Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4018–4022, May 1992
Genetics

P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche*

WOLFGANG J. MILLER*, SYLVIA HAGEMANN†, ELEONORE REITER†, AND WILHELM PINSKER†‡

*Institut für Botanik, Abteilung Cytologie und Genetik, Rennweg 14, A-1030 Vienna, Austria; and †Institut für Allgemeine Biologie, Abteilung Genetik, Währingerstrasse 17, A-1090 Vienna, Austria

Communicated by Bruce Wallace, December 23, 1991

ABSTRACT **In *Drosophila guanche*, *P*-homologous sequences were found to be located in a tandem repetitive array (copy number: 20–50) at a single genomic site. The cytological position on the polytene chromosomes was determined by *in situ* hybridization (chromosome O: 85C). Sequencing of one complete repeat unit (3.25 kilobases) revealed high sequence similarity between the central coding region comprising exons 0 to 2 and the corresponding section of the *Drosophila melanogaster* *P* element. The rest of the sequence has diverged considerably. Exon 3 has no coding function and the inverted repeats have disappeared. The *P* homologues of *D. guanche* apparently have lost their mobility but have retained the coding capacity for a protein similar to the 66-kDa *P*-element repressor of *D. melanogaster*. Divergence between different repeat units indicates early amplification of the sequence at this particular genomic site. The presence of a common *P*-element site at 85C in *Drosophila subobscura*, *Drosophila madeirensis*, and *D. guanche* suggests that clustering of the sequence at this location took place before the phylogenetic radiation of the three species.**

The *P* elements of *Drosophila melanogaster* are perhaps the best characterized eukaryotic transposons (for review, see refs. 1 and 2). For the evolutionary biologist the *P*-element family provides an excellent system for studies of the evolutionary dynamics of mobile DNA (3–5). In general the autonomous propagation of transposable elements in a genome can have deleterious side effects for the host, such as high mutability, chromosomal rearrangements, and sterility (6). Consequently, the following mechanisms favoring the elimination of these genomic parasites may counterbalance their proliferation (7): (*i*) random processes (e.g., deletions), (*ii*) regulating mechanisms caused by the element itself, or (*iii*) active defense mechanisms developed by the host genome. Thus, in its most extreme form, the evolution of a mobile sequence can be seen as a succession of replicative bursts, dispersion, and finally elimination, a pattern that resembles the evolutionary behavior of organisms more than the evolutionary behavior of stationary DNA sequences. Although this simplified model is certainly not the only way transposons could evolve, it may apply to *P* elements and related systems.

In *D. melanogaster*, transposition activity of *P* elements is regulated in a tissue-specific manner (for review, see ref. 8). Tissue specificity is achieved by differential splicing of the full-length transcript (9). The third intron is removed only in the germ line, giving rise to an 87-kDa protein that functions as a transposase necessary for genomic mobility. In somatic cells, the third intron remains and a shorter 66-kDa protein is produced, acting as a repressor of transposition (10). Therefore, propagation of *P* elements is confined to the germ line and the damage to the host organism resulting from *P*-element-induced mutations is minimized.

In germ-line cells, the copy number of *P* elements is genetically regulated by a complex mixture of chromosomal and cytoplasmic factors (11). At least two mechanisms can be distinguished: the maternally transmitted cytotype (12) and a chromosomally inherited control system (13). In the P cytotype, the cytoplasmic state of P-strain individuals, transposition is repressed in somatic cells and in the germ line. In general the P cytotype develops gradually as a result of the increasing number of *P*-element copies. M-strains, on the other hand, have genomes devoid of *P* elements. In the M cytotype, which is found in eggs derived from M-strain females, transposition of paternally introduced *P* elements is not inhibited in the germ line. An intermediate repression potential is found in M′- and Q-strains (14). The chromosomally inherited system of transpositional regulation is thought to be based on a particular class of modified *P* elements, the *KP* element (13, 15), but other factors may be also involved (11). The molecular mechanisms underlying transposition control in the germ line are not completely elucidated, but defective *P*-element copies are supposed to play an important role in the repression of transpositional activity. About two-thirds of the *P*-element copies present in a true P-strain genome carry internal deletions possibly generated by imprecise double-strand gap repair after transposition events (16) and are capable of passive transposition only. The defective *P*-element copies that could influence the regulation of transpositional activity in the germ line fall into at least three distinct functional classes (8): (*i*) copies lacking open reading frame (ORF) 3 sequences, thus producing the 66-kDa *P*-element repressor protein exclusively (17, 18); (*ii*) small nonautonomous *P* elements with large internal deletions that could dilute transposase by titration, thereby reducing the frequency of transposition (19); (*iii*) *P* elements coding for truncated proteins that interfere with the dimerization of the transposase (8). The potential of a defective *P* element to suppress transposition depends not only on the sequence but also on the genomic position (18, 20, 21). Recent results show that the P cytotype can be elicited by only two *P*-element copies (22).

