Genetics and Molecular Biology, 27, 3, 373-377 (2004)
Copyright by the Brazilian Society of Genetics. Printed in Brazil
www.sbg.org.br

Research Article

Characterization of two full-sized *P* elements from *Drosophila sturtevanti* and *Drosophila prosaltans*

Juliana Polachini de Castro and Claudia M.A. Carareto

Universidade Estadual Paulista, Departamento de Biologia, São José do Rio Preto, SP, Brazil.

Abstract

Previously, only partial *P* element sequences have been reported in the *saltans* group of *Drosophila* but in this paper we report two complete *P* element sequences from *Drosophila sturtevanti* and *Drosophila prosaltans*. The divergence of these sequences from the canonical *P* element of *Drosophila melanogaster* is about 31% at the nucleotide level. Phylogenetic analysis revealed that both elements belong to a clade of divergent sequences from the *saltans* and *willistoni* groups previously described by other authors.

Key words: D. sturtevanti, *D. prosaltans*, full-size *P* element, phylogeny.

Received: May 28, 2003; Accepted: February 16, 2004.

Introduction

The *P* elements were first discovered in *Drosophila melanogaster* because of their ability to induce hybrid dysgenesis (Kidwell *et al.*, 1977). Autonomous *P* elements are 2.9 kb in length and have four open reading frames which encode two polypeptides, an 87 kDa transposase enzyme necessary for transposition (Rio *et al.*, 1986) and a 66 kDa repressor protein (Robertson and Engels, 1989). Also required for transposition are the element termini, which include flanking 31-bp perfect inverted repeats (O'Hare and Rubin, 1983), 11-bp subterminal repeats and unique terminal sequences comprising approximately 150 bp (see Engels, 1989 for a review).

Sequences belonging to the *P* family are particularly common in the four principal species groups (*melanogaster, obscura*, *saltans* and *willistoni*) which make up the subgenus *Sophophora* (Daniels *et al.*, 1990) but have also been described in drosophilid species such as *Drosophila mediopunctata* which is not part of the *Sophophora* subgenus (Loreto *et al.*, 2001) and also in *Scaptomyza pallida*, a drosophilid which does not belong to the genus *Drosophila* (Anxolabéhère *et al*., 1985; Simonelig and Anxolabéhère, 1991, 1994). Transposable elements similar to the *P* elements of the Drosophilidae have also been isolated from members of a few other Diptera families, *e.g. Lucilia cuprina* from the Calliphoridae (Perkins and Howells, 1992), *Musca domestica* from the Muscidae (Lee *et al*., 1999) and seven species *Anopheles* from the Culicidae (Sarkar *et al.*, 2003). More divergent and rudimentary sequences related to *P*-transposable elements have also been described using 'in silico' searches such as *Hoppel* (Reiss *et al*., 2003) and *Proto P* (Kapitonov and Jurka, 2003) for the *Drosophila melanogaster* genome and *Phsa* (Hagemann and Pinsker, 2001) for the human genome.

Phylogenetic studies based on nucleotide sequences (Clark and Kidwell, 1997; Hagemann *et al.*, 1994, 1996; Silva and Kidwell, 2000) indicated that the more than 200 *P* element sequences obtained to date fall into 16 distinct clades or subfamilies (Figure 1). Four of these subfamilies have been well characterized. The canonical subfamily appears to be restricted to the sophophoran New World species groups *saltans* and *willistoni* (Clark *et al.*, 1995), with the notable exception of *Drosophila mediopunctata,* which contains *P* elements due to horizontal transfer (Loreto *et al*., 2001). Three *P* element subfamilies (M-, O- and T-type) are found in the Old World *obscura* species group (Hagemann *et al.*, 1992, 1994, 1996), with the T-type appearing to be restricted to the *obscura* lineage (Hagemann *et al.*, 1998) while the M- and O-types also occur in the *saltans* and *willistoni* groups. A new subfamily, the K-type (restricted to the *montium* subgroup species), has recently been described by Nouaud *et al*. (2003).

