Heredity (2004) 92, 418–427

www.nature.com/hdy

Complex structural features of satellite DNA sequences in the genus *Pimelia* (Coleoptera: Tenebrionidae): random differential amplification from a common 'satellite DNA library'

J Pons[1,2], B Bruvo[3], E Petitpierre[1,4], M Plohl[3], D Ugarkovic[3] and C Juan[1]

[1]*Laboratori de Genètica, Departament de Biologia, Universitat de les Illes Balears, 07122 Palma de Mallorca, Balearic Islands, Spain;* [2]*Molecular Systematics Laboratory, Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK;* [3]*Department of Molecular Genetics, Ruder Boskovic Institute, Bijenicka 54, PO Box 1016, HR-Zagreb, Croatia;* [4]*Departament de Recursos Naturals, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), C/. Miquel Marqués, 21, 07190 Esporles, Balearic Islands, Spain*

The major satellites of the nine species of the subgenera *Pimelia* s. str. and *Amblyptera* characterised in this paper are composed of longer monomers (500 and 700 bp) than those described previously in 26 *Pimelia* s. str. taxa (357 bp, a sequence called PIM357). Sequence analysis reveals partial similarity among these satellites and with the PIM357 monomers. The discrepancy between the phylogeny obtained based on three mitochondrial and two nuclear markers and that deduced from satellite DNA (stDNA) sequences suggests that the different *Pimelia* satellites were already present in a common ancestor forming what has been called a 'satellite DNA library'. Thus, the satellite profiles in the living species result from a random amplification of sequences from that 'library' during diversification of the species. However, species-specific turnover in the sequences has occurred at different rates. They have included abrupt replacements, a gradual divergence and, in other cases, no apparent change in sequence composition over a considerable evolutionary time. The results also suggest a common evolutionary origin of all these *Pimelia* satellite sequences, involving several rearrangements. We propose that the repeat unit of about 500 bp has originated from the insertion of a DNA fragment of 141 bp into the PIM357 unit. The 705-bp repeats have originated from a 32-bp direct duplication and the insertion of a 141-bp fragment in inverted orientation relative to a basic structure of 533 bp.
Heredity (2004) **92,** 418–427, advance online publication, 3 March 2004; doi:10.1038/sj.hdy.6800436

Keywords: satellite DNA; *Pimelia*; Coleoptera; repeat unit rearrangements

Introduction

The species of the beetle genus *Pimelia* (Tenebrionidae, Coleoptera) are flightless, saprophagous insects occurring in xerophilic habitats in the southern regions of the Northern hemisphere. *Pimelia* s. str., one of the five recognised subgenera of the genus, is a polyphyletic lineage originating in central Asia and colonising Northeast Africa (Kwieton, 1977). From there, at about 6–12 Mya, a lineage radiated in the North Western region of Africa, colonising Morocco, the Iberian Peninsula, and the Canary Islands. The subgenus *Amblyptera* is thought to be derived from a lineage of the *Pimelia* s. str. Moroccan radiation (Kwieton, 1977).

From previous work, we know that a large number of congeneric *Pimelia* s. str. taxa possess a common major satellite DNA (stDNA) family, which constitutes a remarkably high percentage of their genomes (30–40%) and it is localised in the heterochromatic regions of the chromosomes (Pons *et al*, 1997, 2000a, b). The tandem repeats of this family (PIM357) are similar in sequence showing few insertions and deletions (consensus unit 357-bp long) and very similar nucleotide composition ($69 \pm 1.4\%$ A+T). PIM357 monomers show an intrinsic bending related to a fairly conserved periodical distribution of runs composed of three or more adenines (Barceló *et al*, 1997; Pons *et al*, 2002a). These repeats have evolved gradually as predicted by the molecular drive model (Dover, 2002): that is, sequences are accumulating single mutations and some of the new mutated sequences are homogenised and fixed replacing completely the old ones, in a continuous turnover, leading to species-specific or group-specific stDNAs. Such a model of evolution could explain as to why DNA divergence studies clearly cluster PIM357 satellites into three sequence groups, Iberian-Balearic, Moroccan, and Canarian, which are mostly in accordance with the species biogeography and mitochondrial phylogeny (Pons *et al*, 2002a).

The data presented in this paper concern nine species of the Moroccan lineage of subgenera *Pimelia* s. str. and *Amblyptera* having different major stDNA families with longer repeats than the canonical PIM357 unit (500–700 bp), but which still represent a large fraction of the genome (27–43%, Pons, 1999; this paper). The monomers deduced for these longer families reveal marked divergences from each other and also in

Correspondence: J Pons, Department of Entomology, The Natural History Museum, Cromwell Road, London SW7 5BD, UK.
E-mail: joap@nhm.ac.uk
Published online 3 March 2004

comparison to PIM357 sequences, with long insertions and deletions, but nevertheless show partial sequence similarities. The homogeneous length and nucleotide composition, and experimental evidence of gradual evolution within PIM357 sequences seems to indicate that they are a different stDNA family from the divergent and longer ones. For this reason, those longer sequences were not included in the analysis of PIM357 sequences (Pons *et al*, 2002a).

Dot blot analysis reveals that the monomer of the major stDNA family of *P. monticola* (PMON 533-bp long) is also found at low copy numbers in the genome of other *Pimelia* species having different major stDNAs (Bruvo *et al*, 2003). Six *Pimelia* species, from both subgenera, were screened using primers specifically designed to amplify PMON sequences by PCR. The phylogenetic analysis of the amplified sequences revealed the presence of two groups of sequences: PIM357 in Iberian-Balearic species plus *P. monticola*, and the other PIM357, in African species plus *P. fornicata*. These results from *Pimelia* were additional evidence of the 'library hypothesis' postulated by Fry and Salser (1977) and confirmed experimentally by Mestrovic *et al* (1998). This hypothesis suggests that major stDNAs found in a related group of living species were already present in the common ancestor of those species, forming a pool of stDNA sequences at low copy number ('a satellite DNA library'). During the speciation process, an occasional amplification of a stDNA repeat from that low copy number pool would drive this repeat to form major stDNAs in one species, while the other repeats from the pool would remain in all the others species at low copy number.

In the present paper, we characterise the new major stDNA families in nine *Pimelia* species, comparing them to the PIM357 stDNA family previously described (Pons *et al*, 2002a). The goal of this paper is to understand the mechanisms involved in the formation and persistence of these stDNA sequences. The phylogenetic analysis of the stDNA sequences very likely does not reflect the diversification of *Pimelia* species, due to their marked sequence divergences, and the possibility of random amplification from a pool of stDNA sequences. Therefore, an independent phylogeny, based on mitochondrial and nuclear markers, was obtained to give a broader picture of the evolution of stDNA sequences and distribution of particular stDNA families among species-groups.

