

# Evolutionary rates of mammalian telomere-stability genes correlate with karyotype features and female germline expression

Chiara Pontremoli<sup>1,\*</sup>, Diego Forni<sup>1</sup>, Rachele Cagliani<sup>1</sup>, Uberto Pozzoli<sup>1</sup>, Mario Clerici<sup>2,3</sup> and Manuela Sironi<sup>1</sup>

<sup>1</sup>Bioinformatics, Scientific Institute IRCCS E. MEDEA, 23842 Bosisio Parini, Italy, <sup>2</sup>Department of Physiopathology and Transplantation, University of Milan, 20090 Milan, Italy and <sup>3</sup>Don C. Gnocchi Foundation ONLUS, IRCCS, 20148 Milan, Italy

Received January 30, 2018; Revised May 17, 2018; Editorial Decision May 18, 2018; Accepted May 22, 2018

## ABSTRACT

**Telomeres protect the ends of eukaryotic chromosomes and are essential for cell viability. In mammals, telomere dynamics vary with life history traits (e.g. body mass and longevity), suggesting differential selection depending on physiological characteristics. Telomeres, in analogy to centromeric regions, also represent candidate meiotic drivers and subtelomeric DNA evolves rapidly. We analyzed the evolutionary history of mammalian genes implicated in telomere homeostasis (TEL genes). We detected widespread positive selection and we tested two alternative hypotheses: (i) fast evolution is driven by changes in life history traits; (ii) a conflict with selfish DNA elements at the female meiosis represents the underlying selective pressure. By accounting for the phylogenetic relationships among mammalian species, we show that life history traits do not contribute to shape diversity of TEL genes. Conversely, the evolutionary rate of TEL genes correlates with expression levels during meiosis and episodes of positive selection across mammalian species are associated with karyotype features (number of chromosome arms). We thus propose a telomere drive hypothesis, whereby (sub)telomeres and telomere-binding proteins are engaged in an intra-genomic conflict similar to the one described for centromeres.**

## INTRODUCTION

Telomeres, the protein/DNA complexes that preserve the ends of eukaryotic chromosomes, are essential for genomic stability and cell viability. Telomeres have two main functions: to protect the chromosome terminus from nuclease and DNA repair activities and to provide a mechanism of

compensation for the inability of global genome DNA polymerases to replicate the 5' end of a linear chromosome (1).

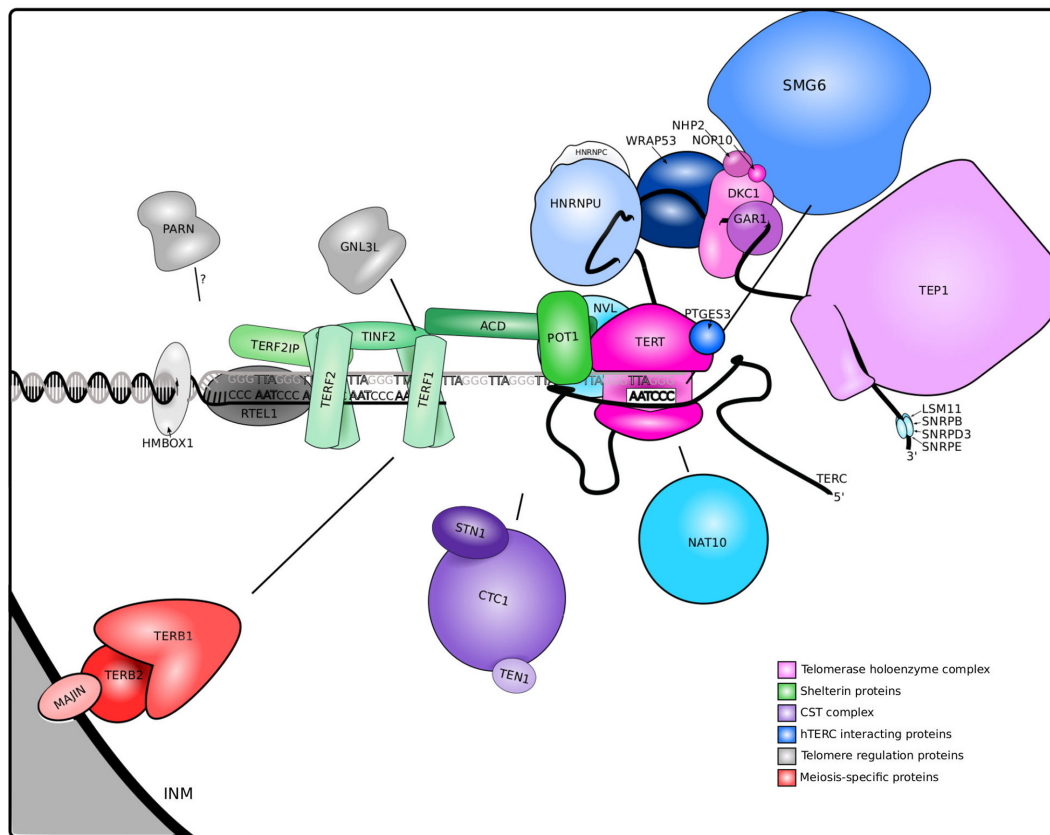
Mammalian telomeres consist of hundreds to thousands of tandem repeats of the sequence 5'-TTAGGG-3'. The majority of these repeats are double-stranded, but the very end of each chromosome has a single-stranded G-rich 3' overhang, which is important for telomerase-driven telomere extension (2). Mammalian telomerase is a ribonucleoprotein, with a catalytic core formed by the TERT protein component and by the TERC RNA component. In the holoenzyme, telomerase associates with several proteins, which contribute to the regulation, maturation, assembly, and localization of the telomerase complex (Figure 1).

Telomere function is also critically dependent on a complex of telomere-associated proteins known as shelterin, whose major function is to protect the chromosome ends (Figure 1).

Telomere homeostasis is essential for cell viability, as telomere dysfunction leads to senescence, apoptosis and malignant transformation. In fact, mutations in genes encoding proteins required for telomere elongation, repair, and maintenance are associated with human genetic disorders such as diskeratosis congenita, one of the primary complex telomere biology disorders (TBDs) (3). Notably, genome instability is often a hallmark of TBDs (3).

In most human somatic cells, telomerase activity is undetectable and telomere length is progressively shortened during cell replication (4). As a result, somatic cells undergo replicative senescence, which plays an important role in suppressing tumorigenesis (5). Telomerase repression in somatic tissues is not an universal phenomenon among mammals, and several small-sized species, which are characterized by a short lifespan, express telomerase in somatic cells (6). In a study performed on >60 mammalian species, telomerase expression was found to co-evolve with body mass, whereas telomere length inversely correlated with lifespan (which is, in turn, strongly correlated with mass)

\*To whom correspondence should be addressed. Tel: +39 031877826; Fax: +39 031877499; Email: chiara.pontremoli@bp.lnf.it



**Figure 1.** Mammalian telomeric complex. The TEL proteins we analyzed are shown to provide a general overview of their function. Gene products are colour-coded. INM = inner nuclear membrane.

(6). A commonly accepted explanation for these observations is that replicative senescence evolved to mitigate the increased cancer risk conferred by a large number of cells (large body mass) and long lifespans (5). Because lifespan is related to several life-history traits such as extrinsic mortality risk and reproductive strategies (5,7), telomere dynamics were suggested to be subject to natural selection depending on the ecological characteristics of different mammalian species (7). This clearly raises the possibility that genes involved in telomere elongation, maintenance, and repair evolved at different rates in distinct mammals and that natural selection shaped their diversity across species.

Recent evidence indicated that the evolution of genes involved in the maintenance of telomere integrity was dominated by positive selection in *Drosophila* (8). Because *Drosophila* telomeres are not composed of 5'-TTAGGG-3' repeats but, rather, they are generated by insertion of specific retrotransposons at the chromosome termini, the rapid evolution of telomere-integrity proteins may be driven by the need to avoid retrotransposon over-proliferation (8). Alternatively, an intra-genomic conflict with telomeric selfish elements that distort meiotic segregation in females may explain the fast evolution of *Drosophila* proteins involved in telomere homeostasis (8). The latter possibility is reminiscent of the proposed evolutionary scenario for kinetochore proteins, which are engaged in a conflict with selfish centromeric DNA (centromere drive hypothesis) (9–12).

Telomeres play an essential role in meiosis: at the early stages of meiosis I all the telomeres in the cell attach to the internal nuclear membrane (INM) and move to the centrosome in a 'bouquet' configuration. Bouquet formation is important to ensure homolog pairing and recombination (13). In mammals, three specialized telomere-associated proteins specifically expressed at meiosis (TERB1, TERB2 and MAJIN) determine attachment to the INM (14) (Figure 1).

Although mammalian telomeres are invariably composed of 5'-TTAGGG-3' repeats, subtelomeric DNA evolves rapidly (15), suggesting that proteins involved in telomere maintenance and elongation may take part in some form of intra-genomic conflict.

Herein, we analyzed the evolutionary history of mammalian genes implicated in telomere protection, elongation and maintenance (TEL genes). We found pervasive positive selection at TEL genes and we used different approaches to determine the most likely underlying selective pressures.

## MATERIALS AND METHODS

### Gene selection

TEL genes were selected on the basis of the following GO terms (as of 1 June 2016): 'Telomerase holoenzyme complex' (GO:0005697) and 'Telomere cap complex' (GO:0000782) filtering with 'direct assay' as evidence and 'Homo

*sapiens*' as organism. We added *RTEL1* and *PARN* as these genes were associated to telomere dysfunction and human genetic diseases (3) (Figure 1). Mammalian homologs of human TEL genes were included only if they represented one-to-one orthologs, as reported in the EnsemblCompara GeneTrees (16). The majority of *HNRNPC* mammalian orthologs represented one-to-many or many-to-many orthologs: for this reason *HNRNPC* was not considered. For the same reason, sequences from rodents and lagomorpha were not included for *POT1*, as these animals have two *Pot1* paralogues (*Pot1a* and *Pot1b*) (17,18). The final list included 32 TEL genes (Supplementary Table S1). To obtain a list of 'control' (CTR) genes, for each TEL gene we randomly selected one human gene with a GC content and coding sequence length differing less than 5% from those of the human TEL gene. Mammalian homologs of CTR genes were included only if they represented one-to-one orthologs (Supplementary Table S1).

As a comparison, genes involved in centromere function (CEN genes) were selected from a recent review (19). In particular, we included *CENPA* and genes coding for proteins involved in the correct localization and stabilization of *CENPA* at centromeric nucleosomes.

For all genes, coding sequence information for at least 53 mammalian species (Supplementary Tables S1 and S2) were retrieved from the NCBI database (<http://www.ncbi.nlm.nih.gov/>), from the Ensembl website (<http://www.ensembl.org/index.html>), and from the UCSC server (<http://genome.ucsc.edu/>).

### Evolutionary analysis in mammals

Coding sequence alignments were generated using the RevTrans 2.0 utility (<http://www.cbs.dtu.dk/services/RevTrans/>), MAFFT v6.240 as an aligner (20), a software which uses the protein sequence alignment as a scaffold for constructing the corresponding DNA multiple alignment. This latter was checked and edited using TrimAl (automated1 mode) (21); manual editing was used to correct a few misalignments in proximity of small gaps (<http://phylemon.bioinfo.cipf.es/utilities.html>).