In natural populations of *D. melanogaster*, there is a tendency to accumulate particular defective *P*-element copies preventing transpositional activity. Two cases have been described in detail. Nitasaka *et al.* (23) found a defective *P* element responsible for the P cytotype in a strain from a Japanese population. Due to a 58-base-pair (bp) deletion near the end of ORF2, this deficient *P* element codes for a truncated protein with repressor function. Another deletion derivative, the *KP* element (13), is present in many copies in the genome of some strains. *KP* elements are thought to be involved in the suppression of transpositional activity (15).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M81221).

Defective *P* elements also have been detected in the genomes of *Drosophila nebulosa* (24) and *Drosophila bifasciata* (25).

In the present study *P*-homologous sequences were isolated and sequenced from a genomic library of *Drosophila guanche*, a member of the *Drosophila obscura* group. Anxolabéhère *et al.* (26) report that *P*-element sequences are widespread in this species group. The most intense hybridization signals in squash blots probed with the *D. melanogaster P* element were obtained with *D. guanche*, indicating either a high copy number or close similarity in sequence. *D. guanche* is a species endemic to the island of Tenerife (Canary Islands) and is considered to be a primitive representative of the group (27). The closest relatives in the *D. obscura* group are the species *Drosophila subobscura* and *Drosophila madeirensis* (28–30). Our investigation provides sequence data to elucidate the complicated evolutionary history of the *P*-element family in the genus *Drosophila*. Furthermore, the present status of the *D. guanche P* element may contribute to the understanding of the evolutionary life cycle of transposable elements in the gene pool of a species.

MATERIAL AND METHODS

Fly Stocks. Wild strains of *D. guanche* (Tenerife 1985), *D. madeirensis* (Madeira 1987), and *D. subobscura* (Sweden 1977) were derived from mass cultures of the original population samples collected in the years indicated. The two cytological marker strains of *D. subobscura* (O_{3+4+6} and O_{3+4+2}) were established from inbred lines maintained at the University of Tübingen (Federal Republic of Germany).

Cloning Procedures. High molecular weight DNA for the genomic library of *D. guanche* was prepared according to Hagemann *et al.* (25). The genomic DNA was partially digested with *Sau*3A and ligated into phage λEMBL3 (31). Screening was done by the methods of Lansman *et al.* (24), using the probe pπ25.1, which contains a complete *P* element of *D. melanogaster* (32). A second *P*-element probe used for control experiments was bif1, a *P*-homologous sequence isolated from the genome of *D. bifasciata* (25), which consists only of 5′ and 3′ noncoding sequences. Procedures for restriction enzyme digestion, agarose gel electrophoresis, gel blot analysis, nick translation of probes, and filter hybridization followed standard methods (33).

DNA Sequencing. The *P*-homologous DNA fragments selected for sequencing were subcloned into the vector pBluescript I KS(+). Sequencing was accomplished by the dideoxynucleotide chain-termination technique (34) using T7 DNA polymerase (Pharmacia).

***In Situ* Hybridization.** Larval salivary gland chromosomes were prepared and hybridized to biotin-labeled probes according to Engels *et al.* (35). Detection was carried out as described in the Detek instruction manual (Enzo Biochem).

RESULTS

Genomic Organization. According to preliminary squash-blot and dot-blot assays (data not shown), the genome of *D. guanche* was expected to harbor multiple copies, 20–50, of *P*-homologous sequences. However, gel-blot analysis of genomic DNA digested with *Bgl* II and probed with the *D. melanogaster P* element pπ25.1 revealed only a single *P*-homologous fragment of 3.25 kilobases (kb). From the intensity of the hybridization signal, it was clear that this 3.25-kb fragment could not be derived from a single *P*-element copy in the *D. guanche* genome. Thus the observed pattern could be explained in the following two ways. (*i*) The *D. guanche P* element is somewhat longer than that of *D. melanogaster* and contains two restriction sites for *Bgl* II. The genomic distribution could be either dispersed or tandem repetitive. (*ii*) The *P* element of *D. guanche* has only one *Bgl* II site and multiple copies are tandemly arranged in a cluster.

An answer to this question is given by the analysis of *P*-homologous EMBL3 clones isolated from a nonamplified genomic library of *D. guanche*. Fifteen clones (EMBL3-Gp1 to EMBL3-Gp15) were digested with *Bam*HI, *Xho* I, or *Bgl* II. With each of the enzyme digests, the fragment pattern was found to be homogenous among all the clones tested. A single 3.25-kb fragment was observed in the *Xho* I and *Bgl* II digests. With *Bam*HI one large fragment of ≈23 kb was obtained, indicating that there are no restriction sites for this enzyme in the inserts of the 15 clones. From these results it can be concluded that the *P*-homologous sequences of *D. guanche* are tandemly repeated in the genome, one repeat unit being 3.25 kb long. This interpretation is supported by a restriction experiment carried out with the clone EMBL3-Gp1. After partial digestion with *Bgl* II, an additional fragment of 6.5 kb representing the dimer of the repeat unit can be detected (Fig. 1). Since rearrangements within phage clones are known to arise sometimes as artifacts of cloning, the experiment was repeated using genomic DNA. The additional multimers shown in Fig. 2 clearly prove that tandemly repeated *P* homologues actually occur in the genome.