Material and methods

Sampling and genomic DNA extraction

The individuals were collected at localities in the Iberian Peninsula and North Africa. Five species are classified as part of the subgenus *Pimelia* s. str.: *P. monticola* (Pico Veleta, Granada, Spain), *P. cordata* (Kraatg, Morocco), *P. echidna* (Kelaa des Sraghna, Morocco), and two were unclassified species from Nefta, Tunisia (*Pimelia sp 1* and *Pimelia sp 2*). Another four species are of the subgenus *Amblyptera*: *P. fornicata* (Zahara de los Atunes, Cádiz, Spain), *P. scabrosa* (Zahara de los Atunes, Cádiz, Spain), *P. rotundipennis* (Kelaa des Sraghna, Morocco), and *P. rugosa* (Tizi-N-Tichka, Morocco). DNA was isolated from adults by standard phenol extraction and ethanol precipitation procedures.

Isolation, cloning, and sequencing of satellite DNA

Digestions of genomic DNA with restriction enzymes were performed according to the instructions of the manufacturer (Roche), and the fragments separated by electrophoresis on 1.5% agarose gels. The DNA bands corresponding to putative monomeric sequences were cut from the agarose gel, purified with the Gene Clean Kit (Bio 101 Inc.), ligated into the *Sma* I site of plasmid pUC18 vector (Amersham) and used to transform *Escherichia coli* DH5α. Clones were screened using the β-galactosidase blue-white colour system. Positive clones were sequenced on both strands by the dideoxy sequencing method using the Dig *Taq* DNA Sequencing Kit for Cycle Sequencing (Boehringer Mannheim) and the semiautomatic sequencing system GATC 1500-System Direct Blotting Electrophoresis (Boehringer Mannheim). The sequenced stDNA repeat units have been deposited in EMBL databank under Accession Numbers AJ247374-AJ247404, AJ307969, and AJ307973.

Southern blot and estimation of stDNA percentage

For Southern analysis, 5 μg of genomic DNA from each species was digested with different restriction enzymes and blotted on nylon membranes. *P. monticola* monomers, from genomic DNA or clones, were purified from the gel, labelled with digoxigenin, and used as probes in a Southern hybridisation under high (85–90% similarity) and low (60% similarity) stringency conditions. Digoxigenin labelling of the probe, filter hybridisations, and detection of the hybridisation signals were performed as described in the manual for the DIG High Prime DNA Labelling and Detection Starter Kit I (Roche). The relative amount of stDNA was determined from digestion of genomic DNA (*Hae* III or *Rsa* I) electrophoresed on an agarose gel. The digitisation and densitometric measurements from the gel photographs were performed with the help of the Sun View program (Pharmacia).

Sequence analysis of monomers

Multiple alignment was performed using Clustal W v. 1.7 (Higgins *et al*, 1996). Details of the nucleotide composition, sequence divergences (measured as the proportion of nucleotide sites at which two sequences are different or *p* distance), and parsimony tree searches (1000 random replicates) were performed using PAUP* 4.05 (Swofford, 2002). Bremer support values were determined performing 200 random replicate searches with constraint files obtained with TreeRot.v2b (Sorenson, 1999). The nucleotide diversity was calculated using DnaSP v. 3.51 (Rozas and Rozas, 1997). Curvature analysis of monomers, searches for A or T $\geqslant$3 runs and mobility on nondenatured polyacrylamide gel electrophoresis, were performed as described in Barceló *et al* (1997).

PCR amplification and sequencing of mitochondrial and nuclear fragments

Mitochondrial fragments were obtained using primers and methods described elsewhere: a 376-bp fragment of cytochrome oxidase subunit I (COI) (Juan *et al*, 1995), a 358-bp segment of cytochrome *b* (Vogler and Welsh, 1997), and a fragment of about 510 bp from the large ribosomal (16S) subunit gene (Funk, 1999). A 328-bp

fragment of the nuclear Histone 3 gene was obtained as described elsewhere (Colgan *et al*, 1998). Finally, a partial region of about 650 bp of the rRNA 28S containing the variable domains D3–D6 was amplified under standard PCR conditions using the following primer combination: forward 5′-GGG ACC CGT CTT GAA ACA C-3′ and reverse 5′-TTA CAC ACT CCT TAG CGG AT-3′. Sequencing of both strands in each case was performed with the dideoxy sequencing method using the Big Dye™ Terminator Cycle Sequencing Kit and an ABI PRISM™ 3700 DNA Analyzer (Applied Biosystems). Sequences have been deposited in EMBL databank under Accession Numbers: COI (X97209–X97223, AJ248198–AJ248217, and Z71727), 28S (AJ565939–AJ565974), cytochrome *b* (AJ565975–AJ566009), Histone 3 (AJ566010–AJ566044), and 16S (AJ566045–AJ566080). The corresponding sequences from *Tenebrio molitor* (X88966, U65184, and AJ438153) and *Tribolium castaneum* (NC003081) were obtained from Gene Bank.

Phylogenetic analysis of mitochondrial and nuclear sequences

Our preferred approach for tree searches was the combined analysis with all characters equally weighted and gaps treated as fifth character under the parsimony criterion. Parsimony tree searches (1000 random replicates), and incongruence length difference test (ILD test, Farris *et al*, 1994; with 200 random replicates) to estimate the incongruence between the different markers were carried out in PAUP* v. 4.05 (Swofford, 2002). Bremer support and partitioned Bremer support (PBS) values were determined performing 200 random replicate searches with constraint files obtained with TreeRot.v2b (Sorenson, 1999). ML branch lengths based on the MP tree and model parameters calculated using MODELTEST (Posada and Crandall, 1998) were modified using the nonparametric rate smoothing (NPRS) algorithm as implemented in TreeEdit v. 1.0a9 (Rambaut and Charleston, 2002).

Results

Cloning and sequence analysis of satellite DNAs

Genomic DNAs of *P. fornicata*, *P. scabrosa*, *P. rotundipennis*, *P. rugosa*, *P. monticola*, *P. cordata*, *P. echidna*, and the two unclassified species *P. sp 1* and *P. sp 2* were digested with several restriction enzymes. *Hae* III and, in the case of *P. cordata*, *Rsa* I were selected for further studies. Seven out of the nine species checked showed conspicuous bands of about 500 bp. Moreover, *P. scabrosa* and *P. rotundipennis* had longer prominent bands of about 700 bp (Figure 1a). Some of the species, such as *P. rugosa*, *P. scabrosa*, *P. rotundipennis*, and *P. sp 2*, also revealed DNA bands twice as large as the monomers interpreted as uncut contiguous units. Densitometric quantification of the corresponding electrophoretic bands in the DNA digestions revealed that these stDNAs represent a high proportion of the genome in the species studied, ranging from 26.1 to 43.6% of the total DNA (Table 1). Given the DNA content of the species screened (Pons, 1999), the stDNA sequences obtained correspond to from 2.1 to 7.5×10^5 copies per haploid genome, depending on the species (Table 1).