To evaluate the level of substitution saturation, we applied the Xia's index implemented in DAMBE (22) to all gene alignments. No statistical evidence of substitution saturation was detected for any alignment (Supplementary Table S3).

All alignments were also screened for the presence of recombination using GARD (Genetic Algorithm Recombination Detection) (23). Evidence of recombination was detected only for *WRAP53*, *HJURP* and *EHD4*, whereas no breakpoints were detected for all the remaining genes. The coding alignments of *WRAP53*, *HJURP* and *EHD4* were split on the basis of the recombination breakpoints and sub-regions were used as the input for molecular evolution analyses.

The average nonsynonymous substitution (dN)/synonymous substitution (dS) rate and the dN–dS parameter were calculated using the single-likelihood ancestor counting (SLAC) method (24).

Gene trees were generated using the phyML program with gamma-distributed rates, four substitution rate categories, and estimation of transition/transversion ratio and proportion of invariable sites (25).

Evidence of positive selection was searched for using the codon-based codeml program implemented in the PAML (Phylogenetic Analysis by Maximum Likelihood) suite (26). We applied different site (NSsite) models. In particular, M7 is a null model that assumes that  $0 < dN/dS < 1$  and is beta distributed among sites; M8 is a positive selection model: it is the same as M7 but also includes a category of sites with  $dN/dS > 1$ ; M8a is the same as M8, except that it does not allow positive selection, but only neutral evolution. To assess statistical significance, twice the difference of the likelihood ( $\Delta \ln L$ ) for the models (M7 vs M8 and M8a vs M8) is compared to a  $\chi^2$  distribution (2 degrees of freedom for the M7 versus M8 comparison, 1 degree of freedom for M8a versus M8). LRT tests were run with both F3  $\times$  4 and F61 codon frequency models, with an initial value of  $\omega = 0.4$  (Supplementary Tables S4–S6). To further investigate whether substitution saturation was responsible for the high fraction of TEL and CEN positively selected genes, we used the PAML Free Ratio (FR) model to estimate dS for all branches of the gene/region phylogenies (27). A minority of branches had  $dS > 0.25$  and really few had  $dS > 0.5$  (Supplementary Tables S4 and S5). Only two gene phylogenies (*TERT* and *DKC1*) had branches with  $dS > 1$ . Analyses were thus re-run after removing 1 branch both for *DKC1* and for *TERT*. Very similar results to those with the full dataset were obtained (Supplementary Table S4). For positively selected genes, results were validated using a different initial value for  $\omega$  ( $\omega = 1$ ) (not shown).

Positively selected sites were identified using the BEB (Bayes Empirical Bayes) method (with a posterior probability  $\geq 0.90$ ) (28), the Random effects likelihood (REL) (24) and the Fast Unconstrained Bayesian AppRoximation (FUBAR) (29) methods from the HyPhy package (30). Sites were declared as positively selected by REL or FUBAR if they showed a Bayes Factor  $\geq 50$  or a posterior probability  $\geq 0.90$ , respectively. Finally, only sites identified by at least two methods were considered as positively selected.

The PAML Free Ratio (FR) model was also used to estimate the value of dN/dS on the branches of the phylogenies of all gene sets (27). The FR model assumes different dN/dS for each lineage and can be compared with a null model with one dN/dS for the entire phylogeny. Statistical significance is assessed by comparing twice the  $\Delta \ln L$  of the two models with a  $\chi^2$  distribution with degrees of freedom equal to the difference in model parameters.

The branch model Clade model C (CmC) (31) was used to identify codons under divergent selection pressure among different clades in the phylogeny. In the CmC model, three classes of sites are assumed in the alignment: the first two represent evolutionary conserved and neutral evolving codons, while the third class represents codons under different selective pressures in the different clades selected *a priori*. This model is then compared against a null model proposed by Weadick and Chang that does not allow variation among lineages (32).

### TEL genes mutations

The list of TEL genes mutations was obtained from the Human Gene Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/>, last accessed December 2017) and from the Online Mendelian Inheritance in Man (OMIM, <https://www.omim.org/>, last accessed December 2017) database.

### Phylogenetic regression analysis

Life-history traits were retrieved for all mammalian species of the AnAge database ( $n = 933$ ). In particular, we focused on information about adult weight as a proxy of body mass and maximum recorded lifespan as a proxy for longevity. We also retrieved information about male and female sexual maturity (expressed in days) and interbirth intervals (called inter litter). A mammal supertree was retrieved from Bininda-Emonds *et al.* (33), and then pruned based on the AnAge species.

We performed a continuous regression analysis within a maximum-likelihood PGLS (Phylogenetic Generalized Least Squares) framework with the BayesTrait software (34) using  $\log_{10}$  transformed body mass and longevity values. BayesTrait incorporates the phylogenetic signal into the regression model by taking into account the shared ancestry specified by the phylogenetic tree. As expected, a positive and significant ( $P < 0.001$ ) correlation between body mass and longevity was obtained (6).

The regression coefficient and the intercept were used to plot the regression line, and the life-history traits (LHT) for the species analyzed in the positive selection analyses were then plotted as points. We assigned these species to a specific category if they fell inside, over, or under the plot region defined by the regression line plus its standard errors. These three categories were then used to create the groups for the CmC test.

We applied the Coevol software (35) to evaluate the correlation of molecular variables with life history traits. This software models the correlated evolution of these traits by assuming a multivariate Brownian diffusion process. All parameters in the model are estimated in a Bayesian framework and the relationship between variables is returned as a posterior probability of covariation. We run Coevol with all LHTs simultaneously along with dS and dN/dS, and we analyzed the covariation among these parameters after controlling for the effect of body mass and longevity. All life-history trait values were  $\log_{10}$ -transformed before running the analyses.

### Correlation with meiotic gene expression

Gene expression changes (fold-change) during female and male mouse meiosis were retrieved from previous works (36,37). The correlation between dN/dS and fold-changes was evaluated using Kendall's correlation. Kendall's correlation is a non-parametric test based on ranks. Ties were not present in any variable (either dN/dS or fold changes) for any dataset (TEL and CTR genes). Under these conditions, the power of Kendall's correlation depends uniquely on the sample size. Thus, irrespective of the expression level and range of fold changes we had equal power to detect correlation for the TEL and CTR datasets.

### Comparison of the selective pattern of CEN and TEL genes

To analyze the pattern of selection at CEN and TEL genes and to compare it to that of CTR genes, we calculated pairwise dN/dS correlations for a set of internal branches and for all tip branches across the phylogenies. In particular, Kendall's correlation coefficient (tau) was calculated for all possible gene pairs. Next, we evaluated tau distribution among different gene sets: for instance, the TEL-TEL distribution was obtained by calculating tau values for all possible pairwise combinations of TEL genes, whereas the TEL+CEN - CTR distribution was obtained by calculating tau for all possible combinations of a TEL or CEN gene with a CTR gene. Distributions were compared by applying Kolmogorov-Smirnov tests.

### Correlation with karyotype features

Information on karyotype features was retrieved from literature sources (38–43) and was available for 60 of the mammalian species that also had molecular evolutionary data. For these taxa, the median dN/dS for TEL, CEN and CTR genes with significant FR test was calculated and correlated with different karyotype features (i.e. aFN, diploid chromosome count, percentage of acrocentric/telocentric chromosomes, percentage of metacentric chromosomes). To assess the role of phylogenetic inertia in driving the correlation between dN/dS and aFN, we generated a concatenated alignment of TEL genes (with significant FR test) for the 60 species with available karyotype information. The alignment and aFN data were used as the input for Coevol, with body mass and longevity entered as covariates.

### Evolutionary analysis in ciliates and yeasts

*Tetrahymena* genes that encode proteins involved in telomere elongation or maintenance were identified by literature searches (44–50). *Tetrahymena thermophila* coding sequences were derived from the NCBI database (<http://www.ncbi.nlm.nih.gov/>) or from the *Tetrahymena* genome database (<http://ciliate.org/index.php/home/>) (51). Orthologs in the *Tetrahymena malaccensis*, *Tetrahymena eliotti* and *Tetrahymena borealis* genomes were identified via the BLAST utility at the *Tetrahymena* genome database website. Genes with no orthologs in one or more *Tetrahymena* species were discarded. For 11 TEL genes, one single ortholog with high similarity to the *T. thermophila* sequence was identified in the other species. To obtain an equal number of control genes, we randomly selected one *T. thermophila* gene with the same coding sequence length and having orthologs in the three other *Tetrahymena* species. Coding sequences were aligned as described for mammalian sequences. Positive selection was tested by applying two methods, M7/M8 contrast, as described above (the more conservative M8a/M8 test was not performed as its power is expected to be very limited when four sequences only are included) and BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) (52). BUSTED allows  $\omega$  to vary from branch to branch and it tests whether a gene has experienced positive selection at at least one site on at least one branch. Genes that showed significant evidences of positive selection (with either method) were analyzed for

the presence of recombination using GARD. No evidence of recombination was detected.

Yeast genes involved in telomere homeostasis were retrieved through Gene Ontology annotations for *Saccharomyces cerevisiae* (GO:0005697 and GO:0000782). Yeast genes that encode meiosis-specific telomere-associated proteins and proteins involved in centromere specification were identified via literature searches (19,53). Evidence of positive selection (M7/M8 contrast) was derived from a previous genome-wide analysis (54).

## RESULTS

### TEL genes evolve at different rates depending on function

We analyzed the evolutionary history of 32 TEL genes and a matched list of CTR genes (see Methods for gene selection procedures) (Figure 1 and Supplementary Table S1) in a large phylogeny of placental mammals. Because they represent an epitome of fast-evolving essential genes that encode repetitive DNA-binding proteins, ten CEN genes were also included (Supplementary Table S1).

We first calculated the average non-synonymous substitution/synonymous substitution rate (dN/dS, also referred to as  $\omega$ ). Comparison of CTR and TEL genes indicated that the latter have significantly higher average dN/dS (Wilcoxon Rank Sum Test for paired samples, two tailed,  $P = 0.025$ ). In turn, average dN/dS was significantly higher for CEN genes than for TEL genes (Wilcoxon Rank Sum Test, two tailed,  $P = 0.037$ ) (Supplementary Table S1).

To gain further insight into the evolutionary pattern of TEL genes, a codon-wise measure of natural selection was obtained by calculating the dN–dS parameter (24). This metric was preferred over the conventional dN/dS because it is not rendered to infinite for dS values equal to 0. TEL genes were divided into six groups, based on their function (Figures 1, 2A and Supplementary Table S1). The distribution of dN–dS values was significantly different across gene groups (Kruskal–Wallis rank sum test,  $P < 10^{-15}$ ). CEN genes and meiosis-specific TEL genes had significantly higher dN–dS compared to CTR genes (Nemenyi post-hoc test,  $P < 10^{-15}$  for both comparisons) (Figure 2A). The same applied to genes encoding shelterin proteins (Nemenyi post-hoc test,  $P = 8.6 \times 10^{-10}$ ), although fewer codons had dN–dS  $> 0$  (Figure 2A). Conversely, genes belonging to other functional categories displayed a definitely higher proportions of constrained sites (dN–dS  $< 0$ ) and the median dN–dS was either significantly lower than that of CTR genes (CTS complex, hTERC interaction, telomere regulation, Nemenyi post-hoc test,  $P < 10^{-8}$  for all tests) or not significantly different (Telomerase holoenzyme complex, Nemenyi post-hoc test,  $P = 0.92$ ). Thus, TEL genes evolve at different rates, depending on their function, and they are generally more constrained than CEN loci, partially explaining their lower average dN/dS (Figure 2A).

