

# Evolution of the Metazoan Mitochondrial Replicase

Marcos T. Oliveira<sup>1,2</sup>, Jani Haukka<sup>1</sup>, and Laurie S. Kaguni<sup>1,3,\*</sup>

<sup>1</sup>Institute of Biosciences and Medical Technology, University of Tampere, Finland

<sup>2</sup>Departamento de Tecnologia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista "Júlio de Mesquita Filho," Jaboticabal, SP, Brazil

<sup>3</sup>Department of Biochemistry and Molecular Biology and Center for Mitochondrial Science and Medicine, Michigan State University

\*Corresponding author: E-mail: [lskaguni@msu.edu](mailto:lskaguni@msu.edu)

Accepted: February 26, 2015

## Abstract

The large number of complete mitochondrial DNA (mtDNA) sequences available for metazoan species makes it a good system for studying genome diversity, although little is known about the mechanisms that promote and/or are correlated with the evolution of this organellar genome. By investigating the molecular evolutionary history of the catalytic and accessory subunits of the mtDNA polymerase, pol  $\gamma$ , we sought to develop mechanistic insight into its function that might impact genome structure by exploring the relationships between DNA replication and animal mitochondrial genome diversity. We identified three evolutionary patterns among metazoan pol  $\gamma$ s. First, a trend toward stabilization of both sequence and structure occurred in vertebrates, with both subunits evolving distinctly from those of other animal groups, and acquiring at least four novel structural elements, the most important of which is the HLH-3 $\beta$  (helix-loop-helix, 3  $\beta$ -sheets) domain that allows the accessory subunit to homodimerize. Second, both subunits of arthropods and tunicates have become shorter and evolved approximately twice as rapidly as their vertebrate homologs. And third, nematodes have lost the gene for the accessory subunit, which was accompanied by the loss of its interacting domain in the catalytic subunit of pol  $\gamma$ , and they show the highest rate of molecular evolution among all animal taxa. These findings correlate well with the mtDNA genomic features of each group described above, and with their modes of DNA replication, although a substantive amount of biochemical work is needed to draw conclusive links regarding the latter. Describing the parallels between evolution of pol  $\gamma$  and metazoan mtDNA architecture may also help in understanding the processes that lead to mitochondrial dysfunction and to human disease-related phenotypes.

**Key words:** mitochondria, mitochondrial DNA replication, structural evolution, mitochondrial replicase, pol  $\gamma$ .

## Introduction

Mitochondrial DNA (mtDNA) replication is accomplished by the sole DNA polymerase found in animal mitochondria, DNA polymerase  $\gamma$  (pol  $\gamma$ ) (reviewed in Kaguni 2004), which functions as part of a larger replication machinery called the mtDNA replisome. The identity of all of the components of the mtDNA replisome is still unknown, but biochemical studies (Korhonen et al. 2003, 2004; Oliveira and Kaguni 2010, 2011) have shown that a group of mitochondrial proteins can interact functionally to promote DNA synthesis in vitro, forming the minimal mtDNA replisome: The mtDNA helicase, also known as Twinkle in humans, unwinds the duplex DNA at the replication fork; pol  $\gamma$  catalyzes nascent DNA synthesis on both the leading and lagging DNA strands; and the mitochondrial single-stranded DNA-binding protein (mtSSB) coordinates their functions while binding and stabilizing the single-stranded DNA template (Korhonen et al. 2004; Oliveira and Kaguni 2011). This scenario, as in most DNA replication

events, requires the presence of short RNA molecules for priming of DNA synthesis by the DNA polymerase, which in animal mitochondria might be achieved by the action of the mitochondrial RNA polymerase (Wanrooij et al. 2008; Fuste et al. 2010; Reyes et al. 2013) and/or the newly identified primase PrimPol (Garcia-Gomez et al. 2013).

Animal pol  $\gamma$  is a hetero-oligomeric enzyme, in most known cases: The catalytic core, pol  $\gamma$ - $\alpha$  (also known as POLG or PolGA), contains both the 5'–3' DNA polymerase and 3'–5' exonuclease activities of the holoenzyme, whereas the accessory subunit, pol  $\gamma$ - $\beta$  (or POLG2, PolGB), serves as a processivity factor, enhancing the interactions between the holoenzyme and the DNA substrate (Lewis et al. 1996; Wang et al. 1997; Carrodegua et al. 1999; Lim et al. 1999). The catalytic core is a member of the family A DNA polymerase group, to which bacterial DNA polymerase I and the catalytic core of bacteriophage T7 DNA polymerase (gp5) also belong (reviewed in Kaguni 2004); pol  $\gamma$ - $\alpha$  shares

substantial structural and functional properties with both Pol I and T7 Pol. Notably, Pol I is not the replicative DNA polymerase in bacteria. It is a single-subunit enzyme that lacks the high fidelity and processivity of the holoenzyme form of T7 Pol, which acquires these properties as a result of the association of the catalytic core with the bacterial protein thioredoxin that serves as its accessory subunit. Although the presence of pol  $\gamma$  as the mitochondrial replicase appears to be conserved among the metazoans (and other eukaryotes), important variations in its structure have been described that may impact the mode in which mtDNA is replicated. It has long been known that pol  $\gamma$  of *Drosophila melanogaster*, one of the best studied insect model organisms, is a heterodimer comprising one catalytic subunit and a single accessory subunit (Wernette and Kaguni 1986; Olson et al. 1995; Wang and Kaguni 1999). On the other hand, the human and mouse holoenzymes have a heterotrimeric conformation, consisting of one pol  $\gamma$ - $\alpha$  and a dimeric pol  $\gamma$ - $\beta$  (Carrodeguas et al. 2001; Yakubovskaya et al. 2006; Lee et al. 2009). Interestingly, the mtDNA of the nematode *Caenorhabditis elegans* appears to be replicated by a machinery containing only a single subunit pol  $\gamma$  (only pol  $\gamma$ - $\alpha$ ) (Bratic et al. 2009; Addo et al. 2010), which resembles the catalytic core in other eukaryotes, such as that of the yeast *Saccharomyces cerevisiae* (Foury 1989). The accessory subunit of pol  $\gamma$ , which indeed appears to be present only in Metazoa, has a remarkable evolutionary origin, because amino acid sequence alignments, phylogenetic inferences, and general protein structure demonstrate its homology to class II aminoacyl-tRNA synthetases (Fan et al. 1999, 2006; Carrodeguas et al. 2001; Wolf and Koonin 2001).

We sought to explore the sequence and structural diversity of pol  $\gamma$  in the animal kingdom, taking advantage of the current increase in nuclear genomes and transcriptomes for which complete sequences are available in public databases. We retrieved as many animal pol  $\gamma$ - $\alpha$  and - $\beta$  gene sequences as are available and performed in silico analyses to infer their molecular evolutionary history, taking into account the substantial biochemical and structural data reported by our group and others. Here we report the oligomeric plasticity of animal pol  $\gamma$ , the finding of new structural elements, and the distinct rates of molecular evolution for different taxa, which may reflect differences in the fundamental mechanisms of mtDNA replication. We discuss our findings in the context of mitochondrial genome diversity, structure, replication, and evolution.

## Materials and Methods

### Searches for Animal pol $\gamma$ - $\alpha$ and - $\beta$ Homologs and Multiple Sequence Alignments

TBLASTN searches (Altschul et al. 1990) in the NCBI (National Center for Biotechnology Information) nonredundant sequence database were performed using the translated

mRNA reference sequences from *Homo sapiens* (pol  $\gamma$ - $\alpha$ , NM\_001126131.1; pol  $\gamma$ - $\beta$ , NM\_007215.3) and *D. melanogaster* (pol  $\gamma$ - $\alpha$ , NM\_057473.3; pol  $\gamma$ - $\beta$ , FJ635829.1) as queries. To retrieve sequences from Porifera, Placozoa, Cnidaria, Mollusca, and Hemichordata species, complementary HMMR3 BLAST (Basic Local Alignment Search Tool) searches (Eddy 2011) were performed (<http://toolkit.lmb.uni-muenchen.de/hmmer3>, last accessed August 2014), followed by BLAST searches against the Ensembl Metazoa database (<http://metazoa.ensembl.org>, last accessed August 2014), allowing the inclusion of missing exons. Most sequences from Nematoda species were retrieved from genomic scaffolds with no gene models deposited in the 959 Nematode Genomes databank (<http://www.nematodes.org/nematodegenomes>, last accessed August 2014), followed by manual processing of the exons. A Python script utilizing functions of the Biopython library (Cock et al. 2009) was developed to extract the coding sequences and other specific information from the original files retrieved from the diverse databanks. The script can be provided by the authors upon request. Highly divergent sequences were tested for true orthology by using them as queries for BLAST searches against the human genome database and by predicting possible mitochondrial localization using the TargetP service (Emanuelsson et al. 2007). The complete set of sequences retrieved is shown in **supplementary table S1, Supplementary Material** online.

Because of the overrepresentation by mammalian and *Drosophila* sequences, this data set was reduced before the alignments were performed to diminish bias in interpretation of the results. The multiple amino acid sequence alignments were performed with the software MAFFT (Katoh and Toh 2008), using the G-INSi and E-INSi algorithms for the pol  $\gamma$ - $\alpha$  and - $\beta$  sequences, respectively, and are shown in **supplementary figures S1 and S2, Supplementary Material** online.

### Phylogenetic Inferences

To create the input files for the phylogenetic inferences, the software PAL2NAL (Suyama et al. 2006) was used to convert the amino acid sequence alignments into codon-based nucleotide alignments, which were then converted to NEXUS format. The phylogenetic trees were inferred using the Bayesian algorithm built in the software MrBayes, version 3.2.2 (Ronquist et al. 2012). The consensus tree, run with 200,000 cycles for pol  $\gamma$ - $\alpha$  and 1 million cycles for pol  $\gamma$ - $\beta$ , was set to the 50-majority rule, gamma variation was expected among the sites and Generalized Time Reversible model was used; other parameters were kept as default.

### Modeling of Protein Structure

The structure of the disordered regions in the human pol  $\gamma$ - $\alpha$  crystallography data (PDB accession number 3IKM, chain A) was predicted with the software I-TASSER (Bazzoli et al. 2011), using default parameters. I-TASSER was also used

with default parameters to model the whole structure of pol  $\gamma$ - $\alpha$  from *C. elegans* (using the human pol  $\gamma$  crystal structure 3IKM:A), and of pol  $\gamma$ - $\beta$  from *Strongylocentrotus purpuratus*, *Ciona intestinalis*, and *Trichoplax adhaerens* (using the human pol  $\gamma$ - $\beta$  crystal structure 2G4C:A). *Drosophila melanogaster* pol  $\gamma$ - $\beta$  structure was modeled with the software MODELLER, version 9.12 (Sali and Blundell 1993), using the 2G4C:A file as template and the multiple sequence alignment (MSA) shown in [supplementary figure S2, Supplementary Material](#) online, as a parameter. The selected models were evaluated by the Z-scores, the discrete optimized protein energy values, and the residue error plot (SwissProt website, [http://swissmodel.expasy.org/workspace/index.php?func=tools\\_structure-assessment1](http://swissmodel.expasy.org/workspace/index.php?func=tools_structure-assessment1), last accessed August 2014). Structures and models were analyzed and figures were produced using Pymol ([www.pymol.org](http://www.pymol.org), last accessed August 2014).