Southern blots of genomic DNA digested with restriction enzymes were hybridised using monomers from the genomic DNA of *P. monticola* as a probe. The repeat units of *P. rugosa*, *P. scabrosa*, *P. cordata*, *P. echidna*, *Pimelia sp. 1*, and *Pimelia sp. 2*, and several species whose major stDNA is PIM357 displayed cross-hybridisation even at high stringency conditions (85–90% similarity, Figure 1). *P. fornicata* and *P. rotundipennis* also revealed cross-hybridisation (not shown) indicating that all the *Pimelia* stDNA sequences studied to date are related in sequence. We cloned the monomeric bands of each taxon and randomly sequenced between 3 and 8 units (see Table 1). The Southern mentioned above shows identical results when a cloned monomer of *P. monticola* was used as a probe. The sequenced repeats showed striking intraspecific similarities in length, nucleotide composition, and nucleotide diversity (Table 1).

At the interspecific level, the repeat units displayed differences in nucleotide sequences (Table 2 and Figure 2) except *P. rotundipennis* and *P. scabrosa*, which share almost identical repeat units (PROT and PSCA respectively). An alignment of the consensus monomeric sequences of the nine species described in this work with the PIM357 consensus sequence characterised previously, reveals the presence of four sequence segments conserved in the monomers of all species (segments I–IV, Figure 2). Interestingly, these conserved regions comprise almost the complete sequence of monomer PIM357. Moreover, all the repeat units, with the exception of PIM357, share a common central region of about 140 bp (Figure 2). The unrooted tree based on parsimony retrieved three different groups of sequences

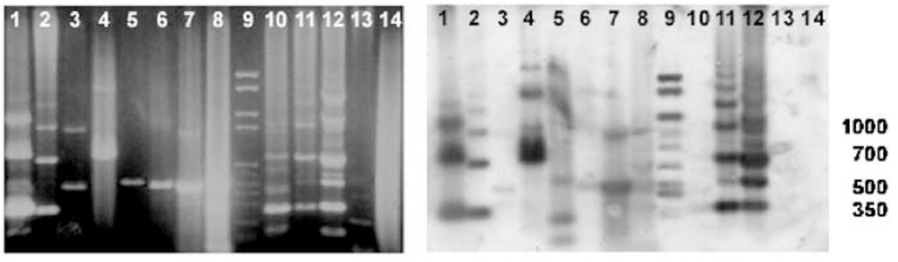

Figure 1 (**a**) Agarose gel electrophoresis of genomic DNA restricted with *Hae* III, except *P. lutaria*, *P. baetica*, *P. atlantis frigioides*, *P. sparsa serrimargo*, *P. laevigata validipes*, and *Tenebrio molitor*, which were cut with *Eco* RI: *Pimelia lutaria* (1), *P. baetica* (2), *P. rugosa* (3), *P. scabrosa* (4), *P. cordata* (5), *P. echidna* (6), *Pimelia sp. 2* (7) and *Pimelia sp. 1* (8), *P. atlantis frigioides* (10), *P. sparsa serrimargo* (11), *P. laevigata validipes* (12), *Tenebrio molitor* (13), and *Timarcha balearica* (14) as negative control. Lane 9 shows a DNA molecular standard (Marker VI Roche) with bands ranging from 2100 to 150 bp. (**b**) Southern blot of the gel after hybridisation using monomers from genomic DNA of *P. monticola* as a probe.

Table 1 List of specimens, species-clones code, restriction enzyme used to isolate the repeats, number of sequenced repeats, percentages of satellite DNA, copy number of the repeat units ($\times 10^3$), intraspecific consensus repeat unit length (bp), percentages of A+T, and the nucleotide diversity (π) of the repeat units

Species	*Code*	*Restriction enzyme*	*Repeats sequenced*	% stDNA	*Copy number*	*Length*	*% A+T*	π
P. echidna	*PECH*	*Hae* III	3	33.3±2.9	380	502	69.32	0.016
P. rugosa	*PRUG*	*Hae* III	3	32.8±3.1	360	502	64.44	0.027
P. sp 2	*P2NF*	*Hae* III	3	35.9±0.6	460	509	57.96	0.013
P. sp 1	*P1NF*	*Hae* III	3	31.2±2.1	250	512	60.16	0.025
P. cordata	*PCOR*	*Rsa* I	2	34.4±3.7	—	515	57.09	0.017
P. rotundipennis	*PROT*	*Hae* III	3	26.1±2.5	—	705	68.75	0.021
P. scabrosa	*PSCA*	*Hae* III	8	30.7±3.0	210	705	68.86	0.023
P. monticola	*PMON*	*Hae* III	5	28.1±3.1	170	533	71.18	0.039
P. fornicata	*PFOR*	*Hae* III	3	43.7±2.9	750	502	64.28	0.023
26 taxa	*PIM357*	*Eco* RI *Hae* III	156	27.1–43.6	450	357	69.00	0.02–0.2

Average, or maximum and minimum values, for 26 previously described taxa which repeat units are about 357 bp long (Pons *et al*, 2002a) is also included.

Table 2 Matrix of pairwise divergence values of stDNA consensus sequences based on alignment of Figure 2 (above diagonal), and pairwise divergences based on the two first more conserved segments only (I and II)

	PSCA	*PROT*	*PMON*	*357AF*	*357IB*	*357CI*	*PRUG*	*PFOR*	*PECH*	*PCOR6*	*P1NF*	*P2NF*
PRUG	0.318	0.318	0.270	0.327	0.285	0.295		0.020	0.173	0.361	0.354	0.308
PSCA		0.003	0.221	0.353	0.320	0.290	0.382	0.382	0.362	0.483	0.479	0.467
PROT	0.003		0.219	0.350	0.317	0.287	0.382	0.382	0.361	0.485	0.481	0.468
PMON	0.179	0.173		0.300	0.283	0.238	0.366	0.371	0.349	0.483	0.474	0.477
357AF	0.278	0.273	0.220		0.154	0.209	0.378	0.382	0.344	0.514	0.473	0.472
357IB	0.236	0.231	0.210	0.151		0.202	0.366	0.367	0.333	0.510	0.460	0.460
357CI	0.218	0.213	0.149	0.208	0.188		0.348	0.348	0.320	0.502	0.481	0.448
PFOR	0.312	0.312	0.275	0.327	0.279	0.289	0.016		0.181	0.363	0.360	0.323
PECH	0.286	0.279	0.254	0.284	0.274	0.271	0.190	0.201		0.363	0.376	0.329
PCOR6	0.421	0.426	0.416	0.461	0.428	0.421	0.376	0.366	0.344		0.440	0.415
P1NF	0.400	0.406	0.400	0.420	0.378	0.421	0.328	0.328	0.339	0.410		0.246
P2NF	0.396	0.401	0.390	0.429	0.382	0.403	0.324	0.335	0.351	0.406	0.151	

Species codes are as indicated in Table 1.