Clearly, a major selective constraint acting on protein coding genes is related to the disease state or loss of fitness that amino acid replacements can determine by altering protein function. Among TEL genes, 13 have been associated with human TBDs. Analysis of dN–dS indicated that disease genes are generally more constrained than the

other TEL genes (Wilcoxon Rank Sum Test, two-tailed,  $P < 10^{-16}$ ) and, within these genes, sites where causative mis-sense mutations were identified ( $n = 151$ ) have lower dN–dS than sites where mutation have never been described (Wilcoxon Rank Sum test, two tailed,  $P = 0.00051$ ) (Figure 2B).

An important contribution of negative selection is expected for genes that play essential cellular functions. To formally test whether positive selection also drove the evolution of TEL genes in placental mammals, we applied likelihood ratio tests (LRT) implemented in the PAML suite (26,55). Null site models (M7 and M8a) were rejected in favor of the positive selection model (M8) for 18 genes. As a comparison, evidence of positive selection was detected for 9 CEN genes and for 8 CTR genes out of 32 (Supplementary Tables S4–S6). The proportion of CTR genes showing evidence of selection is consistent with a previous genome-wide analysis of 29 mammals that, using less stringent criteria than those we applied, reported  $\sim 34\%$  of genes with at least one positively selected codon (56).

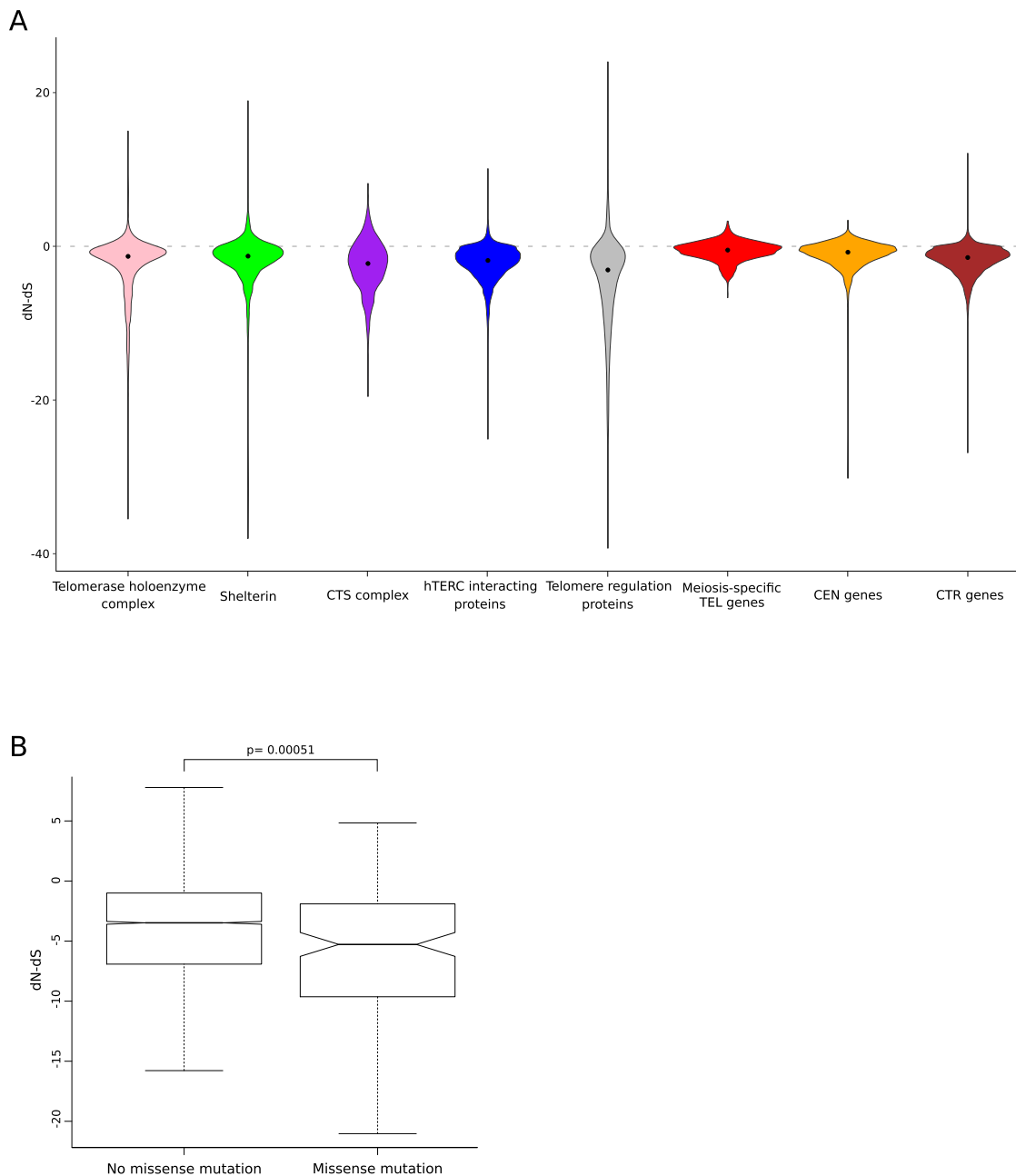
These results indicate that the proportion of TEL genes targeted by positive selection is significantly higher than that of CTR genes (Fisher's Exact Test,  $P = 0.021$ ).

In positively selected genes, the fraction of sites that were called as positively selected (see methods) was 1.00% for TEL, 1.47% for CEN and 0.57% for CTR (Figure 3, Supplementary Figure S1, Supplementary Tables S4 and S5). Thus, both TEL and CEN genes display a significantly higher fraction of positively selected sites compared to CTR genes (TEL: Fisher's Exact Test,  $P = 0.0025$ ; CEN: Fisher's Exact Test,  $P = 9 \times 10^{-7}$ ).

Analysis of selected sites in TEL proteins indicated that they often fall within DNA- and RNA-interacting domains. This is the case of sites in TERF1, TERF2 and SMG6 (Figure 3 and Supplementary Figure S1). In CTC1, STN1 and TEP1, several selected sites map within the OB-fold domains or within the TROVE domain (Supplementary Figure S1), which are responsible for the binding to ssDNA and RNA, respectively. Sites in telomeric DNA-binding domains were also detected in TERB1 (MYB domain) and MAJIN (Figure 3).

### Life history traits do not explain positive selection at TEL genes

As mentioned above, telomerase activity and telomere length correlate with body mass and longevity in mammals (6). We thus tested whether the signatures of positive selection at TEL genes were driven by changes in these life history traits. To this aim, we applied a Phylogenetic Generalized Least Squares (PGLS) framework, which accounts for the shared ancestry among species in a phylogeny, to analyze mass and longevity (6). We retrieved body mass and longevity data for all the mammalian species present in the AnAge database (<http://genomics.senescence.info/species/>) and we calculated the regression line by incorporating the phylogenetic signal (Figure 4). We then defined whether each of the species analyzed for TEL genes deviated from this correlation (plus confidence intervals): we thus generated three categories of species based on the mass~lifespan



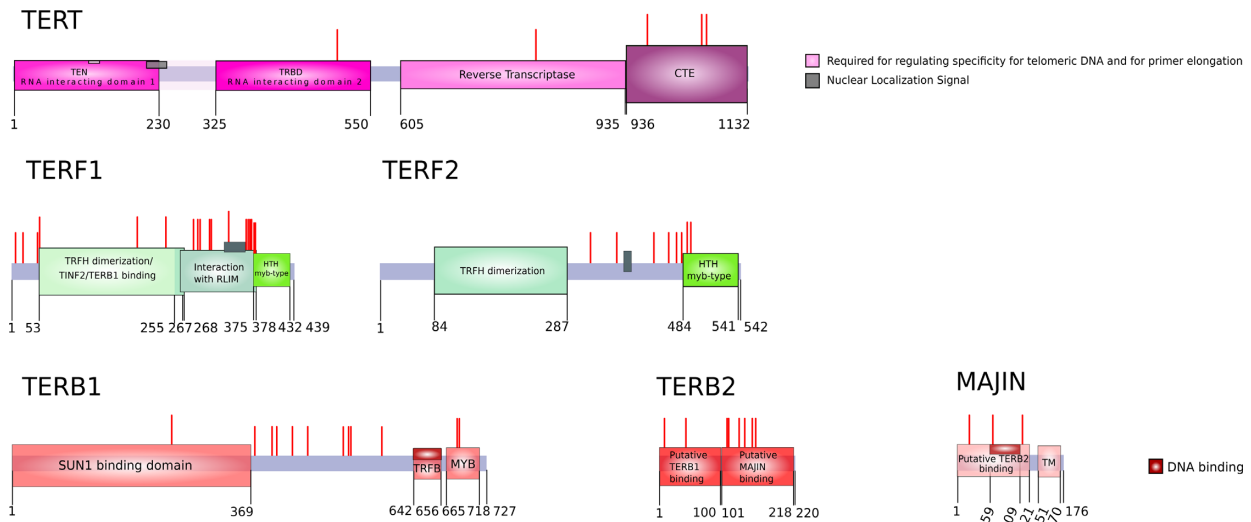
**Figure 2.** Distribution of dN-dS values. (A) Violin plots of dN-dS values calculated for all TEL, CEN and CTR genes. TEL genes are grouped based on their function, as in Figure 1. The dotted line represents a value dN-dS = 0 (neutral evolution). (B) Boxplot representation of dN-dS: TEL disease gene codons carrying missense mutations responsible for TBDs are compared with codons where no missense mutation has been described.

expectation (increased, unchanged, or decreased longevity based on body mass) (Figure 4).

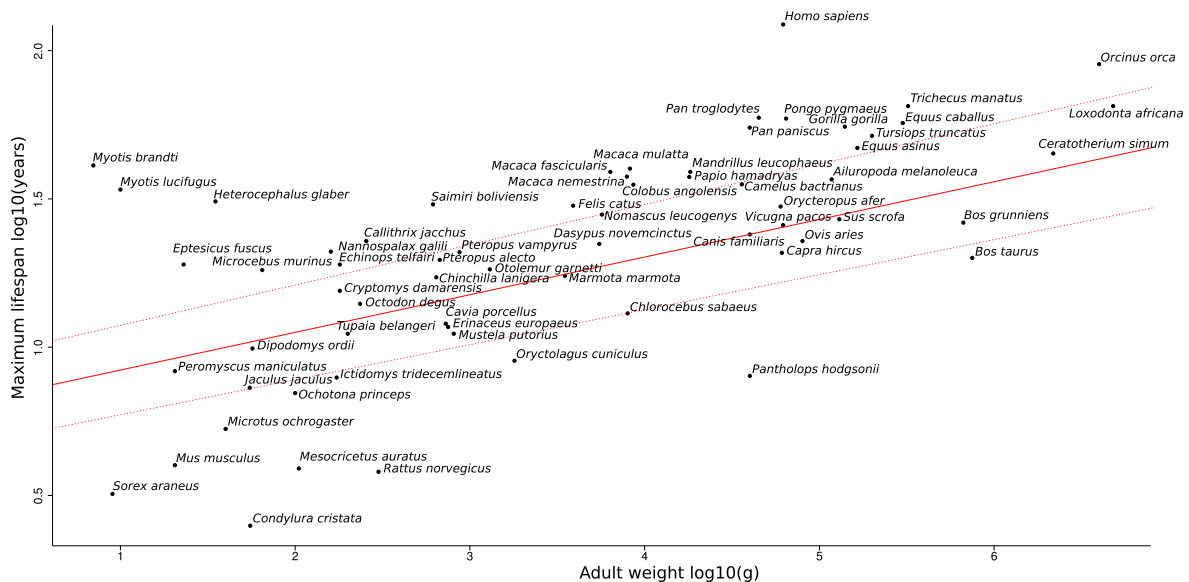
We next applied the *codeml* clade models, which were developed to test for variation in selection pressure among ecologically divergent species (32). Specifically, these models allow site-specific divergence in selective constraint among clades of a phylogeny. We analyzed the different selective patterns in all the TEL genes found under positive selection ( $n = 18$ ) in the three categories defined by the PGLS method. Results showed that, although there were some significant differences in dN/dS among groups, no group had

a class of codons evolving with dN/dS higher than 1 for any gene (Supplementary Table S7).

