The Clock Gene period of the Housefly, Musca domestica, Rescues Behavioral Rhythmicity in Drosophila melanogaster: Evidence for Intermolecular Coevolution?

Alberto Piccin,*1 Martin Couchman,*2 Jonathan D. Clayton,* David Chalmers,*3 Rodolfo Costa1 and Charalambos P. Kyriacou*

*Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom and 1Dipartimento di Biologia, Università di Padova, 35131 Padova, Italy

Manuscript received April 25, 1999
Accepted for publication October 22, 1999

ABSTRACT

In Drosophila, the clock gene period (per), is an integral component of the circadian clock and acts via a negative autoregulatory feedback loop. Comparative analyses of per genes in insects and mammals have revealed that they may function in similar ways. However, in the giant silkworm, Antheraea pernyi, per expression and that of the partner gene, tim, is not consistent with the negative feedback role. As an initial step in developing an alternative dipteran model to Drosophila, we have identified the per orthologue in the housefly, Musca domestica. The Musca per sequence highlights a pattern of conservation and divergence similar to other insect per genes. The PAS dimerization domain shows an unexpected phylogenetic relationship in comparison with the corresponding region of other Drosophila species, and this appears to correlate with a functional assay of the Musca per transgene in Drosophila melanogaster per-mutant hosts. A simple hypothesis based on the coevolution of the PERIOD and TIMELESS proteins with respect to the PER PAS domain can explain the behavioral data gathered from transformants.

Corresponding author: C. P. Kyriacou, Department of Genetics, University of Leicester, University Rd., Leicester LE1 7RH, United Kingdom. E-mail: cpk@leicester.ac.uk

Present address: Dipartimento di Biologia, Università di Padova, 35131 Padova, Italy.

Present address: Department of Biology, Imperial College, Silwood Park, Ascot, Berks SL5 7PY, England.

Present address: ETS/ FC 1Bd A. Fleming, 25020 Besançon, France.
tive clock model, and here we describe the initial cloning of the per homologue from the housefly, the unusual phylogenetic position of the PAS dimerization domain, and how the behavioral study of various interspecific per transfectants, including those carrying the Musca per ortholog, indicates a possible case of intermolecular coevolution between PER and one of its partners, the clock protein TIM.

MATERIALS AND METHODS

Fly strains: D. melanogaster and M. domestica strains were reared in a light-dark (LD) 12:12 cycle at 25°C. Drosophila adults and larvae were fed on sugar medium (6.5% sucrose, 11.5% baker’s yeast, 1% agar, 0.2% nipagin). Adult houseflies’ diet consisted of sucrose and dry milk. Eggs were laid in larval medium (prepared with 50 g bran, 0.1 g dried yeast, 80 ml milk, 30 ml H2O, and 1.5 ml of 20% nipagin), where they developed into adults.

Isolation and sequencing of the housefly per gene: The per homologue from Musca was cloned using a PCR-DOP strategy based on the sequences from Colot et al. (1988). The housefly probe was amplified from Musca genomic DNA using the degenerate primers 5′-CCGAAATCTAGGARCNTRYNA TGGAYGGA-3′ and 5′-CCGAAATTCAATGCCGATTATC-3′ (binding at position corresponding to amino acids 560-566 and 598-603 of the D. melanogaster protein sequence depicted in Figure 2, respectively), both carrying EcoRI-cleavable extensions. The amplified 132-bp fragment, which encodes the per1 region, was then used to screen an EMBL3 Musca genomic library. A 16-kb positive clone was isolated and subcloned into pUC19. Various different subclones were subsequently obtained and sequenced. The resulting 16-kb genomic clone was characterized, partially sequenced, and compared with cDNA sequence obtained from reverse transcriptase (RT)-PCR and 5′-RACE fragments performed on RNA isolated from Musca heads.

