

Highly repeated DNA sequences in three species of the genus *Pteropus* (Megachiroptera, Mammalia)

MJL Barragán¹, S Martínez¹, JA Marchal¹, M Bullejos¹, R Díaz de la Guardia² and A Sánchez Bullejos¹

¹Departamento de Biología Experimental, Facultad de Ciencias Experimentales y de la Salud, Universidad de Jaén, E-23071 Jaén, Spain; ²Departamento de Genética, Facultad de Ciencias, Universidad de Granada, E-18071 Granada, Spain

Bat genomes are characterised by an A-T richness and by a small C-value compared with other mammalian groups. It has been suggested that the small C-value is mainly due to the lack of repetitive DNA sequences. However, little information about repetitive DNA sequences in this mammalian group is available at the molecular level. Here we describe a *Pst*I family of repetitive DNA sequences present in three species of the genus *Pteropus*. These repetitive sequences

are 54.97% G-C rich, organised in tandem and with a unit length of 744 bp. Methylation analysis indicates that some of the CCGG target sites present in these repetitive DNA sequences have methylated cytosines and that there are small differences in the methylation pattern between species. Several features of this family of repetitive sequences suggest that they evolved by concerted evolution.

Heredity (2002) **88**, 366–370. DOI: 10.1038/sj/hdy/6800064

Keywords: *Pteropus*; megachiroptera; repetitive DNA; satellite DNA

Introduction

Mammalian genomes are characterised by a large number of repetitive sequences, which are highly variable from one species to another. These sequences constitute families of repetitive DNA that represent a substantial component of many eukaryotic genomes. In fact, these types of sequences can represent values from less than 10% to more than 80% of the DNA content (Ridley, 1996). Two main kinds of highly repetitive sequences are known in mammalian genomes: interspersed DNA, in which the repeated DNA sequences are dispersed throughout the genome; and satellite DNA which is characterised by long tandem arrays and consistent association with constitutive heterochromatin (Singer, 1982). The localised repeats, or satellite DNAs, are typically found at C-band-positive regions of chromosomes, such as centromeres, telomeres and a portion of the Y chromosomes (Modi *et al*, 1988; Hamilton *et al*, 1990). The satellite DNAs, which are generally A-T rich, are organised in tandem arrays of repeat units and are highly variable with respect to the nucleotide sequence, the size of units, number of repeated units, and genome and chromosome organisation and localisation (Charlesworth *et al*, 1994).

The biological function of these DNA sequences remains elusive, although it has been proposed that they are involved in centromeric condensation, sister chromatid pairing, chromosome association with the mitotic spindle, karyotypic evolution and chromosome arrangement (Miklos and John, 1979; Singer, 1982; Vogt, 1992).

Bats possess a small C value (genome size is approximately 50–87% the genome size of other eutherian mammals: Burton *et al*, 1989) possibly due to control of copy number of repetitive DNA sequences. This decrease in DNA content is accompanied by a high overall A-T content (Pettigrew, 1994). Van Den Bussche *et al* (1995) analysed the frequency of three types of repeat sequences (dinucleotide microsatellites, ribosomal DNA cistrons and a family of repetitive DNA sequences) in the bat *Macrotus waterhousii* demonstrating that there is a lower copy number of interspersed and tandemly repetitive elements. These authors suggest that the difference in the nuclear content of bats in relation to other mammalian groups is mainly due to the lack of interspersed repetitive DNA sequences. Although repetitive DNA sequences have been studied in almost all mammalian groups, in general little information is available at the molecular level for these sequences in bats, the second largest mammalian order, which includes nearly 1000 species in existence.

In this study we present the characterisation of a *Pst*I family of repetitive DNA sequences present in the genome of three bat species from the genus *Pteropus*.

Materials and methods

Species sampled and DNA extraction

Individuals of three species from the genus *Pteropus* were analysed. These species are *P. scapulatus*, *P. alecto* and *P. poliocephalus*. Genomic DNA was extracted from ethanol-fixed tissues according to standard phenol-chloroform procedures (Sambrook *et al*, 1989).