We next analyzed additional life history traits, namely age at female maturity, age at male maturity, and inter litter interval. Analyses were performed using Coevol, a Bayesian method that corrects for phylogenetic inertia (35). Body mass and longevity were used as covariates. Out of eighteen analyzed genes, only *ACD* showed a significant correlation of dN/dS with age at female and male maturity (Supplementary Table S8).



**Figure 3.** Localization of positively selected sites. Domain representation of a subset of positively selected TEL proteins; positively selected sites are indicated in red. Colour-codes are as in Figure 1.



**Figure 4.** Correlation plot between body mass and longevity. A regression line (plus confidence intervals) between maximum lifespan and adult weight (log10-transformed) for 933 mammalian species is plotted in red. Species analyzed in this work are shown with black dots.

Overall, these data suggest that positive selection at TEL genes is not mainly related to changes in life-history traits.

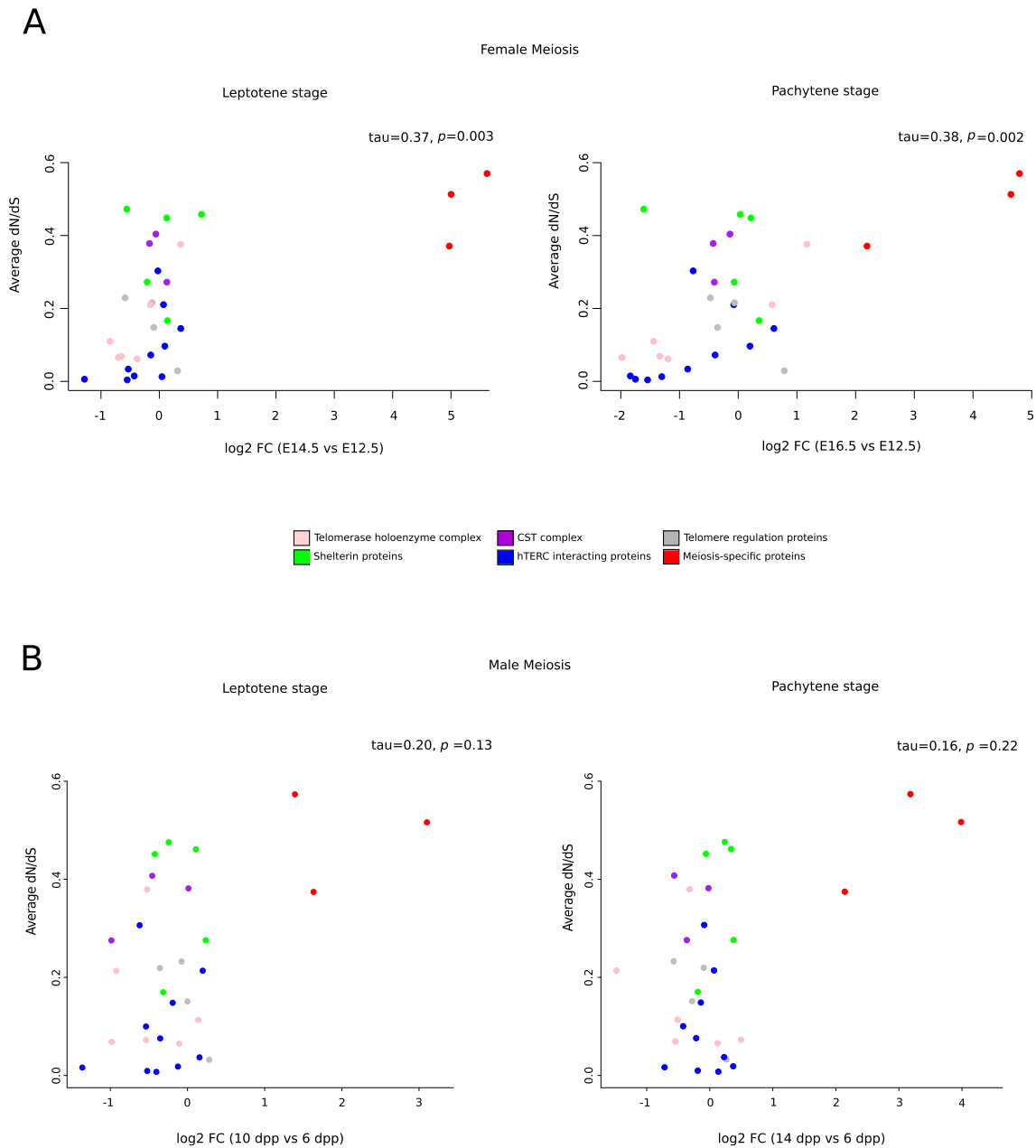
### Evolutionary rates of TEL genes correlate with expression during female meiosis

Because meiosis-specific TEL genes displayed high average dN–dS values and were found to be positively selected, we investigated the relationship between evolutionary rates and TEL gene meiotic expression. Indeed, besides *TERB1*, *TERB2* and *MAJIN*, several other TEL genes are expressed during gametogenesis.

We thus used genome-wide RNA-seq data for fetal mouse ovaries to retrieve information on the expression of TEL and CTR genes before and during meiosis (36). Specifically,

we obtained expression level changes (fold-change) for the leptotene (E14.5) and pachytene (E16.5) stages compared to a pre-meiotic (E12.5) stage. These values were correlated to average dN/dS. For TEL genes, a positive correlation was obtained for both the leptotene and the pachytene stages (Figure 5A). Positive and significant (or almost significant) correlations were also observed when *TERB1*, *TERB2* and *MAJIN* were excluded from the analysis (leptotene stage, Kendall’s tau = 0.25,  $P = 0.06$ ; pachytene stage, Kendall’s tau = 0.27,  $P = 0.045$ ). No significant correlation was observed between dN/dS and increased meiotic expression for CTR genes (leptotene stage, Kendall’s tau = 0.076,  $P = 0.55$ ; pachytene stage, Kendall’s tau = 0.12,  $P = 0.34$ ).

To assess whether the same pattern was observed at the male meiosis, we retrieved expression changes of TEL and



**Figure 5.** Evolutionary rates and expression in meiosis. Average dN/dS for all TEL genes is plotted against the log2 fold-change (FC) of gene expression in the leptotene or pachytene stages versus the pre-meiotic stage of mouse oogenesis (**A**) or spermatogenesis (**B**). Dots are colored based on the gene function (see Figure 1). Kendall's correlation coefficients are also reported.

CTR genes during different stages of mouse male meiosis compared to pre-meiotic stages (6 days post partum, dpp). In particular, time periods that roughly correspond to the leptotene/zygotene stage (10 dpp) and pachytene stage (14 dpp) were analyzed (37). No correlation between average dN/dS and expression changes was observed for either TEL (Figure 5B) or CTR genes (leptotene/zygotene stage, Kendall's  $\tau = -0.018$ ,  $P = 0.89$ ; pachytene stage, Kendall's  $\tau = 0.032$ ,  $P = 0.81$ ).

Overall, these data indicate that TEL genes that are up-regulated in female meiosis evolve faster than TEL genes which show no or limited expression changes. Conversely,

TEL genes that are highly expressed during male gametogenesis have similar evolutionary rates as those showing low expression.

### Similar selective patterns for CEN and TEL genes across the mammalian phylogeny

Because a high proportion of both TEL and CEN genes showed fast evolutionary rates, we investigated their selection pattern across the whole phylogeny of placental mammals. To this aim, we applied the free ratio (FR) model implemented in the PAML software (27). This model estimates a value of dN/dS for each lineage in the phylogeny



and is compared with a null model that estimates a single dN/dS for all lineages. Results indicated that the FR model fitted the data better than the null model for 17 TEL and for 9 CEN genes (Supplementary Table S9). Thus, for these genes, the selective pressure has been acting differently across the phylogeny, with some branches showing dN/dS values higher than 1 (Supplementary Table S9). Twenty CTR genes also had a significant FR test (Supplementary Table S9).

To formally test whether the 17 TEL and 9 CEN genes showed a similar selective pattern and whether such pattern is different from that of CTR genes, we calculated pairwise dN/dS correlations for tip branches and a sub-set of internal branches for gene pairs (Figure 6A). In particular, Kendall's correlation coefficients (tau) were calculated. Results indicated no difference among the distribution of tau for TEL-TEL and CEN-CEN gene pairs (Kolmogorov-Smirnov test,  $P = 0.84$ ). These distributions were also not different from that obtained for TEL-CEN gene pairs (Kolmogorov-Smirnov test,  $P > 0.60$  for TEL-TEL versus CEN-TEL and for CEN-CEN versus CEN-TEL). CTR-CTR gene pairs also had a similar distribution. This is not surprising, though, as features such as body mass and longevity influence dN/dS across mammalian species (57). However, the distribution obtained by considering TEL and CEN genes as a single group (TEL+CEN in Figure 6B) was significantly different from that calculated for TEL+CEN - CTR gene pairs (Kolmogorov-Smirnov test,  $P = 0.009$ ) (Figure 6B), suggesting that TEL and CEN genes show similar selection patterns and that such patterns are different from those of CTR genes.

We next overlaid the signals of selection (proportion of genes showing dN/dS > 1) over the phylogenetic tree to obtain a glimpse of whether positive selection acted on specific lineages. Most of the branches leading to superorders/orders showed at least one gene with dN/dS > 1, both for TEL and for CEN genes (Figure 6A). In particular, the Xenarthra/Afrotheria and the Boreoeutheria branches showed a relatively high number of CEN and TEL selected genes. The same observation applies to the branches leading to Ferungulata and primates. As for tip branches, selection appeared strong in equids and macaques. In general, weak selection signals were detected in rodents, especially for TEL genes (Figure 6A).

### The strength of selection at TEL genes correlates with karyotype features

Mammalian karyotypes evolve rapidly, with the number of chromosomes/chromosome arms (and therefore of telomeres and centromeres) changing widely among species (58). We thus reasoned that karyotype features may correlate with the rate of evolution at TEL and/or CEN genes.

