

On the Origin of the Eukaryotic Chromosome: The Role of Noncanonical DNA Structures in Telomere Evolution

Miguel Garavís^{1,2}, Carlos González², and Alfredo Villasante^{1,*}

¹Centro de Biología Molecular “Severo Ochoa” (CSIC-UAM), Universidad Autónoma de Madrid, Nicolás Cabrera 1, 28049 Madrid, Spain

²Instituto de Química Física Rocasolano, CSIC, C/Serrano 119, 28006 Madrid, Spain

*Corresponding author: E-mail: avillasante@cbm.uam.es.

Associate editor: Bill Martin

Accepted: May 16, 2013

Abstract

The transition of an ancestral circular genome to multiple linear chromosomes was crucial for eukaryogenesis because it allowed rapid adaptive evolution through aneuploidy. Here, we propose that the ends of nascent linear chromosomes should have had a dual function in chromosome end protection (capping) and chromosome segregation to give rise to the “proto-telomeres.” Later on, proper centromeres evolved at subtelomeric regions. We also propose that both noncanonical structures based on guanine–guanine interactions and the end-protection proteins recruited by the emergent telomeric heterochromatin have been required for telomere maintenance through evolution. We further suggest that the origin of *Drosophila* telomeres may be reminiscent of how the first telomeres arose.

Key words: telomeres, centromeres, G-quadruplexes, non-LTR retrotransposons.

The Necessity of Both End Protection and Segregation Functions at the End of Nascent Linear Chromosomes

It has been hypothesized that after the endosymbiosis of an α -proteobacteria into an archaeobacterial host, the massive invasion of the symbiont’s mobile group II introns into the circular genome of the host gave rise to spliceosomal introns (Koonin 2006) and that this was the driving force for the origin of the nucleus (fig. 1). The invention of the nuclear membrane was necessary to physically separate slow splicing from fast translation (Martin and Koonin 2006) (fig. 1). It has also been proposed that this invasion eventually led to the origin of linear eukaryotic chromosomes (Villasante, Abad, et al. 2007; Villasante, Méndez-Lago, et al. 2007). The host’s tolerance to the mobile element invasion and to other eukaryotic innovations could be facilitated by a low effective population size and the consequent weak purifying selection (Lynch and Conery 2003; Lynch 2007; Koonin 2011).

Mobile group II introns are retroelements of eubacterial origin that contain a catalytic RNA and a multifunctional protein with reverse transcriptase (RT) activity (Lambowitz and Zimmerly 2004). These retroelements are thought to be the ancestors of both spliceosomal introns and nonlong terminal

repeat (non-LTR) retrotransposons (Sharp 1985). The evolutionary relationship between group II introns and non-LTR-retrotransposons is based on the similarity of their RT sequences (Xiong and Eickbush 1990; Blocker et al. 2005) and retrotransposition mechanisms (Luan et al. 1993; Zimmerly et al. 1995). After the initial proliferation of group II introns within the protoeukaryotic nuclear genome, their RNA domains degenerated and evolved into spliceosomal snRNAs that functioned in *trans* in a common splicing apparatus (Sharp 1991; Mohr et al. 2010). Although most group II introns evolved as eukaryotic introns, some lost their splicing capability and gave rise to non-LTR-retrotransposons.

It is likely that the continuous breakage of the presumed circular chromosome activated all the mechanisms of DNA repair, including the one mediated by non-LTR retrotransposons (Moore and Haber 1996; Morrish et al. 2002). In this evolutionary scenario, it has been hypothesized that the repetitive capture of non-LTR retrotransposons, with a G/C strand bias, at the ends of DNA double-strand breaks (DSBs) could have eventually resulted in end protection (capping), instead of repair, giving rise to the “proto-telomeres” of the first linear chromosomes (fig. 1) (Villasante, Abad, et al. 2007). The biased distribution of guanine and cytosine between the two strands could have been selected because G-rich

