



HAL
open science

Elucidating the molecular mechanisms of genome evolution in the budding yeast *Saccharomyces cerevisiae*

Melania Jennifer d'Angiolo

► **To cite this version:**

Melania Jennifer d'Angiolo. Elucidating the molecular mechanisms of genome evolution in the budding yeast *Saccharomyces cerevisiae*. Molecular biology. Université Côte d'Azur, 2021. English. NNT : 2021COAZ6001 . tel-03917511

HAL Id: tel-03917511

<https://theses.hal.science/tel-03917511>

Submitted on 2 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

Étude des mécanismes moléculaires de
l'évolution du génome chez la levure
bourgeonnante *Saccharomyces cerevisiae*

Melania Jennifer D'ANGIOLO

IRCAN CNRS UMR7284 INSERM U1081

**Présentée en vue de l'obtention
du grade de docteur en** Sciences de
la vie et de la santé

d'Université Côte d'Azur

Dirigée par : Gianni Liti / Eric Gilson

Soutenue le : 18/01/2021

Devant le jury, composé de :

Miguel GODINHO FERREIRA, DR2
CNRS, IRCAN Nice

Pei-Yun Jenny WU, DR2 CNRS, IGDR
Rennes

Delphine Sicard, DR2 INRA, INRAE
Montpellier

Bernard Dujon, Professeur émérite,
Institut Pasteur Paris

Gianni Liti, DR2 CNRS, IRCAN Nice

Eric Gilson, PU-PH, IRCAN Nice



**Étude des mécanismes moléculaires de l'évolution du
génomme chez la levure bourgeonnante *Saccharomyces
cerevisiae***

Elucidating the molecular mechanisms of genome evolution in the
budding yeast *Saccharomyces cerevisiae*

Jury :

Président du jury

Miguel GODINHO FERREIRA, Directeur de recherche au CNRS,
IRCAN Nice

Rapporteurs

Pei-Yun Jenny WU, Directrice de recherche au CNRS, IGDR Rennes

Delphine Sicard, Directrice de recherche a l'INRA, INRAE Montpellier

Examineurs

Bernard Dujon, Professeur émérite, Institut Pasteur Paris

Étude des mécanismes moléculaires de l'évolution du génome chez la levure bourgeonnante *Saccharomyces cerevisiae*

Résumé

L'évolution du génome consiste en la modification progressive de ce dernier au fil du temps et résulte de la variation des gènes dans leur ensemble, des mutations, des recombinaisons et échanges génétiques entre populations. L'essor des technologies de séquençage de nouvelle génération ainsi que la réduction de leur coût ont permis d'augmenter le nombre de génomes disponibles, permettant d'élucider les mécanismes moléculaires impliqués dans leur évolution. Dans ce travail, j'ai utilisé la levure bourgeonnante *Saccharomyces cerevisiae* pour étudier deux aspects fondamentaux de l'évolution du génome: l'origine des introgressions inter-espèces et la diversité des télomères.

Une introgression est une insertion de matériel génétique provenant d'une population dans une autre. Ce phénomène naît d'événements d'hybridation suivis de rétrocroisements répétés avec l'une des populations parentales. Dans la première partie de ma thèse, j'ai étudié une lignée de *S. cerevisiae* isolée à partir des eaux usées de la production d'huile d'olive (Alpechin), qui comprend d'abondantes introgressions de l'espèce sœur *S. paradoxus*, ainsi qu'une souche hybride *S. cerevisiae/S. paradoxus* caractérisée par des nombreuses régions de perte d'hétérozygotie (LOH). J'ai établi une carte génétique détaillée des LOHs dans la souche hybride et comparé leur position aux introgressions dans les souches d'Alpechin pour déterminer leurs relations. J'ai constaté que LOHs et introgressions se chevauchaient et provenaient de l'ascendance de *S. paradoxus*, indiquant que les introgressions dans la lignée d'Alpechin découlent directement des LOHs. J'ai proposé un modèle pour expliquer l'origine des introgressions chez la levure selon lequel les LOHs permettent aux hybrides inter-espèces de surmonter leur stérilité et j'ai validé la fiabilité de ce postulat à l'aide d'approches expérimentales et informatiques.

Dans la deuxième partie de ma thèse, j'ai caractérisé la diversité des télomères chez *S. cerevisiae* et l'effet d'un stress télomérique sur le fitness cellulaire. Premièrement, j'ai estimé la longueur des télomères dans plus de 900 souches isolées à travers le monde et constaté une vaste hétérogénéité, bien que les souches issues d'habitats naturels présentent des télomères plus courts que celles issues

de l'agroalimentaire. J'ai ensuite réalisé une étude d'association pangénomique qui a permis d'identifier des variants génétiques susceptibles de moduler la longueur des télomères. De plus, j'ai identifié des mutations délétères dans des gènes connus pour influencer la longueur des télomères. J'ai aussi utilisé un ensemble de données phénotypiques pour déterminer si certains facteurs non génétiques sont associés à la variation de longueur des télomères, et j'ai ainsi pu constater une connexion entre le métabolisme mitochondrial et les télomères dans les souches naturelles.

Deuxièmement, j'ai étudié l'effet d'un stress télomérique chronique chez des levures modifiées avec des répétitions télomériques humaines. J'ai fait évoluer ces levures «humanisées» sur plusieurs générations par le biais de lignées cellulaires non-sélectives et j'ai observé un ralentissement de la vitesse de division et de la longévité, parallèlement à une augmentation du taux de mutations. Enfin, j'ai procédé à une expérience d'évolution adaptative pour permettre l'émergence de mutations bénéfiques qui contrecarrent le déclin de fitness des levures «humanisées». Après évolution, la plupart des lignées ont retrouvé leurs caractéristiques originelles grâce à l'apparition de mutations spécifiques en lien avec la réponse aux dommages de l'ADN.

Dans l'ensemble, mes travaux ont permis d'établir une nouvelle hypothèse expliquant l'origine des introgressions chez les espèces reproductivement isolées et de même ont permis de caractériser la diversité et l'instabilité des télomères à une échelle sans précédent, contribuant à l'élucidation des mécanismes moléculaires impliqués dans l'évolution des génomes.

Mots clés : introgression, hybridation, levure, évolution, génome, télomères, sélection.

Elucidating the molecular mechanisms of genome evolution in the budding yeast *Saccharomyces cerevisiae*

Abstract

Genomes are progressively modified during their evolution leading to gene content variation, recombination, mutation and genetic exchange among species/subpopulations. The advent of next-generation sequencing technologies and their cost reduction increased the number of genomes available for evolutionary studies, opening the way to understand the molecular mechanisms involved in genome evolution. In this work, I used the budding yeast *Saccharomyces cerevisiae* as model organism to investigate two important aspects of genome evolution: the origin of interspecies introgressions and telomere evolution.

An introgression is the flow of genetic material between populations and it results from ancient hybridization events followed by repeated backcrossings with one of the parental populations. In the first part of my PhD, I studied a lineage of *S. cerevisiae* strains isolated from the wastewater of olive oil production (Alpechin), carrying abundant introgressions from the sister species *S. paradoxus*, and a natural *S. cerevisiae/S. paradoxus* hybrid, with 50% genome contribution from each parent, carrying abundant regions of loss-of-heterozygosity (LOH). I derived an accurate genetic map of LOHs in the hybrid and compared their position to the introgressions in the Alpechin strains, to infer their evolutionary relations. I found that LOH and introgressions overlapped and shared the same *S. paradoxus* ancestry, indicating that LOHs are the direct origins of introgressions in the Alpechin lineage. I proposed a model for the origin of yeast introgressions in which LOH regions allow interspecies hybrids to overcome sterility, which constitutes the main barrier to introgressions' onset in reproductively isolated species, such as yeasts, and validated the reliability of my model using experimental and computational techniques.

In the second part of my PhD, I studied the extent of telomere diversity in *S. cerevisiae* and the outcome of chronic telomeric stress on cellular fitness. In a first project, I estimated telomere length in over 900 strains isolated around the world and observed remarkable variation. Strains isolated in wild habitats had shorter telomeres than domesticated ones. I performed a genome-wide association study that revealed novel genetic variants possibly regulating telomere length. I also pinpointed

private loss-of-function mutations in known telomere length maintenance genes that could explain the very long/short telomeres of certain lineages. Moreover, I used multiple phenotypic datasets available for this collection to look for non-genetic factors associated to telomere length variation, and discovered an association between mitochondrial metabolism and telomeres in wild strains.

In a second project, I performed experimental evolution of engineered yeasts synthesizing human telomeric DNA repeats at their chromosome-ends. I evolved telomere-humanized strains through mutation accumulation lines (MALs) to minimize selection, and I characterized the detrimental effects caused by telomeres' reshaping. During MALs, humanized yeasts gradually slowed their growth, shortened chronological lifespan and had higher mutation rate and genome instability. Next, I submitted MALs to adaptive evolution by multiple serial transfers (STs) of large population sizes, to map mutations that counteract their fitness decline. After multiple STs, most humanized lines recovered fitness thanks to the independent occurrence of mutations in the DNA-damage response pathway.

Overall, my work contributed to elucidate the molecular mechanisms driving genome evolution, by providing a plausible model for introgression evolution in reproductively isolated species and by giving an unprecedented overview of the impact of the variation of telomere DNA length and sequence on global organismal fitness.

Keywords : introgression, hybridization, yeast, evolution, genome, telomeres, selection.

Table of contents

Chapter 1 - The plasticity of genome evolution: insights from the budding yeast *Saccharomyces cerevisiae*

1.1) What is <i>Saccharomyces cerevisiae</i> ?	2
1.2) What is a genome?	3
1.3) How do genomes evolve?	4
1.3.1) Mutation	4
1.3.2) Genome content variation	7
1.3.3) Balanced structural variation	8
1.3.4) Recombination	10
1.3.5) Other mechanisms of genome evolution	11
1.3.6) Environmental factors and selection	11
References	13

Chapter 2 - The role of hybridization and introgression in genome evolution

2.1) What is an introgression?	20
2.2) The extent of introgressions across the tree of life	21
2.2.1) Introgressions in insects	21
2.2.2) Introgressions in plants	22
2.2.3) Introgressions in mammals	22
2.3) The impact of introgressions in the evolution of the human genome	23
2.3.1) Distribution of archaic introgressions in human populations	23
2.3.2) The phenotypic impact of archaic introgressions in human populations	24
2.4) The importance of reticulated evolution in yeast	26
2.4.1) Hybridization in yeast	26
2.4.2) Introgressions in yeast	27
2.5) Hybrid sterility and the paradox of the origin of yeast introgressions	29
2.6) Possible mechanisms explaining the emergence of yeast introgressions	31
2.6.1) Mechanisms involving the passage through a sporulation stage in the initial hybrid and through a backcrossing phase	31
2.6.2) Mechanisms involving the passage through a sporulation stage in the initial hybrid but not involving a backcrossing phase	32
2.6.3) Mechanisms not involving the passage through a sporulation stage in the initial hybrid	33
2.7) Conclusion	33
References	34

Chapter 3 - The role of telomere diversity in genome evolution and the role of evolution in shaping telomeres	49
3.1) What are telomeres?	49
3.2) Telomere maintenance and replication	52
3.3) Telomere length is a complex trait	53
3.4) The structure of subtelomeres in yeast	55
3.5) The evolutionary contribution of telomere dynamics	58
3.5.1) The evolutionary contribution of telomere length dynamics	58
3.5.1.1) Telomeres as objects of evolution	58
3.5.1.2) Telomere length and lifespan	59
3.5.1.3) Telomere length and growth	61
3.5.1.4) Telomere length and reproduction	62
3.5.1.5) Telomere dynamics in health and disease	65
3.5.1.6) Telomere dynamics and life-history trade-offs: the big picture	66
3.5.2) Telomeres as drivers of evolution	67
3.5.3) The evolutionary contribution of telomere sequence dynamics	70
3.5.3.1) Telomere sequence dynamics in yeasts	74
3.5.3.2) A case-study of telomere evolution: the humanized yeast model	75
References	78
Scientific problems addressed and objectives of this PhD	97
Chapter 4 - A yeast living ancestor reveals the origin of genomic introgressions	101
Abstract	101
Introduction	102
Results	103
Discussion	113
Supplementary discussion	115
Methods	139
References	151
Chapter 5 - Telomeres are shorter in wild <i>S. cerevisiae</i> populations than in domesticated ones	158
Abstract	158
Introduction	159
Results	160
Discussion	171
Supplementary discussion	174
Methods	194

References	198
Chapter 6 - Inactivation of the DNA damage response rescues organismal fitness in telomere-humanized yeast	204
Abstract	204
Introduction	206
Results	207
Methods	230
References	237
Conclusions and perspectives	241
Annex: other projects	245
Curriculum vitae	249

Introduction

Chapter 1

The plasticity of genome evolution: insights from the budding yeast *Saccharomyces cerevisiae*

1.1) What is *Saccharomyces cerevisiae*?

During my PhD, I used *Saccharomyces cerevisiae* as a model organism to investigate two important aspects of genome evolution: the origin of interspecies introgressions and telomere evolution. *Saccharomyces cerevisiae* is a unicellular fungus commonly known as “baker’s yeast” or “brewer’s yeast”, due to its wide use as leavening agent in the production of bread and as fermenting agent in the production of a large variety of beverages (D’Angiolo, 2016). Its close association with human activities has a very ancient origin, as it has been demonstrated by the discovery of remainings of alcoholic beverages on the surface of pottery jars from the early Neolithic village of Jiahu in Henan province in China, dated to more than 7000 years ago (McGovern et al., 2004). *Saccharomyces cerevisiae* is also known with the common name of “budding yeast” because daughter cells spontaneously bud from the surface of the mother cell (**Figure 1**).

S. cerevisiae has been the first eukariotic organism to have its genome completely sequenced (Goffeau et al., 1996). It has long been considered as an exclusively human-associated organism and it was thought to exist only in anthropic environments, mainly associated to fermentation facilities, like breweries, distilleries and bakeries (Vaughan-Martini & Martini, 1995). However,

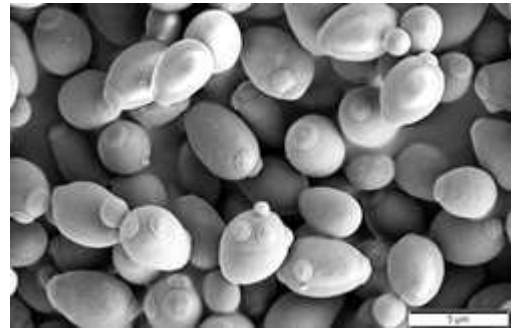


Figure 1: Cells of the budding yeast Saccharomyces cerevisiae.

this concept changed when *S. cerevisiae* was repeatedly isolated from many natural habitats (**Figure 2**) (Liti, 2015). Wild *Saccharomyces* yeasts are most often found in tree bark, soil and leaf surfaces (Boynton & Greig, 2014). The most frequent *Saccharomyces* hosts are oak trees (*Quercus* spp.) and other trees of the Fagaceae family, like chestnut, in the northern hemisphere, and southern beech trees (*Nothofagus* spp.) in the southern hemisphere (Sampaio & Gonçalves, 2008). Recently, *S. cerevisiae* has been isolated from tree bark of various native trees in Brazil, including *Tapirira guianensis* (Barbosa et al., 2016), and from insects (Stefanini & Dapporto, 2012). It can occasionally be isolated from human body niches, as an opportunistic pathogen from immuno-compromised patients, or as non-pathogenic colonizer. A surprising case of this type was recently reported from an Amerindian community living in a remote

area of French Guiana (Angebault et al., 2013).



Figure 2: Wild (top) and human-made (bottom) environments where *S. cerevisiae* has been isolated. From top-left to bottom right: *Quercus* tree, tree bark, *Nothofagus* tree, beer, wine, sake.

1.2) What is a genome?

As defined in the Chambers English dictionary, a genome is “the complete set of genetic material in the cell of a living organism”. The term derives from the German word “genom” and it was first coined in the 1920s by the German botanist Hans Winkler (1877-1945). It is thought that the word was a fusion of the terms “gene” and “chromosome”, which already existed at that time (Cristescu, 2019). Despite an initial delay, the usage of the word constantly increased with time (Figure 3) (Goldman & Landweber, 2016). Currently, the term does not only indicate the genetic material of a living organism, but also the complex and multilayered relations existing among its components. The genome of a contemporary individual does not only give us information about the present phenotype of that very individual, but it also keeps traces of the evolutionary steps that led that genome to its current form.

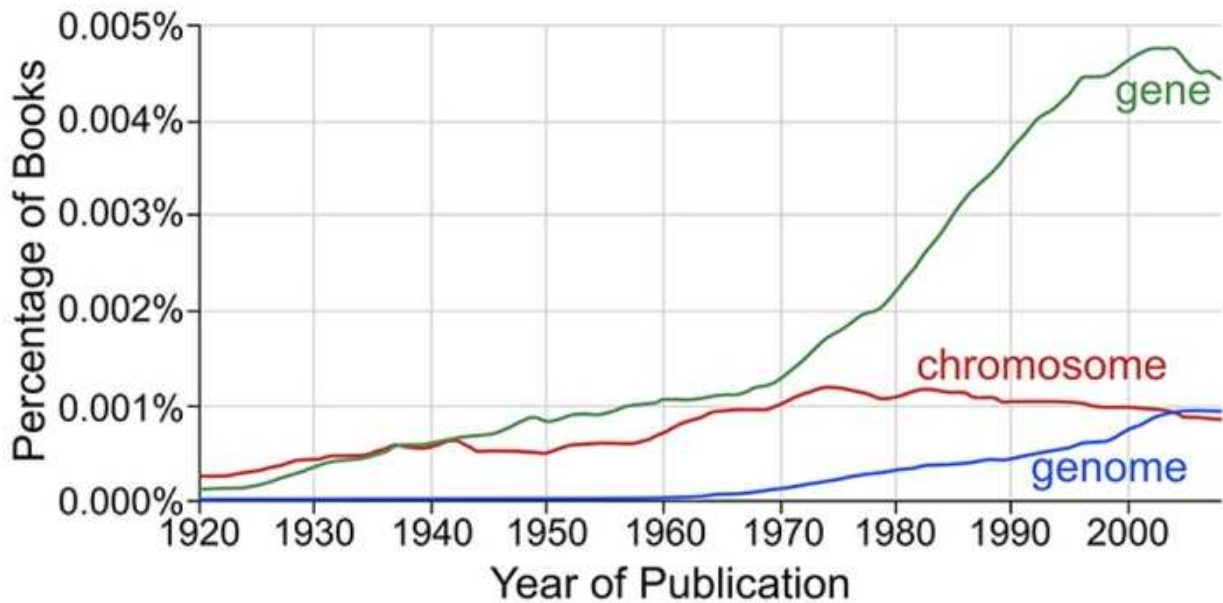


Figure 3: Percentage of English books, published between 1920 and 2008, in which the word "genome" occurs, compared to the other terms "gene" and "chromosome". Adapted from Goldman & Landweber, 2016.

Genomes existed and kept evolving since a very long time, creating the vast array of species that currently populate our planet and shaping the tree of life. The process of progressive modification of a genome during time is defined as “genome evolution”, and it can happen through multiple molecular mechanisms.

1.3) How do genomes evolve?

Genomes are the source of information that tells every individual how to function. They are composed by a long and unique combination of ATCG nucleotides that is different for every individual. The combination of nucleotides constitutes the DNA, which is in turn packed in linear molecules (the chromosomes) in eukaryotic organisms, and the chromosomes reside in the nucleus of the cell.

In humans, the complete genome sequence is only available since ~20 years and it was the result of a huge effort made by the Human Genome Sequencing Consortium and Celera Corporation that culminated with the publication of the complete sequence in 2001 (Lander et al., 2001; Venter et al., 2001). The size of the human genome in its haploid form is over 3 billion basepairs, organized in 23 pairs of double-stranded linear chromosomes. Among these 23 pairs, 22 are autosomes while 2 are the sex chromosomes X and Y.

1.3.1) Mutation

Modifications of one or a small number of nucleotides in an organism's genome are what

differentiates each individual from one another. When these modifications are rare in the population, e.g. present in less than 1% of the individuals, they are called “mutations”. On the contrary, when they are more common, e.g. present in more than 1% of the individuals, they are called “polymorphisms” (Karki et al., 2015). Every polymorphism initially manifested as a mutation, and spread into the population by means of genetic drift or positive selection. When mutations occur in multicellular organisms, like humans, their diffusion potential depends on the type of cell carrying the mutation: if they appear in somatic cells, mutations can only affect those specific cells and their derivatives. On the contrary, when they happen in the germ line, they can be passed on to the offspring and spread into the population.

Mutations can have variable impact on the fitness of their host depending on which position they occur (within genes or non-coding regions) and which specific nucleotide substitutions they involve. Missense (aminoacid substitutions) and non-sense (stop gain) mutations are thought to have a deeper effect on fitness, whereas synonymous mutations and mutations in non-coding regions are generally thought to have a lower impact. However, this is not always true as non-coding regions comprise regulatory elements, like promoters and enhancers, that have important functions for the cell. On the same line, synonymous mutations are not always neutral as they can induce the usage of rare tRNAs and therefore change the expression of a gene (She & Jarosz, 2018). The number of single nucleotide variants (SNV) between two individuals' genomes informs about the degree of their relatedness on an evolutionary scale. In fact, if we assume that mutations occur at a constant rate in time and selection is not acting on them, the number of pairwise nucleotide variations between two genomes is directly proportional to the chronological time passed since the individuals split from their last common ancestor (Fay & Benavides, 2005; Kimura, 1968; Rolland & Dujon, 2011). This principle has been applied to study the evolutionary history of multiple species, including humans and yeasts. Comparative genomics studies between the human genome (*Homo sapiens*) and those of the great apes revealed that chimpanzees (*Pan troglodytes*) and bonobos (*Pan paniscus*) are our closest living relatives, followed by gorillas and orangutans (**Figure 4**) (Wall, 2013).

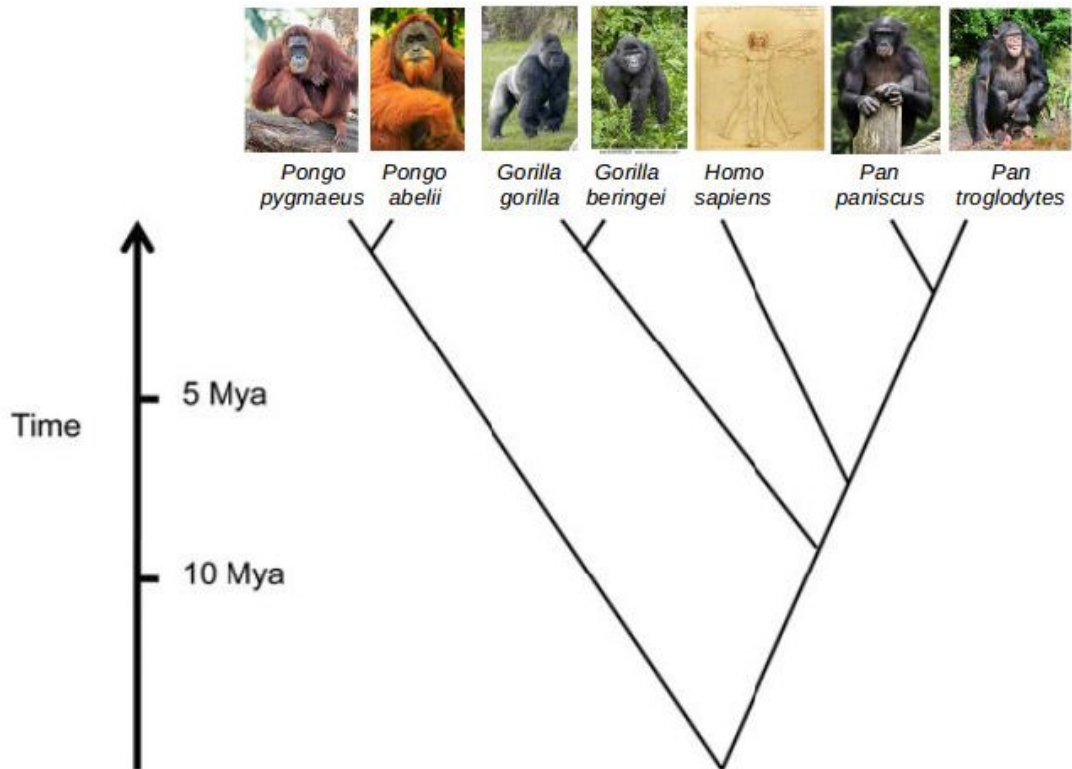


Figure 4: Phylogenetic tree showing the evolutionary relationships between humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), bonobos (*Pan paniscus*), gorillas (*Gorilla gorilla* and *Gorilla beringei*), orangutans (*Pongo pygmaeus* and *Pongo abelii*). Adapted from Wall, 2013.

If we compare orthologous genes, there is on average 1.2% sequence divergence between the human genome and that of chimpanzee or a bonobo. At a chronological level, humans and chimpanzees/bonobos diverged between 5 and 7 million years ago, while chimpanzees and bonobos diverged approximately two million years ago (Prüfer et al., 2012).

Saccharomyces cerevisiae is part of the *Saccharomyces sensu stricto* complex, which contains 8 species: *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *S. kudriavzevii*, *S. arboricola*, *S. eubayanus*, *S. uvarum* plus the recently discovered *S. jurei* (Figure 5) (Naseeb et al., 2017). Despite the similar morphology, *S. cerevisiae* has a sequence divergence of

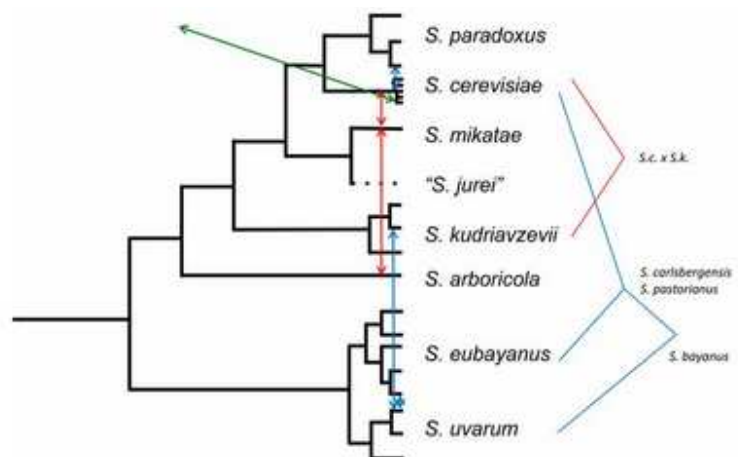


Figure 5: Phylogenetic tree showing evolutionary relationships among species of the *Saccharomyces sensu stricto* complex. Adapted from Dujon & Louis, 2017. Plain lines=hybrids. Arrows= gene flow among species.

approximately 15% with its phylogenetically closest relative *S. paradoxus*, and the divergence goes up to 30-35% with the more distant species of the complex: *S. bayanus* and *S. kudriavzevii* (Morales & Dujon, 2012). A recent phylogenetic study of the *Saccharomycotina* subphylum (comprising budding yeasts) placed the split between *S. cerevisiae* and its closest relative *S. paradoxus* to 4.9 million years ago (Shen et al., 2018).

1.3.2) Genome content variation

Every change in genome content that ranges from 50 base pairs to whole chromosomes is defined as genome content variation (GCV) (Steenwyk & Rokas, 2018). Presence/absence of single genes or groups of genes, together with their variation in copy number (CNV), are part of the GCV. Segmental duplications are also part of GCV. Aneuploidy and whole genome duplication are particular cases of GCV involving the gain or loss of an entire chromosome and the duplication of the entire genome, respectively. Ancient whole genome duplications have reshaped the genomes of yeasts, angiosperm plants and teleost fishes (Levasseur & Pontarotti, 2011; Wolfe & Shields, 1997; Wolfe, 2015). GCV can derive from non-disjunction events during mitotic and/or meiotic divisions, recombination between non-homologous sequences and chromosome breakage. Telomeric dysfunction can also be a powerful driver of GCV, as it generates breakage-fusion-bridge cycles that lead to genomic instability (O'Sullivan & Karlseder, 2010).

Generally, the phenotypic impact of GCV correlates positively with the size of the involved fragment, even though this is not always true and the consequences can vary depending on many environmental and genetic factors, like the ploidy of the organism and the relative importance of the involved gene or genome fragment in the environment where the organism lives. As an example, whole genome duplications never result in living humans (but can be present in cancer cells), although they occur in fungi and plants (Albertin et al., 2009; Albertin & Marullo, 2012; Otto & Whitton, 2000), and only three cases of trisomies of autosomes can arrive to birth: the trisomy of chromosome 13 causing the Patau syndrome, the trisomy of chromosome 18 causing the Edwards syndrome and the trisomy of chromosome 21, which is responsible for the onset of the Down syndrome. Among these, only the Down syndrome is compatible with life while the other two cause the death of the individual in a few weeks (Campbell & Eichler, 2013). Instead, GCVs involving single genes are frequently observed and can sometimes be useful in the adaptation of an organism to its living environment, as this often translates in a variation of abundance of the produced protein (Iskrow et al., 2012; Lye & Purugganan, 2019; Sicard & Legras, 2011; Steenwyk & Rokas, 2018). As an example, human populations with high-starch diets carry multiple copies of *AMY1*, encoding a salivary amylase, resulting in an improvement of their digestion (Perry et al., 2007; Steenwyk & Rokas, 2018).

GCV is more tolerated and more frequent at higher ploidies, where the presence of multiple copies of the same genome can buffer the loss of some of its parts (Todd et al., 2017). In addition, the duplication of genetic material is better tolerated than its loss. GCV is frequent in *S. cerevisiae*, where it happens more extensively in subtelomeric regions, which are rich in genes involved in the interaction with the external environment (Bergstrom et al., 2014; Dunn et al., 2012), and has been associated to adaptive functions. As an example, European and Sake lineages of *S. cerevisiae* acquired multiple copies of the gene *CUP1* (Strope et al., 2015; Warringer et al., 2011), encoding a copper-binding metallothionein which confers elevated tolerance to copper, and it has been hypothesized that this trait evolved to cope with the use of copper sulfate in vineyards (Almeida et al., 2015; Fay et al., 2004; Marsit & Dequin, 2015). Duplications in the *MAL* loci, regulating maltose utilization, have also been found in domesticated yeast strains (Duan et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016). Other gene copy number variants have been observed both in *S. cerevisiae* (Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk & Rokas, 2018; Strope et al., 2015) and other species, including *S. pombe* (Jeffares et al., 2015).

As an example, European and Sake lineages of *S. cerevisiae* acquired multiple copies of the gene *CUP1* (Strope et al., 2015; Warringer et al., 2011), encoding a copper-binding metallothionein which confers elevated tolerance to copper, and it has been hypothesized that this trait evolved to cope with the use of copper sulfate in vineyards (Almeida et al., 2015; Fay et al., 2004; Marsit & Dequin, 2015). Duplications in the *MAL* loci, regulating maltose utilization, have also been found in domesticated yeast strains (Duan et al., 2018; Gallone et al., 2016; Gonçalves et al., 2016). Other gene copy number variants have been observed both in *S. cerevisiae* (Gallone et al., 2016; Gonçalves et al., 2016; Steenwyk & Rokas, 2018; Strope et al., 2015) and other species, including *S. pombe* (Jeffares et al., 2015).

1.3.3) Balanced structural variation

Balanced structural variation indicates every variation in the genome structure that does not involve change in copy number of genes or chromosome segments. Examples of this type of variation are inversions of a genetic fragment or translocations of a stretch of DNA within or between chromosomes. Structural variations (SV) are usually mediated by repetitive sequences, like Ty elements in yeast (Rachidi et al., 1999).

Structural genomic variants (SVs) are ubiquitous and play a major role in adaptation and speciation. This can happen through the generation of chimeric genes with novel functions (Rogers et al., 2009), which can be either adaptive or lead to the onset of disease, as in the famous case of the Philadelphia chromosome which causes leukemia through a fusion gene between *BCR* and *ABL1* (Heisterkamp et al., 1985). Rearrangements can also associate genes with stronger or weaker promoters than their native ones and change their expression level (Stewart & Rogers, 2019). Two

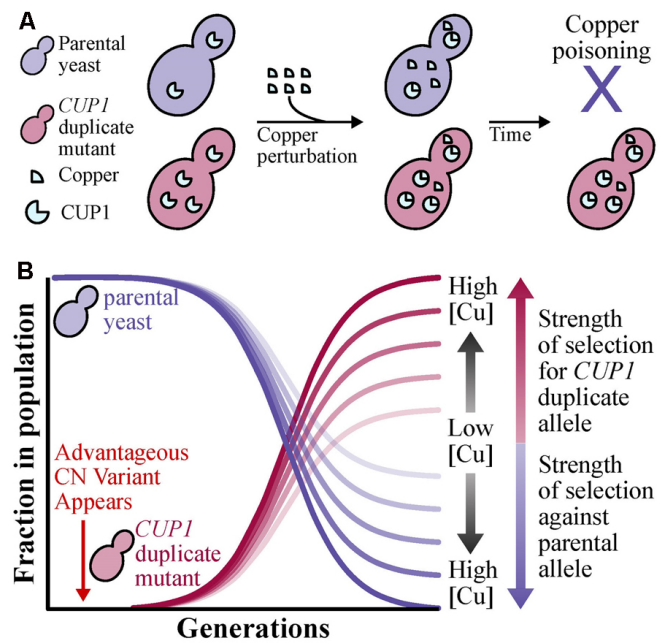


Figure 6: Duplication of *CUP1* gene helps adaptation to the environment. Adapted from Steenwyk & Rokas, 2018.

examples of ectopic recombination between chromosomes VIII and XVI (García-Ríos et al., 2019; Pérez-Ortín et al., 2002), and chromosomes XV and XVI (Zimmer et al., 2014), in wine yeasts, result in the gain of a stronger promoter by the gene *SSU1*, encoding a sulfite pump, which leads to a better survival during wine fermentation in the presence of sulfites (**Figure 7**).

SVs have been important for the evolution of the human genome and frequently occur at telomeric sites: all great apes have a diploid genome content of $n=48$ (Yunis & Prakash, 1982), while humans contain 46 chromosomes. Genetic studies evidenced that the human chromosome II derived from an archaic telomeric

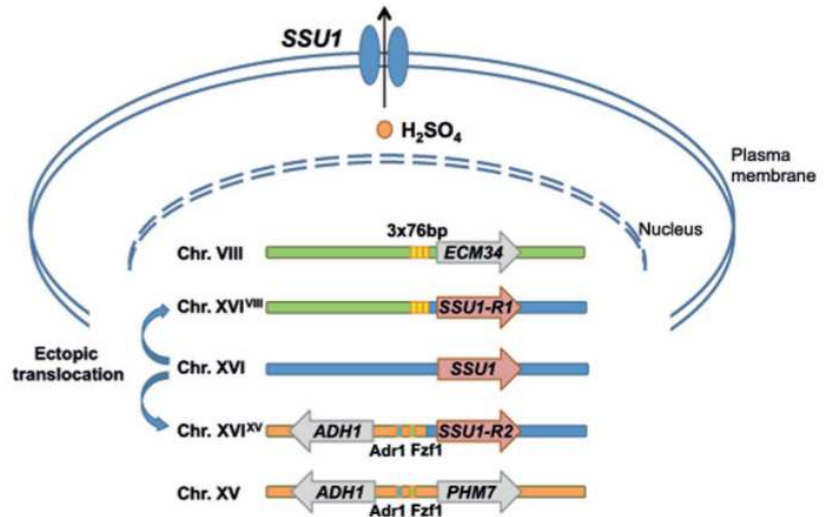


Figure 7: Mechanisms of sulfite resistance through reciprocal translocations. Adapted from Marsit & Dequin, 2015.

fusion of two autosomes, decreasing the global chromosome count of two units (Dreszer et al., 2007; Ijdo et al., 1991; Ventura et al., 2012).

SVs interfere with recombination and result in low gamete viability, due to the inappropriate pairing of homologous chromosomes that can trigger the onset of aneuploidies, chromosome loss and segmental duplications in the offspring. As a consequence, SVs are often at the origin of incipient speciation as they promote reproductive isolation between populations. For example, SVs led to the rapid emergence of reproductive isolation of a *Saccharomyces cerevisiae* subpopulation isolated from the nectar of the bertam palm in Malaysia (Liti, 2015; Naumov et al., 2006) that carries multiple chromosomal rearrangements (Marie-Nelly et al., 2014; Yue et al., 2017). The same phenomenon is observed in the previously designated *S. cariocanus* species, which is in fact a *S. paradoxus* South American subpopulation carrying multiple SVs (**Figure 8**) (Naumov et al., 2000; Yue et al., 2017).

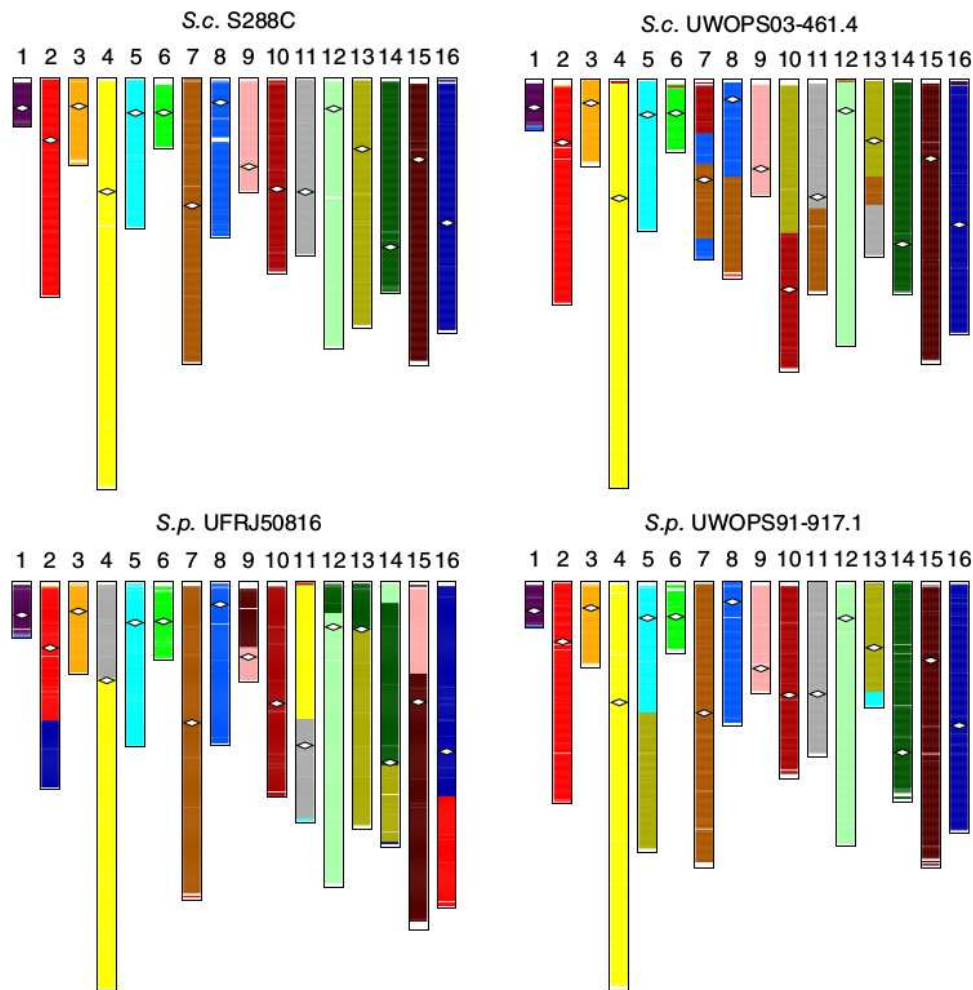


Figure 8: Upper panel: karyotype of the *S. cerevisiae* Malaysian strain (UWOPS034614) compared to the laboratory strain (S288C). Bottom panel: karyotypes of the *S. paradoxus* South-American (UFRJ50816) and Hawaiian (UWOPS919171) strains. Adapted from Yue et al., 2017.

1.3.4) Recombination

Recombination is a pervasive source of genome evolution and the main driver of genetic variability in sexually reproducing organisms. It is a fundamental step of meiosis, where it guarantees the exchange of genetic material between chromosomes and the production of gametes with novel allelic combinations. Meiotic recombination does not always occur randomly in the genome. It has been shown that, in some genomic regions, it happens more frequently than expected by chance. Therefore these regions constitute recombination hot-spots. (Mancera et al., 2008). Recombination does not only occur in meiosis, where it involves homologous chromosomes, but it can also happen during mitotic divisions, although at low frequency (Lafave & Sekelsky, 2009). In this case it involves sister chromatids instead of homologous chromosomes.

Novel allelic combinations can either confer advantages to the individuals who carry them, and help them adapt to the environment (Chu et al., 2017), or they can be detrimental and lead to the death of the individual, as in the case of Dobzhansky-Muller genetic incompatibilities, which are especially

relevant in the evolution of hybrids (Cutter, 2012). Even though recombination risks to produce deleterious combinations, it is globally advantageous (Kondrashov, 1988; Mortimer, 2000), and this is supported by the observation that genomic regions refractory to recombination experience a higher mutational load and eventually deteriorate (Charlesworth & Charlesworth, 2000; Webster & Hurst, 2012). In yeast, adaptive evolution experiments have shown that strains capable of sex have a higher potential of adaptation to harsh environment than asexual ones (Goddard et al., 2005). This is also supported by the observation that wild *S. cerevisiae* isolates retained the ability to recombine their genomes and sporulate, allowing them to survive in nutrient scarcity. On the contrary, domesticated strains lost their sporulation capacity resulting in a massive decline in their capacity to survive nutrient depletion (De Chiara et al., 2020).

1.3.5) Other mechanisms of genome evolution

Other powerful sources of genome variation and drivers of genome evolution and speciation are hybridizations, introgressions and telomere diversity. The contribution of hybridizations and introgressions to genome evolution will be treated in chapter 2 of the introduction. The evolution of telomeric sequences and their impact on the evolution of the rest of the genome will be reviewed in chapter 3 of the introduction.

1.3.6) Environmental factors and selection

All the molecular mechanisms explained above have a huge impact on the evolution of the genome. However, their outcome is shaped mainly by their phenotypic impact and how the resulting phenotype of the carrying organism fits with the environmental condition in which it lives (Chakravarti, 1999). If the genomic changes confer some fitness advantage and make the organisms more fit to their habitat, these organisms will have higher offspring than the rest and in the long term the changes will increase in frequency in the population, possibly reaching fixation (positive selection). On the contrary, if the genomic changes have detrimental effect they will soon be purged from the population (negative selection).

The probability of new genetic variants to increase their frequency in a population and eventually reach fixation is also influenced by other parameters. Genetic drift plays an important role in modulating the destiny of new mutations. In a given population of n individuals, not all of them will reproduce and generate offspring. If only a few individuals contribute genetically to the next generation and those very individuals happen to carry a new genetic variant, that variant will increment its frequency in the next generation just by chance, even if it is phenotypically neutral or even deleterious. As a consequence, population size has a huge effect on the probability of new mutations to reach fixation, and as a general rule small populations have a higher chance of fixing

new mutations just by genetic drift.

In conclusion, the evolution of a genome is ultimately the result of a complex network of interactions between the molecular mechanisms generating genomic variations, their effect on the individuals, the interaction of the individuals with their natural environment...and chance.

References

- Albertin, W., Marullo, P., Aigle, M., Bourgeois, A., Bely, M., Dillmann, C., De Vienne, D., & Sicard, D. (2009). Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: Towards a new domesticated species. *Journal of Evolutionary Biology*, 22(11), 2157–2170. <https://doi.org/10.1111/j.1420-9101.2009.01828.x>
- Albertin, Warren, & Marullo, P. (2012). Polyploidy in fungi: Evolution after whole-genome duplication. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 279, Issue 1738, pp. 2497–2509). Royal Society of London. <https://doi.org/10.1098/rspb.2012.0434>
- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J. L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, 24(21), 5412–5427. <https://doi.org/10.1111/mec.13341>
- Angebault, C., Djossou, F., Abélanet, S., Permal, E., Soltana, M. Ben, Diancourt, L., Bouchier, C., Woerther, P. L., Catzeflis, F., Andremont, A., D'enfert, C., & Bounoux, M. E. (2013). *Candida albicans* is not always the preferential yeast colonizing humans: A study inwayampi amerindians. *Journal of Infectious Diseases*, 208(10), 1705–1716. <https://doi.org/10.1093/infdis/jit389>
- Aronin, N., DiFiglia, M., Liotta, A., Martin, J. B., Glazer, E. J., Basbaum, A. I., Carstens, E., Trevino, D. L., Comp, J., Willis, W. D., Kenshalo, D. R., Leon, R. B., Burton, H., Craig, A. D., Maunz, R. A., Bryan, R. N., Neurol, E., Coulter, J. D., Price, D. D., ... Hoffert, M. (1980). The Origin of Man: A Chromosomal Pictorial Legacy. In *Soc. Neurosci. Abstr* (Vol. 6, Issue 2). Wiley.
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, evv263. <https://doi.org/10.1093/gbe/evv263>
- Bergstrom, A., Simpson, J. T., Salinas, F., Barr??, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, 31(4), 872–888. <https://doi.org/10.1093/molbev/msu037>
- Boynton P. (2014). The ecology and evolution of non-domesticated *Saccharomyces* species. *Yeast*, 31, 449–462.
- Campbell, C. D., & Eichler, E. E. (2013). Properties and rates of germline mutations in humans. In *Trends in Genetics* (Vol. 29, Issue 10, pp. 575–584). <https://doi.org/10.1016/j.tig.2013.04.005>
- Chakravarti (1999). Population genetics – Making sense out of sequence. *Nature Genetics Supplement*. Volume 21, 56-60.
- Charlesworth, B., & Charlesworth, D. (2000). The degeneration of Y chromosomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 355(1403), 1563–1572.