Another line of evidence for the clustered array of *P* homologues in the *D. guanche* genome comes from *in situ* hybridization experiments. When the *Bgl* II fragment of the clone EMBL3-Gp1 was used as a probe, only a single hybridization signal was detected on the polytene chromosomes of *D. guanche*: this signal was localized at position 85C of chromosome O on the cytological map of Moltó *et al.* (36). Analogous experiments carried out with the closely related species *D. subobscura* and *D. madeirensis* revealed that both species possess *P*-homologous sequences at cytologically equivalent positions. In *D. subobscura* the site at 85C is the only one in the genome. The exact chromosomal position in *D. subobscura* was further confirmed by *in situ* hybridization experiments with inversion strains. Since the *P*-element site is located inside of inversion O_6 but outside of O_2, it must be situated between the proximal breakpoints of the two inver-

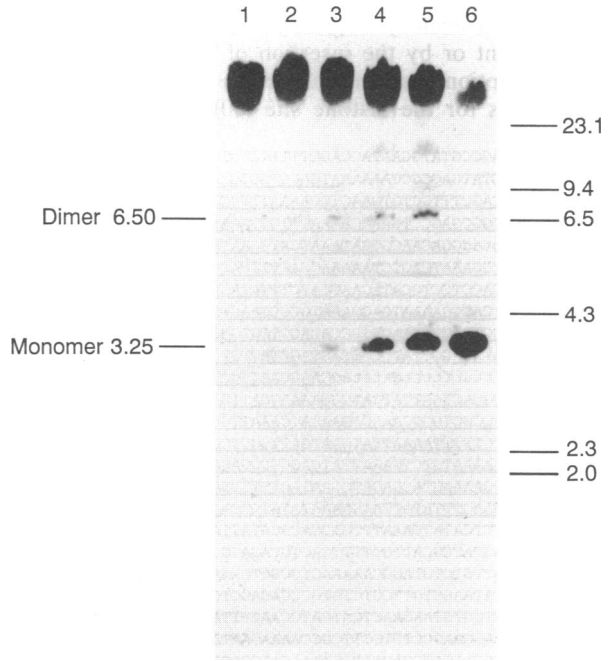

FIG. 1. Southern blot of EMBL3-Gp1 partially digested with *Bgl* II and probed with pπ25.1. Lanes: 1, undigested control; 2, *Bgl* II at 0.625 unit/μg; 3, 1.25 units/μg; 4, 2.5 units/μg; 5, 5 units/μg; 6, complete digestion. The positions of the monomer and the dimer of the 3.25-kb repeat unit are indicated. The additional weak bands probably represent fragments containing vector DNA.

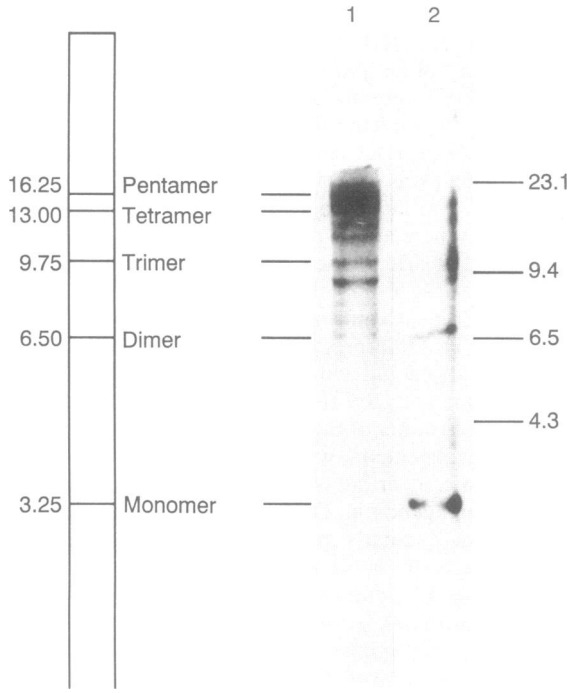

FIG. 2. Southern blot analysis of genomic DNA of *D. guanche* partially digested with *Bgl* II and probed with the *D. guanche P* homologue (insert of EMBL3-Gp1). Lane 1 contains a *Bgl* II (0.0625 unit/μg) digest; multimers of the 3.25-kb repeat unit (from the dimer to the pentamer) can be identified. Additional bands probably result from length variation in the cluster. Lane 2 contains a *Bgl* II (1 unit/μg) digest; the 3.25-kb monomer and the dimer can be seen.

sions indicated on the map of Kunze-Mühl and Müller (37). In *D. madeirensis*, which has the identical banding pattern in this chromosomal segment (29), an additional signal, although of weaker intensity, is observed at position 85A, close to the main *P*-element site. This second site may have been generated by a small cytologically undetectable chromosome rearrangement or by the insertion of a mobile element. A similar disruption of a gene cluster has been described in the same species for the histone site (30). The presence of a common *P*-element site at 85C in the three species suggests that clustering of the sequence at this location took place before the phylogenetic radiation of *D. subobscura, D. madeirensis*, and *D. guanche*.