Cloning and sequencing

In order to detect the repetitive DNA sequences genomic DNAs from *Pteropus* species were digested with *Nhe*I,

Correspondence: Dr A Sánchez Baca, Departamento de Biología Experimental, Facultad de Ciencias Experimentales, Universidad de Jaén, Paraje Las Lagunillas s/n, E-23071 Jaén, Spain. E-mail: abaca@ujaen.es
Received 12 June 2001; accepted 3 January 2002

*Pst*I, *Eco*RI and *Alu*I restriction endonucleases. Only when genomic DNAs were digested with *Pst*I was a prominent band of repetitive DNA found. *Pst*I bands from the three species analysed were eluted from agarose and ligated to *Pst*I digested pUC19 vector.

Escherichia coli JM109 competent cells were transformed with the ligation reactions and the recombinant bacteria containing the sequences of interest were selected after screening, using as a probe the band from *P. alecto*, digoxigenin-labelled by random priming (Roche).

The positive clones were sequenced in both directions using the ThermoSequenase fluorescent cycle sequencing kit from Amersham. The sequence reactions were analysed on a 6.5% polyacrylamide gel in a LICOR-400L automated sequencer.

Southern blot

Genomic DNAs were digested with the appropriate restriction endonucleases. The fragments were separated in a 1% agarose gel and blotted onto nylon membranes (Amersham) according to Sambrook *et al* (1989). Membranes were probed with the *P. alecto* eluted band or with the clone P.ale 2.2 at 55°C overnight. Alkaline phosphatase detection was carried out according to the supplier's recommendations (Roche).

Sequence analysis

Pairwise sequence alignment and multiple alignment were carried out with the program CLUSTAL W 1.6 (Thompson *et al*, 1994). Genetic distances were calculated according to Kimura's (1980) two-parameter method, and the resulting distance matrices were subjected to neighbour-joining analysis (Saitou and Nei, 1987). A consensus tree was constructed, and the significance of the phylogenetic lineages was assessed by bootstrap analysis with 1000 replications. The phylogenetic analyses were carried out using the program MEGA Version 1.02 (Kumar *et al*, 1993).

Results and discussion

Digestion of genomic DNA from *P. scapulatus*, *P. alecto* and *P. poliocephalus* with *Pst*I yielded a prominent band at about 744 bp (Figure 1a). When genomic DNA of these three species was digested with *Pst*I and probed with the digoxigenin-labelled 744 bp band from *P. alecto*, a regular ladder pattern was seen (Figure 1b). In general, the size of each band corresponds to the size of multimers of the monomer unit of 744 bp. Nevertheless, some bands do not correspond to monomer multimers. In *P. scapulatus* and *P. poliocephalus* there are also bands of 372 bp and 186 bp, which could be due to the presence in some monomer units of target sites for *Pst*I. In fact several incomplete target sites for this enzyme are present in the sequences of some monomers. These sites need only one base substitution change to give rise to the complete target site.

The presence of a ladder pattern in Southern blots probed with a digoxigenin-labelled 744 bp band clearly indicates that the repeated sequences in *Pteropus* are arrayed in tandem. It is likely, as occurs with other repeated DNA sequences present in eukaryotic genomes, that the bands containing several monomer units are originated by one base substitution change that modifies the target sequences for *Pst*I.