<https://doi.org/10.1098/rstb.2000.0717>

- Chu, H. Y., Sprouffske, K., & Wagner, A. (2017). The role of recombination in evolutionary adaptation of *Escherichia coli* to a novel nutrient. *Journal of Evolutionary Biology*, *30*(9), 1692–1711. <https://doi.org/10.1111/jeb.13132>
- Cristescu, M. E. (2019). The concept of genome after one century of usage. In *Genome* (Vol. 62, Issue 10, pp. iii–v). Canadian Science Publishing. <https://doi.org/10.1139/gen-2019-0129>
- Cutter, A. D. (2012). The polymorphic prelude to Bateson-Dobzhansky-Muller incompatibilities. *Trends in Ecology and Evolution*, *27*(4), 209–218. <https://doi.org/10.1016/j.tree.2011.11.004>
- D'Angiolo, M. (2016). <https://etd.adm.unipi.it/t/etd-06222016-153506/>. Master's Degree Thesis.
- David Goldman, A., & Landweber, L. F. (2016). *What Is a Genome?* <https://doi.org/10.1371/journal.pgen.1006181.g001>
- De Chiara, M., Barré, B., Persson, K., Chioma, A. O., Irizar, A., Schacherer, J., Warringer, J., & Liti, G. (2020). *Domestication reprogrammed the budding yeast life cycle*. 1–28. <https://doi.org/10.1101/2020.02.08.939314>
- Dreszer, T. R., Wall, G. D., Haussler, D., & Pollard, K. S. (2007). Biased clustered substitutions in the human genome: The footprints of male-driven biased gene conversion. *Genome Research*, *17*(10), 1420–1430. <https://doi.org/10.1101/gr.6395807>
- Duan, S. F., Han, P. J., Wang, Q. M., Liu, W. Q., Shi, J. Y., Li, K., Zhang, X. L., & Bai, F. Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, *9*(1). <https://doi.org/10.1038/s41467-018-05106-7>
- Dunn, B., Richter, C., Kvitek, D. J., Pugh, T., & Sherlock, G. (2012). Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Research*, *22*(5), 908–924. <https://doi.org/10.1101/gr.130310.111>
- Fay, J. C., & Benavides, J. A. (2005). Evidence for domesticated and wild populations of *saccharomyces cerevisiae*. *PLoS Genetics*, *1*(1), 0066–0071. <https://doi.org/10.1371/journal.pgen.0010005>
- Fay, J. C., McCullough, H. L., Sniegowski, P. D., & Eisen, M. B. (2004). Population genetic variation in gene expression is associated with phenotypic variation in *Saccharomyces cerevisiae*. *Genome Biology*, *5*(4), 1–14. <https://doi.org/10.1186/gb-2004-5-4-r26>
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, *166*(6), 1397-1410.e16. <https://doi.org/10.1016/j.cell.2016.08.020>
- García-Ríos, E., Nuévalos, M., Barrio, E., Puig, S., & Guillamón, J. M. (2019). A new chromosomal rearrangement improves the adaptation of wine yeasts to sulfite. *Environmental Microbiology*, *21*(5), 1771–1781. <https://doi.org/10.1111/1462-2920.14586>

- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287), 546, 563–567. <http://www.ncbi.nlm.nih.gov/pubmed/8849441>
- Gonçalves, M., Pontes, A., Almeida, P., Barbosa, R., Serra, M., Libkind, D., Hutzler, M., Gonçalves, P., & Sampaio, J. P. (2016). Distinct Domestication Trajectories in Top-Fermenting Beer Yeasts and Wine Yeasts. *Current Biology*, 26(20), 2750–2761. <https://doi.org/10.1016/j.cub.2016.08.040>
- Heisterkamp N., Stam K., Groffen J. (1985). Structural organization of the bcr gene and its role in the Ph' translocation. *Nature*. Volume 315, 758-761.
- Ijdo, J. W., Baldinit, A., Wardt, D. C., Reeders, S. T., & Wells, R. A. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. In *Proc. Nadl. Acad. Sci. USA* (Vol. 88).
- Iskow, R. C., Gokcumen, O., & Lee, C. (2012). Exploring the role of copy number variants in human adaptation. In *Trends in Genetics* (Vol. 28, Issue 6, pp. 245–257). <https://doi.org/10.1016/j.tig.2012.03.002>
- Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Převorovský, M., Mourier, T., Marsellach, F. X., Iqbal, Z., Lau, W., Cheng, T. M. K., Pracana, R., Müllleder, M., Lawson, J. L. D., Chessel, A., Bala, S., Hellenthal, G., O'Fallon, B., Keane, T., Simpson, J. T., ... Bähler, J. (2015). The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nature Genetics*, 47(3), 235–241. <https://doi.org/10.1038/ng.3215>
- Karki, R., Pandya, D., Elston, R. C., & Ferlini, C. (2015). Defining “mutation” and “polymorphism” in the era of personal genomics. In *BMC Medical Genomics* (Vol. 8, Issue 1). BioMed Central Ltd. <https://doi.org/10.1186/s12920-015-0115-z>
- Kondrashov, A. S. (1988). Deleterious mutations and the evolution of sexual reproduction. *Nature*. Volume 336, 435-440.
- Lafave, M. C., & Sekelsky, J. (2009). Mitotic recombination: why? when? how? where? In *PLoS Genetics* (Vol. 5, Issue 3). <https://doi.org/10.1371/journal.pgen.1000411>
- Lander, S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., ... Yeh, R.-F. (2001). Initial sequencing and analysis of the human genome International Human Genome Sequencing Consortium* The Sanger Centre: Beijing Genomics Institute/Human Genome Center. In *NATURE* (Vol. 409). www.nature.com
- Levasseur, A., & Pontarotti, P. (2011). The role of duplications in the evolution of genomes highlights the need for evolutionary-based approaches in comparative genomics. In *Biology Direct* (Vol. 6). <https://doi.org/10.1186/1745-6150-6-11>
- Liti, G. (2015). The fascinating and secret wild life of the budding yeast *S. cerevisiae*. *ELife*, 4, e05835. <https://doi.org/10.7554/eLife.05835>

- Lye, Z. N., & Purugganan, M. D. (2019). Copy Number Variation in Domestication. In *Trends in Plant Science* (Vol. 24, Issue 4, pp. 352–365). Elsevier Ltd. <https://doi.org/10.1016/j.tplants.2019.01.003>
- M. Kimura. (1968). Evolutionary Rate at the Molecular Level. *Nature*, 217, 624–626. <https://www-nature-com.remote.library.osaka-u.ac.jp:8443/articles/217624a0.pdf>
- Mancera, E., Bourgon, R., Brozzi, A., Huber, W., & Steinmetz, L. M. (2008). High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature*, 454(7203), 479–485. <https://doi.org/10.1038/nature07135>
- Marie-Nelly, H., Marbouty, M., Cournac, A., Flot, J.-F., Liti, G., Parodi, D. P., Syan, S., Guillén, N., Margeot, A., Zimmer, C., & Koszul, R. (2014). High-quality genome (re)assembly using chromosomal contact data. *Nature Communications*, 5, 5695. <https://doi.org/10.1038/ncomms6695>
- Marsit, S., & Dequin, S. (2015). Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. *FEMS Yeast Research*, 15(7), 67. <https://doi.org/10.1093/femsyr/fov067i>
- McGovern, P. E., Zhang, J., Tang, J., Zhang, Z., Hall, G. R., Moreau, R. A., Nuñez, A., Butrym, E. D., Richards, M. P., Wang, C.-S., Cheng, G., Zhao, Z., & Wang, C. (2004). Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51), 17593–17598. <https://doi.org/10.1073/pnas.0407921102>
- Morales, L., & Dujon, B. (2012). Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and Molecular Biology Reviews: MMBR*, 76(4), 721–739. <https://doi.org/10.1128/MMBR.00022-12>
- Mortimer, R. K. (2000). Evolution and variation of the yeast (*Saccharomyces*) genome. *Genome Research*, 10(4), 403–409. <https://doi.org/10.1101/gr.10.4.403>
- Naseeb, S., James, S. A., Alsammar, H., Michaels, C. J., Gini, B., Nueno-Palop, C., Bond, C. J., McGhie, H., Roberts, I. N., & Delneri, D. (2017). *Saccharomyces jurei* sp. Nov., isolation and genetic identification of a novel yeast species from *Quercus robur*. *International Journal of Systematic and Evolutionary Microbiology*, 67(6), 2046–2052. <https://doi.org/10.1099/ijsem.0.002013>
- Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J., & Roberts, I. N. (2000). Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International Journal of Systematic and Evolutionary Microbiology*, 50(5), 1931–1942. <https://doi.org/10.1099/00207713-50-5-1931>
- Naumov, G. I., Serpova, E. V., & Naumova, E. S. (2006). A genetically isolated population of *saccharomyces cerevisiae* in Malaysia. *Microbiology*, 75(2), 201–205. <https://doi.org/10.1134/S0026261706020147>
- O’Sullivan, R. J., & Karlseder, J. (2010). Telomeres: Protecting chromosomes against genome instability. *Nature Reviews Molecular Cell Biology*, 11(3), 171–181.

<https://doi.org/10.1038/nrm2848>

- Otto, S. P., & Whitton, J. (2000). Polyploid incidence and evolution. *Annu Rev Genet* 2000;34:401-437. doi: 10.1146/annurev.genet.34.1.401.
- Pérez-Ortín, J. E., Querol, A., Puig, S., & Barrio, E. (2002). Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Research*, 12(10), 1533–1539. <https://doi.org/10.1101/gr.436602>
- Perry, G. H., Dominy, N. J., Claw, K. G., Lee, A. S., Fiegler, H., Redon, R., Werner, J., Villanea, F. A., Mountain, J. L., Misra, R., Carter, N. P., Lee, C., & Stone, A. C. (2007). Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*, 39(10), 1256–1260. <https://doi.org/10.1038/ng2123>
- Prüfer, K., Munch, K., Hellmann, I., Akagi, K., Miller, J. R., Walenz, B., Koren, S., Sutton, G., Kodira, C., Winer, R., Knight, J. R., Mullikin, J. C., Meader, S. J., Ponting, C. P., Lunter, G., Higashino, S., Hobolth, A., Dutheil, J., Karakoç, E., ... Pääbo, S. (2012). The bonobo genome compared with the chimpanzee and human genomes. *Nature*, 486(7404), 527–531. <https://doi.org/10.1038/nature11128>
- Rachidi, N., Barre, P., & Blondin, B. (1999). Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. *Molecular and General Genetics*, 261(4–5), 841–850. <https://doi.org/10.1007/s004380050028>
- Rogers, R. L., Bedford, T., & Hart, D. L. (2009). Formation and longevity of chimeric and duplicate genes in *Drosophila melanogaster*. *Genetics*, 181(1), 313–322. <https://doi.org/10.1534/genetics.108.091538>
- Rolland, T., & Dujon, B. (2011). Yeasty clocks: Dating genomic changes in yeasts. *Comptes Rendus - Biologies*, 334(8–9), 620–628. <https://doi.org/10.1016/j.crvi.2011.05.010>
- Sampaio, J. P., & Gonçalves, P. (2008). Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with Oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*. *Applied and Environmental Microbiology*, 74(7), 2144–2152. <https://doi.org/10.1128/AEM.02396-07>
- She, R., & Jarosz, D. F. (2018). Mapping Causal Variants with Single-Nucleotide Resolution Reveals Biochemical Drivers of Phenotypic Change. *Cell*, 172(3), 478-490.e15. <https://doi.org/10.1016/j.cell.2017.12.015>
- Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Ichiroh, Čadež, N., Libkind, D., ... Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, 175(6), 1533-1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- Sicard, D., & Legras, J. L. (2011). Bread, beer and wine: Yeast domestication in the *Saccharomyces sensu stricto* complex. In *Comptes Rendus - Biologies* (Vol. 334, Issue 3, pp. 229–236). Elsevier Masson SAS. <https://doi.org/10.1016/j.crvi.2010.12.016>
- Steenwyk, J. L., & Rokas, A. (2018). Copy number variation in fungi and its implications for wine

- yeast genetic diversity and adaptation. In *Frontiers in Microbiology* (Vol. 9, Issue FEB). Frontiers Media S.A. <https://doi.org/10.3389/fmicb.2018.00288>
- Stefanini, I., & Dapporto, L. (2012). Role of social wasps in *Saccharomyces cerevisiae* ecology and evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(33), 13398–13403. <https://doi.org/10.1073/pnas.1208362109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1208362109>
- Stewart, N. B., & Rogers, R. L. (2019). Chromosomal rearrangements as a source of new gene formation in *Drosophila yakuba*. *PLoS Genetics*, *15*(9). <https://doi.org/10.1371/journal.pgen.1008314>
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, *125*(5), 762–774. <https://doi.org/10.1101/gr.185538.114>
- Structural and temporal requirements for geomagnetic field reversal deduced from lava flows. (2005). *Nature*, *434*(7033), 633–636. <https://doi.org/10.1038/nature03431>
- Todd. (2017). Ploidy Variation in Fungi: Polyploidy, Aneuploidy, and Genome Evolution. In *The Fungal Kingdom* (pp. 599–618). American Society of Microbiology. <https://doi.org/10.1128/microbiolspec.funk-0051-2016>
- Vaughan-Martini, A., & Martini, A. (1995). Facts, myths and legends on the prime industrial microorganism. *Journal of Industrial Microbiology*, *14*(6), 514–522. <https://doi.org/10.1007/BF01573967>
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., ... Zhu, X. (2001). *The Sequence of the Human Genome Downloaded from* (Vol. 291). <http://science.sciencemag.org/>
- Ventura, M., Catacchio, C. R., Sajjadian, S., Vives, L., Sudmant, P. H., Marques-Bonet, T., Graves, T. A., Wilson, R. K., & Eichler, E. E. (2012). The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Research*, *22*(6), 1036–1049. <https://doi.org/10.1101/gr.136556.111>
- Wall, J. D. (2013). Great ape genomics. *ILAR Journal*, *54*(2), 82–90. <https://doi.org/10.1093/ilar/ilt048>
- Warringer, J., Z??rg??, E., Cubillos, F. A., Zia, A., Gjuvsland, A., Simpson, J. T., Forsmark, A., Durbin, R., Omholt, S. W., Louis, E. J., Liti, G., Moses, A., & Blomberg, A. (2011). Trait variation in yeast is defined by population history. *PLoS Genetics*, *7*(6). <https://doi.org/10.1371/journal.pgen.1002111>
- Webster, M. T., & Hurst, L. D. (2012). Direct and indirect consequences of meiotic recombination: Implications for genome evolution. In *Trends in Genetics* (Vol. 28, Issue 3, pp. 101–109).

<https://doi.org/10.1016/j.tig.2011.11.002>

- Wolfe, K H, & Shields, D. C. (1997). Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, 387(6634), 708–713. <https://doi.org/10.1038/42711>
- Wolfe, Kenneth H. (2015). Origin of the yeast whole-genome duplication. *PLoS Biology*, 13(8), 1–7. <https://doi.org/10.1371/journal.pbio.1002221>
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Zimmer, A., Durand, C., Loira, N., Durrens, P., Sherman, D. J., & Marullo, P. (2014). QTL dissection of lag phase in wine fermentation reveals a new translocation responsible for *Saccharomyces cerevisiae* adaptation to sulfite. *PLoS ONE*, 9(1), 37–39. <https://doi.org/10.1371/journal.pone.0086298>

Chapter 2

The role of hybridization and introgression in genome evolution

2.1) What is an introgression?

An introgression is the flow of genetic material between populations or species. The term was coined by Edgar Anderson and Leslie Hubricht during their studies on the hybridization of spiderworts (Dunn et al., 2013; Suarez-Gonzalez et al., 2018). Genome introgressions are a pervasive source of genome evolution across the tree of life, with examples in the animal (Edelman et al., 2019), plant (Arnold et al., 2016) and fungal (Sun et al., 2012) kingdoms. Their importance in the processes of adaptation and genome evolution has been increasingly studied throughout the 20th century and more recently (Heliconius Genome Consortium, 2012; Edelman et al., 2019; Enard & Petrov, 2018; Huerta-Sánchez et al., 2014; Prüfer et al., 2014). Introgressions are thought to initiate from ancient hybridization events followed by repeated backcrossing to one parental population/species. They are characterized by allelic replacement, meaning that the genetic fragment which is transferred from one species/population to another will not insert in a random part of the recipient genome, but it will replace its homologous sequence.

Introgressions can often distort our perception of the evolutionary relationships among species.

These relationships are usually represented by bifurcating trees, where the branching nodes represent speciation events (Nichols, 2001). Species are thought to diverge from their last common ancestor by gradually accumulating genetic mutations that subsequently translate in morphological differences. The absence of sexual admixture between the two branching species guarantees that the genetic differences they accumulate are maintained separated. As a consequence, the comparison of gene sequences from different species should give a correct inference of their true evolutionary history, assuming that the rate of admixture between them is very low or even absent.

However, hybridization between related species occurs frequently in nature, affecting around 25%

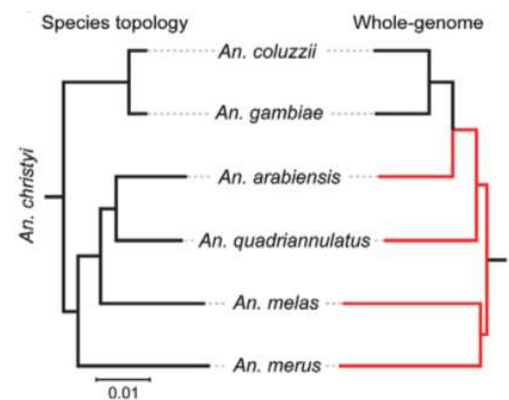


Figure 1: “Whole genome” versus “species” tree topologies of the *Anopheles gambiae* complex in Africa. The “species tree” derives from analysis of the X chromosome only, which is refractory to inter-species gene flow. Adapted from Mallet et al., 2016, Fontaine et al., 2015.

of flowering plants and about 10% of animals (Mallet et al., 2016), and this can conduct to conflicting phylogenies. An example of this phenomenon can be observed in the mosquitoes of the *Anopheles gambiae* complex, comprising 8 species which differentiated within the last two million years. In the phylogeny derived from the comparison of the whole genome, *A. arabiensis* is evolutionarily closer to *A. coluzzii* and *A. gambiae*, while in reality its true sister species is *A. quadriannulatus* (**Figure 1**) (Fontaine et al., 2015). Therefore, the true species history might be better represented by reticulated phylogenies which take into account the timing and frequency of admixture between lineages, rather than by simple bifurcating ones. This concept is known as “reticulated evolution”.

2.2) The extent of introgressions across the tree of life

Introgressions are widespread across the eukaryotic tree of life and have been extensively documented in scientific literature. Several examples of introgressions can be found in insects, plants, mammals and even among hominins.

2.2.1) Introgressions in insects

The genus *Heliconius* comprises butterfly species with beautiful and very diverse wing colour patterns, which live in Central and South-America. Their colour pattern serves as a deterrent for predators and evolved to mimic that of unpalatable species (Brower, 2013). *Heliconius* species often live in partial sympatry across South-America and can be recognized by their colour pattern, although unrelated species can occasionally show similar wing characteristics. Even

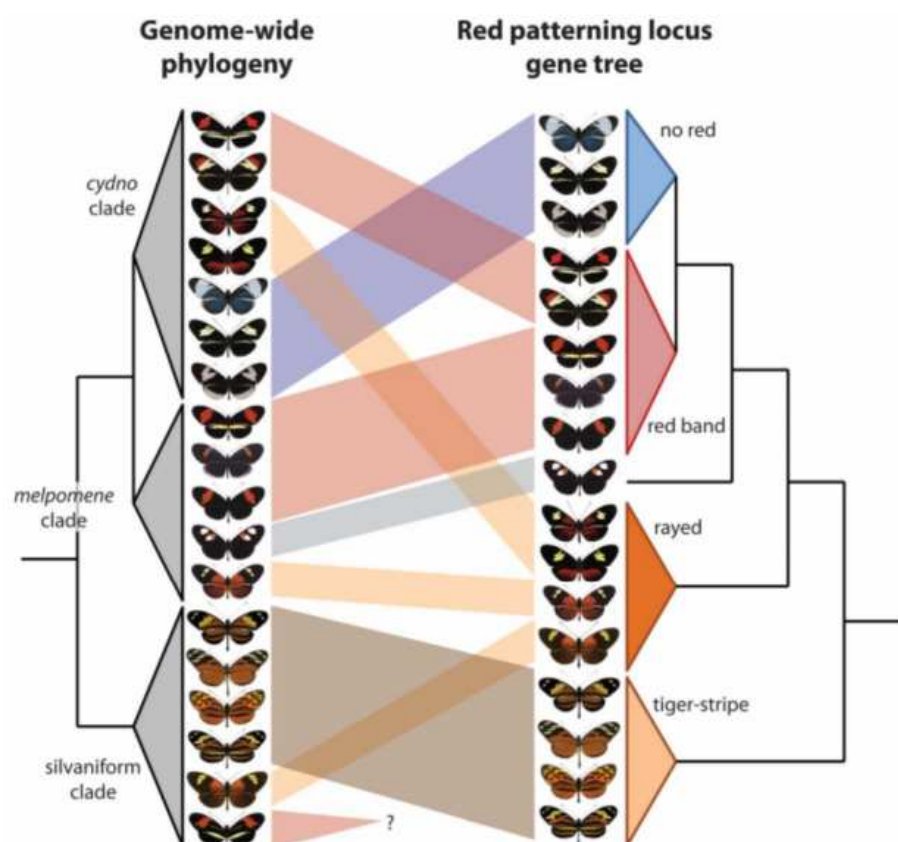


Figure 2: “Whole genome” versus “red-patterning locus” tree topologies of *Heliconius* butterflies. Adapted from Kronforst, 2012.

colour pattern, although unrelated species can occasionally show similar wing characteristics. Even

though these species show assortative mating based on their colours, cases of hybridization have been repeatedly isolated in the wild. Recently, it has been shown that similar colour patterns can result from introgression (Heliconius Genome Consortium, 2012; Kronforst, 2012; Martin et al., 2018; Pardo-Diaz et al., 2012; Zhang et al., 2016). A genomic region surrounding the gene *optix*, responsible for the presence of red spots on the wings, is frequently exchanged among species, and this is demonstrated by the fact the a phylogenetic tree built upon this region is different from the true species phylogeny and clusters the species based on their colour patterns (**Figure 2**).

2.2.2) Introgressions in plants

Introgression of adaptive genetic loci has been documented also in multiple plant species. For example, the flood-tolerant flower species *Iris fulva* donated genetic material to the dry-adapted *Iris brevicaulis*, and these regions conferred the ability to survive extreme flooding conditions (Martin et al., 2006). On the same hand, introgressions have been detected between the tree sister species *Populus trichocarpa* and *P. balsamifera* (**Figure 3**). *P. trichocarpa* is distributed throughout the western United States and Canada, and is adapted to humid and moist conditions. *P. balsamifera*, instead, has a more boreal distribution and is dry- and frost-resistant. Introgressed genes from *P. balsamifera* into *P. trichocarpa* are more frequent in individuals inhabiting the northern range of *P. trichocarpa* geographical distribution, and confer resistance to lower temperatures and drier habitats (Geraldes et al., 2014; Suarez-Gonzalez et al., 2016).



Figure 3: From top-left to bottom-right: *Iris fulva*, *Iris brevicaulis*, *Populus balsamifera*, *Populus trichocarpa*.

2.2.3) Introgressions in mammals

Introgression is frequent and has been documented extensively also in animals. One interesting case of introgression driving ecological adaptation can be observed in snowshoe hares (*Lepus americanus*). Snowshoe hares change their coat during the passage from autumn to winter. Hares living in zones with high snow cover change to a white colour coat, while the ones living in warmer regions maintain a brown coat. The colour of the coat is controlled by the *Agouti* gene in a Mendelian way, with the allele conferring a white colour being dominant. The promoter region of *Agouti* is introgressed from the black-tailed jackrabbit in individuals inhabiting warmer regions, and this causes a decrease in expression of the *Agouti* gene, resulting in a brown coat (Jones et al., 2018). To further support that the *cis*-regulatory region derives from an introgression, the topology

of the snowhares phylogeny constructed upon genome-wide variation and upon the local *Agouti* region are different, with the *Agouti* topology clustering individuals based on their coat colour rather than established evolutionary relationships (**Figure 4**).

The ones described here are just a limited set of examples that I found particularly interesting. However, for a complete review of cases of adaptive introgressions refer to (Hedrick, 2013; Taylor & Larson, 2019).

2.3) The impact of introgressions in the evolution of the human genome

2.3.1) Distribution of archaic introgressions in human populations

Currently, the only hominid species populating the planet is *Homo sapiens*, that is us. However, back in the past other species of the

genus *Homo* inhabited the earth and co-existed with anatomically modern humans, like Neanderthals and Denisovans (Higham et al., 2014; Hublin et al., 2017). The lack of genomic data initially prevented to address the biological question if ancient hominins ever admixed. The advent of next generation sequencing (NGS) technologies and the improvement of techniques to extract DNA from ancient specimens opened the way to the birth of the paleogenomics field. The first draft sequences of the mitochondrial and nuclear Neanderthal genomes were published in the last 15 years (Green et al., 2008; 2016) immediately followed by the publication of more genomes coming from Neanderthal remains found in other parts of the world (Prüfer et al., 2014, 2017). The comparison of these genome sequences with those of present-day *Homo sapiens* individuals of multiple ethnicities revealed the presence of genomic regions introgressed from the Neanderthals, which were more prevalent in non-African individuals and were the long-term remains of ancient hybridization events between the two populations. This discovery was followed by the publication of the genome sequence of other ancient hominins, the Denisovans (Meyer et al., 2012; Reich et al., 2010; Slon et al., 2017). Analogously, the comparison of Denisovan archaic DNA with that of modern humans revealed the presence of introgressed Denisovan DNA, which is only found in modern Melanesian populations. The Neanderthal and Denisovan introgression pattern in modern humans coincides with the geographical distribution of ancient hominid populations: anatomically modern humans appeared in Africa around 200000 years ago and started to migrate out of Africa around 50000 years ago. No fossil record currently exists placing Neanderthals and Denisovans

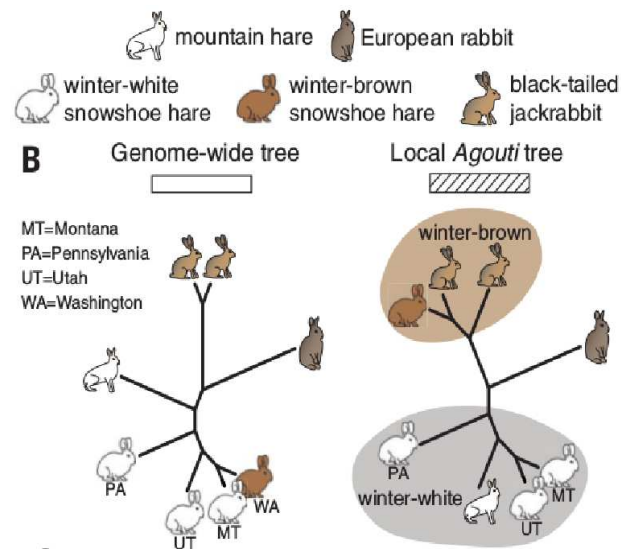


Figure 4: “Whole genome” versus “Agouti” tree topologies of snowshoe hares. Image adapted from Jones et al., 2018.

inside the African continent, and their geographical distribution ranges from Europe to Asia. Neanderthals resided in the Middle East, Europe and parts of Asia and left an abundant collection of fossil and manufacture records (Hublin, 2009). On the contrary, Denisovans are believed to have resided in parts of East and Southeast Asia, although the only fossil documenting their existence was retrieved in the Denisova cave in Siberia (Reich et al., 2010). Modern non-African and Oceanian human populations contain ~2% Neanderthal and 2-4% of Denisovan DNA, respectively. The model which best explains these data starts with the out-of-Africa migration of modern humans to Europe, where they met the Neanderthals and interbred with them. Subsequently, a group of humans continued the migration towards South-East Asia, where it interbred with Denisovans, and finally reached Oceania (Vattathil & Akey, 2015). The absence of both Neanderthal and Denisovan introgressions from African populations confirms that the admixture occurred after the out-of-Africa migration. The exclusive presence of Neanderthal introgressions in European individuals indicates that the admixture with Neanderthals occurred before the one with Denisovans, which is further supported by the presence of both Neanderthal and Denisovan introgressions in Oceanians (**Figure 5**) (Wolf & Akey, 2018). The outstanding recent discovery of a first-generation Neanderthal-Denisovan hybrid, resulting from the cross of a Neanderthal mother and a Denisovan father, provided a direct proof of admixture (Slon et al., 2018). Neanderthal and Denisovan sequences were progressively purged from the human genome by rounds of successive backcrossing within the human population, and are now present only in small traces. Interestingly, introgressions were not unilateral and ancient admixture left traces of modern humans also in the genome of Neanderthals, but not Denisovan individuals (Kuhlwilm et al., 2016).

2.3.2) The phenotypic impact of archaic introgressions in human populations

Introgressions are not evenly distributed in the human genome. As an example, Neanderthal contribution is very low on the X chromosome respect to the autosomes, and there are vast regions completely depleted of Neanderthal ancestry all across the genome (Sankararaman et al., 2014; Vernot & Akey, 2014), which are enriched in functionally relevant genes. One explanation for this phenomenon could be the action of negative selection to purge introgressions from the genome, and is supported by the observation that introgression deserts are significantly enriched for genes which are essential for speech and language development (Vernot & Akey, 2014; Wolf & Akey, 2018). Moreover, associations of Neanderthal alleles with neurological, psychiatric, immunological, and dermatological diseases, along with height, sleeping patterns, mood, and smoking status, have been found in individuals of European ancestry (Dannemann & Kelso, 2017; Simonti et al., 2016). In the middle of the 2020 world COVID19 pandemic, researchers have surprisingly discovered that a genomic region introgressed from Neanderthals constitutes the major risk factor for developing a

severe form of SARS-CoV2 respiratory disease, and it is present in about 16% of the European population (Zeberg & Pääbo, 2020).

On the other hand, Neanderthal and Denisovan introgression have also been shown to have adaptive functions and be under positive selection (**Figure 5**). Regions with Neanderthal ancestry are enriched in genes affecting keratin constitution, skin pigmentation (*OCA2*, *BNC2*) and immune response (*OAS1/2/3*, *TLR1/6/10*, *TNFAIP3*, *STAT2*) supporting the hypothesis that archaic admixture would provide advantageous genetic variants adapted to colder climates, lower sun exposure and endemic pathogens, which are characteristic of habitats inhabited by ancient Neanderthals (Dannemann et al., 2017; Gittelman et al., 2016; Mendez et al., 2012a, 2012b; Sankararaman et al., 2014). More recently, researchers found that Neanderthal introgressions segments in Europeans are enriched for genes encoding proteins that interact with viruses, especially the ones interacting with RNA viruses like HIV, influenza A virus and hepatitis C virus, confirming that Neanderthal DNA introgressed in modern humans helped them adapt against locally-present viruses (Enard & Petrov, 2018). Neanderthal sequences contribute to the human phenotype also at a regulatory level (McCoy et al., 2017).

Denisovan introgressed regions also exhibit evidence of positive selection. Tibetan individuals adapted to live at an altitude of more than 4000 meters through the fixation of a variant of the *EPAS1* gene, coding for a transcription factor modulating haemoglobin concentration at high altitude. This variant is not present in other populations and recent studies showed that it was inherited from Denisovans and maintained in the Tibetan population (Huerta-Sánchez et al., 2014). Another example of adaptive Denisovan introgression involves the genes *WARS2* and *TBX15*, which regulate fat-tissue metabolism and body-fat distribution. These genes are at high frequency and adaptive in the Inuit population of Greenland, and they show Denisovan ancestry (Racimo et al., 2017). An exhaustive list of adaptive introgressions in humans can be found in (Racimo et al., 2015).

In conclusion, recent literature revealed a surprising and fascinating history of interbreeding between ancient hominins and modern humans, that left traces in the genome and can have a deep phenotypic impact also nowadays, shaping our response to rapidly-changing environments and to diseases.

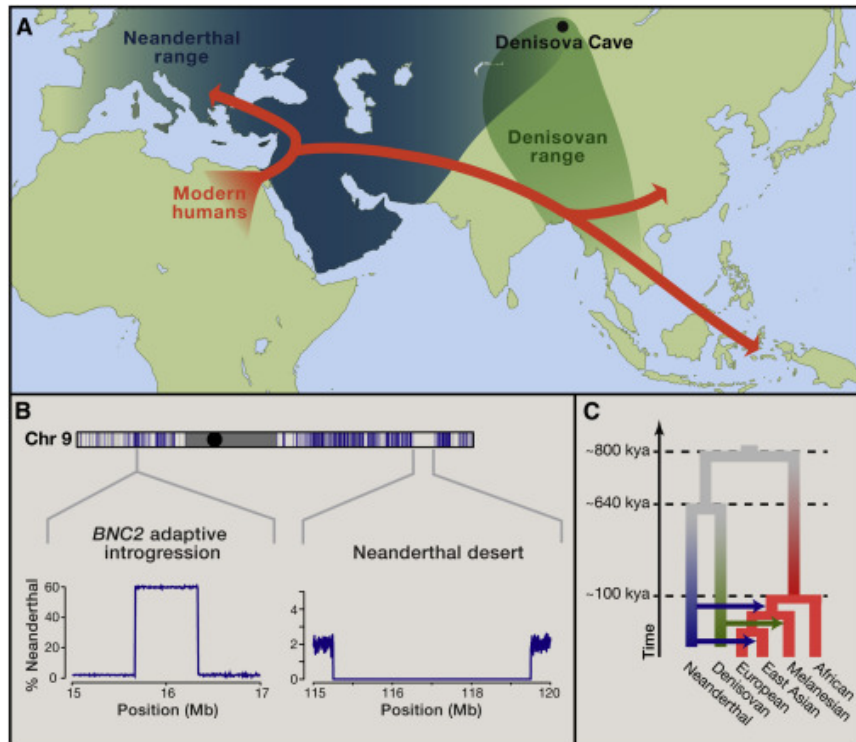


Figure 5: a) Schematic illustration of the out of Africa dispersal of modern humans. b) Distribution of introgressions on chromosome IX in Europeans and East Asians. c) Timing and direction of introgressions between hominin populations. Adapted from Vattathil & Akey, 2015.

2.4) The importance of reticulated evolution in yeast

2.4.1) Hybridization in yeast

The complete genome sequence of *S. cerevisiae* has been obtained in 1996 and it has been the first eukariotic organism to have its genome completely sequenced (Goffeau et al., 1996). Subsequently, many more strains of this species have been isolated and *S. cerevisiae* officially entered the population genomics era in 2009, with the first global-scale study of its population genomics (Liti et al., 2009), followed by other studies which deepened our understanding of *S. cerevisiae* ecology and diversity (Almeida et al., 2015; Peter & Schacherer, 2016; Ayoub et al., 2006; Bergstrom et al., 2014; Cromie et al., 2013; Duan et al., 2018; Fay & Benavides, 2005; Gallone et al., 2016; Liti & Schacherer, 2011; Peter et al., 2018; Ramazzotti et al., 2019; Bigey et al., 2020; Strope et al., 2015). After *S. cerevisiae*, other species of the *Saccharomyces sensu stricto* complex entered the public genomic databases (Almeida et al., 2014; Cliften et al., 2003; Kellis et al., 2003; Langdon et al., 2020; Libkind et al., 2011; Liti et al., 2013; Naseeb et al., 2017; Nespolo et al., 2020). As many more species were sequenced and their genomes compared, it became clear that there was a remarkable degree of genetic flow among them, manifesting in the form of hybridizations and introgressions. In fact, taxa that were considered as pure species, like *S. bayanus*, turned out to be hybrids (Nguyen et al., 2011).

Yeast hybrids are numerous and are especially relevant in the industrial world (**Figure 6**). The most known hybrid is probably *S. pastorianus*, the main fermenting agent used in the production of lager beers. The parental origin of *S. pastorianus* has long been debated: while its hybrid nature was known since the 80s (Martini & Kurtzman, 1985; Vaughan Martini & Martini, 1987), the identity of its second parent (other than *S. cerevisiae*) was unknown until recently (Dunn & Sherlock, 2008). The isolation and sequencing of new yeast isolates from Southern beeches in Patagonia identified a new species, *S. eubayanus*, as the missing parent of *S. pastorianus* (Libkind et al., 2011). Other researchers later identified a Tibetan *S. eubayanus* isolate which is even closer to the genome of *S. pastorianus* (Bing et al., 2014). Strains of *S. pastorianus* can be divided into 2 groups on the basis of their genome content and structure (Dunn & Sherlock, 2008). The hybrid ancestor of group I, which includes Saaz-type bottom-fermenting yeast, was triploid, with a diploid *S. eubayanus* and a haploid *S. cerevisiae* genome content. Group II strains, which include Froberg-type bottom-fermenting yeasts, have actually arisen from the fusion of a homozygous diploid *S. cerevisiae* with a haploid *S. eubayanus* (D'Angiolo, 2016).

Other industrially relevant hybrids are those between *S. cerevisiae* and *S. kudriavzevii* and between *S. cerevisiae* and *S. uvarum*, mainly used for cider, wine and beer fermentation (Belloch et al., 2009; Borneman et al., 2012; Christine et al., 2007; González et al., 2006; Krogerus et al., 2018; Pérez-Torrado et al., 2015; Peris et al., 2012). A complete list of wine hybrids can be found in (Sipiczky 2008). Hybrids sometimes present several advantages and higher fitness and evolvability compared to their parents, a phenomenon known as “hybrid vigor” (Bernardes et al., 2017; Dunn et al., 2013; Gallone et al., 2019; Herbst et al., 2017; Shapira et al., 2014; Stelkens et al., 2014). For example, *S. kudriavzevii* and *S. uvarum* are better adapted to grow at low temperatures, while *S. cerevisiae* is more alcohol tolerant. Hybrids between these species combine these useful traits and can thus grow under high concentrations of ethanol and at low temperature, conditions which are usually present during fermentation processes. Natural hybrids between other *Saccharomyces* species have been described, indicating that this process occurs in the wild and is not limited to conditions of human influence (Barbosa et al., 2016; Morales & Dujon, 2012; Stefanini et al., 2016; H. Zhang et al., 2010).

2.4.2) Introgressions in yeast

The high frequency of natural yeast hybrids translates into an equally abundant frequency of introgressed isolates (**Figure 6**). Cases of introgressions among species of the *Saccharomyces sensu stricto* complex have been widely documented. Introgressions from *S. paradoxus* to *S. cerevisiae* have been described in natural strains isolated in Brazil (Barbosa et al., 2016), in European strains isolated from oaks in the mediterranean region (Almeida et al., 2017; Barbosa et al., 2016), in

strains associated with olive-related environments (Pontes et al., 2019), and in other instances (Dunn et al., 2012; Muller & McCusker, 2009). A *S. paradoxus* introgression on chromosome I is present in the clinical strain YJM789, isolated from a pneumonia infection in an AIDS-affected patient (Wei et al., 2007), and in the strain EM93, which is an ancestor of the lab strain S288C (Esberg et al., 2011). Peter, De Chiara et al. reported the discovery of over 800 *S. paradoxus* introgressed ORFs in more than 1000 *S. cerevisiae* isolates. Some of them were shared among all the isolates, indicating a very ancient origin, while others were private to one or few phylogenetic lineages. Most of these ORFs derived from two different populations of *S. paradoxus*, while ORFs introgressed into a Taiwanese lineage came from a an unknown yeast (Peter et al., 2018).

Introgressions in the opposite direction, that is from *S. cerevisiae* to *S. paradoxus*, have also been observed. For example, an introgression of 25 kb in the left subtelomere of chromosome XIV is fixed in the *S. paradoxus* European population, but absent in the Far-Eastern and American ones (Liti et al., 2006). Frequent introgressions have been documented also between other species of the *S. sensu stricto*

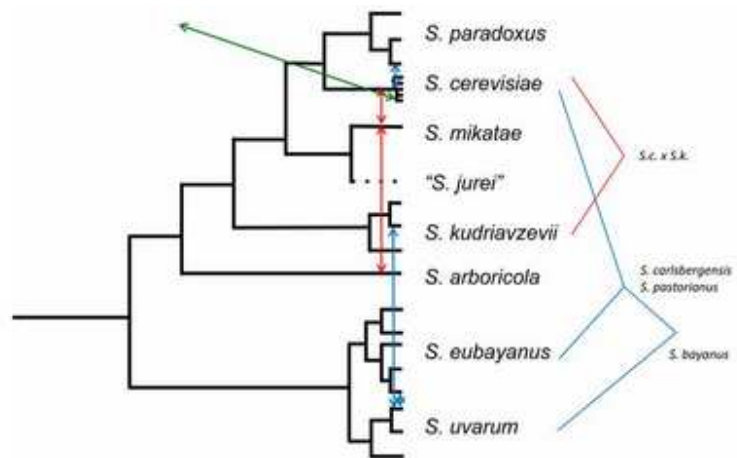


Figure 6: Phylogenetic tree showing hybrids (plain lines) and introgressions (arrows) among species of the *Saccharomyces sensu stricto* complex. Adapted from Dujon & Louis, 2017.

complex: introgressions from *S. eubayanus* (Almeida et al., 2014) and *S. kudriavzevii* (Albertin et al., 2018) into European *S. uvarum* strains associated with human-driven fermentations have been reported (Dujon & Louis, 2017). Gene flow in the form of introgressions can be also detected within subpopulations of the same species: Leducq *et al.* reported the discovery of an introgressed population resulting from the ancient admixture of two indigenous *S. paradoxus* populations living in partial sympatry in North-American forests (Leducq et al., 2016). Introgressions are not exclusive of the nuclear genome, but can be found, although in rare instances, also in the mitochondrial genome (De Chiara et al., 2020; Leducq et al., 2017; Peris et al., 2017).

There is indication that introgressions may have a potential adaptive role, as their underlying genes are often enriched in GO terms related to metabolic processes and the interaction with the external environment, like nitrogen metabolism (Almeida et al., 2014) and transmembrane transporters (Barbosa et al., 2016; Pontes et al., 2019). Galeote *et al.* reported that the gene *FSY1*, transferred to wine yeasts from an unknown *Saccharomyces* species and encoding a high-affinity fructose/H⁺ symporter, may confer a significant advantage to *S. cerevisiae* during wine fermentation (Galeote et

al., 2010a). However, a conclusive evidence of the adaptive role of introgressions in yeast is yet to be reported. Outside the *Saccharomyces* genus, a case of adaptive introgression has been reported in *Kluyveromyces lactis*, where the ability to ferment lactose was acquired through introgression of *LAC* genes from a dairy strain of *K. marxianus* (Varela et al., 2019) (**Figure 7**).

On the other side, the role of gene flow among more distantly related species, hereafter called “horizontal gene transfer (HGT)”, in the adaptation to new environments is known and well documented. Interesting examples can be found in wine yeasts harbouring 3 regions derived from *Zygosaccharomyces bailii* (Galeote et al., 2011; Novo et al., 2009), *Torulaspota microellipsoides* (Marsit et al., 2015) and an unknown donor from the *Saccharomyces* genus (Devia et al., 2020; Galeote et al., 2010b). The *Torulaspota*-derived region harbors *FOT* genes encoding oligopeptide transporters which confer fitness advantage during fermentation through modifications in the metabolism of aminoacids (Marsit et al., 2015). Another interesting example of adaptive HGT regards *S. cerevisiae* strains from spontaneously fermenting milk. These isolates exchanged their native *GAL* pathway with one from a unknown donor close to the *Saccharomyces* genus. The new *GAL* pathway enabled them to grow using both glucose and galactose as carbon sources, and made them perfectly adapted to live in a galactose-rich substrate like milk (Duan et al., 2019). Other examples of adaptive HGT can be found outside the *Saccharomyces* genus (Gonçalves & Gonçalves, 2019; Kominek et al., 2019; Milner et al., 2019; Ropars et al., 2015).