(Figure 3). One group includes the sequences of 705 bp of *P. scabrosa* and *P. rotundipennis*, plus the sequence of 533 bp of *P. monticola*. The second contains PIM357 monomers, with the sequences from Canary Islands appearing as basal. The last group joins the repeats of about 500 bp where the sequences of *P. echidna* are basal relative to two sister groups: (1) *P. rugosa* and *P. fornicata*, and (2) *P. cordata* and the two unclassified *Pimelia* species from Tunisia. A parsimony analysis based on the most conserved segments (I and II), revealed identical sequence groupings to those obtained from the complete sequences.

The consensus sequences of *P. rotundipennis* and *P. scabrosa* (both 705-bp long) differ by two nucleotide substitutions only (positions 124 and 467; Figure 2) but neither of these is completely fixed within the species. The shorter consensus sequence of *P. monticola* (PMON), of 533 bp, shows sequence identity with the 532-bp-long segment of the two former species given that there is an indel at position 441 (Figure 2). The related segments shared by PMON and PSCA–PROT show about 80% similarity (see Table 2) with point mutations distributed throughout the monomers. The remaining 173 bp of the PSCA and PROT sequences are made of partial duplications of the common 533-bp segment: a 32-bp-long direct duplication (90.6% similarity, Figure 2), and a 141-bp-long duplicated segment in an inverted orientation (78.7% similarity, Figure 2).

The species *P. rugosa*, *P. fornicata* and *P. echidna* show related sequences of 502 bp. The sequences of *P. echidna* are more dissimilar to the remaining repeats (differing by around 20%, see Table 2) due to many single diagnostic (fixed) substitutions distributed throughout the monomeric sequence. However, the sequences of *P. fornicata* and *P. rugosa* are very similar (the interspecific divergence is very similar to the intraspecific divergence, see Table 2). In fact, monomers from these species differ by 10 nucleotide substitutions, but only four of these are fixed at the species level (positions 214, 227, 248, and 417; Figure 2). Finally, the repeat units of the two unclassified species from Tunisia (P1NF and P2NF) and those from *P. cordata*, are different in sequence from the ones described above and also from that of PIM357 family. The monomers P1NF and P2NF differ by two gaps of 10 and 12 bp, showing 75.4% sequence similarity (Figure 2). *P. cordata* monomers are the most divergent relative to the repeats of any other species (Figure 2 and Table 2).

Analysis of sequence-induced curvature

All the species have A + T-rich stDNA sequences (Table 1) and high frequencies of periodically distributed runs, composed of three or more adenines or thymines, throughout the monomers. All the isolated repeat units migrated slower than expected, based on their length, on nondenaturing polyacrylamide gels. The repeat units