Construction of the M. domestica and D. yakuba transgenes: The M. domestica per construct P M 1 for P-element transformation was prepared using the D. melanogaster per promoter and 5′ UTR, fused to the coding sequences and 3′ UTR of the Musca gene. To reduce the size of the transgene, the large (~5 kb) Musca intron 2 was removed from the construct; with this exception, the construct was assembled using genomic DNA. A PCR strategy was adopted to fuse the untranslated D. melanogaster portion of exon 2 to the Musca gene at the initial methionine codon. A 422-bp D. melanogaster DNA fragment, containing per sequence from 422 to 101 with respect to the translation start, was amplified from a D. melanogaster per clone, using primers 55d5 (5′-CCGAAATCTACGGAATTTG-3′) and 53d5 (5′-ATTCACTTTCCATGGTGTACG-3′). 53d5 carries a tail (underlined) complementary to part of the Musca coding sequence. A 160-bp Musca cDNA fragment (from 0 to +160 with respect to the starting methionine, corresponding to the genomic region encompassing the large intron 2) was amplified using 35mus (5′-AACCTAAAGCAGCATGGAGGTAATACGGATGAA101-3′) in conjunction with amp6 (5′-CCGCGGATCCGATTCTGACTGAA101-3′). The underlined region of 35mus does not bind to the Musca sequence but represents a region of complementarity to the 5′ UTR D. melanogaster sequence. Comparable amounts of the amplified fragments were then pooled together and again subjected to PCR so that the two complementary tracts would allow the fusion of the two fragments to give the chimeric DNA, and the two primers (55d5 and amp6) would only allow amplification of the chimeric fragment. A proof-reading DNA polymerase (Vent polymerase; New England Biolabs, Beverly, MA) was used to minimize the risk of mutagenesis and the amplified product was then sequenced to ensure no mutations were incorporated. The PCR product was then cut with the enzymes XbaI and BglII to give the resulting 0.5-kb chimeric exon 2 fragment, which was then joined to a 5-kb BglI-Sall genomic fragment (the remaining Musca sequences) and to a 6-kb BamHI-XbaI fragment (the D. melanogaster 5′ regulatory region), and inserted into the transformation vector pW8 (Klemenz et al. 1987) linearized with BamHI and XhoI.

The D. yakuba per orthologue (Thackeray and Kyriacou 1990) was cloned into the Carnegie 20 transformation vector (Spradling 1986), after replacing D. yakuba's upstream sequences with those of D. melanogaster, by swapping a 4-kb XhoI-Sall fragment between the two species. The resultant transgene, pMY1, retains the D. melanogaster 5′ sequence up to the Sall site in the first untranslated exon (Citri et al. 1987). All coding sequences are therefore from D. melanogaster, whereas nearly all the regulatory material is from D. melanogaster.

The published results from the transfectants carrying the following per transgenes are also cited in this study: mps1, which carries the D. pseudoobscura per coding sequences and 3′ UTR, fused to the D. melanogaster 5′ regulatory region at a point close to the 3′ end of the large first intron (Peterson et al. 1988); mps3, a chimeric D. melanogaster/D. pseudoobscura per, in which D. melanogaster provides the 5′ regulatory region and N-terminal coding sequences up to just before the Thr-Gly encoding repeat, and D. pseudoobscura contributes the C-terminal half of the coding sequence and 3′ UTR (Peixoto et al. 1998); per− transgenes from D. melanogaster, carrying the 13.2-kb per transcription unit and either rosy+ (rosy or white) (w+) eye markers (Citri et al. 1987; Sawyer et al. 1997; Peixoto et al. 1998); Ap, a transgene carrying the Antheraea peryi per cDNA and 3′ UTR fused to the D. melanogaster 5′ regulatory regions (Levine et al. 1995).

P-element transformation: Transformation of Drosophila embryos was carried out according to Spradling (1986). The strain used for the microinjections of P M 1 was w; +/++; Sb, e; Δ2-3/TM6, which contains a stable P element (Δ2-3) on the third chromosome as a source of transposase (Robertson et al. 1988). The transformation vector used was pW8, which carries the mini-white gene. Plasmid for injections was purified with the QIAGEN bp-500 (QIAGEN, Chatsworth, CA) columns, using manufacturer’s instructions. PCR on the transfectants’ genomic DNA showed that the Δ2-3 element was successively crossed out of the transformed lines.

Microinjection of the pMY1 (D. yakuba) transgene, carrying a rosy− selectable marker, was performed using per−; rosy+ hosts and the helper plasmid p25.7wc (Karess and Rubin 1984). The chromosomal location of pM1 and pMY1 inserts was determined using appropriate balancer stocks, while the number of inserts within each line was checked by means of Southern blotting.

