Letter to the Editor

Codon Usage and the Origin of *P* Elements

Emmanuelle Lerat,† Christian Biémont,† and Pierre Capy**

*Laboratoire Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, Gif/Yvette, France; and †Laboratoire Biométrie et Biologie Évolutive, Centre National de la Recherche Scientifique, Université de Lyon I, Villeurbanne, France

The authors of a recent comparison of the *P* transposable element and three genes of *Drosophila melanogaster* and *Drosophila willistoni* suggested that the codon usage of the *D. melanogaster P* element is similar to that of *D. willistoni* genes (Powell and Gleason 1996). They concluded that this could be further evidence of the recent horizontal transfer of the *P* element from *D. willistoni* to *D. melanogaster* indicated by several previous findings (Clark and Kidwell 1997). More specifically, it was shown that *D. willistoni* genes tend to be T-ending-codon genes, whereas those of *D. melanogaster* tend to be C-ending-codon genes. The transposase genes of the *P* elements from both species was found to be AT-ending-codon genes, and one explanation for this may be that the *P* element of *D. melanogaster* originated from *D. willistoni.* This hypothesis assumes that the codon usage in transposable elements (TEs) and that in the host genome are similar. However, analysis of a large number of genes and TEs in *D. melanogaster* suggests that the T-ending-codon feature of the *P* element could be a general characteristic of all TEs in *Drosophila* species and independent of the host genome (Shields and Sharp 1989).

We extracted from the GenBank DNA sequence database the sequences of six genes common to *D. melanogaster* and *D. willistoni*: *Alcohol dehydrogenase* (*Adh,* M11290 and L08648), *Amylase*-related gene (*amyrel,* U69607 and AF039560), *superoxide dismutase* (*SOD,* Y00367 and L13281), *xanthine dehydrogenase* (*Xdh,* Y00307 and AF058985), *glycerol 3 phosphate dehydrogenase* (*Gpdh,* X80204 and L37038), and *period* (*per,* M30114 and U51055). The sequences of the last three genes have been only partially determined for *D. willistoni.* We therefore used only the homologous regions in both species to avoid bias due to differences in gene length (Moryiama and Powell 1998; Duret and Mouchiroud 1999). *P* elements, which have been reported in distantly related *Drosophila* species, were also added. These elements were described in *Drosophila bifasciata* (Hagemann, Miller, and Pinkser 1992), *Drosophila subobscura* (Paricio et al. 1991), and *Scaptomyza pallida* (Simonelig and Anxolabéhère 1991). All sequences of RNA (class I) and DNA (class II) elements described in the species previously mentioned were also used to compare *P* element features with those of other elements.

Key words: *P* element, transposable elements, *Drosophila,* codon usage, horizontal transfer.

Address for correspondence and reprints: Pierre Capy, Laboratoire Populations, Génétique et Evolution, Centre National de la Recherche Scientifique, UPR 9034, Bât. 13, 91198 Gif/Yvette Cedex, France. E-mail: capy@pge.cnrs-gif.fr.

Mol. Biol. Evol. 17(3):467–468. 2000

The relative codon frequencies were estimated for all sense codons (59 codons) for each gene and transposable element according to the formula

$$F_j = \frac{\mathrm{n_{ij}}}{\sum_{j=1}^{S_i} n_{ij}},$$

where n_{ij} is the number of codon j observed for the amino acid i, and s_i is the number of synonymous codons for the the amino acid i. The 59 columns of the matrix are the variables of a factorial correspondence analysis (FCA). Because relative and absolute codon frequencies can be sensitive to several biases (Perrière and Thioulouze, personal communication), FCA was also performed using absolute codon frequencies and relative synonymous codon usage (RSCU), frequently used for such analyses (see, e.g., Shields and Sharp 1989). In all cases, the FCA gave similar topologies on the first two axes. Figure 1 shows a factor map crossing the first two axes when the relative frequencies given above are used.

The first two axes accounted for 44% of the total variance (34% on the first axis and 10% on the second). The percentage of the variance explained by the remaining axes was very low, close to 1% for the third and fourth axes, respectively, and $<1\%$ for the other ones. *Drosophila melanogaster* and *D. willistoni* genes are clearly separated. The projection of the codons shows that GC-ending codons (gray ellipses) are more frequent in the *D. melanogaster* genes than in those of *D. willistoni,* which display several T-ending codons. The TEs (black spots) of different *Drosophila* species clearly display a higher frequency of AT-ending codons. Codon usage patterns of *P* elements from different species are similar and clearly differ from those of the host genes. Moreover, the codon usage variability among genes from different species is lower than that between genes and TEs within the same species. A MANOVA (Statistica, version 3.0b, StatSoft) using the coordinates of each point on the first two axes as variables shows that the difference observed between genes of *D. melanogaster* and *D. willistoni* is significant, with a *P* value of 7.34×10^{-4}, while the difference between transposable elements and genes of *D. melanogaster* is significant, with a *P* value of $<10^{-7}$.