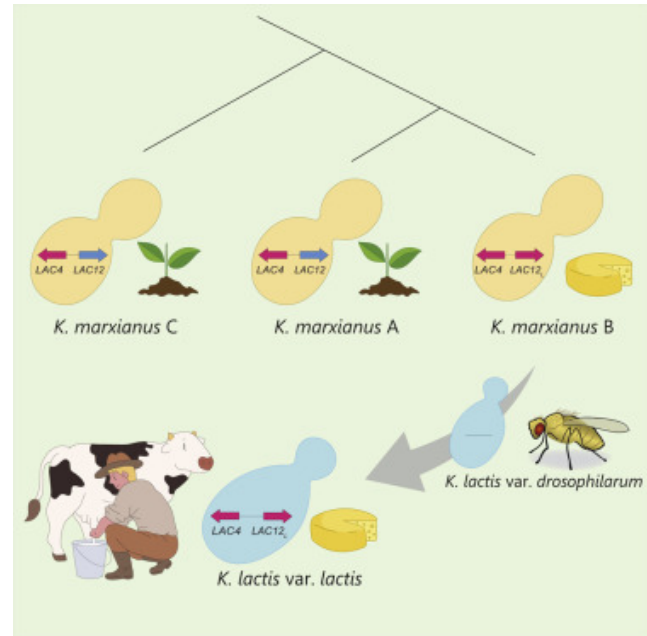


Figure 7: Example of introgression between *Kluyveromyces marxianus* and *Kluyveromyces lactis*. Adapted from Varela et al., 2019.

2.5) Hybrid sterility and the paradox of the origin of yeast introgressions

As we have seen in the previous paragraphs, gene flow is widespread across the tree of life and is particularly abundant in yeasts, where we can find cases of both interspecies hybridizations and introgressions (Dujon & Louis, 2017). The classic model for the origin of introgressions starts with the formation of an initial hybrid followed by the formation of gametes and their repeated unidirectional backcrossing to one parental population/species. This mechanism is widely accepted by the scientific community and can explain the origin of introgressions in humans, or more in

general, in organisms with an obligate sexual reproduction, but it relies on the assumption that the initial hybrid must be fertile and able to generate viable offspring.

As shown in the first chapter of this introduction, the *Saccharomyces sensu stricto* complex comprises 8 species and two hybrids. All these species have largely colinear genomes, where synteny is only disrupted at subtelomeric sites (Yue et al., 2017). Despite of that, sequence divergence among them can be pretty high and goes up to 30-35% among the most distant species of the complex. To make a comparison with humans, in terms of aminoacid changes in orthologous proteins, the diversity present in the *Saccharomyces sensu stricto* complex is on the same range of that present between a human and a chicken (Dujon, 2006; Scannell et al., 2011). We have seen before that hybrids between *sensu stricto* species are frequently found in nature and readily form in the laboratory when strains of opposite mating type are placed together in a tube. These hybrids are viable and are able to propagate clonally by mitotic divisions, and do not present extensive phenotypic defects relative to their parents (Greig, 2009), indicating the absence of pre-zygotic reproductive barriers. However, hybrids are generally sterile and only <1% of their spores are viable (Kao et al., 2010; Liti et al., 2006). The main mechanism causing hybrid sterility is the anti-recombination activity of the mismatch repair system (MMR). During meiosis, recombination initiates with the formation of heteroduplex DNA structure between the homologous parental chromosomes of the hybrid. If these chromosomes are highly diverged, the MRR will detect the mismatches and abort recombination. The abortion of recombination results in a high rate of aneuploidies in the gametes, which eventually will be inviable (Hunter et al., 1996). The deletion of essential MMR genes like *MSH2*, *SGS1*, and *PMS1* effectively results in an increase of hybrid fertility and gamete viability, confirming the preponderant role of anti-recombination in post-zygotic reproductive isolation (Greig et al., 2003; Bozdag et al., 2019; Ono et al., 2020; Rogers et al., 2018). Even though hybrid sterility is mostly explained by the MMR, incompatibilities between genes from different species (Bateson-Dobzhansky-Muller incompatibilities (BDMIs) might also play a role. Currently, no dominant BDMIs have been described in yeast (Greig et al., 2002), but small effect recessive ones have been repeatedly detected (Kao et al., 2010; Bozdag et al., 2019). Moreover, a recessive BDMI involving a nuclear-encoded gene and a mitochondrially-encoded one has been described (Chou et al., 2010; Chou & Leu, 2010; Jhuang et al., 2017; Lee et al., 2008).

This observations lead to an evolutionary paradox in which, on one hand, we observe frequent hybridizations and introgressions between *Saccharomyces* species, but on the other hand the extensive hybrid sterility is difficult to reconcile with a model in which introgressions arise by hybridization followed by multiple backcrossings, because obviously that requires the presence of viable hybrid gametes, and it opens up the question: how do introgressions arise in reproductively isolated species, such as yeasts?

2.6) Possible mechanisms explaining the emergence of yeast introgressions

2.6.1) Mechanisms involving the passage through a sporulation stage in the initial hybrid and through a backcrossing phase

The reliability of the hybridization-backcross model to explain yeast introgressions requires that the initial hybrid overcomes the sterility barrier. This can be achieved through two main molecular mechanisms: genome instability and whole-genome duplication.

Genome instability - Genome instability refers to a group of genetic alterations which can go from point mutations to macroscopic chromosomal rearrangements (Aguilera & Gómez-González, 2008). For the scope of this introduction chapter, when I will refer to genome instability I will only talk about the latter category of alterations. After hybridization, hybrids undergo extensive processes of genome stabilization, including chromosomal rearrangements, gain and loss of various chromosomes and mitotic recombination resulting in the appearance of blocks of sequence homology or loss of heterozygosity (LOH) (Gabaldón, 2020; Marsit et al., 2017). The gradual acquisition of homology blocks through uniparental chromosome loss and resynthesis or LOH formation creates sites where the local sequence divergence is lower than in the rest of the genome, resulting in a buffering of the anti-recombination effect of the mismatch-repair system. This could eventually lead to an increased recombination rate, decreased aneuploidy rate and higher spore viability, therefore neutralizing the sterility barrier in the hybrid. Genome instability progressively accumulates by mitotic recombination in artificial intra- and interspecific hybrids propagating clonally in laboratory conditions (Pankajam et al., 2020; Smukowski Heil et al., 2017; Tattini et al., 2019), and it could explain the recovery of fertility, but currently no study estimated spore viability before and after LOH onset to assess whether there is a change in hybrid fertility.

Whole genome duplication - A second mechanism to recover fertility is whole genome duplication (WGD). Hybrid sterility derives from the absence of recombination between homologous chromosomes that further results in high aneuploidy rate and death of the spores. This can be avoided through the endoreduplication of the genome, which guarantees the presence of a suitable partner for recombination in each chromosome. WGD can occur rapidly and spontaneously during experimental evolution, as observed in *Saccharomyces* hybrids (Charron et al., 2019; Marsit et al., 2020) and also in *Zygosaccharomyces parvibailii* (Ortiz-Merino et al., 2017) and results as expected in a full recovery of fertility.

The clonal expansion of the hybrid and its sporulation at the population level can also constitute a

mean to overcome hybrid sterility, as the massive parallel sporulation of millions of cells increments the number of viable spores obtained.

Proven that hybrids can restore fertility and generate viable offspring, the gametes have to be able to backcross with one of their parents. The unidirectional and repeated backcross seems unlikely given the frequency of sexual reproduction of yeasts in nature, which is estimated in 0.1%, that is one sporulation every 1000 mitotic generations (Hittinger, 2013). Among the spores resulting from meiosis, the majority mates with partners inside the same tetrad, while only 1% performs outcrossing with an unrelated partner (Hittinger, 2013b; Tsai et al., 2008). However, spore viability in hybrids is very low, meaning that the majority of the tetrads will contain a single viable spore. In this type of scenario, outcrossing will be favoured. Recent studies point out that the rate of outcrossing in nature might be underestimated (Murphy & Zeyl, 2010), since strain heterozygosity is high (Kelly & Wickner, 2013; Magwene et al., 2011). The explanation for this phenomenon might be that some ecological niches, like insect guts, favour outcrossing (Reuter et al., 2007; Stefanini et al., 2016), or the variation in germination and mating timing (Maclean & Greig, 2008; McClure et al., 2018; Miller & Greig, 2015; Murphy & Zeyl, 2012; Stelkens et al., 2016).

2.6.2) Mechanisms involving the passage through a sporulation stage in the initial hybrid but not involving a backcrossing phase

Other alternative models for the origin of introgressions have been proposed, which do not involve the backcrossing phase. One of them is the segregation of rare spores with a genome composition strongly biased towards one species. This would greatly reduce the number of backcrosses needed to resect the remaining genome contribution from the other species (Morales & Dujon, 2012), or even nullify it if the segregation is complete. A similar mechanism has been also proposed by Dunn *et al.*, who evolved *S. cerevisiae*/*S. uvarum* hybrids in ammonium-limiting conditions and observed the recurrent appearance of a LOH in a region containing the gene *MEP2*, encoding an ammonium permease. The authors propose that hybrids may produce rare viable spores containing only *S. cerevisiae* chromosomes including LOH regions from *S. uvarum*, that would therefore become the introgressions once spores germinate and self-diploidize. Alternatively, a preferential loss of the *S. uvarum* chromosomes from the hybrid could also result in a *S. cerevisiae* strain containing introgressed *S. uvarum* segments (Dunn et al., 2013). Both mechanisms would not require the hybrid to recover fertility and could therefore be more rapid as the first only needs one sporulation step, and the second does not even need to pass through sporulation. A retention of a specific set of chromosomes in hybrids has been observed also in other evolution experiments with the same type of hybrids (Antunovics et al., 2005; Piotrowski et al., 2012) and it is possible that selection and

environmental conditions might also play a role in guiding hybrid genome evolution.

2.6.3) Mechanisms not involving the passage through a sporulation stage in the initial hybrid

Finally, the homogenization of hybrid genomes can not only result from directional loss of one subgenome, but might also derive from the accumulation of LOH events in one species direction. The final result of this process would be an almost completely homozygous genome, with some small introgressed fragments. Assuming that the LOH onset is gradual and constant, this would require an extensive number of generations since the LOH rate is estimated to be ~250 bp per generation in intraspecific hybrids and ~50 bp in interspecific ones (Tattini et al., 2019).

However, other experimental evolution studies report that LOH rate can be significantly higher in some lines (Dutta et al., 2017; Pankajam et al., 2020), suggesting that LOH accumulation might not follow a linear model, but LOH could appear in bursts (Sampaio et al., 2020).

A possible mechanism explaining LOH bursts could be the ability of *Saccharomyces* strains to exit from meiosis and re-enter into mitosis, a process called “Return to Growth (RTG)”, which results in an extensive map of LOH across the genome (Laureau et al., 2016; Mozzachiodi et al., 2020). As said before, it is possible that selection and environmental conditions might also play a role in determining the direction and tempo of LOH appearance, as it has been demonstrated in several studies (Lancaster et al., 2019; Smukowski Heil et al., 2017).

2.7) Conclusion

All along the last paragraph, I have described many routes that can bring from hybrids to pure strains with introgressions. However, these routes do not reciprocally exclude each other and they can be sometimes intertwined. It is possible that the “true” mechanism of origin of introgressions is a mixture of all these routes and involves all of them at different times, or maybe it does not even exist a “true” mechanism, and introgressions arise following different paths depending on the environmental conditions and the species considered. It is clear that additional research needs to be conducted to clarify how and to what extent each of these possible paths contributes to the emergence of yeast genomic introgressions.

References

- Aguilera, A., & Gómez-González, B. (2008). Genome instability: A mechanistic view of its causes and consequences. *Nature Reviews Genetics*, 9(3), 204–217. <https://doi.org/10.1038/nrg2268>
- Albertin, W., Chernova, M., Durrens, P., Guichoux, E., Sherman, D. J., Masneuf-Pomarede, I., & Marullo, P. (2018). Many interspecific chromosomal introgressions are highly prevalent in Holarctic *Saccharomyces uvarum* strains found in human-related fermentations. *Yeast*, 35(1), 141–156. <https://doi.org/10.1002/yea.3248>
- Almeida, P., Barbosa, R., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2017). Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. *Molecular Ecology*, 26(7), 2167–2182. <https://doi.org/10.1111/mec.14071>
- Almeida, P., Barbosa, R., Zalar, P., Imanishi, Y., Shimizu, K., Turchetti, B., Legras, J. L., Serra, M., Dequin, S., Couloux, A., Guy, J., Bensasson, D., Gonçalves, P., & Sampaio, J. P. (2015). A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular Ecology*, 24(21), 5412–5427. <https://doi.org/10.1111/mec.13341>
- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarede, I., Albertin, W., Durrens, P., Sherman, D. J., Marullo, P., Todd Hittinger, C., Gonçalves, P., & Sampaio, J. P. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature Communications*, 5(May). <https://doi.org/10.1038/ncomms5044>
- Antunovics, Z., Nguyen, H. V., Gaillardin, C., & Sipiczki, M. (2005). Gradual genome stabilisation by progressive reduction of the *Saccharomyces uvarum* genome in an interspecific hybrid with *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 5(12), 1141–1150. <https://doi.org/10.1016/j.femsyr.2005.04.008>
- Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., Bomblies, K., & Yant, L. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8320–8325. <https://doi.org/10.1073/pnas.1600405113>
- Laadan B., Almeida J., Radstrom P., Hahn-Hagerdal B., Gorwa-Grauslund M. (2008). Identification of a NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Yeast*, January, 191–198. <https://doi.org/10.1002/yea>
- Ayoub, M. J., Legras, J. L., Saliba, R., & Gaillardin, C. (2006). Application of Multi Locus Sequence Typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. *Journal of Applied Microbiology*, 100(4), 699–711. <https://doi.org/10.1111/j.1365-2672.2006.02817.x>
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, evv263. <https://doi.org/10.1093/gbe/evv263>

- Belloch, C., Pérez-Torrado, R., González, S. S., JoséE.Pérez-Ortín, García-Martínez, J., Querol, A., & Barrio, E. (2009). Chimeric genomes of natural hybrids Of *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *Applied and Environmental Microbiology*, *75*(8), 2534–2544. <https://doi.org/10.1128/AEM.02282-08>
- Bergstrom, A., Simpson, J. T., Salinas, F., Barré, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, *31*(4), 872–888. <https://doi.org/10.1093/molbev/msu037>
- Bernardes, J. P., Stelkens, R. B., & Greig, D. (2017). Heterosis in hybrids within and between yeast species. *Journal of Evolutionary Biology*, *30*(3), 538–548. <https://doi.org/10.1111/jeb.13023>
- Bing, J., Han, P. J., Liu, W. Q., Wang, Q. M., & Bai, F. Y. (2014). Evidence for a far east asian origin of lager beer yeast. *Current Biology*, *24*(10), 380–381. <https://doi.org/10.1016/j.cub.2014.04.031>
- Borneman, A. R., Desany, B. A., Riches, D., Affourtit, J. P., Forgan, A. H., Pretorius, I. S., Egholm, M., & Chambers, P. J. (2012). The genome sequence of the wine yeast VIN7 reveals an allotriploid hybrid genome with *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii* origins. *FEMS Yeast Research*, *12*(1), 88–96. <https://doi.org/10.1111/j.1567-1364.2011.00773.x>
- Brower, A. V. Z. (2013). Introgression of wing pattern alleles and speciation via homoploid hybridization in *Heliconius* butterflies: A review of evidence from the genome. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1752). <https://doi.org/10.1098/rspb.2012.2302>
- Charron, G., Marsit S., Henault M., Martin H., Landry, C. R. (2019). Spontaneous whole-genome duplication restores fertility in interspecific hybrids. *Nature Communications*, 2019. <http://dx.doi.org/10.1038/s41467-019-12041-8>
- Chou, J. Y., Hung, Y. S., Lin, K. H., Lee, H. Y., & Leu, J. Y. (2010). Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology*, *8*(7). <https://doi.org/10.1371/journal.pbio.1000432>
- Chou, J. Y., & Leu, J. Y. (2010). Speciation through cytonuclear incompatibility: Insights from yeast and implications for higher eukaryotes. *BioEssays*, *32*(5), 401–411. <https://doi.org/10.1002/bies.200900162>
- Christine, L. J., Marc, L., Catherine, D., Claude, E., Jean-Luc, L., Michel, A., & Isabelle, M. P. (2007). Characterization of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces bayanus* var. *uvarum*. *FEMS Yeast Research*, *7*(4), 540–549. <https://doi.org/10.1111/j.1567-1364.2007.00207.x>
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B. A., & Johnston, M. (2003). Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science*, *301*(5629), 71–76. <https://doi.org/10.1126/science.1084337>
- Heliconius Genome Consortium. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94–98. <https://doi.org/10.1038/nature11041>.Butterfly

- Cromie, G. A., Hyma, K. E., Ludlow, C. L., Garmendia-Torres, C., Gilbert, T. L., May, P., Huang, A. A., Dudley, A. M., & Fay, J. C. (2013). Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3: Genes, Genomes, Genetics*, 3(12), 2163–2171. <https://doi.org/10.1534/g3.113.007492>
- D'Angiolo, M. (2016). <https://etd.adm.unipi.it/t/etd-06222016-153506/>. Master's degree thesis.
- Dannemann, M., & Kelso, J. (2017). The Contribution of Neanderthals to Phenotypic Variation in Modern Humans. *American Journal of Human Genetics*, 101(4), 578–589. <https://doi.org/10.1016/j.ajhg.2017.09.010>
- Dannemann, M., Prüfer, K., & Kelso, J. (2017). Functional implications of Neandertal introgression in modern humans. *Genome Biology*, 18(1), 1–11. <https://doi.org/10.1186/s13059-017-1181-7>
- De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., & Liti, G. (2020). Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biology*, 18(1), 1–15. <https://doi.org/10.1186/s12915-020-00786-4>
- Devia, J., Bastías, C., Kessi-Pérez, E. I., Villarroel, C. A., De Chiara, M., Cubillos, F. A., Liti, G., Martínez, C., & Salinas, F. (2020). Transcriptional Activity and Protein Levels of Horizontally Acquired Genes in Yeast Reveal Hallmarks of Adaptation to Fermentative Environments. *Frontiers in Genetics*, 11(April), 1–12. <https://doi.org/10.3389/fgene.2020.00293>
- Duan, S. F., Han, P. J., Wang, Q. M., Liu, W. Q., Shi, J. Y., Li, K., Zhang, X. L., & Bai, F. Y. (2018). The origin and adaptive evolution of domesticated populations of yeast from Far East Asia. *Nature Communications*, 9(1). <https://doi.org/10.1038/s41467-018-05106-7>
- Duan, S. F., Shi, J. Y., Yin, Q., Zhang, R. P., Han, P. J., Wang, Q. M., & Bai, F. Y. (2019). Reverse Evolution of a Classic Gene Network in Yeast Offers a Competitive Advantage. *Current Biology*, 29(7), 1126–1136.e5. <https://doi.org/10.1016/j.cub.2019.02.038>
- Dujon, B. (2006). Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. In *Trends in Genetics* (Vol. 22, Issue 7, pp. 375–387). <https://doi.org/10.1016/j.tig.2006.05.007>
- Dujon, B. A., & Louis, E. J. (2017). Genome diversity and evolution in the budding yeasts (Saccharomycotina). *Genetics*, 206(2), 717–750. <https://doi.org/10.1534/genetics.116.199216>
- Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., Louis, E. J., Liti, G., Sherlock, G., & Rosenzweig, F. (2013). Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. *PLoS Genetics*, 9(3). <https://doi.org/10.1371/journal.pgen.1003366>
- Dunn, B., Richter, C., Kvitek, D. J., Pugh, T., & Sherlock, G. (2012). Analysis of the *Saccharomyces cerevisiae* pan-genome reveals a pool of copy number variants distributed in diverse yeast strains from differing industrial environments. *Genome Research*, 22(5), 908–924. <https://doi.org/10.1101/gr.130310.111>
- Dunn, B., & Sherlock, G. (2008). Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. *Genome Research*, 18(10), 1610–1623. <https://doi.org/10.1101/gr.076075.108>

- Dutta, A., Lin, G., Pankajam, A. V., Chakraborty, P., Bhat, N., Steinmetz, L. M., & Nishant, K. T. (2017). Genome dynamics of hybrid *Saccharomyces cerevisiae* during vegetative and meiotic divisions. *G3: Genes, Genomes, Genetics*, 7(11), 3669–3679. <https://doi.org/10.1534/g3.117.1135>
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-accinelli, G., Belleghem, S. M. Van, & Patterson, N. (2019). Butterfly Radiation. *Science*, 599(November), 594–599.
- Enard, D., & Petrov, D. A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell*, 175(2), 360-371.e13. <https://doi.org/10.1016/j.cell.2018.08.034>
- Fay, J. C., & Benavides, J. A. (2005). Evidence for domesticated and wild populations of *saccharomyces cerevisiae*. *PLoS Genetics*, 1(1), 0066–0071. <https://doi.org/10.1371/journal.pgen.0010005>
- Gabaldón, T. (2020). Hybridization and the origin of new yeast lineages. *FEMS Yeast Research*, 20(5), 1–8. <https://doi.org/10.1093/femsyr/foaa040>
- Galeote, V., Bigey, F., Beyne, E., Novo, M., Legras, J. L., Casaregola, S., & Dequin, S. (2011). Amplification of a *Zygosaccharomyces bailii* DNA segment in wine yeast genomes by extrachromosomal circular DNA formation. *PLoS ONE*, 6(3), 1–10. <https://doi.org/10.1371/journal.pone.0017872>
- Galeote, V., Novo, M., Salema-Oom, M., Brion, C., Valério, E., Gonçalves, P., & Dequin, S. (2010a). FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H⁺ symporter. In *Microbiology* (Vol. 156, Issue 12, pp. 3754–3761). <https://doi.org/10.1099/mic.0.041673-0>
- Galeote, V., Novo, M., Salema-Oom, M., Brion, C., Valério, E., Gonçalves, P., & Dequin, S. (2010b). FSY1, a horizontally transferred gene in the *Saccharomyces cerevisiae* EC1118 wine yeast strain, encodes a high-affinity fructose/H⁺ symporter. *Microbiology*, 156(12), 3754–3761. <https://doi.org/10.1099/mic.0.041673-0>
- Gallone, B., Steensels, J., Mertens, S., Dzialo, M. C., Gordon, J. L., Wauters, R., Theßeling, F. A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prahl, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., & Verstrepen, K. J. (2019). Interspecific hybridization facilitates niche adaptation in beer yeast. *Nature Ecology and Evolution*, 3(11), 1562–1575. <https://doi.org/10.1038/s41559-019-0997-9>
- Gallone, B., Steensels, J., Prahl, T., Soriaga, L., Saels, V., Herrera-Malaver, B., Merlevede, A., Roncoroni, M., Voordeckers, K., Miraglia, L., Teiling, C., Steffy, B., Taylor, M., Schwartz, A., Richardson, T., White, C., Baele, G., Maere, S., & Verstrepen, K. J. (2016). Domestication and Divergence of *Saccharomyces cerevisiae* Beer Yeasts. *Cell*, 166(6), 1397-1410.e16. <https://doi.org/10.1016/j.cell.2016.08.020>
- Geraldes, A., Farzaneh, N., Grassa, C. J., Mckown, A. D., Guy, R. D., Mansfield, S. D., Douglas, C. J., & Cronk, Q. C. B. (2014). Landscape genomics of *populus trichocarpa*: The role of

- hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, 68(11), 3260–3280. <https://doi.org/10.1111/evo.12497>
- Gittelman, R. M., Schraiber, J. G., Vernet, B., Mikacenic, C., Wurfel, M. M., & Akey, J. M. (2016). Archaic Hominin Admixture Facilitated Adaptation to Out-of-Africa Environments. *Current Biology*, 26(24), 3375–3382. <https://doi.org/10.1016/j.cub.2016.10.041>
- Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., & Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287), 546, 563–567. <http://www.ncbi.nlm.nih.gov/pubmed/8849441>
- Gonçalves, C., & Gonçalves, P. (2019). Multilayered horizontal operon transfers from bacteria reconstruct a thiamine salvage pathway in yeasts. *Proceedings of the National Academy of Sciences of the United States of America*, 116(44), 22219–22228. <https://doi.org/10.1073/pnas.1909844116>
- González, S. S., Barrio, E., Gafner, J., & Querol, A. (2006). Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Research*, 6(8), 1221–1234. <https://doi.org/10.1111/j.1567-1364.2006.00126.x>
- Green, R. E., Malaspina, A. S., Krause, J., Briggs, A. W., Johnson, P. L. F., Uhler, C., Meyer, M., Good, J. M., Maricic, T., Stenzel, U., Prüfer, K., Siebauer, M., Burbano, H. A., Ronan, M., Rothberg, J. M., Egholm, M., Rudan, P., Brajković, D., Kučan, Ž., ... Pääbo, S. (2008). A Complete Neandertal Mitochondrial Genome Sequence Determined by High-Throughput Sequencing. *Cell*, 134(3), 416–426. <https://doi.org/10.1016/j.cell.2008.06.021>
- Greig, D., Travisano, M., Louis, E. J., & Borts, R. H. (2003). A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *Journal of Evolutionary Biology*, 16(3), 429–437. <https://doi.org/10.1046/j.1420-9101.2003.00546.x>
- Greig, D. (2009). Reproductive isolation in *Saccharomyces*. *Heredity*, 102(1), 39–44. <https://doi.org/10.1038/hdy.2008.73>
- Greig, Duncan, Borts, R. H., Louis, E. J., & Travisano, M. (2002). Epistasis and hybrid sterility in *Saccharomyces*. *Proceedings. Biological Sciences / The Royal Society*, 269(1496), 1167–1171. <https://doi.org/10.1098/rspb.2002.1989>
- Hedrick, P. W. (2013). Adaptive introgression in animals: Examples and comparison to new mutation and standing variation as sources of adaptive variation. *Molecular Ecology*, 22(18), 4606–4618. <https://doi.org/10.1111/mec.12415>
- Herbst, R. H., Bar-Zvi, D., Reikhav, S., Soifer, I., Breker, M., Jona, G., Shimoni, E., Schuldiner, M., Levy, A. A., & Barkai, N. (2017). Heterosis as a consequence of regulatory incompatibility. *BMC Biology*, 15(1), 1–15. <https://doi.org/10.1186/s12915-017-0373-7>
- Higham, T., Douka, K., Wood, R., Ramsey, C. B., Brock, F., Basell, L., Camps, M., Arrizabalaga, A., Baena, J., Barroso-Ruiz, C., Bergman, C., Boitard, C., Boscato, P., Caparrós, M., Conard, N. J., Draily, C., Froment, A., Galván, B., Gambassini, P., ... Jacobi, R. (2014). The timing and spatiotemporal patterning of Neanderthal disappearance. *Nature*, 512(7514), 306–309.

<https://doi.org/10.1038/nature13621>

- Hittinger, C. T. (2013a). Saccharomyces diversity and evolution: A budding model genus. *Trends in Genetics*, 29(5), 309–317. <https://doi.org/10.1016/j.tig.2013.01.002>
- Hittinger, C. T. (2013b). Saccharomyces diversity and evolution: A budding model genus. *Trends in Genetics*, 29(5), 309–317. <https://doi.org/10.1016/j.tig.2013.01.002>
- Hublin, J. J. (2009). The origin of Neandertals. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38), 16022–16027. <https://doi.org/10.1073/pnas.0904119106>
- Hublin, Jean Jacques, Ben-Ncer, A., Bailey, S. E., Freidline, S. E., Neubauer, S., Skinner, M. M., Bergmann, I., Le Cabec, A., Benazzi, S., Harvati, K., & Gunz, P. (2017). New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. *Nature*, 546(7657), 289–292. <https://doi.org/10.1038/nature22336>
- Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., ... Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512(7513), 194–197. <https://doi.org/10.1038/nature13408>
- Hunter, N., Chambers, S. R., Louis, E. J., & Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO Journal*, 15(15), 1726–1733. <http://europepmc.org/abstract/MED/8612597>
- Jhuang, H., Lee, H., & Leu, J. (2017). Mitochondrial–nuclear co-evolution leads to hybrid incompatibility through pentatricopeptide repeat proteins. *EMBO Reports*, 18(1), 87–101. <https://doi.org/10.15252/embr.201643311>
- Jones, M. R., Mills, L. S., Alves, P. C., Callahan, C. M., Alves, J. M., Lafferty, D. J. R., Jiggins, F. M., & Jensen, J. D. (2018). in Snowshoe Hares. *Science*, 1358(June), 1355–1358.
- Kao, K. C., Schwartz, K., & Sherlock, G. (2010). A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genetics*, 6(7), 1–12. <https://doi.org/10.1371/journal.pgen.1001038>
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., & Lander, E. S. (2003). Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, 423(6937), 241–254. <https://doi.org/10.1038/nature01644>
- Kelly, A. C., & Wickner, R. B. (2013). Saccharomyces cerevisiae: A sexy yeast with a prion problem. *Prion*, 7(3), 215–220. <https://doi.org/10.4161/pri.24845>
- Kominek, J., Doering, D. T., Opulente, D. A., Shen, X. X., Zhou, X., DeVirgilio, J., Hulfachor, A. B., Groenewald, M., Mcgee, M. A., Karlen, S. D., Kurtzman, C. P., Rokas, A., & Hittinger, C. T. (2019). Eukaryotic Acquisition of a Bacterial Operon. *Cell*, 176(6), 1356-1366.e10. <https://doi.org/10.1016/j.cell.2019.01.034>
- Krogerus, K., Preiss, R., & Gibson, B. (2018). A unique saccharomyces cerevisiae saccharomyces

- uvarum hybrid isolated from norwegian farmhouse beer: Characterization and reconstruction. *Frontiers in Microbiology*, 9(SEP), 1–15. <https://doi.org/10.3389/fmicb.2018.02253>
- Kronforst, M. R. (2012). Mimetic butterflies introgress to impress. *PLoS Genetics*, 8(6), 7–9. <https://doi.org/10.1371/journal.pgen.1002802>
- Kuhlwilm, M., Gronau, I., Hubisz, M. J., De Filippo, C., Prado-Martinez, J., Kircher, M., Fu, Q., Burbano, H. A., Lalueza-Fox, C., De La Rasilla, M., Rosas, A., Rudan, P., Brajkovic, D., Kucan, Ž., Gušić, I., Marques-Bonet, T., Andrés, A. M., Viola, B., Pääbo, S., ... Castellano, S. (2016). Ancient gene flow from early modern humans into Eastern Neanderthals. *Nature*, 530(7591), 429–433. <https://doi.org/10.1038/nature16544>
- Lancaster, S. M., Payen, C., Heil, C. S., & Dunham, M. J. (2019). Fitness benefits of loss of heterozygosity in *Saccharomyces* hybrids. *Genome Research*, 29(10), 1685–1692. <https://doi.org/10.1101/gr.245605.118>
- Langdon, Q. K., Peris, D., Eizaguirre, J. I., Oplente, D. A., Buh, K. V., Sylvester, K., Jarzyna, M., Rodríguez, M. E., Lopes, C. A., Libkind, D., & Todd Hittinger, C. (2020). Postglacial migration shaped the genomic diversity and global distribution of the wild ancestor of lager-brewing hybrids. *PLoS Genetics*, 16(4), 1–22. <https://doi.org/10.1371/journal.pgen.1008680>
- Laureau, R., Loeillet, S., Salinas, F., Bergström, A., Legoix-Né, P., Liti, G., & Nicolas, A. (2016). Extensive Recombination of a Yeast Diploid Hybrid through Meiotic Reversion. *PLoS Genetics*, 12(2), e1005781. <https://doi.org/10.1371/journal.pgen.1005781>
- Leducq, J. B., Henault, M., Charron, G., Nielly-Thibault, L., Terrat, Y., Fiumera, H. L., Shapiro, B. J., & Landry, C. R. (2017). Mitochondrial recombination and introgression during speciation by hybridization. *Molecular Biology and Evolution*, 34(8), 1947–1959. <https://doi.org/10.1093/molbev/msx139>
- Leducq, J. B., Nielly-Thibault, L., Charron, G., Eberlein, C., Verta, J. P., Samani, P., Sylvester, K., Hittinger, C. T., Bell, G., & Landry, C. R. (2016). Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology*, 1(1). <https://doi.org/10.1038/nmicrobiol.2015.3>
- Lee, H. Y., Chou, J. Y., Cheong, L., Chang, N. H., Yang, S. Y., & Leu, J. Y. (2008). Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell*, 135(6), 1065–1073. <https://doi.org/10.1016/j.cell.2008.10.047>
- Libkind, D., Hittinger, C. T., Valério, E., Gonçalves, C., Dover, J., Johnston, M., Gonçalves, P., & Sampaio, J. P. (2011). Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 108(35), 14539–14544. <https://doi.org/10.1073/pnas.1105430108>
- Liti, G., Barton, D. B. H., & Louis, E. J. (2006). Sequence diversity, reproductive isolation and species concepts in *saccharomyces*. *Genetics*, 174(2), 839–850. <https://doi.org/10.1534/genetics.106.062166>
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. a, Davey, R. P., Roberts, I. N., Burt, A., Tsai, I. J., Bergman, C. M., Bensasson, D., Michael, J. T., Kelly, O., Oudenaarden,

- A. Van, Barton, D. B. H., Bailes, E., Alex, N., Ba, N., ... Louis, E. J. (2009). *UKPMC Funders Group Population genomics of domestic and wild yeasts*. 458(7236), 337–341. <https://doi.org/10.1038/nature07743>. Population
- Liti, G., Nguyen Ba, A. N., Blythe, M., Müller, C. a, Bergström, A., Cubillos, F. a, Dafhnis-Calas, F., Khoshraftar, S., Malla, S., Mehta, N., Siow, C. C., Warringer, J., Moses, A. M., Louis, E. J., & Nieduszynski, C. a. (2013). High quality de novo sequencing and assembly of the *Saccharomyces arboricolus* genome. *BMC Genomics*, 14, 69. <https://doi.org/10.1186/1471-2164-14-69>
- Liti, G., & Schacherer, J. (2011). The rise of yeast population genomics. *Comptes Rendus - Biologies*, 334(8–9), 612–619. <https://doi.org/10.1016/j.crvi.2011.05.009>
- Maclean, C. J., & Greig, D. (2008). Prezygotic reproductive isolation between *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. *BMC Evolutionary Biology*, 8, 1. <https://doi.org/10.1186/1471-2148-8-1>
- Magwene, P. M., Kayikçi, Ö., Granek, J. A., Reininga, J. M., Scholl, Z., & Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5), 1987–1992. <https://doi.org/10.1073/pnas.1012544108>
- Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, 38(2), 140–149. <https://doi.org/10.1002/bies.201500149>
- Marsit, S, Hénault, M., Charron, G., Fijarczyk, A., Landry, C. R., & Affiliations, *. (2020). The rate of whole-genome duplication can be accelerated by hybridization. *BioRxiv*, 2020.10.14.339820. <https://doi.org/10.1101/2020.10.14.339820>
- Marsit, S., Leducq, J. B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nature Reviews Genetics*, 18(10), 581–598. <https://doi.org/10.1038/nrg.2017.49>
- Marsit, S., Mena, A., Bigey, F., Sauvage, F. X., Couloux, A., Guy, J., Legras, J. L., Barrio, E., Dequin, S., & Galeote, V. (2015). Evolutionary advantage conferred by an eukaryote-to-eukaryote gene transfer event in wine yeasts. *Molecular Biology and Evolution*, 32(7), 1695–1707. <https://doi.org/10.1093/molbev/msv057>
- Martin, N. H., Bouck, A. C., & Arnold, M. L. (2006). Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics*, 172(4), 2481–2489. <https://doi.org/10.1534/genetics.105.053538>
- Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2018). Recombination rate variation shapes barriers to introgression across butterfly genomes. *BioRxiv*, 1–28. <https://doi.org/10.1101/297531>
- Martini, A. V., & Kurtzman, C. P. (1985). Deoxyribonucleic acid relatedness among species of the genus *Saccharomyces sensu stricto*. *International Journal of Systematic Bacteriology*, 35(4), 508–511. <https://doi.org/10.1099/00207713-35-4-508>
- McClure, A. W., Jacobs, K. C., Zyla, T. R., & Lew, D. J. (2018). Mating in wild yeast: Delayed

- interest in sex after spore germination. *Molecular Biology of the Cell*, 29(26), 3119–3127. <https://doi.org/10.1091/mbc.E18-08-0528>
- Mendez, F. L., Watkins, J. C., & Hammer, M. F. (2012a). A haplotype at STAT2 introgressed from neanderthals and serves as a candidate of positive selection in Papua New Guinea. *American Journal of Human Genetics*, 91(2), 265–274. <https://doi.org/10.1016/j.ajhg.2012.06.015>
- Mendez, F. L., Watkins, J. C., & Hammer, M. F. (2012b). Global genetic variation at OAS1 provides evidence of archaic admixture in Melanesian populations. *Molecular Biology and Evolution*, 29(6), 1513–1520. <https://doi.org/10.1093/molbev/msr301>
- Meyer, M., Kircher, M., Gansauge, M. T., Li, H., Racimo, F., Mallick, S., Schraiber, J. G., Jay, F., Prüfer, K., De Filippo, C., Sudmant, P. H., Alkan, C., Fu, Q., Do, R., Rohland, N., Tandon, A., Siebauer, M., Green, R. E., Bryc, K., ... Pääbo, S. (2012). A high-coverage genome sequence from an archaic Denisovan individual. *Science*, 338(6104), 222–226. <https://doi.org/10.1126/science.1224344>
- Miller, E. L., & Greig, D. (2015). Spore germination determines yeast inbreeding according to fitness in the local environment. *American Naturalist*, 185(2), 291–301. <https://doi.org/10.1086/679347>
- Milner, D. S., Attah, V., Cook, E., Maguire, F., Savory, F. R., Morrison, M., Müller, C. A., Foster, P. G., Talbot, N. J., Leonard, G., & Richards, T. A. (2019). Environment-dependent fitness gains can be driven by horizontal gene transfer of transporter-encoding genes. *Proceedings of the National Academy of Sciences of the United States of America*, 116(12), 5613–5622. <https://doi.org/10.1073/pnas.1815994116>
- Morales, L., & Dujon, B. (2012). Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and Molecular Biology Reviews : MMBR*, 76(4), 721–739. <https://doi.org/10.1128/MMBR.00022-12>
- Mozzachiodi, S., Tattini, L., Llored, A., Irizar, A., Škofljanc, N., D'Angiolo, M., De Chiara, M., Barré, B. P., Yue, J., Lutazi, A., Loillet, S., Laureau, R., Marsit, S., Stenberg, S., Persson, K., Legras, J., Dequin, S., Warringer J., Nicolas A., Liti G. (2020). Aborting meiosis overcomes hybrid sterility. Submitted.
- Muller, L. A. H., & McCusker, J. H. (2009). A multispecies-based taxonomic microarray reveals interspecies hybridization and introgression in *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 9(1), 143–152. <https://doi.org/10.1111/j.1567-1364.2008.00464.x>
- Murphy, H. A., & Zeyl, C. W. (2010). Yeast sex: Surprisingly high rates of outcrossing between asci. *PLoS ONE*, 5(5). <https://doi.org/10.1371/journal.pone.0010461>
- Murphy, H. A., & Zeyl, C. W. (2012). Prezygotic isolation between *saccharomyces cerevisiae* and *saccharomyces paradoxus* through differences in mating speed and germination timing. *Evolution*, 66(4), 1196–1209. <https://doi.org/10.1111/j.1558-5646.2011.01516.x>
- Bozdag G., Ono J., Denton J., Karakoc E., Hunter N., Leu J. Y., Greig D. (2019). Engineering recombination between diverged yeast species reveals genetic incompatibilities. Bioarxiv. <https://www.biorxiv.org/content/10.1101/755165v1.abstract>

- Naseeb, S., James, S. A., Alsammar, H., Michaels, C. J., Gini, B., Nueno-Palop, C., Bond, C. J., McGhie, H., Roberts, I. N., & Delneri, D. (2017). *Saccharomyces jurei* sp. Nov., isolation and genetic identification of a novel yeast species from *Quercus robur*. *International Journal of Systematic and Evolutionary Microbiology*, *67*(6), 2046–2052. <https://doi.org/10.1099/ijsem.0.002013>
- Neafsey, D. E., Sharakhov, I. V., Jiang, X., Hall, A. B., Kakani, E., Mitchell, S. N., Wu, Y., Smith, H. A., Hahn, M. W., & Besansky, N. J. (2016). Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science*, *347*(6217), 1–20. <https://doi.org/10.1126/science.1258524>.Extensive
- Nespolo, R. F., Villarroel, C. A., Oporto, C. I., Tapia, S. M., Vega-Macaya, F., Urbina, K., de Chiara, M., Mozzachiodi, S., Mikhalev, E., Thompson, D., Larrondo, L. F., Saenz-Agudelo, P., Liti, G., & Cubillos, F. A. (2020). An Out-of-Patagonia migration explains the worldwide diversity and distribution of *Saccharomyces eubayanus* lineages. *PLoS Genetics*, *16*(5), 1–23. <https://doi.org/10.1371/journal.pgen.1008777>
- Nguyen, H. V., Legras, J. L., Neuvéglise, C., & Gaillardin, C. (2011). Deciphering the hybridisation history leading to the lager lineage based on the mosaic genomes of *Saccharomyces bayanus* strains NBRC1948 and CBS380 T. *PLoS ONE*, *6*(10). <https://doi.org/10.1371/journal.pone.0025821>
- Nichols, R. (2001). Gene trees and species trees are not. *Science Direct*, *16*(7), 358–364.
- Novo, M., Bigey, F., Beyne, E., Galeote, V., Gavory, F., Mallet, S., Cambon, B., Legras, J. L., Wincker, P., Casaregola, S., & Dequin, S. (2009). Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(38), 16333–16338. <https://doi.org/10.1073/pnas.0904673106>
- Ono, J., Greig, D., & Boynton, P. J. (2020). Defining and Disrupting Species Boundaries in *Saccharomyces*. *Annual Review of Microbiology*, *74*, 477–495. <https://doi.org/10.1146/annurev-micro-021320-014036>
- Ortiz-Merino, R. A., Kuanyshev, N., Braun-Galleani, S., Byrne, K. P., Porro, D., Branduardi, P., & Wolfe, K. H. (2017). Evolutionary restoration of fertility in an interspecies hybrid yeast, by whole-genome duplication after a failed mating-type switch. *PLoS Biology*, *15*(5), 1–25. <https://doi.org/10.1371/journal.pbio.2002128>
- Pankajam, A. V., Dash, S., Saifudeen, A., Dutta, A., & Nishant, K. T. (2020). Loss of heterozygosity and base mutation rates vary among *Saccharomyces cerevisiae* hybrid strains. *G3: Genes, Genomes, Genetics*, *10*(9), 3309–3319. <https://doi.org/10.1534/g3.120.401551>
- Pardo-Diaz, C., Salazar, C., Baxter, S. W., Merot, C., Figueiredo-Ready, W., Joron, M., McMillan, W. O., & Jiggins, C. D. (2012). Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genetics*, *8*(6). <https://doi.org/10.1371/journal.pgen.1002752>
- Pérez-Torrado, R., González, S. S., Combina, M., Barrio, E., & Querol, A. (2015). Molecular and enological characterization of a natural *Saccharomyces uvarum* and *Saccharomyces cerevisiae*