## Results

Our searches of public genomic sequence databases found nonredundant sequences for pol  $\gamma$ - $\alpha$  and - $\beta$  of 62 and 52 animal species, respectively ([supplementary table S1, Supplementary Material](#) online). The resulting data set is over-represented by sequences from insect and vertebrate (especially mammalian) species due to the bias in the databases. Animal groups in the Protostomia clade, other than Arthropoda and Nematoda, were most often absent, except for one pol  $\gamma$ - $\alpha$  sequence from Mollusca. Nonetheless, we obtained sequences from key species of basal groups of Metazoa, such as Porifera, Placozoa and Cnidaria, and basal and sister groups of Chordata, such as Tunicata, Cephalochordata, Hemichordata and Echinodermata, which are crucial for the analyses described below.

We excluded several sequences of mammalian and *Drosophila* species from our pol  $\gamma$ - $\alpha$  and - $\beta$  data sets to balance the taxa representation, and also a few of the BLAST results because they contained only partial gene sequences (see [supplementary table S1, Supplementary Material](#) online). [Supplementary figure S1, Supplementary Material](#) online, shows the alignment of 43 pol  $\gamma$ - $\alpha$  amino acid sequences plus the outgroup (the sequence from *S. cerevisiae* pol  $\gamma$ - $\alpha$ , accession number NM\_001183750.1). The alignment of 35 pol  $\gamma$ - $\beta$  amino acid sequences plus the outgroup (the glycyl-tRNA synthetase sequence from the bacterium *Thermus thermophilus*, accession number AJ222643.1) is shown in [supplementary figure S2, Supplementary Material](#) online. The number of pol  $\gamma$ - $\beta$  sequences retrieved was lower than that for pol  $\gamma$ - $\alpha$  mainly because of the absence of the pol  $\gamma$ - $\beta$  gene in the genome of nematode species, as discussed below. In addition, we were unable to find complete pol  $\gamma$ - $\beta$  gene sequences for the poriferan *Amphimedon queenslandica*, the mollusk *Crassostrea gigas*, and the crustacean *Daphnia pulex*. Considering that the pol  $\gamma$ - $\alpha$  sequences from these species do have a potential to form an accessory-interacting determinant

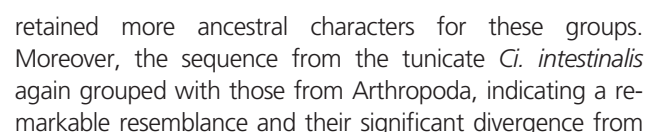
(AID) structure (see below), our failure to find the corresponding pol  $\gamma$ - $\beta$  might be because of the current low coverage of their genome/transcriptome sequences. A summary of our most interesting findings is presented in [figure 1](#), along with schematics of the catalytic and accessory subunit polypeptides. We identified specific features for all sequences retrieved to indicate that our data set most likely consists of true orthologs of the catalytic and accessory subunits of the mitochondrial replicase, and not random genes coding for other family A DNA polymerases or aminoacyl-tRNA synthetases, respectively: 1) Conserved active site motifs in the exonuclease (Exo I–III) and polymerase (Pol A–C) domains of pol  $\gamma$ - $\alpha$  (reviewed in Kaguni 2004) ([supplementary fig. S1, Supplementary Material](#) online) that are shared among family A DNA polymerases; 2) conserved pol  $\gamma$ - $\alpha$ -specific sequences in the spacer region and polymerase domains of pol  $\gamma$ - $\alpha$  (Lewis et al. 1996; Lecrenier et al. 1997; Kaguni 2004); 3) conserved sequence for the AID subdomain of pol  $\gamma$ - $\alpha$  (except in nematodes), which is a pol  $\gamma$ - $\alpha$ -specific feature of metazoans (reviewed in Kaguni 2004) (see below); 4) conserved hydrophobic residues in the C-terminal region of pol  $\gamma$ - $\beta$  that are most relevant for the interactions with the pol  $\gamma$ - $\alpha$  AID, and are found almost exclusively in pol  $\gamma$ - $\beta$ s (Fan et al. 1999; Carrodeguas et al. 2001; Kaguni 2004; Lee et al. 2009); and 5) absence in pol  $\gamma$ - $\beta$  of the residues required for dimerization of class II aminoacyl-tRNA synthetase (Logan et al. 1995; Arnez et al. 1999).

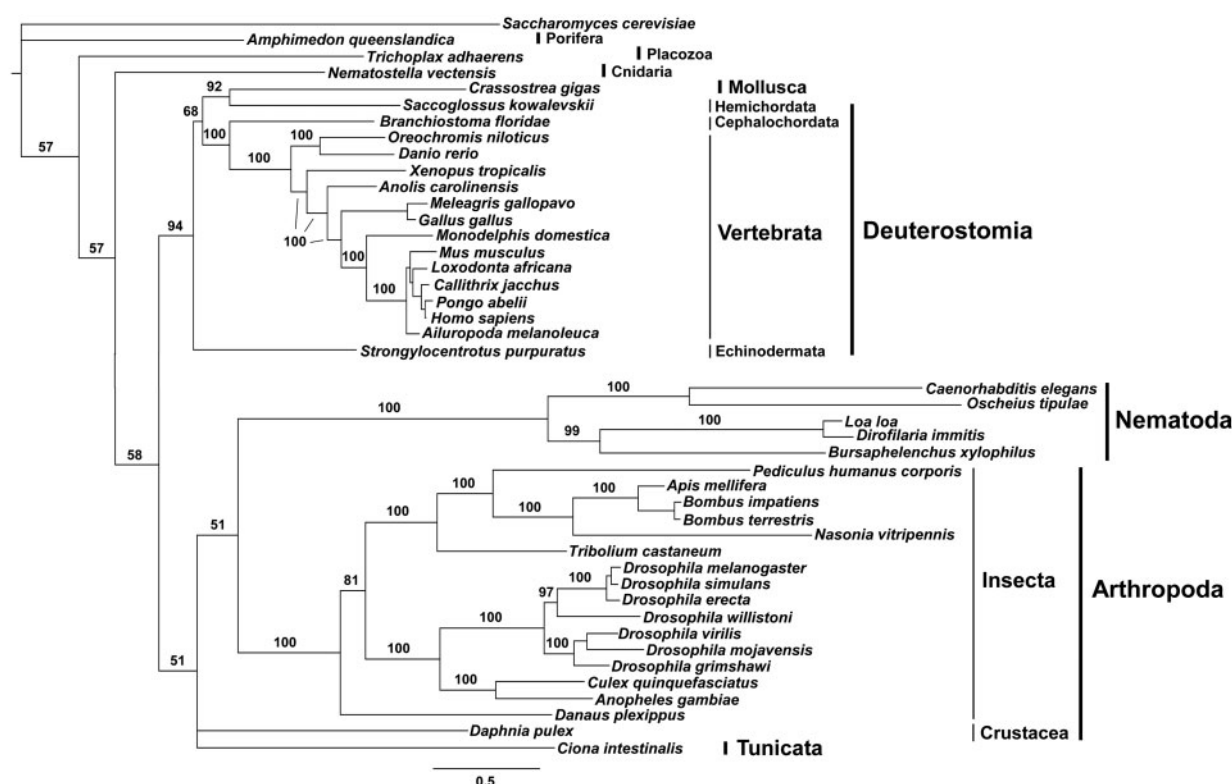
## Distinct Rates of Evolutionary Changes in pol $\gamma$ - $\alpha$ and - $\beta$ Sequences

Phylogenetic inferences with the pol  $\gamma$ - $\alpha$  nucleotide sequences using Bayesian analysis ([fig. 2](#)) reproduced moderately well the currently accepted relationships among animal taxa (Philippe et al. 2009), with few exceptions, and provided important findings regarding the evolution of the gene. First, it is clear that the sequences from nematodes have evolved at a higher rate than those from insect species, which in turn have accumulated more substitutions than the sequences from vertebrates. Second, Tunicata, which is represented only by the sequence from *Ci. intestinalis* (Ascidacea: Cionidae), was grouped within the pol  $\gamma$ - $\alpha$  sequences from Arthropoda. Finally, the branches for the pol  $\gamma$ - $\alpha$  sequences from Deuterostomia (vertebrates and sister groups), excluding *Ci. intestinalis*, and from the mollusk *Cr. gigas* were as short as those for the sequences from basal animal groups, such as Porifera (*A. queenslandica*), Placozoa (*T. adhaerens*), and Cnidaria (*Nematostella vectensis*). In combination, this may indicate that most deuterostome pol  $\gamma$ - $\alpha$ s retained more ancestral characters, whereas pol  $\gamma$ - $\alpha$  sequences from nematodes, arthropods (especially insects), and tunicates have diverged considerably from the original enzyme.

Bayesian phylogenetic analysis using pol  $\gamma$ - $\beta$  nucleotide sequences ([fig. 3](#)) also shows that the insect proteins have







**FIG. 2.**—Bayesian phylogenetic inference for animal pol  $\gamma$ - $\alpha$  nucleotide sequences. The outgroup sequence used was the pol  $\gamma$ - $\alpha$  (*mip-1*) from the yeast *S. cerevisiae* (accession number NM\_001183750.1). The 50% majority-rule consensus tree was inferred using MrBayes 3.2, as described under Materials and Methods. Bayesian posterior probability values are indicated for almost all nodes. The scale bar indicates substitutions per site.

the ancestral accessory subunit polypeptide. We were unable to find any pol  $\gamma$ - $\beta$  coding sequence from species of nematodes, as discussed below.