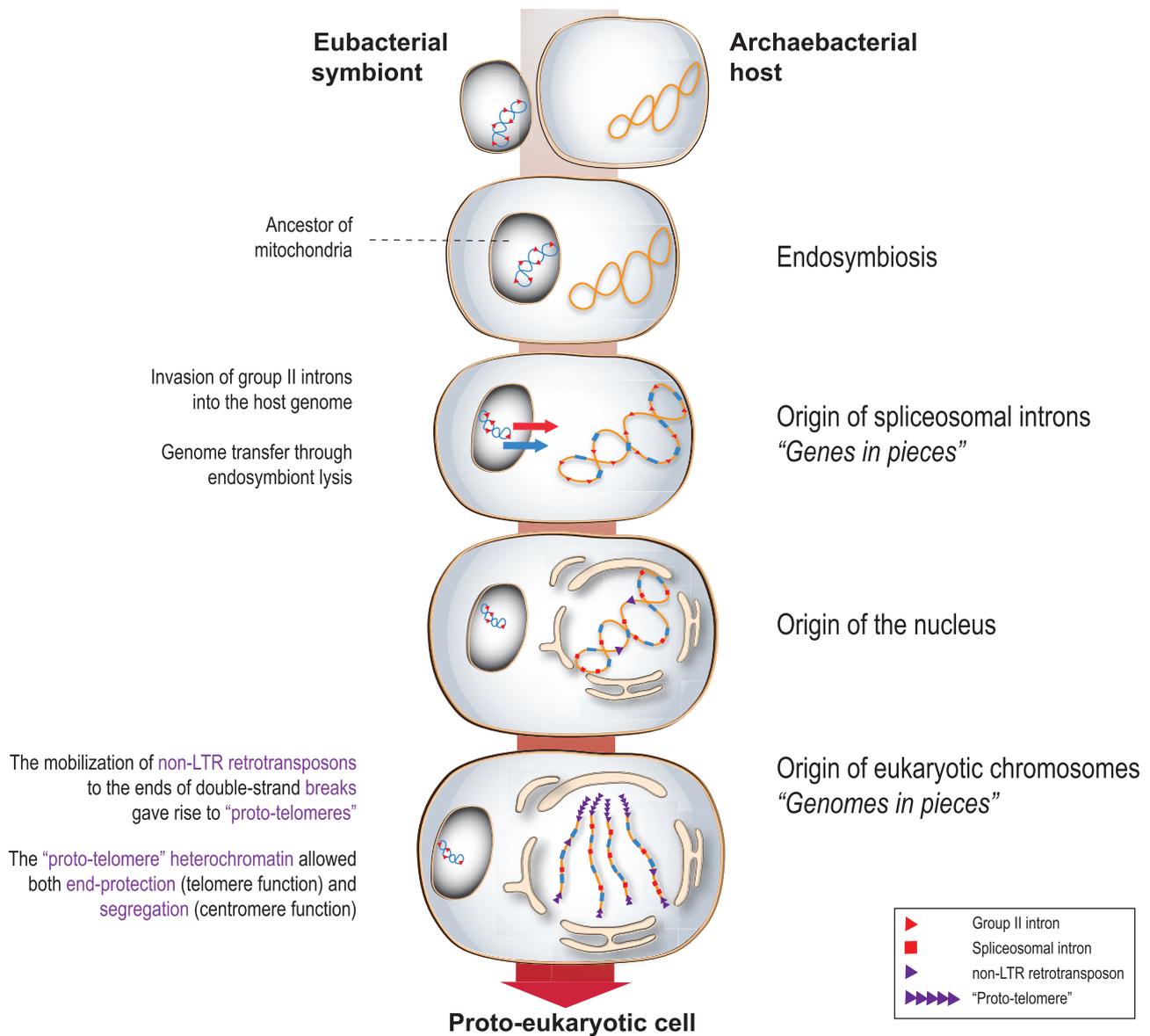


FIG. 1.—Schematic representation of a possible evolutionary scenario for the origin of eukaryotic chromosomes. The scheme shows the origin of spliceosomal introns and the origin of the nucleus as hypothesized (Koonin 2006; Martin and Koonin 2006) and also the proposed origin of the ancestral "proto-telomere" with capping and segregation properties. The eubacterial genome appears in blue and the archaeobacterial genome in orange.

sequences have the intrinsic capacity to fold into noncanonical secondary structures that were utilized for capping or sequestering chromosome ends (Villasante, Abad, et al. 2007; Villasante, Méndez-Lago, et al. 2007). In addition, the iterative transposition generated the first terminal repeats and that also allowed the elongation of chromosome ends by the existing mechanisms of homologous recombination (de Lange 2004). As will be described later, a similar situation occurs in *Drosophila*, where telomeres are maintained by retrotransposition of G/C strand-biased non-LTR elements and by terminal recombination (Mason et al. 2008; Villasante et al. 2008).

The universal 5'–3' nuclease-mediated cleavage of DNA ends created 3'-single-stranded overhangs that were coated by an abundant single-stranded DNA (ssDNA)-binding protein (replication protein A or RPA) via its characteristic oligosaccharide/oligonucleotide-binding fold (OB fold) domains. This nonsequence-specific ssDNA-binding protein is required for multiple processes such as DNA replication, DNA repair, and DNA damage signaling. Therefore, to protect the DNA ends and to avoid their repair, a specialized sequence-specific ssDNA-binding protein should evolve to coat the 3'-G-rich overhangs (Gelinas et al. 2009) and to promote their folding into non-B DNA conformations, which could be (but not

necessarily) quadruplex-like structures. Recent studies have shown that molecular crowding (Heddi and Phan 2011; Xu et al. 2011) or high viscosity conditions (Lannan et al. 2012) stabilize G-quadruplexes, suggesting that cell environment may facilitate the formation of quadruplex-like structures.

On the other hand, a cell-cycle-regulated switch of those ssDNA-binding proteins was needed to re-establish telomere capping after DNA replication, and now it is known that telomeric repeat-containing RNA (TERRA) contributes to induce that switch (Flynn et al. 2011). TERRA is also required to organize a special chromatin structure: the telomeric heterochromatin (Azzalin et al. 2007; Schoeftner and Blasco 2008; Deng et al. 2009; Shpiz et al. 2011).

Here, it is fundamental to notice that “proto-telomeres” with dual function in capping and segregation were required to ensure accurate inheritance of the first linear eukaryotic chromosomes (fig. 1). Thus, the formation of the first heterochromatin at nascent chromosome ends should have facilitated both the recruitment of end-protection proteins and the attachment of spindle microtubules, most likely by means of ribonucleoprotein complexes (Villasante, Abad, et al. 2007; Villasante, Méndez-Lago, et al. 2007). Later on in eukaryogenesis, a mature segregation function evolved at subtelomeric regions. The mechanism of unequal exchange and gene conversion led inevitably to the divergence of the internal subtelomeric repeats, and the strand asymmetry of the repeats provided the potential to form the sequence-independent secondary structures that gave rise to the centromeres (Villasante, Abad, et al. 2007; Villasante, Méndez-Lago, et al. 2007). In this scenario, the recurrent appearance of unstable dicentric chromosomes, through the formation of new centromeres from telomeres, provided an additional mechanism of genome fragmentation. The birth of multiple eukaryotic linear chromosomes was the key innovation that allowed adaptive evolution by means of transient aneuploidy (chromosomal duplications) (Chen et al. 2012; Yona et al. 2012).

If primitive centromeres evolved from “proto-telomeres,” it would be reasonable to expect that telomeric regions may also have some centromere-like properties. Indeed, there are already results that seem to support this assertion.

1) In *Schizosaccharomyces pombe*, deletion of an endogenous centromere leads to neocentromere formation at subtelomeric regions (Ishii et al. 2008), and it has been shown that their centromeric and subtelomeric DNA sequences must possess particular features that promote the incorporation of the centromere-specific histone 3 variant (CENP-A) (Choi et al. 2012).

2) In *Drosophila melanogaster*, the centromere of the Y chromosome contains a large array of *HeT-A*- and *TART*-derived telomeric retrotransposons (Agudo et al. 1999), and the sequence of this satellite DNA has revealed that this centromeric region evolved from a telomere (Méndez-Lago et al. 2009). Furthermore, overexpression of the *Drosophila* CENP-A

induces preferential formation of neocentromeres near telomeres (Heun et al. 2006; Olszak et al. 2011).

3) In some plants and animals, neocentromere activity appears at subtelomeric heterochromatin during meiosis (reviewed in Puertas and Villasante 2013).