The descriptions of the *P* element subfamilies in the *saltans* and *willistoni* species groups have been based so far mainly on partial sequences (Clark *et al.*, 1995; Clark and Kidwell, 1997; Haring *et al.*, 2000; Silva and Kidwell, 2000). The work described in this paper compared two complete sequences obtained from two different *saltans* subgroups (*sturtevanti* and *saltans*) to some of the complete sequences from different *P* element subfamilies as well as

Send correspondence to Claudia M.A. Carareto. Universidade Estadual Paulista, Departamento de Biologia, Rua Cristóvão Colombo 2265, 15054-000 São José do Rio Preto, SP, Brazil. E-mail: carareto@bio.ibilce.unesp.br.

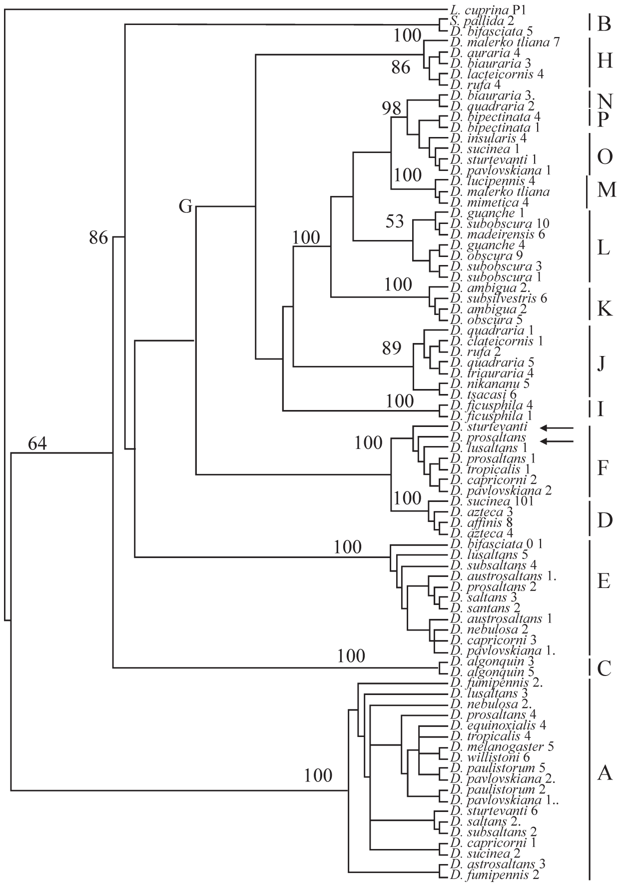

Figure 1 - Cladogram of *P* element nucleotide sequences obtained in this study and those published by Clark and Kidwell (1997). Species names are given followed by the total number of clones used to obtain a consensus sequence or, for the sequences obtained in our study, an arrow. This is a consensus tree of 1,000 equally parsimonious trees each requiring 2,193 steps. Of 449 characters, 34 were constant, 60 uninformative, and 355 parsimony informative. The consistency index is 0.40 and the retention index 0.815. Bootstrap analysis of 500 replicates was performed on the data. Clades names were given by Clark and Kidwell (1997) and are shown on the right.

to a data set consisting of 80 partial consensus sequences provided by Clark and Kidwell (1997) in order to establish their phylogenetic relationships.

Materials and Methods

Fly stocks

We used *Drosophila sturtevanti* collected in the Mexican town of Matlapa and *Drosophila prosaltans* collected in the town of Eldorado in the Brazilian state of Rio Grande do Sul both strains having been recently derived from natural populations and subsequently maintained in the laboratory, under standard conditions.