We thus compared the karyotype fundamental number for autosomes (aFN, a measure of the number of chromosome arms) in different mammalian species with the strength of selective pressure. In particular, for all taxa shown in Figure 6A, we calculated the median dN/dS across all genes showing a significant free ratio model. For those same taxa we retrieved aFN information from literature sources (60 species had available information) (38–

43). We found a positive significant correlation (Kendall's tau = 0.23,  $P = 0.012$ ) between aFN and dN/dS for TEL genes (Figure 6C). No significant correlation was found when CEN and CTR genes were analyzed (CEN: Kendall's tau = 0.07,  $P = 0.40$ ; CTR: Kendall's tau = -0.081,  $P = 0.37$ ). Also, no significant correlation for either TEL, CEN or CTR genes was detected with the diploid number of chromosomes, the percentage of acrocentrics/telocentrics, or the percentage of metacentrics (not shown). Although the correlation between aFN and median dN/dS does not seem to be due to clade/order-specific effects (Figure 6C), we used Coevol on a concatenated alignment of TEL genes to correct for phylogenetic inertia (35). A correlation coefficient of 0.195 was obtained with good statistical support (posterior probability = 0.83).

Thus, the overall rate of evolution is faster for TEL genes in species showing a high number of chromosome arms (i.e. a higher number of telomeres).

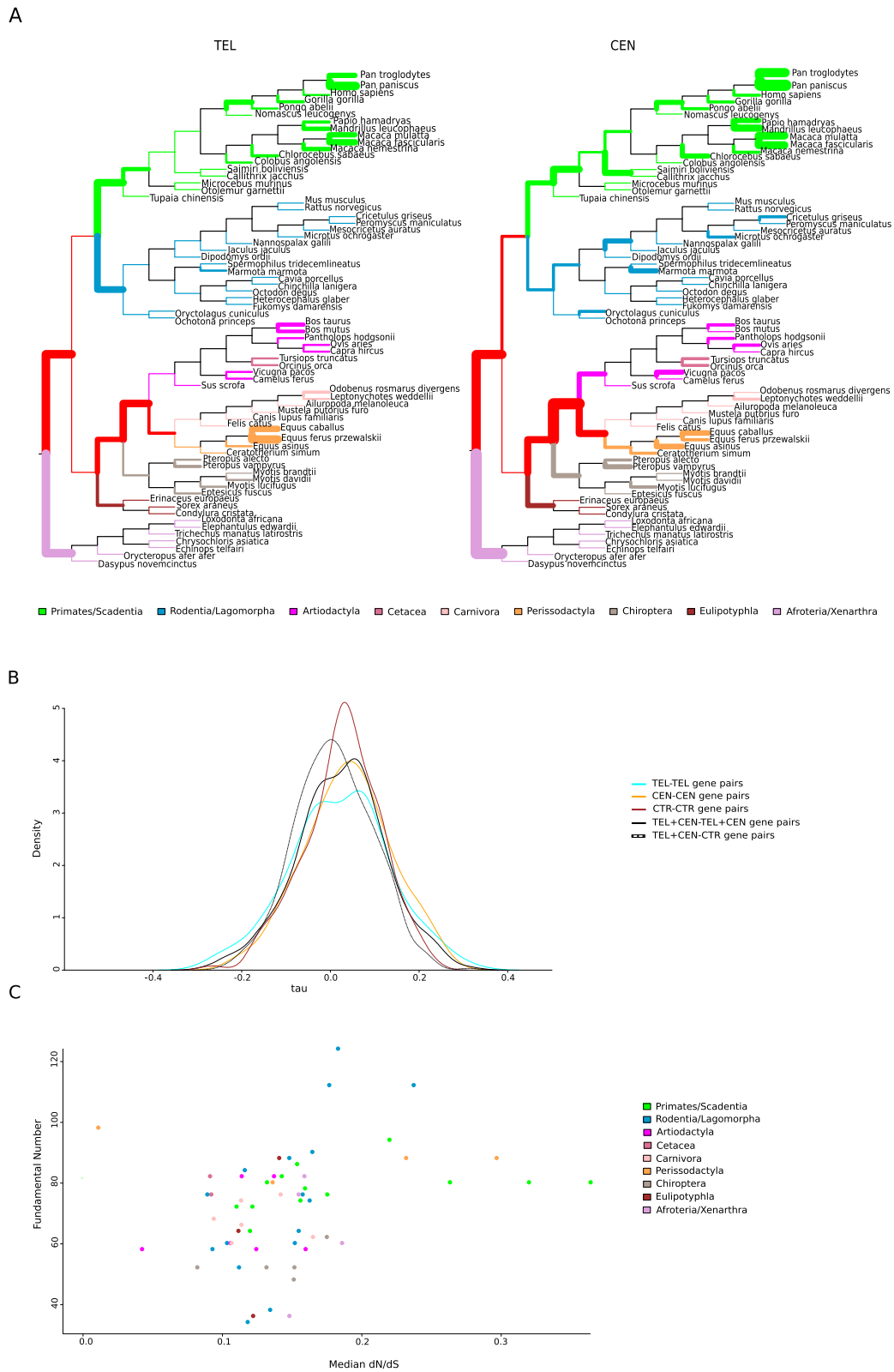
### TEL and CEN genes in eukaryotes with diverse meiotic programs

If the selective pressure responsible for TEL and CEN gene evolution mainly derives from their roles during gametogenesis, eukaryotes with different meiotic programs may display distinct selection patterns, as previously noted for CENPA in ciliates and yeasts (59,60).

Telomere biology has been deeply investigated in *Tetrahymena* species (ciliates that only carry out asymmetric meiosis) and several proteins that associate with the telomerase and/or promote telomere maintenance have been characterized (44–50). We exploited this information and the availability of four fully sequenced *Tetrahymena* genomes to search for orthologs of 18 *Tetrahymena thermophila* genes that encode proteins involved in telomere biology. For 11 genes, orthologs were identified in the three other *Tetrahymena* genomes (*T. borealis*, *T. malaccensis* and *T. elliotti*). An equal number of control genes were also selected (see methods).

Positive selection tests have little power when only few sequences are analyzed, but they are not prone to false positives (61). We however used two methods, that are based on different assumptions of dN/dS variation, to test for positive selection: the M7/M8 LRT and BUSTED. Both methods detected positive selection at four TEL genes (three in common) (Table 1). In contrast, none of the control genes showed evidence of positive selection with either the M7/M8 LRT or with BUSTED (Supplementary Table S10). Overall, these data suggest that positive selection is relatively common at TEL genes in *Tetrahymena*.

Finally, we analyzed TEL and CEN genes in yeasts, organisms with symmetric meiosis only. In particular, we used data from a genome-wide analysis of positive selection across five *Saccharomyces sensu stricto* species (54). We retrieved 19 genes involved in telomere integrity and elongation from Gene Ontology classifications for *Saccharomyces cerevisiae* (see methods and Supplementary Table S11): 17 of these were included in the genome-wide analysis and none showed signals of positive selection (based on the M7/M8 LRT). The same applied to three genes that encode telomere-associated proteins specifically expressed in



**Figure 6.** Selective patterns across the mammalian phylogeny. **(A)** Phylogenetic trees are drawn to summarize the selection patterns of TEL and CEN genes. Colored branches were analyzed for genes showing a significant free ratio model. Branch thickness is proportional to the number of genes showing a  $dN/dS > 1$ . Branches are colored based on the mammalian orders as shown in the legend. Internal branches that were analyzed but do not define orders are in red. **(B)** Distributions of tau correlation coefficients. The distributions of pairwise  $dN/dS$  correlation coefficients calculated for gene pairs from different groups are reported, as described in the text. **(C)** Relationship between selection strength and karyotype. Correlation plot between species median  $dN/dS$  calculated for all genes selected in the free ratio model and the fundamental number of those species. Species are colored as in panel (A).

**Table 1.** Likelihood ratio test statistics for models of variable selective pressure among sites for *Tetrahymena* TEL genes

Systematic gene name	Standard gene name	Description	$-2\Delta\ln L^a$	M7 vs M8 $P$ value <sup>b</sup>	BUSTED $P$ value	References
<b>TTHERM_00658760</b>	<i>TAP19</i> (p19)	Telomerase- (or TERT-) associated protein of 19 kDa	$9.20 \times 10^{-5}$	0.999	1	(48)
<b>TTHERM_00429730</b>	<i>TTP80</i> (p80)	Telomerase- (or TERT-) associated protein of 80 kDa	3.09	0.214	0.12	(50)
<b>TTHERM_00218760</b>	<i>Teb1</i> (p82)	TEB1 (Telomerase DNA Binding subunit 1; also known as Telomerase- (or TERT-) associated protein of 82 kDa)	7.79	<b>0.020</b>	0.84	(48)
<b>TTHERM_00985160</b>	<i>TTP95</i> (p95)	Telomerase- (or TERT-) associated protein of 95 kDa	8.73	<b>0.013</b>	<b>0.001</b>	(50)
<b>TTHERM_00013120</b>	<i>PAT1</i>	Pot1 associated in <i>Tetrahymena</i> 1	0.001	0.999	0.68	(47,91,92)
<b>TTHERM_000378989</b>	<i>POT1</i>	Protection of telomeres 1	1.71	0.426	1	(47,49)
<b>TTHERM_00726370</b>	<i>RLP1</i>	Replication factor-A-like protein 1	7.23	<b>0.027</b>	<b>0.016</b>	(46)
<b>TTHERM_00459400</b>	<i>RLP2</i>	RPA-Like Protein 2	2.35	0.309	0.31	(45)
<b>TTHERM_00106890</b>	<i>RFA1</i>	Homolog of budding yeast RFA1 and mouse RPA1	10.86	<b>0.004</b>	<b>0.0001</b>	(48)
<b>TTHERM_00439320</b>	<i>Teb3</i>	TELomerase DNA Binding subunit 3	3.41	0.182	<b>0.005</b>	(93)
<b>TTHERM_000523058</b>	<i>Tpt1</i>	TPP1/TPz1 in <i>Tetrahymena thermophila</i> , POT1 associated protein	4.46	0.107	0.076	(47,91,92)

<sup>a</sup>Twice the difference of likelihood for the two models compared.

<sup>b</sup> $P$  value of rejecting the neutral model (M7) in favor of the positive selection model (M8).

meiosis (*Ndj1*, *Csm4*, and *Mps3*) (53) (Supplementary Table S11). As for CEN genes, positive selection was not detected for the functional homolog of mammalian CENPA (encoded by *Cse4*), in agreement with previous observations (60). However, *Scm3*, encoding the functional homolog of the HJURP chaperone (19), showed statistically significant evidence of positive selection (M7/M8 LRT,  $P$  value = 0.0081) (Supplementary Table S11).

## DISCUSSION

Herein we show that TEL genes evolve at different rates depending on their function and we detect a significant correlation between the evolutionary rates of TEL genes and their up-regulation during mouse female (but not male) meiosis. No evidence that selection at TEL genes is driven by life history traits was obtained. Conversely, episodes of positive selection across mammalian species were found to be associated with karyotype features.