- hybrid. *International Journal of Food Microbiology*, 204, 101–110. <https://doi.org/10.1016/j.ijfoodmicro.2015.03.012>
- Peris, D., Arias, A., Orlić, S., Belloch, C., Pérez-Través, L., Querol, A., & Barrio, E. (2017). Mitochondrial introgression suggests extensive ancestral hybridization events among *Saccharomyces* species. *Molecular Phylogenetics and Evolution*, 108, 49–60. <https://doi.org/10.1016/j.ympev.2017.02.008>
- Peris, D., Lopes, C. a, Belloch, C., Querol, A., & Barrio, E. (2012). Comparative genomics among *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* natural hybrid strains isolated from wine and beer reveals different origins. *BMC Genomics*, 13(1), 407. <https://doi.org/10.1186/1471-2164-13-407>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates Species-wide genetic and phenotypic diversity. *Nature*. 556(7701):339-344.
- Piotrowski, J. S., Nagarajan, S., Kroll, E., Stanbery, A., Chiotti, K. E., Kruckeberg, A. L., Dunn, B., Sherlock, G., & Rosenzweig, F. (2012). Different selective pressures lead to different genomic outcomes as newly-formed hybrid yeasts evolve. *BMC Evolutionary Biology*, 12(1). <https://doi.org/10.1186/1471-2148-12-46>
- Pontes, A., Čadež, N., Gonçalves, P., & Sampaio, J. P. (2019). A Quasi-Domesticated Relic Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxus* Adapted to Olive Brine. *Frontiers in Genetics*, 10(MAY), 1–14. <https://doi.org/10.3389/fgene.2019.00449>
- Prüfer, K., De Filippo, C., Grote, S., Mafessoni, F., Korlević, P., Hajdinjak, M., Vernot, B., Skov, L., Hsieh, P., Peyrégne, S., Reher, D., Hopfe, C., Nagel, S., Maricic, T., Fu, Q., Theunert, C., Rogers, R., Skoglund, P., Chintalapati, M., ... Pääbo, S. (2017). A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science*, 358(6363), 655–658. <https://doi.org/10.1126/science.aao1887>
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., De Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., ... Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, 505(7481), 43–49. <https://doi.org/10.1038/nature12886>
- Racimo, F., Gokhman, D., Fumagalli, M., Ko, A., Hansen, T., Moltke, I., Albrechtsen, A., Carmel, L., Huerta-Sanchez, E., & Nielsen, R. (2017). Archaic adaptive introgression in TBX15/WARS2. *Molecular Biology and Evolution*, 34(3), 509–524. <https://doi.org/10.1093/molbev/msw283>
- Racimo, F., Sankararaman, S., Nielsen, R., & Huerta-Sánchez, E. (2015). Evidence for archaic adaptive introgression in humans. *Nature Reviews Genetics*, 16(6), 359–371. <https://doi.org/10.1038/nrg3936>

- Rajiv McCoy, A. C., Wakefield, J., Akey Correspondence, J. M., McCoy, R. C., & Akey, J. M. (2017). Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human Gene Expression Neanderthal Modern Human Introgression Allele-Specific Expression Phenotypic Variation Theory Impacts of Neanderthal-Introgressed Sequences on the Landscape of Human G. *Cell*, 168(23), 916–927. <http://dx.doi.org/10.1016/j.cell.2017.01.038>
- Ramazzotti, M., Stefanini, I., Di Paola, M., De Filippo, C., Rizzetto, L., Berná, L., Dapporto, L., Rivero, D., Tocci, N., Weil, T., Lenucci, M. S., Lionetti, P., & Cavalieri, D. (2019). Population genomics reveals evolution and variation of *Saccharomyces cerevisiae* in the human and insects gut. *Environmental Microbiology*, 21(1), 50–71. <https://doi.org/10.1111/1462-2920.14422>
- Reich, D., Green, R. E., Kircher, M., Krause, J., Patterson, N., Durand, E. Y., Viola, B., Briggs, A. W., Stenzel, U., Johnson, P. L. F., Maricic, T., Good, J. M., Marques-Bonet, T., Alkan, C., Fu, Q., Mallick, S., Li, H., Meyer, M., Eichler, E. E., ... Pääbo, S. (2010). Genetic history of an archaic hominin group from Denisova cave in Siberia. *Nature*, 468(7327), 1053–1060. <https://doi.org/10.1038/nature09710>
- Reuter, M., Bell, G., & Greig, D. (2007). Increased outbreeding in yeast in response to dispersal by an insect vector. *Current Biology*, 17(3), 81–83. <https://doi.org/10.1016/j.cub.2006.11.059>
- Green R., Krause J., Briggs A., Maricic T., Stenzel U., Kircher M., Patterson N., Li H., Zhai W., Hsi-Yang Fritz M., Hansen N. (2016). A Draft sequence of the Neandertal Genome. *Science*. 118(24), 6072–6078. <https://doi.org/10.1126/science.1188021.A>
- Rogers, D. W., McConnell, E., Ono, J., & Greig, D. (2018). Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast. *PLoS Biology*, 16(11), 1–17. <https://doi.org/10.1371/journal.pbio.2005066>
- Ropars, J., Rodríguez De La Vega, R. C., López-Villavicencio, M., Gouzy, J., Sallet, E., Dumas, É., Lacoste, S., Debuchy, R., Dupont, J., Branca, A., & Giraud, T. (2015). Adaptive horizontal gene transfers between multiple cheese-associated fungi. *Current Biology*, 25(19), 2562–2569. <https://doi.org/10.1016/j.cub.2015.08.025>
- Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, 507(7492), 354–357. <https://doi.org/10.1038/nature12961>
- Scannell, D. R., Zill, O. A., Rokas, A., Payen, C., Dunham, M. J., Eisen, M. B., Rine, J., Johnston, M., & Hittinger, C. T. (2011). The awesome power of yeast evolutionary genetics: New genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3: Genes, Genomes, Genetics*, 1(1), 11–25. <https://doi.org/10.1534/g3.111.000273>
- Segond, D., Friedrich, A., Guezenec, S., Huyghe, L., Agier, N., Nidelet, T., Sicard, D., Supagro, M., & Biology, Q. (2020). *Keywords*.
- Shapira, R., Levy, T., Shaked, S., Fridman, E., & David, L. (2014). Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models. *Heredity*, 113(4), 316–326. <https://doi.org/10.1038/hdy.2014.33>

- Simonti, C. N., Vernot, B., Bastarache, L., Bottinger, E., Carrell, D. S., Chisholm, R. L., Crosslin, D. R., Hebbbring, S. J., Jarvik, G. P., Kullo, I. J., Li, R., Pathak, J., Ritchie, M. D., Roden, D. M., Verma, S. S., Tromp, G., Prato, J. D., Bush, W. S., Akey, J. M., ... Capra, J. A. (2016). The phenotypic legacy of admixture between modern humans and Neandertals. *Science*, *351*(6274), 737–741. <https://doi.org/10.1126/science.aad2149>
- Slon, V., Hopfe, C., Weiß, C. L., Mafessoni, F., Rasilla, M. De, Lalueza-Fox, C., Rosas, A., Soressi, M., Knul, M. V., Miller, R., Stewart, J. R., Derevianko, A. P., Jacobs, Z., Li, B., Roberts, R. G., Shunkov, M. V., de Lumley, H., Perrenoud, C., Gusic, I., ... Meyer, M. (2017). Pleistocene sediments. *Science*, *608*(May), 605–608.
- Slon, V., Mafessoni, F., Vernot, B., Filippo, C. De, Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Kelso, J., Meyer, M., Prüfer, K., & Pääbo, S. (2019). The genome of the offspring of a Neandertal mother and a Denisovan father. *Nature*. *561*(7721), 113–116. <https://doi.org/10.1038/s41586-018-0455-x>
- Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. *Molecular Biology and Evolution*, *34*(7), 1596–1612. <https://doi.org/10.1093/molbev/msx098>
- Stefanini, I., Dapporto, L., Berní, L., Polsinelli, M., Turillazzi, S., & Cavalieri, D. (2016). Social wasps are a *Saccharomyces* mating nest. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), 2247–2251. <https://doi.org/10.1073/pnas.1516453113>
- Stelkens, R. B., Brockhurst, M. A., Hurst, G. D. D., & Greig, D. (2014). Hybridization facilitates evolutionary rescue. *Evolutionary Applications*, *7*(10), 1209–1217. <https://doi.org/10.1111/eva.12214>
- Stelkens, R. B., Miller, E. L., & Greig, D. (2016). Asynchronous spore germination in isogenic natural isolates of *Saccharomyces paradoxus*. *FEMS Yeast Research*, *16*(3), 1–11. <https://doi.org/10.1093/femsyr/fow012>
- Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., Dietrich, F. S., & McCusker, J. H. (2015). The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Research*, *125*(5), 762–774. <https://doi.org/10.1101/gr.185538.114>
- Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C. B., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, *25*(11), 2427–2442. <https://doi.org/10.1111/mec.13539>
- Suarez-Gonzalez, A., Lexer, C., & Cronk, Q. C. B. (2018). Adaptive introgression: A plant perspective. *Biology Letters*, *14*(3). <https://doi.org/10.1098/rsbl.2017.0688>
- Sun, Y., Corcoran, P., Menkis, A., Whittle, C. A., Andersson, S. G. E., & Johannesson, H. (2012). Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. *PLoS Genetics*, *8*(7). <https://doi.org/10.1371/journal.pgen.1002820>

- Tattini, L., Tellini, N., Mozzachiodi, S., D'Angiolo, M., Loeillet, S., Nicolas, A., & Liti, G. (2019). Accurate Tracking of the Mutational Landscape of Diploid Hybrid Genomes. *Molecular Biology and Evolution*, 36(12), 2861–2877. <https://doi.org/10.1093/molbev/msz177>
- Taylor, S. A., & Larson, E. L. (2019). Importance and Prevalence of Hybridization in Nature. *Nature Ecology & Evolution*, 3(February), 170–177. <http://dx.doi.org/10.1038/s41559-018-0777-y>
- Tsai, I. J., Bensasson, D., Burt, A., & Koufopanou, V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 105(12), 4957–4962. <https://doi.org/10.1073/pnas.0707314105>
- Varela, J. A., Puricelli, M., Ortiz-Merino, R. A., Giacomobono, R., Braun-Galleani, S., Wolfe, K. H., & Morrissey, J. P. (2019). Origin of Lactose Fermentation in *Kluyveromyces lactis* by Interspecies Transfer of a Neo-functionalized Gene Cluster during Domestication. *Current Biology*, 29(24), 4284–4290.e2. <https://doi.org/10.1016/j.cub.2019.10.044>
- Vattathil, S., & Akey, J. M. (2015). Small Amounts of Archaic Admixture Provide Big Insights into Human History. *Cell*, 163(2), 281–284. <https://doi.org/10.1016/j.cell.2015.09.042>
- Vaughan Martini, A., & Martini, A. (1987). Three newly delimited species of *Saccharomyces sensu stricto*. *Antonie van Leeuwenhoek*, 53(2), 77–84. <https://doi.org/10.1007/BF00419503>
- Vernot, B., & Akey, J. M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. *Science*, 343(6174), 1017–1021. <https://doi.org/10.1126/science.1245938>
- Wei, W., McCusker, J. H., Hyman, R. W., Jones, T., Ning, Y., Cao, Z., Gu, Z., Bruno, D., Miranda, M., Nguyen, M., Wilhelmy, J., Komp, C., Tamse, R., Wang, X., Jia, P., Luedi, P., Oefner, P. J., David, L., Dietrich, F. S., ... Steinmetz, L. M. (2007). Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proceedings of the National Academy of Sciences of the United States of America*, 104(31), 12825–12830. <https://doi.org/10.1073/pnas.0701291104>
- Wolf, A. B., & Akey, J. M. (2018). Outstanding questions in the study of archaic hominin admixture. *PLoS Genetics*, 14(5), 1–14. <https://doi.org/10.1371/journal.pgen.1007349>
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Zeberg, H., & Pääbo, S. (2020). *The major genetic risk factor for severe COVID-19 is inherited from Neandertals*. <https://doi.org/10.1101/2020.07.03.186296>
- Zhang, H., Skelton, A., Gardner, R. C., & Goddard, M. R. (2010). *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: Evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Research*, 10(7), 941–947. <https://doi.org/10.1111/j.1567-1364.2010.00681.x>
- Zhang, W., Dasmahapatra, K. K., Mallet, J., Moreira, G. R. P., & Kronforst, M. R. (2016). Genome-

wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology*, 17(1). <https://doi.org/10.1186/s13059-016-0889-0>

Chapter 3

The role of telomere diversity in genome evolution and the role of evolution in shaping telomeres

3.1) What are telomeres?

Telomeres are ribonucleoprotein structures situated at the end of chromosomes (**Figure 1**). Their DNA structure was first described in 1978 by Elizabeth Blackburn and Joseph Gall in the model organism *Tetrahymena thermophila* (Blackburn & Gall, 1978). They are constituted by tandem repeats of a DNA sequence whose motif and length varies across species. Their

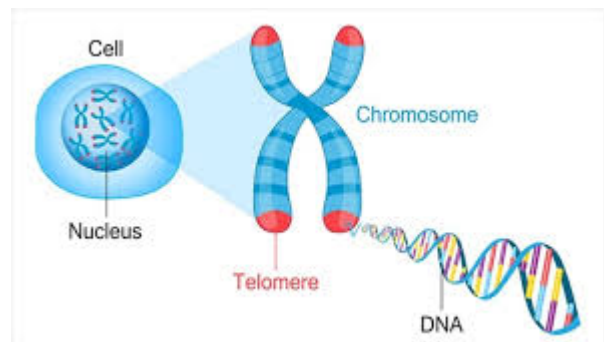


Figure 1: Image of telomeres inside the nucleus of a eukaryotic cell.

essential functions are to stabilize chromosomes, to distinguish chromosome-ends from double-stranded breaks (DSBs), as well as to guarantee the correct segregation of chromosomes at the opposite poles during cell divisions. The presence of telomeres is an effective solution to solve the so-called “end replication problem” (Watson, 1972; Olovnikov, 1973). In fact, at each cell division the conventional DNA replication machinery fails to completely duplicate the parental DNA extremities. This results in the progressive loss of terminal repeats. Telomeres act as a sort of buffer preventing internal genes to be deleted as a result of the chromosome end resection. Moreover, they provide a protective cap to linear chromosomes, and consequently chromosomes with dysfunctional telomeres are highly unstable and tend to be lost (Lundblad & Szostak, 1989; Shampay et al., 1984; Szostak & Blackburn, 1982).

In the yeast *S. cerevisiae*, telomeres are constituted by the tandem repetition of the degenerate sequence $C_{1-3}A/TG_{1-3}$ and have a length of approximately 300+-75 bp (Wellinger & Zakian, 2012). Telomeres are double-stranded for most of their length, but they terminate with a single-stranded 3' G-rich tail, which is only 12 to 15 nucleotides long (Larrivé et al., 2004). As a ribonucleoprotein structure, telomeres are constituted by DNA, RNA and a series of associated proteins providing a peculiar chromatin structure protecting the chromosome ends from unwanted DNA damage response and serving as initiating sites for heterochromatin formation (**Figure 2**). The telomere-associated proteins can be subdivided in two groups according to the regions in which they bind: the proteins binding to the double-stranded area or the ones binding to the G tail.

Double-stranded area proteins – The most important protein of the double-stranded area is Rap1p. Rap1p binding sites are found more or less every 20 bp and we expect to have one bound Rap1p molecule every 18⁺-4 bp (Gilson et al., 1993). Rap1p has very high affinity for telomeric sequences (Conrad et al., 1990; Lustig et al., 1990) and the number of Rap1p molecules present at telomeric sites at a given time (together with Rif1/Rif2) functions as a sort of “counting machinery” to regulate the homeostasis of telomere length (Marcand et al., 1997). Rap1p binding sites are not only found at telomeres but are distributed across the whole genome and are quite heterogeneous in sequence. Around 300 genomic loci are thought to be bound by Rap1p (Lieb et al., 2001). In addition to its telomeric functions, Rap1p acts as an activator and a repressor of gene expression (Shore & Nasmyth, 1987). Interestingly, Rap1p regulates the expression of a network of genes involved in metabolism, translation and stress response (Ye et al., 2014) and behaves as a stress response factor by coordinating its spectrum of regulated genes to starvation, genotoxic exposure and telomere shortening (Buck & Lieb, 2009; Platt et al., 2013; Tomar et al., 2008). Partial inactivation of Rap1p results in telomere shortening and chromosome fusions (Conrad et al., 1990; Lustig et al., 1990; Marcand et al., 2008). Rap1p functions as a hub to recruit other important proteins to telomeres, which are not directly bound to the telomeric DNA: Rif proteins, the Yku complex and the Sir complex.

Rif1 and Rif2 associate directly with the Rap1 C-terminal domain (RCT). Rif1 and Rif2 prevent telomeres from being erroneously recognized as DSBs and contribute to telomerase regulation. Together with Rap1p, Rif1 and Rif2 assure that telomeres maintain a regulated equilibrium size and their deletion results in telomere lengthening. The double mutant *rif1/rif2* exhibits additional telomere lengthening than the single mutants, indicating that the two proteins are part of two separate telomere length regulation pathways (Wotton & Shore, 1997). Rif2, but not Rif1, inhibits the non-homologous end joining pathway (NHEJ) and the access of the Tel1/MRX complex to telomeres. Instead, Rif1 is needed to maintain cells viable in the absence of other important telomere-associated proteins, like the CST complex (Kupiec, 2014).

The Yku complex is composed of Yku70 and Yku80. Both proteins are involved in the DNA repair pathway of non-homologous end joining (Gilson et al., 1993; Palladino et al., 1993), but counterintuitively, their presence at telomeric sites is essential for telomere maintenance and the tethering of telomeres to the nuclear envelope (Laroche et al., 1998). The Yku complex is recruited at telomeres indirectly through the association with other telomere-binding proteins, but is also directly bound to telomeric DNA (Martin et al., 1999; Roy et al., 2004). The separate roles of the

Yku complex in NHEJ and telomere maintenance were established by separation of function mutations (Lopez et al., 2011; Ribes-Zamora et al., 2007). The correct interaction between Yku70 and Yku80 is essential for their role, as it has been observed that some allelic combinations do not perform well together and result in very short telomeres (Liti et al., 2009).

The Sir complex comprises three proteins: Sir2, Sir3 and Sir4. Among them, Sir2 is a NAD-dependent histone deacetylase. The Sir complex is recruited to telomeres by interactions with Rap1 where it is responsible for the initiation and spreading of heterochromatin in subtelomeres. This results in the transcriptional repression of genes found up to 15 kb from the telomere ends (Pryde & Louis, 1999), a phenomenon called “telomere position effect” (TPE) (Gottschling et al., 1990). In addition to its silencing role, the Sir complex is also involved in the spatial organization of telomeres within the nucleus through the interaction with nuclear envelope proteins (Hediger et al., 2002).

Single-stranded G-tail proteins – The single-stranded G-tail is bound by the CST complex, composed of three proteins: Cdc13, Stn1 and Ten1. This complex is structurally similar to the Replication Protein A complex (RPA), which is also a single-stranded DNA binder. However, RPA is involved in DNA replication and DNA repair and its function at telomeres is unknown, although it can be detected there during DNA replication (Grandin & Charbonneau, 2013; Luciano et al., 2012; Schramke et al., 2004). Functional subdomains can be readily interchanged between the RPA and CST complex without any phenotypic effect (Gao et al., 2007; Gelinis et al., 2009).

Inactivation of Cdc13 results in cell-cycle arrest and cell death, but overexpression of the other subunits, Stn1 and Ten1, can rescue this phenotype (Petreaca et al., 2006; Sun et al., 2009). Interestingly, Cdc13 can also be recruited at internal DSB sites and promote the addition of telomeric repeats there (Mandell et al., 2011).

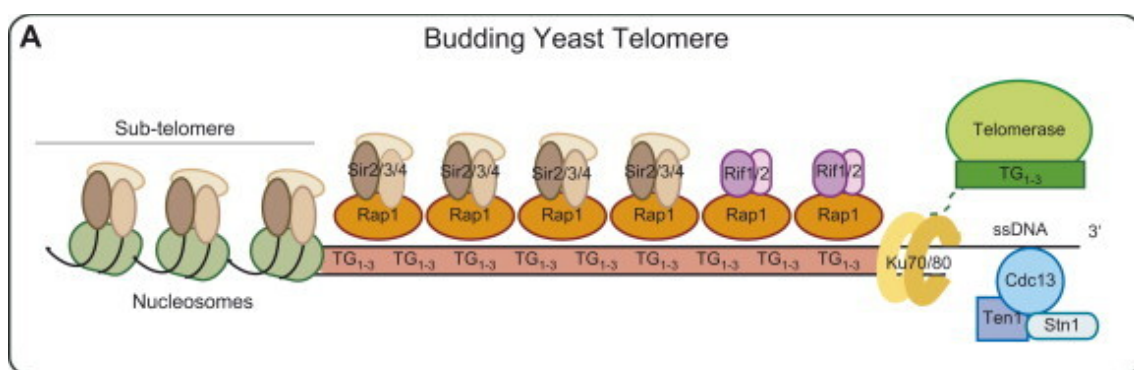


Figure 2: Telomeric structure in the budding yeast *S. cerevisiae*. Adapted from Auriche et al., 2008.

3.2) Telomere maintenance and replication

Most telomeric DNA is replicated by the classical DNA replication machinery and telomeres are replicated very late during the S phase of the cell cycle, probably due to the late activation of replication origins (ARS) near telomeres (Ferguson & Fangman, 1992). The late activation is not a consequence of the sequence of the ARS, as it is demonstrated by the fact that placing an early activating ARS near a telomere results in its late activation, but rather of the telomere exerting a position effect on the activation timing (Ferguson & Fangman, 1992; Wellinger et al., 1993; Wellinger & Zakian, 2012). The timing of firing is also influenced by the length of the telomere, in fact ARS associated to short telomeres fire earlier than the ones associated to long telomeres (Bianchi & Shore, 2007).

DNA replication is carried out by DNA polymerases, which can synthesize DNA only in the 5'-3' direction and require a RNA primer to start the synthesis. The RNA primer is usually 8-12 nt long and needs to be removed once the replication is over. This leads to the loss of 8-12 nt at the terminus of the DNA molecule at each DNA replication, and was initially dubbed as the “end replication problem” (Watson, 1972; Olovnikov, 1973). The discovery of the 3' overhang natural structure of chromosomal DNA ends implied to revisit this lagging strand problem into a leading strand end replication problem due to the inability of the DNA polymerase to reproduce on the leading strand the 3' overhang of the parental DNA (Lingner et al., 1995). However, nature invented an interesting mechanism to circumvent this issue: the specialised enzyme telomerase. Telomerase was first discovered in the ciliate *Tetrahymena* by Elizabeth Blackburn and Carol Greider (Greider & Blackburn, 1985, 1987). Thanks to this discovery, Blackburn, Greider and Jack Szostak were awarded the 2009 Nobel Prize in Medicine. Telomerase is a reverse transcriptase which uses a RNA moiety as template. The RNA template's sequence is complementary to the G-tail of telomeres, and binds to that. After that, telomerase extends the G-tail and then conventional DNA polymerases fill in the gap by synthesizing the complementary DNA (Greider & Blackburn, 1989).

Telomerase is a protein complex composed by three subunits (EST1, EST2, and EST3) plus the template RNA TLC1 (**Figure 3**). EST2 is the catalytic reverse transcriptase and the central core of the complex. Its function is highly dependent on its basic N-terminal domain (TEN) which mediates interaction with the other subunits and with the RNA template (Friedman & Cech, 1999; Friedman et al., 2003; Talley et al.,

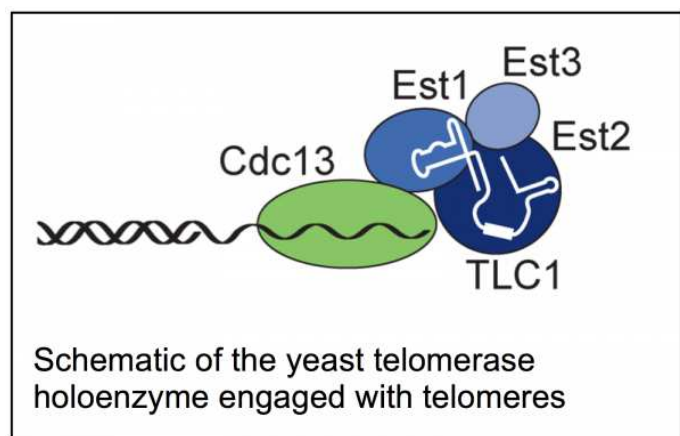


Figure 3: Schematic of the yeast telomerase holoenzyme.

2011). EST1 is a subunit that mediates the interaction between the enzymatic subunit (EST2) and the telomere. It is shown in fact that EST1 is able to bind both RNA and single-stranded TG₁₋₃ sequences *in vitro* (DeZwaan & Freeman, 2009; Virta-pearlman & Lundblad, 1996). EST1 is not essential, as it has been shown that fusion proteins between CDC13 and EST2 can guarantee the extension of telomeres without the need of EST1, and suggests that its function is to bring the enzymatic subunit to its target (Evans & Lundblad, 1999). In addition, the same study also evidenced that EST1 stimulates the activity of telomerase, as the strains carrying the fusion protein had abnormally elongated telomeres, probably due to the constant presence of telomerase at chromosome ends (Evans, 1999). The function of EST3 in the holoenzyme is not completely clear, but it has probably a regulatory activity on telomerase. EST3 is only present in budding yeast, but not in mammals, and its role in telomere homeostasis is essential *in vivo* but not *in vitro* (Lendvay et al., 1996; Lingner et al., 1997).

The last component of the telomerase holoenzyme is the RNA template, which is encoded by the gene TLC1. Its transcription is carried out by RNA polymerase II and generates two types of transcripts: a polyadenylated one and a non-polyadenylated one. The non-polyadenylated version is the one present in telomerase (Bosoy et al., 2003; Chapon et al., 1997). Yeast TLC1 RNA is very long (>1000 nt) but the template region used by telomerase is much shorter (only 16 nt), and, however, it has been shown that a RNA of 1/3 of the total length is sufficient to guarantee the maintenance of telomeres, suggesting that most of the RNA is not essential (Singer & Gottschling, 1994; Zappulla et al., 2005).

Telomerase activity is regulated by the cell cycle and by telomere length itself (Marcand et al., 1997). Short telomeres, in fact, are preferentially elongated by telomerase (Marcand et al., 1999). Experiments carried out in yeast diploid cells showed that telomerase does not act at every telomere at every cell cycle, but its probability of elongating a telomere increases as the telomere gets shorter (Teixeira et al., 2004). Overall, the regulated mean telomere length of dividing yeast cells results from a similar rate between telomere shortening and lengthening. During the return to regulated length of an abnormally shortened telomere, the probability of telomerase to act at telomeres decreases progressively with increasing telomere length while during the return to the regulated length of abnormally lengthened telomeres, the probability of telomerase to act at telomeres increases progressively with decreasing telomere length.

3.3) Telomere length is a complex trait

Telomere length is a complex and quantitative trait, that is shaped by the combination of both genetic and environmental determinants. Its rate of genetic heritability is estimated between 34% and 50% in humans. The first telomere length genes to be identified were the subunits of the

telomerase holoenzyme: EST (Lendvay et al., 1996; Lundblad & Szostak, 1989). Their name derives from the phenotype of their deletion mutants. EST stands for “Ever Shorter Telomeres” and was coined because EST mutants are viable but have extremely short telomeres and chromosomal instability, and most of them die within 50-100 generations unless they are able to lengthen their telomeres through alternative mechanisms.

The yeast genome contains around 6000 genes, among which 4700 are non-essential and can be deleted without causing the death of the organism. This deletion collection (YKOC – Yeast KnockOut Collection) was used by two independent studies to find genes involved in telomere length maintenance (TLM) (Askree et al., 2004; Gatbonton et al., 2006). In these studies, DNA was extracted from the mutants and submitted to Southern blot using the restriction enzyme *XhoI*, which cuts in a fixed site at the end of a Y' subtelomeric repetitive element and generates a telomeric restriction fragment (TRF). All genes which caused a shift in the migration of the TRF (reflecting a change in telomere length) were annotated as TLM genes. All together, these studies identified more than 270 genes involved in TLM. Among these, some of them shorten telomeres while others lengthen them. The collection of TLM genes comprises almost every cellular pathway, including DNA, RNA and chromatin metabolism, cellular trafficking, protein turnover and organelles functions, indicating that telomere biology is at the very center of the cellular life (Harari & Kupiec, 2014). Interestingly, some genes did not affect telomere length directly, but their action resulted from the perturbation of the expression of a neighboring gene. This has been observed by complementation assays in which the reinsertion of the knocked-out gene through a plasmid did not rescue the phenotype, but the insertion of its neighbor brought back the telomere length to its normal state (Ben-Shitrit et al., 2012). In addition, some genes exerted their effect by modulating the level of the RNA template (Mozdy et al., 2008).

The YKOC was later amplified to contain a subset of essential genes which had been made hypomorphic, and the screening of these genes revealed additional 87 TLM genes (Ungar et al., 2009). More recently, new methods have been set up which allow an estimation of telomere length from whole-genome sequencing data (Ding et al., 2014; Farmery et al., 2018; Feuerbach et al., 2019; Lee et al., 2017; Nersisyan & Arakelyan, 2015). These methods all rely on the assumption that a higher number of telomeric-repeat-containing reads translates into longer telomeres. The comparison of the number of these reads among strains can therefore give an estimate of their telomere length. The complete YKOC was sequenced by NGS technologies in 2019 and this study identified new TLM genes, further validated by traditional Southern blot (Puddu et al., 2020). Even though these studies have provided invaluable knowledge about the nature of the cellular processes influencing telomere length, their power is limited due to the inability to assess the effect of combinations of multiple gene deletions/inactivations. Moreover, the deletion of an entire gene

limits the resolution of these studies, as it does not tell us anything about the effect of single polymorphisms inside the gene itself.

Telomere length is not only controlled by genes, but is determined also by environmental factors. In yeast, it is known that starvation shortens telomeres, and a similar phenotype is observed treating cells with the TORC1 inhibitor rapamycin. This effect is mediated by the Yku complex (Ungar et al., 2011). Other environmental factors can have contrasting effects on telomeres. For instance, ethanol and acetic acid elongate telomeres, while caffeine and heat shorten them (Romano et al., 2013). The effect of ethanol is mediated by Rif1 and the Nonsense mediated decay pathway (NMD), which regulates the amount of Stn1 and other RNA molecules in the cell. The effect of caffeine instead is mediated by the Tel1/Mec1 DNA-damage response pathway.

3.4) The structure of subtelomeres in yeast

Subtelomeres are chromosomal regions which are rich in repetitive sequences called TAS (telomere-associated sequences). These include the X and Y' elements, which differ in terms of position and frequency at chromosome-ends. The X element is present in the majority of telomeres and is centromere-proximal to the Y'. The X element is very heterogeneous in sequence, although it contains a conserved sequence called the “core X”, which is about 475 bp long. Its size ranges from 300 bp to 3 kb. The Y' element is located next to the telomeric repeats and its sequence is much more conserved than that of the X element. It is present in about 70% of the telomeres and can be found as a single copy or as tandem repeats of multiple copies. The identity of telomeres lacking Y' element can vary from strain to strain (Horowitz et al., 1984). The Y' elements are divided in two types depending on their size: Y'-short (5.2 kb) and Y' long (6.7 kb). The two types differ by the presence of insertions/deletions in their sequence (Chan & Tye, 1983; Louis & Haber, 1992).

The paradigm of the yeast subtelomeric structure composed by one X element present at all chromosomal ends and possibly one or more Y' elements has been recently revised. The complete genome assembly of 12 *Saccharomyces* strains revealed the existence of unusual subtelomeric structures deprived of the usual X element or containing more than one copy of it (Yue et al., 2017). Subtelomeres in nature are therefore more variable than what we thought (**Figure 4**).

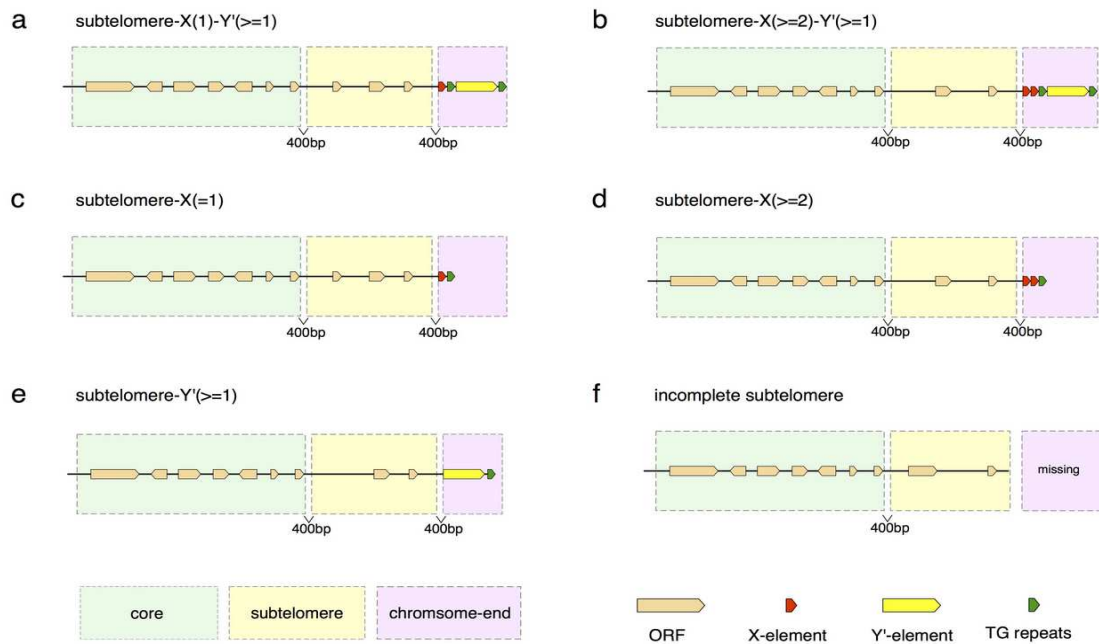


Figure 4: Observed subtelomeric structures in *Saccharomyces* yeasts. Adapted from Yue et al., 2017.

Both X and Y' sequences contain binding sites for transcription factors (Mak et al., 2009) and can act as insulators which block the spreading of heterochromatin and as silencing relays (protosilencers) regulating the spreading extent of the heterochromatin emanating from telomeres (Fourel et al., 1999). Most of the subtelomeric DNA is organized in nucleosomes, while the core X is depleted of histones and is transcriptionally repressed. On the contrary, the Y' element contains histones and is actively expressed (Zhu & Gustafsson, 2009). Thus, the transcriptional profile of a strain can change depending on its subtelomeric repeat content which regulates the expression of nearby genes.

The regions between the X and Y' elements can contain interstitial telomeric sequences (ITS) (Claussin & Chang, 2015). These are stretches of telomeric repeats which are virtually indistinguishable from the actual telomeres, except for their position (Walmsley et al., 1984). ITS are always associated and immediately preceding a Y' element. It has been observed that ITS are very prone to expand and retract, they constitute fragile sites in the genome and can promote chromosomal rearrangements and other types of structural variations (Aksenova et al., 2013, 2015). In addition, they are able to interact with telomeres to generate intrachromosomal recombinations and can modulate the expression of nearby genes through the recruitment of telomere-bound proteins (reviewed in (Aksenova & Mirkin, 2019)). ITS expansion, accompanied by that of Y' elements, is used as an alternative pathway to compensate for the lack of telomerase. Alternative lengthening of telomeres (ALT) is a mechanism widely used from yeasts to mammals to overcome the absence of an active telomerase enzyme, and it relies on

recombination. The existence of such mechanisms was first postulated when rare *est1* KO cells escaped death (**Figure 5**). These rare cells were called survivors and analysis of their genome revealed that they could be of two possible types: type I and type II survivors (Lundblad & Blackburn, 1993; Teng & Zakian, 1999). Type I survivors contain multiple tandem Y' sequences intertwined by ITS, but their telomeres are very short (Larrivé & Wellinger, 2006; Lundblad & Blackburn, 1993). The appearance of type I survivors is strictly dependent upon recombination pathways, and therefore cannot happen when RAD genes are inactivated (Chen et al., 2001; Le et al., 1999).

Type II survivors do not contain amplifications of subtelomeric elements and ITS but can have both very short and extremely long telomeres in the same cell, which can arrive up to 12 kb (Teng & Zakian, 1999; Teng et al., 2000). These survivors contain extrachromosomal circles of telomeric DNA, and their unstable telomeres are maintained by rolling circle replication (Larrivé & Wellinger, 2006; Lin et al., 2005). These circles could be generated by intrachromosomal recombinations between telomeric sequences (Lustig, 2003). The onset of this type of survivors is dependent on the *MRX* complex, *RAD59* and *SGS1* (Chen et al., 2001; Huang et al., 2001; Johnson et al., 2001; Le et al., 1999; Teng et al., 2000; Wellinger & Zakian, 2012).

Although the pattern of ITS distribution is extremely variable in yeast strains, nothing is currently known about the genes and environmental factors which contribute to this diversity.

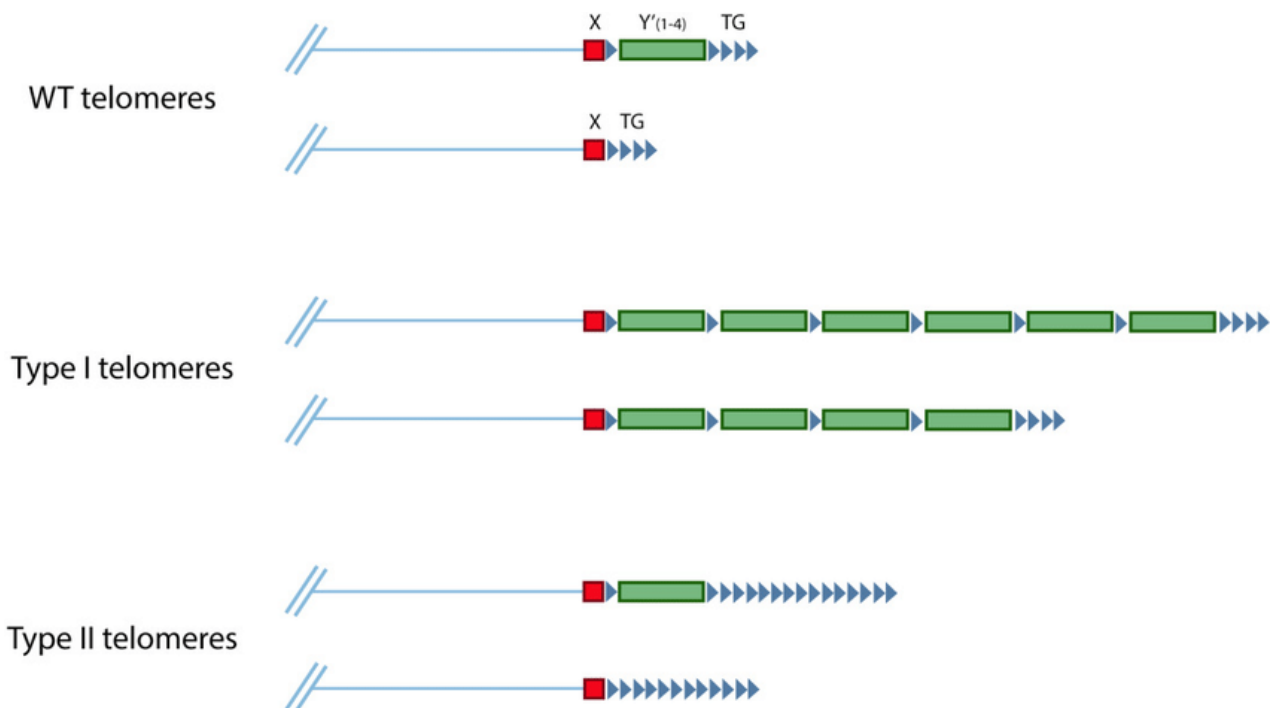


Figure 5: Schematic representation of telomeres and subtelomeres in Type I and Type II survivors. Adapted from Claussin & Chang 2015.

3.5) The evolutionary contribution of telomere dynamics

Telomeres are extremely variable among taxa. Species typically have very different average telomere lengths, and even among the same species, telomere length can vary among individuals and populations. However, telomeres are not only different in terms of length, but also in terms of sequence, and a great diversity of telomeric DNA motifs was reported among organisms. For the rest of this chapter, I will illustrate the contribution of telomere diversity to evolution, focusing first on telomere length diversity and then on telomere sequence diversity.

Telomeres are ribonucleoprotein structures located at the end of linear chromosomes, and, as such, are an integrative part of the genome and coevolve with it under the action of the same selective pressures. In parallel, it is also true that modifications of the sequence, length and structure of telomeres can have profound consequences on the rest of the genome. In this sense, the contribution of telomere dynamics to genome evolution can be seen by two different angles. On one side we can consider telomeres as whatever other part of the genome and describe the factors that lead to their modification in space and time: in this sense telomeres are the objects on which selection acts upon. On another side, the state of telomeres, and in particular their dysfunction, is known to provoke other genomic alterations: in this other sense telomeres are the drivers of genome evolution.

3.5.1) The evolutionary contribution of telomere length dynamics

3.5.1.1) Telomeres as objects of evolution

In many adult vertebrates, in contrast to the embryonic life, telomerase expression is low in somatic cells with a notable higher expression in stem cells that is, nevertheless, insufficient to fully replenish telomere length during cell division but can extend the proliferative capacities of these cells. As a consequence, at each cell division mammalian telomeres get progressively shortened and, as soon as telomere length reaches a critical threshold, the cell perceives the presence of an uncapped chromosome and activates the DNA damage response, which results in the permanent block of the cell cycle, a state known as “replicative senescence” (Bodnar et al., 1998). The progressive shortening of telomeres at each cell division acts as a “molecular clock” to count how many divisions a cell has undergone, and therefore how big is its mutational burden, and it is thought to have evolved as a barrier against the onset of cancer (Campisi, 1996). The length of telomeres is variable in different species. Human telomeres can range between 5-15 kb (Canela et al., 2007; Kong et al., 2013; Okuda, 2002), while mice have extremely long telomeres (>25 kb) (Zhu et al., 1998). Yeast telomeres are relatively short and range between 300+-75 bp (Wellinger & Zakian, 2012).

In humans, short telomeres have a profound impact on health and have been associated to different pathological conditions induced by the absence of a functional telomerase or telomere-bound

proteins, like dyskeratosis congenita. These syndromes appear very early in life and their phenotypical outcome is an accelerated aging associated with an extremely short lifespan (Savage & Alter, 2009).

Even though the causal role of telomere dysfunction in various age-related pathologies is clear, there are still open questions whether telomere traits, and especially the length, can influence other fitness traits. Until recently, almost all of the work regarding telomere biology was targeted at understanding the molecular mechanisms of telomere homeostasis and the link between telomere biology and the onset of disease. However, there has been a renewal of interest in investigating variation in telomere dynamics in non-model organisms and the association between telomere length and life history traits like growth, reproduction and longevity. Can we establish if selection acts on telomere length, directly or indirectly, in order to induce phenotypic benefits?

3.5.1.2) Telomere length and lifespan

Telomere length varies both among chromosomes in the same cell and among individuals of the same species, while still being maintained within a certain range which is specific of each species. It varies even more among different species. In humans, telomere length can range between 0.5-15 kb. In rodents 10-72 kb, in ungulates 7–23 kb, in dogs 12–23 kb, in lagomorphs 2–50 kb, and in yeast 100-400 bp (Monaghan, 2010).

A key question in the field of evolutionary biology is whether telomere length can predict lifespan across individuals of the same species and even among species. This hypothesis is based on the reasoning that if telomere attrition progresses constantly during a lifetime, individuals or species with an initial shorter telomere length at birth should die sooner. Much of the work trying to give an answer to this question has been done in birds, and performed in wild and captive populations. In many bird species, as in humans, telomeres shorten with age, although some species exist that show an opposite trend (Hausmann et al., 2003). Heidinger et al. confirmed that telomere length early in life can be used as a predictor of longevity. They measured telomere length in red blood cells of zebra finches (*Taeniopygia guttata*) at the nestling stage and at multiple other time-points in life, and found that telomere length at 25 days of life was positively and significantly correlated with lifespan, and individuals who lived the longest carried long telomeres at all timepoints (Heidinger et al., 2012). In another study with a similar approach but conducted on a different species, the tree swallow (*Tachycineta bicolor*), Hausmann et al. showed that individuals with longer telomeres at 1 year of age have longer survival (Hausmann et al., 2005). Other studies found similar results (Asghar et al., 2015; Barrett et al., 2013; Bize et al., 2009; Fairlie et al., 2016; Pauliny et al., 2006; Salomons et al., 2009). It is good to know, however, that another study realized with barn swallows (*Hirundo rustica*) did not find significant correlation between the two variables (Caprioli et al.,

2013). A meta-analysis of 27 studies investigating the relationship between telomere length and lifespan in non-human vertebrates confirmed that, in the majority of the species, shorter telomeres are associated with increased mortality risk (Wilbourn et al., 2018). Outliers are constituted by reptiles, which have different telomere dynamics throughout life respect to birds and humans.