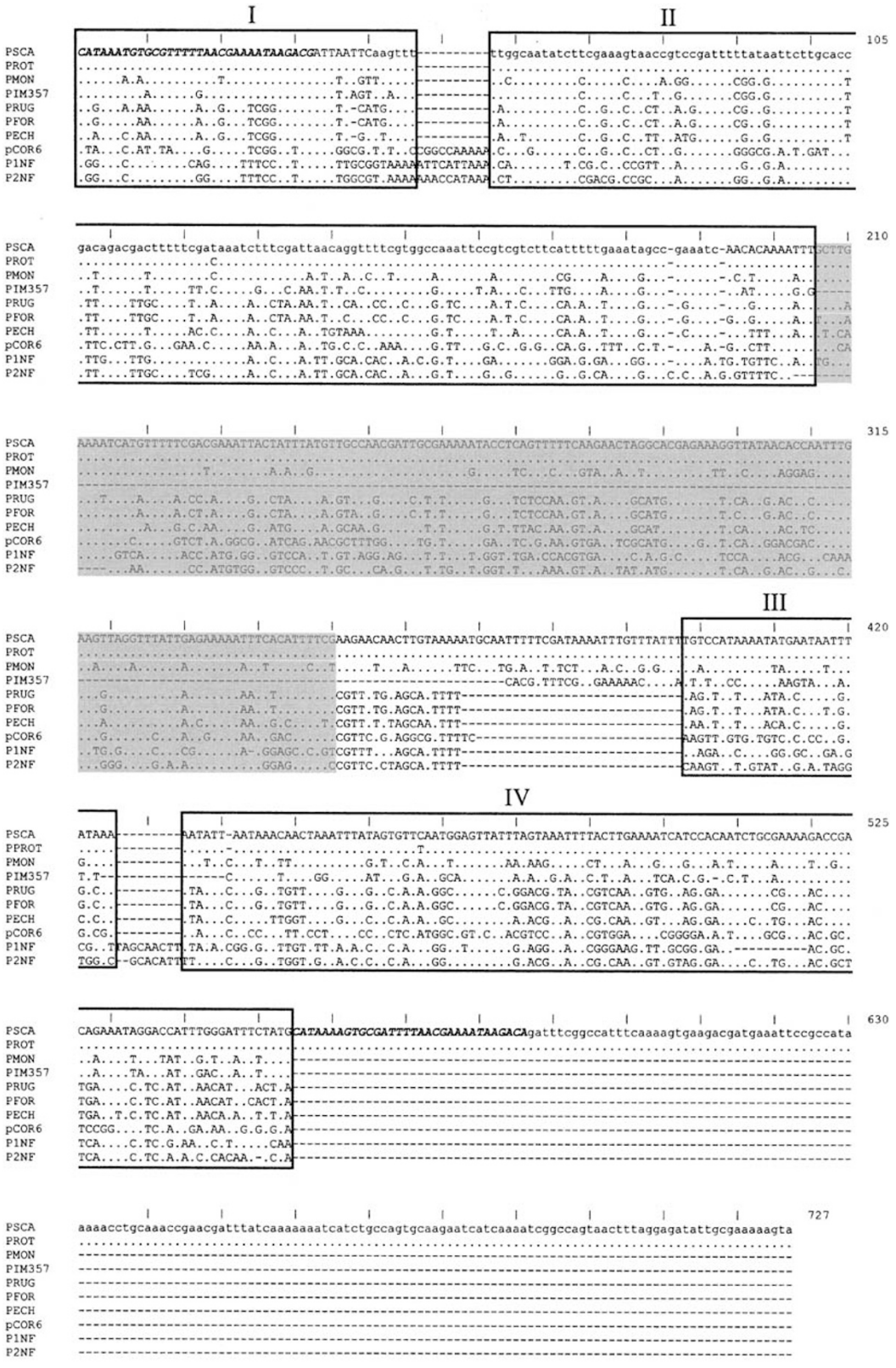

Figure 2 Nucleotide sequence alignment of the consensus monomeric sequences from: *Pimelia scabrosa* (PSCA), *P. rotundipennis* (PROT), *P. monticola* (PMON), stDNA family from 26 *Pimelia* taxa characterised elsewhere (Pons *et al*, 2002a, PIM357), *P. rugosa* (PRUG), *P. fornicata* (PFOR), *P. echidna* (PECH), clone pCOR6 from *P. cordata*, and *Pimelia sp. 1* (P1NF) and *Pimelia sp. 2* (P2NF) from Tunisia. Dots denote nucleotides that are the same as the first sequence; dashes denote gaps. Boxes denote the four fragments conserved in all nine consensus monomeric sequences (I–IV). The central segment (shadowed) is not present in the consensus PIM357 unit. Repeat units of *P. scabrosa* and *P. rotundipennis* exhibit a 32-bp-long direct repeat (indicated in bold and italics), and inverted one of 141 bp in length (indicated in lower case).

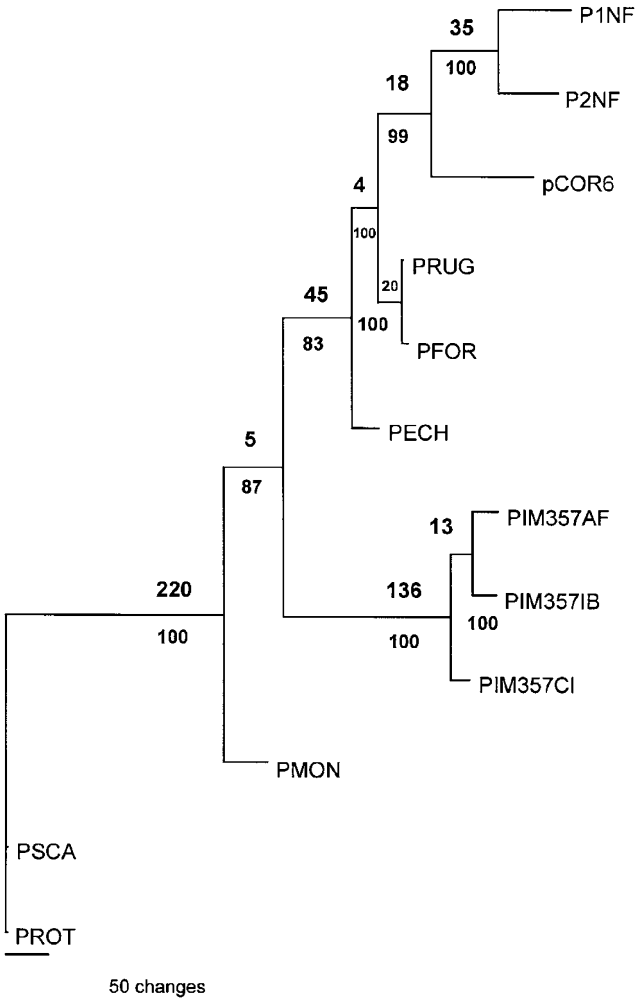

Figure 3 Unrooted phylogram representing the single shortest tree obtained from the parsimony analysis of the *Pimelia* monomeric sequences (1264 steps). Branch lengths represent the average number of parsimony changes (gaps treated as 5th character). Numbers at each node indicate Bremer support value out of 200 replicates (above node) and the percentage of trees representing the particular node out of 1000 bootstrap replicates (below node). Bootstrap values under 50 are not indicated.

showed similar R_L values ranging from 1.71–2.17 (R_L is a ratio of apparent to actual length in an acrylamide gel at 4°C, and it is an estimation of the retarded mobility).

Pimelia phylogeny based on mitochondrial and nuclear markers

Alignment of the five genes (COI, cytochrome *b*, 16S, 28S and Histone 3) was straightforward with few gap sites (seven in 28S and 16 in 16S), mostly nonambiguous. The combined analysis of the five markers in tree searches resulted in a single tree of 2633 steps (Figure 4). The two subgenera, *Pimelia* s. str. and *Amblyptera*, appear as monophyletic sister groups. Within *Pimelia* s. str., there is one clade composed of African species whose major stDNAs belong to the PIM357 family with *P. monticola* basal to them. The other one is composed of two clades: (1) Iberian-Balearic species whose major stDNAs belong to the PIM357 family; and (2) endemic Canarian species (with the major stDNAs being the PIM357 family) with African species whose stDNAs are about 500-bp long basal to them (*P. echidna*, *P. cordata*, and the two unclassified species from Tunisia). Within *Amblyptera*, two clades are resolved: one composed of species with the major satellite of 705 bp (*P. rotundipennis* and *P. scabrosa*), and the other with species *P. fornicata* and *P. rugosa*, sharing the highly similar major satellite of 502 bp.

The five partitions were studied separately to test for the phylogenetic signal and potential conflict. Character incongruence between the different markers as determined by the ILD test was nonsignificant for all possible combinations of the five genes (data not shown), except when 16S gene was included ($P<0.01$). Nevertheless, the incongruence between the three mitochondrial partitions was reduced when they were analysed together with the nuclear data (ILD/change = 0.0362 *vs* 0.0072), demonstrating that combining all data in a simultaneous analysis produced a more consistent character distribution than any of the partitions separately. Most of the signal and support was provided by COI, 16S, and Histone 3 (PBS 177, 87, and 61, respectively). Despite its signal, cytochrome *b* supported only 40% of the nodes with no global PBS. Finally, 28S showed no global positive PBS probably due to its low number of informative sites (14), supporting four nodes only (the *Amblyptera* clade, the African PIM357 clade, the *P. radula radula* and *P. radula granulata* clade, and the outgroups *T. molitor* and *T. castaneum*).