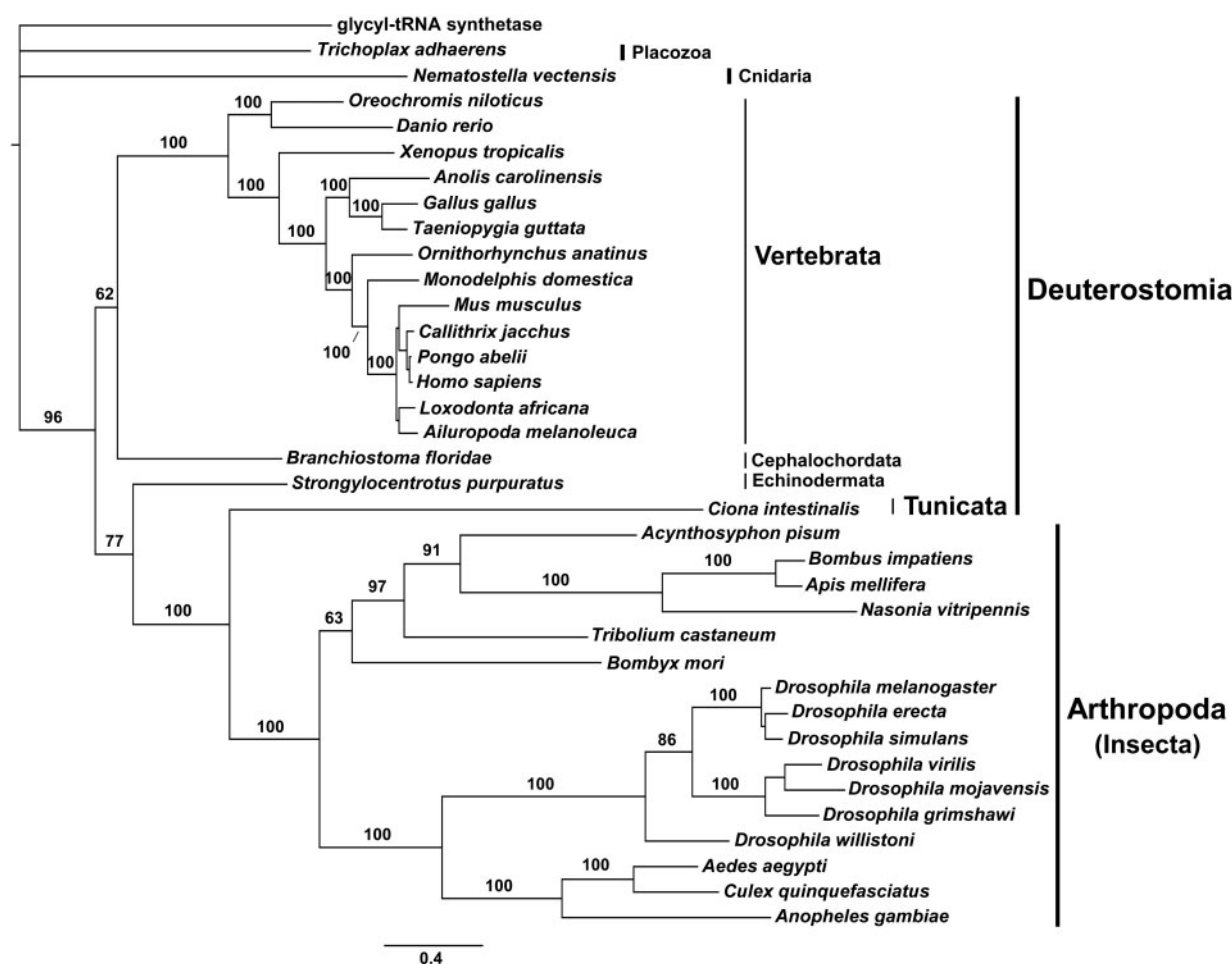
The fast evolutionary rates observed for the pol  $\gamma$ - $\alpha$  and  $\beta$  genes in insects, tunicates, and nematodes appear to follow a general tendency described for the majority of nuclear-encoded genes from these groups (Li 1997; Mitreva et al. 2005; *Drosophila* 12 Genomes Consortium et al. 2007; Denoeud et al. 2010). Interestingly, our preliminary analyses of the genes encoding mtSSB and mtDNA helicase show substitution rates that differ from those of the pol  $\gamma$ - $\alpha$  and  $\beta$  genes among animal groups (Oliveira MT, Haukka J, Kaguni LS, unpublished data). Thus, it appears that the mtSSB and mtDNA helicase genes may have atypical evolutionary constraints, but this requires further validation.

### New Motif in the Exonuclease Domain of Vertebrate pol $\gamma$ - $\alpha$ : Implications for DNA Binding

The MSA identified an insertion of approximately 17 amino acid residues between motifs Exo II and III of the pol  $\gamma$ - $\alpha$  exonuclease domain (H320–A336 in humans; A308–A309 in *D. melanogaster*) that is present in all species of Vertebrata, and possibly other Deuterostomia species (except *Ci. intestinalis*; fig. 4A). Similarly long regions in

the same position are present in *Da. pulex* (Arthropoda: Crustacea), *Oscheius tipulae* (Nematoda: Rhabditida), *Cr. gigas* (Mollusca: Bivalvia), and *A. queenslandica* (Porifera: Demospongiae), but these have little sequence conservation and may represent independent insertion events into the pol  $\gamma$ - $\alpha$  genes of these species. The insertion is several residues downstream of the orienter, a structural module that is highly conserved within eukaryotes, and which has been reported to coordinate the balance between the polymerase and exonuclease functions (Szczepanowska and Foury 2010).

Locating the new vertebrate Exo motif in the crystal structure of the human pol  $\gamma$ - $\alpha$  (PDB: 3IKM; Lee et al. 2009) revealed that the insertion is part of a disordered region (K319–S344) for which we have no structural information, but which may indicate a flexible domain involved in transient interactions. Modeling the missing residues in the human pol  $\gamma$ - $\alpha$  structure based upon the secondary structure prediction algorithm built into the I-TASSER server resulted in two short alpha helices connected by a short loop, which orient this element toward the DNA binding cleft of the enzyme (fig. 4B). In particular, residues H320, K327, K331, and K335 are in close proximity to the minor groove of the primed DNA template modeled onto the putative DNA binding cleft of pol  $\gamma$ - $\alpha$  (Euro et al. 2011). We postulate that the



**Fig. 3.**—Bayesian phylogenetic inference for animal pol  $\gamma$ - $\beta$  nucleotide sequences. The outgroup sequence used was the glycyl-tRNA synthetase from the bacterial species *Thermus thermophilus* (accession number AJ222643.1). The 50% majority-rule consensus tree was inferred using MrBayes 3.2, as described under Materials and Methods. Bayesian posterior probability values are indicated for almost all nodes. The scale bar indicates substitutions per site.

new Exo motif may represent a module that serves to enhance DNA binding in vertebrate pol  $\gamma$ s.

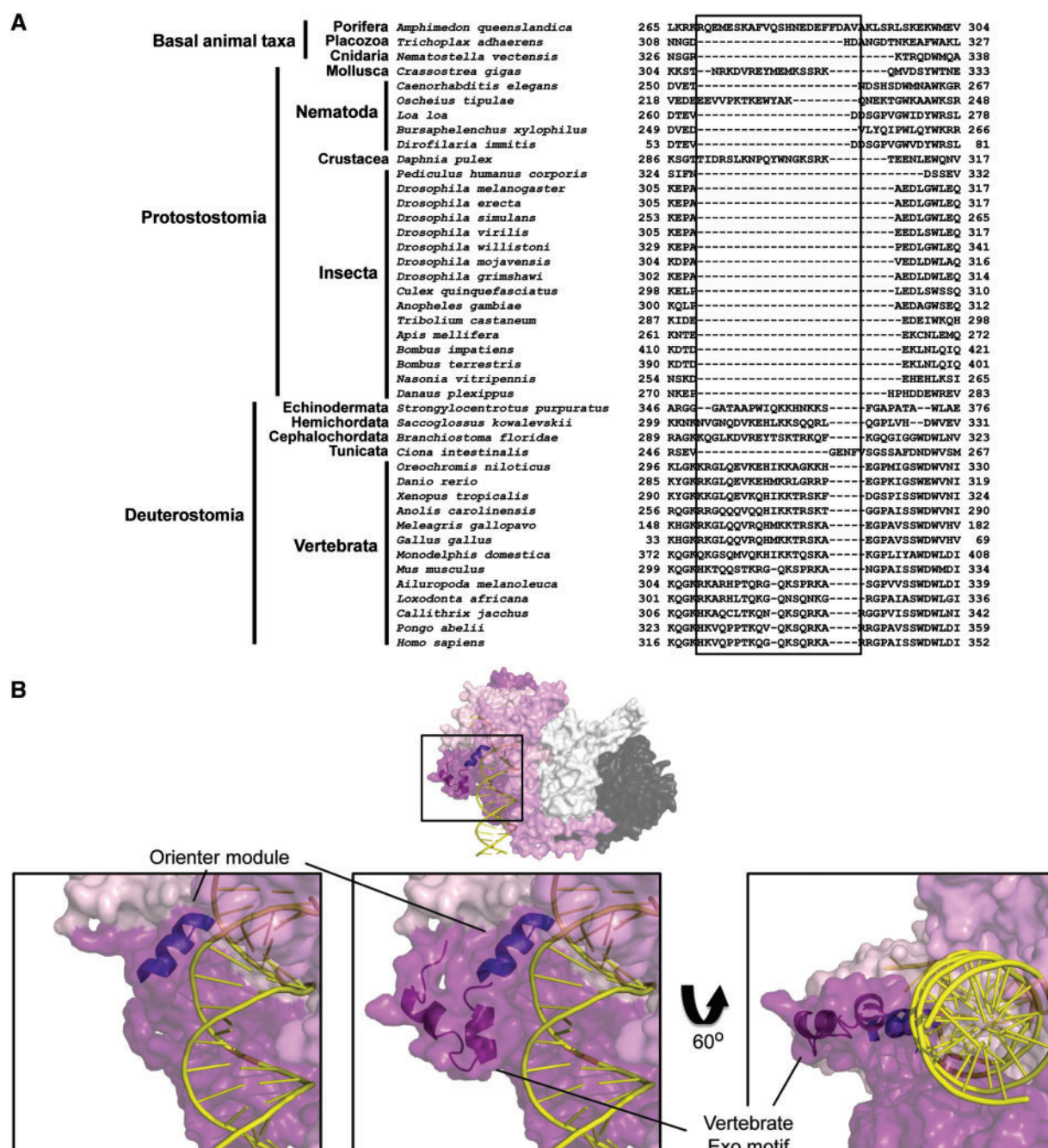
## New Motif in the IP Subdomain of Vertebrate pol $\gamma$ - $\alpha$ : Implications for pol $\gamma$ - $\beta$ Interactions

The spacer region of pol  $\gamma$ - $\alpha$ , which connects the N-terminal exonuclease domain to the C-terminal polymerase domain, contains intrinsic processivity (IP) and AID subdomains (fig. 1). As their names suggest, these subdomains provide structural platforms for supporting both the intrinsic processivity of pol  $\gamma$ - $\alpha$  alone and the enhanced processivity of the holoenzyme, respectively (Lee et al. 2009). Our MSA also identified an insertion of 30 amino acid residues on average in the IP subdomain (E692–R722 in humans; L640–S641 in *D. melanogaster*) that is present consistently in all species of Vertebrata (fig. 5A). Other deuterostome species, such as *St. purpuratus* (Echinodermata) and *Branchiostoma floridae*

(Cephalochordata), and some other noninsect animal species also have insertions of varying sizes in this position. However, the low sequence similarity does not provide enough evidence for homology among the new vertebrate IP motif and the regions to which it aligns in these other animals. In fact, the N-terminal region of this motif in vertebrates is highly conserved, but its C-terminus shows high variability. Again, the sequence from *Ci. intestinalis* resembles significantly those of insect species due to the lack of any amino acid residues in this region. In summary, this element is a distinct and conserved, derived feature in vertebrate pol  $\gamma$ - $\alpha$ ; for other metazoan groups, there is no clear indication of its evolutionary constraints.