Behavioral analyses: Fly locomotor activity was monitored with the use of an activity event recorder (e.g., see Hamblen et al. 1986) produced by Biodata Ltd. (Manchester, United Kingdom), consisting of many individual activity units, each sandwiched between two infrared photocells. Single flies were loaded into glass tubes and each glass tube was clamped between the diodes of the photocells. The flies were entrained in a 12:12 LD photoperiod for 2 days prior to the start of the activity recording in constant darkness (DD). Data were collected over 7 days in a 30-min bin format. The periodicity was calculated by spectral analyses, performed with the CLEAN algorithm of Robert et al. (1987), which was run on a Silicon Graphics platform. Significance levels were determined by Monte Carlo simulation as described in Peixoto et al. (1998).
In addition, all activity data were analyzed by autocorrelation, from which significant periods, at least at the 5% level, were extracted. Only flies with significant periods from both the spectral and autocorrelation procedures were judged as “rhythmic” (see Sawyer et al. 1997; Peixoto et al. 1998).

**Computer analyses:** DNA and protein data analyses were performed using various programs of the GCG package for molecular biology (version 8; University of Wisconsin Genetics Computer Group, Madison, WI; Deverux et al. 1984). Multiple sequence alignments were performed with the program ClustalW (Higgins and Sharp 1988) and corrected by eye. The phylogenetic analyses were computed with the PHYLIP (Phylogeny Inference Package, version 3.57c) package provided by J. Felsenstein (University of Washington, Seattle, Washington). DNA phylogeny was performed applying Kimura’s two-parameter model (Kimura 1980), while protein distance matrices were calculated using either the PAM (Dayhoff 1979) or Kimura (1983) methods, and the phylogenetic trees were generated with the UPGMA algorithm (Sneath and Sokal 1973). PEST sequence analyses were performed with the PEST-FIND program (Rogers et al. 1986).

**RESULTS**

**Cloning of the *M. domestica* per homologue:** *M. domestica* per spans 9 kb from the starting codon to the putative polyadenylation signal (GenBank accession nos. AF142662, AF142663, and AF142664). This dramatic increase in size compared to its Drosophila orthologues (e.g., Citri et al. 1987; Colot et al. 1988; Thackeray and Kyriacou 1990) is deceptive, in that the encoded 1048-residue protein is shorter than that of *D. melanogaster*. The size of the gene is increased by a modification in its intron-exon structure, with four additional introns (Figure 1). Apart from the large intron 2, which has expanded to 5 kb compared to its positional homologues in *D. melanogaster* and *D. virilis*, which are 60 and 70 bp, respectively (Colot et al. 1988), the other introns falling within the coding sequence are small, ranging in size between 50 and 72 bp. Intron 2 is located at exactly the same position in all dipteran *per* genes and the intron/exon boundaries are well conserved. The available 1.6-kb sequence from Musca intron 2 shows traces of at least one duplication event involving 84 bp, indicating how this intron may have expanded to its final size from a shorter ancestor. All the remaining introns of the *Musca per* gene are short, resembling those in *D. melanogaster*, including introns 3, 4, 5, and 9, which do not have a Drosophila counterpart. Interestingly, one RT-PCR product carried the sequence corresponding to amino acids RLKLSPPFYSETNNGGTS TQ, which is found in the *Musca* genomic sequence, inserted just N terminal to the putative *Musca NLS*. However, subsequent RT-PCRs failed to confirm this cDNA, and so this sequence corresponds either to a very rare head transcript or is an RT-PCR artefact. In any case, it represents the novel intron number 3 (see Figure 1). In the conserved region known as “c2” (Colot et al. 1988) are introns 5, 6, and 7, the latter two having a Drosophila positional homologue occurring in the same codon and in the same phase. Introns 8, 10, and 11 of *Musca* have a Drosophila equivalent even though the divergence between the two *per* genes does not allow an alignment of the neighbouring exon sequence. All the introns in the *Musca* gene are rich in A + T (66%) as in *D. pseudoobscura* and *D. virilis* (62 and 60%, respectively) but in contrast with *D. melanogaster* and *D. yakuba* (52–53%).