When genomic DNA from other bat species, including

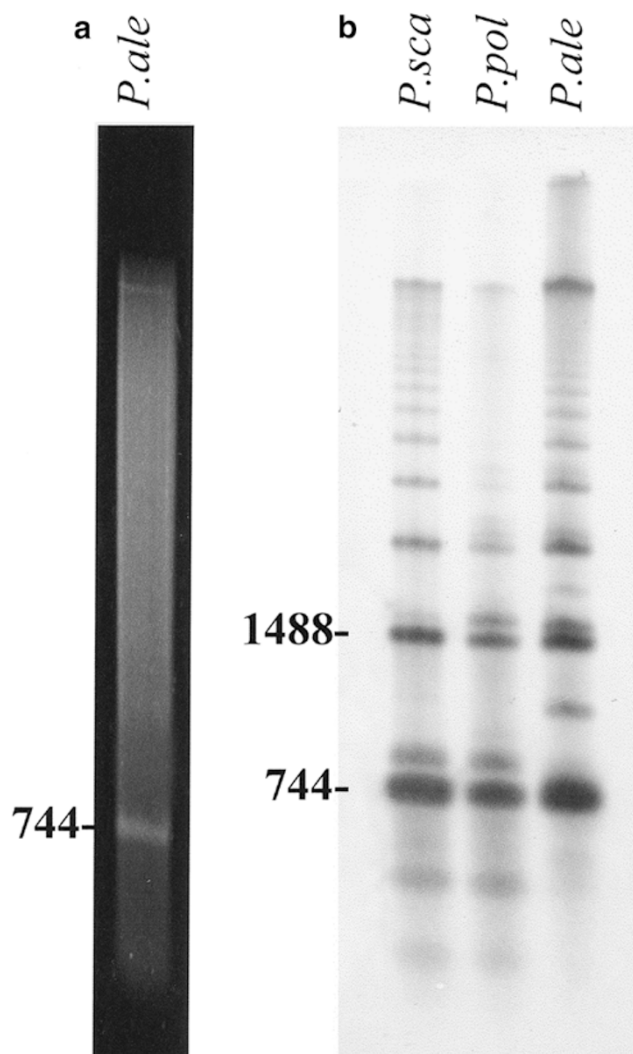


Figure 1 (a) Gel electrophoresis of *Pst*I digested genomic DNA from *Pteropus alecto*. (b) Southern blot of *Pst*I digested genomic DNAs from *P. scapulatus* (*P.sca*), *P. poliocephalus* (*P.pol*) and *P. alecto* (*P.ale*) probed with the random primer digoxigenin-labelled 744 bp band of *P. alecto*.

a megabat (*Rousettus aegyptiacus*) and three microbats (*Myotis myotis*, *Rhinolophus ferrumequinum* and *R. hipposideros*), were digested with *Pst*I and probed with the repetitive sequence of *Pteropus*, no band was observed (data not shown). This result implies that these sequences are absent in these micro- and megabat species and hence could be exclusive to the genus *Pteropus*.

After ligation with *Pst*I digested pUC19 vector, transformation of *E. coli* and screening, positive clones were obtained from all the species of the genus *Pteropus* analysed: five positive clones from *P. scapulatus*: P.sca 19, 21, 27, 36 and 50; one from *P. poliocephalus*: P.pol 66 and two from *P. alecto*: P.ale 4.15 and 2.2 (with two monomer units, P.ale 2.2a and b).

Sequences from different clones showed that the satellite DNA cloned in *Pteropus* species belong to the same repetitive family. Sequence alignment allowed the determination of the consensus sequence for the monomer units. The length of the consensus sequence was 744 bp, although the length of cloned sequences ranged between