These data may be consistent with a telomere drive hypothesis, whereby telomeres and telomere-binding proteins are engaged in an intra-genomic conflict similar to the one described for centromeres. The centromere drive hypothesis posits that selfish centromeric DNA elements promote their preferential inclusion in the oocyte through the recruitment of kinetochore components (9–12,62). Because the effects of centromere drive are deleterious during the symmetric male gametogenesis, CEN genes are thought to be under selective pressure to evolve mutations that restore meiotic balance (9–12,62). However, as noted elsewhere (59), deleterious effects of meiotic drive may also occur at the female meiosis itself. Indeed, our results on germline expression suggest that in the case of TEL genes the whole conflict occurs at oogenesis. This is also supported by the observation that TEL genes evolve under positive selection in eukaryotes with asymmetric meiosis only, but not in those with symmetric meiosis alone. In this respect, we note that *Tetrahymena* species differ from most eukaryotes as these ciliates carry out programmed somatic chromosome breakage and *de novo* telomere addition (63). We cannot there-

fore conclude that the selective forces acting on *Tetrahymena* TEL genes are only related to meiotic drive. We also mention that the detection of positive selection at the yeast *Scm3* gene is not necessarily in contrast with the centromere-drive hypothesis. For instance, the 2  $\mu$ m plasmid of *S. cerevisiae* recruits Cse4 to its partitioning locus *STB* to stably propagate itself (64). Because the association of Cse4 with *STB* is dependent on *Scm3* (65), the latter may be involved in a genetic conflict with extrachromosomal selfish DNA elements (9,12).

Clearly, a major difference between telomeres and centromeres is that centromeric DNA evolves rapidly whereas mammalian telomeres are invariably composed of 5'-TTAGGG-3' repeats. The centromere drive hypothesis indicates that positive selection at CENPA (and other CEN proteins) affects its centromeric DNA binding preferences, thus suppressing centromere drive (9,62). However, this model was recently questioned by the observation that selected sites in the *Drosophila* CENPA are unlikely to affect DNA binding specificity, but rather modulate its association with the CAL1 chaperone (66). Thus, positive selection at CENPA was proposed to modulate the efficiency of its centromeric deposition by CAL1 (66). Whatever the underlying mechanism, the role of centromeric sequences was recently supported by the observation that expanded satellite repeats act as selfish elements at mouse meiosis by incorporating more CENPA containing nucleosomes (67).

Telomeric DNA is invariant in sequence but, in analogy to centromeric/pericentromeric chromosome regions, subtelomeric DNA is a preferential site for segmental duplications (68). Subtelomeres are enriched of subtelomeric repeat elements (SREs) and large variant alleles, mainly composed of SREs, have been described for many human chromosomes (15,68). Analysis of primate genomes indicated rapid evolution of subtelomeric regions, with intrachromosomal recombination playing an important role in this process (15). Notably, large blocks of subterminal heterochromatin are common in chimpanzee and gorilla chromosomes but are absent in human chromosomes. These blocks, which are composed primarily of specific sets of segmental dupli-

cations and subterminal satellites (69,70), were suggested to prolong the bouquet stage during meiotic prophase in chimpanzee (71).

In humans, growing evidence also suggests that, at least in somatic cells, different chromosome arms display distinct telomere dynamics and that subtelomere-specific and haplotype-specific effects influence telomere length, indicating the presence of *cis*-acting regulators of telomere dynamics (72–76). Although such regulators are presently uncharacterized, they are likely to reside in the highly variable subtelomeric regions and may act as selfish elements by promoting their preferential segregation during female meiosis.

Bouquet formation at meiosis is highly conserved across eukaryotes and it is necessary to assist homolog pairing, to resolve entanglements between non-homologs, and to ensure recombination (13). The recombination events are in turn crucial to ensure proper segregation of homologous chromosomes during the first meiotic metaphase. In mammals, shelterin is released during prophase and the TERB1/2-MAJIN complex establishes a direct link between telomeric DNA and the inner nuclear membrane (14). Mice lacking any of the three proteins of the complex display no overt somatic phenotype but are infertile due to meiotic arrest (14). Chromosome movements and bouquet formation are impaired in these mutants (14).

Clearly, it remains to be clarified whether and how bouquet formation or other features associated to meiotic telomeres can be exploited by selfish DNA elements and how this may have negative effects on oogenesis. One possibility is that differences in telomere length (determined by subtelomeric sequences) influence the directionality of chromosome movements along the nuclear envelope, and thus affect chromosome positioning at later stages, eventually skewing segregation. Altered chromosome topology at the bouquet stage may in turn increase the probability of segregation errors for other homolog pairs (Figure 7A).

Alternatively, telomere length/subtelomere sequences may influence remodeling during bouquet formation and alter the rigidity of the telomeric region or modify the chromosome's biophysical properties. Indeed, recent evidence suggest that TERB1 recruits cohesin to telomeres, resulting in rigid structures reminiscent of centromeres (77), and two of the TERB1 selected sites we detected are located within the cohesin-binding region. In this scenario, selfish elements may hoard cohesin (or some rate-limiting protein component) away from other homologs (Figure 7B). Interestingly, decreased cohesion not only at centromeric regions, but also along distal chromosome arms is thought to contribute to cross-over terminalization, bivalent hypostretching, and eventually mis-segregation during aged mouse oogenesis (78).

Another possible mechanism is that telomeres directly interact with spindle components. In fact, evidence from mouse oocytes indicates that telomere dysfunction leads to disruption of the meiotic spindle, misalignment of metaphase chromosomes, and a high incidence of lagging chromosomes (79).

A recent study in *Drosophila* indicated that extreme telomere-length variations do not cause meiotic drive, and identified a weak driver linked to a centromere region (80). It should however be noted that *Drosophila* telomeres

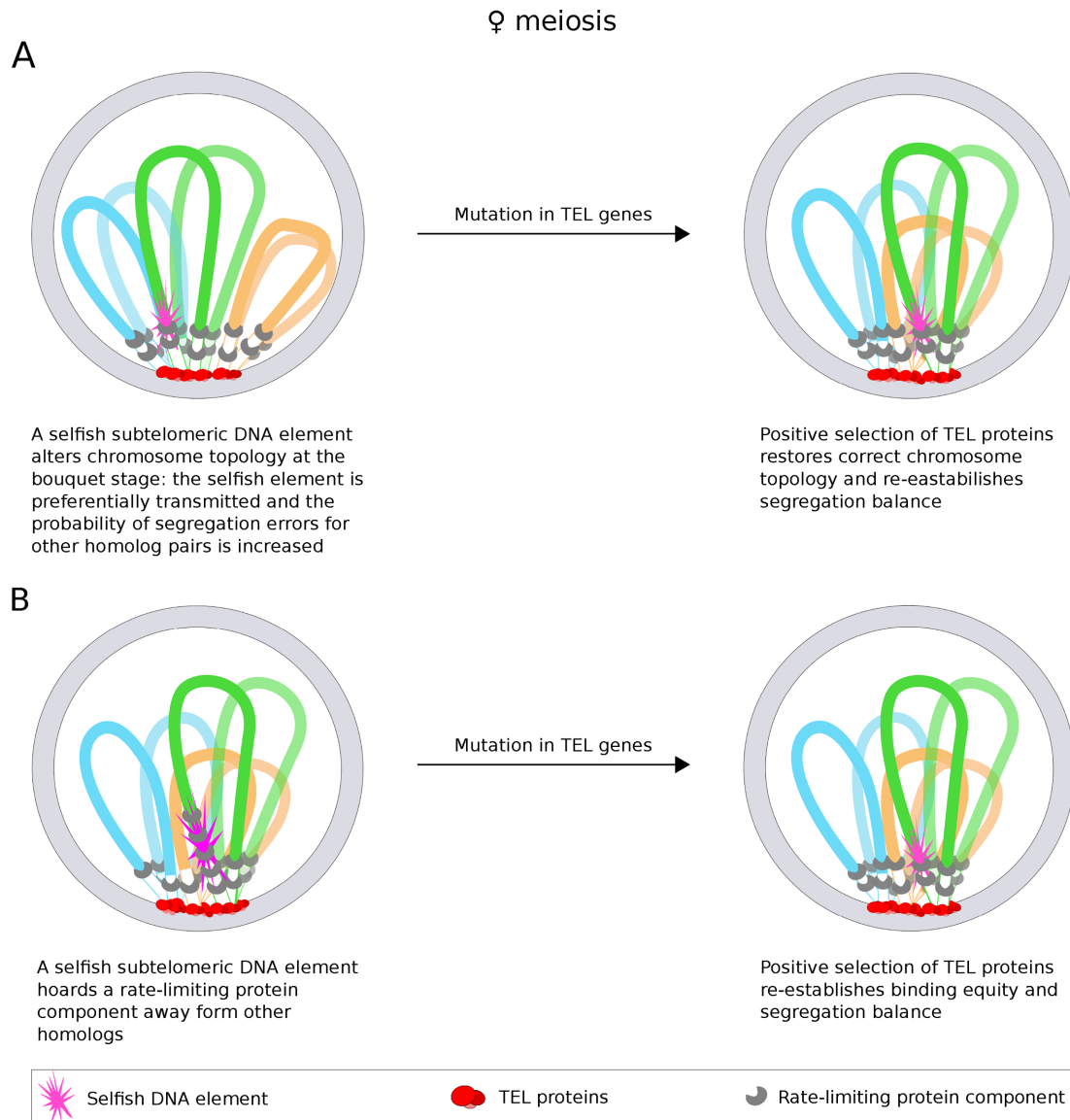
are not maintained by telomerase and fruitfly lineages have extremely variable and labile telomere lengths (80). Some aspects of telomere biology may thus totally differ from organisms that canonically rely on telomerase. Indeed, whereas crosses of *Drosophila* lines with short and long telomeres were fertile, experiments in mice indicated that large differences in telomere lengths at homologous chromosomes compromise meiotic pairing, synapsis, and recombination, especially at the female meiosis (81).

In contrast to the *Drosophila* study, evidence of transmission ratio distortion at a telomeric (and at a centromeric) region was previously obtained in chicken (82). Interestingly, these animals, like most other birds, have extremely long telomeres (83).

Finally, we mention that selfish elements may affect telomere features different from length. Like centromeres, telomeres and subtelomeres are assembled into heterochromatin domains (84). Epigenetic features may thus act as selfish elements, again a situation reminiscent of centromeric regions, which are mainly specified epigenetically (85).

Comparison of TEL and CEN genes indicated that their evolutionary patterns across the mammalian phylogeny are similar and differ from those of control genes. This suggests that features directly (e.g. frequency of polyandry) or indirectly (e.g. climatic changes) connected to the ecology of meiotic drive (86) have played a role in shaping the diversity of these loci over evolutionary times. In fact, limited studies of simple organisms have indicated that, irrespective of their nature and molecular mechanisms, interaction between drivers and environmental factors can modulate the fitness of carrying populations/species (86). For TEL genes, however, we found that evolutionary rates are also related to the number of chromosome arms. A very simplistic interpretation of this finding is that more chromosome arms and, therefore more telomeres, in a karyotype imply a higher possible number of selfish DNA elements and stronger selection for driver suppressors. Another possibility is that the correlation we observed is accounted for by variations in recombination rate (which in turn correlates with the number of chromosomes/chromosome arms in mammals (87,88)). Finally, the correlation between chromosome arms and TEL gene evolutionary rates may depend on chromosome average size. Because genome size is relatively constant in mammals, a high number of chromosome arms implies that the karyotype is dominated by small chromosomes (88). At least in humans, small chromosomes also tend to display the highest probability of meiotic segregation errors at oogenesis, especially in aged women (89). Finally, we note that mammalian karyotypes evolve rapidly, partially because of meiotic drive that favors or disfavors the inclusion of Robertsonian fusions versus metacentrics in the oocyte (40,85,90). The correlation between chromosome arms and evolutionary rates at TEL genes may thus reflect more complex dynamics.

