

Inverted Repeats and Genome Architecture Conversions of Terrestrial Isopods Mitochondrial DNA

Vincent Doublet · Quentin Helleu · Roland Raimond · Catherine Souty-Grosset · Isabelle Marcadé

Received: 17 April 2013 / Accepted: 18 September 2013 / Published online: 26 September 2013
© Springer Science+Business Media New York 2013

Abstract Mitochondrial DNA (mtDNA) is usually depicted as a circular molecule, however, there is increasing evidence that linearization of mtDNA evolved independently many times in organisms such as fungi, unicellular eukaryotes, and animals. Recent observations in various models with linear mtDNA revealed the presence of conserved inverted repeats (IR) at both ends that, when they become single-stranded, may be able to fold on themselves to create telomeric-hairpins involved in genome architecture conversions. The atypical mtDNA of terrestrial isopods (Crustacea: Oniscidea) composed of linear monomers and circular dimers is an interesting model to study genome architecture conversions. Here, we present the mtDNA control region sequences of two species of the genus *Armadillidium*: *A. vulgare* and *A. pelagicum*. All features of arthropods mtDNA control regions

are present (origin of replication, poly-T stretch, GA and TA-rich blocks and one variable domain), plus a conserved IR. This IR can potentially fold into a hairpin structure and is present in two different orientations among the *A. vulgare* populations: either in one sense or in its reverse complement. This polymorphism, also observed in a single individual (heteroplasmy), might be a signature of genome architecture conversions from linear to circular monomeric mtDNA via successive opening and closing of the molecules.

Keywords Linear mtDNA · Dimers · Telomere · Heteroplasmy · Control region · Oniscidea

Introduction

Mitochondria are organelles derived from an alpha-proteobacterium that lived in symbiosis in the first eukaryote cells (Andersson et al. 1998). From this bacterial ancestor, mitochondria kept their own genome: the mitochondrial DNA (mtDNA). As for bacteria, the genome of the first mitochondria were probably circular, but after 1.5 billion years of evolution, mtDNA architecture followed various trends in the eukaryote lineages (Gray et al. 1999). Although mtDNA is generally depicted as a circular molecule, a growing number of examples shows that linear molecules appeared independently in organisms such as plants (Ryan et al. 1978; Coleman et al. 1991; Pérez-Brocal et al. 2010; Smith et al. 2010), fungi (Fukuhara et al. 1993; Forget et al. 2002; Kosa et al. 2006; Valach et al. 2011; Ma et al. 2013) and unicellular eukaryotes (Burger et al. 2000; Burger et al. 2003; de Graaf et al. 2009; Hikosaka et al. 2010; Sesterhenn et al. 2010; Barth and Berendonk 2011; Burger et al. 2013). In animals, only a few examples of

Electronic supplementary material The online version of this article (doi:10.1007/s00239-013-9587-7) contains supplementary material, which is available to authorized users.

V. Doublet · Q. Helleu · R. Raimond · C. Souty-Grosset · I. Marcadé

Equipe Ecologie Evolution Symbiose, Laboratoire Ecologie et Biologie des Interactions, UMR CNRS 7267, Université de Poitiers, 40 Avenue du Recteur Pineau, 86022 Poitiers Cedex, France

Present Address:

V. Doublet (✉)
Institut für Biologie, Allgemeine Zoologie, Martin-Luther-Universität Halle-Wittenberg, Hoher Weg 8, 06099 Halle, Germany
e-mail: vincent.doublet@zoologie.uni-halle.de

Present Address:

Q. Helleu
Laboratoire Evolution, Génomes et Spéciation, CNRS, 91198 Gif-sur-Yvette, France

fully linear mtDNA have been observed, in sponges and cnidarians (Voigt et al. 2008; Kayal et al. 2012; Smith et al. 2012; Lavrov et al. 2013).

Mechanisms of linearization of circular mtDNA are still not fully understood, however, there is evidence that mitochondrial genome architecture conversion to a linear form might be due to similar mechanisms in unrelated organisms. Indeed, the majority of mitochondrial linear molecules sequenced so far from diverse organisms revealed the presence of inverted terminal repeats (ITR) on linear molecules ends (Morin and Cech 1988; Dinouël et al. 1993; Vahrenholz et al. 1993; Kairo et al. 1994; Burger et al. 2000; Forget et al. 2002; Kosa et al. 2006; Shao et al. 2006; Smith and Lee 2008; Voigt et al. 2008; Fricova et al. 2010; Hikosaka et al. 2010; Pérez-Brocail et al. 2010; Smith et al. 2010; Valach et al. 2011; Kayal et al. 2012; Smith et al. 2012; Burger et al. 2013)—but there is a counterexample in *Paramecium caudatum* (Barth and Berendonk 2011). These ITRs are palindromic repeats thought to form telomeric-hairpins (t-hairpins) structures which covalently close the double-stranded DNA linear molecules. T-hairpins permit to overcome the problem of the entire replication of linear mtDNA molecules (Nosek et al. 2004). In some organisms, such as jakobid protists and maize, the inverted repeats (IR) originate from foreign plasmid DNA inserted into circular mtDNA, inducing the genome linearization (Schardl et al. 1984; Burger et al. 2013). Although the origin of such plasmids is not known, Nosek et al. (2006) described this foreign DNA as selfish elements that convert circular chromosomes into linear, providing a telomere that solves the problems associated with linearization.

Few organisms within plants and yeasts have both linear and circular mtDNA molecules co-occurring in mitochondria (Dinouël et al. 1993; Backert et al. 1997). In animals, only one model combines linear and circular mtDNA molecules: the atypical mitochondrial genome of terrestrial isopods (Crustacea: Oniscidea) (Raimond et al. 1999; Doublet et al. 2012). This genome is composed of two molecular forms: linear monomeric molecules and “head-to-head” palindromic circular dimers (Fig. 1). Both molecular forms present the same nucleotides and genes sequence [except one point mutation (Marcadé et al. 2007; Doublet et al. 2008)] and are believed to be generated from one to another via a permanent mechanism of linearization and circularization. In order to find what facilitates genome architecture conversions in the atypical mtDNA of isopods, the sequencing of the complete mtDNA of the common backyard pill bug *Armadillidium vulgare* has recently been undertaken (Marcadé et al. 2007). Although 99 % of the genome has been successfully sequenced, the palindromic junctions of dimers and ends of linear monomers (where linearization and

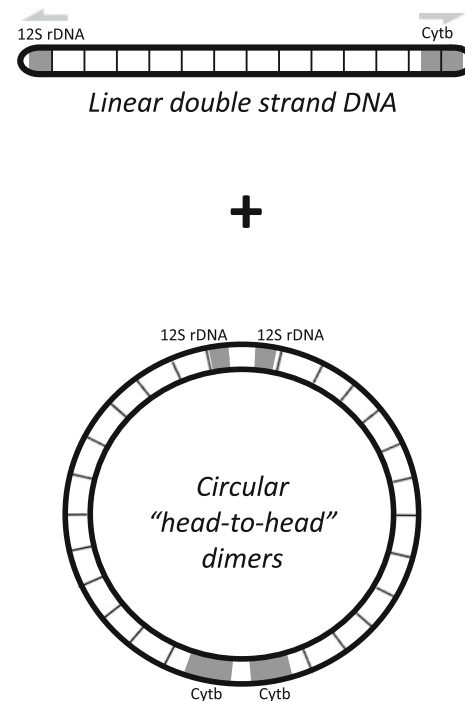


Fig. 1 Architecture of the atypical mtDNA of terrestrial isopods composed of linear monomers and “head-to-head” palindromic circular dimers. Gray arrows show the positions of the primers used in this study. Gray boxes show the positions of the genes surrounding the control region: 12S rDNA and cytochrome *b* (*Cytb*). Distance between thin lines = 1 kb

circularization is believed to occur) were not completely sequenced due to the formation of secondary structures during amplification processes—a problem also encountered recently by Kilpert et al. (2012) by re-sequencing *A. vulgare* mtDNA.