The new IP motif was localized in a region of the human pol  $\gamma$ - $\alpha$  crystal structure for which most of the residues were also disordered (G674–R709) (Lee et al. 2009). Although the C-terminus of the motif contains a residue mutated in some cases of human patients with Alpers disease (R722H) and

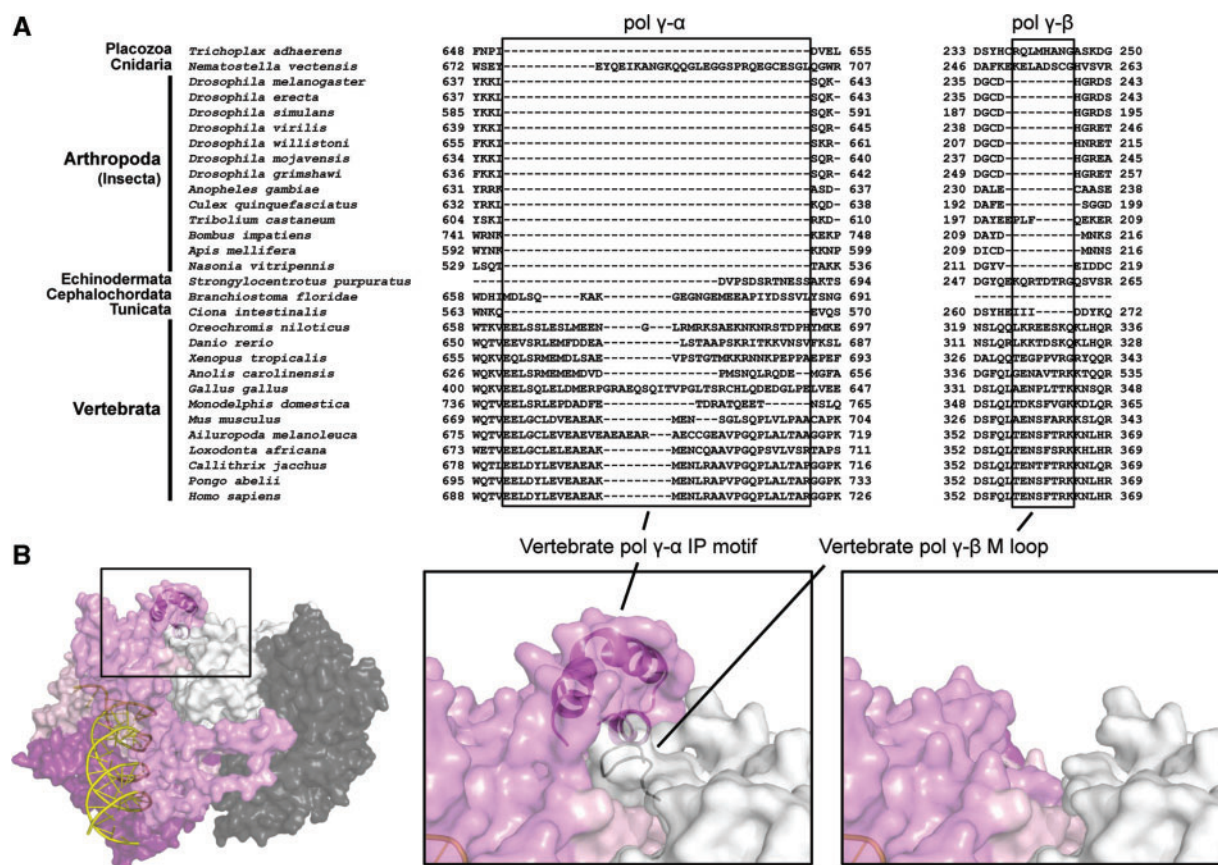




**FIG. 4.**—Identification of a new Exo motif in vertebrate pol  $\gamma$ - $\alpha$ , potentially implicated in primer–template DNA binding. (A) Amino acid sequence alignment indicates the presence of the extra residues (boxed) between pol  $\gamma$ - $\alpha$  motifs Exo II and III in all species of Vertebrata, and a few other animal groups, including other deuterostome species. These residues are disordered in the crystal structure of the human pol  $\gamma$  apo-holoenzyme (3IKM; Lee et al. 2009). (B) Structural model of the human residues indicated in (A) showing their proximity to the primer–template DNA binding cleft. The left panel shows the DNA binding cleft structure without the new vertebrate Exo motif, as it appears in the PDB data file 3IKM. Primer–template DNA binding to pol  $\gamma$ - $\alpha$  was modeled by Euro et al. (2011) and the orienter module (see text for details) is shown as described by Szczepanowska and Foury (2010). Colors are as in figure 1.

possibly implicated in DNA binding (Euro et al. 2011), modeling the missing 36 amino acids resulted in a helix–turn–helix structure located in close proximity to a loop at the end of the Middle domain of the proximal pol  $\gamma$ - $\beta$  protomer (fig. 5B). Part

of this loop in the proximal pol  $\gamma$ - $\beta$ , which to our knowledge has not been described previously and is hereafter called the “M loop” (T357–K364 in humans; D238–H239 in *D. melanogaster*), is again almost exclusive to vertebrates (fig. 5A),



**FIG. 5.**—Identification of a new, putative interacting surface between the catalytic and accessory subunits of vertebrate pol  $\gamma$ . (A) Amino acid sequence alignments indicate the presence of new motifs in the pol  $\gamma$ - $\alpha$  IP subdomain (boxed, left panel) and in the pol  $\gamma$ - $\beta$  middle domain (boxed, right panel) of all species of Vertebrata. Only the species for which both pol  $\gamma$ - $\alpha$  and  $\beta$  sequences were retrieved are shown. The indicated residues in pol  $\gamma$ - $\alpha$  are disordered in the crystal structure of the human pol  $\gamma$  holoenzyme (3IKM; Lee et al. 2009); likewise those in pol  $\gamma$ - $\beta$  are disordered in the crystal structure of the human pol  $\gamma$ - $\beta$  dimer (2G4C; Fan et al. 2006). (B) Model of the human residues indicated in (A), suggesting that the predicted structural elements are in close proximity to each other. The right panel shows the possible interacting region in the absence of the newly identified motifs, as they appear in the PDB files 3IKM and 2G4C. Primer-template DNA binding to pol  $\gamma$  was modeled by Euro et al. (2011). Colors are as in figure 1.

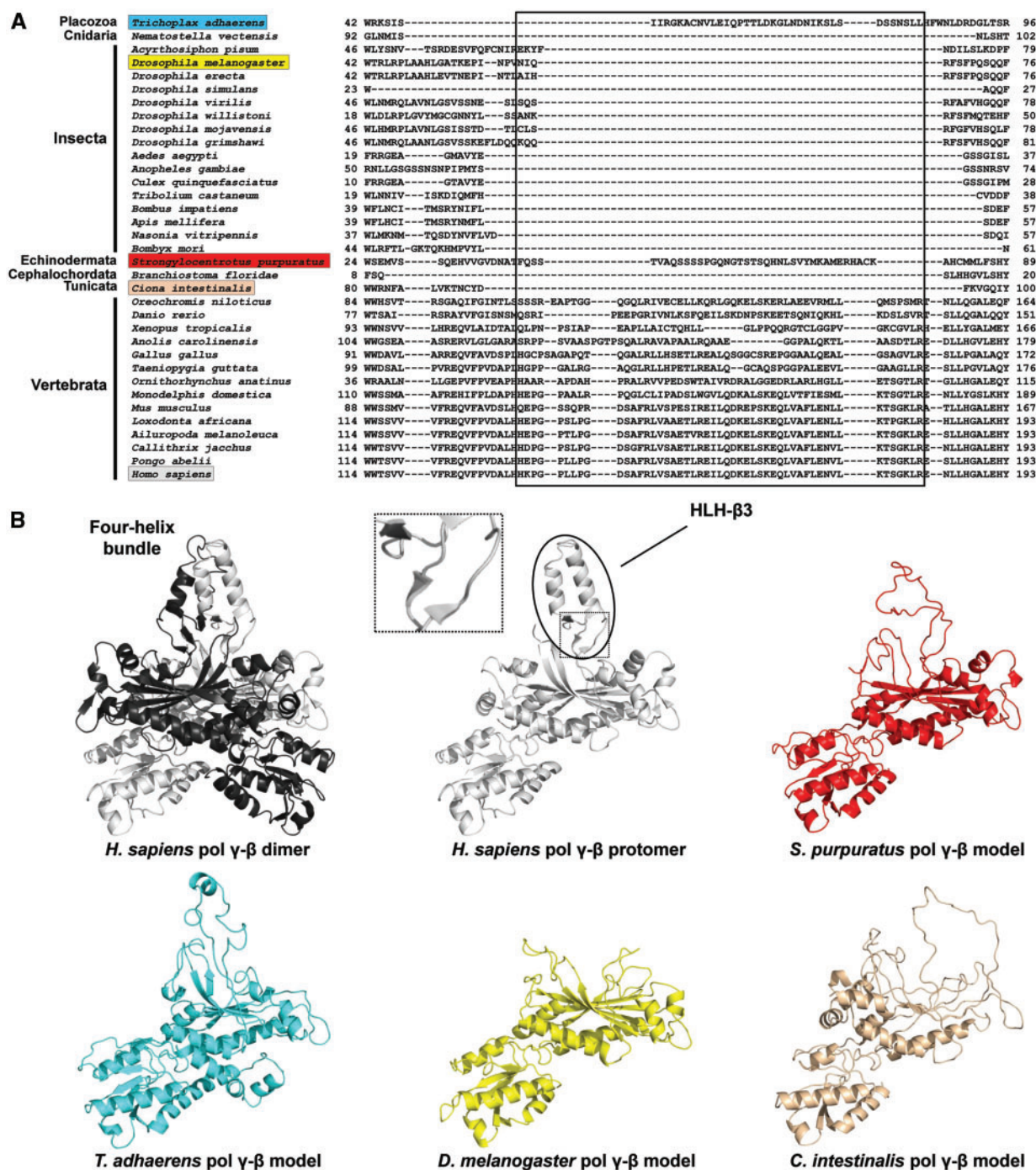
although Echinodermata (*St. purpuratus*), Cnidaria (*N. vectensis*), and Placozoa (*T. adhaerens*) do possess several residues in this same region. A firm correlation between the conserved new IP motif in pol  $\gamma$ - $\alpha$  and the M loop of pol  $\gamma$ - $\beta$  occurs only for vertebrate species; the absence of both structural elements is firmly correlated only for insects and tunicates. Unfortunately, without a better taxa representation, we are unable to conclude whether both elements have been lost in insects and tunicates or gained in vertebrates.

### Dimerization of pol $\gamma$ - $\beta$

Structural and biochemical data have documented that the human and mouse pol  $\gamma$ - $\beta$ s form homodimers in solution and that the homodimer form associates with pol  $\gamma$ - $\alpha$  to constitute a functional heterotrimeric holoenzyme (Carrodegus et al. 2001; Fan et al. 2006; Yakubovskaya et al. 2006; Lee et al. 2009). The major pol  $\gamma$ - $\beta$  dimerization interface is provided by

the HLH- $\beta$ 3 [helix-loop-helix, 3  $\beta$ -sheets] domain (H133–R182 in humans; N63–Q65 in *D. melanogaster*), which is present consistently across all vertebrate species (fig. 6A). Interestingly, the echinoderm *St. purpuratus* and the placozoan *T. adhaerens* also have amino acid residues that could potentially fold into a partial HLH- $\beta$ 3 structure. We tested this hypothesis by modeling the structure of pol  $\gamma$ - $\beta$  from these two species and that from *D. melanogaster* and *Ci. intestinalis* (control species that lack completely the HLH- $\beta$ 3 element). Neither the helix-loop-helix structure, responsible for the formation of the four-helix bundle with the adjacent pol  $\gamma$ - $\beta$ , nor the three  $\beta$  sheets found at the base of the four-helix bundle are clearly observed for *St. purpuratus* and *T. adhaerens* (fig. 6B). As expected, the pol  $\gamma$ - $\beta$  models for *D. melanogaster* and *Ci. intestinalis* also have none of the structural elements necessary for the HLH- $\beta$ 3 folding. All these proteins are, therefore, most likely unable to homodimerize using the same structural features adopted by





**FIG. 6.**—Dimerization of vertebrate pol  $\gamma$ - $\beta$  through the formation of the four-helix bundle structure. (A) Amino acid sequence alignment indicates the presence of the HLH- $\beta$ 3 domain (boxed) in all species of Vertebrata and possibly in few other animal groups. (B) Comparison of the crystal structure of the human pol  $\gamma$ - $\beta$  dimer and structural models for pol  $\gamma$ - $\beta$  of *Trichoplax adhaerens*, *Strongylocentrotus purpuratus*, *Drosophila melanogaster*, and *Ciona intestinalis*, showing that only vertebrate pol  $\gamma$ - $\beta$  can fold into a HLH- $\beta$ 3 structure and therefore form the four-helix bundle dimerization interface. The inset shows the three short  $\beta$ -sheets at the base of the HLH- $\beta$ 3 structure. For more information about the structural features of the four-helix bundle fold, see Kamtekar and Hecht (1995) and Carrodegus et al. (2001).

the vertebrate pol  $\gamma$ - $\beta$ . We propose here that the ability to form homodimers is exclusive to vertebrate pol  $\gamma$ - $\beta$ , and that this feature appears to have emerged through the acquisition of a structural evolutionary novelty, the HLH- $\beta$ 3 domain.