DNA Amplification and Sequencing

Polymerase chain reactions were carried out using a total volume of 50 µL containing 100 ng of genomic DNA, 1 mM $MgCl_2$, 400 µM of each deoxynucleotide, 0.5 µM of M-IR primer (5'CATAAGGTGGTCCCGTCG3', corresponding to nt. 14-31 and 2877-2894, within the TIRs - terminal inverted repeats - of the *D. melanogaster* canonical *P* element; Haring *et al.*, 1995) and 2 units of *Taq* polymerase in 1x polymerase buffer. Amplification consisted of 30 cycles of 45 s denaturation at 94 °C, 45 s of primer annealing at 57 °C and 1.5 min of primer extension at 72 °C. The first cycle was preceded by a step of 7 min at 94 °C for denaturation and the last cycle was followed by a final extension at 72 °C for 10 min. The PCR products were cloned into a TOPO TA cloning vector (Invitrogen) and both strands of a randomly-chosen single clone were sequenced for each species. The primers $Stu1_{369}$ (5'GTTCCGTATCG AGAC CCGA C3'), $Stu2_{2402}$ (5'AATGACGAAGACTC GTCGC G3'), $Stu5_{1091}$ (5' GGAAGCAACCAGTTTTCT TT3') and $Stu6_{1598}$ (5'CACATCAAACCAATCATTTA3') were designed based on the clone sequences and used to complete the sequencing of the element.

Sequence alignment and phylogenetic analysis

For alignment we used a set (Clark and Kidwell, 1997) of 80 partial (429 bp) *P*-element consensus sequences mapped to positions 1328-1757 in open reading frame 2 (ORF 2) of the canonical *P* element and the following full-length *P* element nucleotide sequences obtained from the literature: *D. melanogaster* p25.1 (O'Hare and Rubin, 1983); *Drosophila nebulosa* N10 (Lansman *et al.*, 1987); *Drosophila willistoni* (Daniels *et al.*, 1990); *Drosophila guanche, Drosophila bifasciata* and *D. bifasciata* jbifM3 (Hagemann *et al.*, 1992); *Drosophila ambigua* T-type (Hagemann *et al.*, 1998); *Drosophila helvetica* M-type (Haring *et al.*, 2000); *D. mediopunctata* (Loreto *et al.*, 2001); *Scaptomyza pallida* (Simonelig and Anxolabéhère, 1991) and *M. domestica* (Lee *et al.*, 1999). The *P* element DNA sequences obtained by us in this study have been deposited in the GenBank under accession number AF530052 for *D. sturtevanti* and AF530053 for *D. prosaltans*.

Alignments of sequences were done by the CLUSTAL W program (Thompson *et al.*, 1994) and the phylogenetic relationships between the *P* sequences constructed using the maximum parsimony method as implemented in PAUP program version 4.0b10 (Swofford, 1997). Branch support was calculated using bootstrap analysis with 500 replicates. The distance matrix was constructed according to the Kimura two-parameter model of nucleotide substitution (Kimura, 1980).

Results and Discussion

Sequence features

The M-IR primer corresponds to positions 14-31 and 2877-2894 of the canonical *P* element, and so cannot amplify the first and last 13 bp of a complete *P* element. The *D. sturtevanti P* element is 2829 bp and the *D. prosaltans* 2828 bp, if the TIRs of these elements are complete they should be 2854 bp for *D. sturtevanti* and 2855 bp for *D. prosaltans*, which is about 50 bp less than the *D. melanogaster* canonical *P* element. The alignment of the two sequences against the *D. melanogaster* sequence showed that the *D. sturtevanti* and *D. prosaltans P* elements are similar in structure and sequence to each other (88%) but strongly divergent from the *D. melanogaster* canonical *P* element (31% different). Table 1 shows the main differences between the alignments from which it can be seen that, in general, *D. sturtevanti* has the same deletions and insertions as the *D. prosaltans*.

Even though the TIRs were not completely sequenced, PCR amplification with primers specific to the TIR regions indicates that at least the second half of the TIRs are present and well conserved both in *D. sturtevanti* and *D. prosaltans*. However, the transposase binding sites, located at positions 48-68 and 2855-2871 in *D. melanogaster* (Kaufman *et al.*, 1989)*,* the TATA box and the 11-bp subterminal inverted repeats are not well conserved. In all four exons the translational reading frame is interrupted by stop codons and frameshift mutations, suggesting that these sequences do not encode a functional protein. Indeed, leucine-zipper and helix-turn-helix motifs were not detected, supporting the suggestion that these sequences might be non-autonomous in the genome.

A nucleotide differentiation and genetic difference matrix based on Kimura's two-parameter method was calculated for the full-length *P* element sequences from the literature and the two sequences described here (Table 2) and it was found that the *D. sturtevanti* and *D. prosaltans* sequences present an overall divergence of 31% as compared to the canonical sequences described in *D. melanogaster, D. willistoni, D. mediopunctata* and *D. nebulosa*.