To obtain the missing sequence of *A. vulgare* mtDNA, we then tried to amplify it with a new primer pair, assuming the presence of circular monomers with “head-to-tail” junctions, the “typical” architecture of animal mtDNA, never observed in terrestrial isopods mtDNA (Raimond et al. 1999; Doublet et al. 2012). With this new approach, we successfully amplified and sequenced what constitutes the mtDNA control region. Here, we present the mtDNA control region sequences of two species of the genus *Armadillidium*, from 10 populations of *A. vulgare* and one population of *A. pelagicum*. In these two species, all features generally present in arthropods mtDNA control regions have been observed (origin of replication, poly-T stretch, GA- and TA-rich blocks and one variable domain), plus a conserved IR present in two orientations: either in one sense or in its reverse complement. We discuss the composition of this control region and the implication of the IR in genome architecture conversions in terrestrial isopods mtDNA.

Table 1 Samples origin. Type “A” and type “B” referred to the inverted repeat (IR) orientation

Species	Name	Origin	Type	Nb of sequences (tested samples)	Genbank Accession number
<i>Armadillidium vulgare</i>	PA	Porto Alegre, Brazil	A	2 (2)	JX840157
	SP	São Paulo, Brazil	A	1 (6)	JX840158
	TU	Tunis, Tunisia	A	1 (3)	JX840159
	ME	Mentrida, Spain	A	2 (16)	JX840160
	TE	Tenerife, Spain	A	3 (5)	JX840161
	VA	Vancouver, Canada	A	1 (5)	JX840162
	HE	Helsingør, Denmark	B	2 (53)	JX840163
	CB	Celles-sur Belle, France	B	1 (19)	JX840164
	AR	Archigny, France	B	1 (2)	JX840165
	MT	Mas Thibert, France	B	3 (6)	JX840166
				Total = 17 (117)	
<i>Armadillidium pelagicum</i>		Tunis, Tunisia		2 (2)	JX840167
					JX840168

Materials and Methods

Sampling, mtDNA Amplifications and Cloning

Animals were sampled in the field or from populations reared in the laboratory (Table 1). Total DNA was extracted as previously described (Bouchon et al. 1998). MtDNA control regions were amplified by PCR with the primers 12SCR (5'-GAGATAAGTCGTAACAAAGTAG) and *Cytb*CR (5'-CTACCTTGAGGTCAAATATC), designed from conserved regions of the 12S rDNA and Cytochrome *b* (*Cytb*) gene sequences. In *A. vulgare* mtDNA, these two genes are present at both ends of the linear molecules and the junctions of the palindromic dimers (Fig. 1, Marcadé et al. 2007). Primers were obtained by an alignment of *A. vulgare* mtDNA partial genome sequence (NCBI accession number EF643519.2), with two other isopods mitochondrial genomes: *Ligia oceanica* (DQ442914.1) and *Idotea baltica* (DQ442915.1). To overcome the problem of PCR artefacts, two independent PCR methods were performed. Amplifications of control regions were first realized with GoTaq (Promega) using the following program: 3 min at 95 °C, followed by 25 cycles of 30 s at 95 °C, 1 min at 54 °C and 1 min 30 at 72 °C, and a final step of 5 min at 72 °C. To confirm our results, a subset of samples were then amplified using the Phusion High-Fidelity PCR Master Mix (New England Biolabs) with the following program: 30 s at 98 °C, followed by 35 cycles of 15 s at 98 °C, 30 s at 55 °C and 30 s at 72 °C, and a final step of 10 min at 72 °C. Cloning was performed for one heteroplasmic sample using pGEM-T Easy Vector Systems kit (Promega). Purified PCR products and clones were sequenced with the BigDye Terminator kit (Applied Biosystems) and analyzed on an ABI Prism 3130 Genetic Analyzer.

Sequence Analysis

For each sequence obtained, we first checked and assembled both strand sequences using Staden (Staden 1996). We aligned these sequences with the annotated partial genome sequence of *A. vulgare* (EF643519.2) and removed the part of the sequence that belongs to the *Cytb* gene. Likewise, we used the software ARWEN (Laslett and Canbäck 2008) to localize the tryptophan transfer RNA (*tRNA*^{Trp}) gene, flanking the 12sRNA gene, and removed its sequence. The control region sequence was then defined between the last nucleotide of *tRNA*^{Trp} and the stop codon of *Cytb* (these two genes are in opposite strands).

Sequence alignments and dot plot analyses were made using BioEdit software (Hall 1999). Secondary structures of DNA have been detected with RNAfold software (Hofacker 2003), using DNA parameters. Only the most likely secondary structures indicated by a color ranking probability have been considered in this study. Other mtDNA control region features have been identified by eye inspection of sequences, following description of Zhang and Hewitt (1997) and Kuhn et al. (2008). A search for restriction enzyme on the IR sequence has been done using the platform NEBcutter V2.0 (New England Biolabs, <http://tools.neb.com/NEBcutter2/>).

Location and Orientation of the Origin of Replication

To locate the origin of replication, cumulative GC-skew of the now completed *A. vulgare* mtDNA sequence of has been calculated using the GenSkew software (<http://genskew.csb.univie.ac.at/>) with a window size and a step size of 100 and 20 nucleotides, respectively. This GC-skew

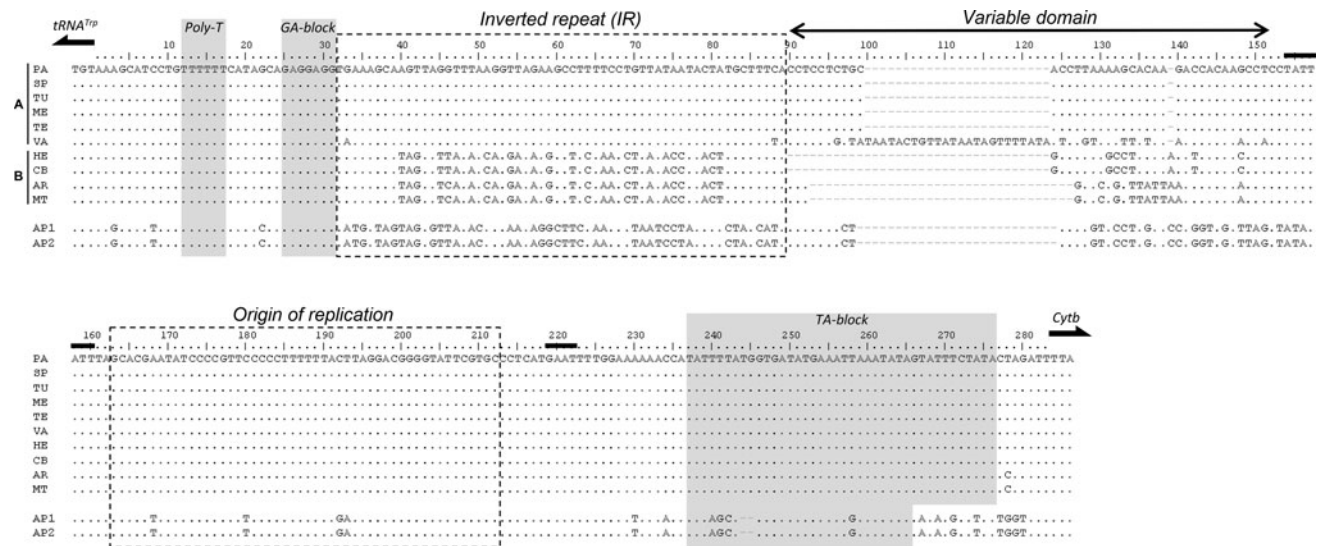


Fig. 2 Alignment of *A. vulgare* and *A. pelagicum* mtDNA control region sequences. Flanking arrows with tRNA^{Trp} and Cytb show the localisation of these respective genes (note the 12S rDNA gene is located next to the tRNA^{Trp}). The IR and the origin of replication are shown in frames. Black lines show the conserved flanking sequences