Sequence Data. The base sequence of one complete repeat unit (G1) from the EMBL3-Gp1 insert is shown in Fig. 3. The *P*-homologous region was aligned to the sequence of the canonical *P* element of *D. melanogaster*, and the results of this interspecific sequence comparison are shown in Table 1. A good match between the two sequences is obtained only in the central segment extending from the 5′ end of exon 0 to the 3′ end of exon 2 (positions 146–1993 in Fig. 3). In the sequences upstream and downstream of this apparently conserved subregion, the similarity is significantly lower (χ^2 = 182.9; degree of freedom = 1; $P < 0.001$). As a consequence, the delimitation of the *P* homologue (2938 bp) in the repeat unit suggested by the computer alignment may be questionable. The 8-bp target sequences flanking each *P* element in *D. melanogaster* as direct repeats are completely missing, and the entire sequence between the conserved *P*-element homologues (including the intervening spacer segment) does not contain stretches of DNA resembling direct repeats. Therefore, the target site duplications have either disappeared by multiple mutations or were located outside of the boundaries of the amplified segment. Neither the 31-bp terminal inverted repeats (positions 1–30 and 2908–2938, according to the alignment) nor the 11-bp subterminal inverted repeats (positions 123–132 and 2790–2802) have retained their complementary structures, and there are no inverted repeats elsewhere in the sequence. It can be assumed that these *P* homologues have lost the capability for passive mobilization by transposase (provided from intact *P* elements) in the remote past. Since then the recognition sites for the transposase have been further eroded by random mutations.

In contrast to the noncoding flanking regions, the protein-coding sequences of exons 0 to 2 appear intact. If only nucleotide substitutions are considered, the average similarity to the *D. melanogaster P* element is 70%. Small deletions and insertions of up to 7 bp are found in all three exons. Since these mutations occur either as triplet changes or are compensated by mutations in the adjoining codons, the ORF is not disrupted. The splicing signals for introns 1 (positions 430–493) and 2 (positions 1141–1261) are conserved and

```
CAGAATTATAGCCGTATGGAATACCAGTTTTACATGTCTGGTCACACTCTGTCGCCAGCTTTTCAATTGCACATCTACATAGATTCGGAAAAGGTTGAACAAATTTTCTATTTTTTTCAA   120
TTGCTAACCGTATGAGCGCCAAAAAATGACGTGGTGCAGTGTATGCGGCAAAGTCGCTAACCATGTGAAATTGGTTCATGTCCCGGTATGTTTAGAGAAGAGAAAATTGTGGGAGCAAAT   240
TCTGGACTGCAGCTTTGCTGTGAACTCGAAAATTTGCGATTCACATTTCGATGCTTCACAGTGGAGGTCGCCACCAAAGGAAGGGCAGATATATAAAAGAAGGCGACTGAAAGCTGATGC   360
TGTTCCTCATGGCGAGCCTGAACCAAAATTTGTTAAATTAGGCTTCGCAAACTCGAGTACGCAAACAGAgtaagctaaacatgctcaaataaatatttacacaaatgccaatactaatttt   480
gtttgtaatgtagGGACAACGTGATAAATCATGCAATACGTGTGGAGAACGAGAGTCTGAGGAAACAGAATCGTCGAATGCAGAAAGAAATGCATTCTTTACGTCAGCAACTTGAGGACT   600
TTAAGGAATTGGAAATCAGTCTAAAAACCATTTTCACCGAGACCCAAATAAATATATTGAAGAGTGGAGGCAAGAGAGCCGTATTTAATGCAACAGACATGTCTGCTGCTATTTGCCTTC   720
ATACCGCGGGACCTCCTGCATACAATCATCTTTATAGAAAAGGATTTCCTTTGCCGAGTCGAGCAACATTATACAGATGGTTGGCAGATGTTAACATCAGTACAGGTACCCTTGATGTCG   840
TTATAGATCTCATGGAAAATGAGGAAATGCCCGAAGTGGATAAGCTATGCGTTTTGTCCTTTGATGAAATGAAGGTCGCTGCTGCATTCGAGCATGACAGCTCAGCGGATGTTGATTATG   960
AGCCAAGCACCTATGTTCAACTGGCCATAGCTCGTGGCTTGAATAAGTCTTGGGAGCAGCCAGTATTCTTCGATTTCAGTACTCTAATGGACGCGGACACTCTGCATTCAATAATAAACA  1080
AACTGCACAAAAGAGGCTATCCCGTAGTAGCTATTGTCTCTGATTTGGGCGCTGGAAATCAAACATTATGGACAGAGCTTGGCATATCGGAGAgtaagtttcttatacgagttaaaatcc  1200
gaaatttaatttatttttgttttagCAAGGAATTGGTTTACCCATCCAGCGGATGAAGATTTGAAAATATTCGTTTTCTCGGATACGCCAACTCTGATAAAGTTGGTCCGTGACCAGTA  1320
TGTTGACTCCGGACTAATCATCAACGGAAAACGATTGACTAAATCGACAGTCCAGCAGACAATTAGCCACTGTGCTAAGCCAGATGTGTCGATGTCATTCAATATTACCGATAATCACCT  1440
TAATATTGGACCGCTGGCCAAACAGAATATCAAATTGGCAACGCAACTGTTCTCCAATACAACCGGCAGCTTCATAAGACGATGCAATGCATTGGGTTATAACGTACAGAACGCATCCGA  1560
AACCGCTGATCTTTTTAAAATTATCAATGATTGGTTCGGGGTCTTCAACTCAAAGTCTTCAACATCAAATTCCATCGAACCGACGCAGCCTTATGGCAAACAAATCGAAATCCAACGAGG  1680
CATTTTGGCCAAAATGTCTGAGATTATGAGCTCGGAGATACTTGGAGTGGGAGCCCACAGCCTGCCCTTCCAAAAAGGCATTTTAGTCAACAATGCATCCCTTGAAGGTTTGTATTGCTA  1800
TCTATCGTCGGAAAAGTACGAGATCGAATACATTTTTACAAGTCGTCTCTCCCAAGACATTGTTGAGAACTTTTTTATGCCCATGCGGCCTAAAGGTGAACAGTTTGAGCATCCGACTCC  1920
ACTTCAATTTAAGTTTATGTTAAGAAAATATATATCAGGTATGACAAAATTAAATAAACCAATTGATAAATAATTGATTTGAACAATACAAATTGTGACATAGAAATAATTGTTTTAAGT  2040
TATTTTATATTTCATATTAAATTTGAGGACTCATATTATCAGAGGAATGATATGGTTGTATATCTAGGTTCCTGCTCCATGGAGCTGTATTAAAGAGCTGTAGCATGTGCTCAAGTCTAA  2160
GTTTGCTAAAGTACCGCATTGTTTTTCACTGCAGACGCTCGCGTATGATCGCTGTGCGGCTGAAATTTCCTTGGCCGCTGCTGGCAGTACCTATGGGATCAGGATGCGCACATTGGCCGC  2280
ATGGTATAAGTTGTCTGTAGGCAAAAAGTGCGCTGAAAGTCCTCGAAATAGAACTGCATTTCCAGAGTGATGAACAGTCGGATGAGGTCCAGGGCCAGCTTGGCATCGTGCCTGGCCAGC  2400
TTGGCCGCGTATAAGCTGTTCGTCTGTGCCCAGAGCTCCCTCAAAGACCTGATAGTTCTCGCAATTGGAGCGAAGGAGCTGCAGCAGGTCCAGCAGGAACTTGGGTCGCTGTTCGAGCGT  2520
GTTGGAAATGTCGTGTAAGAACTCATCATCCAAGTTTAGCGGTATTTTGAGATTAATTATACCCGTTACTCGATATACACGAATATACTAGGGTATATTGTATTTGTGGGGAAACAATGT  2640
AATTCACAGAAGGAAGCCTTTCCTTCGCCAAAAGAATATATATTTTGATCAGCATGAATAGCCGAGTCAATGTGTCAACCTGTCTATTTTTTTTTTAATTTCTGTACGCCCCCGCAAAACA  2760
GCGAAAATCTGCTATATCCACAATTTTAAAGATCGAGAAAACTAAAAATGCCATTCCGTAGGGAATGAATATAGCTATCAGATCAACGAATTGGGATCAGATTGGATCATTATTACAGCC  2880
AGAATGAAGAAATTAATTTGCAGTGGCTAAACCCTCCCCGCCCCGCAGCTTTATTATGtttgcacattcgaaaatgcgagctaaaagagcgtgttggcgtaagaagtgatgtacatgta  3000
gaagtagatgtaaatgacattgtagaaaacaattttaaatggcaaatgtaaaagttgaaagttgaaagaaaaaaaaaaaggtacacatatactaactgtgtaccgggtataaaagttgtg  3120
acgctaagaagcgtcttacacatccccttctcgttttaccttcggttcttatggaatcttcatccggactgttttgcaaagcaaacaaaacaaaacagctcacagctgttt  3232
```