The characteristics of the *P* element in *D. willistoni* described by Powell and Gleason (1996) could be a general feature of TEs in *Drosophila,* and not attributable to the host. This was suggested by a previous analysis of *D. melanogaster* (Shields and Sharp 1989) and is confirmed by similarities in *P* elements from different hosts. Moreover, these elements and other DNA and RNA el-

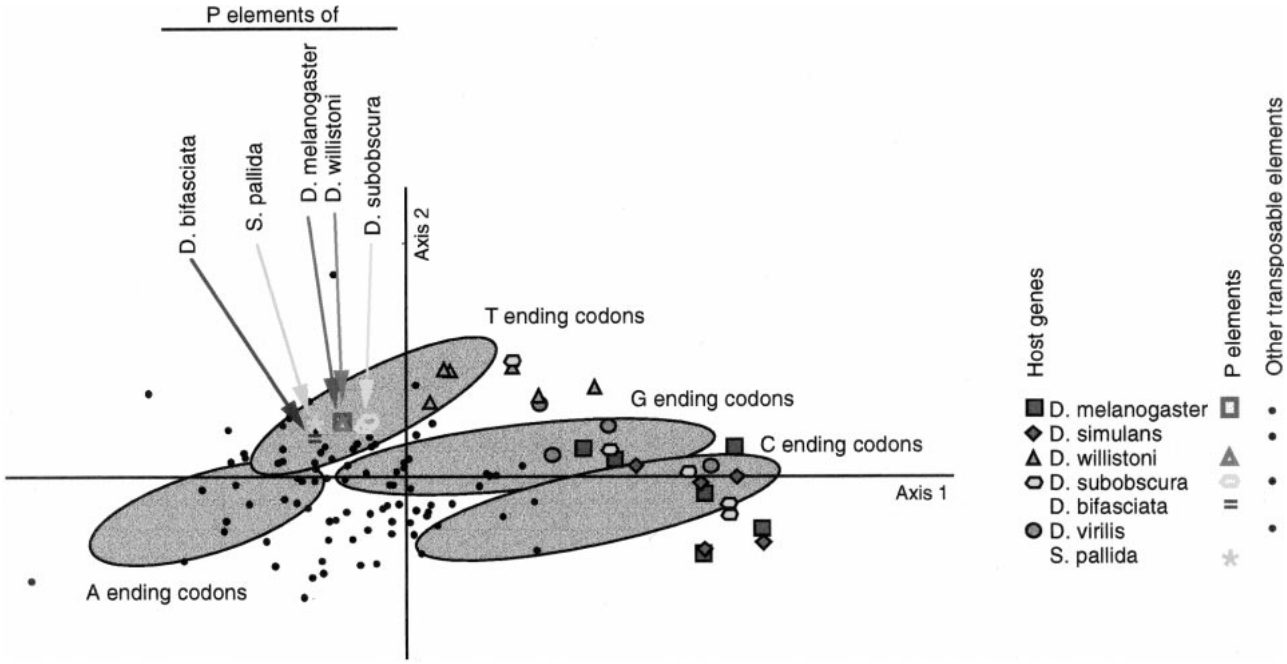

FIG. 1.—Projection on the first two axes on an FCA. The accession numbers of the *P* element sequences are X60990 for *jbifM3* of *Drosophila bifasciata* (Hagemann, Miller, and Pinkser 1992), S74793 for *A1* and *A2* of *Drosophila subobscura* (Paricio et al. 1991), and M63341 and M63342 for *PS2* and *PS18* of *Scaptomyza pallida* (Simonelig and Anxolabéhère 1991), respectively. Black spots represent the other DNA and RNA transposable elements described for different *Drosophila* species. Ellipses indicate the projections of the codons grouped according to their third bases.

ements described for several *Drosophila* species group together in our analysis (black spots in fig. 1).

These findings strongly suggest that codon usage cannot be employed to demonstrate horizontal transfers of TEs between *Drosophila* species. TEs and host genes may not be subject to the same constraints, but the possibility that the evolutions of these two entities are linked cannot be ruled out. For instance, the most frequent codon for the host gene could be the least frequent for TEs and vice versa. The only way to check this will be to analyze species using different codon usage strategies.

LITERATURE CITED

CLARK, J. B., and M. G. KIDWELL. 1997. A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. Proc. Natl. Acad. Sci. USA **94**:11428–11433.

DURET, L., and D. MOUCHIROUD. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila,* and *Arabidopsis.*. Proc. Natl. Acad. Sci. USA **96**:4482–4487.

HAGEMANN, S., W. J. MILLER, and W. PINSKER. 1992. Identification of a complete *P* element in the genome of *Drosophila bifasciata.* Nucleic Acids Res. **20**:409–413.

MORIYAMA, E. N., and J. R. POWELL. 1998. Gene length and codon usage bias in *Drosophila melanogaster, Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. **26**:3188–3193.

PARICIO, N., A. M. PEREZ-ALONSO, M. J. MARTINEZ-SEBASTIAN, and R. FRUTOS. 1991. *P* sequences of *Drosophila subobscura* lack exon 3 and may encode a 66 kd repressor-like protein. Nucleic Acids Res. **19**:6713–6718.

POWELL, J. R., and J. M. GLEASON. 1996. Codon usage and the origin of *P* elements. Mol. Biol. Evol. **13**:278–279.

SHIELDS, D. C., and P. M. SHARP. 1989. Evidence that mutation patterns vary among *Drosophila* transposable elements. J. Mol. Biol. **207**:843–846.

SIMONELIG, M., and D. ANXOLABÉHÈRE. 1991. A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **88**:6102–6106.

MANOLO GOUY, reviewing editor

Accepted November 18, 1999