TATT and G(A)nT surrounding the origin of replication. The arrows locate the variable domain. Poly-T stretch, GA-block, and TA-block are shown in gray boxes. Numeration of bases is restricted to the control region sequence. A and B refer to the type “A” and type “B” orientations of the IR (see Fig. 5)

statistics reflect the strand-bias of nucleotide composition in the genome, using the following formula: $(G-C)/(G+C)$ which gives the relative proportion of Guanine (G) compare to Cytosine (C) in a sequence (Perna and Kocher 1995). This strand-bias varies gradually along a genome and the region with the lower GC-skew value indicates where the origin of replication is located. The GC-skew is generally negative for the heavy strand of arthropod mtDNA, but variants exist, reflecting a reverse orientation of the origin of replication (Hassanin et al. 2005). We compare the results obtained for the mitochondrial genomes of *A. vulgare* and *Daphnia pulex* (NC_000844.1), an aquatic crustacean species.

Results

Amplification Success and Nucleotide Composition

Control region sequences have been obtained from *A. vulgare* samples from 10 populations and from two *A. pelagicum* samples from one population (Table 1). However, we successfully amplified and sequenced the mtDNA control region of only 17 individuals of *A. vulgare* out of 117 tested, which represent an amplification success of only 14.5 %. No polymorphism have been observed within populations, except for the population from Mas Thibert (MT), France (see below), but no more than three individuals have been successfully sequenced for each population. The Figure 2 shows an alignment of control

region sequences obtained from each population of *A. vulgare* and the two individuals of *A. pelagicum*. The mtDNA control region sequences from *A. vulgare* have a size ranging from 249 to 282 bp (base pairs) and a size of 255 bp for *A. pelagicum*. Proportions of A and T are of 64.2 % (± 0.02) for *A. vulgare* and 63.1 % for *A. pelagicum*. However, compared to the proportion of AT in the complete genome of *A. vulgare* (71.2 %), the control region is less AT-rich. Control region sequences are very similar within *A. vulgare* populations with a degree of similarity ranging from 74 to 100 % (Table 2). The degree of similarity of the control region sequence of *A. pelagicum* with the different mitotypes of *A. vulgare* varies from 62 to 68 %.

Origin of Replication

The origins of replication of *A. vulgare* and *A. pelagicum* mtDNAs have been identified via a secondary structure analysis of the sequences. They present a GC-rich sequence and are characterized by two conserved flanking motifs, a TATT-like sequence on 5' and a GA(n)T motif in 3' (Fig. 3). Sequences of the origins of replication are fully conserved among *A. vulgare* populations (Fig. 2). The origin of replication of *A. pelagicum* mtDNA presents a similar secondary structure and location in the genome compared to *A. vulgare* (Figs. 2, 3). Analysis of the cumulative GC-skew along *A. vulgare* mtDNA sequence confirmed the location of the origin of replication (Fig. 4). The positive value of the cumulative GC-skew in *A. vulgare*

Table 2 Similarity matrix between control region sequences of *A. vulgare* populations

	Type “A”						Type “B”			
	PA	SP	TU	ME	TE	VA	HE	CB	AR	MT
PA	x	–	–	–	–	–	–	–	–	–
SP	1	x	–	–	–	–	–	–	–	–
TU	1	1	x	–	–	–	–	–	–	–
ME	1	1	1	x	–	–	–	–	–	–
TE	1	1	1	1	x	–	–	–	–	–
VA	0.87	0.87	0.87	0.87	0.87	x	–	–	–	–
HE	0.83	0.83	0.83	0.83	0.83	0.74	x	–	–	–
CB	0.83	0.83	0.83	0.83	0.83	0.74	1	x	–	–
AR	0.80	0.80	0.80	0.80	0.80	0.74	0.87	0.87	x	–
MT	0.80	0.80	0.80	0.80	0.80	0.74	0.87	0.87	1	x

Type “A” and type “B” referred to the inverted repeat (IR) orientation

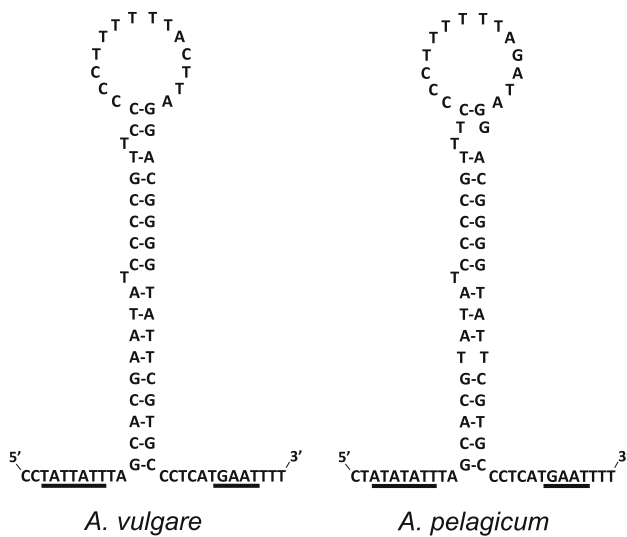


Fig. 3 Secondary structures of the origins of replication of *A. vulgare* and *A. pelagicum* and their conserved flanking sequences TATT and G(A)nT *underlined*

also shows that the origin of replication is in a reverse orientation, compared to other crustacean *D. pulex*.

Poly-T stretch, GA- and TA-block, and the Variable Domain

Three other common motifs of arthropod mtDNA control regions have been identified in *A. vulgare* and *A. pelagicum*. These motifs are presented in gray boxes in the Fig. 2 and are: (i) a poly-T stretch, here composed of 6 thymine in both species, (ii) a GA-rich block (GA-block), fully conserved between the two species *A. vulgare* and *A. pelagicum*, and (iii) a long TA-block located nearby the *Cytb* gene, with a conserved sequence within *A. vulgare* but a different length and sequence in *A. pelagicum*. A

variable domain is also present in the control regions of the two species, located between the IR (see below) and the origin of replication (Fig. 2). This region concentrates the maximum of variability within *A. vulgare* and between *A. vulgare* and *A. pelagicum*, both in sequence and size. The control region of the individual from Vancouver (VA) presents the largest variable domain, due to the three repeats of the sequence CTGTTATAATA.