**The *Musca PER* protein:** The *Musca per* transcript contains an ORF encoding for the 1048-amino-acid long polypeptide depicted in Figure 2, with a putative molecular weight of 116 kD. The division of per into conserved (c) and nonconserved (nc) regions was introduced upon comparison of the gene in three different species, *D. melanogaster*, *D. pseudoobscura*, and *D. virilis* (Colot et al. 1988). A fourth Drosophila *per* gene, cloned from *D. yakuba* (Thackeray and Kyriacou 1990), gives the short evolutionary distance of this species from *melanogaster* (6–15 million years; Lachaise et al. 1988; Russo et al. 1995), does not show much variation, even in the so-called nonconserved regions, when compared to its closely related homologue. This overall pattern of variation is largely preserved in the housefly gene; the six conserved blocks are clearly apparent upon comparison of *Musca per* with any of the Drosophila homologues. As in Drosophila, c1 and c2 constitute most of the N-terminal half of the protein, while c3, c4, c5, and c6 are localized in the C-terminal half and are generally less well conserved (Figure 2). In Musca the similarity of c1, c2, and c3 to the Drosophila proteins is very high, between 80 and 94%, slightly lower in c6 (75–82%), and considerably lower in c4 and c5 (57–71%), as scored by the GCG program Bestfit.

The N-terminal block c1 contains the NLS (Vosshall et al. 1994; Saez and Young 1996). Interestingly, a second conserved putative NLS is found in c3, but a functional analysis of this signal has not been reported. The longest conserved block is c2, representing almost half the length of the entire *Musca PER* protein and containing the PAS dimerization region (Huang et al. 1993) and the cytoplasmic localization domain (CLD; Saez and Young 1996). Our definition of PAS includes residues 238–496 (in the *D. melanogaster* sequence; e.g., see Pelléquer et al. 1998), and begins a few amino acids upstream of the first 51-residue PASA degenerate repeat, and ends downstream of the PASB repeat after
Figure 1.—Intron/exon comparison of D. melanogaster (top) and M. domestica per (bottom). Open boxes represent the 5′ UTR. The regions of high homology (conserved regions) are depicted in black, while the nonconserved regions are in gray. The NLS, PAS (including PAC/CLD), and Thr-Gly domains are indicated. Introns are numbered. Musca intron 1 is depicted by a broken line to indicate that its length and complete sequence have not been determined.

the PAC domain (Ponting and Aravind 1997). The PAC domain includes the CLD as defined by the deletion studies of Saez and Young (1996), except for a few C-terminal residues that cannot be unaligned between the species. This broad definition of PAS encompasses all the N-terminal regions that physically interact with TIM (Saez and Young 1996). The sites to which the per1, per3, and per10 mutations have been mapped (Baylies et al. 1987; Yu et al. 1987) are included within this region and are perfectly conserved in all PER proteins (Figure 2).

Secondary structure analyses of the predicted Musca protein sequence with the PHDsec program (EMBL) identified an HLH domain located in c2 [amino acids (aa) 450–512], at the end of the CLD (Figure 2). The same structural motif was also found in the D. melanogaster sequence (aa 525–571). The Musca candidate HLH lies in a different region of the protein from the one suggested in the mammalian mper1 homologue (Sun et al. 1997). Musca c5 also contains an opa repeat (CAG), which generates a cluster of glutamines at positions 907–917, a feature associated with transcriptional activators (Courey and Tjian 1988; Emili et al. 1994); despite this poly-Q stretch being localized in a conserved region, none of the other Drosophila orthologues display a similar motif. A poly-Q stretch is also found in nc2 of D. virilis. In nc2 lies the Thr-Gly repeat, and as reported in other non-Drosophilid dipterans (Nielsen et al. 1994), the Musca repeat of two Thr-Gly pairs has not undergone the dramatic expansion in size observed in the Drosophila genus (Costa et al. 1991; Peixoto et al. 1992, 1993). Various PEST sequences (Rogers et al. 1986) and phosphorylation sites are also found within the PER proteins (Figure 2). One putative site for casein kinase II phosphorylation is found in all the Dipteran sequences within the C-terminal conserved PEST motif (see Figure 2).