FIG. 3. Sequence of one *P*-homologous repeat unit (G1) of *D. guanche*. The sequence consists of two subunits: the *P* homologue (roman type, positions 1–2938) and the intervening sequence (lowercase italic type, positions 2939–3232). Presumptive protein coding sequences of exons 0, 1, and 2 are underlined. Introns 1 and 2 are given in lowercase type. The four additional stop codons after exon 2 are also underlined.

Table 1. Interspecific sequence comparison

Region	Length, bp	Sequence differences, bp				% sim
		Id	Sub	Del	Ins	
5′ IR	31	15	15	1	0	48.4
5′ NCR	122	57	57	7	1	46.7
Exon 0	290	192	92	6	0	66.2
Intron 1	64	36	22	0	6	56.3
Exon 1	680	468	200	0	12	68.8
Intron 2	54	35	17	1	1	64.8
Exon 2	732	515	211	0	6	70.4
Intron 3	203	105	83	2	13	51.7
Exon 3	582	236	325	8	13	40.5
3′ NCR	174	69	101	0	4	39.7
3′ IR	31	12	19	0	0	38.7
Total	2963	1740	1142	25	56	58.7

Sequence differences between the *D. guanche P* homologue (G1) and the canonical *P* element of *D. melanogaster* in various subregions from the 5′ end to the 3′ end. Id, identical base-pair positions; Sub, substitutions; Del, deletions; Ins, insertions; % sim, percent sequence similarity including insertions and deletions; IR, terminal inverted repeats; NCR, noncoding regions.

located at positions homologous to those in the *D. melanogaster* sequence. These two introns resemble the corresponding regions in the *D. melanogaster P* element in length and in sequence. In exons 0 to 2, the sequence changes in *D. guanche* did not generate stop codons and the ORFs are preserved. Altogether the section comprising exons 0 to 2 is more like a functional coding sequence than a degenerated pseudogene derived from an immobilized transposon.