Inverted Repeat (IR)

In all control region sequences from *A. vulgare* and *A. pelagicum* samples, a conserved IR has been found between the GA-block and the variable domain. This IR can potentially fold into a hairpin structure according to the software RNAfold (Fig. 2). Two different sequences of IR have been observed in *A. vulgare*: a type “A”, present in six populations, and a type “B”, present in the other four populations and the two samples of *A. pelagicum* (Fig. 5). A dot-plot analysis of the control region sequences shows that the two types of IR present nevertheless the same nucleotide sequence, but differ only by their orientation (Fig. 6). Indeed, sequences of type “A” correspond exactly to the reverse complements of the type “B”. The secondary structure analysis shows that type “A” IR can be potentially longer, including the GA-block (data not shown). Almost no variability has been observed within IR sequences between *A. vulgare* populations, except for the population of VA in which two point mutations do not affect the secondary structure, and for the French populations Archigny (AR) and Mas Thibert (MT) which present a point mutation in the inside loop of the IR.

To account for possible PCR artefact, the control region sequence was amplified independently with a high-fidelity DNA polymerase for two individuals of the population Tenerife (TE). Identical sequences were obtained,

Fig. 4 Graphical representation of the cumulative GC-skew along the mitochondrial genome sequences of *A. vulgare* (EF643519.3) and *D. pulex* (NC_000844.1), using the software GenSkew (<http://genskew.csb.univie.ac.at/>), with a window size and a step size of 100 and 20 nucleotides, respectively

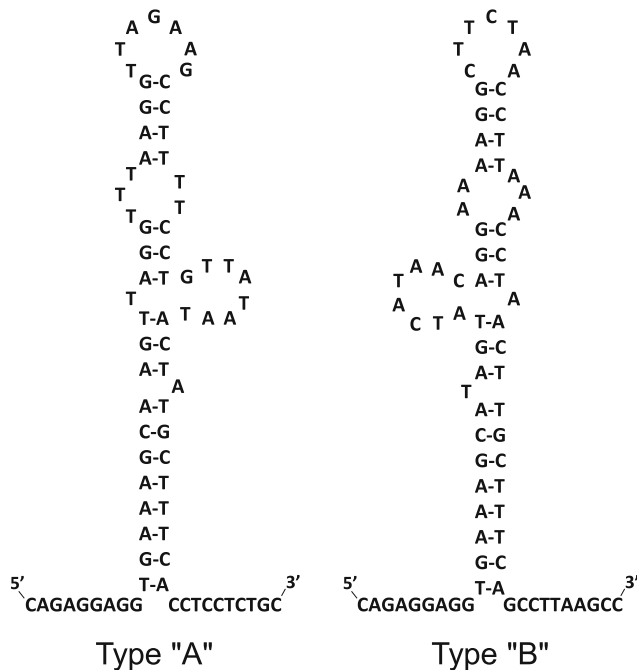
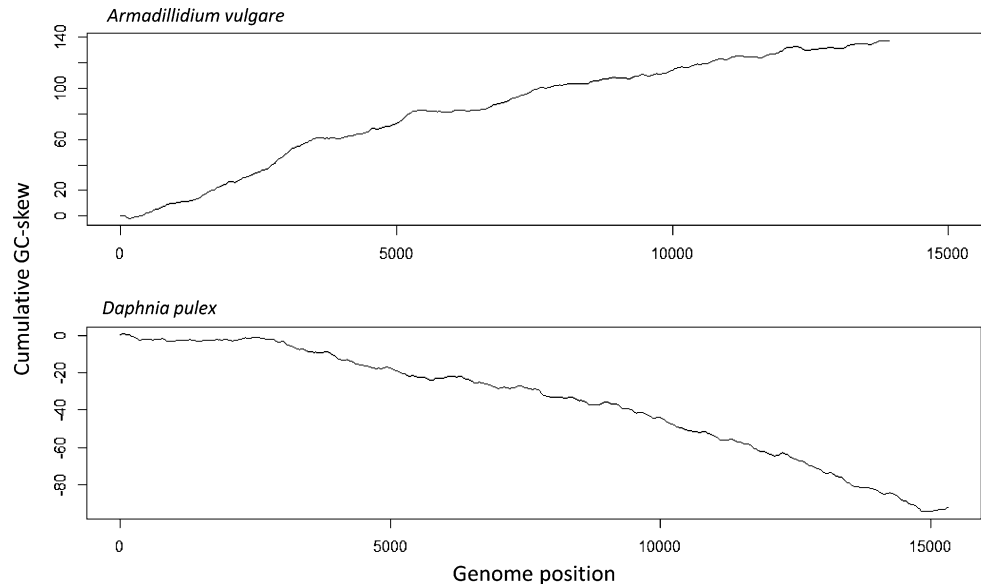


Fig. 5 Secondary structures of the IR of type “A” and type “B”. Sequences from the samples of Porto Alegre (PA) and Archigny (AR) have been used for type “A” and type “B”, respectively

confirming the presence of the IR. However, this method does not rule out the hypothesis that IR sequences obtained here were generated from an overlap of two linear molecules during PCR, giving the impression of a continuous sequence.

Heteroplasmy

A heteroplasmy has been detected in the control region of a sample from the population of MT. Out of the 14

sequenced clones from the PCR products, 13 control region sequences were type “B” (as MT is represented in Fig. 2) and one clone was type “A” with a 43 bases repeated sequence inserted into the variable domain (Supplemental material S1).

