Evolution of the Threonine-Glycine Repeat Region of the *period* Gene in the *melanogaster* Species Subgroup of *Drosophila*

A. A. Peixoto,1 R. Costa,2 D. A. Wheeler,3,* J. C. Hall,3 and C. P. Kyriacou1

1 Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK
2 Dipartimento di Biologia, Universita' di Padova, Via Trieste 75, 35121 Padova, Italy
3 Department of Biology, Brandeis University, Waltham, MA 02254, USA

Summary. The Threonine-Glycine (Thr-Gly) region of the *period* gene (*per*) in *Drosophila* was compared in the eight species of the *D. melanogaster* subgroup. This region can be divided into a diverged variable-length segment which is flanked by more conserved sequences. The number of amino acids encoded in the variable-length region ranges from 40 in *D. teissieri* to 69 in *D. mauritiana*. This is similar to the range found within natural populations of *D. melanogaster*. It was possible to derive a Thr-Gly “allele” of one species from that of another by invoking hypothetical Thr-Gly intermediates. A phylogeny based on the more conserved flanking sequences was produced. The results highlighted some of the problems which are encountered when highly polymorphic genes are used to infer phylogenies of closely related species.

Key words: *Drosophila* — *per* gene — Threonine-Glycine — repeat sequence — *melanogaster* subgroup phylogeny

Introduction

The *period* (*per*) gene in *Drosophila* determines biological rhythmicity in the circadian (Konopka and Benzer 1971), ultradian (Kyriacou and Hall 1980, 1989) and infradian (Kyriacou et al. 1990) time domains. It has been studied at the behavioral, genetic, neurobiological, molecular, and evolutionary levels. (See Kyriacou and Hall 1993 for review.) The *per* coding sequences have been reported in *D. melanogaster* (Jackson et al. 1986; Citri et al. 1987), *D. yakuba* (Thackeray and Kyriacou 1990), and *D. pseudoobscura* and *D. virilis* (Colot et al. 1988). A general feature appears to be that the 3' end of the gene is more variable between species than the 5' end.

Within the fifth exon, there is, in *D. melanogaster* (Jackson et al. 1986; Citri et al. 1987) and the closely related *D. yakuba* (Thackeray and Kyriacou, 1990), a repeat motif which encodes a number of Threonine-Glycine (Thr-Gly) pairs. The number of Thr-Gly pairs in *D. melanogaster* is highly polymorphic among laboratory strains (Yu et al. 1987a) and in natural populations (Costa et al. 1991). In *D. pseudoobscura* the Thr-Gly repeat is replaced by a five-amino-acid degenerate motif (Colot et al. 1988) which is also polymorphic in length (Costa et al. 1991). *Drosophila virilis* has a very short Thr-Gly region (Colot et al. 1988). The evolution of the Thr-Gly encoding length variants within *D. melanogaster* appears to be due to slippage-like events involving deletion and duplication of Thr-Gly pairs (Costa et al. 1991).

The Thr-Gly length polymorphism within *D. melanogaster* shows a robust north-south clinal distribution in natural populations from Northern Eu-
per yakuba.

However, deflanking regions can be used for a traditional phylogeny of the subgroup. This view is supported by the observation that removing the perThr-Gly repeat from D. melanogaster (Yu et al. 1987a) and transforming this mutated construct into amorphic per mutants (Baylies et al. 1987; Yu et al. 1987b), leads to a loss of temperature stability in the circadian phenotype (Ewer et al. 1990). Temperature compensation is a fundamental feature of circadian cycles in that the period of any circadian rhythm remains close to 24 h in spite of large variations in temperature (Pittendrigh 1954). Perhaps then the Thr-Gly region plays a role in the thermostability of the per molecule.

The region around the Thr-Gly repeat has also been implicated in the control of species-specific ultradian courtship song cycles (Kyriacou and Hall 1986; Wheeler et al. 1991). These song cycles may play a role in species recognition, an important evolutionary mechanism for maintaining species boundaries. Consequently this repetitive region of per appears to play a prominent functional role within the gene.