Molecular phylogeny of the PER proteins: The phylogeny of the six species D. melanogaster, D. yakuba, D. pseudoobscura, D. virilis, M. domestica, and A. pernyi is well known from traditional taxonomic approaches. A. pernyi belongs to the order Lepidoptera, which was already well differentiated at the end of the Triassic era 200 mya (Boudreaux 1978). The group Calyptratae (to which M. domestica belongs) diverged from the group Acalyptratae (which includes the Drosophilidae) 100 mya (Hennig 1981); the time of divergence of D. melanogaster and D. virilis is estimated to be ~40 mya (Schlotterer et al. 1994), the obscura group (to which D. pseudoobscura belongs) separated from the melanogaster group between 25 mya (Russo et al. 1995) and 30 mya (Schlotterer et al. 1994), and the phylogenetic dis-
Figure 2.—Alignment of the conserved regions of the Per proteins from *D. melanogaster*, *D. yakuba*, *D. pseudoobscura*, *D. virilis*, *M. domestica*, and *A. pernyi* (see materials and methods). An alignment is also attempted in nonconserved region 1 (nc1), in nc4, and in nc5. The ends of the sequences are indicated by asterisks. Dots represent equal amino acids and dashes represent deletions. The positions of introns are indicated by arrows on the *melanogaster* and *Musca* sequences. The nuclear translocation signals (NLS) are boxed in gray, with the label NLS below. The putative HLH domain is indicated by a double underline and the CLD is boxed on the *melanogaster* sequence. Shown in boldface type are also the polyQ stretches. PEST regions are underlined; the casein kinase II site within the C-terminal PEST is boxed in gray. The PAS region (from residue 238 to 496 of the *D. melanogaster* sequence and that includes PAC/CLD) is shown with a thick underline, with the two degenerate PAS repeats underlined in gray.
Figure 2.—Continued.
Figure 2.—Continued.
DNA sequence coding for the PAS domain (Figure 3A). We used Kimura's two-parameter method (Kimura 1980) to estimate the evolutionary distance. As can be seen, there is no ambiguity in that the species tree is faithfully reproduced. We then generated a phylogeny based on the alignable amino acid sequence from non-PAS regions c1 + c3 (Figure 3B), employing the PAM distance matrix (Dayhoff 1979), which takes into account the fact that some amino acid replacements occur at higher frequencies than others, irrespective of the necessary number of nucleotide substitutions. A tree similar to the DNA tree from Figure 3A was obtained, except that D. virilis and D. pseudoobscura had swapped positions.

The third fragment of PER used in this analysis was the PAS region from c2 (including PAC/CLD, residues 238–496 of the D. melanogaster sequence), which represents a functional domain of PER. The PAS tree places Musca PAS closer to D. melanogaster than D. pseudoobscura and D. virilis (see Figure 3C), contradicting the species tree drawn from per DNA (Figure 3A). A similar switching of positions of the Musca and D. pseudoobscura/virilis groups, with similarly high bootstrap values, was also observed using Kimura's (1983) protein distance matrix (data not shown), so the tree is reasonably robust. This unusual PAS phylogeny is reflected in the smaller number of differences between the D. melanogaster/M. domestica pairwise comparison (29 aa changes + 1 aa deletion) compared to that of D. melanogaster/D. pseudoobscura (33 aa changes) or D. melanogaster/D. virilis (44 replacements).

**Rescue of circadian rhythmicity in transgenic D. melanogaster carrying the housefly per.** To study any possible functional implications of the unusual phylogenies observed for the PAS domain, we transformed the Musca per gene into arrhythmic Drosophila per mutants. The transgene, pM 1, carries the 5′ D. melanogaster regulatory sequences until the first coding methionine, and the coding sequence and 3′ UTR of Musca per. Two transgenic lines, each carrying one autosomal copy of pM 1, were studied for rescue of free-running circadian locomotor activity. In addition, we also studied one transgenic line, pMY1-M7, which carried an autosomal copy of the D. yakuba per coding sequences (Thackeray and Kyriacou 1990) fused to the D. melanogaster 5′ region. Both the pM 1 and pMY1 transgenes were studied in per−/+ males. Table 1 reveals the patterns of rescue observed in these various Musca and D. yakuba transgenic lines. Included in Table 1 are the results obtained in our laboratory for transgenic lines carrying a single autosomal copy of the D. pseudoobscura per coding sequences fused to the 5′ regions of D. melanogaster per (the mps1 transgene; see materials and methods). In addition, the data are also illustrated from transformants carrying a single copy of the chimeric transgene, mps3, also from our laboratory, in which the 5′ regulatory and coding regions of D. melanogaster were fused to the 3′ sequences of D. pseudoobscura at a position