4) The evolutionary history of chromosome 3 in primates shows at least three examples of telomere–centromere functional interchange (Ventura et al. 2004). Similarly, other telomere-to-centromere conversions have been described after the comparative analysis of eight mammalian genomes (Murphy et al. 2005). Because the subtelomeric repeats could have a role in these conversions, this chromosomal behavior could be due to the ancestral centromeric competence of a telomeric region.

Similarly, if primitive centromeres began at DSBs, one could wonder whether the dynamic chromatin formed around breakage sites could have centromere-like features. Here, too, there are results in favor of this consideration.

1) It has been shown that the centromeric proteins CENP-A, CENP-N, CENP-T, and CENP-U are rapidly recruited to DSBs (Zeitlin et al. 2009) and has been hypothesized that, under certain circumstances, this recruitment could generate a neocentromere (Zeitlin et al. 2009).

2) Strikingly, it had been previously noticed that several human neocentromeres were located near breakpoints and had been hypothesized that these breaks could induce the emergence of neocentromeres (Ventura et al. 2003; Marshall et al. 2008).

The previous hypothesis for the origin of the eukaryotic chromosome proposed that centromeres arose before telomeres and that probably evolved from the origin of replication region of the bacterial chromosome (Cavalier-Smith 1981). Recently, Cavalier-Smith (2010) has still suggested that centromeres arose first and has proposed that they originated from the partitioning locus, a region proximal to the bacterial origin of replication implicated in bacterial chromosome partitioning/segregation. But he did not say how the fragmented prokaryotic genome could give rise to a centromere on each nascent linear chromosome and what was the hypothetical process that led to the formation of regional centromeres containing repetitive DNA. In support of an ancestral regional centromere, a recent study in *Saccharomyces cerevisiae* has found centromere-like regions (without a specific DNA sequence) in close proximity to the native point centromere (Lefrançois et al. 2013). Because these small regions promote proper segregation, possibly through sequence-independent centromeric structures, they seem to be evolutionary remnants derived from a regional centromere rather than from a point centromere (Lefrançois et al. 2013).

To recapitulate, in this section, we have proposed that the origin of linear chromosomes (genomes in pieces) was a eukaryotic innovation generated by the mobilization of group II intron-derived retroelements as a response to endosymbiosis stress (McClintock 1984; Koonin 2011). Specifically, we have

hypothesized that the repetitive capture of G/C strand biased non-LTR retrotransposons at the ends of DSBs gave rise to “proto-telomeres,” a primitive terminal heterochromatic structure with a dual function: end protection (telomeric function) and segregation (centromeric function) (fig. 1).

Noncanonical DNA Structures Based on Guanine–Guanine Interactions Seem to Have Played a Role in Telomere Origin and Evolution

In most eukaryotic chromosomes, telomere DNA sequences are arrays of short guanine-rich repetitive sequences that terminate in a 3′-single-strand G-rich overhang (150–200 nucleotides). The G-rich strand is synthesized by a telomere-specific RT, called telomerase, using a small region of its RNA subunit as template and the 3′-OH on the end of the chromosome as a primer (Blackburn 1992). Found in animals, fungi, and Amoebozoa, TTAGGG was the telomeric simple repeat sequence present in the ancestral Unikont. Moreover, its occurrence in some species of the supergroups Plantae, Chromalveolata, Excavata, and Rhizaria suggests that TTAGGG could be the ancestral telomeric repeat sequence for eukaryotes (Fulnecková et al. 2013).

On the other hand, the use of prokaryotic retroelements to root a RT phylogenetic tree shows that the telomerase seems to have evolved from the RT of an ancestral non-LTR retrotransposon (Eickbush 1997). Furthermore, it is believed that the ability of non-LTR RTs to use the 3′-OH of chromosome ends to prime reverse transcription was crucial for the birth of early telomerases (Moore and Haber 1996; Morrish et al. 2002, 2007; Curcio and Belfort 2007).

Telomerase-based telomeres brought two principal advantages: facilitated telomere homeostasis and a greater structural protection by the incorporation of simple G-rich repeats

with the inherent ability to form G-quadruplex structures (Henderson et al. 1987; Arthanari and Bolton 2003; Teixeira and Gilson 2005). G-quadruplexes consist of stacked G-quartets, which are planar arrangements of four guanines held together by Hoogsteen hydrogen bonds (Neidle 2009) (fig. 2A). G-quadruplex formation may occur within the terminal G-rich 3′-overhang (fig. 2B) or when the overhang invades the adjacent double-stranded region of the telomere to form T-loop structures (fig. 2C) (Maizels 2006; Rhodes 2006; Xu et al. 2008; Bochman et al. 2012). Nevertheless, it has been hypothesized that after the appearance of telomerase, the maintenance of telomeres by the primitive T-loop-replication mechanism becomes less relevant (de Lange 2004). The first visualization of telomeric G-quadruplex formation in vivo was performed in the ciliate *Stylonychia* (Schaffitzel et al. 2001; Paeschke et al. 2005). Most recently, a highly specific DNA G-quadruplex antibody has been employed to visualize G-quadruplex structures at human telomeres (Biffi et al. 2013).

It is important to point out that the putative ancestral telomerase-synthesized sequence, $(TTAGGG)_n$, is not only capable of folding into a G-quadruplex structure but is the best one at doing so in vitro (Tran et al. 2011). In addition, recent biophysical studies on the folding of these telomeric G-quadruplexes have shown that structure formation occurs in milliseconds. These folding kinetics are biologically relevant because they are comparable to those of transcription and DNA replication (Zhang and Balasubramanian 2012).

During evolution, mutations in the telomerase RNA template have given rise to repeat variants with different lengths of guanine motifs (G_2 , G_4 , and more). Recent experiments have found that the G-quadruplexes formed by telomeric repeats with only two consecutive guanines (TTAGG in arthropods and TTAGGC in nematodes) are in equilibrium with G-hairpins and other noncanonical structures (Tran et al. 2011).

