, Gimenez S, Aury JM, Duvic B, Hilliou F, et al. Identification and characterization of the butterfly. Curr Biol. 2005;15(17):R714–7.
51. Kaiser TS, Neumann D, Heckel DG. Timing the tides: Genetic control of diurnal and lunar emergence times is correlated in the marine midden Cunio marinus. BMC Genet. 2011;12:49.
52. Cyran SA, Buchsbaum AM, Reddy KL, Lin MC, Glossop NR, Hardin PE, Young MW, Storti RV, Blau J. vrille, and DCl:clock form a second feedback loop in the Drosophila circadian clock. Cell. 2003;112(3):329–41.
53. Zhu H, Gegear RJ, Casselman A, Kanginakudru S, Reppert SM. Defining behavioral and molecular differences between summer and migratory monarch butterflies. BMC Biol. 2009;7:14.
54. Zhu H, Yuan Q, Briscoe AD, Foq O, Briscoe AD, Foq O, Casselman A, Reppert SM. The two CRYs of the butter fly. Curr Biol. 2005;15(23):R953–954.
55. Yuan Q, Metterville D, Briscoe AD, Reppert SM. Insect cryptochromes: gene duplication and loss define diverse ways to construct insect circadian clocks. Mol Biol Evol. 2007;24(4):948–55.
56. Trang LTD, Sehadeva H, Ichihara N, Iwai S, Mita K, Takeda M. Casein kinases I of the silkworm, Bombyx mori. Their possible roles in circadian timing and developmental determination. J Biol Rhythms. 2006;21(5):335–49.
57. Sandrelli F, Costa R, Kyriacou CP, Rosato E. Comparative analysis of circadian clock genes in insects. Insect Mol Biol. 2008;17(5):447–63.
58. Tohmoka K, Matsumoto A. Circadian molecular clocks in non-model insects. Curr Opin Insect Sci. 2015;5:78–84.
59. Goto SG, Denlinger DL. Short- and long-day expression patterns of genes involved in the flight cycle: synchronisation mechanisms, efficiency of cycle and phase control. J Insect Physiol. 2002;48(8):803–16.
60. Reppert SM, Tsal T, Roca A, Sauman I. Cloning of a Structural and Functional Homolog of the Circadian Clock Gene Product from the Giant Silkmoth Antheraea-perynii. Poultry. 1994;13(5):1167–76.
61. Froy O, Gotter AL, Casselman AL, Reppert SM. Illuminating the circadian clock in monarch butterfly migration. Science. 2003;300(5623):1303–5.
62. Iwai S, Fukui Y, Fujwara Y, Takeda M. Structure and expressions of two circadian clock genes, period and timeless in the commercial silkworm, Bombyx mori. J Insect Physiol. 2006;52(63):625–37.
63. Zhu H, Sauman I, Yuan Q, Casselman A, Emery-Le M, Emery P, Reppert SM. Cryptochromes define a novel circadian clock mechanism in monarch butterflies that may underlie sun compass navigation. PLoS Biol. 2008;6(1):e4.
73. Wittkopp PJ, Haerum BK, Clark AG. Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet. 2008;40(3):346–50.

74. Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8(3):206–16.

75. Merlin C, Gegear RJ, Reppert SM. Antennal circadian clocks coordinate sun compass orientation in migratory monarch butterflies. Science. 2009;325(5948):1700–4.

76. Martin-Fairey CA, Ramanathan C, Stowie A, Walaszczuk E, Smale L, Nunez AA. Plastic oscillators and fixed rhythms: Changes in the phase of clock-gene rhythms in the PVN are not reflected in the phase of the melatonin rhythm of grass rats. Neuroscience. 2015;288:178–86.

77. Ramanathan C, Stowie A, Smale L, Nunez AA. Phase preference for the display of activity is associated with the phase of extra-suprachiasmatic nucleus oscillators within and between species. Neuroscience. 2010;170(3):758–72.

78. Angeles-Castellanos M, Mendoza J, Escobar C. Restricted feeding schedules phase shift daily rhythms of c-Fos and protein Per1 immunoreactivity in corticolimbic regions in rats. Neuroscience. 2007;144(1):344–55.

79. Feillet CA, Mendoza J, Albrecht U, Pevet P, Challet E. Forebrain oscillators ticking with different clock hands. Mol Cell Neurosci. 2008;37(2):209–21.

80. Hänniger S, Dumas P, Schöff G, Gebauer-Jung S, Vogel H, Unbehend M, Heckel DG, Groot A. Data from: Genetic basis of allochronic differentiation in the fall armyworm. http://dx.doi.org/10.17617/3.k