ML branch lengths based on the parameters selected by MODELTEST (GTR + I + G) and the MP tree were used to estimate node ages. Likelihood ratio tests between rate-constant and rate variable models, for the combined data and for each one of the partitions, revealed deviation from the molecular clock (data not shown). Therefore, ML branch lengths of the combined data were fitted using the NPRS algorithm. Absolute node ages were estimated by constraining the split of *P. laevigata laevigata* and *P. laevigata costipennis* to 1 Mya since the island El Hierro, from where the latter species is endemic, is not older than 1 Mya (Juan *et al*, 1995). This age represents a maximum estimate since these beetles not necessarily colonised El Hierro immediately after its emergence. An older precolonisation age also seems to be unlikely because there is no evidence of an earlier split of the lineage in La Palma Island (with an age of about 2 Mya; Juan *et al*, 1995) from where *P. laevigata laevigata* is endemic. Using that split of 1 Mya as a calibration point, we can conclude that the maximum estimate for the split between the two subgenera *Pimelia* s. str. and *Amblyptera* was about 11 Mya. On this basis, *Pimelia* s. str. lineages originated about 9–6 Mya: the African lineage harbouring the major stDNA PIM357 about 9 Mya, the Iberian and Canarian lineages at about 7 Mya, and the species with major stDNA sequences of 500 bp about 8–6 Mya. However, estimation based on mitochondrial ML branch lengths, fitted on NPRS, and using the considered standard insect mtDNA clock of around 2% per My (Brower, 1994) nearly doubled node ages.

Discussion

The phylogeny based on mitochondrial (COI, cytochrome *b*, and 16S) and nuclear markers (28S and

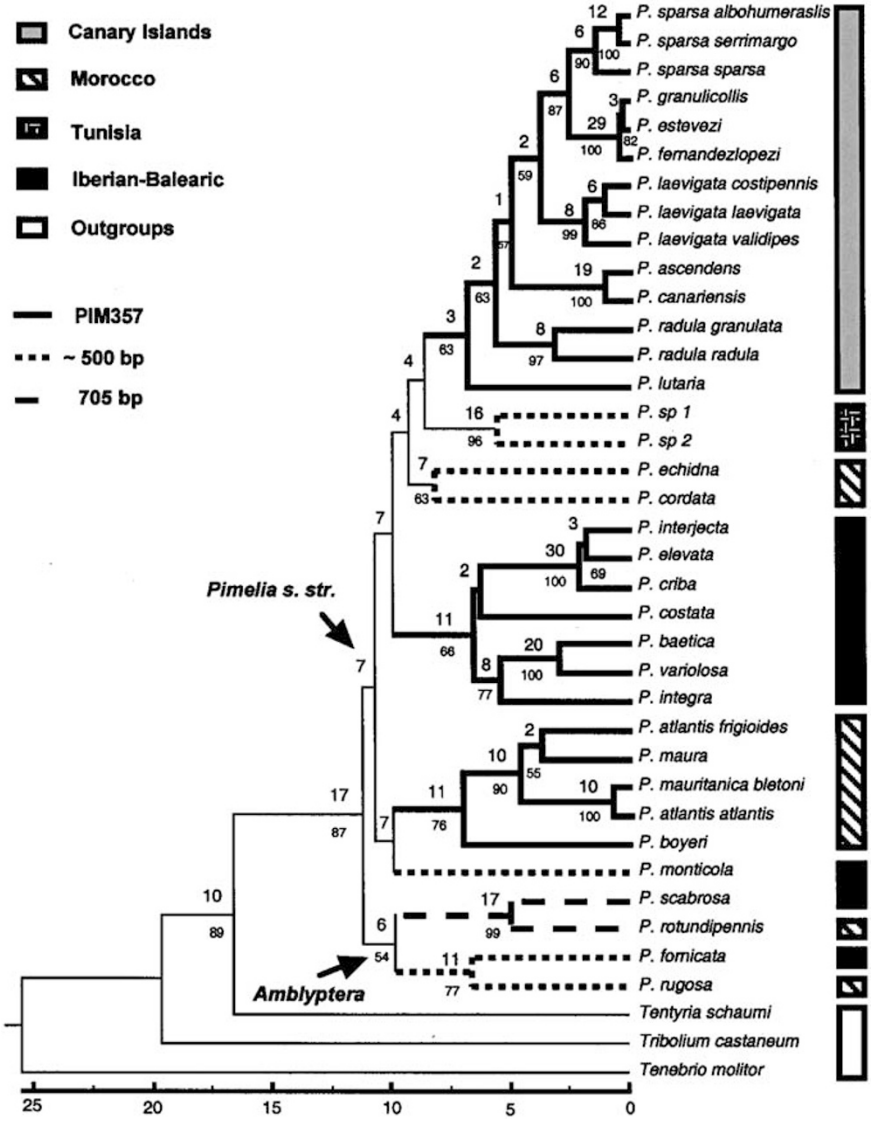

Figure 4 Parsimony tree representing the single shortest tree (2633 steps) obtained from simultaneous analysis of COI, cytochrome *b*, 16S, 28S, and Histone 3 (2236 characters). Numbers at each node indicate Bremer support values out of 200 replicates (above node) and the percentage of trees representing the particular node out of 1000 bootstrap replicates (below node). Bootstrap values under 50 are not indicated. Relative node ages were estimated calculating the ML branch lengths (GTR + G + I) on the MP tree and then made ultrametric by NPRS. The scale bar shows the calibration of absolute ages (My). Absolute node ages were estimated by constraining the split of *P. laevigata laevigata* and *P. laevigata costipennis* to 1 Mya since the island El Hierro, from where is endemic the last subspecies, is not older than 1 Mya (Juan *et al*, 1995). The biogeographic regions and major satellites are indicated for each species.

Histone 3) reveals a different scenario of diversification of *Pimelia* species than that based on the major stDNA sequences. The presence of units similar in length (502 bp) and sequence, in species from two different subgenera *P. rugosa* and *P. fornicata* (*Amblyptera*), and *P. echidna* (*Pimelia* s. str.) could be explained by an independent amplification from the same pool of satellite sequences, that is, the 'library hypothesis' (Fry and Salser, 1977; Ugarkovic and Plohl, 2002). The same applies for the case of *P. monticola* (*Pimelia* s. str.) and the *P. scabrosa–P. rotundipennis* pair (*Amblyptera*) whose major stDNA sequences, though different in length (533 and 705 bp, respectively), are clearly related. The existence of monophyletic groups having PIM357 as the major stDNA in three independent nodes of the phylogeny (Moroccan, Iberian-Balearic, and Canarian) could also be explained by independent amplification. Nevertheless, the phylogeny based on mitochondrial and nuclear markers could, depending on character transformations, support any of three different hypotheses, one, two or three independent amplifications from the 'library'.