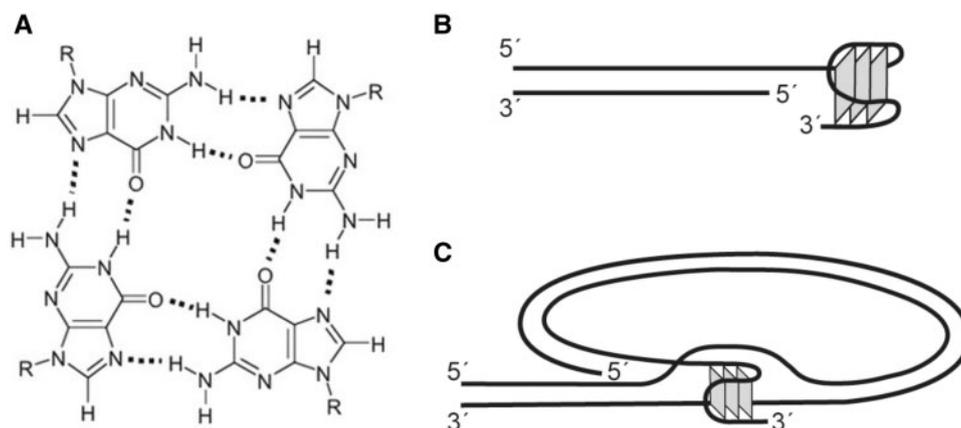


FIG. 2.—Schematic diagrams of a G-quartet and two telomeric G-quadruplexes. (A) Four guanines assemble in a planar arrangement to form a G-quartet. Hydrogen bonds are in dashed lines. (B) Diagram of an intramolecular G-quadruplex at a telomere end. (C) Diagram of a G-quadruplex at a T-loop. The G-quadruplexes in the figure are composed of three stacked G-quartets (shaded squares).

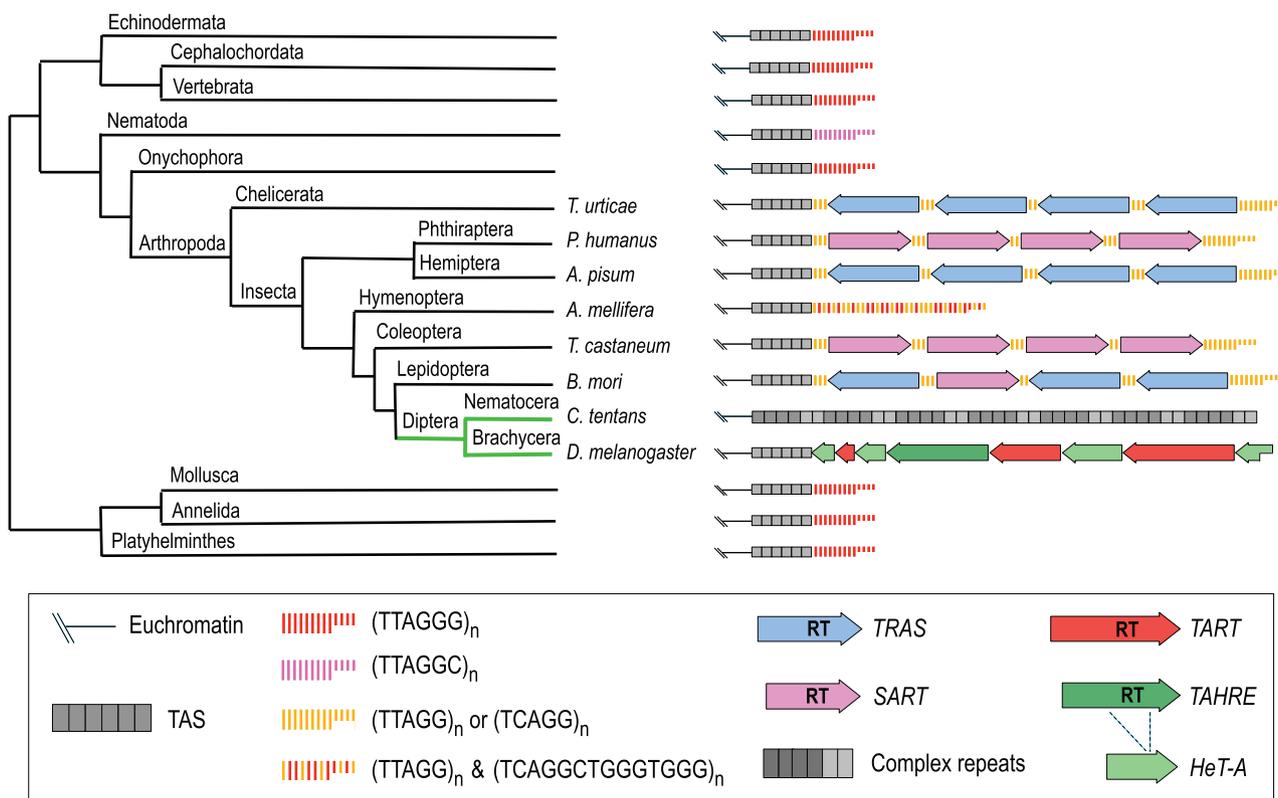


Fig. 3.—Distribution of telomeric sequences within Bilateria. Most eukaryotes have G-rich telomerase-synthesized repeats with adjacent complex subtelomeric repeats called telomere-associated sequences (TAS). In most arthropods, telomere-specific retrotransposons are inserted into telomerase-synthesized repeats. As can be seen in the diagram, *TRAS* elements insert in reverse orientation to that of the *SART* elements. In an ancestor of diptean insects, the telomerase gene was lost (green line). In *Chironomus tentans* (lower Diptera), the telomeric sequences consist of complex tandem repeats maintained by recombination. However, *Drosophila* species have multiple telomere-specific retrotransposons (autonomous and nonautonomous) that transpose to chromosomal ends. The deletion event in the ancestral *TAHRE* element is shown with dashed lines.

In the silk moth *Bombyx mori* (Lepidoptera) and the flour beetle *Tribolium castaneum* (Coleoptera), the telomerase activity is weak, and telomere-specific non-LTR retroelements (*TRAS* and *SART* family elements in *B. mori* and *SART* family elements in *T. castaneum*) are inserted into the telomeric repeats in a specific manner (Fujiwara et al. 2005; Osanai et al. 2006) that preserves the G/C strand bias (fig. 3). The massive integration of these elements into the proximal regions of the *TTAGG* repeat arrays of *B. mori* and the *TCAGG* arrays of *T. castaneum* (an alternative telomere variant in insects) gives rise to huge telomeres with sizes larger than 200 kb.