Even though initial telomere length is a good biomarker for longevity, other factors can contribute to this result. In particular, the assumption that the rate of telomere attrition is constant throughout life might not be exact. Telomeres can shorten at different rates depending on genetic and environmental conditions, not only in birds but also in humans and reptiles. Noteworthy, the presence of strong stressors like predators, limited resources and oxidative stress can have an important impact (Axelsson et al., 2020; Dupoué et al., 2017). Therefore, the rate of telomere attrition can be even a better marker for longevity than initial telomere length per se. Sudyka et al. did a meta-analysis of previous literature on birds and reported a positive correlation between telomere length and lifespan across ten bird species (Sudyka et al., 2016). Tricola et al. confirmed these results on a larger scale using 19 bird species and also found that the rate of telomere loss is conserved among phylogenetic families (Tricola et al., 2018). Other studies expanded the phylogenetic range of species and confirmed that the same trend is observed also in mammals (Dantzer & Fletcher, 2015; Hausmann et al., 2003; Whittemore et al., 2019). It is important to keep in mind that while within species the initial telomere length is a predictor of longevity, this relationship does not exist anymore across species, and the only predictive variable is the telomere attrition rate (**Figure 6**) (Monaghan & Hausmann, 2006).

The fact that telomere length correlates with mortality is a strong indication that telomeres are under selective pressure (Monaghan et al., 2018), where selection operates to disfavour the reproduction of individuals with shorter telomeres, and constitutes an example of how selection can shape genome evolution.

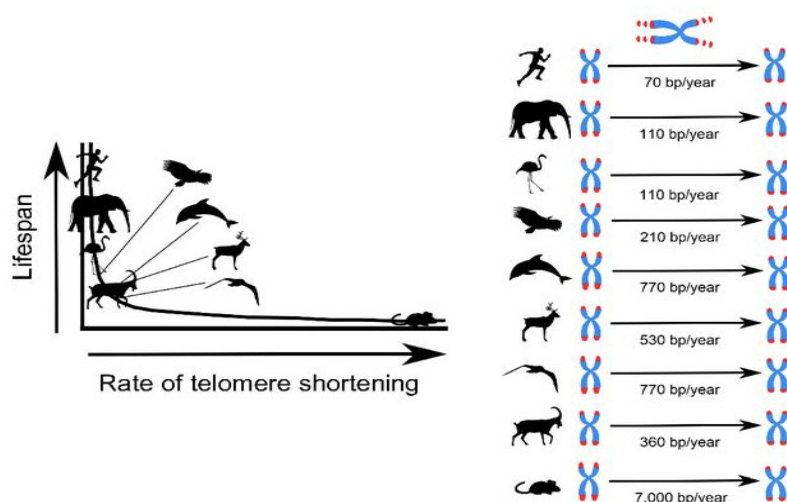


Figure 6: Relationship between telomere attrition rate and lifespan across species. Adapted from Whittemore et al., 2019.

3.5.1.3) Telomere length and growth

Species are different in the rate at which they grow, and this rate can influence their lifespan (Metcalf & Monaghan, 2001, 2003). Species with a faster growth rate usually die younger, as a result of life history trade-offs between growth and longevity. Since an organism has a limited amount of resources, it can preferentially allocate them to a pathway at the expense of another one. A classical example is when energy is devoted to growth at the expenses of self-maintenance and longevity. What are the molecular mechanisms that mediate this trade-off? A possible answer may reside in telomere maintenance.

Species vary in their body size, and this difference is primarily due to different cell division rates. Every time a cell divides it needs to replicate its genome and this brings to a loss of part of its telomeres. On top of this, cell division is an expensive process which needs energy produced by mitochondria. A high mitochondrial metabolic activity produces high levels of free radicals and reactive oxygen species, to whom telomeres are particularly sensitive (Alonso-Alvarez et al., 2007). Species like mammals and birds grow very much during early life in order to reach their final body size. In humans, that coincides more or less to sexual maturity and after that point very little growth takes place. As a consequence, the rate

of telomere loss is maximum during this first period of intense growth (Daniali et al., 2013; Geiger et al., 2012). Faster growth is associated to higher telomere loss not only early in life, but also in other periods. For example, a period of starvation resulting in reduced body weight can be followed by a period of higher growth to compensate for the earlier loss, and this is correlated with a reduced lifespan (Lee et al., 2013). This reduction in lifespan is usually accompanied by telomere length shortening (Tarry-Adkins et al., 2009).

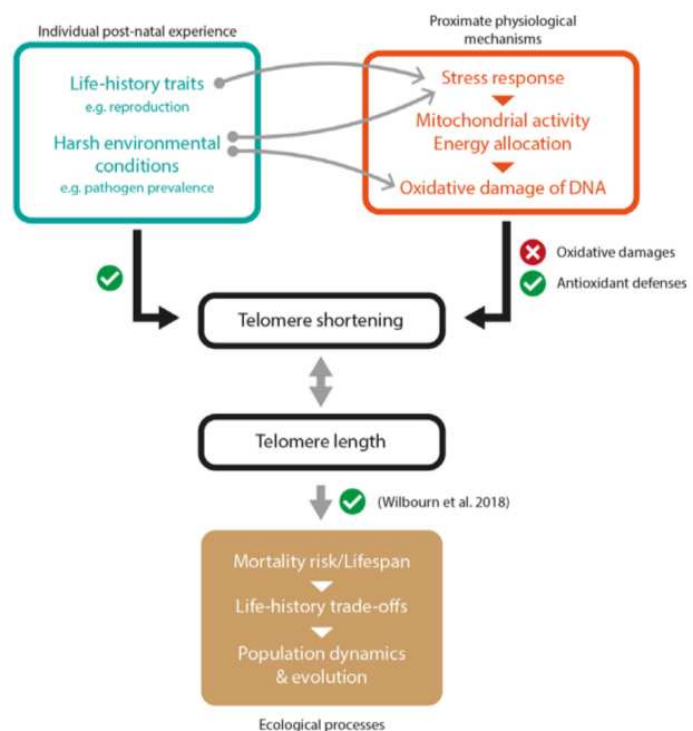


Figure 7: Relationship between telomere length, stress and ROS. Adapted from Chatelain et al., 2020.

Telomere attrition after intense growth can very likely be a consequence of the damage caused by oxidative stress on DNA, if the antioxidant cellular defenses are not sufficient to buffer its effect. Telomeres are a preferential target of ROS, because of their high content in guanine, which is more prone to oxidation and can be transformed into 8-oxo-7,8-dihydroguanine (Coluzzi et al., 2019).

Moreover, the presence of telomere-associated proteins makes it difficult for the components of the DNA repair machinery to reach telomeric DNA and fix it (Monaghan & Ozanne, 2018). At the cellular level, oxidative damage to telomeric regions can activate the DNA damage response that in turn induces cell cycle arrest and replicative senescence in cell cultures, and this effect is independent of telomere length (Fumagalli et al., 2013). Oxidative stress also actively shortens telomeres while the reduction of mitochondrially-produced ROS can reduce this attrition (Reichert & Stier, 2017; Von Zglinicki, 2002). The induction of DNA damage only at telomeric sites arrests the cell cycle, ruling out the hypothesis that replicative senescence upon oxidative stress is just the result of generic genomic damage (Sun et al., 2015).

At the organismal level, individuals with higher exposure to oxidative stress have increased telomere loss (Glade & Meguid, 2015), and circulating ROS have been correlated with telomere loss in red blood cells of king penguin chicks (*Aptenodytes patagonicus*). The administration of antioxidant molecules is able to counteract these effects: indeed, this has been observed in cell cultures, and in whole organisms when antioxidants were administered by diet (Badás et al., 2015; Prasad et al., 2017). In rats born from undernourished mothers, the administration of ubiquinone by diet prevented telomere shortening in the aorta and the heart tissues (Tarry-Adkins et al., 2013). Finally, a recent meta-analysis of 109 relevant studies revealed that telomere shortening is correlated not only with growth rate, but with environmental stressors in general (Chatelain et al., 2020). It is possible that ROS are the central hub of all these relationships, as response to stressors requires ready-to-use energy mainly generated by mitochondria (**Figure 7**).

3.5.1.4) Telomere length and reproduction

A particular and fascinating “stressor” is reproduction. Just like growth, reproduction requires energy, that will in turn not be available for other organismal needs, like survival (Westendorp & Kirkwood, 1998; Sudyka, 2019). Telomeres could constitute an interesting biomarker for reproductive success as their erosion is associated to the cumulative action of various intrinsic and extrinsic conditions, like the abundance and quality of nutrition and stress exposure. This effect goes beyond the single generation, as it has been shown that mothers which are undernourished or exposed to stress deliver



Figure 8: Image of one female (big head at the center) surrounded by many male red garter snakes during breeding. Adapted from Olsson, 2018.

babies with shorter telomeres (Send et al., 2017). Various studies confirmed the costly nature of reproduction. One study on zebra finches reported that telomeres are shorter after the breeding season, but this effect is transient and telomeres are then restored to their normal length (Heidinger et al., 2012). In other studies, the authors manipulated the brood size in order to modulate the stress on the parents, and discovered that parents caring for a bigger brood had shorter telomeres in comparison with control birds attending unaltered broods (Reichert et al., 2014; Sudyka et al., 2014).

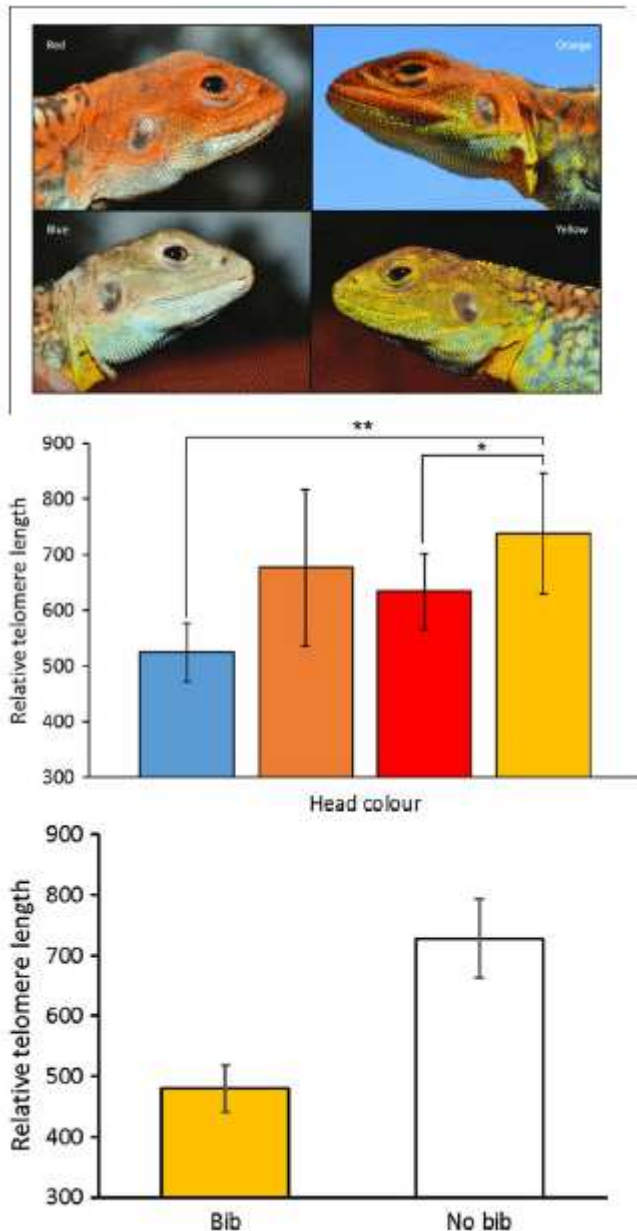


Figure 9: Relationship between telomere length and body morphs in Australian dragon lizards. Adapted from Rollings et al., 2017.

Pauliny et al., 2015; Plot et al., 2012; Ujvari et al., 2017; Ujvari & Madsen, 2009).

In Atlantic silversides (*Menidia menidia*) more fecund fish had shorter lifespans and shorter

Many insights on the trade-offs between telomere maintenance and other life-history traits comes from research on ectotherm organisms, such as reptiles. Differently from what happens in mammals and birds, where telomere length decreases in time, telomere length follows more complex dynamics in reptiles. As an example, american and chinese alligators (*Alligator mississippiensis* and *Alligator sinensis*) have a telomeric behaviour similar to endotherms, with telomeres shortening across time (Scott et al., 2006; Xu et al., 2009). Siberian sturgeons (*Acispenser baerii*) and garter snakes (*Thamnophis elegans*) showed similar behaviour (Bronikowski, 2008; Simide et al., 2016). However, other ectotherms, like sand lizards (*Lacerta agilis*), frillneck lizards (*Chlamydosaurus kingii*), Atlantic salmon (*Salmo salar*), leatherback turtles (*Dermochelys coriacea*) and water pythons (*Liasis fuscus*) show different patterns, with differences in telomere length through life which are a consequence of variation in telomerase activity (Olsson et al., 2011;

telomeres (Gao & Munch, 2015). Telomere length differences can also arise from different investment in reproduction between the sexes. Interesting examples of this phenomenon can be observed in reptiles. Red-sided garter snakes (*Thamnophis sirtalis parietalis*) show sexual dimorphism, with females being much larger than males. Females have a much more passive behaviour in terms of reproduction and just wait for males to find them and fecund them, while males engage in competition with each other for courtship. In addition, female snakes reproduce biennially, whereas males reproduce yearly. As expected, this translates in females having longer telomeres than males (**Figure 8**) (Olsson et al., 2018; Rollings et al., 2017).

The maintenance of body ornaments which are used to attract potential partners can also be considered as a cost of reproduction. In the Australian painted dragon lizard (*Ctenophorus pictus*), the skin colour is used as a marker of health and better ornaments can attract more partners. Red males defeat yellow ones in fights, while yellow males have a faster and more efficient sex and their sperms are of better quality. The presence of bibs can guarantee a better protection of partners against rivals, but is very costly. Overall, yellow-headed and bib-less males have longer telomeres than red, blue and bibbed males, as a result of less energy expenditure in reproduction (**Figure 9**) (Rollings et al., 2017).

Other studies reported different results. In a study of barn swallows (*Hirundo rustica*) telomere length was positively correlated with plumage coloration and reproductive success. An explanation of this is that telomere length reflects plumage that in turn reflects the general quality and performance of the individual (Parolini et al., 2017).

All these results are not necessarily mutually exclusive: in fact, much of the variance observed in telomere length depends on when the measure has been taken. If the measure is taken before reproduction, telomere length can be positively correlated with reproduction because its costs have not been paid yet. On the other hand, if telomere length is measured after the breeding, the true costs of reproduction can be detected. Another factor that needs to be considered is that individuals of differential quality can vary in the efficiency at which they find nutritional and other types of resources, and this may have profound consequences on telomere length upon reproduction, with individuals having better performance paying less costs or being more efficient in repairing them (Sudyka, 2019).

3.5.1.5) Telomere dynamics in health and disease

Telomere length shortens with age in human somatic cells. In humans, telomere length is typically measured in leukocytes (LTL), because erythrocytes do not contain a nucleus. It has been shown that, although there is variation in telomere length across tissues in the same individual, with highly proliferative tissues having shorter telomeres, this variation is little compared to the difference between individuals and LTL

can be used as an indicator of an individual's telomere length (Daniali et al., 2013).

There is evidence that LTL differs among human populations. Sub-saharan Africans have a LTL 400 bp longer than African-Americans,

and these in turn have a LTL which is 200 bp longer than in Europeans (Figure 10) (Hansen et al., 2016; Hunt et al., 2008). Variation of LTL among populations suggests that selection might be at play and that there is past and ongoing telomere evolution.

Since LTL has been shown to correlate with organismal fitness, this correlation might

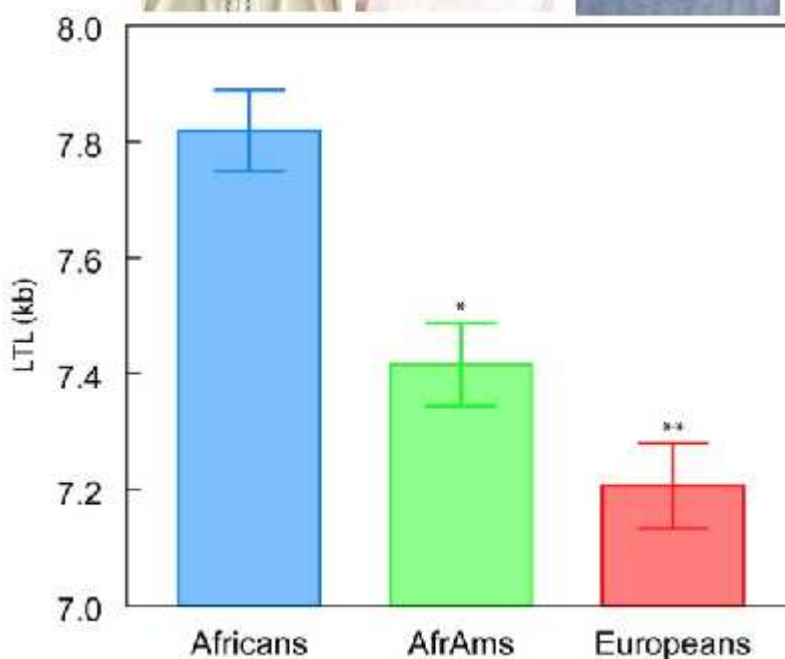


Figure 10: Telomere length in men of different ancestry. Adapted from Hansen et al., 2016.

pass through a major or minor propensity to develop diseases. In human populations, LTL variation correlates with skin pigmentation, and this is thought to have arisen to counteract the onset of melanoma. Individuals with dark skin are naturally more protected from melanoma, while lighter ones are not. The shortening of telomeres in Europeans has been selected for to protect them from a higher risk of developing melanomas (Mangino et al., 2015). This is further strengthened by the fact that Europeans with melanoma have longer LTL than other Europeans without the disease (Caini et al., 2015; Nan et al., 2011).

The fact that reduced telomere length evolved as a barrier against the onset of cancer is not new (Campisi, 1996). Cancer is a genetic disease that is triggered by the onset of mutations in

oncogenes/oncosuppressors that result in an uncontrolled cell proliferation. Therefore, the higher is the body mass of an organism, the higher will be its mutational burden and its risk of cancer. It is expected to observe a higher cancer incidence in big mammals than in small ones, but it is not the case, and this is known as the Peto's paradox (Risques & Promislow, 2018). The inconsistency between theoretical predictions and reality stands up because big size species evolved mechanisms to protect themselves from disease. Typically, big size mammals inactivate telomerase and have shorter telomere length. Some of them, which are particularly big, invented even additional mechanisms of protection. For instance, the African and Asian elephants duplicated their tumour suppressors genes, like TP53 (Abegglen et al., 2015; Gaughran, 2016). The bow-head whale chose to follow another strategy and duplicated its DNA repair genes (Keane et al., 2015).

Overall, longer LTL might expose individuals to a higher risk of elongating the proliferative potential of cells, and to cancer as a consequence of that. The acquisition of cellular immortalization is in fact an essential step in tumour progression and passes through the reactivation of telomerase or alternative telomere-lengthening pathways. Carrying short telomeres, however, is not safe as well. Short telomeres limit the maximum proliferative potential of cells, leading to early tissue decline and reduced regenerative potential (Aviv & Shay, 2018). The final outcome is a higher risk of developing degenerative and cardio-vascular diseases. Thus, there appears to be a trade-off between the risk associated to maintaining both long and short telomeres and the “best telomere length” probably depends on the environment where the organism lives and the type of stresses to which it is exposed.

3.5.1.6) Telomere dynamics and life-history trade-offs: the big picture

In the previous paragraphs we have seen how telomeres are at the center of the life of a cell. Telomeres can influence the maximum lifespan of a cell and, in consequence, that of the whole organism, as a result of both their initial state after birth and their rate of attrition during life. The initial state and attrition rate are influenced by hereditary factors, the stress of the mother during pregnancy and the environmental condition in early life. Environmental conditions continue to exert a strong effect on telomeres throughout life, determining the rate at which telomeres shorten. Stressors do not have an effect on telomeres only when they are permanently present, but also sporadic events like a spike in growth following a period of undernourishment, and reproductive effort, can have an important impact on telomere attrition. The action of telomerase can restore telomeres and its level is different across species but also during life-history within the same individual.

Telomere length per se can be considered as a risk factor for many diseases, including cancer (when it is long) and degenerative pathologies (when it is short). Telomeres are generally maintained

shorter when the risk of cancer is high, like in species with a high body mass. Evolutionary strategies exist to circumvent this problem, and, in theory, if the amount and efficiency of oncosuppressors guarantees an accurate protection against cancer, telomerase inactivation would not even be necessary.

Overall, the interaction of organisms with their environment is a complex and multi-levelled relationship, which is characterized by the interplay of many intrinsic and extrinsic factors. Reproductive success, health and lifespan can have different outcomes depending on how energy and limited resources are distributed, and telomeres are often the mediators of these outcomes.

In conclusion, the “best” telomere length of an individual, population or species probably does not exist, and the actual telomere length is the result of a complex interplay between many factors. The final result will depend on which of these factors weighs the most in the environment where the organism lives.

3.5.2) Telomeres as drivers of evolution

As I had anticipated in one of the first paragraphs of this chapter, the contribution of telomere dynamics to genome evolution can be seen by two different angles. On one side we can consider telomeres as whatever other part of the genome and as objects on which selection acts upon. On another side, telomeres can act as drivers of genome evolution, and the dynamic change in their structure can influence the state of the rest of the genome. Paradoxically, the action of telomeres as drivers of genome evolution often passes through their dysfunction.

Human cultured fibroblasts can only replicate a finite number of

times (usually between 40 and 60), a phenomenon known as the “Hayflick limit”, in honour of its discoverer Leonard Hayflick, who first described it (Shay & Wright, 2000). The cessation of cell

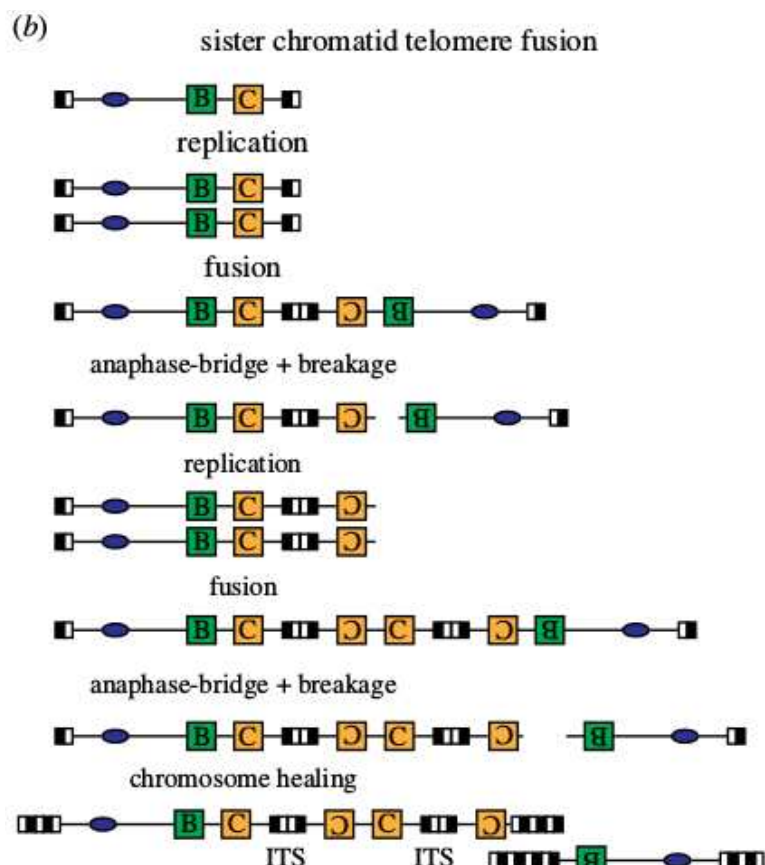


Figure 11: Schematic representation of bridge-fusion-breakage cycle between sister chromatids. Adapted from Baird, 2018.

divisions is known as “replicative senescence” and is mainly caused by the progressive shortening of telomeres upon genome replication. Once telomeres become critically short, cellular checkpoints like the DNA damage response are activated and trigger a permanent cell cycle arrest. Cells can remain in this non-dividing state for many years before finally dying (Baird, 2018). However, this control can be circumvented by inactivating important checkpoints mediated by oncosuppressor proteins. In humans, mutations or deletions of key effectors of the cell cycle checkpoints, like Tp53 and RB, often result in the ability of cells to escape these checkpoints and continue dividing, even in the presence of critically short telomeres (Baird, 2018). This eventually leads to the presence of completely uncapped chromosome-ends, which are recognized as double-strand breaks (DSBs) and processed by the DNA repair machinery. The most obvious outcome of this process is the fusion of telomeres from different chromosomes. This leads to the origin of fused chromosomes carrying two centromeres. At the next cell cycle, these chromosomes will be subject to breakage and generate other uncapped ends which will in turn be fused and exacerbate the genome reshuffling, generating even more complex chromosomal rearrangements, in a process known as bridge-fusion-breakage (BFB) cycle, first described in *Zea mays* by Barbara McClintock (**Figure 11**) (McClintock, 1941). The reiteration of telomere fusions and BFB cycles generates interstitial telomeric sequences (ITS) which can further promote genome instability by mediating recombination with other ITS (Aksenova & Mirkin, 2019). Several rounds of BFB cycles, such as those induced at telomere crisis, can lead to complex chromosome rearrangements named “chromothripsis” (Cleal & Baird, 2020; Maciejowski et al., 2015).

These types of genome rearrangements can trigger the onset of cancer (Artandi et al., 2000), and genome instability is in fact a hallmark of cancer cells (Hanahan & Weinberg, 2011). The reiteration of BFB cycles in cancer cells can lead to a high level of heterogeneity among cells in the same tumour. High tumour heterogeneity is usually associated to a higher chance of developing resistance to chemotherapeutics, due to the high rate of genome evolution which increases the chance of duplicating or generating new genes able to counteract the effect of drugs, and to a worse prognosis. However, such a catastrophic and unstable cellular state cannot survive long in the absence of chromosome-end maintenance pathways. Inside a tumour, the majority of cells undergoes death and apoptosis after entering into this “crisis” stage (Hayashi et al., 2015), but some clones can escape crisis by restoring the ability to maintain their chromosome-ends. The acquisition of cellular immortalization is an essential step in tumour progression and passes through the reactivation of telomerase or alternative telomere-lengthening pathways (Mathieu et al., 2004). In a study of over 18000 human cancer and non-neoplastic samples, the authors discovered that telomere length was often shorter in cancer cells, and 73% of cancers reactivated telomerase by various mechanisms, including point mutations, gene duplications, gene chimerisation and modulation of its expression

(Barthel et al., 2017).

Telomere dysfunction is not only observed in cancer, but is associated with other pathologies like chronic inflammation states, neurodegenerative diseases and cardio-vascular diseases. However, these studies are generally correlational, and cannot inform whether telomere dysfunction is at the origin or a consequence of the pathological state. Additional research is needed to get an insight into the contribution of telomere-driven genetic instability to the onset of these diseases.

Telomeres can also have an important impact on species evolution. Reproductive isolation between populations can arise as a consequence of structural variations between their genomes and initiate speciation. In chapter 1, I described two examples of reproductive isolation between *Saccharomyces* species. In one example, a Malaysian *Saccharomyces cerevisiae* subpopulation was reproductively isolated from the rest of the species because of multiple chromosomal rearrangements (Marie-Nelly et al., 2014; Yue et al., 2017). The same happened to a *S. paradoxus* South-American subpopulation carrying multiple SVs (Naumov et al., 2000; Yue et al., 2017) (**Chapter 1 - Figure 8**). If the reproductive isolation between the subpopulations continues for enough evolutionary time, the populations will continue to diverge and will eventually become different species.

If genome rearrangements are a consequence of telomere dysfunction, we can affirm that we are in front of a case of telomere-driven evolution. This can easily be tested by the inspection of ITS distribution in internal parts of the genome, which are a direct result of telomere fusions. An interesting example of this phenomenon can be found comparing the human and other great apes' genomes: while all great apes have a diploid genome content of $n=48$ (Yunis & Prakash, 1982), humans contain only 46 chromosomes, because chromosome II derived from an archaic telomeric fusion of two autosomes (Dreszer et al., 2007; Ijdo et al., 1991; Ventura et al., 2012). The presence of ITS is not characteristic of humans, but has been found in other organisms (Meyne et al., 1990; Rovatsos et al., 2015). An extreme example is the muntjac deer, where the Indian population contains 6–7 chromosomes and multiple ITS, whereas the Chinese population contains 46 chromosomes. Other populations contain an intermediate number of chromosomes (Benirschke & Wurster, 1970; Mudd et al., 2020; Tsipouri et al., 2008; Wang & Lan, 2000). The fixation of these new karyotype structures in a population may be facilitated by selective advantage and constitute the first step of telomere-driven speciation. In this sense, the acquisition of ITS can promote adaptation to the environment as ITS constitute active binding sites for transcription factors like Rap1p and Trf2, and therefore their presence near a gene can modulate its expression (Yang et al., 2011).

All the mechanisms discussed so far involve the onset of structural variation upon telomere dysfunction. However, telomeres can exert an action on their host genome also at other, less

macroscopic levels. In a study that used *S. cerevisiae* as model organism, the authors inactivated telomerase by deleting *EST1* and placed *CAN1* at a subtelomeric site to measure its mutation rate. They found that *est1delta* strains had 50-fold higher chance of inactivating the gene by terminal deletions. However, the authors did not detect an increase in point mutation rate upon telomerase inactivation, suggesting that the genomic effects of telomere dysfunction are mediated by structural variation (Hackett et al., 2001).

Overall, there are sufficient examples in literature to conclude that telomeres have an important role in driving genome and species evolution, but additional research is needed to unravel the molecular mechanisms and pathways underlying this process.

3.5.3) The evolutionary contribution of telomere sequence dynamics

Telomeres are ribonucleoproteins essential for the stability and maintenance of chromosomes. As their role is fundamental for the health of the cell and their perturbation can cause severe damage, one would expect their sequence to be highly conserved across the tree of life. In contrast, one would remain easily surprised by the astonishing variety present in such tree in terms of telomeric sequences. Even though telomere sequences all have conserved characteristics, like being highly repetitive, G-rich and short, there are many variants of telomeric motifs.

So far, the most comprehensive analysis of telomere sequences by an evolutionary perspective has been carried out in 2013 by Fulneckova et al. The authors examined a set of 143 publicly available genome sequences of under-represented taxa and determined their telomeric repeats using bioinformatic methods (Fulnečková et al., 2013).

Organisms of the Metazoa taxa, comprising all animals, have mostly TTAGGG repeats, with some exceptions (Fajkus et al., 2005; Gomes et al., 2011). The canonical TTAGGG is characteristic of vertebrates. Other Metazoa taxa have variants of this motif. Nematodes like the model organism *Caenorabditis elegans* carry a variant which contains a mismatch at the level of the last base (TTAGGC) (Zetka & Müller, 1996). Most insects contain a partially deleted repeat which has lost the last base (TTAGG) (Vítková et al., 2005). Inside the insect taxa, however, additional modifications have appeared at the level of subtaxa and ultimately modified the telomeric sequence. For instance, some families of the order Coleoptera carry telomeres with TCAGG sequence (Frydrykova & Marec, 2002; Mravinac et al., 2011). Other insects, like the order Diptera which comprises the model organism *Drosophila melanogaster*, carry alternative mechanisms of chromosome capping, involving the use of retrotransposons and long terminal repeats (LTR) (Kuznetsova et al., 2020).

In the plants kingdom (Archaeplastida), the widespread telomeric repeat carries one additional base and became TTTAGGG (Peska & Garcia, 2020), and it is also called *Arabidopsis*-type telomeric

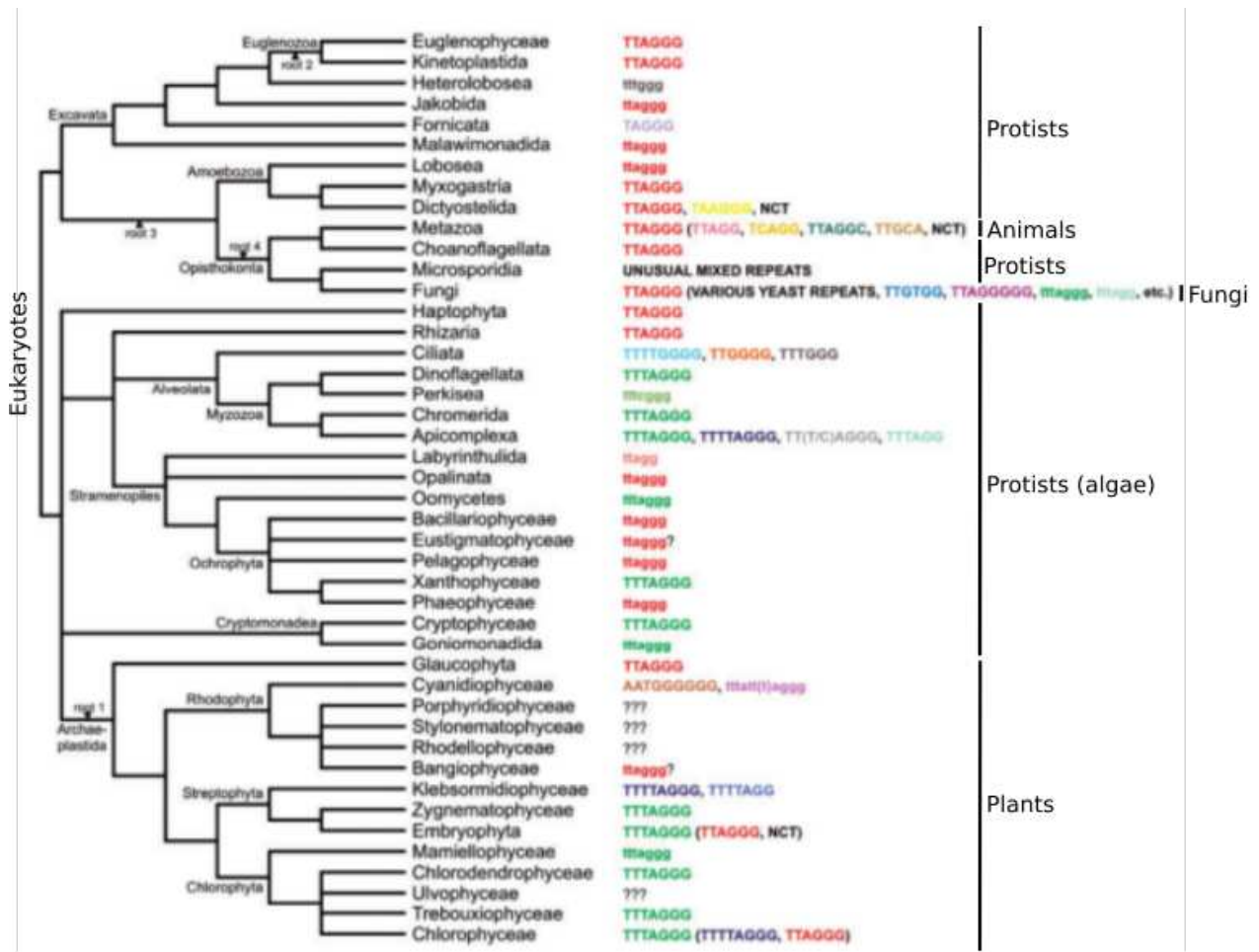
repeat since it was discovered in the plant model organism *Arabidopsis thaliana* (Richards & Ausubel, 1988). As per what happens in the animal kingdom, there are variations of this motif. In the red algae (Rhodophyta), the few known telomeric repeats are AATGGGGGG for *Cyanidioschyzon merolae* (Cyanidiophyceae) (Nozaki et al., 2007), TTATT(T)AGGG for *Galdieria sulphuraria* (Cyanidiophyceae) (Fulnečková et al., 2013), while the canonical TTAGGG has been found in *Porphyra umbilicalis* (Bangiophyceae) (Fulnečková et al., 2013). The nature of telomeric repeats in other Rhodophyta is still unknown. In the green algae (Chlorophyta), the TTTAGGG motif is widespread and it has been described in the genera *Ostreococcus* (Mamiellophyceae) (Derelle et al., 2006) and *Chlorella* (Trebuxiophyceae) (Higashiyama et al., 1995), but exceptions are present in the class Chlorophyceae, where the genus *Dunaliella* and *Stephanosphaeria* went back to the canonical TTAGGG (Fulnečková et al., 2012), and the motif TTTTAGGG is present in *Chlamydomonas* (Petraček et al., 1990). In the clade Streptophyta, which comprises also the land plants (Embriophyta), the majority of the classes contains the plant-like motif TTTAGGG, while the class *Klebsormidiophyceae* carries the alternative motifs TTTTAGGG and TTTTAGG (Cox et al., 1993; Fuchs et al., 1995). Among the land plants, the unusual vertebrate-type TTAGGG was observed in Aloe and in some other Asparagales (Puizina et al., 2003; Sýkorová et al., 2003; Weiss & Scherthan, 2002).

Most protists carry the canonical TTAGGG as telomeric repeat, but a non-negligible number of protist taxa have telomeres more similar to those of plants. For example, the brown algae (Ochrophyta) almost exclusively contain TTAGGG motifs, except in *Xantophyceae* which carry the plant-like motif. Other variations to the TTAGGG motifs are also present in Alveolates, a group of protists with characteristic flattened sacs positioned under their cellular membrane.


Most fungi contain the canonical TTAGGG repeat, but some important variations are present, especially in yeasts which carry degenerated telomeric sequences and will be described in the next paragraph.

Overall, the majority of taxa carry a TTAGGG telomeric repeat, which is the one present in humans. This pattern suggests that the TTAGGG motif was the ancestral one present in the last common ancestor of all eukaryotes (LECA), and other groups of organisms modified their telomeric repeat later in evolution (Fulnečková et al., 2013). Among the main changes is the one happened in the ancestor of plants, which added one T. Some specific plants carried on the evolution of their telomeric repeat and even added a second T, later followed by a deletion of a terminal G, like in *Klebsormidiophyceae*, or others went back to the ancestral TTAGGG repeat, like in some *Chlorophyceae*. Among the Metazoa (animals), the common ancestor of insects deleted the terminal G, and other subgroups of insects continued in their evolution to experiment other motifs, like TCAGG or TTGCA. Most protists remained with the ancestral-type repeat but others took the same


evolutionary path of plants and added one T, like in some Alveolates. More complicated variations evolved from this plant-like motif by additional mutation/deletion/addition of nucleotides in Ciliates and Apicomplexans (Sohanpal et al., 1995). This remarkable variability indicates that, although telomeres are essential for the life of the cell, they show a considerable level of plasticity and they can be modified in sequence as long as they maintain certain important characteristics, like G-richness (**Figure 12**).




Excavata

 unicellular eukaryotes with ventral feeding groove (*Giardia intestinalis*)


Euglenozoa

 photosynthetic, flagellate excavates (*Euglena gracilis*)


Alveolata

 Eukaryotes with cortical alveoli: flattened vesicles packed under the membrane (*Paramecium caudatum*)

Myzozoa

 Alveolates feeding by myzocytosis: perforing membrane of cells and sucking the content (*Plasmodium falciparum*)


Stramenopiles

 Various types of algae


Ochrophyta

 Golden-brown algae


Amoebozoa

 unicellular protozoa with jellylike cytoplasm moving through body extensions (*Amoeba proteus*)

Opisthokonta

 Large group of eukaryotes with a posterior flagellum. Comprises species of fungi, animals and amoebas.


Cryptomonadea

 Algae with ejectosomes

Archaeplastida

 Plants *sensu lato*

Rhodophyta

 Red algae

Chlorophyta

 Green algae

Streptophyta


 Land plants and other types of plants

Figure 12: Phylogenetic tree illustrating telomeric repeats in the major kingdoms of life. Capital letters indicate telomeres discovered by teloblot, while non-capital letters indicate telomeres discovered by in silico prediction. Adapted from Fulneckova et al., 2013.

3.5.3.1) Telomere sequence dynamics in yeasts

Even among Fungi, the most prevalent telomeric sequence is the canonical TTAGGG, which is present in the model organism *Neurospora crassa* (Schechtman, 1990). In yeasts, instead, there is a remarkable variation of telomeric repeats, which are even irregular (Qi et al., 2013). In the yeast *Saccharomyces cerevisiae*, for instance, telomeres are repetitive

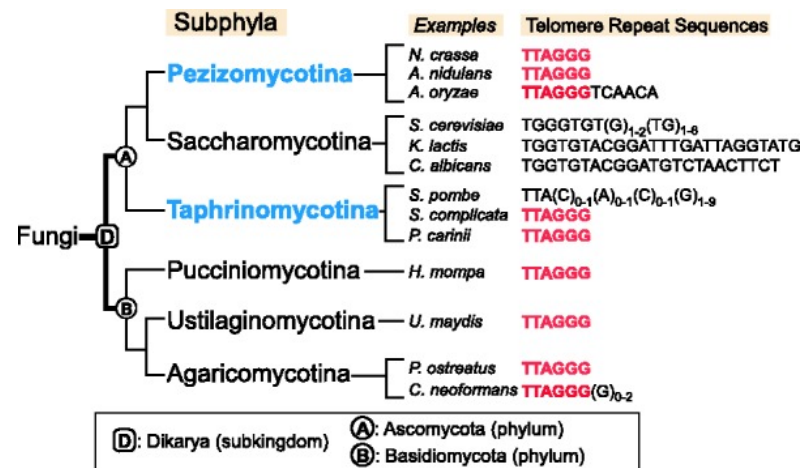


Figure 13: Representation of telomeric repeat variation in the kingdom Fungi. Adapted from Qi et al., 2013.

blocks of a sequence which always has one T but can be followed by 1, 2 or 3 G. The adenine residue has been deleted from the telomeric unit. The combination of these irregular motifs makes every telomere in yeast cells unique in its sequence. In the genus *Candida*, the evolution of telomeres has been even more pronounced and these yeasts carry very long telomeric units which are very far from the ancestral TTAGGG motif (Figure 13). Yeasts of the genus *Yarrowia* carry intermediately divergent telomeres, which inserted a variable sequence between T and A, and whose motif can be summarized by the formula TTNNNNAGGG (Červenák et al., 2019).

Yeasts evolved from a canonical TTAGGG repeat and this is confirmed by the presence of degenerated TTAGGG repeats in the centrosome-proximal regions of their telomeres, and are reminiscent of the ancient telomeric structure (Brun et al., 1997). The evolution of telomeric sequences happens by progressive accumulation of mutations in the RNA template of telomerase and it is confirmed by the fact that telomeric repeats are complementary to the RNA template. Thanks to this mechanism, a unique modification in one genome site can cause the modification of all the telomeres at once.

The modification of telomeric sequences must be followed by an adaptation of telomere-binding proteins. The presence of ancient TTAGGG repeats at the beginning of *S. cerevisiae* telomeres suggests that the proteins which are binding them might also be the ancestral telomere-binding proteins of the last yeast common ancestor (LYCA). In *S. cerevisiae*, these units are bound by Tbf1p, a transcription factor which is essential for mitotic growth (Brigati et al., 1993; Koering et al., 2000). Additional evidence that Tbf1p might be the ancestral telomere-binding protein comes from the fact that it contains a DNA-binding domain more closely related than Rap1p to the one of TRF1 and TRF2, which are the principal telomere-binding proteins in humans. All Tbf1-related proteins possess a very similar Myb-like binding domain called “telobox” (Bilaud et al., 1996,

1997). Tbf1-like proteins are found in the majority of yeasts, and are very similar in their structure. However, most of the telomere length in *S. cerevisiae* is composed of the derivative TG₁₋₃ motifs and is bound by Rap1p. The switch from TTAGGG to TG₁₋₃ has been accompanied and allowed by a shift between Tbf1-bound to Rap1-bound telomeres (Červenák et al., 2017; Teixeira & Gilson, 2005). Rap1p is characterized by a much more flexible binding to DNA, which allows more variable sequence in the binding sites (Lieb et al., 2001). At the same time, Tbf1p lost its binding to telomeres but remained bound to subtelomeric sites, where it continued to regulate telomere length and gene expression (Berthiau et al., 2006).

3.5.3.2) A case-study of telomere evolution: the humanized yeast model

The evolution of telomeric sequences occurs by spontaneous and progressive accumulation of mutations in the RNA template of telomerase. This is possible because telomerase is not extremely strict and can tolerate to work with a slightly different RNA partner. Naturally-occurring telomere mutants have been described in *Kluyveromyces lactis* and caused abnormal enlargement of cells (Smith & Blackburn, 1999). Other spontaneous mutations have also been described in *Tetrahymena* and *S. cerevisiae* (Blackburn, 2000). Modified telomeres in budding yeast can also be produced artificially. Lin et al. produced a set of 63 mutant yeasts with mutated telomeres by engineering three basepairs in the template region of *TLC1*, coding the telomerase RNA template. These mutants exhibited defects in chromosome segregation and activation of cell-cycle checkpoints (Lin et al., 2004). A particular case of telomere modification is when the RNA template of *S. cerevisiae* is modified to contain vertebrate-type TTAGGG telomeric repeats. This can be achieved just by the modification of 16 nt in the template region of *TLC1* (Henning et al., 1998). This strain is called “humanized yeast” although, as discussed above, TTAGGG repeats can be found at the telomeres of a large variety of eukaryotic cells. The first purpose of this strain was initially to completely reconstitute human telomeres in a more amenable system to study human telomere biology (Bah et al., 2004). However, the humanized yeast constitutes an invaluable model to investigate the evolutionary mechanisms of telomere evolution and how organisms can adapt to new telomeric sequences. In addition, it constitutes a sort of “reverse evolution” experiment in which yeasts go back to their ancestral telomeric units (Teixeira & Gilson, 2005). What did we learn so far about humanized yeast?