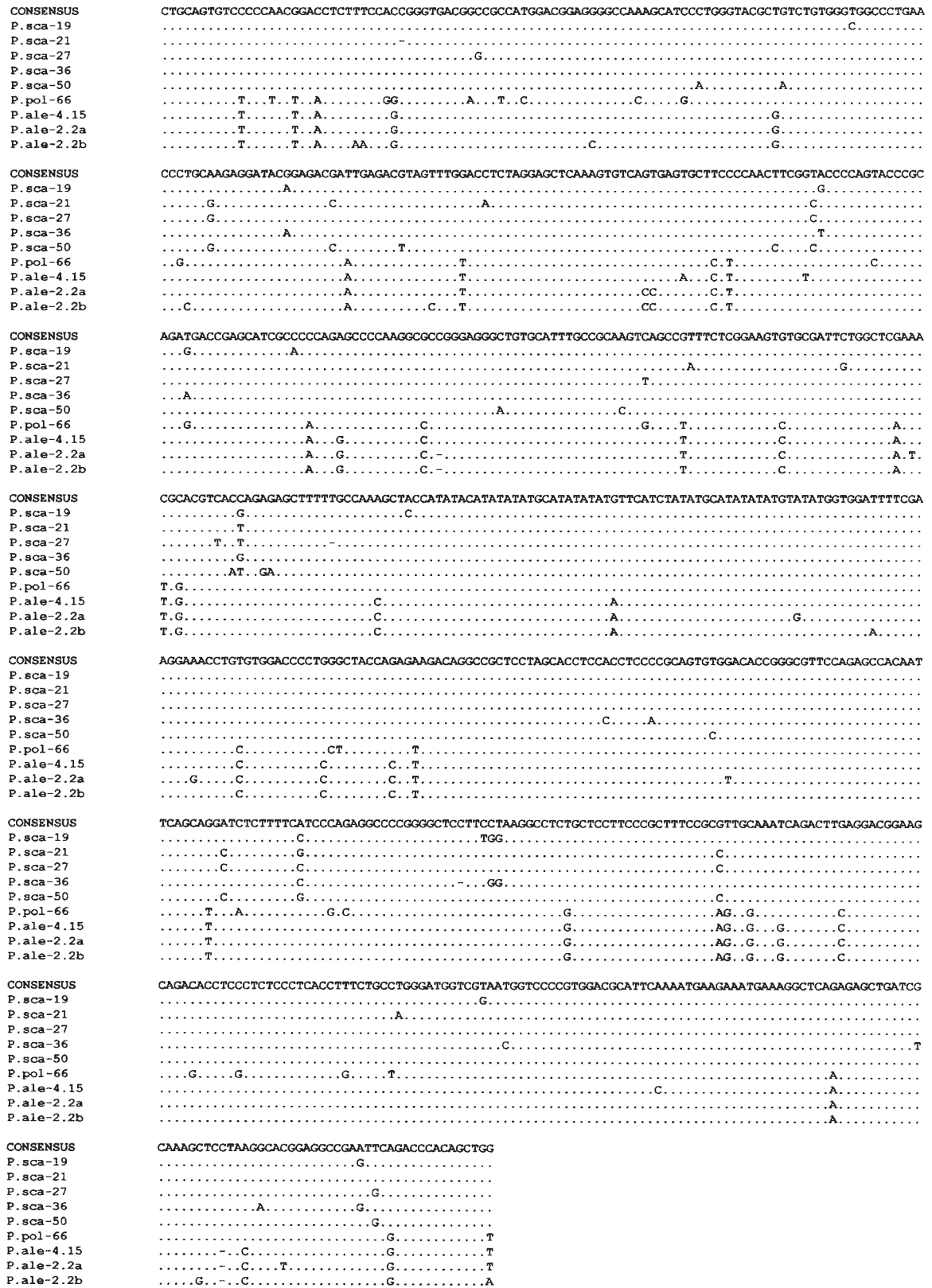


Figure 2 Sequence alignment of the consensus sequence with the different monomer units cloned from *Pteropus scapulatus* (P.sca), *P. poliocephalus* (P.pol) and *P. alecto* (P.ale). These sequences have been submitted to GenBank and have been assigned the accession numbers AJ311379 to AJ311387.

744 and 742 bp (Figure 2). The consensus sequence is 54.97% G-C rich. A high A-T content is considered to be a general feature of most of satellite DNAs, as is the case for several rodent species which range from 60% to 63% (Singer, 1982). Furthermore, this higher G-C content for these sequences is interesting because the genomes of bat species are characterised by a high A-T content. In fact, the genomes of pteropodid bats have over 70% A-T content in their DNA (Petigrew and Kirsch, 1995).

Figure 2 shows the alignment of the consensus sequence with the cloned monomer units. Comparative analysis demonstrated that base substitution mutations are randomly spread along the sequences and that there is a high sequence identity, with the percentage between different monomer units ranging between 96 and 98%. The same substitutions have occurred in several clones of this satellite DNA at the same position in the sequences. Such a mutational pattern could be explained as the result of homogenisation processes: by unequal crossing-over and amplification some particular monomer variations can be spread throughout the whole satellite (Dover, 1986). The evolution of satellite DNA is mainly determined by processes of concerted evolution. The concept 'concerted evolution' describes processes by which a sequence can be amplified, homogenised throughout the genome and distributed among both homologous and non-homologous chromosomes (Dover, 1982). Several mechanisms have been proposed to be active in certain genomes, such as unequal crossing over, gene conversion, sequence transposition and rolling circle replication (Dover, 1982; Walsh, 1987), which can act together or individually during the process of concerted evolution.

The consensus sequence from *P. scapulatus* monomers shows percentages of identity of 94.89% with the consensus sequence from *P. alecto* and of 93.41% with the single monomer unit from *P. poliocephalus*. The identity between *P. alecto* consensus sequence and *P. poliocephalus* monomer unit is 96.10%.