The telomeres of the honey bee *Apis mellifera* (Hymenoptera) are exceptional among the arthropods because they do not have non-LTR elements inserted into their telomeric repeats. Instead, the telomere sequence consists of *TTAGG* repeat arrays (Robertson and Gordon 2006) interspersed with *TCAGGCTGGG*, *TCAGGCTGGGTGGG*, and *TCAGGCTGGGTGAGGATGGG* higher order repeat arrays (Garavís M, Villasante A, unpublished results) (fig. 3). These higher order repeats arose by amplification of the mutated repeats present in proximal telomeric regions and the

interspersed pattern developed by further amplifications of the 5-bp repeat arrays together with higher order repeat arrays. However, the *TTAGG* repeats of *Acyrtosiphon pisum* (Hemiptera) and *Pediculus humanus* (Phthiraptera) contain insertions of non-LTR retrotransposons of the *TRAS* and *SART* family, respectively (International Aphid Genomics Consortium 2010; Kirkness et al. 2010) (fig. 3). As Hemiptera and Phthiraptera are basal to Hymenoptera, Coleoptera, Lepidoptera, and Diptera (fig. 3), the telomeres of *A. mellifera* seem to represent a case where the *TRAS* and/or *SART* retrotransposons were lost at a later stage in evolution. It is tempting to speculate that the appearance of those higher order repeat arrays with propensity to form 3-quartet G-quadruplexes caused the decay, and eventual loss, of the telomeric retrotransposons.

Because the telomeres of the spider mite *Tetranychus urticae* (from the basal branch Chelicerata) are also a mosaic of short *TTAGG* repeats interrupted by non-LTR retrotransposons closely related to *TRAS* (Grbić et al. 2011) (fig. 3), the telomeres of the arthropods seem to be maintained by telomerase, by insertion of specific non-LTR retrotransposons into the

TTAGG repeat array and by recombination. The same system of telomere maintenance has also been found in some nonarthropod species (Arkhipova and Morrison 2001; Yamamoto et al. 2003; Gladyshev and Arkhipova 2007; Starnes et al. 2012). It has not escaped our notice that the appearance of this apparently suboptimal mechanism of telomere maintenance, which might have created chromosome instability, seems to have coincided with the great arthropod radiation into Chelicerata and Mandibulata.

On the other hand, certain yeasts from the *Ascomycota* phylum have telomeric repeats that are diverse in terms of their sequence, length, and homogeneity (McEachern and Blackburn 1994). In these yeasts, the degenerate repeats result from the nonprocessivity of their telomerases (Prescott and Blackburn 1997). Importantly, these repeats, despite their TG-richness, are less prone to fold into G-quadruplexes (Tran et al. 2011), and it has been shown that in these organisms, the telomere-binding proteins are fast evolving (Teixeira and Gilson 2005). Therefore, it is possible that these yeasts are using an ancestral system of chromosome end protection where the ssDNA-binding proteins facilitate the folding of the 3'-overhangs into G-quadruplex-like structures. This yeast-capping mechanism likely arose de novo by convergent evolution.

Do Noncanonical Secondary Structures Have a Role in the Maintenance of Telomeres without Telomerase?

Once telomerase becomes completely dysfunctional, the gene encoding telomerase could be lost if telomeres are maintained by the ancestral alternative mechanism of homologous recombination. Apparently, this is what happened in the ancestor of Diptera about 260 Ma (Wiegmann et al. 2011). In the lower Diptera, *Anopheles*, *Rhynchosciara*, and *Chironomus*, long tandem repeats are present at chromosome ends, suggesting that telomere maintenance takes place by homologous recombination (Nielsen and Edstrom 1993; Biessmann et al. 1998; Madalena et al. 2010) (fig. 3). In *Drosophila*, however, telomere maintenance occurs primarily by transposition of telomere-specific retrotransposons to receding chromosome ends (fig. 3). In addition to retrotransposition, *Drosophila* telomeres are also maintained, as in any eukaryote, by recombination/gene conversion (Kahn et al. 2000).

In *D. melanogaster*, three telomeric retrotransposons *TART*, *TAHRE*, and *HeT-A* (a nonautonomous element derived from an ancestral *TAHRE* that lost its RT), transpose occasionally to chromosome ends using the free 3'-OH at chromosome termini to prime reverse transcription (Biessmann et al. 1990, 1992; Sheen and Levis 1994; Abad et al. 2004a, 2004b). In agreement with this mechanism, the telomeric elements appear randomly mixed in head-to-tail arrangements and variably truncated at the 5'-end (Mason et al. 2008; Villasante et al. 2008; Pardue and DeBaryshe 2011) (fig. 3). It is noteworthy that deletion of the RT coding region of the telomeric

elements has occurred recurrently during *Drosophila* evolution, and multiple nonautonomous elements appear at the telomeres of the *Drosophila* species examined. As an example, up to four nonautonomous elements along with their corresponding autonomous elements have been found in *D. mojavensis* telomeres (Villasante et al. 2007). Interestingly, similar situations occur with group II introns where their RTs also act in *trans* to mobilize multiple deleted introns (Mohr et al. 2010).

Because *Drosophila* telomeres consist of retrotransposon arrays in constant flux, there is not a specific terminal sequence and their telomere-capping proteins (the "terminin" complex) have evolved to bind chromosome ends independently of the primary DNA sequence (Raffa et al. 2009, 2010). The "terminin" complex is functionally analogous to the "shelterin" complex (human telomere-capping proteins), but their components are not evolutionarily conserved (Palm and de Lange 2008; Raffa et al. 2009, 2010). Thus, *Drosophila* telomeres are made of rapidly evolving telomeric retrotransposons (Villasante et al. 2007) and telomere-capping proteins (Gao et al. 2010; Raffa et al. 2010). Moreover, as Verrochio is a telomere-capping protein with one OB-fold domain and all telomeric proteins containing OB folds are 3'-overhang binding proteins, *Drosophila* telomeres also seem to have single-strand overhangs (Raffa et al. 2010).

It is noteworthy that, despite the complexity of telomeric sequences in the genus *Drosophila* and *Chironomus*, the *Drosophila* telomeric retrotransposon arrays and the *Chironomus* telomeric complex repeats also have the telomeric G/C strand bias (Nielsen and Edstrom 1993; Danilevskaia et al. 1998). The conservation of this G/C strand bias may indicate that telomere capping depends on the formation of noncanonical structures based on guanine-guanine interactions. In agreement with this idea, it has been shown that the 3'-untranslated region of the abundant *D. melanogaster* telomeric element *HeT-A* contains sequences with propensity to form G-quadruplexes (Abad and Villasante 1999).