Humanized yeasts can survive for long periods of time (up to 2500 generations) and their native TG₁₋₃ telomeres are progressively replaced by human ones. However, a core of 30-50 nt is always maintained in a TG₁₋₃ state at the beginning of telomeres. This is due to the action of telomerase which replenishes telomeres before they can get completely eroded. As a consequence, telomeres in humanized yeasts are always mixed with a first sequence composed by TG₁₋₃ units and a terminal

one composed by the newly synthesized human units (Bah et al., 2004). Humanized telomeres are generally shorter than wild-type ones, while their terminal single-stranded G-tail is longer (Bah et al., 2004). The humanization of yeast telomeres is accompanied by a reshaping of telomere-binding proteins. Humanized telomeres are not bound anymore by Rap1p, but they are bound by the ancestral telomere-binding protein Tbf1p (Alexander & Zakian, 2003; Brevet, 2003; Ribaud et al., 2012). Ku80p is also able to bind humanized telomeres (**Figure 14**) (Auriche et al., 2008).

Humanized yeasts are strictly dependent on the presence of an active and functional telomerase for their survival. Their telomere turnover is very high and when telomerase is inactivated they cannot grow (Bah et al., 2011).

However, humanized telomeres are not harmless and they trigger a chronic activation of the DNA damage response (DDR) pathway, which can be deduced by the detection of Rad53 phosphorylation. The activation of DDR induces a delay in cell cycle progression and slowing of the growth rate (Di Domenico et al., 2009), which is even exacerbated by DNA replication problems as the replication fork is frequently stalled at the level of these telomeres (Bah et al., 2011). Interestingly, the presence of a humanized single-stranded G-tail is sufficient to activate the DDR, while the presence of humanized telomeres “sandwiched” between yeast telomeres does not trigger any DDR and is perfectly tolerated (Bah et al., 2011). The chronic activation of DDR prevents chromosome end-to-end fusions. When main DDR effectors like *TEL1* and *MEC1* are deleted, yeasts start to accumulate genome instability (Di Domenico et al., 2009). Humanized yeasts have also an altered transcriptional profile as they are incapable of spreading transcriptional silencing in nearby subtelomeric sites (TPE). This is due by their inability to recruit Sir proteins due to the loss of the Rap1p binding. Most telomere-binding proteins, like Rap1 and Tbf1, are transcription factors. As a consequence, the restructuring of telomeres will induce a reshape in the location and abundance of transcription factors over genomic sites. It is expected therefore that telomere humanization will have a profound effect on the transcriptional profile of the cell, not only limited to the de-repression of subtelomeric genes. Unfortunately, this experimental model system has fallen into disuse and currently no transcription profiles are available.

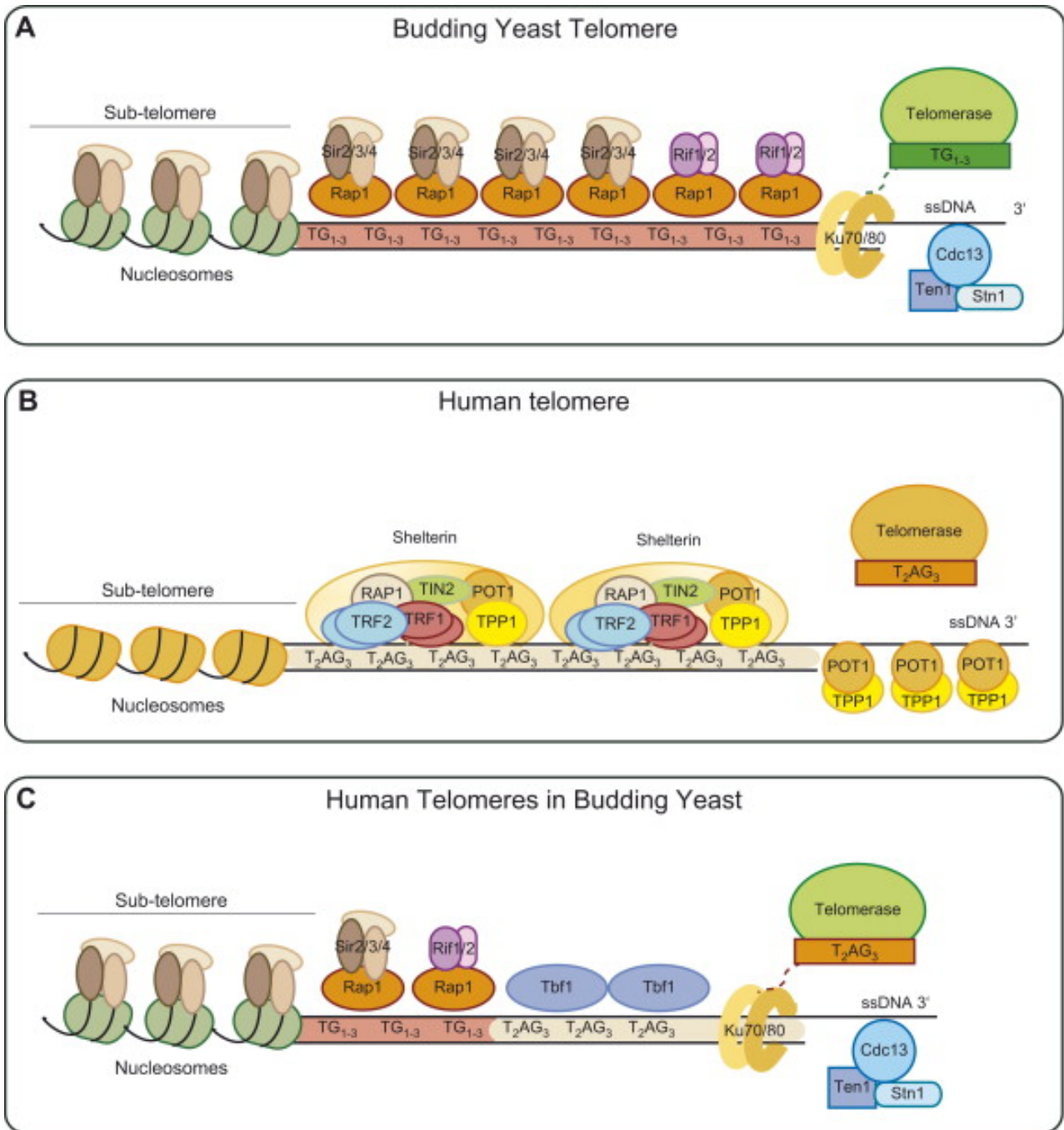


Figure 14: Comparison of telomere structures of: a) wild-type yeasts, b) humans, c) humanized yeasts. Adapted from Auriche et al., 2008.

References

- Abegglen L., Caulin A., Chan A., Lee K., Robinson R., Campbell M., Kiso W., Schmitt D., Waddell P., Bhaskara S., Jensen S., Maley C., Schiffman J. (2015). Potential mechanisms for cancer resistance in elephants and comparative cellular response to DNA damage in humans. *JAMA*. 176(1), 100–106. <https://doi.org/10.1126/science.1166426>.
- Aksenova, A. Y., Greenwell, P. W., Dominska, M., Shishkin, A. A., Kim, J. C., Petes, T. D., & Mirkin, S. M. (2013). Genome rearrangements caused by interstitial telomeric sequences in yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49), 19866–19871. <https://doi.org/10.1073/pnas.1319313110>
- Aksenova, A. Y., Han, G., Shishkin, A. A., Volkov, K. V., & Mirkin, S. M. (2015). Expansion of Interstitial Telomeric Sequences in Yeast. *Cell Reports*, 13(8), 1545–1551. <https://doi.org/10.1016/j.celrep.2015.10.023>
- Aksenova, A. Y., & Mirkin, S. M. (2019). At the beginning of the end and in the middle of the beginning: Structure and maintenance of telomeric dna repeats and interstitial telomeric sequences. *Genes*, 10(2). <https://doi.org/10.3390/GENES10020118>
- Alexander, M. K., & Zakian, V. A. (2003). Rap1p telomere association is not required for mitotic stability of a C₃TA₂ telomere in yeast. *EMBO Journal*. 22(7): 1688–1696.
- Alonso-Alvarez, C., Bertrand, S., Faivre, B., & Sorci, G. (2007). Increased susceptibility to oxidative damage as a cost of accelerated somatic growth in zebra finches. *Functional Ecology*, 21(5), 873–879. <https://doi.org/10.1111/j.1365-2435.2007.01300.x>
- Artandi. (2000). *Telomere dysfunction promotes non-reciprocal translocations*. 20070(1999), 641–645.
- Asghar, M., Hasselquist, D., Hansson, B., Zehndjiev, P., Westerdahl, H., & Bensch, S. (2015). Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. *Science*, 347(6220), 436–438. <https://doi.org/10.1126/science.1261121>
- Askree, S. H., Yehuda, T., Smolikov, S., Gurevich, R., Hawk, J., Coker, C., Krauskopf, A., Kupiec, M., & McEachern, M. J. (2004). A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.0401263101>
- Auriche, C., Di Domenico, E. G., & Ascenzioni, F. (2008). Budding yeast with human telomeres: A puzzling structure. In *Biochimie*. <https://doi.org/10.1016/j.biochi.2007.09.009>
- Aviv, A., & Shay, J. W. (2018). Reflections on telomere dynamics and ageing-related diseases in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1741). <https://doi.org/10.1098/rstb.2016.0436>
- Axelsson, J., Wapstra, E., Miller, E., Rollings, N., & Olsson, M. (2020). Contrasting seasonal patterns of telomere dynamics in response to environmental conditions in the ectothermic sand lizard, *Lacerta agilis*. *Scientific Reports*, 10(1), 1–9. <https://doi.org/10.1038/s41598-019-57084-5>

- Badás, E. P., Martínez, J., Rivero de Aguilar Cachafeiro, J., Miranda, F., Figuerola, J., & Merino, S. (2015). Ageing and reproduction: Antioxidant supplementation alleviates telomere loss in wild birds. *Journal of Evolutionary Biology*, 28(4), 896–905. <https://doi.org/10.1111/jeb.12615>
- Bah, A., Bachand, F., Clair, É., Autexier, C., & Wellinger, R. J. (2004). Humanized telomeres and an attempt to express a functional human telomerase in yeast. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh511>
- Bah, A., Gilson, E., & Wellinger, R. J. (2011). Telomerase is required to protect chromosomes with vertebrate-type T2AG33' ends in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M111.220186>
- Baird, D. M. (2018). Telomeres and genomic evolution. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1741). Royal Society Publishing. <https://doi.org/10.1098/rstb.2016.0437>
- Barrett, E. L. B., Burke, T. A., Hammers, M., Komdeur, J., & Richardson, D. S. (2013). Telomere length and dynamics predict mortality in a wild longitudinal study. *Molecular Ecology*, 22(1), 249–259. <https://doi.org/10.1111/mec.12110>
- Barthel, F. P., Wei, W., Tang, M., Martinez-Ledesma, E., Hu, X., Amin, S. B., Akdemir, K. C., Seth, S., Song, X., Wang, Q., Lichtenberg, T., Hu, J., Zhang, J., Zheng, S., & Verhaak, R. G. W. (2017). Systematic analysis of telomere length and somatic alterations in 31 cancer types. *Nature Genetics*. <https://doi.org/10.1038/ng.3781>
- Ben-Shitrit, T., Yosef, N., Shemesh, K., Sharan, R., Ruppín, E., & Kupiec, M. (2012). Systematic identification of gene annotation errors in the widely used yeast mutation collections. *Nature Methods*, 9(4), 373–378. <https://doi.org/10.1038/nmeth.1890>
- Benirschke & Wurster. (1970). Indian Momtjac, *Muntiacus muntiac*: A Deer with a Low Diploid Chromosome Number. *Science*. Vol. 168, Issue 3937, pp. 1364-1366. DOI: 10.1126/science.168.3937.1364.
- Berthiau, A. S., Yankulov, K., Bah, A., Revardel, E., Luciano, P., Wellinger, R. J., Géli, V., & Gilson, E. (2006). Subtelomeric proteins negatively regulate telomere elongation in budding yeast. *EMBO Journal*. <https://doi.org/10.1038/sj.emboj.7600975>
- Bianchi, A., & Shore, D. (2007). Early Replication of Short Telomeres in Budding Yeast. *Cell*, 128(6), 1051–1062. <https://doi.org/10.1016/j.cell.2007.01.041>
- Bilaud, T., Brun, C., Ancelin, K., Koering, C. E., Laroche, T., & Gilson, E. (1997). *telobox* protein. 17(october), 155–158.
- Bilaud, T., Koering, C. E., Binet-Brasselet, E., Ancelin, K., Pollice, A., Gasser, S. M., & Gilson, E. (1996). The telobox, a Myb-related telomeric DNA binding motif found in proteins from yeast, plants and human. *Nucleic Acids Research*, 24(7), 1294–1303. <https://doi.org/10.1093/nar/24.7.1294>
- Bize, P., Criscuolo, F., Metcalfe, N. B., Nasir, L., & Monaghan, P. (2009). Telomere dynamics rather than age predict life expectancy in the wild. *Proceedings of the Royal Society B: Biological Sciences*, 276(1662), 1679–1683. <https://doi.org/10.1098/rspb.2008.1817>

- Blackburn, E. H. (2000). *The end of the (DNA) line*. 7(10), 847–850.
- Blackburn, E. H., & Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*. *Journal of Molecular Biology*, 120(1), 33–53. [https://doi.org/10.1016/0022-2836\(78\)90294-2](https://doi.org/10.1016/0022-2836(78)90294-2)
- Bodnar, A. G., Ouellette, M., Frolkis, M., Holt, S. E., Chiu, C. P., Morin, G. B., Harley, C. B., Shay, J. W., Lichtsteiner, S., & Wright, W. E. (1998). Extension of life-span by introduction of telomerase into normal human cells. *Science*, 279(5349), 349–352. <https://doi.org/10.1126/science.279.5349.349>
- Bosoy, D., Peng, Y., Mian, I. S., & Lue, N. F. (2003). Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo. *Journal of Biological Chemistry*, 278(6), 3882–3890. <https://doi.org/10.1074/jbc.M210645200>
- Brevet V., Berthiau A., Civitelli L., Donini P., Schramke V., Geli V., Acenzioni F., Gilson E. (2003). The number of vertebrate repeats can be regulated at yeast telomeres by Rap1-independent mechanisms. *EMBO Journal*. 22(7): 1697–1706.
- Brigati C., Kurtz S., Balderes D., Vidali G., Shore D. (1993). An Essential Yeast Gene Encoding a TTAGGG repeat-binding protein. *Mol Cell Biol*. 13(2): 1306–1314.
- Bronikowski, A. M. (2008). The evolution of aging phenotypes in snakes: A review and synthesis with new data. *Age*, 30(2–3), 169–176. <https://doi.org/10.1007/s11357-008-9060-5>
- Brun, C., Marcand, S., & Gilson, E. (1997). Proteins that bind to double-stranded regions of telomeric DNA. *Trends in Cell Biology*, 7(8), 317–324. [https://doi.org/10.1016/S0962-8924\(97\)01092-1](https://doi.org/10.1016/S0962-8924(97)01092-1)
- Buck, M. J., & Lieb, J. D. (2009). Transcription Factor Targets. *October*, 38(12), 1446–1451. <https://doi.org/10.1038/ng1917.A>
- Caini, S., Raimondi, S., Johansson, H., De Giorgi, V., Zanna, I., Palli, D., & Gandini, S. (2015). Telomere length and the risk of cutaneous melanoma and non-melanoma skin cancer: A review of the literature and meta-analysis. *Journal of Dermatological Science*, 80(3), 168–174. <https://doi.org/10.1016/j.jdermsci.2015.08.003>
- Campisi, J. (1996). Replicative senescence: An old lives' tale? *Cell*, 84(4), 497–500. [https://doi.org/10.1016/S0092-8674\(00\)81023-5](https://doi.org/10.1016/S0092-8674(00)81023-5)
- Canela, A., Vera, E., Klatt, P., & Blasco, M. A. (2007). High-throughput telomere length quantification by FISH and its application to human population studies. *Proceedings of the National Academy of Sciences of the United States of America*, 104(13), 5300–5305. <https://doi.org/10.1073/pnas.0609367104>
- Caprioli, M., Romano, M., Romano, A., Rubolini, D., Motta, R., Folini, M., & Saino, N. (2013). Nestling telomere length does not predict longevity, but covaries with adult body size in wild barn swallows. *Biology Letters*, 9(5). <https://doi.org/10.1098/rsbl.2013.0340>
- Červenák, F., Juríková, K., Devillers, H., Kaffe, B., Khatib, A., Bonnell, E., Sopkovičová, M., Wellinger, R. J., Nosek, J., Tzfati, Y., Neuvéglise, C., & Tomáška, L. (2019). Identification of

- telomerase RNAs in species of the *Yarrowia* clade provides insights into the co-evolution of telomerase, telomeric repeats and telomere-binding proteins. *Scientific Reports*, 9(1), 1–15. <https://doi.org/10.1038/s41598-019-49628-6>
- Červenák, F., Juríková, K., Sepšiová, R., Neboháčová, M., Nosek, J., & Tomáška, L. (2017). Double-stranded telomeric DNA binding proteins: Diversity matters. *Cell Cycle*, 16(17), 1568–1577. <https://doi.org/10.1080/15384101.2017.1356511>
- Chan, C. S. M., & Tye, B. K. (1983). Organization of DNA sequences and replication origins at yeast telomeres. *Cell*, 33(2), 563–573. [https://doi.org/10.1016/0092-8674\(83\)90437-3](https://doi.org/10.1016/0092-8674(83)90437-3)
- Chapon, C., Cech, T. R., & Zaugg, A. J. (1997). Polyadenylation of telomerase RNA in budding yeast. *Rna*, 3(11), 1337–1351.
- Chatelain, M., Drobniak, S. M., & Szulkin, M. (2020). The association between stressors and telomeres in non-human vertebrates: a meta-analysis. In *Ecology Letters* (Vol. 23, Issue 2, pp. 381–398). Blackwell Publishing Ltd. <https://doi.org/10.1111/ele.13426>
- Chen, Q., Ijima, A., & Greider, C. W. (2001). Two Survivor Pathways That Allow Growth in the Absence of Telomerase Are Generated by Distinct Telomere Recombination Events. *Molecular and Cellular Biology*, 21(5), 1819–1827. <https://doi.org/10.1128/mcb.21.5.1819-1827.2001>
- Claussin, C., & Chang, M. (2015). The many facets of homologous recombination at telomeres. *Microbial Cell*, 2(9), 308–321. <https://doi.org/10.15698/mic2015.09.224>
- Cleal, K., & Baird, D. M. (2020). Catastrophic Endgames: Emerging Mechanisms of Telomere-Driven Genomic Instability. *Trends in Genetics*, 36(5), 347–359. <https://doi.org/10.1016/j.tig.2020.02.001>
- Coluzzi, E., Leone, S., & Sgura, A. (2019). Oxidative Stress Induces Telomere Dysfunction and Senescence by Replication Fork Arrest. *Cells*, 8(1), 19. <https://doi.org/10.3390/cells8010019>
- Conrad, M. N., Wright, J. H., Wolf, A. J., & Zakian, V. A. (1990). RAP1 protein interacts with yeast telomeres in vivo: Overproduction alters telomere structure and decreases chromosome stability. *Cell*, 63(4), 739–750. [https://doi.org/10.1016/0092-8674\(90\)90140-A](https://doi.org/10.1016/0092-8674(90)90140-A)
- Cox A., Bennett S., Parokony A., Kenton A., Callimassia M., Bennett M. (1993). Comparison of Plant Telomere Locations using a PCR-generated Synthetic Probe. *Annals of botany*. Volume 72, Issue 3, September 1993, Pages 239–247.
- Daniali, L., Benetos, A., Susser, E., Kark, J. D., Labat, C., Kimura, M., Desai, K., Granick, M., & Aviv, A. (2013). Telomeres shorten at equivalent rates in somatic tissues of adults. *Nature Communications*, 4, 1–7. <https://doi.org/10.1038/ncomms2602>
- Dantzer, B., & Fletcher, Q. E. (2015). Telomeres shorten more slowly in slow-aging wild animals than in fast-aging ones. *Experimental Gerontology*, 71, 38–47. <https://doi.org/10.1016/j.exger.2015.08.012>
- Derelle E., Ferraz C., Rombauts S., Rouze P., Worden A., Robbens S., Partensky F., Degroevé S., ... Moreau H. (2006). Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *PNAS*. 103(31): 11647–11652.

- DeZwaan, D. C., & Freeman, B. C. (2009). The conserved Est1 protein stimulates telomerase DNA extension activity. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(41), 17337–17342. <https://doi.org/10.1073/pnas.0905703106>
- Di Domenico, E. G., Auriche, C., Viscardi, V., Longhese, M. P., Gilson, E., & Ascenzioni, F. (2009). The Mec1p and Tel1p checkpoint kinases allow humanized yeast to tolerate chronic telomere dysfunctions by suppressing telomere fusions. *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2008.10.005>
- Ding, Z., Mangino, M., Aviv, A., Spector, T., & Durbin, R. (2014). Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gku181>
- Dreszer, T. R., Wall, G. D., Haussler, D., & Pollard, K. S. (2007). Biased clustered substitutions in the human genome: The footprints of male-driven biased gene conversion. *Genome Research*, *17*(10), 1420–1430. <https://doi.org/10.1101/gr.6395807>
- Dupoué, A., Rutschmann, A., Le Galliard, J. F., Clobert, J., Angelier, F., Marciaud, C., Ruault, S., Miles, D., & Meylan, S. (2017). Shorter telomeres precede population extinction in wild lizards. *Scientific Reports*, *7*(1), 1–8. <https://doi.org/10.1038/s41598-017-17323-z>
- Evans, S. K., Lundblad V. (1999). Est1 and Cdc13 act as comediators of telomerase access. *286*(5437):117-20.
- Fairlie, J., Holland, R., Pilkington, J. G., Pemberton, J. M., Harrington, L., & Nussey, D. H. (2016). Lifelong leukocyte telomere dynamics and survival in a free-living mammal. *Aging Cell*, *15*(1), 140–148. <https://doi.org/10.1111/accel.12417>
- Fajkus, J., Sýkorová, E., & Leitch, A. R. (2005). Telomeres in evolution and evolution of telomeres. *Chromosome Research*, *13*(5), 469–479. <https://doi.org/10.1007/s10577-005-0997-2>
- Farmery, J. H. R., Smith, M. L., Huissoon, A., Furnell, A., Mead, A., Levine, A. P., Manzur, A., Thrasher, A., Greenhalgh, A., Parker, A., Sanchis-Juan, A., Richter, A., Gardham, A., Lawrie, A., Sohal, A., Creaser-Myers, A., Frary, A., Greinacher, A., Themistocleous, A., ... Lynch, A. G. (2018). Telomerecat: A ploidy-agnostic method for estimating telomere length from whole genome sequencing data. *Scientific Reports*. <https://doi.org/10.1038/s41598-017-14403-y>
- Ferguson, B. M., & Fangman, W. L. (1992). A position effect on the time of replication origin activation in yeast. *Cell*, *68*(2), 333–339. [https://doi.org/10.1016/0092-8674\(92\)90474-Q](https://doi.org/10.1016/0092-8674(92)90474-Q)
- Feuerbach, L., Sieverling, L., Deeg, K. I., Ginsbach, P., Hutter, B., Buchhalter, I., Northcott, P. A., Lichter, P., Pfister, S. M., Jones, D. T. W., Rippe, K., & Brors, B. (2019). TelomereHunter: telomere content estimation and characterization from whole genome sequencing data. *BMC Bioinformatics*. <https://doi.org/10.1101/065532>.
- Fourel, G., Revardel, E., Koering, C. E., & Gilson, É. (1999). Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO Journal*, *18*(9), 2522–2537. <https://doi.org/10.1093/emboj/18.9.2522>
- Friedman, K. L., & Cech, T. R. (1999). Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants. *Genes and Development*, *13*(21), 2863–2874. <https://doi.org/10.1101/gad.13.21.2863>

- Frydrykova. (2002). Costs of secondary parasitism in the facultative hyperparasitoid *Pachycrepoideus dubius*: Does host size matter? *Entomologia Experimentalis et Applicata*, 103(3), 239–248. <https://doi.org/10.1023/A>
- Fuchs, J., Brandes, A., & Schubert, I. (1995). Telomere sequence localization and karyotype evolution in higher plants. *Plant Systematics and Evolution*, 196(3–4), 227–241. <https://doi.org/10.1007/BF00982962>
- Fulnečková, J., Hasíková, T., Fajkus, J., Lukešová, A., Eliáš, M., & Sýkorová, E. (2012). Dynamic evolution of telomeric sequences in the green algal order Chlamydomonadales. *Genome Biology and Evolution*, 4(3), 248–264. <https://doi.org/10.1093/gbe/evs007>
- Fulnečková, J., Ševčíková, T., Fajkus, J., Lukešová, A., Lukeš, M., Vlček, Č., Lang, B. F., Kim, E., Eliáš, M., & Sýkorová, E. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution*, 5(3), 468–483. <https://doi.org/10.1093/gbe/evt019>
- Fumagalli, M., Rossiello, F., Clerici, M., Barozzi, S., Kaplunov, J. M., Bucci, G., Dobрева, M., Matti, V., Beausejour, C. M., Herbig, U., & Longhese, M. P. (2013). Telomeric DNA damage is irreparable and causes persistent DNA damage response activation. *Nat Cell Biol.* 14(4): 355–365.
- Gao, H., Cervantes, R. B., Mandell, E. K., Otero, J. H., & Lundblad, V. (2007). RPA-like proteins mediate yeast telomere function. *Nature Structural and Molecular Biology*, 14(3), 208–214. <https://doi.org/10.1038/nsmb1205>
- Gao, J., & Munch, S. B. (2015). Does reproductive investment decrease telomere length in *Menidia menidia*? *PLoS ONE*, 10(5), 1–13. <https://doi.org/10.1371/journal.pone.0125674>
- Gatbonton, T., Imbesi, M., Nelson, M., Akey, J. M., Ruderfer, D. M., Kruglyak, L., Simon, J. A., & Bedalov, A. (2006). Telomere length as a quantitative trait: Genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.0020035>
- Gaughran. (2016). TP53 copy number expansion is associated with the evolution of increased body size and an enhanced DNA damage response in elephants. *ELife*, 5(September2016), 3–5. <https://doi.org/10.7554/eLife.11994>
- Geiger, S., Le Vaillant, M., Lebard, T., Reichert, S., Stier, A., Le Maho, Y., & Criscuolo, F. (2012). Catching-up but telomere loss: Half-opening the black box of growth and ageing trade-off in wild king penguin chicks. *Molecular Ecology*, 21(6), 1500–1510. <https://doi.org/10.1111/j.1365-294X.2011.05331.x>
- Gelinas, A. D., Paschini, M., Reyes, F. E., Héroux, A., Batey, R. T., Lundblad, V., & Wuttke, D. S. (2009). Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain. *Proceedings of the National Academy of Sciences of the United States of America*, 106(46), 19298–19303. <https://doi.org/10.1073/pnas.0909203106>
- Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., & Gasser, S. M. (1993). Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites. In *Journal of*

Molecular Biology (Vol. 231, Issue 2, pp. 293–310). <https://doi.org/10.1006/jmbi.1993.1283>

- Glade, M. J., & Meguid, M. M. (2015). A glance at ... telomeres, oxidative stress, antioxidants, and biological aging. *Nutrition*, 31(11–12), 1447–1451. <https://doi.org/10.1016/j.nut.2015.05.018>
- Gomes, N. M. V., Ryder, O. A., Houck, M. L., Charter, S. J., Walker, W., Forsyth, N. R., Austad, S. N., Venditti, C., Pagel, M., Shay, J. W., & Wright, W. E. (2011). Comparative biology of mammalian telomeres: Hypotheses on ancestral states and the roles of telomeres in longevity determination. *Aging Cell*, 10(5), 761–768. <https://doi.org/10.1111/j.1474-9726.2011.00718.x>
- Gottschling, D. E., Aparicio, O. M., Billington, B. L., & Zakian, V. A. (1990). Position effect at *S. cerevisiae* telomeres: Reversible repression of Pol II transcription. *Cell*, 63(4), 751–762. [https://doi.org/10.1016/0092-8674\(90\)90141-Z](https://doi.org/10.1016/0092-8674(90)90141-Z)
- Grandin, N., & Charbonneau, M. (2013). RPA provides checkpoint-independent cell cycle arrest and prevents recombination at uncapped telomeres of *Saccharomyces cerevisiae*. *DNA Repair*, 12(3), 212–226. <https://doi.org/10.1016/j.dnarep.2012.12.002>
- Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell*, 43(2 PART 1), 405–413. [https://doi.org/10.1016/0092-8674\(85\)90170-9](https://doi.org/10.1016/0092-8674(85)90170-9)
- Greider, C. W., & Blackburn, E. H. (1987). The telomere terminal transferase of tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, 51(6), 887–898. [https://doi.org/10.1016/0092-8674\(87\)90576-9](https://doi.org/10.1016/0092-8674(87)90576-9)
- Greider, C. W., & Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature*, 337(6205), 331–337. <https://doi.org/10.1038/337331a0>
- Hackett, J. A., Feldser, D. M., & Greider, C. W. (2001). Telomere dysfunction increases mutation rate and genomic instability. *Cell*. [https://doi.org/10.1016/S0092-8674\(01\)00457-3](https://doi.org/10.1016/S0092-8674(01)00457-3)
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell*, 144(5), 646–674. <https://doi.org/10.1016/j.cell.2011.02.013>
- Hansen, M. E. B., Hunt, S. C., Stone, R. C., Horvath, K., Herbig, U., Ranciaro, A., Hirbo, J., Beggs, W., Reiner, A. P., Wilson, J. G., Kimura, M., Vivo, I. De, Chen, M. M., Kark, J. D., Levy, D., Nyambo, T., Tishkoff, S. A., & Aviv, A. (2016). Shorter telomere length in Europeans than in Africans due to polygenetic adaptation. *Human Molecular Genetics*, 25(11), 2324–2330. <https://doi.org/10.1093/hmg/ddw070>
- Harari, Y., & Kupiec, M. (2014). Genome-wide studies of telomere biology in budding yeast. *Microbial Cell*, 1(3), 70–80. <https://doi.org/10.15698/mic2014.01.132>
- Hausmann, M. F., Winkler, D. W., O'Reilly, K. M., Huntington, C. E., Nisbet, I. C. T., & Vleck, C. M. (2003). Telomeres shorten more slowly in long-lived birds and mammals than in short-lived ones. *Proceedings of the Royal Society B: Biological Sciences*, 270(1522), 1387–1392. <https://doi.org/10.1098/rspb.2003.2385>
- Hausmann, M. F., Winkler, D. W., & Vleck, C. M. (2005). Longer telomeres associated with higher

survival in birds. *Biology Letters*, 1(2), 212–214. <https://doi.org/10.1098/rsbl.2005.0301>

- Hediger, F., Neumann, F. R., Van Houwe, G., Dubrana, K., & Gasser, S. M. (2002). Live Imaging of Telomeres. *Current Biology*, 12(24), 2076–2089. [https://doi.org/10.1016/s0960-9822\(02\)01338-6](https://doi.org/10.1016/s0960-9822(02)01338-6)
- Heidinger, B. J., Blount, J. D., Boner, W., Griffiths, K., Metcalfe, N. B., & Monaghan, P. (2012). Telomere length in early life predicts lifespan. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), 1743–1748. <https://doi.org/10.1073/pnas.1113306109>
- Henning, K. A., Moskowitz, N., Ashlock, M. A., & Liu, P. P. (1998). Humanizing the yeast telomerase template. *Proceedings of the National Academy of Sciences of the United States of America*, 95(10), 5667–5671. <https://doi.org/10.1073/pnas.95.10.5667>
- Higashiyama, T., Maki, S., & Yamada, T. (1995). Molecular organization of *Chlorella vulgaris* chromosome I: presence of telomeric repeats that are conserved in higher plants. *Mgg Molecular & General Genetics*, 246(1), 29–36. <https://doi.org/10.1007/BF00290130>
- Horowitz, H., Thorburn, P., & Haber, J. E. (1984). Rearrangements of highly polymorphic regions near telomeres of *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 4(11), 2509–2517. <https://doi.org/10.1128/mcb.4.11.2509>
- Huang, P. H., Pryde, F. E., Lester, D., Maddison, R. L., Borts, R. H., Hickson, I. D., & Louis, E. J. (2001). SGS1 is required for telomere elongation in the absence of telomerase. *Current Biology*, 11(2), 125–129. [https://doi.org/10.1016/S0960-9822\(01\)00021-5](https://doi.org/10.1016/S0960-9822(01)00021-5)
- Hunt, S. C., Chen, W., Gardner, J. P., Kimura, M., Srinivasan, S. R., Eckfeldt, J. H., Berenson, G. S., & Aviv, A. (2008). Leukocyte telomeres are longer in African Americans than in whites: The National Heart, Lung, and Blood Institute Family Heart Study and the Bogalusa Heart Study. *Aging Cell*, 7(4), 451–458. <https://doi.org/10.1111/j.1474-9726.2008.00397.x>
- Ijdo, J. W., Baldinit, A., Wardt, D. C., Reeders, S. T., & Wells, R. A. (1991). Origin of human chromosome 2: An ancestral telomere-telomere fusion. In *Proc. Nadl. Acad. Sci. USA* (Vol. 88).
- Johnson, F. B., Marciniak, R. A., McVey, M., Stewart, S. A., Hahn, W. C., & Guarente, L. (2001). The *Saccharomyces cerevisiae* WRN homolog Sgs1p participates in telomere maintenance in cells lacking telomerase. *EMBO Journal*, 20(4), 905–913. <https://doi.org/10.1093/emboj/20.4.905>
- Keane, M., Semeiks, J., Thomsen, B., Pedro, J., Keane, M., Webb, A. E., Li, Y. I., Quesada, V., Craig, T., Bruhn Madsen, L., van Dam, S., Brawand, D., Marques, P. I., Michalak, P., Kang, L., Bhak, J., Yim, H.-S., Grishin, N. V., Hjort Nielsen, N., ... Pedro de Magalhães, J. (2015). Insights into the Evolution of Longevity from the Bowhead Whale Genome Resource Insights into the Evolution of Longevity from the Bowhead Whale Genome. *Cell Reports*, 10, 112–122. <http://dx.doi.org/10.1016/j.celrep.2014.12.008><http://dx.doi.org/10.1016/j.celrep.2014.12.008>This is an open access article under the CC BY license <http://creativecommons.org/licenses/by/3.0/>
- Kirkwood. (1973). [Letters to nature]. *Nature*, 246(5429), 170. <https://doi.org/10.1038/246170a0>

- Koering, C. E., Fourel, G., Binet-Brasselet, E., Laroche, T., Klein, F., & Gilson, E. (2000). Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Research*, 28(13), 2519–2526. <https://doi.org/10.1093/nar/28.13.2519>
- Kong, C. M., Lee, X. W., & Wang, X. (2013). Telomere shortening in human diseases. *FEBS Journal*, 280(14), 3180–3193. <https://doi.org/10.1111/febs.12326>
- Kupiec, M. (2014). Biology of telomeres: Lessons from budding yeast. In *FEMS Microbiology Reviews*. <https://doi.org/10.1111/1574-6976.12054>
- Kuznetsova. (2020). Telomere structure in insects: A review. *Journal of Zoological Systematics and Evolutionary Research*, 58(1), 127–158. <https://doi.org/10.1111/jzs.12332>
- Laroche, T., Martin, S. G., Gotta, M., Gorham, H. C., Pryde, F. E., Louis, E. J., & Gasser, S. M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Current Biology*, 8(11), 653–657. [https://doi.org/10.1016/s0960-9822\(98\)70252-0](https://doi.org/10.1016/s0960-9822(98)70252-0)
- Larrivé, M., LeBel, C., & Wellinger, R. J. (2004). The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex. *Genes and Development*, 18(12), 1391–1396. <https://doi.org/10.1101/gad.1199404>
- Larrivé, M., & Wellinger, R. J. (2006). Telomerase- and capping-independent yeast survivors with alternate telomere states. *Nature Cell Biology*, 8(7), 741–747. <https://doi.org/10.1038/ncb1429>
- Le, S., Moore, J. K., Haber, J. E., & Greider, C. W. (1999). RAD50 and RAD51 define two pathways that collaborate to maintain telomeres in the absence of telomerase. *Genetics*, 152(1), 143–152.
- Lee, M., Napier, C. E., Yang, S. F., Arthur, J. W., Reddel, R. R., & Pickett, H. A. (2017). Comparative analysis of whole genome sequencing-based telomere length measurement techniques. *Methods*. <https://doi.org/10.1016/j.ymeth.2016.08.008>
- Lee, W. S., Monaghan, P., & Metcalfe, N. B. (2013). Experimental demonstration of the growth rate-lifespan trade-off. *Proceedings of the Royal Society B: Biological Sciences*, 280(1752). <https://doi.org/10.1098/rspb.2012.2370>
- Lendvay, T. S., Morris, D. K., Sah, J., Balasubramanian, B., & Lundblad, V. (1996). Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes. *Genetics*, 144(4), 1399–1412.
- Lieb, J. D., Liu, X., Botstein, D., & Brown, P. O. (2001). Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nature Genetics*, 28(4), 327–334. <https://doi.org/10.1038/ng569>
- Lin. (2003). Maintaining the Shape of Nerve Cells □. *Molecular Biology of the Cell*, 14(December), 5069–5081. <https://doi.org/10.1091/mbc.E03>
- Lin, C. Y., Chang, H. H., Wu, K. J., Tseng, S. F., Lin, C. C., Lin, C. P., & Teng, S. C. (2005). Extrachromosomal telomeric circles contribute to Rad52-, Rad50-, and polymerase δ -mediated telomere-telomere recombination in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, 4(2), 327–336. <https://doi.org/10.1128/EC.4.2.327-336.2005>

- Lingner, J., Cooper, J. P., & Cech, T. R. (1995). gXA ; 5'. *Science*, 269(September), 533–534.
- Lingner, J., Hughes, T. R., Shevchenko, A., Mann, M., Lundblad, V., & Cech, T. R. (1997). Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science*. 276(5312):561-7.
- Liti, G., Haricharan, S., Cubillos, F. A., Tierney, A. L., Sharp, S., Bertuch, A. A., Parts, L., Bailes, E., & Louis, E. J. (2009). Segregating YKU80 and TLC1 alleles underlying natural variation in telomere properties in wild yeast. *PLoS Genetics*, 5(9). <https://doi.org/10.1371/journal.pgen.1000659>
- Lopez, C. R., Ribes-Zamora, A., Indiviglio, S. M., Williams, C. L., Haricharan, S., & Bertuch, A. A. (2011). Ku must load directly onto the chromosome end in order to mediate its telomeric functions. *PLoS Genetics*, 7(8). <https://doi.org/10.1371/journal.pgen.1002233>
- Louis, E. J., & Haber, J. E. (1992). The structure and evolution of subtelomeric Y' repeats in *Saccharomyces cerevisiae*. *Genetics*, 131(3), 559–574.
- Luciano, P., Coulon, S., Faure, V., Corda, Y., Bos, J., Brill, S. J., Gilson, E., Simon, M. N., & Géli, V. (2012). RPA facilitates telomerase activity at chromosome ends in budding and fission yeasts. *EMBO Journal*, 31(8), 2034–2046. <https://doi.org/10.1038/emboj.2012.40>
- Lundblad, V., & Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell*, 73(2), 347–360. [https://doi.org/10.1016/0092-8674\(93\)90234-H](https://doi.org/10.1016/0092-8674(93)90234-H)
- Lundblad, V., & Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, 57(4), 633–643. [https://doi.org/10.1016/0092-8674\(89\)90132-3](https://doi.org/10.1016/0092-8674(89)90132-3)
- Lustig, A. J. (2003). Clues to catastrophic telomere loss in mammals from yeast telomere rapid deletion. *Nature Reviews Genetics*, 4(11), 916–923. <https://doi.org/10.1038/nrg1207>
- Lustig, A. J., Kurtz, S., & Shore, D. (1990). Involvement of the silencer and UAS binding protein RAP1 in regulation of telomere length. *Science*, 250(4980), 549–553. <https://doi.org/10.1126/science.2237406>
- Maciejowski, J., Li, Y., Bosco, N., Campbell, P. J., & De Lange, T. (2015). Chromothripsis and Kataegis Induced by Telomere Crisis. *Cell*, 163(7), 1641–1654. <https://doi.org/10.1016/j.cell.2015.11.054>
- Mak, H. C., Pillus, L., & Ideker, T. (2009). Dynamic reprogramming of transcription factors to and from the subtelomere. *Genome Research*, 19(6), 1014–1025. <https://doi.org/10.1101/gr.084178.108>
- Mandell, E. K., Gelinas, A. D., Wuttke, D. S., & Lundblad, V. (2011). Sequence-specific binding to telomeric DNA is not a conserved property of the Cdc13 DNA binding domain. *Biochemistry*, 50(29), 6289–6291. <https://doi.org/10.1021/bi2005448>
- Mangino, M., Christiansen, L., Stone, R., Hunt, S. C., Horvath, K., Eisenberg, D. T. A., Kimura, M., Petersen, I., Kark, J. D., Herbig, U., Reiner, A. P., Benetos, A., Codd, V., Nyholt, D. R., Sinnreich, R., Christensen, K., Nassar, H., Hwang, S. J., Levy, D., ... Aviv, A. (2015). DCAF4, a novel gene associated with leucocyte telomere length. *Journal of Medical Genetics*. <https://doi.org/10.1136/jmedgenet-2014-102681>

- Marcand, S., Brevet, V., & Gilson, E. (1999). Progressive cis-inhibition of telomerase upon telomere elongation. *EMBO Journal*, 18(12), 3509–3519. <https://doi.org/10.1093/emboj/18.12.3509>
- Marcand, S., Gilson, E., & Shore, D. (1997). A protein-counting mechanism for telomere length regulation in yeast. *Science*, 275(5302), 986–990. <https://doi.org/10.1126/science.275.5302.986>
- Marcand, S., Pardo, B., Gratiias, A., Cahun, S., & Callebaut, I. (2008). Multiple pathways inhibit NHEJ at telomeres. *Genes and Development*, 22(9), 1153–1158. <https://doi.org/10.1101/gad.455108>
- Marie-Nelly, H., Marbouty, M., Cournac, A., Flot, J.-F., Liti, G., Parodi, D. P., Syan, S., Guillén, N., Margeot, A., Zimmer, C., & Koszul, R. (2014). High-quality genome (re)assembly using chromosomal contact data. *Nature Communications*, 5, 5695. <https://doi.org/10.1038/ncomms6695>
- Martin, S. G., Laroche, T., Suka, N., Grunstein, M., & Gasser, S. M. (1999). Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell*, 97(5), 621–633. [https://doi.org/10.1016/S0092-8674\(00\)80773-4](https://doi.org/10.1016/S0092-8674(00)80773-4)
- Mathieu, N., Pirzio, L., Freulet-Marrière, M. A., Desmaze, C., & Sabatier, L. (2004). Telomeres and chromosomal instability. In *Cellular and Molecular Life Sciences*. <https://doi.org/10.1007/s00018-003-3296-0>
- McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics*, 26(2), 234–282.
- Metcalf, N. B., & Monaghan, P. (2001). Compensation for a bad start: Grow now, pay later? *Trends in Ecology and Evolution*, 16(5), 254–260. [https://doi.org/10.1016/S0169-5347\(01\)02124-3](https://doi.org/10.1016/S0169-5347(01)02124-3)
- Metcalf, N. B., & Monaghan, P. (2003). Growth versus lifespan: Perspectives from evolutionary ecology. *Experimental Gerontology*, 38(9), 935–940. [https://doi.org/10.1016/S0531-5565\(03\)00159-1](https://doi.org/10.1016/S0531-5565(03)00159-1)
- Meyne, J., Baker, R. J., Hobart, H. H., Hsu, T. C., Ryder, O. A., Ward, O. G., Wiley, J. E., Wurster-Hill, D. H., Yates, T. L., & Moyzis, R. K. (1990). Distribution of non-telomeric sites of the (TTAGGG)_n telomeric sequence in vertebrate chromosomes. *Chromosoma*, 99(1), 3–10. <https://doi.org/10.1007/BF01737283>
- Monaghan, P. (2010). Telomeres and life histories: The long and the short of it. In *Annals of the New York Academy of Sciences* (Vol. 1206, pp. 130–142). Blackwell Publishing Inc. <https://doi.org/10.1111/j.1749-6632.2010.05705.x>
- Monaghan, P., Eisenberg, D. T. A., Harrington, L., & Nussey, D. (2018). Understanding diversity in telomere dynamics. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1741). Royal Society Publishing. <https://doi.org/10.1098/rstb.2016.0435>
- Monaghan, P., & Haussmann, M. F. (2006). Do telomere dynamics link lifestyle and lifespan? In *Trends in Ecology and Evolution* (Vol. 21, Issue 1, pp. 47–53).