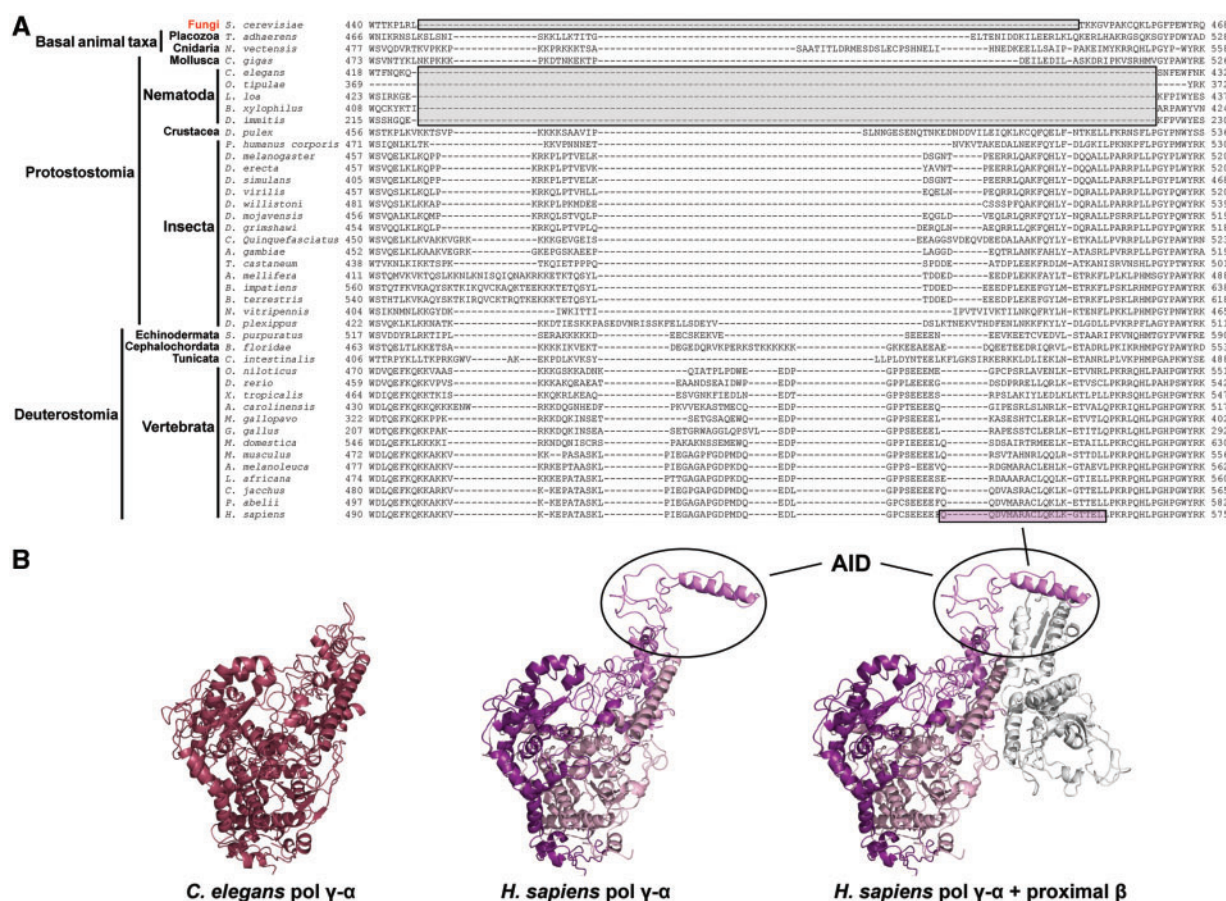
# Deletion of pol $\gamma$ - $\alpha$ AID Subdomain and Absence of the pol $\gamma$ - $\beta$ Gene in Nematoda

Our MSA also shows clearly that the residues of pol  $\gamma$ - $\alpha$  AID subdomain are completely absent in all species of Nematoda (fig. 7). In addition, we searched exhaustively for homologs of pol  $\gamma$ - $\beta$  in diverse sequence databanks, but found only nematode sequences related to aminoacyl-tRNA synthetase genes, with very high *E* value scores (data not shown). Others have indicated previously, using an alignment with a small number of amino acid sequences, that the *C. elegans* pol  $\gamma$ - $\alpha$  may not have the AID subdomain and might function as a monomeric enzyme (Bratic et al. 2009;

Addo et al. 2010); our bioinformatic analyses provide support for this claim, and suggest that this is a common feature for all nematode species.

## Discussion

Although the fields of animal mitochondrial genomics and mtDNA replication have expanded greatly in recent years, insights from these two fields have not yet effectively cross-fertilized our understanding of mitochondrial genome biology. Complete mitochondrial genomes of virtually all animal groups (reviewed in Gissi et al. 2008 and D'Onofrio de Meo et al. 2012) and thousands of human individuals (reviewed in Ye et al. 2014) are currently available, showing strong constraints such as the small size of the molecule (~16 kb), its circular conformation and compact gene organization (~37 genes and few noncoding nucleotides), the strictly orthologous encoded genes, and the very low importance of



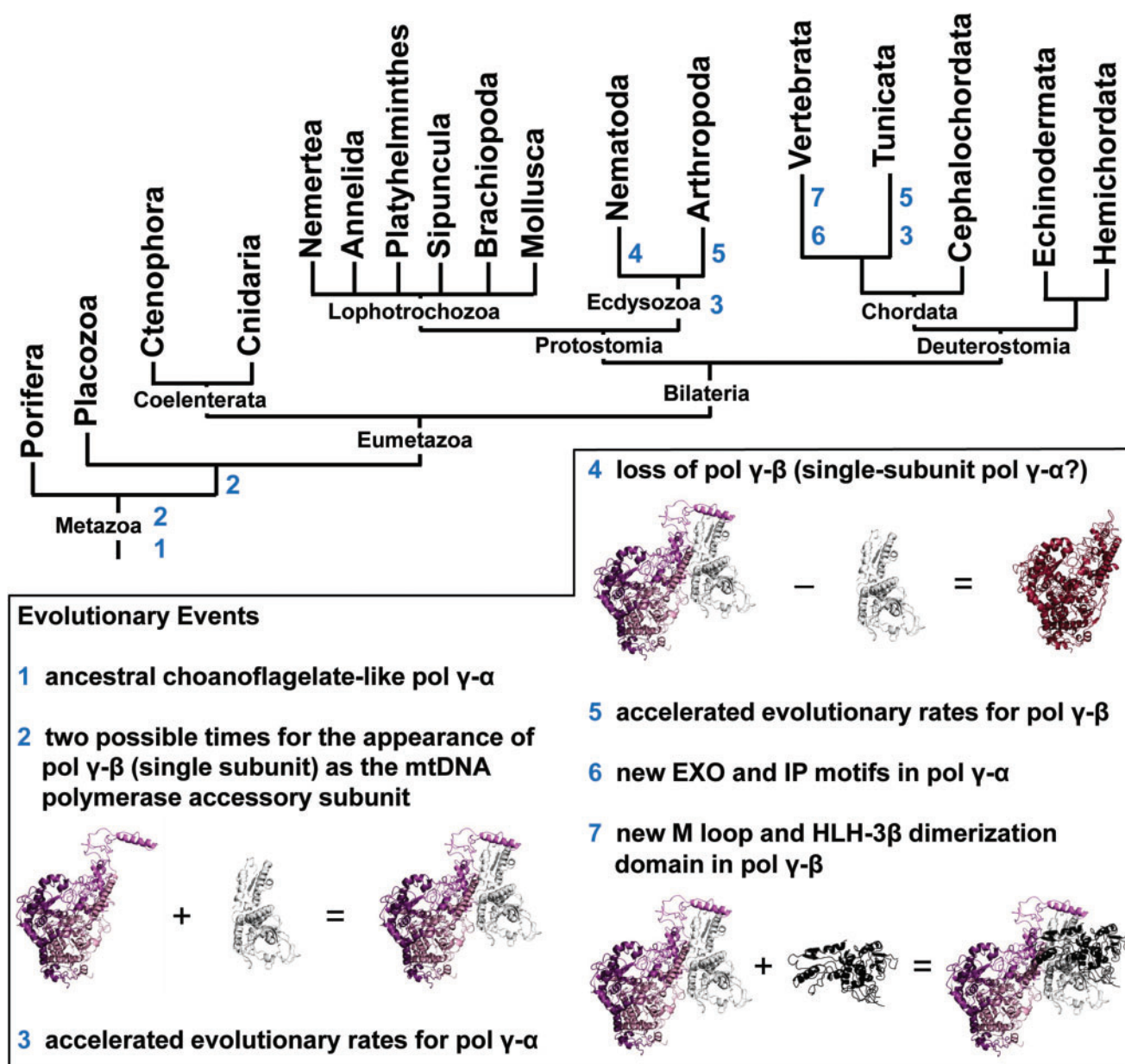
**Fig. 7.**—The likely monomeric nature of nematode pol  $\gamma$ . (A) Amino acid sequence alignment indicates the absence of the pol  $\gamma$ - $\alpha$  AID subdomain (gray boxes) in Nematoda species, which is similar to the well-characterized enzyme from the yeast *Saccharomyces cerevisiae*. The pink box highlights the residues forming the L helix of human pol  $\gamma$ - $\alpha$ , the major point of interaction with the proximal pol  $\gamma$ - $\beta$ . (B) Structural model of the *Caenorhabditis elegans* pol  $\gamma$ - $\alpha$  (left panel) adjacent to the human pol  $\gamma$  structure (3IKM; Lee et al. 2009) that contains a fully folded AID subdomain (middle and right panels). Colors are as in figure 1.



recombination *sensu stricto* for the purpose of genome inheritance (Supplementary fig. S3 and table S2, Supplementary Material online). On the other hand, research has also shown a great variability in nucleotide and amino acid substitution rates, gene rearrangement, number and length of noncoding regions, and genetic codes, whose underlying mechanisms are related generally to mtDNA transaction processes, such as replication, transcription, and repair. Testing the functional importance of such changes has been difficult due to our inability to perform mtDNA site-directed

mutagenesis (reviewed in Moraes et al. 2014); at the same time, this experimental constraint does not apply to the study of mitochondrial proteins encoded in the nucleus, as are all of the proteins involved in mtDNA metabolism.

In one of the first attempts to correlate genomic structural diversity and functional mechanisms, Jiang et al. predicted, based on the variability of nucleotide composition bias along the mitochondrial genome of snakes, that their duplicated major noncoding regions (D-loops) may contain an additional origin of heavy-strand replication, implying a different



**FIG. 8.**—Phylogeny of animal groups indicating the molecular evolutionary events identified in this work for the catalytic and accessory subunits of the mitochondrial replicase. The tree is based on that in Philippe et al. (2009).



replication mode (see vertebrate mtDNA replication mechanisms below) (Jiang et al. 2007). *Cis*-elements that control transcription and replication are most likely present in the mtDNA noncoding regions, and hence their genomic rearrangements would reflect more readily mechanistic differences in these fundamental processes. However, a number of important functional sites have been described within the mtDNA coding region, such as the binding sites for transcription termination factors, which in fact have a documented impact on both transcription and replication, regulating/restricting the progression of mtDNA replication forks to prevent collisions and to facilitate productive interaction between the two machineries (Hyvarinen et al. 2007, 2011; Joers et al. 2013). Therefore, evolutionary changes involving only gene rearrangements of the mitochondrial genome are also likely to correlate with changes in maintenance mechanisms. To develop molecular insights into DNA polymerase structure and function that could impact genome evolution, we have reported here the sequence and structural evolution of the key enzyme in animal mtDNA replication, pol  $\gamma$ , establishing correlations and evaluating their implications at the biochemical/molecular level, in consideration of the features of the mitochondrial genome in different metazoan groups.