In this paper we have analyzed the Thr-Gly region in the eight species of the melanogaster subgroup of species. We observe that the DNA sequences in the repeat are very variable between species due probably to a combination of slippage events and point mutations. The more conserved flanking regions can be used for a traditional phylogenetic analysis of the subgroup. However, detailed analysis of the Thr-Gly encoded repeat further allows us to produce a tentative phylogeny of the different Thr-Gly alleles of D. melanogaster and its sibling species.

Materials and Methods

Drosophila strains. The following Drosophila strains were used in this work: D. simulans Ken (Kenscoff-Haiti), D. simulans Aus (Australia), and D. ananassae (0371.0) obtained from Bowling Green State University, Ohio; D. erecta, obtained from the Umea stock center, Sweden; D. mauritiana, D. teissieri, and D. yakuba strains were obtained from Dr. Barrie Burnet (Sheffield, U.K.); D. sechellia and D. orena were provided by Dr. Jean David (Gif-sur-Yvette, France).

PCR Amplification and DNA Sequencing. The sequence of the D. simulans strains were obtained by cloning and plasmid sequencing as described in Wheeler et al. (1991). The sequences of the remaining species were obtained by PCR and direct DNA sequencing as previously described (Costa et al. 1991). Briefly, the PCR amplification was carried out for 30 cycles (95°C for 1 min, 65°C for 1 min, and 70°C for 1 min) in a Perkin-Elmer-Cetus thermo cycler. We used the same primers as in Costa et al. (1991): 5’ primers: 5'-AACTATAACGAGAACCTGTCTG-3' (4874-4893) and 5'-CCCCTGACCCAGGGCGGACGGGGG-3' (5005-5028); 3’ primers: 5'-TTCTCCATCTCGTCGTTGTTG-3' (5336-5355) and 5'-GCCGGCAGCTCCGGTCTCCTTG-3' (5364-5387). The primer positions (in parentheses) refer to the D. melanogaster sequence published by Citri et al. (1987). Fly DNA used in the PCR reactions was prepared using the method of Gloor and Engels (1990). Single males were ground in 50 μl of buffer (10 mM Tris-HCl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 μg/ml Proteinase K), left for 30 min at 37°C, and then heated above 95°C for 2 min; 1 μl of genomic DNA was used for each 10 μl of PCR reaction. The DNA sequencing was carried out by using the Sequenase version 2.0 kit from Unites States Biochemical.

Analysis of the DNA Sequences. The sequences obtained were analyzed using the University of Wisconsin Genetics Computer Group (UWGCG) software (version 5.0; Devereux et al. 1984) and the program CLUSTAL V (Higgins and Sharp 1988).

Results and Discussion

Evolution of the Thr-Gly Region of the Per Protein

Figure 1 shows the alignment of the putative per protein in the Thr-Gly region of the eight species in the melanogaster subgroup. The amino acids are arbitrarily numbered according to their relative position in the alignment. The sequence of D. melanogaster is from Citri et al. (1987). PCR amplification and direct DNA sequencing of the Thr-Gly encoding region from D. yakuba genomic DNA revealed three differences from the published sequence (Thackeray and Kyriacou 1990). This was further confirmed by PCR amplification and direct sequencing from the cloned D. yakuba gene. With respect to the published sequence the differences are: a Thr instead of a Ser at position 101, a Gly instead of a Ser at position 110, and a Ser instead of an Ala at position 129.

The entire region can be subdivided into a variable segment (positions 33–124) flanked by two well-conserved sequences. The N-terminal conserved region shows only one amino acid substitution, Glu→Gly (E→G) at position 1 in D. erecta and D. orena in relation to the other species. The C-terminal conserved region does not show any amino acid substitutions between the species.

The variable region shows a number of interesting features. First there is extensive evidence for the occurrence of slippage-like events, for example, in the duplication of the initial Gly-Thr-Gly (GTG) motif (position 33–35) which has occurred in D. erecta and D. orena. Furthermore, the variability in the length of this region ranges from 40 amino acids in D. teissieri to 69 amino acids in D. mauritiana. This is fairly similar to the size range observed in