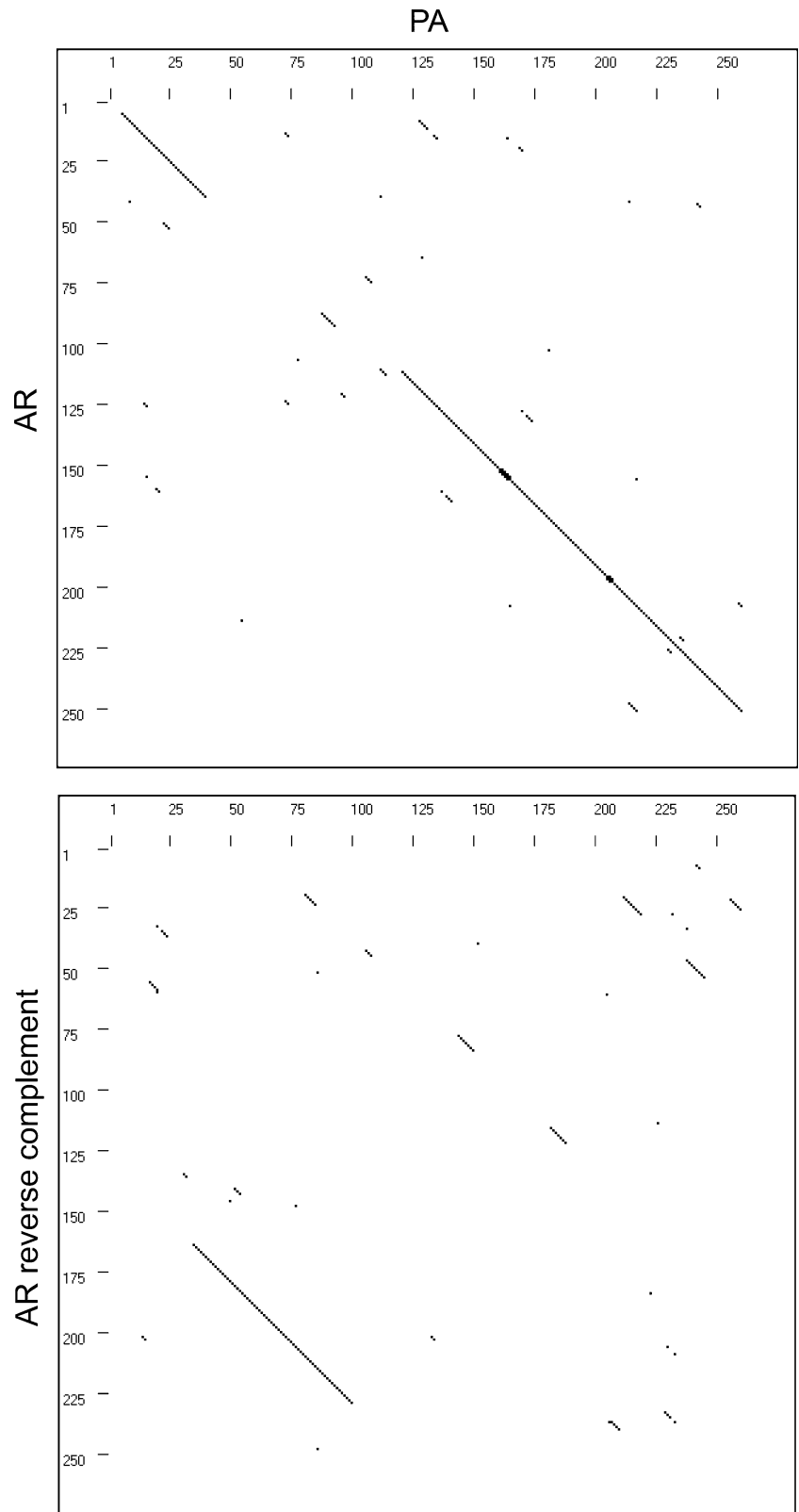
Discussion

Armadillidium vulgare mtDNA Control Region

By sequencing its control region, the mitochondrial genome sequence of *A. vulgare* is now complete. The previously published partial sequence (Marcadé et al. 2007) has been updated in the GenBank database (EF643519.3) with a sequence from a sample of the same population (Celles-sur-Belle, CB). The total size of *A. vulgare* mitochondrial genome is 13,939 bp, the shortest described so far in crustaceans and one of the shortest in arthropods. The size reduction of *A. vulgare* mtDNA is due to short intergenic spacers and overlapping gene sequences but also to the reduced control region, as observed in other short mitochondrial genomes (Yuan et al. 2010; Lin et al. 2012). In *A. vulgare* mtDNA, the control region has a very small size for animal mtDNA, ranging from 249 to 282 bp, due to a size reduction of its variable domain.

All the elements generally present in arthropod mtDNA control regions have been found in *A. vulgare* and *A. pelagicum* control regions (origin of replication, poly-T stretches, GA- and TA-blocks), confirming their function. The analysis of the cumulative GC-skew along *A. vulgare* mtDNA also confirmed the location of the origin of replication in this part of the genome, as well as its reverse orientation in the genome, a characteristic of isopod mtDNA (Kilpert et al. 2012).

Fig. 6 Dot plot analysis showing similarity of control region nucleotide sequences between a samples of type “A” from Porto Alegre (PA, *x* axis) and a sample of type “B” from Archigny (AR, *y* axis, *above plot*) and the reverse complement of the sequence from Archigny (AR, *y* axis, *below plot*). Numbers represent bases positions



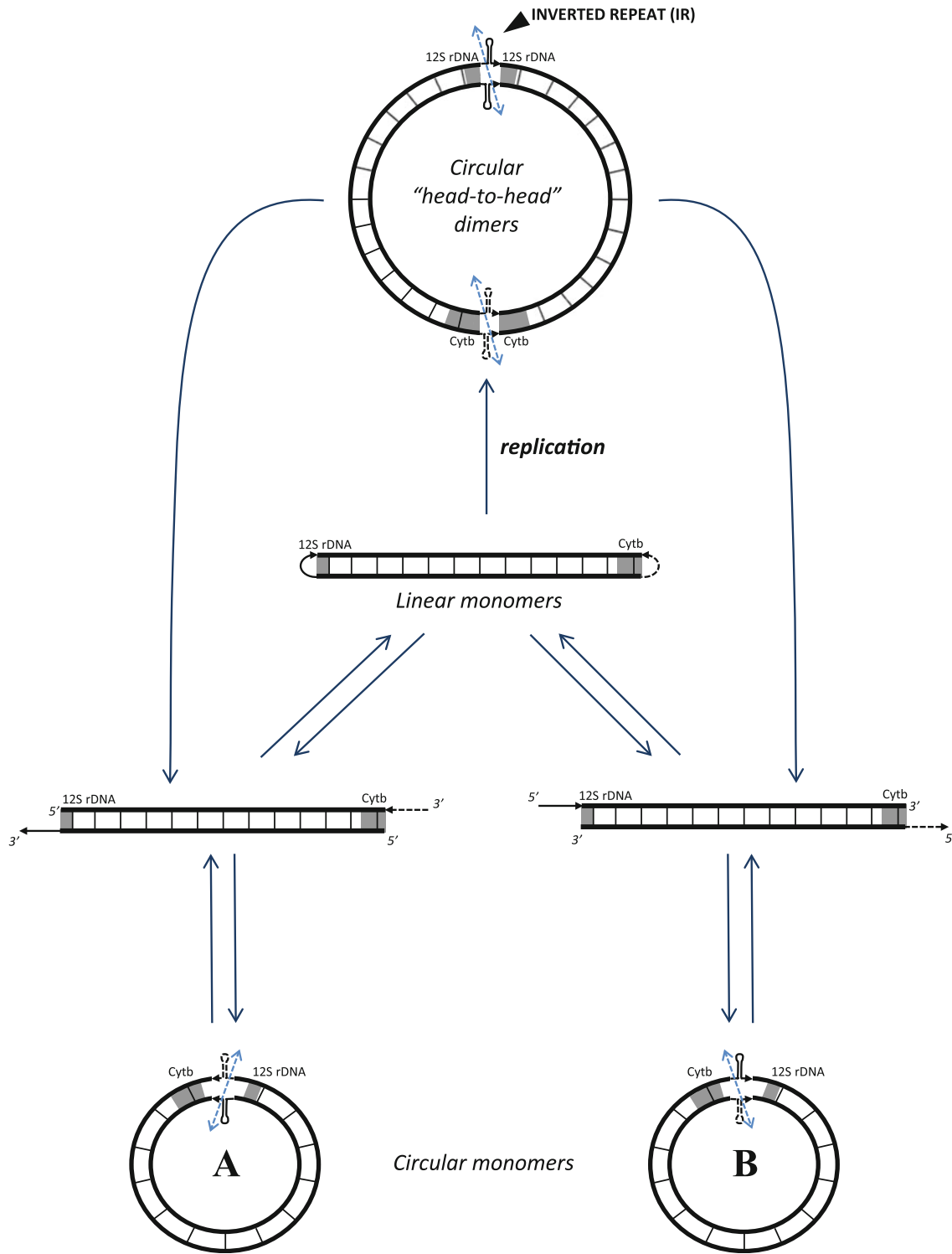


Fig. 7 Genome architecture conversions scheme in the atypical mtDNA of terrestrial isopods. *A* and *B* represents the type “A” and type “B” variations of the control regions. *Solid arrows* show the conversion from one architecture to the other. *Dotted arrows* represent the scission of the palindromic “head-to-head” dimers at the location

of the IR, either by an exonuclease or by double-strand DNA breaks. *Gray boxes* show the positions of the genes surrounding the control region: 12S rDNA and cytochrome *b* (*Cytb*). Distance between thin lines in molecules = 1 kb