As repetitive DNA sequences are known to be subject to hypermethylation, we investigated the methylation status in the CCGG sequences of this repetitive DNA in *P. scapulatus* and *P. poliocephalus*. The genomic DNA was digested independently with the methylation-insensitive enzyme *MspI* and with the methylation-sensitive enzyme *HpaII*. After Southern blot the membrane was probed with the clone P.ale 2.2 labelled with digoxigenin. In both species there are clear differences between the hybridisation pattern obtained with the two enzymes. While with *MspI* only two small bands showed hybridisation signal, with *HpaII* a typical ladder band was observed. These results are consistent with the fact that these repetitive DNA sequences have three or four target sites for these restriction endonucleases. Although the ladder obtained with *HpaII* in both species has the same main pattern, in *P. scapulatus* there are more interstitial bands than in *P. poliocephalus* (Figure 3). This result indicates that some of the CCGG target sites present in the repetitive DNA sequence have methylated cytosines in both species, and that there are small differences in the methylation pattern between them.

A neighbour-joining tree using the Kimura-2 parameter distance was constructed using nine monomer units from *Pteropus* species (Figure 4). Most branches are well supported including those that discriminate between

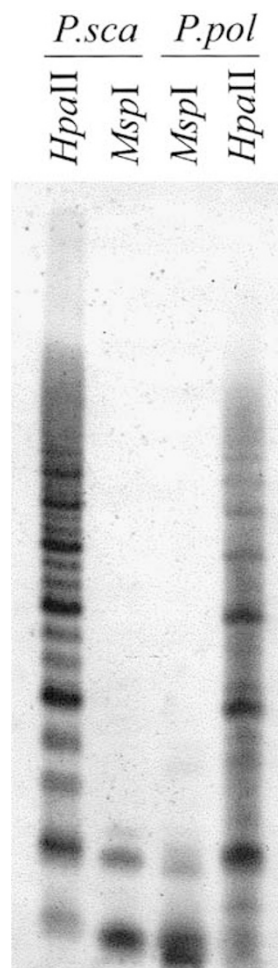


Figure 3 Southern blot of *Pteropus scapulatus* (*P.sca*) and *P. poliocephalus* (*P.pol*) genomic DNA digested with the methylation-insensitive enzyme *MspI* and with the methylation-sensitive enzyme *HpaII*, probed with the digoxigenin-labelled clone *P. ale 2.2*.

species monomers. All monomer units from *P. scapulatus* are clearly located on a branch separated from those belonging to *P. poliocephalus* and *P. alecto* (100% bootstrap value). Furthermore, the sequences for these two later species are also split into two separate branches with a bootstrap value of 99%.

Phylogenetic inference reveals that genetic distances between repeat units of the same species were smaller than genetic distances between repeat units from different species. These data suggest a concerted mode of evolution for this satellite DNA family. Hence, the repetitive DNA sequences described here for the three *Pteropus* species can be considered as paralogous sequences. Those sequences could have arisen by duplication or amplification in the ancestral species and become slightly differentiated in the genome of each species. Concerted evolution has been demonstrated for other repetitive DNA from rodents, as in the case of the harvest mouse *Reithrodontomys* (Hamilton *et al*, 1990). In fact, the alignment of the monomer units from these species showed the existence of multiple diagnostic positions, that is, species-specific changes (Figure 2).

In conclusion, we described a *PstI* family of repetitive DNA sequences present in three megabat species belong-