Our study provides evidence that mammalian genes involved in telomere homeostasis are common targets of positive selection, which might ensue from an intra-genomic conflict akin to that described for centromeres and centromere-associated proteins. Centromeres and telomeres can potentially influence chromosome movement during



**Figure 7.** Possible models of telomere drive. (A) Via binding to TEL proteins, a subtelomeric selfish DNA element influences the directionality of chromosome movements and thus affects chromosome topology at the bouquet stage. This, in turn, modifies chromosome positioning at later stages (not shown), skews segregation in favor of the selfish element, and results in increased probability of segregation errors for other homolog pairs (orange pair). Mutations in TEL proteins that reinstate correct chromosome movements are positively selected so that bouquet topology and segregation balance are restored. (B) A selfish DNA element recruits an excess of a rate-limiting protein component, thus curtailing the availability of this same component for other chromosomes. Protein binding avidity favors the transmission of the selfish elements but increases segregation errors for other homolog pairs. Mutations in TEL proteins that restore equal binding are selected for and segregation balance is re-established.

meiosis and thus represent excellent candidates as meiotic drivers. These findings raise the possibility that telomere dynamics influence karyotype evolution and eventually speciation in mammals.

#### DATA AVAILABILITY

Coding sequence information were retrieved from the NCBI database (<http://www.ncbi.nlm.nih.gov/>), from the Ensemble website (<http://www.ensembl.org/index.html>), and from UCSC server (<http://genome.ucsc.edu/>), which provide access to biomedical and genomic information.

The list of TEL genes mutations was obtained from the Human Gene Mutation Database (HGMD, <http://www.hgmd.cf.ac.uk/ac/>, last accessed December 2017) and from the Online Mendelian Inheritance in Man (OMIM, <https://www.omim.org/>, last accessed December 2017) database.

Life-history traits were retrieved for all mammalian species of the AnAge database (<http://genomics.senescence.info/species/>).

#### SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## FUNDING

The authors received no specific funding for this work. Funding for open access charge: Italian Ministry of Health, 'Ricerca Corrente program' [RC3.06 to M.S.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Linger, B.R. and Price, C.M. (2009) Conservation of telomere protein complexes: shuffling through evolution. *Crit. Rev. Biochem. Mol. Biol.*, **44**, 434–446.
- Schmidt, J.C. and Cech, T.R. (2015) Human telomerase: biogenesis, trafficking, recruitment, and activation. *Genes Dev.*, **29**, 1095–1105.
- Sarek, G., Vannier, J.B., Panier, S., Petrini, J.H. and Boulton, S.J. (2015) TRF2 recruits RTEL1 to telomeres in S phase to promote t-loop unwinding. *Mol. Cell*, **57**, 622–635.
- Chang, J.T., Chen, Y.L., Yang, H.T., Chen, C.Y. and Cheng, A.J. (2002) Differential regulation of telomerase activity by six telomerase subunits. *Eur. J. Biochem.*, **269**, 3442–3450.
- Gorbunova, V., Seluanov, A., Zhang, Z., Gladyshev, V.N. and Vijg, J. (2014) Comparative genetics of longevity and cancer: Insights from long-lived rodents. *Nat. Rev. Genet.*, **15**, 531–540.
- Gomes, N.M., Ryder, O.A., Houck, M.L., Charter, S.J., Walker, W., Forsyth, N.R., Austad, S.N., Venditti, C., Pagel, M., Shay, J.W. *et al.* (2011) Comparative biology of mammalian telomeres: Hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*, **10**, 761–768.
- Olsson, M., Wapstra, E. and Friesen, C.R. (2017) Evolutionary ecology of telomeres: A review. *Ann. N. Y. Acad. Sci.*, doi:10.1111/nyas.13443.
- Lee, Y.C., Leek, C. and Levine, M.T. (2017) Recurrent innovation at genes required for telomere integrity in drosophila. *Mol. Biol. Evol.*, **34**, 467–482.
- Henikoff, S., Ahmad, K. and Malik, H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098–1102.
- Henikoff, S. and Malik, H.S. (2002) Centromeres: selfish drivers. *Nature*, **417**, 227.
- Malik, H.S. (2009) The centromere-drive hypothesis: a simple basis for centromere complexity. *Prog. Mol. Subcell. Biol.*, **48**, 33–52.
- Malik, H.S. and Henikoff, S. (2009) Major evolutionary transitions in centromere complexity. *Cell*, **138**, 1067–1082.
- Ebrahimi, H. and Cooper, J.P. (2016) Finding a place in the SUN: telomere maintenance in a diverse nuclear landscape. *Curr. Opin. Cell Biol.*, **40**, 145–152.
- Shibuya, H., Hernandez-Hernandez, A., Morimoto, A., Negishi, L., Hoog, C. and Watanabe, Y. (2015) MAJIN links telomeric DNA to the nuclear membrane by exchanging telomere cap. *Cell*, **163**, 1252–1266.
- Linardopoulou, E.V., Williams, E.M., Fan, Y., Friedman, C., Young, J.M. and Trask, B.J. (2005) Human subtelomeres are hot spots of interchromosomal recombination and segmental duplication. *Nature*, **437**, 94–100.
- Vilella, A.J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R. and Birney, E. (2009) EnsemblCompara GeneTrees: complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res.*, **19**, 327–335.
- Wu, L., Multani, A.S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J.M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S.M. *et al.* (2006) Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell*, **126**, 49–62.
- Hockemeyer, D., Daniels, J.P., Takai, H. and de Lange, T. (2006) Recent expansion of the telomeric complex in rodents: two distinct POT1 proteins protect mouse telomeres. *Cell*, **126**, 63–77.
- McKinley, K.L. and Cheeseman, I.M. (2016) The molecular basis for centromere identity and function. *Nat. Rev. Mol. Cell Biol.*, **17**, 16–29.
- Wernersson, R. and Pedersen, A.G. (2003) RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.*, **31**, 3537–3539.
- Capella-Gutierrez, S., Silla-Martinez, J.M. and Gabaldon, T. (2009) trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, **25**, 1972–1973.
- Xia, X. (2013) DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. *Mol. Biol. Evol.*, **30**, 1720–1728.
- Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. and Frost, S.D. (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.*, **23**, 1891–1901.
- Kosakovsky Pond, S.L. and Frost, S.D. (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.*, **22**, 1208–1222.
- Guindon, S., Delsuc, F., Dufayard, J.F. and Gascuel, O. (2009) Estimating maximum likelihood phylogenies with PhyML. *Methods Mol. Biol.*, **537**, 113–137.
- Yang, Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.*, **24**, 1586–1591.
- Yang, Z. and Nielsen, R. (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. *J. Mol. Evol.*, **46**, 409–418.
- Yang, Z., Wong, W.S. and Nielsen, R. (2005) Bayes empirical bayes inference of amino acid sites under positive selection. *Mol. Biol. Evol.*, **22**, 1107–1118.
- Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. and Scheffler, K. (2013) FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol. Evol.*, **30**, 1196–1205.
- Pond, S.L., Frost, S.D. and Muse, S.V. (2005) HyPhy: hypothesis testing using phylogenies. *Bioinformatics*, **21**, 676–679.
- Bielawski, J.P. and Yang, Z. (2004) A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution. *J. Mol. Evol.*, **59**, 121–132.
- Weadick, C.J. and Chang, B.S. (2012) An improved likelihood ratio test for detecting site-specific functional divergence among clades of protein-coding genes. *Mol. Biol. Evol.*, **29**, 1297–1300.
- Bininda-Emonds, O.R., Cardillo, M., Jones, K.E., MacPhee, R.D., Beck, R.M., Grenyer, R., Price, S.A., Vos, R.A., Gittleman, J.L. and Purvis, A. (2007) The delayed rise of present-day mammals. *Nature*, **446**, 507–512.
- Pagel, M., Meade, A. and Barker, D. (2004) Bayesian estimation of ancestral character states on phylogenies. *Syst. Biol.*, **53**, 673–684.
- Lartillot, N. and Poujol, R. (2011) A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.*, **28**, 729–744.
- Soh, Y.Q., Junker, J.P., Gill, M.E., Mueller, J.L., van Oudenaarden, A. and Page, D.C. (2015) A gene regulatory program for meiotic prophase in the fetal ovary. *PLoS Genet.*, **11**, e1005531.
- Margolin, G., Khil, P.P., Kim, J., Bellani, M.A. and Camerini-Otero, R.D. (2014) Integrated transcriptome analysis of mouse spermatogenesis. *BMC Genomics*, **15**, 39.
- Kuchta-GA, adysz, M., Grabowska-Joachimak, A., Szeleszczuk, O., Szczerbal, I., Kociucka, B. and NiedbaA, a.P. (2015) Karyotyping of chinchilla lanigera mol. (rodentia, chinchillidae). *Caryologia*, **68**, 138–146.
- Modi, W.S. (1987) Phylogenetic analyses of chromosomal banding patterns among the nearctic arvicolidae (mammalia: Rodentia). *Syst. Zool.*, **36**, 109–136.
- Pardo-Manuel de Villena, F. and Sapienza, C. (2001) Female meiosis drives karyotypic evolution in mammals. *Genetics*, **159**, 1179–1189.
- Arslan, A. and Zima, J. (2015) Heterochromatin distribution and localization of nucleolar organizing regions in the 2n = 56 cytotypes of nannospalax xanthodon and N. ehrenbergi from turkey. *Zool. Stud.*, **54**, 6.
- Sotero-Caio, C., Baker, R.J. and Volleth, M. (2017) Chromosomal evolution in chiroptera. *Genes*, **8**, 272.
- Ben Faleh, A., Ben Othmen, A., Said, K. and Granjon, L. (2010) Karyotypic variation in two species of jerboas jaculus jaculus and jaculus orientalis (rodentia, dipodidae) from tunisia. *Folia. Biol. (Krakow)*, **58**, 229–236.
- Wang, Y. and Feigon, J. (2017) Structural biology of telomerase and its interaction at telomeres. *Curr. Opin. Struct. Biol.*, **47**, 77–87.
- Upton, H.E., Chan, H., Feigon, J. and Collins, K. (2017) Shared subunits of tetrahymena telomerase holoenzyme and replication protein A have different functions in different cellular complexes. *J. Biol. Chem.*, **292**, 217–228.
- Min, B. and Collins, K. (2010) Multiple mechanisms for elongation processivity within the reconstituted tetrahymena telomerase holoenzyme. *J. Biol. Chem.*, **285**, 16434–16443.