The IR and Genome Architecture Conversions

In addition to all the typical elements of mtDNA control region, an IR was also found in all *A. vulgare* and *A. pelagicum* sequences. The nucleotide sequence of this IR and the potential secondary structure conserved across *A. vulgare* populations suggest that this element has an essential function—alternatively, these sequences could also be due to PCR artefact (see “Results” section), or nuclear copies of mtDNA, but the extreme sequence conservation across distant populations of *A. vulgare* makes this latter hypothesis rather unlikely. IRs are common elements in invertebrates mtDNA non coding regions (Mundy et al. 1996; Arunkumar and Nagaraju 2006; Rondan Dueñas et al. 2006; Lukić-Bilela et al. 2008; Pie et al. 2008; Brauer et al. 2012; Krebs and Bastrop 2012). They can be implicated in various functions including DNA replication (Voineagu et al. 2008). Indeed, IRs present in mtDNA control regions have been shown to be associated with the initiation of replication (Tapper and Clayton 1981; Hixson et al. 1986; Gerhold et al. 2010). The IR present in mtDNA control region of *Armadillidium* might also have another function. Among the elements described in the control region of *Armadillidium* mtDNA, the IR also appears as a good candidate for facilitating genome architecture conversions. Thus, using the results of this study and the previous literature about isopod mtDNA structure, we present here a putative model of genome architecture conversions with the IR (Fig. 7). In this model, we have located the IR at both ends of the linear monomers and at the junction of the palindromic “head-to-head” dimers. This is indeed the position of the control region revealed by the sequencing of *A. vulgare* mtDNA (Marcadé et al. 2007; Kilpert et al. 2012). At the ends of the linear molecules, the IR is represented folded into t-hairpins, a secondary structure that covalently closes the double-strand DNA. Although the presence of t-hairpins at both ends of linear mtDNA of isopods has not been directly observed, they are believed to give stability to the molecules. Linear molecules represent almost the half of the mtDNA molecules in *A. vulgare*—the other half being the circular dimers (Raimond et al. 1999). Without t-hairpins, linear molecules would be prone to exonucleases and might not be conserved. The failure to clone, amplify, and sequence the ends of the linear molecules of *A. vulgare* mtDNA (Marcadé et al. 2007) might be a mark of this physical protection that t-hairpins can provide.

T-hairpins present another major advantage to linear molecules. By closing ends of the double-stranded DNA, they permit the complete replication of linear molecules. In our model, we assume that replication of linear molecules of isopod mtDNA facilitated by t-hairpins, would generate palindromic “head-to-head” dimers (Fig. 7). This is

supported by the fact that linear monomers and “head-to-head” dimers are constitutive of isopods mtDNA and present the same nucleotide sequences (Raimond et al. 1999; Marcadé et al. 2007; Doublet et al. 2012). An identical replication model exists in yeasts with linear mtDNA, where t-hairpins provide a substrate for terminus elongation and induce the formation of palindromic “head-to-head” dimers, also called “replicative intermediate” (Dinouël et al. 1993; Nosek et al. 1995; Valach et al. 2011). Although this replication model has not been experimentally observed in isopod mtDNA, the products of replication appear similar in the yeast linear mtDNA and the *Armadillidium* mtDNA.

The maintenance of nucleotide sequence homogeneity between the two molecular forms (linear and dimers) of isopod mtDNA requires a dynamic mechanism generating one architecture from the other. Thus, linear molecules of isopod mtDNA are likely the produce of the scission of the dimers, but in *A. vulgare* mtDNA the mechanism of resolution of palindromic “head-to-head” dimers into linear molecule remains unknown. In our model (Fig. 7), we assume that a specific mechanism associated to the IR would resolve dimers into two linear molecules. These linear molecules would have then opens ends that permit either to close covalently the double-stranded DNA linear molecules and produce new dimers by replication, or the formation of another architecture (see below). IRs are known to produce genetic instability due to their ability to form secondary structure and cruciform motifs causing double-strand DNA breaks (Leach 1994; Akgun et al. 1997; Nasar et al. 2000; Lobachev et al. 2007). Another hypothesis would be the scission of the dimers by a restriction enzyme. Such mechanism has been described in the bacteria spirochete *Borrelia*, where IRs are processed by a telomere resolvase into t-hairpins (Kobryn et al. 2009). Although the nucleotide sequence of the IR present only few mutations across *A. vulgare* populations, no conserved motif has been found to match for a restriction enzyme. Thus, it is difficult to know how the dimers are split in two linear forms and experimental confirmations of one or the other mechanism would be needed.

As discussed above, the “replicative intermediate” model appears to be the principal mechanism of genome architecture conversions of *A. vulgare* mtDNA. However, the sequences obtained during this study suggest that this mtDNA is not constituted only of linear monomers and palindromic “head-to-head” dimers, as previously observed (Raimond et al. 1999). Indeed, the primers designed for the amplification of the control region and the IR bind in opposite directions at the two ends of the linear molecules (Fig. 1). The amplification of the control region with these primers demonstrates that alternative architectures exist in *A. vulgare* mtDNA. Interestingly, the amplification success

of the control region was very low. This suggests that this alternative architecture could be rare and maybe transient, possibly not transmitted to the next generation, and thus not detectable by RFLP analyses of pooled individuals as used in past studies (Raimond et al. 1999; Marcadé et al. 2007; Doublet et al. 2012). The form of this alternative architecture is unknown. However, the amplification shows that this form presents a “head-to-tail” junction. Such junctions can be obtained in two ways. First, a “head-to-tail” junction can be made by concatenating molecules. Concatenated “head-to-tail” molecules are rather rare but have been observed in plants with linear mtDNA, where they are formed after replication of the linear molecules (Backert 2002). The second possibility is the formation of circular monomers with a “head-to-tail” junction. This form is the “typical” architecture of animal mtDNA (Boore 1999), but has never been observed in isopods with linear mtDNA. Both concatemers and circular monomers can occur in *Armadillidium* mtDNA. However, for clarity, although we do not know the form of this alternative molecule, we have represented the alternative architecture in our model by circular monomers (Fig. 7). Whatever form this alternative architecture is, its formation required the presence of linear molecules with open ends. Thus, we believe that the formation of molecules with “head-to-tail” junctions might occur after the scission of dimers in two linear molecules with open ends. The observation of two different orientations of the IR across *A. vulgare* populations supports the open ends hypothesis. In our model (Fig. 7), we show how linear molecules with open ends in two different orientations can be generated by scission of the dimers. When these linear molecules convert into the alternative architecture, two “head-to-tail” junction sequences are generated corresponding to the two types of IR with reversed orientations (A and B, Fig. 7). Interestingly, a reversed orientation of the IR was also observed in yeasts with linear mtDNA (Dinouël et al. 1993). In this case, authors postulated an equivalent mechanism, called “flip-flop” inversion mechanism, which allows switching from “head-to-head” dimers to “head-to-tail” dimers via the formation of transient open ends linear molecules.