The structural and phylogenetic analyses of all *Drosophila* telomeric-specific retrotransposons show that they had a common ancestor and indicate that non-LTR retrotransposons have been recruited to perform the cellular function of telomere maintenance. Therefore, we propose that the recruitment of *Drosophila* telomeric elements may resemble the ancestral mechanism that led to the maintenance of the "proto-telomeres" of the first eukaryotic chromosomes.

On the other hand, it has been found that yeast cells lacking telomerase can survive telomere sequence loss through the formation of terminal blocks of heterochromatin. This happens by amplifying and rearranging either subtelomeric sequences in *S. cerevisiae* and *S. pombe* or rDNA sequences in *S. pombe* (Lundblad and Blackburn 1993; Jain et al. 2010). Significantly, the *S. cerevisiae* subtelomeric Y' repeats also have purine/pyrimidine strand bias (Nickles and McEachern 2004), and the *S. pombe* end-protection protein POT1

(protection of telomeres 1) binds, in a nonsequence-specific manner, to the 3'-overhangs of G-rich rDNA (Jain et al. 2010). Interestingly, adaptive recombination-based mechanisms of telomere maintenance (called ALT for alternative lengthening of telomeres) also occur in tumor cells that lack telomerase (Bryan et al. 1995; Cesare and Reddel 2010).

To summarize, in species that have lost telomerase either during evolution (order Diptera) or through experimental manipulation, the data available suggest a role of structural DNA features in telomere maintenance, reveal the importance of telomeric heterochromatin (regardless of the underlying primary sequence) in the recruitment of end-binding proteins, and show how easily backup mechanisms may have been used to maintain telomeres during evolution.

Conclusions

We have discussed how genomes have exploited their noncoding structural potential to establish primordial innovations during eukaryogenesis. In particular, how the highly polymorphic secondary structures based on guanine-guanine interactions have evolved in concert with proteins to allow the origin and evolution of telomeres. In addition, we have hypothesized that the first linear eukaryotic chromosomes arose by the appearance of "proto-telomeres" at DSBs. This ancestral terminal structure had the dual function of end protection and segregation. Furthermore, we have discussed that the concomitant "proto-telomere" heterochromatin formation was fundamental for this key evolutionary innovation. Once again, the study of a noncanonical mechanism, like the maintenance of *Drosophila* telomeres, has generated new insights into the evolution of eukaryotes.

Acknowledgments

The authors thank Jim Mason for discussions on the origin of *Drosophila* telomeres, Maria J. Puertas for discussions on the heterochromatin of meiotic neocentromeres, and Douglas V. Laurents for revision of the manuscript. They are grateful to the reviewers for their valuable comments and suggestions. They apologize to those whose work was not cited in this review. This work was supported by the FPI fellowship BES-2009-027909 from the Ministerio de Ciencia e Innovación to M.G., by the Ministerio de Ciencia e Innovación (CTQ2010-21567-C02-02) to C.G., the Ministerio de Economía y Competitividad (BFU2011-30295-C02-01) to A.V., and by an institutional grant from the Fundación Ramón Areces to the Centro de Biología Molecular "Severo Ochoa."

Literature Cited

- Abad JP, et al. 2004a. Genomic analysis of *Drosophila melanogaster* telomeres: full-length copies of HeT-A and TART elements at telomeres. *Mol Biol Evol.* 21:1613–1619.
- Abad JP, et al. 2004b. TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol Biol Evol.* 21:1620–1624.
- Abad JP, Villasante A. 1999. The 3' non-coding region of the *Drosophila melanogaster* HeT-A telomeric retrotransposon contains sequences with propensity to form G-quadruplex DNA. *FEBS Lett.* 453:59–62.
- Agudo M, et al. 1999. Centromeres from telomeres? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. *Nucleic Acids Res.* 27:3318–3324.
- Arkhipova IR, Morrison HG. 2001. Three retrotransposon families in the genome of *Giardia lamblia*: two telomeric, one dead. *Proc Natl Acad Sci U S A.* 98:14497–14502.
- Arthanari H, Bolton PH. 2003. Did quadruplex DNA play a role in the evolution of the eukaryotic linear chromosome? *Mini Rev Med Chem.* 3:1–9.
- Azzalin CM, Reichenbach P, Khoraiuli L, Giulotto E, Lingner J. 2007. Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 318:798–780.
- Biessmann H, et al. 1990. Addition of telomere-associated HeT DNA sequences "heals" broken chromosome ends in *Drosophila*. *Cell* 61:663–763.
- Biessmann H, et al. 1992. HeTA, a transposable element specifically involved in "healing" broken chromosome ends in *Drosophila melanogaster*. *Mol Cell Biol.* 12:3910–3918.
- Biessmann H, Kobeski F, Walter MF, Kasravi A, Roth CW. 1998. DNA organization and length polymorphism at the 2L telomeric region of *Anopheles gambiae*. *Insect Mol Biol.* 7:83–93.
- Biffi G, Tannahill D, McCafferty J, Balasubramanian S. 2013. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem.* 5:182–186.
- Blackburn EH. 1992. Telomerases. *Annu Rev Biochem.* 61:113–129.
- Blocker FJ, et al. 2005. Domain structure and three-dimensional model of a group II intron-encoded reverse transcriptase. *RNA* 11:14–28.
- Bochman ML, Paeschke K, Zakian VA. 2012. DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet.* 13:770–780.
- Bryan TM, Englezou A, Gupta J, Bacchetti S, Reddel RR. 1995. Telomere elongation in immortal human cells without detectable telomerase activity. *EMBO J.* 4:4240–4248.
- Cavalier-Smith T. 1981. The origin and early evolution of the eukaryotic cell. In: Carlile MJ, Collins JF, Moseley BEB, editors. *Molecular and cellular aspects of microbial evolution*. Cambridge: Cambridge University Press. p. 33–84.
- Cavalier-Smith T. 2010. Origin of the cell nucleus, mitosis and sex: roles of intracellular coevolution. *Biol Direct.* 5:7.
- Cesare AJ, Reddel RR. 2010. Alternative lengthening of telomeres: models, mechanisms and implications. *Nat Rev Genet.* 1:319–330.
- Chen G, Rubinstein B, Li R. 2012. Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *Bioessays* 34:893–900.
- Choi ES, et al. 2012. Factors that promote H3 chromatin integrity during transcription prevent promiscuous deposition of CENP-A (Cnp1) in fission yeast. *PLoS Genet.* 8(9):e1002985.
- Curcio MJ, Belfort M. 2007. The beginning of the end: links between ancient retroelements and modern telomerases. *Proc Natl Acad Sci U S A.* 104:9107–9108.
- Danilevskaya ON, Lowenhaupt K, Pardue ML. 1998. Conserved subfamilies of the *Drosophila* HeT-A telomere-specific retrotransposon. *Genetics* 148:233–242.
- Deng Z, Norseen J, Wiedmer A, Riethman H, Lieberman PM. 2009. TERRA RNA binding to TRF2 facilitates heterochromatin formation and ORC recruitment at telomeres. *Mol Cell.* 35:403–413.