Analyses of amino acid sequence alignments, Bayesian phylogenetic inferences, and protein modeling allowed us to identify three important evolutionary patterns within the catalytic and accessory subunits of animal pol  $\gamma$  (fig. 8). First, it appears that vertebrate pol  $\gamma$ s show a high level of constraint, with little sequence and structural variability within the group. Both subunits have evolved in a highly similar fashion (figs. 2 and 3) that is distinct from that of other animal groups, and together have acquired at least four structural evolutionary novelties: New Exo and IP motifs in pol  $\gamma$ - $\alpha$  (figs. 4 and 5), and a new M loop and the HLH-3 $\beta$  domain in pol  $\gamma$ - $\beta$  (figs. 5 and 6). The new motif in the Exo domain is potentially present in other deuterostome species, such as echinoderms, hemichordates, and cephalochordates, but it is clearly absent in tunicates (fig. 4A). The amino acid residues involved have the potential to fold into a structure that may enhance and/or stabilize enzyme–DNA interactions (fig. 4B). Interestingly, these amino acids appear disordered in the crystal structure of the human pol  $\gamma$  apo-holoenzyme (Lee et al. 2009), indicating that the motif may adopt a flexible structure that might be stabilized upon primer–template DNA binding. In addition to a proposed role in DNA binding per se, the new motif may function specifically in concert with the conserved element known as the orienter (Szczezanowska and Foury 2010). The orienter lies only three amino acids upstream and within 8.3 Å of the N-terminus of the new motif in the three-dimensional structure (fig. 4B), and has been shown to coordinate the exonuclease and polymerase functions of pol  $\gamma$  to enhance the fidelity of the holoenzyme. Thus, the close proximity of the new motif in the Exo domain and its likely capacity to bind DNA suggests that it may enhance and/

or regulate the function of the orienter. The acquisition of the new motif in the IP subdomain of vertebrate pol  $\gamma$ - $\alpha$  is correlated with the acquisition of the M loop in the middle region of vertebrate pol  $\gamma$ - $\beta$  (fig. 5). These structures may interact, thereby increasing the contact surface between the catalytic and the proximal accessory subunit. However, this putative enhanced interaction may only be stable either upon binding to DNA or to another component of the mtDNA replisome, given that the residues in both subunits are disordered in the crystal structures of the human apo-holoenzyme and in the pol  $\gamma$ - $\beta$  dimer, respectively (Fan et al. 2006; Lee et al. 2009). We believe that these new structural elements found in pol  $\gamma$ - $\alpha$  and - $\beta$  of vertebrates are not only remarkable from an evolutionary point of view but also important from a biomedical perspective, as they lie within the structural and functional regions defined as clusters 2 and 5 by Euro et al., which bear diverse pathogenic mutations of the human pol  $\gamma$ - $\alpha$  (Euro et al. 2011; Farnum et al. 2014), or in close proximity to the disease mutation R369G in human pol  $\gamma$ - $\beta$  (Young et al. 2011; Craig et al. 2012). Future biochemical study of each of these three new elements is warranted to test our predictions, and to determine their specific roles in both vertebrate pol  $\gamma$  function and mtDNA replication.

Perhaps the most striking result regarding the evolution of vertebrate pol  $\gamma$  is the homodimeric nature of pol  $\gamma$ - $\beta$  in this animal group, which we propose to have arisen with the acquisition of the HLH-3 $\beta$  domain (fig. 6). The two alpha-helices within this domain in one pol  $\gamma$ - $\beta$  protomer interact with those in the other protomer to form a four-helix bundle, a common protein fold that is characterized by four helices packed against each other, in a roughly lengthwise configuration, at a specific angle relative to a virtual central axis (Kamtekar and Hecht 1995; Carrodegua et al. 2001). Residues from echinoderm and placozoan species, aligned with the vertebrate HLH-3 $\beta$  domain, are clearly nonhomologous, having evolved independently and unable to fold into a similar structure (fig. 6). We attempted to identify the origin of the vertebrate dimerization domain in BLAST searches of its sequence against the human, mouse, and zebrafish genomes, using relaxed parameters, but only the pol  $\gamma$ - $\beta$  gene sequences themselves were obtained as results (data not shown). Previously published alignments (Fan et al. 2006; Lee et al. 2010) using the amino acid sequences of only a few animal pol  $\gamma$ - $\beta$ s (human, mouse, and *D. melanogaster*) mistakenly suggested that two insertions into mammalian pol  $\gamma$ - $\beta$  created the dimer interface. Our analysis using the highest number of animal pol  $\gamma$ - $\beta$  sequences available to date (35 sequences vs. the commonly used alignment of the human, mouse, frog, and fly sequences [Carrodegua et al. 2001; Fan et al. 2006; Lee et al. 2009]) argues clearly that the residues comprising the HLH-3 $\beta$  domain were inserted as a whole, early in the evolution of Vertebrata.

Deletion variants of the human and mouse pol  $\gamma$ - $\beta$  that lack most of the HLH-3 $\beta$  domain (called  $\Delta$ I4, and comprising

residues R147–G179 in humans) are observed as monomers in solution, though the proteins crystallize as dimers (Carrodeguas et al. 2001; Lee et al. 2009, 2010). Using this pol  $\gamma$ - $\beta$  variant and related proteins, it has been shown that the functions of the two pol  $\gamma$ - $\beta$  protomers that constitute the  $\beta$  dimer in the human holoenzyme are distinct: The proximal subunit strengthens primer–template DNA interactions whereas the distal one accelerates the DNA synthetic rate (Lee et al. 2010). Although the authors argue that the homodimeric nature of human pol  $\gamma$ - $\beta$  provides the molecular basis for these functions of the accessory subunit, our earlier biochemical data show at least in the fly that the monomeric *D. melanogaster* pol  $\gamma$ - $\beta$  imparts the same functions to the cognate fly holoenzyme (Lewis et al. 1996; Wang et al. 1997). Thus, at least in invertebrates, dimerization of pol  $\gamma$ - $\beta$  is not a structural requirement for enhanced DNA binding, or the increased processivity and rate of nucleotide polymerization of the pol  $\gamma$  holoenzyme. If the additional pol  $\gamma$ - $\beta$  protomer indeed provides increased activity to the vertebrate holoenzyme, its effect is likely mitigated by the fact that the stimulation of the human pol  $\gamma$  by its cognate mtSSB is only moderate ( $\sim 3$ -fold), and is observed only at low salt concentrations in vitro (Oliveira and Kaguni 2010). In contrast, *D. melanogaster* pol  $\gamma$  is stimulated up to 20-fold by its cognate mtSSB, and this is observed over a broad range of KCl concentrations (Farr et al. 1999). In this regard, we would speculate that the functions of the distal pol  $\gamma$ - $\beta$  protomer in the vertebrate mtDNA replisome are likely at least in part contributed by mtSSB in insect (invertebrate) systems, either directly by protein–protein interactions, or by its ability to organize the DNA substrate.

An interesting question that has been raised previously (Lee et al. 2010) is whether the homodimeric state of vertebrate pol  $\gamma$ - $\beta$  is a basal or a derived character, taking into account that class II aminoacyl-tRNA synthetases also function as homodimers. Our data argue that the pol  $\gamma$ - $\beta$  from all invertebrate species is monomeric (except in nematodes, which have lost the gene), including basal animal species, and that this is an ancestral character for metazoans. Thus, we propose that the dimerization of modern class II aminoacyl-tRNA synthetases evolved independently from that of vertebrate pol  $\gamma$ - $\beta$  but perhaps for a common reason, to allow more sophisticated mechanistic or regulatory strategies during protein translation and mtDNA replication, respectively. This is supported further by the fact that the sequence and structure of the dimerization domain of the *Th. thermophilus* glycyl-tRNA synthetase (Logan et al. 1995; Arnez et al. 1999), a prototype of prokaryotic class II aminoacyl-tRNA synthetases, is significantly dissimilar to the HLH-3 $\beta$  domain of vertebrate pol  $\gamma$ - $\beta$  (Carrodeguas et al. 2001; Fan et al. 2006).

The shared sequence and structural features described above for vertebrate pol  $\gamma$ , absent in other metazoans, correlate with the fact that vertebrate mitochondrial genomes are extremely invariable in genome architecture, gene content

and strand asymmetry, and in the position and number of noncoding regions (supplementary fig. S3 and table S2, Supplementary Material online; reviewed in Gissi et al. 2008). We hypothesize that the correlation between stabilization of mitochondrial genomic features and shared pol  $\gamma$  structure and sequence evolution may reflect a common mechanism of mtDNA replication in vertebrates. In this regard, it is tempting to speculate that the dimerization of the vertebrate accessory subunit of pol  $\gamma$  is related to its acquired ability to bind to double-stranded DNA: Carrodeguas et al. (2002) showed that the human pol  $\gamma$ - $\beta$  dimer binds specifically to template–primer junctions in vitro and on the basis of physiological experiments, Di Re et al. (2009) have proposed that it serves a role in regulation of nucleoid structure. Holt, Jacobs and collaborators have shown that the mtDNA of human tissues and cells in culture, mouse, and chicken (*Gallus gallus*) are replicated in a similar fashion through one (or two) mechanism(s) that involve(s) the incorporation of extensive tracts of RNA throughout the mtDNA lagging strand (Yang et al. 2002; Yasukawa et al. 2006; Pohjoismaki et al. 2010). This RNA appears to originate from processed transcripts of mitochondrial genes that hybridize to the displaced strand during synthesis of the leading DNA strand, and the authors propose that these transcripts later serve as primers for lagging strand synthesis (Reyes et al. 2013). There is also evidence that at least some fraction of the replicating mtDNA molecules are replicated by a coupled leading and lagging strand synthesis mechanism, resembling that of bacterial species in the replication of their circular chromosomes (Holt et al. 2000; Reyes et al. 2005). Although long-standing alternate mechanisms exist (reviewed in Bogenhagen and Clayton 2003), there has always been a consensus that the mode(s) of mtDNA replication is (are) shared among vertebrate species, and our data corroborate this idea.