The frequency of formation of the alternative architecture with “head-to-tail” junctions, as well as its evolutionary consequences remains unclear. Our sample size is too small to tell if multiple reversions of the IR happened during the evolutionary history of *A. vulgare* mtDNA and additional sampling would be needed. However, the presence of an individual with the two orientations (heteroplasmy) suggests that this reversion might occur relatively frequently. In previous studies, we observed a correlation between the presence of linear mtDNA molecules and the constitutive heteroplasmy in terrestrial isopods (Doublet et al. 2008; Doublet et al. 2012), which might be a reason why the complex mechanism of genome architecture

conversions of terrestrial isopod mtDNA is conserved. However, the relation between the heteroplasmy and the presence of different genome architecture remains to be investigated.

Conclusion

The presence and the maintenance of different molecular architecture co-occurring in the atypical mtDNA of terrestrial isopods make this genome an interesting model to study genome architecture conversions. Genome architecture conversions appears as a key mechanism for replication and evolution of isopod mtDNA, and with the present results we postulate that genome architecture conversions might be possible due to a specific IR sequence conserved in all mtDNA control region sequenced. Similar mechanisms of linearization of circular genomes due to the presence of IR have been observed in many other organisms with linear mtDNA (plants, yeasts, animals). Such convergent evolution toward linearization shows how organelle genome architectures could be flexible.

Acknowledgments We are grateful to Daniel Guyonnet for technical assistance, Yves Caubet, Sébastien Verne, and Maurício P. Almerão (Universidade Federal do Rio Grande do Sul, Brazil) for providing samples, Julian Reynolds for the English proofreading, and anonymous reviewers for their suggestions on the manuscript. VD Ph.D. was granted by the Region Poitou-Charentes.

References

- Akgun E, Zahn J, Baumes S, Brown G, Liang F, Romanienko PJ, Lewis S, Jasin M (1997) Palindrome resolution and recombination in the mammalian germ line. *Mol Cell Biol* 17:5559
- Andersson SGE, Zomorodipour A, Andersson JO, Sicheritz-Pontén T, Alsmark UCM, Podowski RM, Näslund AK, Eriksson AS, Winkler HH, Kurland CG (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396:133
- Arunkumar KP, Nagaraju J (2006) Unusually long palindromes are abundant in mitochondrial control regions of insects and nematodes. *PLoS One* 1:e110
- Backert S (2002) R-loop-dependent rolling-circle replication and a new model for DNA concatemer resolution by mitochondrial plasmid mp1. *EMBO J* 21:3128
- Backert S, Nielsen BL, Börner T (1997) The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. *Trends Plant Sci* 2:477
- Barth D, Berendonk TU (2011) The mitochondrial genome sequence of the ciliate *Paramecium caudatum* reveals a shift in nucleotide composition and codon usage within the genus *Paramecium*. *BMC Genomics* 12:272
- Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767
- Bouchon D, Rigaud T, Juchault P (1998) Evidence for widespread *Wolbachia* infection in isopod crustaceans: molecular identification and host feminization. *Proc Biol Sci* 265:1081

- Brauer A, Kurz A, Stockwell T, Baden-Tillson H, Heidler J, Wittig I, Kaufenstein S, Mebs D, Stöcklin R, Remm M (2012) The mitochondrial genome of the venomous cone snail *Conus consors*. PLoS One 7:e51528
- Burger G, Zhu Y, Littlejohn TG, Greenwood SJ, Schnare MN, Lang BF, Gray MW (2000) Complete sequence of the mitochondrial genome of *Tetrahymena pyriformis* and comparison with *Paramecium aurelia* mitochondrial DNA. J Mol Biol 297:365
- Burger G, Forget L, Zhu Y, Gray MW, Lang BF (2003) Unique mitochondrial genome architecture in unicellular relatives of animals. Proc Natl Acad Sci USA 100:892
- Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-rich mitochondrial genomes throughout Jakobid protists. Genome Biol Evol 5:418
- Coleman AW, Thompson WF, Goff LJ (1991) Identification of the mitochondrial genome in the chrysophyte alga *Ochromonas danica*. J Protozool 38:129
- de Graaf RM, van Alen TA, Dutilh BE, Kuiper JWP, van Zoggel HJAA, Huynh MB, Görtz H-D, Huynen MA, Hackstein JHP (2009) The mitochondrial genomes of the ciliates *Euplotes minuta* and *Euplotes crassus*. BMC Genomics 10:514
- Dinouël N, Drissi R, Miyakawa I, Sor F, Rousset S, Fukuhara H (1993) Linear mitochondrial DNAs of yeasts: closed-loop structure of the termini and possible linear-circular conversion mechanisms. Mol Cell Biol 13:2315
- Doublet V, Souty-Grosset C, Bouchon D, Cordaux R, Marcadé I (2008) A thirty million year-old inherited heteroplasmy. PLoS One 3:e2938
- Doublet V, Raimond R, Grandjean F, Lafitte A, Souty-Grosset C, Marcadé I (2012) Widespread atypical mitochondrial DNA structure in isopods (Crustacea, Peracarida) related to a constitutive heteroplasmy in terrestrial species. Genome 55:234
- Forget L, Ustinova J, Wang Z, Huss VA, Lang BF (2002) *Hyaloraphidium curvatum*: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. Mol Biol Evol 19:310
- Fricova D, Valach M, Farkas Z, Pfeiffer I, Kucsera J, Tomaska L, Nosek J (2010) The mitochondrial genome of the pathogenic yeast *Candida subhashii*: GC-rich linear DNA with a protein covalently attached to the 5' termini. Microbiology 156:2153
- Fukuhara H, Sor F, Drissi R, Dinouël N, Miyakawa I, Rousset S, Viola AM (1993) Linear mitochondrial DNAs of yeasts: frequency of occurrence and general features. Mol Cell Biol 13:2309
- Gerhold JM, Aun A, Sedman T, Jöers P, Sedman J (2010) Strand invasion structures in the inverted repeat of *Candida albicans* mitochondrial DNA reveal a role for homologous recombination in replication. Mol Cell 39:851
- Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. Science 283:1476
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95
- Hassanin A, Leger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. Syst Biol 54:277
- Hikosaka K, Watanabe Y-I, Tsuji N, Kita K, Kishine H, Arisue N, Palacpac NMQ, Kawasu S-I, Sawai H, Horii T, Igarashi I, Tanabe K (2010) Divergence of the mitochondrial genome structure in the apicomplexan parasites, *Babesia* and *Theileria*. Mol Biol Evol 27:1107
- Hixson JE, Wong TW, Clayton DA (1986) Both the conserved stem-loop and divergent 5'-flanking sequences are required for initiation at the human mitochondrial origin of light-strand DNA replication. J Biol Chem 261:2384
- Hofacker IL (2003) Vienna RNA secondary structure server. Nucleic Acids Res 31:3429
- Kairo A, Fairlamb AH, Gobright E, Nene V (1994) A 7.1 kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. EMBO J 13:898
- Kayal E, Bentlage B, Collins AG, Kayal M, Pirro S, Lavrov DV (2012) Evolution of linear mitochondrial genomes in medusozoan cnidarians. Genome Biol Evol 4:1
- Kilpert F, Held C, Podsiadlowski L (2012) Multiple rearrangements in mitochondrial genomes of Isopoda and phylogenetic implications. Mol Phylogenet Evol 64:106
- Kobryn K, Briffotiaux J, Karpov V (2009) Holliday junction formation by the *Borrelia burgdorferi* telomere resolvase, ResT: implications for the origin of genome linearity. Mol Microbiol 71:1117
- Kosa P, Valach M, Tomáška L, Wolfe KH, Nosek J (2006) Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. Nucleic Acids Res 34:2472
- Krebs L, Bastrop R (2012) The mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): a new gene order and non-neutral sequence evolution of tandem repeats in the control region. Comp Biochem Physiol D 7:201–211
- Kuhn K, Streit B, Schwenk K (2008) Conservation of structural elements in the mitochondrial control region of *Daphnia*. Gene 420:107
- Laslett D, Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24:172
- Lavrov DV, Pett W, Voigt O, Wörheide G, Forget L, Lang BF, Kayal E (2013) Mitochondrial DNA of *Clathrina clathrus* (Calcarea, Calcinea): six linear chromosomes, fragmented rRNAs, tRNA editing, and a novel genetic code. Mol Biol Evol 30(4):865–880
- Leach DR (1994) Long DNA palindromes, cruciform structures, genetic instability and secondary structure repair. Bioessays 16:893
- Lin F-J, Liu Y, Sha Z, Tsang LM, Chu KH, Chan T-Y, Liu R, Cui Z (2012) Evolution and phylogeny of the mud shrimps (Crustacea: Decapoda) revealed from complete mitochondrial genomes. BMC Genomics 13:631
- Lobachev KS, Rattray A, Narayanan V (2007) Hairpin- and cruciform-mediated chromosome breakage: causes and consequences in eukaryotic cells. Front Biosci 12:4208
- Lukić-Bilela L, Brandt D, Pojskić N, Wiens M, Gamulin V, Muller WEG (2008) Mitochondrial genome of *Suberites domuncula*: palindromes and inverted repeats are abundant in non-coding regions. Gene 412:1
- Ma L, Huang D-W, Cuomo CA, Sykes S, Fantoni G, Das B, Sherman BT, Yang J, Huber C, Xia Y, Davey E, Kutty G, Bishop L, Sassi M, Lempicki RA, Kovacs JA (2013) Sequencing and characterization of the complete mitochondrial genomes of three *Pneumocystis* species provide new insights into divergence between human and rodent *Pneumocystis*. FASEB J 27(5):1962–1972
- Marcadé I, Cordaux R, Doublet V, Debenest C, Bouchon D, Raimond R (2007) Structure and evolution of the atypical mitochondrial genome of *Armadillidium vulgare* (Isopoda, Crustacea). J Mol Evol 65:651
- Morin GB, Cech TR (1988) Mitochondrial telomeres: surprising diversity of repeated telomeric DNA sequences among six species of *Tetrahymena*. Cell 52:367
- Mundy NI, Winchell CS, Woodruff DS (1996) Tandem repeats and heteroplasmy in the mitochondrial DNA control region of the loggerhead shrike (*Lanius ludovicianus*). J Hered 87:21
- Nasar F, Jankowski C, Nag DK (2000) Long palindromic sequences induce double-strand breaks during meiosis in yeast. Mol Cell Biol 20:3449