---

**Figure 3.**—Phylogeny of different fragments of the per coding sequence or PER protein. Tree A represents the molecular phylogeny of the PAS-encoding DNA, and was obtained by applying the Kimura two-parameter method (Kimura 1980) to the UPGMA tree-drawing method (Sneath and Sokal 1973). The following two trees, based on amino acid sequence, were obtained with the PAM distance matrix (Dayhoff 1979) and the UPGMA method. B represents the phylogeny of the c1 + c3 regions and C that of the PAS region (residues 238–496 of the D. melanogaster sequence). The numbers represent the confidence limits of the branching on their right. The length of each branch is proportional to the value of the distance calculated by the program.

D. melanogaster and D. yakuba is 6–15 mya (Lachaise et al. 1988; Russo et al. 1995). Although there are uncertainties about the exact time of divergence, there are no ambiguities in the branching order of these species (see below). We used the phylogenetic approach to examine whether there is any significant difference between the species tree and the PER protein tree, which could be taken as an indicator of unusual events in the evolution of PER protein sequences. We were particularly interested in analyzing the PAS domain, which has been implicated in the protein-protein interactions between PER and TIM, and comparing it with the evolution of non-PAS sequences (Huang et al. 1993; Gekakis et al. 1995; Saez and Young 1996).

First, a molecular phylogeny was computed on the DNA sequence coding for the PAS domain (Figure 3A).

---

754 A. Piccin et al.
corresponding to D. melanogaster residue 639, ~180 bp upstream of the Thr-Gly-encoding repeat region (see materials and methods). The data from mps1 and mps3 are taken directly from Peixoto et al. (1998). Also from this study, the data are included from control transgene, with either ry− or w− markers. Finally, and also in Table 1, are the results obtained by Levine et al. (1995) for a perz transformant line, M 4-15, which gave the best reported rescue of rhythmicity when carrying a transgene encoding the corresponding A. peryi per cDNA (Ap) fused to the D. melanogaster 5′ regulatory region. All these transgenes therefore carried D. melanogaster regulatory sequences ligated to various species-coding sequences, and all the Drosophila and Musca constructs were derived from genomic DNA, except that the large intron 2 of Musca per was removed.

The results reveal a striking correlation between the level of rescue of rhythmicity and the phylogeny of the PAS region (Figure 3C). The D. pseudoobscura per transgene, mps1, rescues rhythmicity in a perz background relatively poorly, with significant rhythmicity observed in ~50% of individuals, but with longer-than-normal periods. The rescue obtained in our study with the mps1 transgene is better than that reported by Petersen et al. (1988) for the same transgenic strains, in which rhythmicity was ~10%. The difference in our results is due to our use of a more sensitive statistical measure of rhythmicity (see Sawyer et al. 1997; Peixoto et al. 1998).

In contrast, the Musca per transgene, pMM 1, rescues behavior remarkably robustly, with 80–100% of individuals showing statistically significant rhythms, although the periods are ~2 hr shorter than the corresponding D. melanogaster per transformants. Furthermore, our spectral analyses (see Sawyer et al. 1997; Peixoto et al. 1998), mean that the strength of individual rhythms will broadly correlate with the proportion of flies that are rhythmic within each genotype class, and not surprisingly, pMM 1 transformants have much stronger individual rhythms than mps1 (data not shown). The D. yakuba per transgene, pMY 1, also yields very robust rhythms with >80% of individuals giving statistically significant cycles with a free-running period of ~23.5 hr. The line carrying the A. peryi per transgene, which best rescues rhythmicity, generated only ~20% rhythmic individuals with very short periods (Levine et al. 1995). Finally, the mps3 chimeric transgene, in which the N-terminal coding sequences of D. pseudoobscura have been replaced with those of D. melanogaster, generates essentially wild-type rhythms (Peixoto et al. 1998). The D. virilis per gene has yet to be transformed into D. melanogaster hosts. Thus the M. domestica per transgene provides a robust rescue of perz rhythms, which belies its evolutionary position relative to D. pseudoobscura.