The second evolutionary pattern we have identified among the animal pol  $\gamma$  subunits is that both the pol  $\gamma$ - $\alpha$  and - $\beta$  genes of arthropods and tunicates have evolved approximately twice as rapidly as their vertebrate homologs (figs. 2 and 3), despite the fact that the polypeptides they encode tend to be shorter (supplementary figs. S1 and S2, Supplementary Material online). In addition to the phylogenetic clustering of arthropod and tunicate sequences, *Ci. intestinalis* pol  $\gamma$ - $\alpha$  and - $\beta$  have some structural features that are more similar to arthropods than to other deuterostome species: The complete absence of the new Exo and IP motifs in pol  $\gamma$ - $\alpha$  and of the M loop and the HLH-3 $\beta$  domain in pol  $\gamma$ - $\beta$ . Even though these are clearly vertebrate elements, echinoderms do possess residues that align with the new Exo and IP motifs, the M loop and the HLH-3 $\beta$  domain, cephalochordates may also form the new Exo and IP motifs, and hemichordates may possess the new Exo motif. We investigated whether the *Ci. intestinalis* sequences might result from a contamination of an arthropod DNA that was mistakenly assigned to this tunicate, or a product of a possible lateral gene transfer. Serving to exclude the possibility of

contamination, we found that the sequences of both genes were firmly assembled into chromosomes 12 and 7, respectively (data not shown), and not randomly annotated in an unplaced contig/scaffold. The pol  $\gamma$ - $\beta$  gene architecture in the *Ci. intestinalis* genome, for example, is highly similar to that in the human, mouse, chicken, and zebrafish genomes (eight exons with conserved exon/intron boundaries), whereas the homologous gene of *Apis mellifera* and *Drosophila* species has only two exons, with variable exon/intron boundaries (data not shown), indicating that the probability of lateral transfer of an arthropod gene is minimal. Thus, the sequences of *Ci. intestinalis* used in this report are indeed encoded in its genome, and most likely cluster with those of arthropods as a consequence of the phenomenon of long branch attraction (reviewed in Bergsten 2005).

Tunicate mtDNAs exhibit a very high degree of genome variability, which includes frequent protein-coding gene rearrangements, extra copies of tRNA genes, maximal gene strand asymmetry and extremely short noncoding regions (supplementary fig. S3 and table S2, Supplementary Material online; reviewed in Gissi et al. 2008). Among Deuterostomia, this is the taxon that may be the most likely to employ a substantially different mode of replicating its mitochondrial genome; in addition to the features of the mtDNA itself, the monomeric nature of pol  $\gamma$ - $\beta$  and the fast rate of evolutionary change for the catalytic and accessory subunit genes all serve as indicators. However, experimental data to validate this hypothesis are needed. For arthropods, on the other hand, a substantial amount of research toward the understanding of mtDNA replication is available, using *D. melanogaster* as a model system (reviewed in Oliveira et al. 2010). Although this species may not be representative of the entire diverse group of Arthropoda, considering that its mitochondrial genome (Lewis et al. 1995), and that of other dipterans, shares features with basal species of the phylum, such as the horseshoe crab *Limulus polyphemus* (Chelicerata: Xiphosura) (Lavrov et al. 2000) and Amblypygi spiders (Chelicerata: Arachnida) (Fahreïn et al. 2009), one might expect that studies with *D. melanogaster* may reveal a conserved and basal mechanistic model of mtDNA replication. Joers and Jacobs (2013) have described recently that *D. melanogaster* mtDNA is replicated predominately through a strand-coupled DNA synthesis mode, initiating within the major noncoding region and proceeding unidirectionally at an irregular rate of elongation. Unlike the situation in vertebrates, the presence of RNA molecules hybridized to the mtDNA lagging strand was not detected, although DNA replication intermediates containing short stretches of single-stranded DNA was observed for a small subset of the replicating mtDNA molecules. In combination with our findings on the rapid evolution of the arthropod pol  $\gamma$  subunits and the heterodimeric composition of the holoenzyme, the data suggest that arthropods may employ a unique mode of replication among metazoan mtDNAs. Considering that the mtDNA features among arthropod

species are also highly diverse (supplementary fig. S3 and table S2, Supplementary Material online), with groups such as the myriapods and the Acari arachnids showing numerous rearrangements of gene order (reviewed in Gissi et al. 2008), we can also predict that additional mechanisms of replication will be identified.

Remarkably, as the third pattern of animal pol  $\gamma$  evolution, we found that the oligomeric form of the holoenzyme is more variable than the concept of the monomeric versus homodimeric nature of pol  $\gamma$ - $\beta$ . Our data argue that pol  $\gamma$ - $\beta$  is not found in any nematode species, and that this is a derived character: First, we retrieved no gene coding for a homologous protein, despite exhaustive databank searches and second, we observed that nematode pol  $\gamma$ - $\alpha$  lacks the residues needed to form the AID subdomain, the structural platform for the interactions between the catalytic and the accessory subunits in other animal species (fig. 7). Although it may be premature to conclude that nematode pol  $\gamma$ - $\alpha$  is a monomeric enzyme, we suggest that the mtDNA replisome in this animal group is distinct from that of any other metazoan taxa, and it might employ a distinct mtDNA replication mechanism. The facts that nematode pol  $\gamma$ - $\alpha$ s are the fastest evolving sequences among metazoans (fig. 2), and the mtDNA of Nematoda species has unusual features, such as loss of the *atp8* gene, all genes coded only in one strand, and frequent changes in the gene order among species (supplementary fig. S3 and table S2, Supplementary Material online; reviewed in Gissi et al. 2008), are all consistent with this speculation. Indeed, Lewis et al. (2015) have reported that mtDNA replication in the nematode *C. elegans* most likely proceeds via a rolling circle mechanism, which differs from any other mechanism documented in metazoans to date. Alternatively, one might argue that this monomeric pol  $\gamma$ - $\alpha$  is not the processive, replicative DNA polymerase of nematode mitochondria because the gene is not essential for *C. elegans* development and survival (Bratic et al. 2009; Addo et al. 2010); it could be a mtDNA repair enzyme that does not require processive DNA synthesis, therefore eliminating the need for pol  $\gamma$ - $\beta$  (Longley et al. 1998; Copeland and Longley 2014). However, it is important to point out that the yeast *S. cerevisiae* pol  $\gamma$  is a processive enzyme that does not contain a pol  $\gamma$ - $\beta$ -like accessory subunit (Foury 1989). Furthermore, the knockdown of pol  $\gamma$ - $\alpha$  in *C. elegans* does cause strong mtDNA depletion as in *D. melanogaster*, mouse, and humans (Iyengar et al. 1999, 2002; Spelbrink et al. 2000; Hance et al. 2005; Fukuoh et al. 2014). In our view, the pol  $\gamma$ - $\alpha$  sequences we retrieved for nematodes do represent the mitochondrial replicase, and the successful development of *C. elegans* with a substantially lower level of both pol  $\gamma$ - $\alpha$  and mtDNA most likely indicates that mitochondrial oxidative phosphorylation is the feature dispensable for egg and larval viability in this organism.

Although we attempted to correct for the bias in taxa sampling by excluding several species of *Drosophila* and mammals



from our data set, a large overrepresentation of sequences from Insecta and Vertebrata remain, limiting the interpretation of our results for other animal groups. Nonetheless, we consider the correlations presented here regarding pol  $\gamma$  evolution as robust, and in all likelihood are a reflection of distinct mechanistic differences in mtDNA replication. Because of its roles in mitochondrial function, pol  $\gamma$  has been studied extensively from a clinical and biomedical perspective. Our study argues that revealing its interesting evolutionary path(s) among metazoans, which has just recently caught the attention of biochemists and structural biologists, is not only of interest per se, but may also serve to promote the understanding of human disease-related phenotypes caused by pol  $\gamma$ - $\alpha$  and - $\beta$  mutations. Moreover, investigating whether or not the evolutionary paths of the mtDNA helicase and mtSSB throughout metazoan history are consistent with those of pol  $\gamma$  might reveal coevolution toward a more efficient replisome. For example, we have demonstrated recently that the *D. melanogaster* mtDNA helicase contains an iron-sulfur cluster in its N-terminal domain, which appears to be absent in the human homolog (Stiban et al. 2014). What its functions and phylogenetic distribution are, and how this property has evolved represent important new questions. Clearly, molecular evolutionary studies of each of the protein components of the animal mtDNA replisome are warranted to achieve a better understanding both of organelle evolution and of the processes leading to mitochondrial dysfunction.

## Supplementary Material

Supplementary tables S1 and S2 and figures S1–S3 are available at *Genome Biology and Evolution* online (<http://www.gbe.oxfordjournals.org/>).

## Acknowledgments

The authors thank Dr David Arnosti, Dr Grzegorz Ciesielski, and Anssi Nurminen for their critical review of the manuscript. This work was supported by grants from the National Institutes of Health (grant number 45295) to L.S.K.; and the Marie Curie International Incoming Fellowship from the European Union (grant number GA328988) and the Fundação de Amparo à Pesquisa do Estado de São Paulo (grant number 2014/02253-6) to M.T.O. J.H. and L.S.K. were supported partially by the University of Tampere and the Finland Distinguished Professor Programme of the Academy of Finland, respectively.

## Literature Cited

- Addo MG, et al. 2010. *Caenorhabditis elegans*, a pluricellular model organism to screen new genes involved in mitochondrial genome maintenance. *Biochim Biophys Acta*. 1802:765–773.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215:403–410.
- Arnez JG, Dock-Bregeon AC, Moras D. 1999. Glycyl-tRNA synthetase uses a negatively charged pit for specific recognition and activation of glycine. *J Mol Biol*. 286:1449–1459.
- Bazzoli A, Tettamanzi AG, Zhang Y. 2011. Computational protein design and large-scale assessment by I-TASSER structure assembly simulations. *J Mol Biol*. 407:764–776.
- Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21: 163–193.
- Bogenhagen DF, Clayton DA. 2003. The mitochondrial DNA replication bubble has not burst. *Trends Biochem Sci*. 28:357–360.
- Bratic I, et al. 2009. Mitochondrial DNA level, but not active replicase, is essential for *Caenorhabditis elegans* development. *Nucleic Acids Res*. 37:1817–1828.
- Carrodegua JA, Kobayashi R, Lim SE, Copeland WC, Bogenhagen DF. 1999. The accessory subunit of *Xenopus laevis* mitochondrial DNA polymerase gamma increases processivity of the catalytic subunit of human DNA polymerase gamma and is related to class II aminoacyl-tRNA synthetases. *Mol Cell Biol*. 19:4039–4046.
- Carrodegua JA, Pinz KG, Bogenhagen DF. 2002. DNA binding properties of human pol gammaB. *J Biol Chem*. 277:50008–50014.
- Carrodegua JA, Theis K, Bogenhagen DF, Kisker C. 2001. Crystal structure and deletion analysis show that the accessory subunit of mammalian DNA polymerase gamma, Pol gamma B, functions as a homodimer. *Mol Cell*. 7:43–54.
- Cock PJ, et al. 2009. Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25: 1422–1423.
- Copeland WC, Longley MJ. 2014. Mitochondrial genome maintenance in health and disease. *DNA Repair (Amst)*. 19:190–198.
- Craig K, et al. 2012. A p.R369G POLG2 mutation associated with adPEO and multiple mtDNA deletions causes decreased affinity between polymerase gamma subunits. *Mitochondrion* 12:313–319.
- D'Onorio de Meo P, et al. 2012. MitoZoa 2.0: a database resource and search tools for comparative and evolutionary analyses of mitochondrial genomes in Metazoa. *Nucleic Acids Res*. 40:D1168–D1172.
- Denoeud F, et al. 2010. Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. *Science* 330: 1381–1385.
- Di Re M, et al. 2009. The accessory subunit of mitochondrial DNA polymerase gamma determines the DNA content of mitochondrial nucleoids in human cultured cells. *Nucleic Acids Res*. 37: 5701–5713.
- Drosophila* 12 Genomes Consortium, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203–218.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol*. 7: e1002195.
- Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc*. 2: 953–971.
- Euro L, Farnum GA, Palin E, Suomalainen A, Kaguni LS. 2011. Clustering of Alpers disease mutations and catalytic defects in biochemical variants reveal new features of molecular mechanism of the human mitochondrial replicase, Pol gamma. *Nucleic Acids Res*. 39: 9072–9084.
- Fahrein K, Masta SE, Podsiadlowski L. 2009. The first complete mitochondrial genome sequences of Amblypygi (Chelicerata: Arachnida) reveal conservation of the ancestral arthropod gene order. *Genome* 52: 456–466.
- Fan L, et al. 2006. A novel processive mechanism for DNA synthesis revealed by structure, modeling and mutagenesis of the accessory subunit of human mitochondrial DNA polymerase. *J Mol Biol*. 358: 1229–1243.
- Fan L, Kaguni LS. 2001. Multiple regions of subunit interaction in *Drosophila* mitochondrial DNA polymerase: three functional domains in the accessory subunit. *Biochemistry* 40:4780–4791.