- Eickbush TH. 1997. Telomerase and retrotransposons: which came first? *Science* 277:911–912.
- Flynn RL, et al. 2011. TERRA and hnRNP A1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature* 471:532–536.
- Fujiwara H, Osanai M, Matsumoto T, Kojima KK. 2005. Telomere specific non-LTR retrotransposons and telomere maintenance in the silkworm, *Bombyx mori*. *Chromosome Res.* 13:455–467.
- Fulnecková J, et al. 2013. A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biol Evol.* 5: 468–483.
- Gao G, et al. 2010. HipHop interacts with HOAP and HP1 to protect *Drosophila* telomeres in a sequence-independent manner. *EMBO J.* 29:819–829.
- Gelinas AD, et al. 2009. Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain. *Proc Natl Acad Sci U S A.* 106:19298–19303.
- Gladyshev EA, Arkhipova IR. 2007. Telomere-associated endonuclease-deficient Penelope-like retroelements in diverse eukaryotes. *Proc Natl Acad Sci U S A.* 104:9352–9357.
- Grbić M, et al. 2011. The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature* 479:487–492.
- Heddi B, Phan AT. 2011. Structure of human telomeric DNA in crowded solution. *J Am Chem Soc.* 133:9824–9833.
- Henderson E, Hardin CC, Walk SK, Tinoco I Jr, Blackburn EH. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* 51:899–908.
- Heun P, et al. 2006. Mislocalization of the *Drosophila* centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell.* 10:303–315.
- International Aphid Genomics Consortium. 2010. Genome sequence of the pea aphid *Acyrtosiphon pisum*. *PLoS Biol.* 8:e1000313.
- Ishii K, et al. 2008. Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science* 321:1088–1091.
- Jain D, Hebden AK, Nakamura TM, Miller KM, Cooper JP. 2010. HAATI survivors replace canonical telomeres with blocks of generic heterochromatin. *Nature* 467:223–227.
- Kahn T, Savitsky M, Georgiev P. 2000. Attachment of HeT-A sequences to chromosomal termini in *Drosophila melanogaster* may occur by different mechanisms. *Mol Cell Biol.* 20:7634–7642.
- Kirkness EF, et al. 2010. Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc Natl Acad Sci U S A.* 107:12168–12173.
- Koonin EV. 2006. The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate? *Biol Direct.* 1:22.
- Koonin EV. 2011. The logic of chance: the nature and origin of biological evolution. Upper Saddle River (NJ): FT Press.
- Lambowitz AM, Zimmerly S. 2004. Mobile group II introns. *Annu Rev Genet.* 38:1–35.
- de Lange T. 2004. T-loops and the origin of telomeres. *Nat Rev Mol Cell Biol.* 5:323–329.
- Lannan FM, Mamajanov I, Hud NV. 2012. Human telomere sequence DNA in water-free and high-viscosity solvents: G-quadruplex folding governed by Kramers rate theory. *J Am Chem Soc.* 134:15324–15330.
- Lefrançois P, Auerbach RK, Yellman CM, Roeder GS, Snyder M. 2013. Centromere-like regions in the budding yeast genome. *PLoS Genet.* 9:e1003209.
- Luan DD, Korman MH, Jakubczak JL, Eickbush TH. 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72:595–605.
- Lundblad V, Blackburn EH. 1993. An alternative pathway for yeast telomere maintenance rescues est1-senescence. *Cell* 73:347–360.
- Lynch M. 2007. The origins of genome architecture. Sunderland (MA): Sinauer Associates.
- Lynch M, Conery JS. 2003. The origins of genome complexity. *Science* 302:1401–1404.
- Madalena CRG, Amabis JM, Gorab E. 2010. Unusually short tandem repeats appear to reach chromosome ends of *Rhynchosciara americana* (Diptera: Sciaridae). *Chromosoma* 119:613–623.
- Maizels N. 2006. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat Struct Mol Biol.* 13:1055–1059.
- Marshall OJ, Chueh AC, Wong LH, Choo KH. 2008. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. *Am J Hum Genet.* 82:261–282.
- Martin W, Koonin EV. 2006. Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440:41–45.
- Mason JM, Frydrychova RC, Biessmann H. 2008. *Drosophila* telomeres: an exception providing new insights. *Bioessays* 30:25–37.
- McClintock B. 1984. The significance of responses of the genome to challenge. *Science* 226:792–801.
- McEachern MJ, Blackburn EH. 1994. A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts. *Proc Natl Acad Sci U S A.* 91:3453–3457.
- Méndez-Lago M, et al. 2009. Novel sequencing strategy for repetitive DNA in a *Drosophila* BAC clone reveals that the centromeric region of the Y chromosome evolved from a telomere. *Nucleic Acids Res.* 37: 2264–2273.
- Mohr G, Ghanem E, Lambowitz AM. 2010. Mechanisms used for genomic proliferation by thermophilic group II introns. *PLoS Biol.* 8(6):e1000391.
- Moore JK, Haber JE. 1996. Capture of retrotransposon DNA at the sites of chromosomal double-strand breaks. *Nature* 383:644–646.
- Morrish TA, et al. 2002. DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet.* 2:159–165.
- Morrish TA, et al. 2007. Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres. *Nature* 446:208–212.
- Murphy WJ, et al. 2005. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* 309:613–617.
- Neidle S. 2009. The structures of quadruplex nucleic acids and their complexes. *Curr Opin Struct Biol.* 9:239–250.
- Nickles K, McEachern MJ. 2004. Characterization of *Kluyveromyces lactis* subtelomeric sequences including a distal element with strong purine/pyrimidine strand bias. *Yeast* 21:813–830.
- Nielsen L, Edstrom JE. 1993. Complex telomere-associated repeat units in members of the genus *Chironomus* evolve from sequences similar to simple telomeric repeats. *Mol Cell Biol.* 13:1583–1589.
- Olszak AM, et al. 2011. Heterochromatin boundaries are hotspots for *de novo* kinetochore formation. *Nat Cell Biol.* 13:799–808.
- Osanai M, Kojima KK, Futahashi R, Yaguchi S, Fujiwara H. 2006. Identification and characterization of the telomerase reverse transcriptase of *Bombyx mori* (silkworm) and *Tribolium castaneum* (flour beetle). *Gene* 376:281–289.
- Paeschke K, Simonsson T, Postberg J, Rhodes D, Lipps HJ. 2005. Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat Struct Mol Biol.* 12:847–854.
- Palm W, de Lange T. 2008. How shelterin protects mammalian telomeres. *Annu Rev Genet.* 42:301–334.
- Pardue ML, DeBaryshe PG. 2011. Retrotransposons that maintain chromosome ends. *Proc Natl Acad Sci. U S A.* 108:20317–20324.
- Prescott J, Blackburn EH. 1997. Telomerase RNA mutations in *Saccharomyces cerevisiae* alter telomerase action and reveal nonprocessivity *in vivo* and *in vitro*. *Genes Dev.* 11:528–540.
- Puertas MJ, Villasante A. 2013. Is the heterochromatin of meiotic neocentromeres a remnant of the early evolution of the primitive centromere?. In: Jiang J, Birchler JA, editors. *Plant centromere biology* Oxford (UK): Wiley-Blackwell. p. 95–109.
- Raffa GD, et al. 2009. The *Drosophila* modigliani (moi) gene encodes a HOAP-interacting protein required for telomere protection. *Proc Natl Acad Sci U S A.* 106:2271–2276.