The region after the 3′ end of exon 2 has considerably diverged from the *D. melanogaster* sequence. Intron 3 has lost the 3′ splice signal. In the adjacent segment equivalent to exon 3 of the *D. melanogaster P* element, the original reading frame has shifted because of randomly distributed noncompensating deletions and insertions. Moreover, all possible reading frames are blocked by several stop codons and it is rather obvious that this sequence has no coding function.

To estimate the amount of variation among the *P* homologues in the cluster, two additional copies (G2 and G3) derived from a different EMBL3 clone (EMBL3-Gp7) were partially sequenced (260 bp of exon 0). Although the three *P* homologues belong to a tandemly clustered gene family, they have diverged considerably (8.6–11.9%), a finding that suggests early amplification of the sequence at this particular genomic site. This result is consistent with the interspecific comparison of the chromosomal location of *P* homologues in *D. guanche* and its close relatives. In spite of the substitutional divergence, the translational reading frame of G2 and G3 has been preserved.

According to these data the *P* homologues of *D. guanche* can be described as stationary transposon derivatives that are degenerated at their 5′ and 3′ termini but have some residual coding capacity. With respect to the internal coding region, the existence of differently structured *P* elements can be excluded on the basis of restriction experiments. Theoretically, however, it could be that internally deleted *P*-element derivatives with intact terminal repeat structures are present in the genome of *D. guanche* in addition to the tandemly repeated type. To test this possibility, Southern blots of genomic DNA of *D. guanche* were hybridized with the *P*-element probe bif1 of *D. bifasciata* (25), a sequence consisting exclusively of noncoding termini. Since no signals were detected in these control experiments, we assume that the genome of *D. guanche* does not contain sequences homologous to the 31-bp terminal inverted repeats found in all *P* elements studied so far.

DISCUSSION

The rather high degree of sequence divergence between the *P* homologues of *D. guanche* and the *D. melanogaster P* element indicates that the two sequences have evolved independently over a long period of time. This finding provides additional evidence for the theory postulating recent invasion of the *D. melanogaster* gene pool by a *P* element horizontally introduced from *Drosophila willistoni* (38). *D. melanogaster* is phylogenetically closer related to *D. guanche* than to *D. willistoni* (39). Nevertheless the *P* elements of *D. melanogaster* and *D. willistoni* are almost identical, whereas the *D. guanche P* homologue is considerably different from both. It can be assumed that the *P* homologue of *D. guanche* has remained continuously in the same gene pool since the common ancestor of the lineages leading to the *D. willistoni* group and the branch that later separated into the *D. melanogaster* group and the *D. obscura* group. Therefore, the *P* elements of *D. melanogaster* and *D. guanche* have been separated 53 million years, the estimated divergence time between *D. guanche* and *D. willistoni* (40). It would be interesting to compare the sequence of the *D. guanche P* homologue with related *P* sequences from other species of the *D. obscura* group to see whether a group-specific *P* element exists with shared substitutions characteristic of the lineage. Unfortunately, the sequences of the internally deleted *P* elements of *D. bifasciata* (25) do not overlap with the functionally conserved regions of the *D. guanche P* homologue. The degenerated flanking regions of the *D. guanche P* homologue, however, have diverged at a much faster rate, thus obscuring the phylogenetic relationships.

The conservation of exons 0 to 2 suggests that the *P* homologue of *D. guanche* may code for a truncated protein similar to the 66-kDa *P*-element repressor (18) of *D. melanogaster*. The amino acid sequence of the putative protein is given in Fig. 4. Provided that the third intron is not removed because of the mutation at the 3′ splice signal, the translational reading frame is terminated by a stop codon at positions 1991–1993. Four additional stop codons are found among the following 15 base triplets (Fig. 3). The protein predicted from the DNA sequence consists of 575 amino acids and has a molecular weight of 65,341. With 63.6% identical amino acid positions, the homology to the *P*-element repressor of *D. melanogaster* would be rather high. Since 33% of the amino acid replacements are conservative changes, one is inclined to postulate similar functions for the two proteins.

Although transcriptional activity of the *D. guanche P* homologue has not been proved, it is tempting to speculate why this terminally degenerated *P*-element derivative has accumulated in the genome of *D. guanche*. The most plausible assumption is that a *P*-element repressor coding sequence was amplified at position 85C as a molecular defense mechanism to suppress transposition. After switching of the cytotype from the M to the P state, all mobile *P*-element sequences (full-sized transposase producers and internally deleted passive elements) became immobilized and started to degenerate by random mutations. The process of degeneration may have gone so far that these elements are no longer detectable by hybridization techniques.