<https://doi.org/10.1016/j.tree.2005.11.007>

- Monaghan, P., & Ozanne, S. E. (2018). Somatic growth and telomere dynamics in vertebrates: Relationships, mechanisms and consequences. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1741). Royal Society Publishing. <https://doi.org/10.1098/rstb.2016.0446>
- Mozdy, A. D., Podell, E. R., & Cech, T. R. (2008). Multiple Yeast Genes, Including Paf1 Complex Genes, Affect Telomere Length via Telomerase RNA Abundance. *Molecular and Cellular Biology*, 28(12), 4152–4161. <https://doi.org/10.1128/mcb.00512-08>
- Mravinac. (2011). TCAGG, an alternative telomeric sequence in insects. *Chromosoma*, 120(4), 367–376. <https://doi.org/10.1007/s00412-011-0317-x>
- Mudd, A. B., Bredeson, J. V., Baum, R., Hockemeyer, D., & Rokhsar, D. S. (2020). Analysis of muntjac deer genome and chromatin architecture reveals rapid karyotype evolution. *Communications Biology*, 3(1), 1–10. <https://doi.org/10.1038/s42003-020-1096-9>
- Nan, H., Du, M., De Vivo, I., Manson, J. A. E., Liu, S., McTiernan, A., Curb, J. D., Lessin, L. S., Bonner, M. R., Guo, Q., Qureshi, A. A., Hunter, D. J., & Han, J. (2011). Shorter telomeres associate with a reduced risk of melanoma development. *Cancer Research*, 71(21), 6758–6763. <https://doi.org/10.1158/0008-5472.CAN-11-1988>
- Naumov, G. I., James, S. A., Naumova, E. S., Louis, E. J., & Roberts, I. N. (2000). Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus*, *Saccharomyces kudriavzevii* and *Saccharomyces mikatae*. *International Journal of Systematic and Evolutionary Microbiology*, 50(5), 1931–1942. <https://doi.org/10.1099/00207713-50-5-1931>
- Nersisyan, L., & Arakelyan, A. (2015). Computel: Computation of mean telomere length from whole-genome next-generation sequencing data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0125201>
- Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., Takio, S., Tamura, K., Chung, S. J., Nakamura, S., Kuroiwa, H., Tanaka, K., Sato, N., & Kuroiwa, T. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga *Cyanidioschyzon merolae*. *BMC Biology*, 5, 1–8. <https://doi.org/10.1186/1741-7007-5-28>
- Okuda, K. (2002). Telomere Length in the Newborn. *Pediatric Research*, 52(3), 377–381. <https://doi.org/10.1203/01.pdr.0000022341.72856.72>
- Olovnikov, A. M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of Theoretical Biology*, 41(1), 181–190. [https://doi.org/10.1016/0022-5193\(73\)90198-7](https://doi.org/10.1016/0022-5193(73)90198-7)
- Olsson, M., Pauliny, A., Wapstra, E., Uller, T., Schwartz, T., Miller, E., & Blomqvist, D. (2011). Sexual differences in telomere selection in the wild. *Molecular Ecology*, 20(10), 2085–2099. <https://doi.org/10.1111/j.1365-294X.2011.05085.x>
- Olsson, M., Wapstra, E., & Friesen, C. R. (2018). Evolutionary ecology of telomeres: a review. In

Annals of the New York Academy of Sciences (Vol. 1422, Issue 1, pp. 5–28). Blackwell Publishing Inc. <https://doi.org/10.1111/nyas.13443>

- Palladino, F., Laroche, T., Gilson, E., Axelrod, A., Pillus, L., & Gasser, S. M. (1993). SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres. *Cell*, 75(3), 543–555. [https://doi.org/10.1016/0092-8674\(93\)90388-7](https://doi.org/10.1016/0092-8674(93)90388-7)
- Parolini, M., Romano, A., Costanzo, A., Khoriauli, L., Santagostino, M., Nergadze, S. G., Canova, L., Rubolini, D., Giulotto, E., & Saino, N. (2017). Telomere length is reflected by plumage coloration and predicts seasonal reproductive success in the barn swallow. *Molecular Ecology*, 26(21), 6100–6109. <https://doi.org/10.1111/mec.14340>
- Pauliny, A., Devlin, R. H., Johnsson, J. I., & Blomqvist, D. (2015). Rapid growth accelerates telomere attrition in a transgenic fish. *BMC Evolutionary Biology*, 15(1), 1–10. <https://doi.org/10.1186/s12862-015-0436-8>
- Pauliny, A., Wagner, R. H., Augustin, J., Szép, T., & Blomqvist, D. (2006). Age-independent telomere length predicts fitness in two bird species. *Molecular Ecology*, 15(6), 1681–1687. <https://doi.org/10.1111/j.1365-294X.2006.02862.x>
- Peska, V., & Garcia, S. (2020). Origin, Diversity, and Evolution of Telomere Sequences in Plants. *Frontiers in Plant Science*, 11(February), 1–9. <https://doi.org/10.3389/fpls.2020.00117>
- Petracek, M. E., Lefebvre, P. A., Silflow, C. D., & Berman, J. (1990). Chlamydomonas telomere sequences are A+T-rich but contain three consecutive G·C base pairs. *Proceedings of the National Academy of Sciences of the United States of America*, 87(21), 8222–8226. <https://doi.org/10.1073/pnas.87.21.8222>
- Petreaca, R. C., Chiu, H. C., Eckelhoefer, H. A., Chuang, C., Xu, L., & Nugent, C. I. (2006). Chromosome end protection plasticity revealed by Stn1p and Ten1p bypass of Cdc13p. *Nature Cell Biology*, 8(7), 748–755. <https://doi.org/10.1038/ncb1430>
- Platt, J. M., Ryvkin, P., Wanat, J. J., Donahue, G., Ricketts, M. D., Barrett, S. P., Waters, H. J., Song, S., Chavez, A., Abdallah, K. O., Master, S. R., Wang, L. S., & Johnson, F. B. (2013). Rap1 relocalization contributes to the chromatin-mediated gene expression profile and pace of cell senescence. *Genes and Development*, 27(12), 1406–1420. <https://doi.org/10.1101/gad.218776.113>
- Plot, V., Criscuolo, F., Zahn, S., & Georges, J. Y. (2012). Telomeres, age and reproduction in a long-lived reptile. *PLoS ONE*, 7(7), 1–6. <https://doi.org/10.1371/journal.pone.0040855>
- Prasad, K. N., Wu, M., & Bondy, S. C. (2017). Telomere shortening during aging: Attenuation by antioxidants and anti-inflammatory agents. *Mechanisms of Ageing and Development*, 164(April), 61–66. <https://doi.org/10.1016/j.mad.2017.04.004>
- Pryde, F. E., & Louis, E. J. (1999). Limitations of silencing at native yeast telomeres. *EMBO Journal*, 18(9), 2538–2550. <https://doi.org/10.1093/emboj/18.9.2538>
- Puddu, F., Herzog, M., Selivanova, A., Wang, S., & Zhu, J. (2020). Genome architecture and stability in the *S. cerevisiae* knockout collection. *Nature*. 573(7774), 416–420. <https://doi.org/10.1038/s41586-019-1549-9>

- Puizina, J., Weiss-Schneeweiss, H., Pedrosa-Harand, A., Kamenjarin, J., Trinajstić, I., Riha, K., & Schweizer, D. (2003). Karyotype analysis in *Hyacinthella dalmatica* (Hyacinthaceae) reveals vertebrate-type telomere repeats at the chromosome ends. *Genome*, *46*(6), 1070–1076. <https://doi.org/10.1139/g03-078>
- Qi, X., Li, Y., Honda, S., Hoffmann, S., Marz, M., Mosig, A., Podlevsky, J. D., Stadler, P. F., Selker, E. U., & Chen, J. J. L. (2013). The common ancestral core of vertebrate and fungal telomerase RNAs. *Nucleic Acids Research*, *41*(1), 450–462. <https://doi.org/10.1093/nar/gks980>
- Reichert, S., & Stier, A. (2017). Does oxidative stress shorten telomeres in vivo? A review. *Biology Letters*, *13*(12). <https://doi.org/10.1098/rsbl.2017.0463>
- Reichert, S., Stier, A., Zahn, S., Arrivé, M., Bize, P., Massemin, S., & Criscuolo, F. (2014). Increased brood size leads to persistent eroded telomeres. *Frontiers in Ecology and Evolution*, *2*(APR), 1–11. <https://doi.org/10.3389/fevo.2014.00009>
- Ribaud, V., Ribeyre, C., Damay, P., & Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. *EMBO Journal*, *31*(1), 138–149. <https://doi.org/10.1038/emboj.2011.349>
- Ribes-Zamora, A., Mihalek, I., Lichtarge, O., & Bertuch, A. A. (2007). Distinct faces of the Ku heterodimer mediate DNA repair and telomeric functions. *Nature Structural and Molecular Biology*, *14*(4), 301–307. <https://doi.org/10.1038/nsmb1214>
- Richards, E. J., & Ausubel, F. M. (1988). Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, *53*(1), 127–136. [https://doi.org/10.1016/0092-8674\(88\)90494-1](https://doi.org/10.1016/0092-8674(88)90494-1)
- Risques, R. A., & Promislow, D. E. L. (2018). All's well that ends well: Why large species have short telomeres. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *373*(1741). <https://doi.org/10.1098/rstb.2016.0448>
- Rollings, N., Friesen, C. R., Sudyka, J., Whittington, C., Giraudeau, M., Wilson, M., & Olsson, M. (2017). Telomere dynamics in a lizard with morph-specific reproductive investment and self-maintenance. *Ecology and Evolution*, *7*(14), 5163–5169. <https://doi.org/10.1002/ece3.2712>
- Rollings, N., Uhrig, E. J., Krohmer, R. W., Waye, H. L., Mason, R. T., Olsson, M., Whittington, C. M., & Friesen, C. R. (2017). Age-related sex differences in body condition and telomere dynamics of red-sided garter snakes. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1852). <https://doi.org/10.1098/rspb.2016.2146>
- Romano, G. H., Harari, Y., Yehuda, T., Podhorzer, A., Rubinstein, L., Shamir, R., Gottlieb, A., Silberberg, Y., Pe'er, D., Ruppin, E., Sharan, R., & Kupiec, M. (2013). Environmental Stresses Disrupt Telomere Length Homeostasis. *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1003721>
- Rovatsos, M., Kratochvíl, L., Altmanová, M., & Pokorná, M. J. (2015). Interstitial telomeric motifs in squamate reptiles: When the exceptions outnumber the rule. *PLoS ONE*, *10*(8), 1–14. <https://doi.org/10.1371/journal.pone.0134985>
- Roy, R., Meier, B., McAinsh, A. D., Feldmann, H. M., & Jackson, S. P. (2004). Separation-of-function Mutants of Yeast Ku80 Reveal a Yku80p-Sir4p Interaction Involved in Telomeric

- Silencing. *Journal of Biological Chemistry*, 279(1), 86–94. <https://doi.org/10.1074/jbc.M306841200>
- Salomons, H. M., Mulder, G. A., Van De Zande, L., Haussmann, M. F., Linskens, M. H. K., & Verhulst, S. (2009). Telomere shortening and survival in free-living corvids. *Proceedings of the Royal Society B: Biological Sciences*, 276(1670), 3157–3165. <https://doi.org/10.1098/rspb.2009.0517>
- Savage, S. A. (2017). Dyskeratosis Congenita. *Congenital and Acquired Bone Marrow Failure*, 23(2), 225–233. <https://doi.org/10.1016/B978-0-12-804152-9.00018-X>
- Schechtman, M. G. (1990). Characterization of telomere DNA from *Neurospora crassa*. *Gene*, 88(2), 159–165. [https://doi.org/10.1016/0378-1119\(90\)90027-O](https://doi.org/10.1016/0378-1119(90)90027-O)
- Scherl, A., Coute, Y., De, C., Calle, A., Sanchez, J., Greco, A., Hochstrasser, D., Diaz, J., & Laennec, R. T. H. (2002). Functional Proteomic Analysis of Human Nucleolus □. *Molecular Biology of the Cell*, 13(November), 4100–4109. <https://doi.org/10.1091/mbc.E02>
- Schramke, V., Luciano, P., Brevet, V., Guillot, S., Corda, Y., Longhese, M. P., Gilson, E., & Géli, V. (2004). RPA regulates telomerase action by providing Est1p access to chromosome ends. *Nature Genetics*, 36(1), 46–54. <https://doi.org/10.1038/ng1284>
- Scott, N. M., Haussmann, M. F., Elsey, R. M., Trosclair, P. L., & Vleck, C. M. (2006). Telomere length shortens with body length in Alligator mississippiensis. *Southeastern Naturalist*, 5(4), 685–692. [https://doi.org/10.1656/1528-7092\(2006\)5\[685:TLSWBL\]2.0.CO;2](https://doi.org/10.1656/1528-7092(2006)5[685:TLSWBL]2.0.CO;2)
- Send, T. S., Gilles, M., Codd, V., Wolf, I., Bardtke, S., Streit, F., Strohmaier, J., Frank, J., Schendel, D., Sütterlin, M. W., Denniff, M., Laucht, M., Samani, N. J., Deuschle, M., Rietschel, M., & Witt, S. H. (2017). Telomere length in newborns is related to maternal stress during pregnancy. *Neuropsychopharmacology*, 42(12), 2407–2413. <https://doi.org/10.1038/npp.2017.73>
- Shampay, J., Szostakt, J. W., & Blackburn, E. H. (1984). *Maintained in Yeast*. 310(July), 1–4.
- Shay, J. W., & Wright, W. E. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology*, 1(1), 72–76. <https://doi.org/10.1038/35036093>
- Shore, D., & Nasmyth, K. (1987). Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements. *Cell*, 51(5), 721–732. [https://doi.org/10.1016/0092-8674\(87\)90095-X](https://doi.org/10.1016/0092-8674(87)90095-X)
- Simide, R., Angelier, F., Gaillard, S., & Stier, A. (2016). Age and heat stress as determinants of telomere length in a long-lived fish, the siberian sturgeon. *Physiological and Biochemical Zoology*, 89(5), 441–447. <https://doi.org/10.1086/687378>
- Singer, M. S., & Gottschling, D. E. (1994). TLC1: Template RNA component of *Saccharomyces cerevisiae* telomerase. *Science*, 266(5184), 404–409. <https://doi.org/10.1126/science.7545955>
- Smith, C. D., & Blackburn, E. H. (1999). Uncapping and Dereglulation of Telomeres Lead to Detrimental Cellular Consequences in Yeast. *J Cell Biol.* 145(2): 203–214.
- Sohanpal, B. K., Morzaria, S. P., Gobright, E. I., & Bishop, R. P. (1995). Characterisation of the telomeres at opposite ends of a 3 Mb *Theileria parva* chromosome. *Nucleic Acids Research*,

23(11), 1942–1947. <https://doi.org/10.1093/nar/23.11.1942>

- Sudyka, J., Arct, A., Drobniak, S., Dubiec, A., Gustafsson, L., & Cichoń, M. (2014). Experimentally increased reproductive effort alters telomere length in the blue tit (*Cyanistes caeruleus*). *Journal of Evolutionary Biology*, 27(10), 2258–2264. <https://doi.org/10.1111/jeb.12479>
- Sudyka, J. (2019). Does Reproduction Shorten Telomeres? Towards Integrating Individual Quality with Life-History Strategies in Telomere Biology. In *BioEssays* (Vol. 41, Issue 11). John Wiley and Sons Inc. <https://doi.org/10.1002/bies.201900095>
- Sudyka, J., Arct, A., Drobniak, S., Gustafsson, L., & Cichoń, M. (2016). Longitudinal studies confirm faster telomere erosion in short-lived bird species. *Journal of Ornithology*, 157(1), 373–375. <https://doi.org/10.1007/s10336-015-1304-4>
- Sun, J., Yu, E. Y., Yang, Y., Confer, L. A., Sun, S. H., Wan, K., Lue, N. F., & Lei, M. (2009). Stn1-Ten1 is an Rpa2-Rpa3-like complex at telomeres. *Genes and Development*, 23(24), 2900–2914. <https://doi.org/10.1101/gad.1851909>
- Sun, L., Tan, R., Xu, J., LaFace, J., Gao, Y., Xiao, Y., Attar, M., Neumann, C., Li, G. M., Su, B., Liu, Y., Nakajima, S., Levine, A. S., & Lan, L. (2015). Targeted DNA damage at individual telomeres disrupts their integrity and triggers cell death. *Nucleic Acids Research*, 43(13), 6334–6347. <https://doi.org/10.1093/nar/gkv598>
- Sýkorová, E., Lim, K. Y., Kunická, Z., Chase, M. W., Bennett, M. D., Fajkus, J., & Leitch, A. R. (2003). Telomere variability in the monocotyledonous plant order Asparagales. *Proceedings of the Royal Society B: Biological Sciences*, 270(1527), 1893–1904. <https://doi.org/10.1098/rspb.2003.2446>
- Szostak, J. W., & Blackburn, E. H. (1982). Cloning yeast telomeres on linear plasmid vectors. *Cell*, 29(1), 245–255. [https://doi.org/10.1016/0092-8674\(82\)90109-X](https://doi.org/10.1016/0092-8674(82)90109-X)
- Talley, J. M., DeZwaan, D. C., Maness, L. D., Freeman, B. C., & Friedman, K. L. (2011). Stimulation of yeast telomerase activity by the ever shorter telomere 3 (Est3) subunit is dependent on direct interaction with the catalytic protein Est2. *Journal of Biological Chemistry*, 286(30), 26431–26439. <https://doi.org/10.1074/jbc.M111.228635>
- Tarry-Adkins, J. L., Blackmore, H. L., Martin-Gronert, M. S., Fernandez-Twinn, D. S., McConnell, J. M., Hargreaves, I. P., Giussani, D. A., & Ozanne, S. E. (2013). Coenzyme Q10 prevents accelerated cardiac aging in a rat model of poor maternal nutrition and accelerated postnatal growth. *Molecular Metabolism*, 2(4), 480–490. <https://doi.org/10.1016/j.molmet.2013.09.004>
- Tarry-Adkins, J. L., Chen, J. H., Smith, N. S., Jones, R. H., Cherif, H., & Ozanne, S. E. (2009). Poor maternal nutrition followed by accelerated postnatal growth leads to telomere shortening and increased markers of cell senescence in rat islets. *The FASEB Journal*, 23(5), 1521–1528. <https://doi.org/10.1096/fj.08-122796>
- Teixeira, M. T., & Gilson, E. (2005). Telomere maintenance, function and evolution: The yeast paradigm. *Chromosome Research*, 13(5), 535–548. <https://doi.org/10.1007/s10577-005-0999-0>
- Teixeira, M. T., Arneric M., Sperisen P., Lingner J. (2004). Telomere Length Homeostasis Is Achieved via a Switch between Telomerase-Extendible and-Nonextendible States. *Cell*.

117(3):323-335.

- Teng, S. C., & Zakian, V. A. (1999). Telomere-Telomere Recombination Is an Efficient Bypass Pathway for Telomere Maintenance in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 19(12), 8083–8093. <https://doi.org/10.1128/mcb.19.12.8083>
- Teng, S. C., Chang, J., McCowan, B., & Zakian, V. A. (2000). Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational process. *Molecular Cell*, 6(4), 947–952. [https://doi.org/10.1016/S1097-2765\(05\)00094-8](https://doi.org/10.1016/S1097-2765(05)00094-8)
- Tomar, R. S., Zheng, S., Brunke-Reese, D., Wolcott, H. N., & Reese, J. C. (2008). Yeast Rap1 contributes to genomic integrity by activating DNA damage repair genes. *EMBO Journal*, 27(11), 1575–1584. <https://doi.org/10.1038/emboj.2008.93>
- Tricola, G. M., Simons, M. J. P., Atema, E., Boughton, R. K., Brown, J. L., Dearborn, D. C., Divoky, G., Eimes, J. A., Huntington, C. E., Kitaysky, A. S., Juola, F. A., Lank, D. B., Litwa, H. P., Mulder, E. G. A., Nisbet, I. C. T., Okanoya, K., Safran, R. J., Schoech, S. J., Schreiber, E. A., ... Haussmann, M. F. (2018). The rate of telomere loss is related to maximum lifespan in birds. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1741). <https://doi.org/10.1098/rstb.2016.0445>
- Tsipouri, V., Schueler, M. G., Hu, S., Dutra, A., Pak, E., Riethman, H., & Green, E. D. (2008). Comparative sequence analyses reveal sites of ancestral chromosomal fusions in the Indian muntjac genome. *Genome Biology*, 9(10). <https://doi.org/10.1186/gb-2008-9-10-r155>
- Ujvari, B., Biro, P. A., Charters, J. E., Brown, G., Heasman, K., Beckmann, C., & Madsen, T. (2017). Curvilinear telomere length dynamics in a squamate reptile. *Functional Ecology*, 31(3), 753–759. <https://doi.org/10.1111/1365-2435.12764>
- Ujvari, B., & Madsen, T. (2009). Short telomeres in hatchling snakes: Erythrocyte telomere dynamics and longevity in tropical pythons. *PLoS ONE*, 4(10), 1–5. <https://doi.org/10.1371/journal.pone.0007493>
- Ungar, L., Harari, Y., Toren, A., & Kupiec, M. (2011). Tor complex 1 controls telomere length by affecting the level of Ku. *Current Biology*, 21(24), 2115–2120. <https://doi.org/10.1016/j.cub.2011.11.024>
- Ungar, L., Yosef, N., Sela, Y., Sharan, R., Ruppin, E., & Kupiec, M. (2009). A genome-wide screen for essential yeast genes that affect telomere length maintenance. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkp259>
- Ventura, M., Catacchio, C. R., Sajjadian, S., Vives, L., Sudmant, P. H., Marques-Bonet, T., Graves, T. A., Wilson, R. K., & Eichler, E. E. (2012). The evolution of African great ape subtelomeric heterochromatin and the fusion of human chromosome 2. *Genome Research*, 22(6), 1036–1049. <https://doi.org/10.1101/gr.136556.111>
- Virta-pearlman, V., & Lundblad, V. (1996). Estl has the properties of a. *Genes & Development*, 3094–3104.
- Vítková, M., Král, J., Traut, W., Zrzavý, J., & Marec, F. (2005). The evolutionary origin of insect telomeric repeats, (TTAGG)_n. *Chromosome Research*, 13(2), 145–156.

<https://doi.org/10.1007/s10577-005-7721-0>

- Von Zglinicki, T. (2002). Oxidative stress shortens telomeres. *Trends in Biochemical Sciences*, 27(7), 339–344. [https://doi.org/10.1016/S0968-0004\(02\)02110-2](https://doi.org/10.1016/S0968-0004(02)02110-2)
- Walmsley, R. M., Chan, C. S. M., Tye, B. K., & Petes, T. D. (1984). Erratum: Unusual DNA sequences associated with the ends of yeast chromosomes (*Nature* (1984) 310 (157-160)). *Nature*, 311(5983), 280. <https://doi.org/10.1038/311280b0>
- Wang, W., & Lan, H. (2000). Rapid and parallel chromosomal number reductions in muntjac deer inferred from mitochondrial DNA phylogeny. *Molecular Biology and Evolution*, 17(9), 1326–1333. <https://doi.org/10.1093/oxfordjournals.molbev.a026416>
- Watson. (1972). *Nature New Biology*, 238, 37–38.
- Weiss, H., & Scherthan, H. (2002). Aloe spp. - Plants with vertebrate-like telomeric sequences. *Chromosome Research*, 10(2), 155–164. <https://doi.org/10.1023/A:1014905319557>
- Wellinger, R. J., Wolf, A. J., & Zakian, V. A. (1993). Saccharomyces Telomeres Acquire Single-Strand TG 1-3 Tails in Lat S Phase. *Cell*, 72.
- Wellinger, R. J., & Zakian, V. A. (2012). Everything you ever wanted to know about Saccharomyces cerevisiae telomeres: Beginning to end. In *Genetics*. <https://doi.org/10.1534/genetics.111.137851>
- Whittemore, K., Vera, E., Martínez-Nevado, E., Sanpera, C., & Blasco, M. A. (2019). Telomere shortening rate predicts species life span. *Proceedings of the National Academy of Sciences of the United States of America*, 116(30), 15122–15127. <https://doi.org/10.1073/pnas.1902452116>
- Wilbourn, R. V., Moatt, J. P., Froy, H., Walling, C. A., Nussey, D. H., & Boonekamp, J. J. (2018). The relationship between telomere length and mortality risk in non-model vertebrate systems: A meta-analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1741). <https://doi.org/10.1098/rstb.2016.0447>
- Wotton, D., & Shore, D. (1997). A novel Rap1p-interacting factor, Rif2p, cooperates with Rif1p to regulate telomere length in Saccharomyces cerevisiae. *Genes and Development*, 11(6), 748–760. <https://doi.org/10.1101/gad.11.6.748>
- Xu, M., Wu, X. B., Yan, P., & Zhu, H. T. (2009). Telomere length shortens with age in chinese alligators (*Alligator sinensis*). *Journal of Applied Animal Research*, 36(1), 109–112. <https://doi.org/10.1080/09712119.2009.9707042>
- Yang, D., Xiong, Y., Kim, H., He, Q., Li, Y., Chen, R., & Songyang, Z. (2011). Human telomeric proteins occupy selective interstitial sites. *Cell Research*, 21(7), 1013–1027. <https://doi.org/10.1038/cr.2011.39>
- Ye, J., Renault, V. M., Jamet, K., & Gilson, E. (2014). Transcriptional outcome of telomere signalling. *Nature Reviews Genetics*, 15(7), 491–503. <https://doi.org/10.1038/nrg3743>
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6),

913–924. <https://doi.org/10.1038/ng.3847>

Zappulla, D. C., Goodrich, K., & Cech, T. R. (2005). A miniature yeast telomerase RNA functions in vivo and reconstitutes activity in vitro. *Nature Structural and Molecular Biology*, 12(12), 1072–1077. <https://doi.org/10.1038/nsmb1019>

Zetka, M. C., & Müller, F. (1996). Telomeres in nematodes. *Seminars in Cell and Developmental Biology*, 7(1), 59–64. <https://doi.org/10.1006/scdb.1996.0009>

Zhu, L., Hathcock, K. S., Hande, P., Lansdorp, P. M., Seldin, M. F., & Hodes, R. J. (1998). Telomere length regulation in mice is linked to a novel chromosome locus. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15), 8648–8653. <https://doi.org/10.1073/pnas.95.15.8648>

Zhu, X., & Gustafsson, C. M. (2009). Distinct differences in chromatin structure at subtelomeric X and Y' elements in budding yeast. *PLoS ONE*, 4(7). <https://doi.org/10.1371/journal.pone.0006363>

Scientific problems addressed and objectives of this PhD

Genomes are progressively modified during their evolution leading to gene content variation, recombination, mutation and genetic exchange among species/subpopulations. The advent of next-generation sequencing technologies and their cost reduction increased the number of genomes available for evolutionary studies, opening the way to understand the molecular mechanisms involved in genome evolution.

During my university studies in Italy, I developed a deep interest for studying the genome and its alterations. When I had to choose where to perform my Master thesis, I chose Gianni Liti's team at the Institute for Research on Cancer and Aging (IRCAN) for two main reasons: first, his team is involved in genomics studies which make use of vast bioinformatic datasets and whose main goal is to characterize genome instability and population structure in yeast; second, I was happy to be able to learn advanced bioinformatic techniques applied to my main research interest and couple this with an experience abroad. During my Master, I applied state-of-the-art bioinformatic pipelines to study the origin of introgressions in yeast, and finalized this project during my PhD in the same institute.

Genome introgressions are widespread across the tree of life and initiate from archaic admixtures followed by repeated backcrossing to one parental species. However, how introgressions arise in reproductively isolated species, such as yeasts, has remained unclear. The genus *Saccharomyces* itself constitutes a sort of evolutionary paradox, as it comprises species which are reproductively isolated from each other (i.e. their hybrids cannot generate viable offspring) but yet they have a long history of hybridization and reticulate evolution. Thanks to these features, the genus *Saccharomyces* constitutes a perfect model system to study the molecular mechanisms of origin, maintenance and phenotypic consequences of genomic introgressions. Recently, the team was involved in “The 1002 Yeast Genomes Project” which provided the most comprehensive overview of *S. cerevisiae* genome evolution and population structure to date (Peter & De Chiara 2018). This study described, among others, four lineages with abundant introgressions from the sister species *S. paradoxus*. The Alpechin lineage, associated with olive-related environments, carried the largest amount of introgressed material. In the first chapter of this PhD thesis, I report the discovery of a *S. cerevisiae*-*S. paradoxus* hybrid that is the direct clonal descendant of an ancient hybridization event that founded the Alpechin lineage. The access to the yeast hybrid allowed us: i) to probe whether it is the the source of the Alpechin introgressions, ii) to study how introgressions could emerge and be maintained by reconstructing possible evolutionary scenarios in the laboratory.

At the end of my Master, I decided to move the focus of my research to study how telomeric sequence and length evolve and influence the phenotype of an organism. I started two projects which were conducted as a collaboration between the teams of Gianni Liti and Eric Gilson, both at the Institute for research on cancer and aging in Nice, as part of the Labex Signallife PhD programme. The collaboration allowed me and the projects to benefit from the expertise of both teams. Gianni Liti's team, *Population genomics and complex traits*, applies state-of-the-art bioinformatic and experimental techniques to understand general aspects of genome evolution in budding yeast and dissect the genetic architecture of complex phenotypic traits. Eric Gilson's team, *Telomeres, senescence and cancer*, studies the role of telomeres and their associated proteins in aging and in disease onset and progression. The expertise of the teams converged in my PhD projects to study the link between telomere length and sequence variation and organismal fitness using natural populations of budding yeast.

Telomeres are ribonucleoprotein structures situated at the end of chromosomes. Their essential function is to distinguish chromosomes ends from double-strand breaks (DSBs) and their homeostasis is maintained via the specialised enzyme telomerase. They are constituted by tandem repeats of DNA sequence whose length and motif vary across organisms. Telomere length has been measured only on a reduced number of yeast strains but little is known about its variation in wild habitats. Moreover, the quantitative contribution of telomere length to the organismal fitness remains largely unexplored due to the lack of an adequate number of strains and the extreme laboriousness of current measuring techniques. The recent publication of the genome and phenome of more than 1000 yeast strains thanks to the effort of the Liti team in the "1002 Yeast Genomes Project" constitutes an invaluable resource and enabled me to address important biological questions related to telomere biology. In the second chapter of the thesis, I describe a new bioinformatic method that I developed to estimate telomere length from whole-genome sequencing data, which is specifically tailored for yeast genomic data. I used these measurements to study: i) what is the degree of variation in telomere length and interstitial telomeric sequences across the *S. cerevisiae* population, with a special attention on strains isolated from natural environment, ii) what are the genetic and non-genetic factors which contribute to telomere length variation, iii) what is the relationship between telomeres and interstitial telomeric sequences.

Telomeres are variable not only in length, but also in sequence. Across the tree of life, there is a great variety of telomeric sequences, whose evolutionary origin is puzzling. Different taxa evolved variable telomeric sequences, and telomere-associated proteins had to evolve as well to adapt to the

new telomeres. The impact of variation in telomeric sequence on the organismal fitness is currently unknown. Yeast telomeres that have the same sequence as human telomeres can be generated by using a “humanized” version of *TLC1*, the gene encoding the RNA template of telomerase, in which the template region is altered to encode TTAGGG repeats. In the third chapter of the thesis, I carried out experimental evolution in both non-selective and selective conditions to determine: i) the impact of telomeric sequence variation on organismal fitness, ii) if and through which molecular mechanisms yeast adapts to new telomeric sequences.

Results

Chapter 4

A yeast living ancestor reveals the origin of genomic introgressions

This chapter is adapted from:

Melania D'Angiolo, Matteo De Chiara, Jia-Xing Yue, Agurtzane Irizar, Simon Stenberg, Karl Persson, Agnès Llored, Benjamin Barré, Joseph Schacherer, Roberto Marangoni, Eric Gilson, Jonas Warringer and Gianni Liti. A yeast living ancestor reveals the origin of genomic introgressions. *Nature*. <https://doi.org/10.1038/s41586-020-2889-1>.

Author contributions:

M.D., M.D.C., J.X.Y., J.W., G.L. designed the experiments; M.D., M.D.C., J.X.Y., A.I., S.S., K.P., A.L. B.B. performed and analysed the experiments; E.G., R.M., J.S., J.W. and G.L. contributed with resources and reagents; J.W. and G.L. conceived and supervised the project; M.D., J.W. and G.L. wrote the paper.

Abstract

Genome introgressions drive evolution across the animal, plant and fungal kingdoms. Introgressions initiate from archaic admixtures followed by repeated backcrossing to one parental species. However, how introgressions arise in reproductively isolated species, such as yeasts, has remained unclear. Here, we discovered a clonal descendant of the ancestral yeast hybrid that founded the extant *S. cerevisiae* Alpechin lineage, which carries abundant *S. paradoxus* introgressions. We show that this clonal descendant, hereafter defined as “living ancestor”, retained the ancestral genome structure of the first-generation hybrid with contiguous *S. cerevisiae* and *S. paradoxus* subgenomes. The ancestral first-generation hybrid underwent catastrophic genomic instability through more than a hundred mitotic recombination events, mainly manifesting as homozygous genome blocks generated by loss-of-heterozygosity. These homozygous sequence blocks rescue hybrid fertility by restoring meiotic recombination and are the direct origins of the introgressions present in the Alpechin lineage. We suggest a plausible route for introgression evolution through reconstruction of extinct stages and propose genome instability to allow hybrids to overcome reproductive isolation and enable introgressions to emerge.

Introduction

Genome introgressions drive evolution across the animal, plant and fungal kingdoms (Arnold et al., 2016; Edelman et al., 2019; Sun et al., 2012). Introgressions of genomic material between populations and species often reflect key historical and demographical events of the species evolution. For example, non-African and Oceanian human populations contain ~2% Neanderthal and 2-4% of Denisovan DNA respectively as a consequence of archaic admixture with extinct hominid populations (Wolf & Akey, 2018). Introgression begins with an ancestral hybridization followed by successive backcrossings with one parental species, resulting in bulk loss of haplotypes from the other subgenome (Harrison & Larson, 2014; Mallet, 2005; Martin & Jiggins, 2017). However, many biological species pairs may not form hybrids, or show nearly complete hybrid sterility (Greig, 2009). In the model genus *Saccharomyces*, sister species are known to hybridize (Marsit et al., 2017; Morales & Dujon, 2012), but the high sequence divergence between the hybrids' subgenomes inhibits meiotic recombination via the mismatch repair system (Hunter et al., 1996; Kao et al., 2010). The resulting viable gametes are rare and are almost always non-recombined, unfit and reproductively isolated from their parental backgrounds (Greig et al., 2002). Nevertheless, *Saccharomyces* introgressions are frequent and characterised by replacement of one parental haplotype by the other (Almeida et al., 2014; Barbosa et al., 2016; Liti et al., 2006, 2009; Peter et al., 2018) These small introgression blocks are scattered throughout the genome and imply very extensive historical recombination, which is hard to reconcile with reproductive isolation. Hybrids do form between more distantly-related yeast species (Gallone et al., 2019; Langdon et al., 2020), but do not result in introgressions. This is likely to be explained by the hybrids' higher sequence divergence, which reinforces reproductive isolation by abolishing recombination, and also by incompatibility between subsets of genes (Lee et al., 2008).

We recently sequenced the genomes of more than 1000 *S. cerevisiae* strains and described 26 well-defined lineages representing specific ecological niches or geographical areas (Peter et al., 2018). Four of these lineages, Alpechin, Brazilian bioethanol, Mexican agave and French Guiana, present abundant introgressions of highly diverged (~12%) sequence from the sister species *S. paradoxus*, derived from at least two ancient admixture events. The Alpechin lineage, associated with olive oil production, carries the largest amount of introgressed material, ranging from 4% to 5% in individual strains and combined covering 8% of the genome (Peter et al., 2018). Alpechin strains were mostly isolated from olive oil wastewater (Alpechin, in Spanish) in Spain. The Alpechin population has been recently resampled from similar olive-based environments, indicating that the clade is stable and characteristic of olive-domestication environments (Pontes et al., 2019). Here, we report the discovery of a *S. cerevisiae*-*S. paradoxus* hybrid that is the direct clonal descendant of an ancient hybridization event that founded the Alpechin lineage by sexual reproduction and backcrossing to

S. cerevisiae. For shorthand, we refer to this strain as the “living ancestor”, as it retains the genome structure of the ancestor, while noting that the present-day individual is separated from the actual ancestor by some evolutionary time. The ancestral hybrid genome was shaped by extensive genomic instability, resulting in mitotic recombination events that by loss-of-heterozygosity (LOH) generated over a hundred scattered blocks of homozygous DNA. We show that these LOH blocks restore recombination efficiency and gamete viability, explaining how the ancestral hybrid overcame the main barrier to introgression.

Results

A clonal descendant of the ancestral hybrid

We identified a single natural *S. cerevisiae*-*S. paradoxus* living ancestor hybrid that co-exists with the *S. cerevisiae* Alpechin clade in olive oil related niches (**Supplementary Table 1**). We combined long- and short-read deep sequencing and haplotype structure information from meiotic gametes to infer its genome structure. The diploid hybrid contains the complete set of 16 chromosomes from each parent species. First, we inferred the phylogenetic relationship of both the subgenomes of the living ancestor against all major *S. cerevisiae* and *S. paradoxus* subpopulations (Bergstrom et al., 2014; Peter et al., 2018; Yue et al., 2017). We identified the Alpechin lineage as the closest relative to the living ancestor’s *S. cerevisiae* subgenome and the European *S. paradoxus* population as the closest relative to the *S. paradoxus* subgenome (**Fig. 1a**).

Second, we compared the genomic locations of the *S. paradoxus* LOH blocks (with two copies of *S. paradoxus* DNA retained) in the living ancestor to the *S. paradoxus* introgressed regions in the Alpechin isolates (**Fig. 1b, Extended Data Fig. 1, Supplementary Tables 2-3**). Among the 114 Alpechin introgressions, 62 overlapped with the *S. paradoxus* LOH blocks in the living ancestor for 534736 bp (spanning 43872 markers). Given that the living ancestor has 9.7% of its genome in LOH with two *S. paradoxus* copies (as well as 5.5% with two *S. cerevisiae* copies) and the Alpechins have 8% of introgressed *S. paradoxus* genome, the expected number of markers randomly overlapping between LOH and introgressions would be only 8221 (two-tailed χ^2 test, $p < 0.00001$). We further compared the breakpoints of these 62 LOH blocks and the introgressed regions and found them to show remarkable genome-wide concurrence. Among 124 introgression boundaries, 90 were within 1 kb from the closest LOH boundary, with the vast majority sharing the exact same coordinates. These results imply that the *S. paradoxus* LOH blocks were the direct source of most of the Alpechins’ introgressions.

We then explored whether newly arisen heterozygous mutations within *S. paradoxus* LOH blocks, accumulating over time within the two initially identical DNA sequences, were also present in the Alpechins’ introgressions. Identical *de novo* mutations are highly unlikely to arise independently

and are therefore compelling evidence of shared ancestry. We confidently called 207 heterozygous sites within *S. paradoxus* LOH blocks and for each site inferred the mutated and the ancestral state. Of the 156 *de novo* mutations on the *S. cerevisiae* haplotype of the LOH blocks, 125 (80%) were present also on the *S. paradoxus* introgression blocks of the Alpechins (**Fig. 1b, Extended Data Fig. 2**). In contrast, only 3 of the 51 *de novo* mutations present in the *S. paradoxus* haplotype were detected in the Alpechin population. This is uniquely compatible with a model in which most of the retained introgressions are direct descendants of the *S. paradoxus* DNA copied into the *S. cerevisiae* subgenome during the LOH block formation in the ancestral hybrid (**Supplementary Discussion 1**). Introgressions non-overlapping with LOH blocks likely emerged by integration of additional *S. paradoxus* subgenome during meiotic and mitotic recombination after the split with the living ancestor.

Taken together, these results unequivocally show that the discovered hybrid is closely related to the Alpechin lineage, is a good proxy for the ancestral hybrid state and can serve as a “living ancestor” to understand the early evolution of the Alpechin lineage.

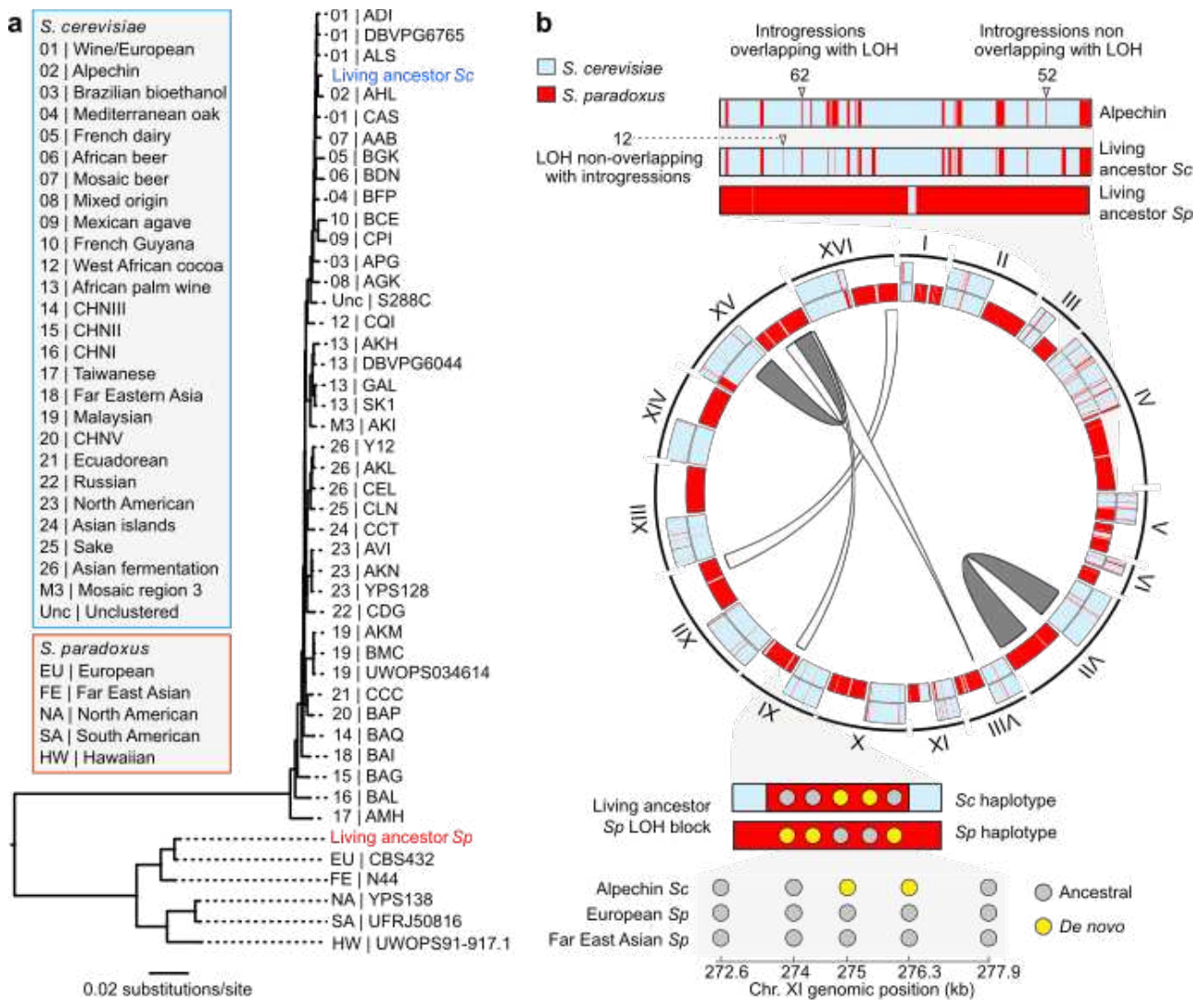


Figure 1

Fig. 1 | The ancestor of the Alpechin lineage. **a**, Phylogenetic tree generated using a matrix of 2258 1-to-1 orthologous genes across 46 *Saccharomyces* genomes. We name and number the *S. cerevisiae* lineages as in the 1002 Yeast Genome Project, and the *S. paradoxus* lineages as in the Yeast Population Reference Panel (left boxes). Lineage codes are reported in the phylogenetic tree, followed by strain name. The living ancestor *S. cerevisiae* subgenome (Living ancestor Sc) tightly clusters with the Alpechin lineage, while the *S. paradoxus* subgenome (Living ancestor Sp) has European ancestry. **b**, Circular plot of the living ancestor's (inner ring) and a pan-introgression Alpechin's (outer ring) genomes. Internal ribbons link regions involved in gross chromosomal rearrangements (grey intra-homolog and white inter-homolog) in the living ancestor, which were reverted to produce colinear maps. Genomic coordinates of *S. cerevisiae* and *S. paradoxus* chromosomes are based on the genome assemblies of the DBVPG6765 and CBS432 strains, respectively. The upper part represents a zoom-in of chromosome IV. Numbers refer to genome-wide overlap between *S. paradoxus* LOH and introgressions. The bottom part represents a zoom-in of the living ancestor's chromosomes XI with analysis of 5 heterozygous sites (circles) within a *S.*

paradoxus LOH block. Yellow and grey circles respectively represent *de novo* mutations and ancestral alleles. Genotypes of the Alpechins (bottom box) indicate inheritance of *de novo* mutations from the *S. cerevisiae* haplotype. In contrast, *de novo* mutations are absent from European and Far East Asian *S. paradoxus*, consistent with their appearance post-dating hybridization and LOH formation.