- Fan L, Sanschagrin PC, Kaguni LS, Kuhn LA. 1999. The accessory subunit of mtDNA polymerase shares structural homology with aminoacyl-tRNA synthetases: implications for a dual role as a primer recognition factor and processivity clamp. *Proc Natl Acad Sci U S A*. 96: 9527–9532.
- Farnum GA, Nurminen A, Kaguni LS. 2014. Mapping 136 pathogenic mutations into functional modules in human DNA polymerase gamma establishes predictive genotype-phenotype correlations for the complete spectrum of POLG syndromes. *Biochim Biophys Acta*. 1837:1113–1121.
- Farr CL, Wang Y, Kaguni LS. 1999. Functional interactions of mitochondrial DNA polymerase and single-stranded DNA-binding protein. Template-primer DNA binding and initiation and elongation of DNA strand synthesis. *J Biol Chem*. 274:14779–14785.
- Foury F. 1989. Cloning and sequencing of the nuclear gene *MIP1* encoding the catalytic subunit of the yeast mitochondrial DNA polymerase. *J Biol Chem*. 264:20552–20560.
- Fukuoka A, et al. 2014. Screen for mitochondrial DNA copy number maintenance genes reveals essential role for ATP synthase. *Mol Syst Biol*. 10:734.
- Fuste JM, et al. 2010. Mitochondrial RNA polymerase is needed for activation of the origin of light-strand DNA replication. *Mol Cell*. 37: 67–78.
- García-Gómez S, et al. 2013. PrimPol, an archaic primase/polymerase operating in human cells. *Mol Cell*. 52:541–553.
- Gissi C, Iannelli F, Pesole G. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity (Edinb)*. 101:301–320.
- Hance N, Ekstrand MI, Trifunovic A. 2005. Mitochondrial DNA polymerase gamma is essential for mammalian embryogenesis. *Hum Mol Genet*. 14:1775–1783.
- Holt IJ, Lorimer HE, Jacobs HT. 2000. Coupled leading- and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100:515–524.
- Hyvarinen AK, et al. 2007. The mitochondrial transcription termination factor mTERF modulates replication pausing in human mitochondrial DNA. *Nucleic Acids Res*. 35:6458–6474.
- Hyvarinen AK, Pohjoismäki JL, Holt IJ, Jacobs HT. 2011. Overexpression of MTERFD1 or MTERFD3 impairs the completion of mitochondrial DNA replication. *Mol Biol Rep*. 38:1321–1328.
- Iyengar B, Luo N, Farr CL, Kaguni LS, Campos AR. 2002. The accessory subunit of DNA polymerase gamma is essential for mitochondrial DNA maintenance and development in *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 99:4483–4488.
- Iyengar B, Roote J, Campos AR. 1999. The *tamas* gene, identified as a mutation that disrupts larval behavior in *Drosophila melanogaster*, codes for the mitochondrial DNA polymerase catalytic subunit (DNApol-gamma125). *Genetics* 153:1809–1824.
- Jiang ZJ, et al. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol*. 7:123.
- Joers P, et al. 2013. Mitochondrial transcription terminator family members mTTF and mTerf5 have opposing roles in coordination of mtDNA synthesis. *PLoS Genet*. 9:e1003800.
- Joers P, Jacobs HT. 2013. Analysis of replication intermediates indicates that *Drosophila melanogaster* mitochondrial DNA replicates by a strand-coupled theta mechanism. *PLoS One* 8:e53249.
- Kaguni LS. 2004. DNA polymerase gamma, the mitochondrial replicase. *Annu Rev Biochem*. 73:293–320.
- Kamtekar S, Hecht MH. 1995. Protein Motifs. 7. The four-helix bundle: what determines a fold? *FASEB J*. 9:1013–1022.
- Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. *Brief Bioinform*. 9:286–298.
- Korhonen JA, Gaspari M, Falkenberg M. 2003. TWINKLE Has 5' → 3' DNA helicase activity and is specifically stimulated by mitochondrial single-stranded DNA-binding protein. *J Biol Chem*. 278: 48627–48632.
- Korhonen JA, Pham XH, Pellegrini M, Falkenberg M. 2004. Reconstitution of a minimal mtDNA replisome in vitro. *EMBO J*. 23:2423–2429.
- Lavrov DV, Boore JL, Brown WM. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol Biol Evol*. 17:813–824.
- Lecrenier N, Van Der Bruggen P, Foury F. 1997. Mitochondrial DNA polymerases from yeast to man: a new family of polymerases. *Gene* 185: 147–152.
- Lee YS, et al. 2010. Each monomer of the dimeric accessory protein for human mitochondrial DNA polymerase has a distinct role in conferring processivity. *J Biol Chem*. 285:1490–1499.
- Lee YS, Kennedy WD, Yin YW. 2009. Structural insight into processive human mitochondrial DNA synthesis and disease-related polymerase mutations. *Cell* 139:312–324.
- Lewis DL, Farr CL, Kaguni LS. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol Biol*. 4:263–278.
- Lewis DL, Farr CL, Wang Y, Lagina AT 3rd, Kaguni LS. 1996. Catalytic subunit of mitochondrial DNA polymerase from *Drosophila* embryos. Cloning, bacterial overexpression, and biochemical characterization. *J Biol Chem*. 271:23389–23394.
- Lewis SC, Joers P, Wilcox S, Griffith JD, Jacobs HT, Hyman BC. 2015. A rolling circle replication mechanism produces multimeric lariats of mitochondrial DNA in *Caenorhabditis elegans*. *PLoS Genet*. 11: e1004985.
- Li W-H. 1997. Molecular evolution. Sunderland (MA): Sinauer Associates.
- Lim SE, Longley MJ, Copeland WC. 1999. The mitochondrial p55 accessory subunit of human DNA polymerase gamma enhances DNA binding, promotes processive DNA synthesis, and confers N-ethylmaleimide resistance. *J Biol Chem*. 274:38197–38203.
- Logan DT, Mazauric MH, Kern D, Moras D. 1995. Crystal structure of glycyl-tRNA synthetase from *Thermus thermophilus*. *EMBO J*. 14: 4156–4167.
- Longley MJ, Prasad R, Srivastava DK, Wilson SH, Copeland WC. 1998. Identification of 5'-deoxyribose phosphate lyase activity in human DNA polymerase gamma and its role in mitochondrial base excision repair in vitro. *Proc Natl Acad Sci U S A*. 95: 12244–12248.
- Mitrevva M, Blaxter ML, Bird DM, McCarter JP. 2005. Comparative genomics of nematodes. *Trends Genet*. 21:573–581.
- Moraes CT, Bacman SR, Williams SL. 2014. Manipulating mitochondrial genomes in the clinic: playing by different rules. *Trends Cell Biol*. 24: 209–211.
- Oliveira MT, Garesse R, Kaguni LS. 2010. Animal models of mitochondrial DNA transactions in disease and ageing. *Exp Gerontol*. 45:489–502.
- Oliveira MT, Kaguni LS. 2010. Functional roles of the N- and C-terminal regions of the human mitochondrial single-stranded DNA-binding protein. *PLoS One* 5:e15379.
- Oliveira MT, Kaguni LS. 2011. Reduced stimulation of recombinant DNA polymerase gamma and mitochondrial DNA (mtDNA) helicase by variants of mitochondrial single-stranded DNA-binding protein (mtSSB) correlates with defects in mtDNA replication in animal cells. *J Biol Chem*. 286:40649–40658.
- Olson MW, Wang Y, Elder RH, Kaguni LS. 1995. Subunit structure of mitochondrial DNA polymerase from *Drosophila* embryos. Physical and immunological studies. *J Biol Chem*. 270:28932–28937.
- Philippe H, et al. 2009. Phylogenomics revives traditional views on deep animal relationships. *Curr Biol*. 19:706–712.
- Pohjoismäki JL, et al. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. *J Mol Biol*. 397:1144–1155.

- Reyes A, et al. 2013. Mitochondrial DNA replication proceeds via a “boot-lace” mechanism involving the incorporation of processed transcripts. *Nucleic Acids Res.* 41:5837–5850.
- Reyes A, Yang MY, Bowmaker M, Holt IJ. 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. *J Biol Chem.* 280:3242–3250.
- Ronquist F, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 61: 539–542.
- Sali A, Blundell TL. 1993. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol.* 234:779–815.
- Spelbrink JN, et al. 2000. In vivo functional analysis of the human mitochondrial DNA polymerase POLG expressed in cultured human cells. *J Biol Chem.* 275:24818–24828.
- Stiban J, Farnum GA, Hovde SL, Kaguni LS. 2014. The N-terminal domain of the *Drosophila* mitochondrial replicative DNA helicase contains an iron-sulfur cluster and binds DNA. *J Biol Chem.* 289: 24032–24042.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34:W609–W612.
- Szczepanowska K, Foury F. 2010. A cluster of pathogenic mutations in the 3′-5′ exonuclease domain of DNA polymerase gamma defines a novel module coupling DNA synthesis and degradation. *Hum Mol Genet.* 19:3516–3529.
- Wang Y, Farr CL, Kaguni LS. 1997. Accessory subunit of mitochondrial DNA polymerase from *Drosophila* embryos. Cloning, molecular analysis, and association in the native enzyme. *J Biol Chem.* 272: 13640–13646.
- Wang Y, Kaguni LS. 1999. Baculovirus expression reconstitutes *Drosophila* mitochondrial DNA polymerase. *J Biol Chem.* 274: 28972–28977.
- Wanrooij S, et al. 2008. Human mitochondrial RNA polymerase primes lagging-strand DNA synthesis in vitro. *Proc Natl Acad Sci U S A.* 105: 11122–11127.
- Wernette CM, Kaguni LS. 1986. A mitochondrial DNA polymerase from embryos of *Drosophila melanogaster*. Purification, subunit structure, and partial characterization. *J Biol Chem.* 261:14764–14770.
- Wolf YI, Koonin EV. 2001. Origin of an animal mitochondrial DNA polymerase subunit via lineage-specific acquisition of a glycyl-tRNA synthetase from bacteria of the *Thermus-Deinococcus* group. *Trends Genet.* 17:431–433.
- Yakubovskaya E, Chen Z, Carrodeguas JA, Kisker C, Bogenhagen DF. 2006. Functional human mitochondrial DNA polymerase gamma forms a heterotrimer. *J Biol Chem.* 281:374–382.
- Yang MY, et al. 2002. Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. *Cell* 111:495–505.
- Yasukawa T, et al. 2006. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. *EMBO J.* 25:5358–5371.
- Ye F, Samuels DC, Clark T, Guo Y. 2014. High-throughput sequencing in mitochondrial DNA research. *Mitochondrion* 17C:157–163.
- Young MJ, et al. 2011. Biochemical analysis of human POLG2 variants associated with mitochondrial disease. *Hum Mol Genet.* 20: 3052–3066.

Associate editor: Sarah Schaack