Phylogenetic analysis

Phylogenetic analyses of *P* elements in the subgenus *Sophophora* (Clark *et al.*, 1995, 1998; Clark and Kidwell, 1997; Silva and Kidwell, 2000) indicate the existence of multiple *P* element subfamilies in lineages of single species that apparently must have entered the genome at different times during the past (Lee *et al.*, 1999; Haring *et al.*, 2000). The aim of our phylogenetic analysis was to determine if the *D. sturtevanti* and *D. prosaltans P* element sequences belonged to some of the well-characterized *P* element subfamilies. Figure 1 summarizes the results of our phylogenetic analysis of *P* element sequences using parsimony, from which it can be seen that in 100% of bootstrap replicates the *D. sturtevanti* and *D. prosaltans* sequences clustered in Clark and Kidwell's (1997) F clade, which contains *P* element sequences of some other *saltans* group species as well as some *willistoni* group species.

Based on an internal portion of the *P* element exon 2, Clark *et al.* (1995) and Clark and Kidwell (1997) placed the *P* elements of *saltans* group in four different clades or subfamilies (A, E, F and O) and found that the divergence within the other three *saltans-willistoni* clades, excluding the canonical *P* element clade, ranges from 17% to 30% up to 46%. Our sequences belong to Clark and Kidwell's (1997) clade F and have a nucleotide divergence varying from 6% to 17% within this clade. Clark and Kidwell thought that the F subfamily may be under represented in the *saltans* group because they had sampled only a few sequences, although they did not discard the hypothesis that this low frequency could be due to the use of PCR primers for sampling. However, our data suggest that the F clade *P* element subfamily might be more widely distributed in the *saltans* group than previously believed because this *P* element was detected in two different species in spite of the fact that canonical sequences (Castro and Carareto, 2004) also existed in these genomes.

Table 1 - Differences observed between *Drosophila sturtevanti* and *D. prosaltans P* element sequences and that of the canonical *P* element of *D. melanogaster.*

Characteristics		*D. sturtevanti*	*D. prosaltans*
Sequence length		2829 bp	2828 bp
Total of substitutions		866	971
Total of deletions		113	99
Total of insertions		58	35
Main indels	E0	+2; +2; -6; +3; -1; -2; -22; -2; -6; -5	+2; +2; -6; +4; -1; -2; -2; -6; -5
	E1	+2; -1; +1; +1; -1	+2; -1; +2; -1; +2; -1
	E2	-13	
	E3	-2; -1; +3; -1; -2; -3; +1; +2; -1; +2; +13; +1; +4; -3; -1; -5; -3; +1; -10; -1	-2; -1; +3; -1; -2; -3; +1; +2; -2; +1; +4; -1; -3; -1; -7; -6; -1; -3; -3; +1; -10; -1
	I1	+3; +5; +1; +5; -1	+3; +5; +1; -1
	I2	+5	+2
	I3	+1; -4; -1; -1; -10; -2	-2; -4; -1; -1; -10; -2

Notes: E0= exon 0; E1= exon 1; E2= exon 2; E3= exon 3; I1= intron 1; I2= intron 2 e I3= intron 3. The size and nature of the insertion (+) or deletion (-) are denoted separately when the indels are noncontiguous.

Table 2 - Proportion of differences in complete *P* element nucleotide sequences (above) and genetic distances calculated using Kimura's two-parameter method (below) among full-length *P* elements described here and those from other *Drosophila* species, *Scaptomyza pallida* and *Musca domestica*.