The existence of a 'library of stDNA sequences' at low copy number has been previously demonstrated experimentally in six *Pimelia* species (Bruvo *et al*, 2003), and also in other tenebrionid species of the genus *Palorus* (Mestrovic *et al*, 1998). The *Pimelia* phylogeny based on mitochondrial and nuclear markers and the results of previous work (Bruvo *et al*, 2003) suggest a common ancestor whose genome harboured all or most of the

major stDNA families present in the living species at low copy numbers. Speciation and concerted evolution of the repeats may have changed dramatically their copy number by differentially amplifying from the 'library' one low copy satellite variant to compose the major satellite. However, in contrast to the data obtained in *Palorus* species, whose major satellite sequences showed no similarity, the monomers of the major stDNA families of *Pimelia* still conserve some regions of sequence similarity despite their marked differences, which suggests a common evolutionary origin.

The persistence of intrinsic bending in *Pimelia* monomers indicate its potential role in heterochromatin condensation (Brutlag, 1980; Martínez-Balbás *et al*, 1990; Barceló *et al*, 1998). Nevertheless, it has been shown that there is a tremendous variation in the magnitude of intrinsic bending among satellite DNAs (Fitzgerald *et al*, 1994), and it is this bending, rather than DNA sequence *per se*, which could be the motif being recognised by specific binding proteins (Lobov *et al*, 2001). The importance of bending could explain the partial conservation of the primary sequence in *Pimelia* monomers, since curvature is linked to a periodical distribution of runs composed of three or more adenines, but not conservation of length (which is 357, 500, or 705 bp). A similar trend has been suggested for other tenebrionid stDNAs (Ugarkovic *et al*, 1995; Barceló *et al*, 1998).

A schematic representation of a plausible evolutionary scenario for the origin of satellite repeats of *Pimelia* in the ancestor, that is, the acquisition of the 'library', is shown in Figure 5. Since repetitive sequences from outgroups have no similarity with *Pimelia* satellites, and the independent phylogeny could support either scenario, we cannot decide whether *Pimelia* repeats originated from 357 or 500 bp sequences. However, most of the cases described to date demonstrate that long and complex repeats have been created from shorter ones (Bigot *et al*, 1990; Rojas-Rousse *et al*, 1993; Ugarkovic *et al*, 1996). Therefore, we suggest that the insertion of a 141-bp DNA fragment into the unit of about 357 bp shared for all species created the ancestral units of 500 bp. The sequences of 500-bp split in two lineages: one has preferentially been amplified in *P. monticola* (533 bp) and the other is present in several species. The repeats of *P. scabrosa* and *P. rotundipennis* (705 bp) have likely originated by a 32-bp direct duplication and the insertion of a 141-bp duplicated segment in inverted orientation with respect to the basic structure of the 533-bp unit currently present in *P. monticola* (Figure 5). Other insertion/deletion events seem to have occurred in a

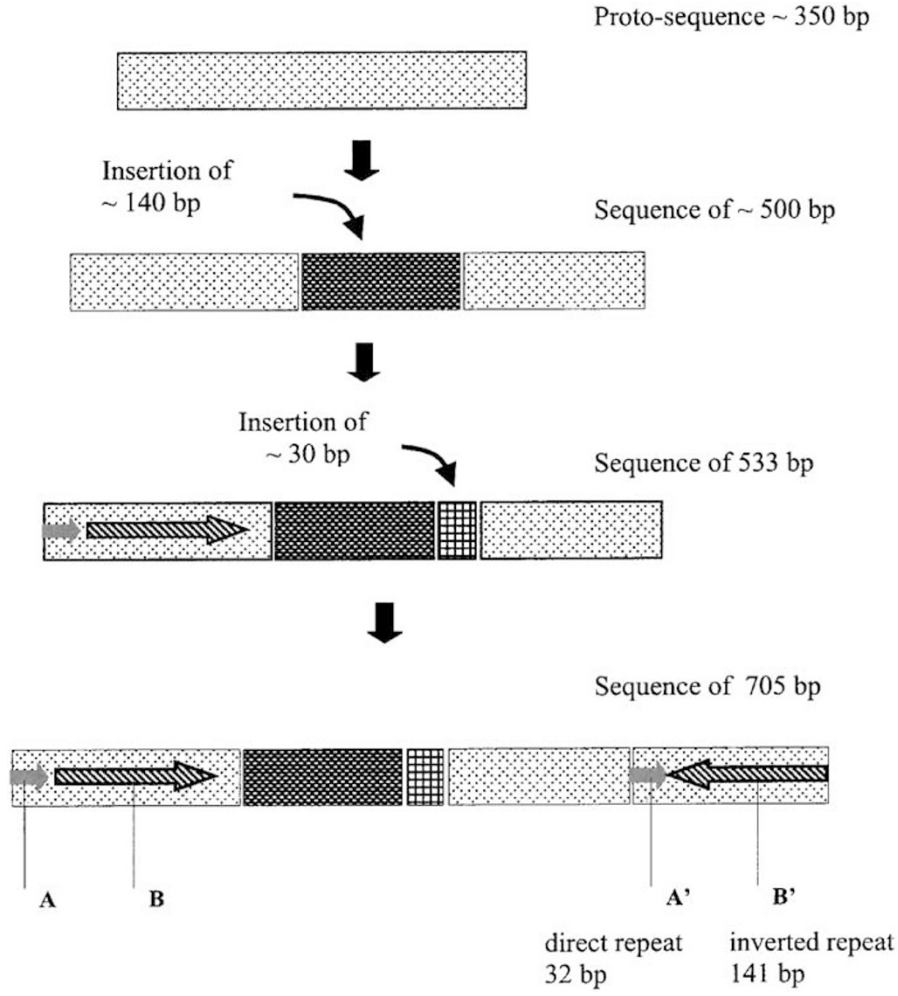

Figure 5 A schematic portrayal of a plausible evolutionary scenario for the satellite repeats of *Pimelia*.

species-specific way, as they do not show similarity with any other region of a repeat unit, although mutational processes might have obscured the similarity between the original and duplicated sequences. Rearrangement of the PIM357 monomers to create 500-bp units has been previously described in some Canarian *Pimelia* (Pons *et al*, 2002a), but these are different from the sequences described in this paper (eg they do not include the 141-bp motif shared by 500 and 700 repeats). Similar rearrangements have been suggested as mechanisms involved in the evolution of two related stDNA families in the coleopteran *Tribolium madens*, which are composed of different subunits of about 100 bp with 60–80% homology (Ugarkovic *et al*, 1996). Recombination is thought to be one of the main forces driving concerted evolution (Dover, 2002), but may be also a force generating new monomers by rearrangement of the repeats.