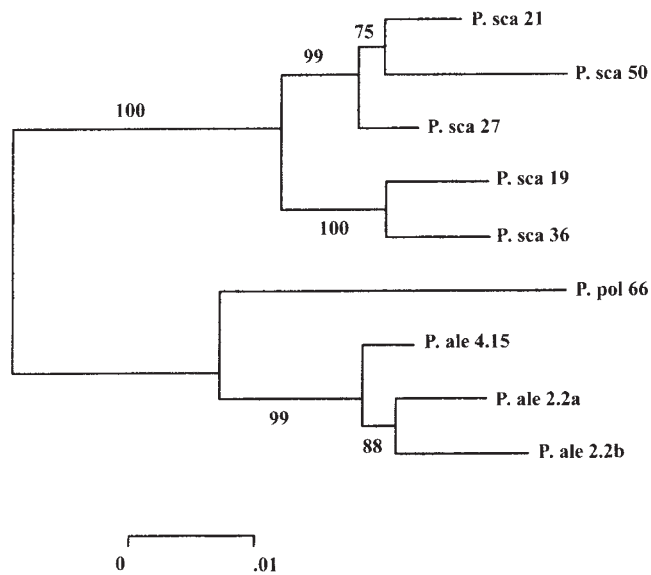


Figure 4 Neighbour-joining dendrogram showing the nine repetitive units of *PstI* satellite DNA from *Pteropus* species. The Kimura-2 parameter distance was used for tree construction. P. sca #, P. pol # and P. ale # are the different monomer units from *P. scapulatus*, *P. poliocephalus* and *P. alecto* respectively. The number in each branch indicates the bootstrap values (based on 1000 replicates). The branch length is drawn to the scale of the genetic distances (shown at bottom of figure). Note that monomers of each species are grouped in separate branches.

ing to the genus *Pteropus*. These repetitive sequences are organised in tandem arrays with a monomer unit of 744 bp and are 54.97% G-C rich. These sequences are clearly methylated in cytosine residues of the sequence CCGG. Furthermore, CLUSTAL alignment and phylogenetic analysis suggest that this family evolved by concerted evolution.

Acknowledgements

We would like to thank Dr Len Martin of Queensland University for generously providing megabat samples. This work was supported by the Spanish DGESIC through project numbers: PB98-1378-C02-01 and PB98-1378-C02-02, and by the Junta de Andalucía through the programme 'Ayudas a grupos de investigación', group numbers: CVI 0109 and CVI 220.

References

- Burton DW, Bickhan JW, Genoways HH (1989). Flow-cytometric analysis of nuclear DNA content in four families of Neotropical bats. *Evolution* **34**: 756-765.
- Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371**: 215-220.
- Dover GA (1982). Molecular drive: a cohesive mode of species evolution. *Nature* **199**: 111-117.
- Dover GA (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet* **168**: 159-165.
- Hamilton MJ, Honeycutt RL, Baker RJ (1990). Intragenomic movement, sequence amplification and concerted evolution in satellite DNA in harvest mice (*Reithrodontomys*): evidence from in situ hybridization. *Chromosoma* **99**: 321-329.
- Kimura M (1980). A simple method for estimating evolutionary rates of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* **16**: 111-120.
- Kumar S, Tamura K, Nei M (1993). MEGA: Molecular Evolutionary Genetics Analysis, Version 1.01. The Pennsylvania State University, University Park, PA 16802.
- Miklos GL, John B (1979). Heterochromatin and satellite DNA in man: properties and prospects. *Am J Hum Genet* **31**: 264-280.
- Modi WS, Fanning TG, Wayne RK, O'Brien SJ (1988). Chromosomal localization of satellite DNA sequences among 22 species of felids and canids (Carnivora). *Cytogenet Cell Genet* **48**: 208-213.
- Pettigrew JD (1994). Flying DNA. *Curr Biol* **4**: 277-280.
- Pettigrew JD, Kirsch JA (1995). Flying primates revisited: DNA hybridization with fractionated, GC-enriched DNA. *S Afr J Sci* **91**: 477-482.
- Ridley M (1996). *Evolution*. Blackwell Science Inc: Cambridge, Mass, USA.
- Saitou N, Nei M (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**: 406-425.
- Sambrook J, Fritsch EF, Maniatis T (1989). *Molecular Cloning. A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory: Cold Spring Harbor, NY, USA.
- Singer ML (1982). Highly repeated sequences in mammalian genomes. *Int Rev Cytol* **76**: 67-112.
- Thompson JD, Higgins DG, Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl Acids Res* **22**: 4673-4680.
- Van Den Bussche RA, Longmire JL, Baker RJ (1995). How bats achieve a small C-value: frequency of repetitive DNA in *Macrotus*. *Mammal Genome* **6**: 521-525.
- Vogt P (1992). Code domains in tandem repetitive DNA sequence structures. *Chromosoma* **101**: 585-589.
- Walsh JB (1987). Persistence of tandem arrays: Implication for satellite and simple-sequence DNAs. *Genetics* **115**: 553-567.