- Raffa GD, et al. 2010. Verrocchio, a *Drosophila* OB fold-containing protein, is a component of the terminin telomere-capping complex. *Genes Dev.* 24:1596–1601.
- Rhodes D. 2006. The structural biology of telomeres. In: de Lange T, Lundblad V, Blackburn EH, editors. *Telomeres*, 2nd ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press. p. 317–343.
- Robertson HM, Gordon KH. 2006. Canonical TTAGG-repeat telomeres and telomerase in the honey bee, *Apis mellifera*. *Genome Res.* 16: 1345–1351.
- Schaffitzel C, et al. 2001. In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with *Styloynchia lemnae* macro-nuclei. *Proc Natl Acad Sci U S A.* 98:8572–8577.
- Schoeftner S, Blasco MA. 2008. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol.* 10:228–236.
- Sharp PA. 1985. On the origin of RNA splicing and introns. *Cell* 42: 397–400.
- Sharp PA. 1991. “Five easy pieces”. *Science* 254:663.
- Sheen FM, Levis RW. 1994. Transposition of the LINE-like retrotransposon TART to *Drosophila* chromosome termini. *Proc Natl Acad Sci U S A.* 91: 12510–12514.
- Shpiz S, et al. 2011. Mechanism of the piRNA-mediated silencing of *Drosophila* telomeric retrotransposons. *Nucleic Acids Res.* 39: 8703–8711.
- Starnes JH, Thornbury DW, Novikova OS, Rehmeyer CJ, Farman ML. 2012. Telomere-targeted retrotransposons in the rice blast fungus *Magnaporthe oryzae*: agents of telomere instability. *Genetics* 2: 389–406.
- Teixeira MT, Gilson E. 2005. Telomere maintenance, function and evolution: the yeast paradigm. *Chromosome Res.* 13:535–548.
- Tran PL, Mergny JL, Alberti P. 2011. Stability of telomeric G-quadruplexes. *Nucleic Acids Res.* 39:3282–3294.
- Ventura M, et al. 2003. Neocentromeres in 15q24–26 map to duplicons which flanked an ancestral centromere in 15q25. *Genome Res.* 13: 2059–2068.
- Ventura M, et al. 2004. Recurrent sites for new centromere seeding. *Genome Res.* 14:1696–1703.
- Villasante A, Abad JP, Méndez-Lago M. 2007. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. *Proc Natl Acad Sci U S A.* 104:10542–10547.
- Villasante A, de Pablos B, Mendez-Lago M, Abad JP. 2008. Telomere maintenance in *Drosophila*: rapid transposon evolution at chromosome ends. *Cell Cycle* 7:2134–2138.
- Villasante A, et al. 2007. *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res.* 17:1909–1918.
- Villasante A, Méndez-Lago M., Abad JP, Montejo de Garcini E. 2007. The birth of the centromere. *Cell Cycle* 6:2872–2876.
- Wiegmann BM, et al. 2011. Episodic radiations in the fly tree of life. *Proc Natl Acad Sci U S A.* 108:5690–5695.
- Xiong Y, Eickbush T. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* 9: 3353–3362.
- Xu L, Feng S, Zhou X. 2011. Human telomeric G-quadruplexes undergo dynamic conversion in a molecular crowding environment. *Chem Commun (Camb).* 47:3517–3519.
- Xu Y, Sato H, Sannohe Y, Shinohara K, Sugiyama H. 2008. Stable lariat formation based on a G-quadruplex scaffold. *J Am Chem Soc.* 130: 6470–6471.
- Yamamoto Y, et al. 2003. Retrotransposon-mediated restoration of *Chlorella* telomeres: accumulation of Zepp retrotransposons at termini of newly formed minichromosomes. *Nucleic Acids Res.* 31: 4646–4653.
- Yona AH, et al. 2012. Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci U S A.* 109: 21010–21015.
- Zeitlin SG, et al. 2009. Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci U S A.* 106:15762–15767.
- Zhang AY, Balasubramanian S. 2012. The kinetics and folding pathways of intramolecular g-quadruplex nucleic acids. *J Am Chem Soc.* 134: 19297–19308.
- Zimmerly S, Guo H, Perlman PS, Lambowitz AM. 1995. Group II intron mobility occurs by target DNA-primed reverse transcription. *Cell* 82: 545–554.