	1	2	3	4	5	6	7	8	9	**10**	**11**	12	13
1 *D. melanogaster*		0.00	0.03	0.03	0.29	0.19	0.21	0.32	0.37	**0.31**	**0.31**	0.20	0.48
2 *D. willistoni*	0.00		0.03	0.03	0.29	0.19	0.21	0.32	0.37	**0.31**	**0.31**	0.20	0.48
3 *D. nebulosa*	0.03	0.03		0.04	0.30	0.20	0.21	0.33	0.38	**0.31**	**0.31**	0.21	0.49
4 *D. mediopunctata*	0.03	0.03	0.04		0.30	0.20	0.21	0.33	0.37	**0.31**	**0.31**	0.21	0.48
5 *D. bifasciata*	0.37	0.37	0.38	0.38		0.28	0.28	0.33	0.38	**0.32**	**0.33**	0.29	0.51
6 *D. bifasciata* M	0.23	0.23	0.24	0.23	0.35		0.06	0.32	0.35	**0.28**	**0.29**	0.07	0.47
7 *D. helvetica*	0.24	0.24	0.25	0.25	0.36	0.07		0.32	0.35	**0.30**	**0.30**	0.09	0.47
8 *D. ambígua*	0.42	0.42	0.44	0.43	0.44	0.41	0.42		0.35	**0.34**	**0.34**	0.33	0.51
9 *D. guanche*	0.51	0.51	0.53	0.50	0.52	0.48	0.48	0.48		**0.38**	**0.39**	0.29	0.54
10 *D. prosaltans*	**0.40**	**0.40**	**0.41**	**0.41**	**0.42**	**0.37**	**0.39**	**0.45**	**0.54**		**0.12**	0.30	0.51
11 *D. sturtevanti*	**0.40**	**0.40**	**0.41**	**0.40**	**0.43**	**0.36**	**0.39**	**0.46**	**0.56**	0.39		**0.29**	**0.51**
12 *S. pallida*	0.24	0.24	0.25	0.25	0.36	0.07	0.10	0.43	0.50	**0.39**	**0.39**		**0.48**
13 *M. domestica*	0.77	0.77	0.78	0.76	0.85	0.75	0.75	0.84	0.96	**0.86**	**0.86**	0.76	

Acknowledgments

We are indebted to Drs. M.G. Kidwell and A.J. Holyoake who kindly provided advice and laboratory facilities at the University of Arizona and for their critical comments on earlier versions of the manuscript. We also thank J.C. Silva (University of Arizona, Tucson, USA) and V.L.S. Valente (Universidade Federal do Rio Grande do Sul, Departamento de Genética, Porto Alegre-RS, Brazil) for supplying the drosophilid strains used in this study and J.B. Clark and M.G. Kidwell for the sequences data set. This work was supported by the Brazilian agencies FAPESP (Grants n. 95/7192-2, 98/08734-1 and 00/11313-0 and a fellowship to J.P.C.) and CNPq.

References

Anxolabéhère D and Périquet G (1987) *P*-homologous sequences in Diptera are not restricted to the Drosophilidae family. Genet Iber 39:211-222.

Castro JP and Carareto CMA (2004). Canonical *P* elements are transcriptionally active in the *saltans* group of *Drosophila*. J Mol Evol 59:31-40.

Clark JB and Kidwell MG (1997) A phylogenetic perspective on *P* transposable element evolution in *Drosophila.* Proc Natl Acad Sci USA 94:11428-11433.

Clark JB, Altheide TK, Schlosser MJ and Kidwell MG (1995) Molecular evolution of *P* transposable elements in the genus *Drosophila*. I. The *saltans* and *willistoni* species group. Mol Biol Evol 12:902-913.

Clark JB, Kim P and Kidwell MG (1998) Molecular evolution of *P* transposable elements in the genus *Drosophila*. III. The *melanogaster* species group. Mol Biol Evol 15:746-755.

Daniels SB, Peterson KR, Strausbaugh LD, Kidwell MG and Chovnick A (1990) Evidence for horizontal transmission of the *P* transposable element between *Drosophila* species. Genetics 124:339-355.

Engels WR (1989) *P* elements in *Drosophila*. In: Berg D and Howe M (eds) Mobile DNA. American Society for Microbiology, Washington, DC, pp 437-484.

Hagemann S and Pinsker W (2001) *Drosophila P* transposons in the human genome? Mol Biol Evol 18:1979-1982.

Hagemann S, Miller WJ and Pinsker W (1992) Identification of a complete *P*-element in the genome of *Drosophila bifasciata.* Nucl Acids Res 20:409-413.

Hagemann S, Miller WJ and Pinsker W (1994) Two distinct *P* element families subfamilies in the genome of *Drosophila bifasciata.* Mol Gen Genet 244:168-175.