**DISCUSSION**

**Conservation of the Musca per homologue:** The cloning of the M. domestica per orthologue has revealed stretches of similarity in all the conserved regions first identified by Colot et al. (1988) in Drosophila per genes. However, the structure of the Musca gene is different from its Drosophila orthologues; even though the full length of the primary transcript is not known, it must be considerably longer than that of D. melanogaster given the increase in the size of intron 2. Also, the
number of introns is increased in the housefly gene, reflecting the size of the genome of _M. domestica_, which is about five times larger than that of _D. melanogaster_ (John and Miklos 1988). It is tempting to say that the increase in genomic complexity must correlate with an increase in both number and size of introns. Unfortunately, most of the available _Musca_ gene sequences come from cDNA libraries, so there are not sufficient genomic data to test this hypothesis.

In _D. melanogaster_, _per_ is sex-linked, being located at the 3B1-2 region, close to the tip of the X chromosome (Young and Judi 1978). In _D. pseudoobscura_, _per_ has also been mapped to the X chromosome (Petersen et al. 1988) while the chromosome location of _per_ in _D. virilis_ is not known. In situ mapping of the _Musca_ _per_ gene to polytene chromosomes was unsuccessful, but as the X chromosome of _M. domestica_ is entirely heterochromatic (Malacrida et al. 1985) it may contain very few, if any, genes. Indications that _per_ might be located on chromosome III of _Musca_ come from Malacrida et al. (1985), who described the correspondence between various linkage groups of _D. melanogaster_ and _M. domestica_. In particular, the _Musca_ genes corresponding to _Drosophila_ yellow and white, which lie close to _per_, and are called brown body and _w_, respectively, are found on the right arm of the housefly chromosome III.

The A + T content is relatively high in intron sequences from _Musca_, _D. pseudoobscura_ and _D. virilis_, but in contrast to _Drosophila_, this high A + T profile is maintained in the _Musca_ coding sequence. This reflects a bias in _Drosophila_ _per_ for the usage of codons that either terminate in C or G and that is seen particularly in highly expressed genes (Sharp et al. 1992). These authors used the term “optimal codons” to describe these frequently used triplets, which may reflect translational selection among synonymous codons, mutational trends (Moriyama and Gojobori 1992), or selection for particular structures in DNA. The few _Musca_ genes analyzed cannot provide much insight for elucidating the codon usage in this organism, but if _Musca_ _per_ is as extensively expressed as _Drosophila_ _per_ at least in the adult (e.g. Plautz et al. 1997), then unlike _Drosophila_, any bias in codon usage toward a higher frequency of optimal codons may be A- and T-ending.

At the protein level, _c1_ and _c2_ regions show high similarity between all species (Figure 2), underscoring the importance of PAS in the biochemical function of PER, and suggesting that an equally important role may be played by _c1_, in which is found the NLS. In _c2_, two sites are found within the PAS domain that, when mutated, decrease dimerization efficiency: the _per_ site and a cluster of amino acids at position 413–419 (in the second PAS repeat) of the _D. melanogaster_ protein (Huang et al. 1993). The effect on protein-protein interactions of this amino acid cluster was assayed because the residues contained in the fragment are highly conserved in the PAS regions of _AHR_, _ARNT_, _SIM_, and _PER_ (Huang et al. 1993). Both of these areas of PAS are highly conserved in _Musca_. Another area in _c2_ has been identified as a short period domain in which mutations consistently shorten the circadian period (Baylies et al. 1987; Rutila et al. 1992). It extends from 3 amino acids upstream to 16 downstream of the _per2_ site. The high degree of similarity suggests that in _Musca_ this area retains its functional importance. An intriguing feature of the C-terminal region of both _Musca_ and _Drosophila_ PAS domains was the suggestion of an HLH motif, which was generated with the use of the PHDsec structural algorithm. The report of a similar motif in mammalian PER, albeit in another region (Sun et al. 1997), seems more than coincidence, so whether these regions can act as an HLH domain should be tested.