- Nosek J, Dinouël N, Kováč L, Fukuhara H (1995) Linear mitochondrial DNAs from yeasts: telomeres with large tandem repetitions. *Mol Gen Genet* 247:61
- Nosek J, Tomáška L, Kucejová B (2004) The chromosome end replication: lessons from mitochondrial genetics. *J Appl Biomed* 2:71
- Nosek J, Kosa P, Tomaska L (2006) On the origin of telomeres: a glimpse at the pre-telomerase world. *Bioessays* 28:182
- Pérez-Brocal V, Shahar-Golan R, Clark CG (2010) A linear molecule with two large inverted repeats: the mitochondrial genome of the stramenopile *Proteromonas lacertae*. *Genome Biol Evol* 2:257
- Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol* 41:353
- Pie MR, Oliveira-Neto JF, Boeger WA, Ostrensky A, Baggio RA (2008) The organization of the mitochondrial control region in 2 Brachyuran Crustaceans: *Ucides cordatus* (Ocypodidae) and *Cardisoma guanhumi* (Gecarcinidae). *J Hered* 99:432
- Raimond R, Marcadé I, Bouchon D, Rigaud T, Bossy JP, Souty-Grosset C (1999) Organization of the large mitochondrial genome in the isopod *Armadillidium vulgare*. *Genetics* 151:203
- Rondan Dueñas JC, Gardenal CN, Albrieu Llinás G, Panzetta-Dutari GM (2006) Structural organization of the mitochondrial DNA control region in *Aedes aegypti*. *Genome* 49:931
- Ryan R, Grant D, Chiang KS, Swift H (1978) Isolation and characterization of mitochondrial DNA from *Chlamydomonas reinhardtii*. *Proc Natl Acad Sci USA* 75:3268
- Schardl C, Lonsdale DM, Pring DR, Rose KR (1984) Linearization of maize mitochondrial chromosomes by recombination with linear episomes. *Nature* 310:292
- Sesterhenn TM, Slaven BE, Keely SP, Smulian AG, Lang BF, Cushion MT (2010) Sequence and structure of the linear mitochondrial genome of *Pneumocystis carinii*. *Mol Genet Genomics* 283:63
- Shao Z, Graf S, Chaga OY, Lavrov DV (2006) Mitochondrial genome of the moon jelly *Aurelia aurita* (Cnidaria, Scyphozoa): a linear DNA molecule encoding a putative DNA-dependent DNA polymerase. *Gene* 381:92
- Smith DR, Lee RW (2008) Mitochondrial genome of the colorless green alga *Polytomella capuana*: a linear molecule with an unprecedented GC content. *Mol Biol Evol* 25:487
- Smith DR, Hua J, Lee RW (2010) Evolution of linear mitochondrial DNA in three known lineages of *Polytomella*. *Curr Genet* 56:427
- Smith DR, Kayal E, Yanagihara AA, Collins AG, Pirro S, Keeling PJ (2012) First complete mitochondrial genome sequence from a box jellyfish reveals a highly fragmented linear architecture and insights into telomere evolution. *Genome Biol Evol* 4:52
- Staden R (1996) The Staden sequence analysis package. *Mol Biotechnol* 5:233
- Tapper DP, Clayton DA (1981) Mechanism of replication of human mitochondrial DNA. Localization of the 5' ends of nascent daughter strands. *J Biol Chem* 256:5109
- Vahrenholz C, Riemen G, Pratje E, Dujon B, Michaelis G (1993) Mitochondrial DNA of *Chlamydomonas reinhardtii*: the structure of the ends of the linear 15.8-kb genome suggests mechanisms for DNA replication. *Curr Genet* 24:241
- Valach M, Farkas Z, Fricova D, Kovac J, Brejova B, Vinar T, Pfeiffer I, Kucsera J, Tomáška L, Lang BF, Nosek J (2011) Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Res* 39:4202
- Voigt O, Erpenbeck D, Wörheide G (2008) A fragmented metazoan organellar genome: the two mitochondrial chromosomes of *Hydra magnipapillata*. *BMC Genomics* 9:350
- Voineagu I, Narayanan V, Lobachev KS, Mirkin SM (2008) Replication stalling at unstable inverted repeats: interplay between DNA hairpins and fork stabilizing proteins. *Proc Natl Acad Sci USA* 405:9936
- Yuan ML, Wei DD, Wang BJ, Dou W, Wang JJ (2010) The complete mitochondrial genome of the citrus red mite *Panonychus citri* (Acari: Tetranychidae): high genome rearrangement and extremely truncated tRNAs. *BMC Genomics* 11:597
- Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem Syst Ecol* 25:99