The two sister clades of the subgenus *Amblyptera*, which split about 10–20 Mya, amplified different repeats: in *P. rugosa* and *P. fornicata* a sequence of 502 bp, and in *P. scabrosa* and *P. rotundipennis*, a 705-bp unit. Interestingly, the sequences show very little variation within the clades despite having diversified about 6–12 Mya. On the other hand, the three paraphyletic clades showing PIM357 sequences that split between 10–20 Mya, Moroccan, Iberian-Balearic, and Canarian, have evolved gradually, homogenising and fixing mutations within each clade during the last 6–12 Mya. Our results show that satellite sequences are evolving at different rates in two manners: both gradually and saltatory. Moreover, the partial conservation of the periodical distribution of runs composed of three or more adenines in *Pimelia* satellites could be related to the role of stDNA in tight heterochromatin condensation.

In summary, the results presented in this paper suggest that most of the major stDNAs of *Pimelia* present in the extant species were already present in a common ancestor composing a 'library of satellite sequences' of common origin characterised by extensive sequence rearrangement. During diversification, different species randomly amplified different examples of these low copy sequences to form the abundant major satellites. This turnover occurred at different rates: in relatively short periods of time (abrupt-saltatory replacement), in a gradual manner (consistent with a molecular drive model, Dover, 2002), or simply with no apparent change for long evolutionary time.

Acknowledgements

We are very grateful to Drs P Oromí, J Gomez-Zurita, M Palmer, and C Garin who provided many of the *Pimelia* samples. Drs Oromí and Gómez-Zurita also helped in the taxonomic determination. This work was supported by the Spanish Research Funds REN2000–0282/GLO, REN2003–00024/GLO to CJ and BOS2000–0822 to EP.

References

Barceló F, Pons J, Petitpierre E, Barjau I, Portugal J (1997). Polymorphic curvature of satellite DNA in three subspecies of the beetle *Pimelia sparsa*. *Eur J Biochem* **244**: 318–324.

Barceló F, Gutiérrez F, Barjau I, Portugal J (1998). A theoretical perusal of the satellite DNA curvature in tenebrionid beetles. *J Biomol Struct Dyn* **16**: 41–50.

Bigot Y, Hamelin MH, Periquet G (1990). Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). *Mol Biol Evol* **7**: 351–364.

Brower AVZ (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-DNA evolution. *Proc Natl Acad Sci USA* **91**: 6491–6495.

Brutlag DL (1980). Molecular arrangement and evolution of heterochromatin DNA. *Ann Rev Genet* **14**: 121–144.

Bruvo B, Pons J, Ugarkovic D, Juan C, Petitpierre E, Plohl M (2003). Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (Coleptera). *Gene* **312**: 85–94.

Colgan DJ, Mclauchlan A, Wilson GDF, Livingston SP, Edgecombe GD, Macaranas J *et al* (1998). Histone H3 and U2 snRNA DNA sequences and arthropod molecular evolution. *Aust J Zool* **46**: 419–437.

Dover G (2002). Molecular drive. *Trends Genet* **18**: 587–589.

Farris JS, Källersjö M, Kluge A G, Bult C (1994). Testing significance of incongruence. *Cladistics* **10**: 315–319.

Fitzgerald G, Dryden EC, Bronson EC, Williams JS, Anderson, JN (1994). Conserved pattern of bending in satellite and nucleosome positioning DNA. *J Biol Chem* **269**: 2312–2314.

Fry K, Salser W (1977). Nucleotide sequence of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterisation of similar sequences in other rodents. *Cell* **12**: 1069–1084.

Funk DJ (1999). Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. *Mol Biol Evol* **16**: 67–82.

Higgins DG, Thompson JD, Gibson TJ (1996). Using clustal for multiple sequence alignments. *Methods Enzymol* **266**: 383–402.

Juan C, Oromí P, Hewitt GM (1995). Mitochondrial DNA phylogeny and sequential colonization of Canary Islands by darkling beetles of the genus *Pimelia* (Tenebrionidae). *Proc R Soc B* **261**: 173–180.

Kwieton E (1977). Esquisse phylogenetique du genre *Pimelia* F. *Acta Entomol Mus Natl Pragae* **39**: 559–589.

Lobov IB, Tsutsui K, Mitchell AR, Podgornaya OI (2001). Specificity of SAF-A and lamin B binding *in vitro* correlates with the satellite DNA bending state. *J Cell Biochem* **83**: 218–229.

Martínez-Balbás A, Rodriguez-Campos A, Garcia-Ramirez M, Sainz J, Carrera P, Aymani J *et al* (1990). Satellite DNAs contain sequences that induce curvature. *Biochemistry* **29**: 2342–2348.

Mestrovic N, Plohl M, Mravinac B, Ugarkovic D (1998). Evolution of satellite DNAs from the genus *Palorus* – experimental evidence for the 'library' hypothesis. *Mol Biol Evol* **15**: 1062–1068.

Pons J (1999). *Evolución del DNA satélite en el género Pimelia*. PhD Thesis,. University of Balearic Islands.

Pons J, Bruvo B, Juan C, Petitpierre E, Plohl M, Ugarkovic D (1997). Conservation of satellite DNA in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Gene* **255**: 183–190.

Pons J, Petitpierre E, Juan C (2002a). Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Mol Biol Evol* **19**: 1329–1340.

Pons J, Petitpierre E, Juan C (2002b). Higher-order organisation and compartmentalisation of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. *Chromosome Res* **10**: 597–606.

Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.

Rambaut A, Charleston M (2002). *TreeEdit. Phylogenetic Tree Editor v1.0 alpha 10*. Oxford University: Oxford.

Rojas-Rousse D, Bigot Y, Periquet G (1993). DNA insertions as a component of the evolution of unique satellite DNA families in two genera of parasitoid wasps: *Diadromus* and *Eupelmus* (Hymenoptera). *Mol Biol Evol* **10**: 383–396.

Rozas J, Rozas R (1997). DnaSP version 2.0 a novel software package for extensive molecular population genetic analysis. *Comput Appl Biosci* **13**: 307–311.

Sorenson MD (1999). *TreeRot, Version 2*. Boston University: Boston, MA.

Swofford DL (2002). PAUP*. *Phylogenetic Analysis using Parsimony, Version 4.0b*. Sinauer Associates: Sunderland, MA.

Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles – causes and effects. *EMBO J* **21**: 5955–5959.

Ugarkovic D, Petitpierre E, Juan C, Plohl M (1995). Satellite DNAs in tenebrionid species: structure, organization and evolution. *Croat Chem Acta* **68**: 627–638.

Ugarkovic D, Durajlija S, Plohl M (1996). Evolution of *Tribolium madens* (Insecta, Coleoptera) satellite DNA through DNA inversion and insertion. *J Mol Evol* **42**: 350–358.

Vogler AP, Welsh A (1997). Phylogeny of North American *Cicindela* tiger beetles inferred from multiple mitochondrial DNA sequences. *Mol Phylogenet Evol* **8**: 225–235.