Hagemann S, Miller WJ and Pinsker W (1996) A new *P* element subfamily from *Drosophila tristis, D. ambigua* and *D. obscura.* Genome 39:978-985.

Hagemann S, Haring E and Pinsker W (1998) Horizontal transmission versus vertical inheritance of *P* elements in *Drosophila* and *Scaptomyza*: Has the M-type subfamily spread from East Asia? J Zool Syst Evol Res 36:75-83.

Haring E, Hagemann S and Pinsker W (1995) Different evolutionary behavior of *P* element subfamilies: M-type and O-type elements in *Drosophila bifasciata* and *D. imaii.* Gene 163:197-202.

Haring E, Hagemann S and Pinsker W (2000) Ancient and recent horizontal invasions of drosophilids by *P* elements. J Mol Evol 51:577-586.

Kapitonov VV and Jurka J (2003) Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc Natl Acad Sci USA 100:6569-6574.

Kaufman PD, Doll RF and Rio DC (1989) *Drosophila P* element transposase recognizes internal *P* element DNA sequences. Cell 59:359-371.

Kidwell MG, Kidwell JF and Sved JA (1977) Hybrid dysgenesis in *Drosophila melanogaster*: A syndrome of aberrant traits including mutation, sterility and male recombination. Genetics 86:813-833.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative analysis of nucleotide sequences. J Mol Evol 16:111-120.

Lansman RA, Shade RO, Grigliatti TA and Brock HW (1987) Evolution of *P* transposable elements: Sequences of *Drosophila nebulosa P* elements. Proc Natl Acad Sci USA 84:6491-6495.

Lee SH, Clark JB and Kidwell MG (1999) A *P* element-homologous sequence in the house fly, *Musca domestica.* Insect Mol Biol 8:491-500.

Loreto ELS, Valente VLS, Zaha A, Silva JC and Kidwell MG (2001) *Drosophila mediopunctata P* elements: A new example of horizontal transfer. The J Hered 92:375-381.

Nouad D, Quesneville H and Axolabéhère D (2003). Recurrent exon shuffling between distant *P*-element families. Mol Biol Evol 20:190-199.

O'Hare K and Rubin GM (1983) Structure of *P* transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell 34:25-35.

Perkins HD and Howells AJ (1992) Genomic sequences with homology to the *P* element of *Drosophila melanogaster* occur in the blowfly *Lucilia cuprina.* Proc Natl Acad Sci USA 89:10753-10757.

Reiss D, Quesneville H, Nouaud D, Andrieu O and Anaxolabéhère D (2003) *Hoppel*, a *P*-like element without introns: A *P*-element ancestral structure or a retrotranscription derivative? Mol Biol Evol 20:869-879.

Rio DC, Laski FA and Rubin GM (1986) Identification and immunochemical analysis of biologically active *Drosophila P* element transposase. Cell 44:21-32.

Robertson HM and Engels WR (1989) Modified P elements that mimic P cytotype in *Drosophila melanogaster*. Genetics 123:815-824.

Sarkar A, Sengupta R, Krzywinski J, Wang X, Roth C and Collins FH (2003) *P* elements are found in the genomes of nematoceran insects of the genus *Anopheles*. Insect Biochemistry and Molecular Biology 33:381-387.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406-425.

Silva JC and Kidwell MG (2000) Horizontal transfer and selection in the evolution of *P* elements. Mol Biol Evol 17:1542-1557.

Simonelig M and Anxolabéhère D (1991) A *P* element of *Scaptomyza pallida* is active in *Drosophila melanogaster*. Proc Natl Acad Sci USA 88:6102-6106.

Simonelig M and Anxolabéhère D (1994) *P* elements are old components of the *Scaptomyza pallida* genome. J Mol Evol 38:232-240.

Swofford D (1997) PAUP: Phylogenetic analysis using parsimony. Version 4.0b10. Smithsonian Institution, Washington, D.C.

Thompson JD, Higgins DG and Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple alignment through sequence weighting, positions-specific gap penalties and weigh matrix choice. Nucl Acids Res 22:4673-4680.

Associate Editor: Louis Bernard Klaczko