An alternative hypothesis for the origin of the *D. guanche P* homologue appears also reasonable. Accordingly, the *P* homologues of *D. guanche* may represent an ancient genomic sequence that predates mobile *P* elements. A copy of this gene was captured by primordial *P* elements and converted into a functional transposase gene by addition of a fourth exon. The formation of a transposon derived from a *P* element by a similar process has been described (41). The interspecific distribution pattern of *P* elements (26, 38, 42, 43) could be interpreted to mean that this transposon family was present in the progenitor of the genus *Drosophila*. On the

```
Dg MTWCSVCGKVANHVKLVHVPVCLEKRKLWEQILDCSFAVNSKICDSHFDASQWRSPPKEG  60
Dm  KY KF C AVTG   I   K AI       S G  LGE  Q   T  ND   KA A K

Dg QIYKRRRLKADAVPHG--EPEPKFVKLGFANSSTQTEDNVINHAIRVENESLRKQNRRMQ 120
Dm  TF     N     SKVI    EKI E YTSG     SCSL---FN-  K   EKI TLE

Dg KEMHSLRQQLEDFKELEISLKTIFTETQINILKSGGKRAVFNATDMSAAICLHTAGPPAY 180
Dm Y  RA E   RESQQ  E  RK   D   R   N  Q  T  SD I T         R

Dg NHLYRKGFPLPSRATLYRWLADVNISTGTLDVVIDLMNEEEMPEVDKLCVLSFDEMKVAA 240
Dm     K       T      S  D KR C EW     DSDGVDDA      A

Dg AFEHDSSADVDYEPSTYVQLAIARGLNKSWEQPVFFDFSTLMDADTLHSIINKLHKRGYP 300
Dm    Y     IV    D      V  K  K       N R  P   NN LR   RK  L

Dg VVAIVSDLGAGNQTLWTELGISETRNWFTHPADEDLKIFVFSDTPHLIKLVRDQYVDSGL 360
Dm          T   K         SKT  S    DH                  NH

Dg IINGKRLTKSTVQQTISHCAKPDVSMSFNITDNHLNIGPLAKQNIKLATQLFSNTTGSFI 420
Dm T    K   K I EALHL N S L IL K NE  I VRS    KV          A S

Dg RRCNALGYNVQNASETADLFKIINDWFGVFNSKSSTSNSIEPTQPYGKQIEIQRGILAKM 480
Dm    YS   DIE  T    F  LM    DI    L    C  CS      LD  ND  NR

Dg SEIMSSEILGVGAHSLPFQKGILVNNASLEGLYCYLSEKYEIEYIFTSRLSQDIVENFFM 540
Dm      RTG  DKPKR-       I      D   K  Q NFSMQ  L    N     H  G

Dg PMRPKGEQFEHPTPLQFKFMLRKYISGMTKL----NKPIDKX                   582
Dm S  SR G  D        YR     I   N KECV  NVIP
```

FIG. 4. Amino acid alignment between the *D. melanogaster* repressor protein sequence (Dm) and the presumed translation product of the *D. guanche P* homologue (Dg). Conservative replacements are shown in uppercase type; nonconservative replacements are underlined. Dashes indicate amino acid deletions.

basis of the "ancient gene" hypothesis, it is, however, difficult to explain why *P*-homologous sequences are completely absent in many species (38, 44, 45). The hypothesis would require the repeated loss of both types of *P*-homologous sequences (mobile *P* elements and the stationary source gene) in separate phylogenetic lineages. Although this possibility cannot be ruled out, it seems less convincing than the "terminal-degeneration" model outlined above.

Parallel investigations in the three related species *D. guanche, D. madeirensis*, and *D. subobscura* suggest that the amplification of the repressor-producing *P* homologue at the cytological site 85C occurred before the phylogenetic radiation of this species cluster. Therefore, any mobile *P* elements should have been eliminated long ago. This, however, raises the question as to why a repressor-producing sequence is still conserved in the absence of mobile *P* elements. Consequently, the now useless *P* repressor genes should have degenerated at a rate similar to that of the immobilized *P* elements. To explain this strange phenomenon, one has to postulate some additional function for the protein that provides a selective advantage to the fly. A change of function in connection with tandem amplification has been described for a rudimentary *P* element in *D. melanogaster* (46). In that special case, the insertion of four *P* elements into the control region of the *Gpdh* gene was found to influence transcriptional activity. The *P* derivative apparently serves as an enhancer binding a regulatory host protein. For the *P* homologues of *D. guanche*, however, we have to postulate a functional change at the protein level. The protein encoded in the cluster possibly has acquired an assignment distinct from its former repressor function. Thus the *P* element of this species may have entered a further stage in the evolutionary life cycle of a transposon: molecular domestication, the transition of a former genomic parasite to a stationary gene beneficial to the host.

We are very much obliged to V. M. Cabrera (Tenerife) who provided the strains of *D. guanche* and *D. madeirensis*. We are grateful to D. Schweizer (Vienna) for the kind hospitality granted to our group in his laboratories throughout the project and for valuable advice. We greatly appreciate the comments on the manuscript by M. Breitenbach (Vienna) and D. Sperlich (Tübingen). The work was supported by a grant from the Fonds zur Förderung der wissenschaftlichen Forschung (Project P6575B).