Phosphorylation of PER by the DBT casein kinase 1ε has been shown to play an important role in contributing to the delay observed in peak levels of _per_ transcript and protein (Kloss et al. 1998; Price et al. 1998). In the yeast fructose-1,6-biphosphatase, phosphorylation transforms a weak PEST region into a strong proteolytic signal (Rechsteiner 1988). Similarly, the degradation of PER appears to occur mainly at the level of the phosphorylated forms, whose appearance triggers the negative feedback on _per_ transcription (Edery et al. 1994).

By analogy to the yeast protein, a conditional PEST region(s) in PER could be activated by phosphorylation. Although many consensus phosphorylation sites are found in the different PER proteins, it is interesting that the conserved C-terminal PEST region also carries a conserved potential phosphorylation site for casein kinase II.

Circadian rhythmicity of locomotor activity is restored in _D. melanogaster_ _per_ mutants expressing one copy of _Musca_ _per_. More than 80% of the transformants display rhythmic behavior. This is in stark contrast with what has been reported by Petersen et al. (1988) and later by Peixoto et al. (1998), where a comparable transgene, _mps1_ containing the _D. pseudoobscura_ _per_ coding sequence fused to the _D. melanogaster_ promoter, generates a significantly weaker rescue. As the processing of the 5′ introns of _mps1_ has been assayed by Petersen et al. (1988) and found to be normal, and the _mps3_ construct, which encodes the C-terminal half of the _D. pseudoobscura_ protein, gives wild-type rescue of behavior (Peixoto et al. 1998), suggesting that the 3′ introns are also processed normally, it is highly unlikely that the poor rescue of _mps1_ is due to problems in producing the transcript. In addition, _mps3_ enhances rhythmic behavior back to wild-type levels, thus mapping the poorer rescue of _mps1_ to the _D. pseudoobscura_ coding sequences in the N-terminal half of _PER_ (Peixoto et al. 1998). Thus, attention is drawn to the N terminus and the PAS domain, implicated in protein-protein interactions with TIM, for explaining the different levels of transgenic rescue (Gekakis et al. 1995; Saez and Young 1996).

Given the unusual phylogeny of PAS (Figure 3C),
the ability of Musca PER to direct efficient rescue of rhythmicity in D. melanogaster aperiodic mutants, in contrast to D. pseudoobscura PER, could mean that PAS-mediated PER-TIM interactions can take place in an almost normal fashion between Musca PER and the host D. melanogaster TIM, as opposed to the melanogaster-pseudoobscura pairing. This idea could also be extended to any PAS-mediated interactions between Musca PER with the hosts CLOCK and BMAL1. This simple but compelling explanation for the functional data may represent an example of intermolecular coevolution between PER and its various partners. Testing this hypothesis experimentally for the TIM interaction would require the simultaneous transformation into D. melanogaster double mutant per<sup>tm1</sup>itim<sup>0</sup> hosts, of both D. pseudoobscura tim and per, with the expectation that the levels of rescue should be significantly improved over those seen with the D. pseudoobscura mps1 transgene (Peterson et al. 1988; Peixoto et al. 1998). In addition, coevolution between PER and TIM (or CLOCK and BMAL1) might also be seen at the sequence level in phylogenetic trees, with a switching of the relative positions of the relevant D. pseudoobscura and M. domestica TIM sequences as observed for PAS (Figure 3C).

In conclusion, the identification and isolation of Musca per has provided some initial surprises. Without the phylogenetic perspective, the results obtained from the transformation experiments would have been difficult to interpret. We suggest an initial, simple, testable hypothesis based on PER-TIM coevolution to explain the differential success of interspecific clock gene transformations to rescue clock function in D. melanogaster, and to this end we are attempting to identify PER partners in both Musca and D. pseudoobscura.

M.C. thanks the Biotechnology and Biological Research Council (BBSRC) for a studentship. C.P.K. and R.C. were supported by a European Community grant under the Human Capital and Mobility programme. In addition, C.P.K. acknowledges grants from BBSRC and Human Frontiers Science Programme and R.C. acknowledges grants from Ministero Università e Ricerca Scientifica e Tecnologica and Ministero delle Risorse Agricole, Alimentari e Forestali.

LITERATURE CITED


Levine, J. D., J. Sauman, M. Imbalzano and S. M. Reppert, 1995 Period protein from the giant silkmoth Antheraea pernyi functions...

Communicating editor: J. J. Loros