47. Cranert, S., Heyse, S., Linger, B.R., Lescasse, R. and Price, C. (2014) Tetrahymena Pot2 is a developmentally regulated paralog of Pot1 that localizes to chromosome breakage sites but not to telomeres. *Eukaryot. Cell.*, **13**, 1519–1529.
48. Min, B. and Collins, K. (2009) An RPA-related sequence-specific DNA-binding subunit of telomerase holoenzyme is required for elongation processivity and telomere maintenance. *Mol. Cell.*, **36**, 609–619.
49. Jacob, N.K., Lescasse, R., Linger, B.R. and Price, C.M. (2007) Tetrahymena POT1a regulates telomere length and prevents activation of a cell cycle checkpoint. *Mol. Cell Biol.*, **27**, 1592–1601.
50. Collins, K., Kobayashi, R. and Greider, C.W. (1995) Purification of tetrahymena telomerase and cloning of genes encoding the two protein components of the enzyme. *Cell*, **81**, 677–686.
51. Stover, N.A., Punia, R.S., Bowen, M.S., Dolins, S.B. and Clark, T.G. (2012) Tetrahymena genome database wiki: a community-maintained model organism database. *Database (Oxford)*, **2012**, bas007.
52. Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M. *et al.* (2015) Gene-wide identification of episodic selection. *Mol. Biol. Evol.*, **32**, 1365–1371.
53. Conrad, M.N., Lee, C.Y., Chao, G., Shinohara, M., Kosaka, H., Shinohara, A., Conchello, J.A. and Dresser, M.E. (2008) Rapid telomere movement in meiotic prophase is promoted by NDJ1, MPS3, and CSM4 and is modulated by recombination. *Cell*, **133**, 1175–1187.
54. Scannell, D.R., Zill, O.A., Rokas, A., Payen, C., Dunham, M.J., Eisen, M.B., Rine, J., Johnston, M. and Hittinger, C.T. (2011) The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)*, **1**, 11–25.
55. Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.*, **13**, 555–556.
56. Lindblad-Toh, K., Garber, M., Zuk, O., Lin, M.F., Parker, B.J., Washietl, S., Kheradpour, P., Ernst, J., Jordan, G., Mauceli, E. *et al.* (2011) A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, **478**, 476–482.
57. Figuet, E., Nabholtz, B., Bonneau, M., Mas Carrio, E., Nadachowska-Brzyska, K., Ellegren, H. and Galtier, N. (2016) Life history traits, protein evolution, and the nearly neutral theory in amniotes. *Mol. Biol. Evol.*, **33**, 1517–1527.
58. Ferguson-Smith, M.A. and Trifonov, V. (2007) Mammalian karyotype evolution. *Nat. Rev. Genet.*, **8**, 950–962.
59. Elde, N.C., Roach, K.C., Yao, M.C. and Malik, H.S. (2011) Absence of positive selection on centromeric histones in tetrahymena suggests un-suppressed centromere: Drive in lineages lacking male meiosis. *J. Mol. Evol.*, **72**, 510–520.
60. Talbert, P.B., Bryson, T.D. and Henikoff, S. (2004) Adaptive evolution of centromere proteins in plants and animals. *J. Biol.*, **3**, 18.
61. Wong, W.S., Yang, Z., Goldman, N. and Nielsen, R. (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. *Genetics*, **168**, 1041–1051.
62. Rosin, L.F. and Mellone, B.G. (2017) Centromeres drive a hard bargain. *Trends Genet.*, **33**, 101–117.
63. Coyne, R.S. and Yao, M.C. (1996) Evolutionary conservation of sequences directing chromosome breakage and rDNA palindrome formation in tetrahymenine ciliates. *Genetics*, **144**, 1479–1487.
64. Hajra, S., Ghosh, S.K. and Jayaram, M. (2006) The centromere-specific histone variant Cse4p (CENP-A) is essential for functional chromatin architecture at the yeast 2-microm circle partitioning locus and promotes equal plasmid segregation. *J. Cell Biol.*, **174**, 779–790.
65. Huang, C.C., Hajra, S., Ghosh, S.K. and Jayaram, M. (2011) Cse4 (CenH3) association with the *Saccharomyces cerevisiae* plasmid partitioning locus in its native and chromosomally integrated states: implications in centromere evolution. *Mol. Cell Biol.*, **31**, 1030–1040.
66. Rosin, L. and Mellone, B.G. (2016) Co-evolving CENP-A and CAL1 domains mediate centromeric CENP-A deposition across drosophila species. *Dev. Cell.*, **37**, 136–147.
67. Iwata-Otsubo, A., Dawicki-McKenna, J.M., Akera, T., Falk, S.J., Chmatal, L., Yang, K., Sullivan, B.A., Schultz, R.M., Lampson, M.A. and Black, B.E. (2017) Expanded satellite repeats amplify a discrete CENP-A nucleosome assembly site on chromosomes that drive in female meiosis. *Curr. Biol.*, **27**, 2365–2373.
68. Mefford, H.C. and Trask, B.J. (2002) The complex structure and dynamic evolution of human subtelomeres. *Nat. Rev. Genet.*, **3**, 91–102.
69. Royle, N.J., Baird, D.M. and Jeffreys, A.J. (1994) A subterminal satellite located adjacent to telomeres in chimpanzees is absent from the human genome. *Nat. Genet.*, **6**, 52–56.
70. Ventura, M., Catacchio, C.R., Sajjadian, S., Vives, L., Sudmant, P.H., Marques-Bonet, T., Graves, T.A., Wilson, R.K. and Eichler, E.E. (2012) The evolution of african great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Res.*, **22**, 1036–1049.
71. Hirai, H., Matsubayashi, K., Kumazaki, K., Kato, A., Maeda, N. and Kim, H.S. (2005) Chimpanzee chromosomes: Retrotransposable compound repeat DNA organization (RCRO) and its influence on meiotic prophase and crossing-over. *Cytogenet. Genome Res.*, **108**, 248–254.
72. Graakjaer, J., Bischoff, C., Korsholm, L., Holstebro, S., Vach, W., Bohr, V.A., Christensen, K. and Kolvraa, S. (2003) The pattern of chromosome-specific variations in telomere length in humans is determined by inherited, telomere-near factors and is maintained throughout life. *Mech. Ageing Dev.*, **124**, 629–640.
73. Graakjaer, J., Der-Sarkissian, H., Schmitz, A., Bayer, J., Thomas, G., Kolvraa, S. and Londono-Vallejo, J.A. (2006) Allele-specific relative telomere lengths are inherited. *Hum. Genet.*, **119**, 344–350.
74. Britt-Compton, B., Rowson, J., Locke, M., Mackenzie, I., Kipling, D. and Baird, D.M. (2006) Structural stability and chromosome-specific telomere length is governed by cis-acting determinants in humans. *Hum. Mol. Genet.*, **15**, 725–733.
75. McCaffrey, J., Young, E., Lassahn, K., Sibert, J., Pastor, S., Riethman, H. and Xiao, M. (2017) High-throughput single-molecule telomere characterization. *Genome Res.*, **27**, 1904–1915.
76. Young, E., Pastor, S., Rajagopalan, R., McCaffrey, J., Sibert, J., Mak, A.C.Y., Kwok, P.Y., Riethman, H. and Xiao, M. (2017) High-throughput single-molecule mapping links subtelomeric variants and long-range haplotypes with specific telomeres. *Nucleic Acids Res.*, **45**, e73.
77. Shibuya, H., Ishiguro, K. and Watanabe, Y. (2014) The TRF1-binding protein TERB1 promotes chromosome movement and telomere rigidity in meiosis. *Nat. Cell Biol.*, **16**, 145–156.
78. Greaney, J., Wei, Z. and Homer, H. (2017) Regulation of chromosome segregation in oocytes and the cellular basis for female meiotic errors. *Hum. Reprod. Update*, doi:10.1093/humupd/dmx035.
79. Liu, L., Blasco, M., Trimarchi, J. and Keefe, D. (2002) An essential role for functional telomeres in mouse germ cells during fertilization and early development. *Dev. Biol.*, **249**, 74–84.
80. Wei, K.H., Reddy, H.M., Rathnam, C., Lee, J., Lin, D., Ji, S., Mason, J.M., Clark, A.G. and Barbash, D.A. (2017) A pooled sequencing approach identifies a candidate meiotic driver in drosophila. *Genetics*, **206**, 451–465.
81. Liu, L., Franco, S., Spyropoulos, B., Moens, P.B., Blasco, M.A. and Keefe, D.L. (2004) Irregular telomeres impair meiotic synapsis and recombination in mice. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 6496–6501.
82. Axelsson, E., Albrechtsen, A., van, A.P., Li, L., Megens, H.J., Vereijken, A.L., Crooijmans, R.P., Groenen, M.A., Ellegren, H., Willerslev, E. *et al.* (2010) Segregation distortion in chicken and the evolutionary consequences of female meiotic drive in birds. *Heredity (Edinb)*, **105**, 290–298.
83. Delany, M.E., Gessaro, T.M., Rodrigue, K.L. and Daniels, L.M. (2007) Chromosomal mapping of chicken mega-telomere arrays to GGA9, 16, 28 and W using a cytogenomic approach. *Cytogenet. Genome Res.*, **117**, 54–63.
84. Blasco, M.A. (2007) The epigenetic regulation of mammalian telomeres. *Nat. Rev. Genet.*, **8**, 299–309.
85. Chmatal, L., Schultz, R.M., Black, B.E. and Lampson, M.A. (2017) Cell biology of cheating-transmission of centromeres and other selfish elements through asymmetric meiosis. *Prog. Mol. Subcell. Biol.*, **56**, 377–396.
86. Lindholm, A.K., Dyer, K.A., Firman, R.C., Fishman, L., Forstmeier, W., Holman, L., Johannesson, H., Knief, U., Kokko, H., Larracuente, A.M. *et al.* (2016) The ecology and evolutionary dynamics of meiotic drive. *Trends Ecol. Evol.*, **31**, 315–326.
87. Pardo-Manuel de Villena, F. and Sapienza, C. (2001) Recombination is proportional to the number of chromosome arms in mammals. *Mamm. Genome*, **12**, 318–322.

88. Dumont, B.L. (2017) Variation and evolution of the meiotic requirement for crossing over in mammals. *Genetics*, **205**, 155–168.
89. Capalbo, A., Hoffmann, E.R., Cimadomo, D., Maria Ubaldi, F. and Rienzi, L. (2017) Human female meiosis revised: New insights into the mechanisms of chromosome segregation and aneuploidies from advanced genomics and time-lapse imaging. *Hum. Reprod. Update*, **23**, 706–722.
90. Chmatal, L., Gabriel, S.I., Mitsainas, G.P., Martinez-Vargas, J., Ventura, J., Searle, J.B., Schultz, R.M. and Lampson, M.A. (2014) Centromere strength provides the cell biological basis for meiotic drive and karyotype evolution in mice. *Curr. Biol.*, **24**, 2295–2300.
91. Linger, B.R., Morin, G.B. and Price, C.M. (2011) The Pot1a-associated proteins Tpt1 and Pat1 coordinate telomere protection and length regulation in tetrahymena. *Mol. Biol. Cell*, **22**, 4161–4170.
92. Premkumar, V.L., Cranert, S., Linger, B.R., Morin, G.B., Minium, S. and Price, C. (2014) The 3' overhangs at tetrahymena thermophila telomeres are packaged by four proteins, Pot1a, Tpt1, Pat1, and Pat2. *Eukaryot. Cell*, **13**, 240–245.
93. Jiang, J., Chan, H., Cash, D.D., Miracco, E.J., Ogorzalek Loo, R.R., Upton, H.E., Cascio, D., O'Brien Johnson, R., Collins, K., Loo, J.A. *et al.* (2015) Structure of tetrahymena telomerase reveals previously unknown subunits, functions, and interactions. *Science*, **350**, aab4070.