1. Engels, W. R. (1983) *Annu. Rev. Genet.* **17,** 315–344.
2. Engels, W. R. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 437–484.
3. Kaplan, N., Darden, T. & Langley, C. H. (1985) *Genetics* **109,** 459–480.
4. Ajioka, J. W. & Hartl, D. L. (1989) in *Mobile DNA*, eds. Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington), pp. 939–958.
5. Charlesworth, B. & Langley, C. H. (1989) *Annu. Rev. Genet.* **23,** 251–287.
6. Kidwell, M. G., Kidwell, J. F. & Sved, J. A. (1977) *Genetics* **86,** 813–833.
7. Orgel, L. E. & Crick, F. H. C. (1980) *Nature (London)* **284,** 604–607.
8. Rio, D. C. (1990) *Annu. Rev. Genet.* **24,** 543–578.
9. Laski, F. A., Rio, D. C. & Rubin, G. M. (1986) *Cell* **44,** 7–19.
10. Rio, D. C., Laski, F. A. & Rubin, G. M. (1986) *Cell* **44,** 21–32.
11. Raymond, J. D., Ojala, T. A., White, J. & Simmons, M. J. (1991) *Genet. Res.* **57,** 227–234.
12. Engels, W. R. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 4011–4015.
13. Black, D. M., Jackson, M. S., Kidwell, M. G. & Dover, G. A. (1987) *EMBO J.* **6,** 4125–4135.
14. Kidwell, M. G. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 1655–1659.
15. Jackson, M. S., Black, D. M. & Dover, G. A. (1988) *Genetics* **120,** 1003–1013.
16. Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. & Sved, J. (1990) *Cell* **62,** 515–525.
17. Robertson, H. M. & Engels, W. R. (1989) *Genetics* **123,** 815–824.
18. Misra, S. & Rio, D. C. (1990) *Cell* **62,** 269–284.
19. Simmons, M. J. & Bucholz, L. M. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 8119–8123.
20. Preston, C. R. & Engels, W. R. (1989) *Prog. Nucleic Acid Res. Mol. Biol.* **36,** 71–85.
21. Simmons, M. J., Raymond, J. D., Rasmusson, K. E., Miller, L. M., McLarnon, C. F. & Zunt, J. R. (1990) *Genetics* **124,** 663–676.
22. Ronsseray, S., Lehmann, M. & Anxolabéhère, D. (1991) *Genetics* **129,** 501–512.
23. Nitasaka, E., Mukai, T. & Yamazaki, T. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 7605–7608.
24. Lansman, R. A., Shade, R. O., Grigliatti, T. A. & Brock, H. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 6491–6495.
25. Hagemann, S., Miller, W. J. & Pinsker, W. (1990) *J. Mol. Evol.* **31,** 478–484.
26. Anxolabéhère, D., Nouaud, D. & Périquet, G. (1985) *Génét. Sél. Evol.* **17,** 579–584.
27. Monclús, M. (1976) *Bol. R. Soc. Esp. Hist. Nat. (Secc. Biol.)* **74,** 197–213.
28. Cabrera, V. M., Gonzales, A. M., Larruga, J. M. & Gullón, A. (1983) *Evolution* **37,** 675–689.
29. Krimbas, C. B. & Loukas, M. (1984) *Heredity* **53,** 469–482.
30. Felger, I. & Pinsker, W. (1987) *Z. Zool. Syst. Evolutionsforsch.* **25,** 127–140.
31. Frischauf, A. M., Lehrach, H., Poustka, A. & Murray, N. (1983) *J. Mol. Biol.* **170,** 827–842.
32. O'Hare, K. & Rubin, G. M. (1983) *Cell* **34,** 25–35.
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY), 2nd Ed.
34. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
35. Engels, W. R., Preston, C. R., Thompson, P. & Eggleston, W. B. (1986) *Focus* **8,** 6–8.
36. Moltó, M. D., DeFrutos, R. & Martinez-Sebastián, M. J. (1987) *Genetica* **75,** 55–70.
37. Kunze-Mühl, E. & Müller, E. (1958) *Chromosoma* **9,** 559–570.
38. Daniels, S. B., Peterson, K. R., Strausbaugh, L. D., Kidwell, M. G. & Chovnick, A. (1990) *Genetics* **124,** 339–355.
39. Throckmorton, L. H. (1975) in *Handbook of Genetics*, ed. King, R. C. (Plenum, New York), Vol. 3, pp. 421–469.
40. Beverly, S. M. & Wilson, A. C. (1984) *J. Mol. Evol.* **21,** 1–13.
41. Tsubota, S. I. & Dang-Vu, H. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 693–697.
42. Lansman, R. A., Stacey, S. N., Grigliatti, T. A. & Brock, H. W. (1985) *Nature (London)* **318,** 561–563.
43. Simonelig, M. & Anxolabéhère, D. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 6102–6106.
44. Brookfield, J. F. Y., Montgomery, E. & Langley, C. H. (1984) *Nature (London)* **310,** 330–332.
45. Stacey, S. N., Lansman, R. A., Brock, H. W. & Grigliatti, T. A. (1986) *Mol. Biol. Evol.* **3,** 522–534.
46. Itoh, M., Iwabuchi, M., Yoshida, K. & Hori, S. H. (1989) *Biochem. Genet.* **27,** 699–718.