Genome instability in the ancestral hybrid

We found the living ancestor's subgenomes to have remained phased, with contiguous *S. cerevisiae* and *S. paradoxus* haplotypes, meaning that this clonal descendant never underwent a single meiosis after the initial, ancient hybridization (**Fig. 1b, Extended Data Fig. 3**). In contrast, the exceptional number of LOH blocks suggested a massive mitotic genome instability event. We detected 102 LOH blocks, together representing 15.2% of its genome, dispersed across the 16 chromosome pairs (**Supplementary Table 2**). The LOH blocks are disproportionately common in some chromosomes (e.g. chr. IV) and have complex patterns consistent with multiple nested recombination events in others (e.g. chr. V). The LOH size varies over 4-orders of magnitude (0.085 to 245 kb), with events shorter than 10 kb being enriched (**Extended Data Fig. 4a**). We found more interstitial ($n=92$) than terminal LOH blocks ($n=10$) and the former were much smaller (11 vs. 67 kb, average), consistent with these two LOH types originating from distinct repair mechanisms (Symington et al., 2014). We detected a strong LOH bias for *S. paradoxus* rather than *S. cerevisiae* homozygosity blocks (78 vs. 24, covering 1092514 and 611623 bp respectively) and investigated this by evolving eight parallel mutation accumulation lines in which we propagated the living ancestor asexually for ~2220 generations through 120 single cell bottlenecks (Dutta et al., 2017; Tattini et al., 2019). Together, the mutation accumulation lines acquired 37 new LOH with no bias for either parental haplotype (**Extended Data Fig. 4b, Supplementary Table 2, Supplementary Discussion 2**). This is consistent with the parental bias towards *S. paradoxus* LOH blocks in the living ancestor being due to selection favouring retention of these blocks (Dunn et al., 2013; Bozdog et al., 2019; Smukowski Heil et al., 2017), rather than to mutational mechanisms promoting their formation.

We also found signatures of genome instability in five gross chromosomal rearrangements and in chromosome I aneuploidy. The rearrangements are mostly restricted to the *S. paradoxus* subgenome and likely postdate the hybridization (**Extended Data Fig. 3, Supplementary Discussion 3**). We conclude that extensive genome instability followed the hybridization, resulting in exceptional levels of LOH and rearrangements.

LOH blocks rescue hybrid fertility

To validate that the living ancestor can overcome the reproductive isolation barrier, we tested

whether it produces viable meiotic offspring. Indeed, the gamete viability, 3.93% (**Supplementary Table 4**), was 5.9-fold higher than the 0.67% observed in experimental *S. cerevisiae*-*S. paradoxus* crosses (Liti et al., 2006). This level of gamete viability is remarkable considering the presence of multiple chromosomal rearrangements that strongly exacerbate the reproductive isolation (Liti et al., 2006; Hou et al., 2014). We estimated a theoretical gamete viability of 13.97%, if the living ancestor did not have the chromosomal rearrangements (**Supplementary Discussion 3**). One attractive explanation is that LOH blocks mediate crossing-over by providing dispersed segments of sequence identity, and thereby promote correct chromosome segregation (Rogers et al., 2018). The genome sequences of 25 viable living ancestor's gametes revealed that 76% of the crossovers (125 of 165; two-tailed χ^2 test, $p=1.56 \times 10^{-104}$) occurred within 1 kb of a LOH block, validating this hypothesis (**Fig. 2a, b, Extended Data Fig. 5a**).

We detected 133 recombinant chromosomes in the 25 gametes, with 106 single, 23 double, 3 triple and 1 quadruple recombination events (**Extended Data Fig. 5a**). This corresponds to 6.6 crossovers per viable gamete (0.4 per chromosome), but it is an underestimation since crossovers on chromosome I, V and IX were undetected due to aneuploidy and large LOH. Nevertheless, it markedly exceeds the 2.7 crossovers per viable gamete previously measured in *S. cerevisiae*-*S. paradoxus* experimental hybrids (Kao et al., 2010). The combination of novel crossovers and pre-existing LOH produces mosaic chromosomes with *S. paradoxus* haplotypes that are comparable in size to the Alpechins' introgression blocks (**Extended Data Fig. 5b-d**).

The crossovers mediated by LOH blocks promoted correct chromosome segregation and strongly limited the emergence of aneuploidies (mean of 1.12 per spore). LOH content and aneuploidy rate were negatively correlated ($r=-0.31$) across the gametes. Chromosomes involved in heterozygous inter-homolog translocations segregated to produce balanced genomes in viable gametes, underscoring their deleterious effect (**Extended Data Fig. 3, 5a, Supplementary Discussion 3**).

We probed whether the LOH blocks in the living ancestor also guide subsequent mitotic recombination events and new LOH events. We revisited the mutation accumulation lines and found that at least 9/37 new LOH were primed by pre-existing ones (**Extended Data Fig. 6a, Supplementary Table 2**). Next, we investigated *de novo* LOH formation in cell populations expanding clonally for a few generations and in cells returned to clonal growth after an aborted meiosis (**Fig. 2c, d**) (Laureau et al., 2016; Mozzachiodi et al., 2020). We observed higher LOH formation rate in the living ancestor compared to an experimental *S. cerevisiae*-*S. paradoxus* first-generation hybrid with subgenomes resembling those of the ancestral hybrid before the genome instability. This is consistent with the pre-existing LOH blocks promoting recombination. Sequencing the genomes of 18 clones derived from the living ancestor, we found that new LOH events were almost always mediated by pre-existing LOH blocks (19/21 events, **Fig. 2c, Extended**

Data Fig. 6a, Supplementary Table 2).

Finally, we investigated if these new LOH blocks further increased the fertility of the living ancestor. An additional large LOH block, emerging in cells having expanded clonally for a few generations, increased the gamete viability by 47% (to a median of 5.77%), in line with that LOH blocks help restore fertility in interspecies hybrids (**Extended Data Fig. 6b, c, Supplementary Tables 4, 5, 10, Supplementary Discussion 4**). We conclude that the massive instability of the ancestral hybrid genome led to exceptional levels of LOH and that these blocks induced and guided both meiotic and mitotic recombination, thereby helping it overcome the sterility barrier.

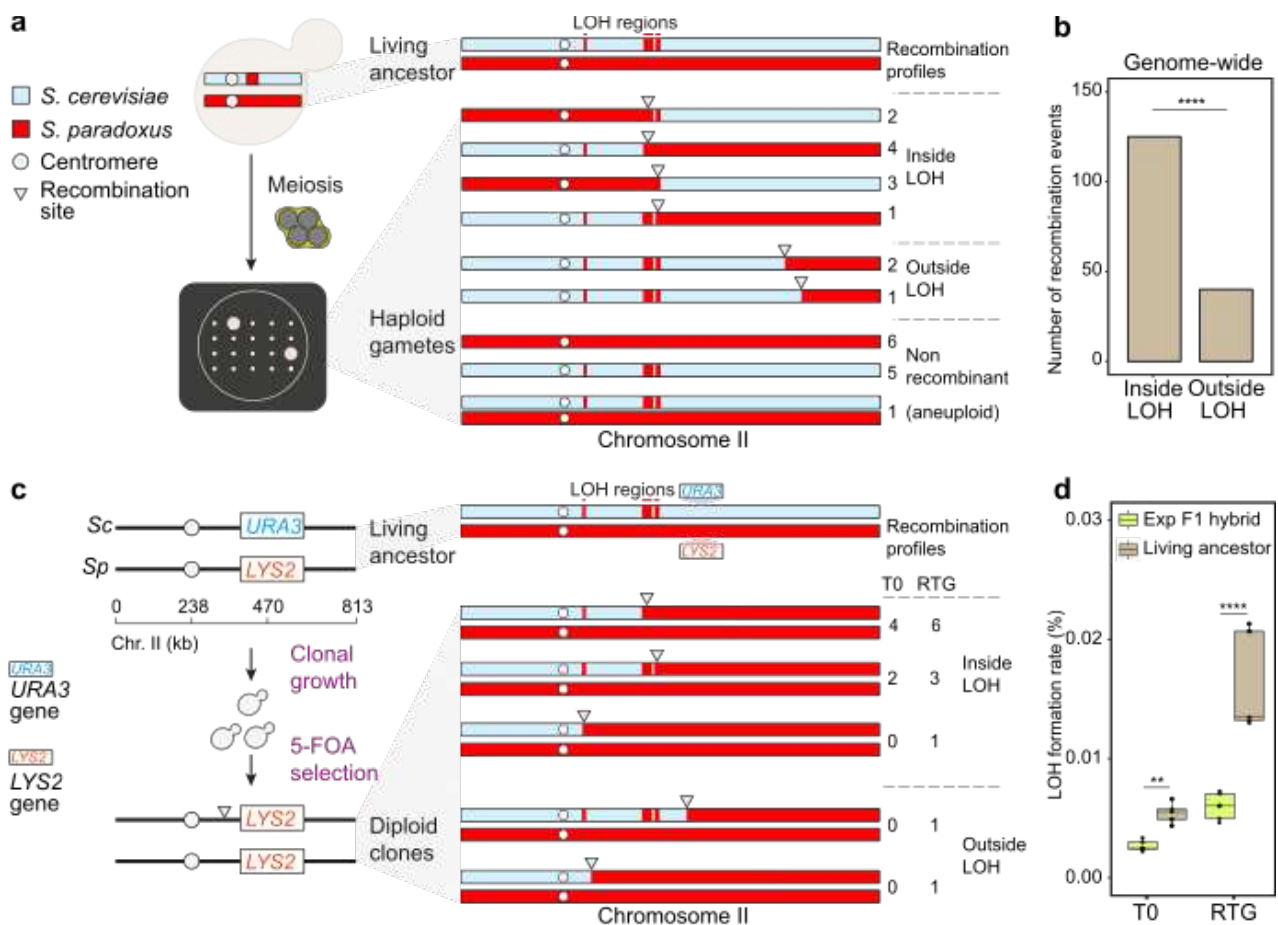


Figure 2

Fig. 2 | LOH blocks guide recombination of the hybrid's subgenomes. a, Twenty-five haploid meiotic gametes were isolated from the living ancestor and whole-genome sequenced. Chromosome II meiotic recombination patterns in haploid gametes indicate that meiotic crossover sites (arrows) are highly enriched within LOH blocks (red lines). The number of spores that share a similar chromosome II recombination profile is reported. **b**, The bar plot represents the total number of recombination sites inside/outside LOH blocks genome-wide (two-tailed χ^2 test, $p=1.56 \times 10^{-104}$), ****: $p < 0.0001$. **c**, The living ancestor chromosome II was engineered with a heterozygous *URA3* gene at the native *LYS2* locus. This system enables us to select LOH events at the *LYS2* locus by

plating diploid clones in 5-Fluoroorotic Acid (5-FOA) media, which selects for *URA3* loss. Six mitotic (T0) and twelve return-to-growth (RTG) clones were isolated and genome-sequenced. Chromosome II analysis of the diploid clones shows that most of the recombination events are mediated by pre-existing LOH homozygous blocks (red lines). The *URA3* gene is embedded in the living ancestor's *S. cerevisiae* chromosome II and therefore selects for events that generate a *S. paradoxus* homozygous region spanning the *LYS2* locus. **d**, LOH formation rate in an experimental *S. cerevisiae*-*S. paradoxus* F1 hybrid and the living ancestor using a *URA3* loss fluctuation assay, with heterozygous *URA3* engineered as illustrated in panel **c** for both strain backgrounds. The living ancestor has a higher rate both in mitotic growth (T0) and in returned-to-growth (RTG) experiments (two-tailed Mann-Whitney test, $p=0.008$ and $p=8.66e^{-05}$, respectively; $n=5$ biologically independent samples for each genetic background and condition). Boxes: horizontal line: median, upper/lower hinge: interquartile range (IQR), whiskers: max/min values. **: $p<0.01$, ****: $p<0.0001$.

Recreating the history of introgressions

The unique availability of a living ancestor and its modern descendants allowed us to reconstruct and probe the key evolutionary stages leading to genome introgressions (**Fig. 3, Extended Data Fig. 7a, Supplementary Tables 4, 5**). First, we investigated the reproductive capacity of an experimental *S. cerevisiae*-*S. paradoxus* first-generation hybrid. As expected (Greig et al., 2002; Hunter et al., 1996; Liti et al., 2006), this hybrid produced gametes, i.e. it sporulated efficiently, but these were almost always inviable (median of 0.5% viability).

Second, we investigated a first generation of gametes derived from the living ancestor, which restored a diploid state by mating-type switching and haplo-selfing (McClure et al., 2018). This F1 generation had variable sporulation efficiency and gamete viability, with the detrimental effects of chromosomal aneuploidies explaining much of this variation (**Extended Data Fig. 7b-d**).

Third, we explored the reproductive capacity of the next evolutionary stage by backcrossing eight haploid-stable F1 living ancestor gametes to each parental species. The living ancestor gametes were more fertile with *S. cerevisiae* than with *S. paradoxus* (median 22.12% vs 13.89%, two-tailed Mann-Whitney test, $p=0.008$). The gamete viability of the backcrosses varied, with negative effects of the chromosomal aneuploidies (**Extended Data Fig. 7e**). The higher fertility with *S. cerevisiae* may help explain why this backcross was favoured, but randomness, *S. cerevisiae* having been more common and its backcross being fitter, may have also played roles.

Fourth, we induced meiosis in four extant Alpechin strains and found gamete viability to be generally high (median 93.17%; one outlier at 25%). One strain (AQA), containing a relatively high amount of heterozygous introgressions (**Extended Data Fig. 8**), produced gametes that were highly viable (89%), but half of these reproduced slowly asexually. By sequencing fit and unfit offspring,

we mapped this monogenic growth defect to a locus located far from the closest introgression. Thus, we found no signs of introgressions impairing meiosis, sporulation or germination in the Alpechin strains.

Finally, we explored whether the Alpechin strains remained fertile with both parental species. We crossed a haploid Alpechin derivative with *S. cerevisiae* and *S. paradoxus*. Gametes of the *S. cerevisiae* (91.03%) but not the *S. paradoxus* backcross (0.74%) were viable, showing that the Alpechins are reproductively isolated from the *S. paradoxus* ancestor, despite retaining 5% of its genome as introgressions, while they are fertile with a *S. cerevisiae* lineage lacking *S. paradoxus* introgressions.

Overall, our reconstructions show how the Alpechin clade and its introgressions could evolve from the ancestral hybrid and its LOH regions. Molecular dating of the divergence indicates that over a million asexual generations have occurred since the LOH began, while a maximum of ~400000 asexual generations have occurred since the living ancestor and the Alpechin lineages split (**Fig. 3, Supplementary Table 6**). We estimated a maximum chronological time of ~4000 years since their last shared common ancestor, meaning that they likely diversified after the emergence of olive oil production (Kaniewski et al., 2012).

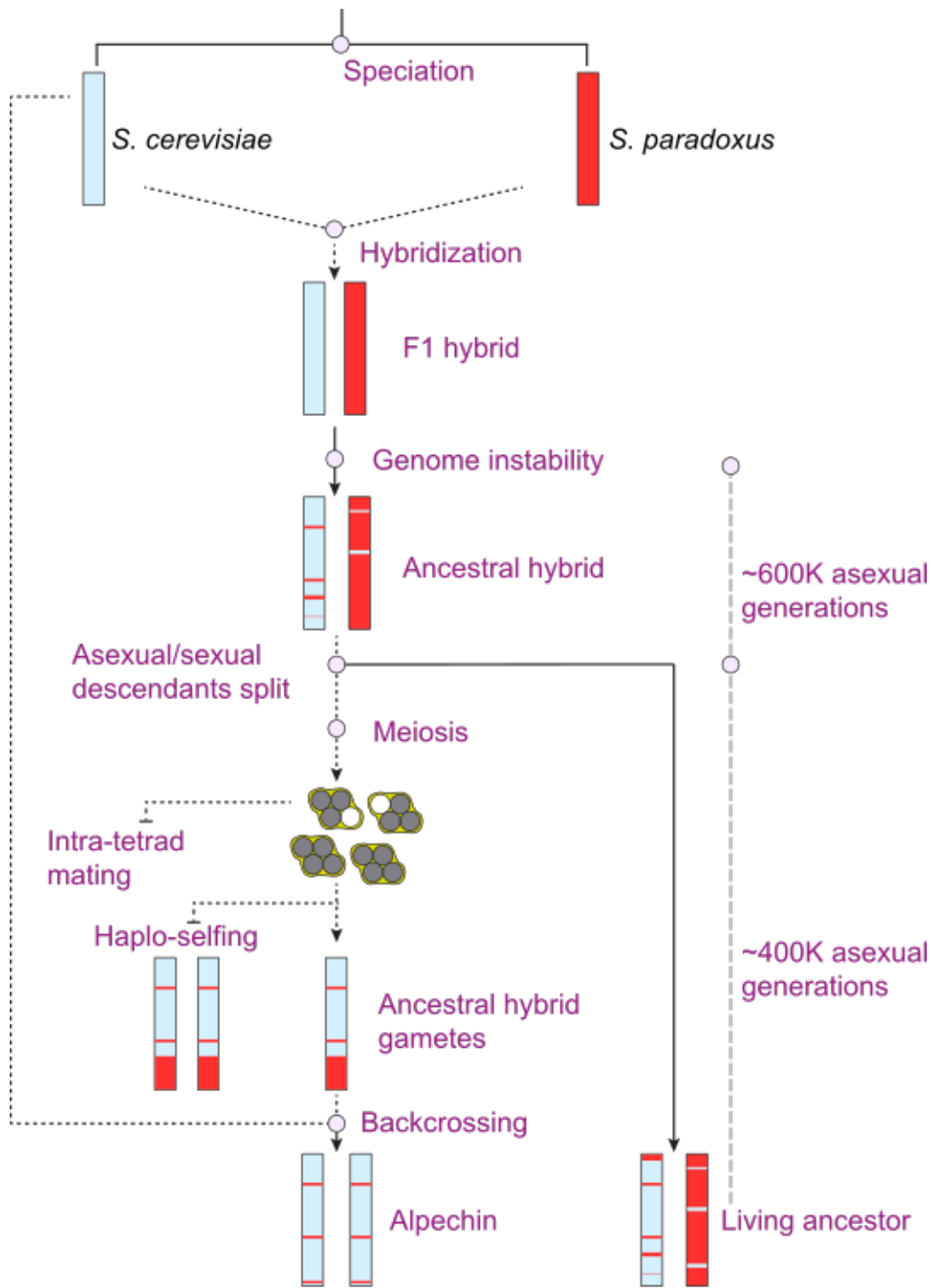


Figure 3

Fig. 3 | Reconstructing the origin of introgressions. Reconstruction of missing stages of the most likely route of the hybridization-to-introgression process. Plain lines point to naturally occurring stages (parental species, living ancestor and Alpechins), while dotted lines point to stages reproduced in the lab (experimental *S. cerevisiae*-*S. paradoxus* F1 hybrid, living ancestor gametes and backcrosses). Red and blue bars represent *S. paradoxus* and *S. cerevisiae* ancestry, respectively. The timespan inferred by molecular dating is represented by dotted grey lines and is expressed as number of asexual generations. White and grey circles respectively depict viable and dead meiotic

gametes held as a tetrad within an ascus. Their low viability prevents intra-tetrad mating while their low fitness after germination disfavours haplo-selfing. The combination of these events promotes the backcrossing process leading to the extant Alpechin lineage.

The living ancestor's asexual fitness

We next sought to understand why the living ancestor and the Alpechin descendants coexist in nature, while intermediate stages have not been isolated. We therefore measured their clonal reproduction in many environments permitting mitosis, and their survival after nutrients depletion. The living ancestor was often fitter than both *S. cerevisiae* and *S. paradoxus* parent species and the experimental F1 hybrid (in 40, 83 and 28% of environments respectively, **Fig. 4a, Extended Data Fig. 9a, Supplementary Table 7**). Gametes derived from the living ancestor suffered widespread (77/82 environments) and severe (33% average increase in doubling time) defects in asexual reproduction. Extensive survival defects amplified, rather than compensated for, these defects (**Extended Data Fig. 9b**). We propose that the fitness defects arise from recessive incompatibilities between genes (Bozdag et al., 2019), which are unmasked in the gametes' mosaic *S. cerevisiae* and *S. paradoxus* genomes (**Extended Data Fig. 5b**).

Mating the haploid gametes of the living ancestor back to the haploid *S. cerevisiae* or *S. paradoxus* parents restored fitness. We found only rare (3/82 environments in *S. cerevisiae* backcross, 10/82 in *S. paradoxus* backcross) and weak (1% average increase of doubling time for both backcrosses) asexual reproduction defects relative to the living ancestor (**Fig. 4a, Extended Data Fig. 9a, Supplementary Table 7**). This is consistent with the recessive incompatibilities being masked by the first round of backcrossing.

Across many environments, the asexual reproduction of the living ancestor closely resembled that of a sub-group of the Alpechins and only rarely and weakly deviated from that of the Alpechin lineage as a whole (worse in 6%, better in 6% of environments; **Fig. 4b, Extended Data Fig. 10a**). The Alpechins better resembled their closest extant *S. cerevisiae* parent than their *S. paradoxus* parent, but they enjoyed some niche-specific advantages relative to the former (in 11% of environments, **Extended Data Fig. 10b**), reflecting adaptations potentially conferred by *S. paradoxus* introgressions. Supporting that introgressed genes were retained due to selection, we found them to overlap non-randomly (83/492, two-tailed χ^2 test, $p < 0.00001$) with genes independently introgressed and retained in other *S. cerevisiae* subpopulations (**Fig. 4c, d, Extended Data Fig. 10c**) (Peter et al., 2018). These overlapping genes represent at least one additional and separate hybridization event, involving the American rather than the European *S. paradoxus* population. Further supporting selection, Alpechin introgressions were enriched in genes mediating interactions with the environment (**Supplementary Tables 8, 9**).

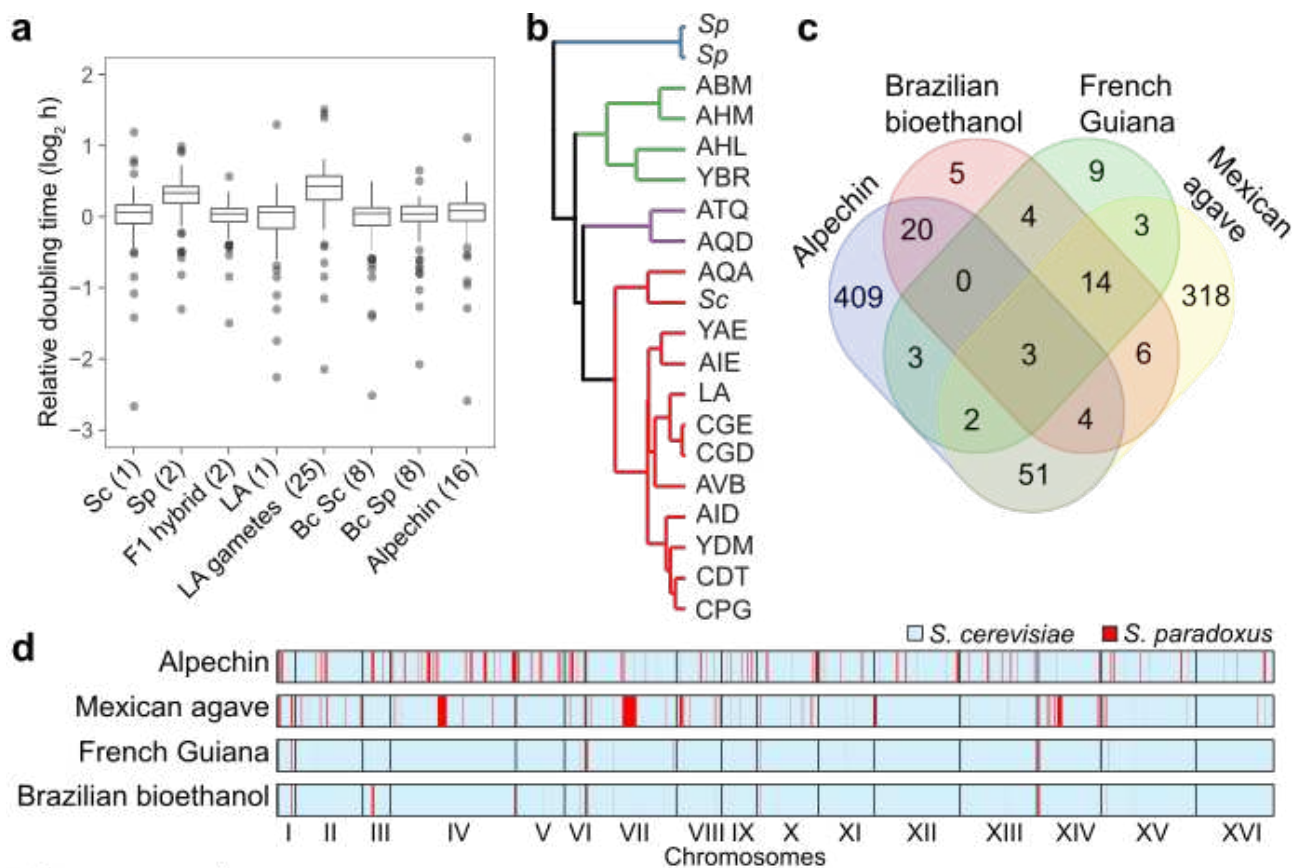


Figure 4

Fig. 4 | The asexual fitness landscape. **a**, Relative doubling time for each hybridization-to-introgression stage resulting from the samples in each group (indicated in brackets) and their 8 replicates (40 for the living ancestor) across 82 environments. *S. cerevisiae* (Sc, $n=656$), *S. paradoxus* (Sp, $n=1312$), experimental first-generation hybrid (F1 hybrid, $n=1312$), living ancestor (LA, $n=3280$), living ancestor gametes (LA gametes, $n=16400$), backcross to *S. cerevisiae* (Bc Sc, $n=5248$), backcross to *S. paradoxus* (Bc Sp, $n=5248$), Alpechin ($n=10496$). Boxes: horizontal line: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR. **b**, Hierarchical clustering (Pearson's r , complete linkage) of the living ancestor (LA), the Alpechin isolates (three-letter names) and the closest living relatives of their *S. cerevisiae* (Sc) and *S. paradoxus* (Sp) parents. Strains were grouped based on doubling time similarity (**Extended Data Fig. 10a**) and tree branches colour indicates the four distinct phenotypic clusters. **c-d**, Overlap of introgressed genes (**c**) and regions (**d**) in four *S. cerevisiae* lineages with abundant *S. paradoxus* genetic material (red segments).

Discussion

We discovered a yeast clonal descendant of an ancient hybridization event that enabled us to understand the early evolution of the *S. cerevisiae* Alpechin lineage and its abundant *S. paradoxus* introgressions. We showed how genome instability in the ancestral hybrid allowed it to by-pass the

hybrid sterility barrier and initiated the formation of introgressions. Specifically, the genome instability promoted recombination, which in turn reduced aneuploidies and substantially rescued the gametes' viability. Most of the asci contained single viable spores, preventing intra-tetrad mating, while the low fitness of self-mated progeny disfavoured haplo-selfing (**Fig. 3**). These conditions directed the gametes toward backcrossing, which restored both asexual fitness and much of the hybrid fertility in a single life cycle step. Additional mechanisms such as whole genome duplication (Charron et al., 2019; Greig et al., 2002) and intercrosses among gametes might have contributed to the late phase of the introgression evolution.

How can an ancient hybridization state compete with its modern descendants in the same ecological niche? Wild yeasts enter meiosis once per 1000 generations (Tsai et al., 2008) but the living ancestor has survived >1000x as long, without sexual re-shuffling of its genome and without entering into a protective spore state, even though it is capable of both sex and sporulation. This life cycle reprogramming likely reflects relaxation of selection for sex and gamete formation due to living in a human-made environment (Magwene et al., 2011). Despite its long asexual evolution, the living ancestor remain competitive with the sexually active descendants, matching the fitness of the Alpechins in most niches and occasionally surpassing them. This is consistent with divergent adaptation, with selection maintaining two distinct stages of the introgression process in different sub-niches.

The *S. paradoxus* genes that introgressed into the *S. cerevisiae* Alpechins are not a random collection of functions, but what we expect given adaptive benefits. Introgressed genes from Denisovans and Neanderthals retained in modern human populations have driven adaptation to high altitude (Huerta-Sánchez et al., 2014) and virus defense (Enard & Petrov, 2018), although widespread depletion of Neanderthal introgressions in functional genomic regions likely reflects a general selection against Neanderthal variants (Wolf & Akey, 2018).

Introgressions can be key to reconstructing species histories. In humans, the restriction of Neanderthal introgressions to out-of-Africa populations and of Denisovan introgressions to Oceanian and Asian populations inform on their historical range, co-existence and contacts. Analogously, the scarcity of *S. paradoxus* introgressions into *S. cerevisiae* before the out-of-China expansion and diversification of the latter, and their abundance after, inform on yeast species range, co-existence and contacts⁵. Introgressions from an unknown hominin into Denisovans (Prüfer et al., 2014) and from unknown yeasts into Taiwanese *S. cerevisiae*⁵ inform on extinct or undiscovered species. The recent discovery of a first-generation Neanderthal-Denisovan hybrid individual provided direct evidence of admixture (Slon et al., 2019), like our yeast living ancestor hybrid. Unlike an ancient genome, access to the yeast living ancestor from which introgressions derive, and the capacity to reconstruct other evolutionary stages along the path, enables direct probing of the

species history. Isolation of additional intermediate hybridization-to-introgression stages could further illuminate the emergence, evolution and function of interspecies introgressions.

Supplementary discussion

1, The origin of the living ancestor and the Alpechin lineage

We propose a scenario that places the living ancestor hybrid as an ancestral state of the Alpechin lineage (model 1). We also probe an alternative route that starts from the extant Alpechins to generate the living ancestor via a recent cross with a *S. paradoxus* (model 2). Model 1 postulates that the LOH blocks were the direct sources of the Alpechins' introgressions. Model 2 instead postulates that introgressions were the direct sources of the living ancestor's *S. paradoxus* LOH. Below, we compare the two models and test them against data on genome structure and single nucleotide variation. We reject model 2, which is incompatible with these data and therefore improbable (**Extended Data Fig. 2**).

Genome structure data. First, we compared the genomic locations of the *S. paradoxus* LOH blocks in the living ancestor to the *S. paradoxus* introgressed regions in the Alpechin isolates. Among the 114 introgressions, 62 overlapped with the living ancestor's *S. paradoxus* LOH blocks, while 52 did not have a *S. paradoxus* LOH counterpart. In model 1, these 52 introgressions that lack LOH counterparts derive from the *S. paradoxus* subgenome of the hybrid. The process first involves meiotic or mitotic recombinations (**Fig. 2a, c**) and then is followed by progressive re-sectioning through repeated backcrossing with the *S. cerevisiae* parent. We also assume that the ancestral hybrid might have existed as a population with various individuals that differed somewhat by their LOH patterns and additional LOH accumulated in the individual/s that directly founded the Alpechins. In model 2, the living ancestor arose when an Alpechin strain not containing *S. paradoxus* DNA at these 52 loci hybridized with a European *S. paradoxus* strain. Both scenarios are entirely plausible and the presence of 52 introgressions without a *S. paradoxus* LOH counterpart does not argue in favour of either model.

Second, we do not find 12 of the living ancestor's *S. paradoxus* LOH blocks in any extant Alpechin isolate (assuming that we sampled the whole Alpechins' introgression diversity). In model 1, these 12 LOH arose after the split of the living ancestor with the descendant that produced the gametes that founded the Alpechin lineage. This is entirely plausible because in model 1 the split between the living ancestor hybrid and the Alpechin lineage is ancient, providing long time for additional genome restructuring events to occur. Alternatively, these 12 LOH blocks could have been lost from the initial Alpechin progenitor during the backcrossing phase. In model 2, the living ancestor must have acquired these LOH blocks after the hybridization of its Alpechin and *S. paradoxus* parents. Since the hybridization event in this scenario is more recent, this model is less likely because the

time available for the occurrence of these 12 rare events is much shorter.

Third, the living ancestor's genome contains five gross chromosomal rearrangements (**Extended Data Fig. 3**). None of these rearrangements are present in any extant *S. paradoxus* or *S. cerevisiae* lineage, making it likely that they post-date the hybridization. Two of the rearrangements resulted in chimeric *S. cerevisiae* and *S. paradoxus* chromosomes, consistent with occurring after the hybridization. Gross chromosomal rearrangements are exceedingly rare events in natural *Saccharomyces* evolution (Fischer et al., 2000). Thus, their presence in the living ancestor becomes challenging to explain if this genome is recent (model 2). They therefore argue in favour of the living ancestor being ancient (model 1).

Single nucleotide variants. First, we described in the main text the analysis of newly arisen heterozygous mutations within the *S. paradoxus* LOH blocks in the living ancestor. A large proportion (125 out of 156) of the *de novo* mutations that are present in the *S. cerevisiae* haplotype of the living ancestor's *S. paradoxus* LOH blocks are also present in the introgressions of the Alpechins. A small fraction (3 of the 51) of the *de novo* mutations present in the *S. paradoxus* haplotype of the living ancestor's *S. paradoxus* LOH blocks are also shared with the Alpechins. In model 1, both types of heterozygous mutations in the *S. paradoxus* LOH blocks post-dated the LOH events and the LOH blocks passed these mutations on to the Alpechins' introgressions. The more prevalent inheritance of *de novo* mutations from the *S. cerevisiae* rather than the *S. paradoxus* haplotype suggests that the ancestral spore/s from which the Alpechin strains derive mostly inherited *S. cerevisiae* chromosomes in regions flanking the *S. paradoxus* LOH blocks, which were easy to integrate by recombination in a *S. cerevisiae* background during backcrosses. However, the inheritance of 3 heterozygous mutations from the *S. paradoxus* haplotype can be explained if a few of the inherited *S. cerevisiae* chromosomes recombined: i.e. they contained regions with *S. paradoxus* DNA, which is fully supported by the meiotic recombination observed in the living ancestor's gametes (**Extended Data Fig. 5a-d**). In model 2, the 125 heterozygous mutations shared between the *S. cerevisiae* haplotype of the living ancestor's *S. paradoxus* LOH blocks and the Alpechins' introgressions must have occurred in the Alpechin parent, before the hybridization. The 3 heterozygous mutations that are shared between the *S. paradoxus* haplotype of the living ancestor's *S. paradoxus* LOH regions and the Alpechins' introgressions cannot be explained in this way: if they occurred in the Alpechin parent before the hybridization it is very hard to explain their presence in the living ancestor's *S. paradoxus* haplotype. Model 2 can explain these mutations only by mitotic recombination within the *S. paradoxus* LOH blocks, subsequent to the hybridization event. Given that the hybridization must be recent and mitotic recombination events are rare, this makes model 2 a less likely scenario and argues in favour of model 1.

The most exclusive evidence that favours model 1 against model 2 is contained in the 24 LOH

blocks of the living ancestor that carry two copies of the parental *S. cerevisiae* genome. In both model 1 and model 2, the origin of these LOH regions must post-date the hybridization event. Consequently, every heterozygous *de novo* mutation present in these LOH blocks must also post-date the LOH event itself. If model 1 is correct, some of these mutations should be present in the Alpechin strains because their origin in this model also post-dates the hybridization. Mutations that are quite recent, i.e. mutations that post-date the origin of the Alpechin strains, will not be shared. On the contrary, if model 2 is correct, none of these heterozygous *de novo* mutations present in *S. cerevisiae* LOH regions should be present in the Alpechin strains, because in this scenario the Alpechin strains pre-date the hybridization. This distinction provides a strict test to check which of the two models best fits the empirical data. We analysed 34 heterozygous mutations present in the 24 *S. cerevisiae* LOH blocks of the living ancestor and found that 20 mutations are shared by the Alpechins. Among these, 17 derived from the *S. cerevisiae* haplotype while 3 derived from the *S. paradoxus* haplotype. This can only be realistically explained by model 1. Together with the indicative data above, it strongly argues that the Alpechins originated from partially recombined living ancestor's spore/s and that both the *S. cerevisiae* and *S. paradoxus* subgenomes of the hybrid contributed to the Alpechin strains' genome. Taken together, we find the available data to be fully consistent with model 1 and partially or completely incompatible with model 2. We therefore conclude that the genome of the living ancestor is indeed an ancient, and not a recent, hybridization event.

2, *De novo* LOH during mitotic laboratory evolution

Evolving the living ancestor under laboratory protocols allowed us to monitor how its genome evolves and investigate the impact of newly acquired mutational events. The consecutive single-cell bottlenecks imposed by the mutation accumulation protocol resulted in the accumulation of mutations consistent with evolution under near-neutral conditions (Sharp et al., 2018; Tattini et al., 2019). We investigated the living ancestor's mutation accumulation lines to probe whether the newly acquired LOH were biased toward the *S. paradoxus* haplotype. The living ancestor has over three times more LOH blocks homozygous for *S. paradoxus* compared to LOH blocks homozygous for *S. cerevisiae* (78 vs 24) (**Supplementary Table 2**). If such bias has a mechanistic origin, the hybrid's *S. cerevisiae* subgenome should accumulate more DNA lesions such as DNA double-strand breaks (DSBs) to bias the LOH toward the *S. paradoxus* genotype. This would require *cis*-acting factors directing the DSBs toward the *S. cerevisiae* subgenome. In a fully mitotic scenario, this could be achieved through heterozygous fragile sites (Chumki et al., 2016). If cells were returned to mitotic growth following an aborted meiosis (Laureau et al., 2016; Mozzachiodi et al., 2020), this could be induced through biased Spo11p DNA cleavage. Alternatively, the bias could be derived

from parental quantitative differences in DSBs repair efficiency. Currently, there are no studies showing parental *cis*-acting factors influencing fragile sites, DSBs formation or recombination dynamics in yeast diploid hybrids that can support this scenario.

If the bias is not mechanistic, both types of LOH formed at similar rates and subsequent selection favoured the maintenance of *S. paradoxus* LOH by two non-mutually exclusive mechanisms. First, punctuated LOH can be adaptive and arise within few generations under strong selective regimes (Gerstein et al., 2014; James et al., 2019; Vázquez-García et al., 2017), even in highly diverged *Saccharomyces* interspecies hybrid subgenomes (Dunn et al., 2013; Smukowski Heil et al., 2017). Second, sequential accumulation of LOH can be constrained by negative genetic interactions (Bozdag et al., 2019) with initial events toward one parent favouring the maintenance of subsequent LOH with the same parental haplotype.

In the mutation accumulation lines, we observed an equal number of novel LOH per parental subgenome (19 in *S. cerevisiae* and 18 in *S. paradoxus*). This is consistent with the rate of *de novo* LOH formation not being biased toward one parent, and instead supports that subsequent selection has driven the LOH parental bias in the living ancestor. The overall size distribution of *de novo* LOH events resembles the size of the pre-existing LOHs, with terminal events that are larger compared to interstitial ones (**Extended Data Fig. 4b, Supplementary Table 2**). Consistently, interstitial LOH are assumed to be short and to originate mainly from the repair of mismatches in heteroduplex DNA tracts, resulting from two-ended homologous recombination events, whereas terminal events can be ascribed to break-induced replication producing long LOH that extend to chromosome ends (Symington et al., 2014). It is interesting to notice that under this near-neutral evolutionary scenario, novel terminal LOH events can be up to three times larger compared to the pre-existing ones, consistent with the hypothesis that large events can emerge but are likely purged during evolution. These results are also consistent with the accumulation of LOH in *S. cerevisiae*-*S. paradoxus* experimental hybrids not being biased toward a parental haplotype (Tattini et al., 2019). We propose that selection has consistently favoured *S. paradoxus* LOH or alternatively the initial accumulation of *S. paradoxus* LOH was driven by adaptation (Smukowski Heil et al., 2017) and that this constrained the maintenance of subsequent events toward the same parent to prevent negative epistatic interactions (Bozdag et al., 2019).

3, Genomic rearrangements in the living ancestor's genome

We discovered five gross chromosomal rearrangements in the living ancestor's genome. We checked the presence of these rearrangements in a panel of reference-quality *S. cerevisiae* and *S. paradoxus* genomes and other outgroup species (Yue et al., 2017). All rearrangements are unique to the living ancestor and therefore likely post-date the hybridization. We catalogued the

rearrangements as intra-homolog and inter-homolog translocations (**Extended Data Fig. 3**). Specifically, there is a recombination between the two homologous chromosomes VII-L with the breakpoint mapping to one LOH event that likely pre-dated and mediated the translocation by providing sequence identity. A second rearrangement is a reciprocal translocation between chromosomes *Sp* XII-R and *Sp* XVI-R. The other events are more complex, involving multiple chromosomes, and likely occurred in multiple steps. This series of events consists of: recombination between the two XV-R homologs, two non-reciprocal translocations (part of the recombined *Sp* XV-R on the *Sp* VIII-L, and *Sp* XI-L on the recombined *Sc* XV-R). The result of this multi-step rearrangement is a highly rearranged triplet of chromosomes derived from chromosomes VIII, XI and XV. The remaining parts of *S. paradoxus* chromosomes XI and XV are stabilized by subtelomeric translocations. More precisely, the translocation of the right subtelomere of chromosome XV stabilizes chromosome XI, and the translocation of the right subtelomere of chromosome VIII stabilizes chromosome XV. We confirmed the global chromosomal structure by pulsed-field gel electrophoresis (data not shown). Rearrangements arise mainly from ectopic recombination events involving repetitive sequences, such as Ty elements or tRNAs (Fischer et al., 2000; Yue et al., 2017). However, repetitive elements did not mediate all the translocations we detected.

We investigated the non-recombinant chromosomes in the living ancestor's F1 gametes and found little evidence of biased chromosome inheritance, except for chromosome I, which mostly inherited the *S. paradoxus* copy, and chromosome XI, which mostly inherited the *S. cerevisiae* copy. This biased chromosome I inheritance is expected, as the living ancestor has two copies of this *S. paradoxus* chromosome. Non-reciprocal translocation in the *S. paradoxus* chromosome can explain the result for chromosome XI. For every translocation, we observed co-segregation of the chromosomes carrying the rearrangements in the F1 gametes, consistent with the constraints imposed by the structural variants to produce balanced complete genomes.

The large inter-chromosomal translocations described above have therefore a considerable impact on gamete viability. In *Saccharomyces* crosses, large translocations cause 50% loss of gamete viability when they are reciprocal, and 25% when they are non-reciprocal, due to the formation of multi-valents during meiosis (Liti et al., 2006). In a theoretical scenario of maximum gamete viability (100%), one reciprocal translocation and two non-reciprocal ones would decrease the maximum viability to 28.125%. Given that we observed a spore viability of 3.93% in the living ancestor, subtracting the additive detrimental impact of the three translocations results in a theoretical 13.97% viability. This value is remarkable considering that experimental *S. cerevisiae*-*S. paradoxus* first-generation hybrids without LOH have only 0.5-1% viability, and underscores the impact of how dispersed blocks of homozygous regions embedded in highly diverged genomes can

restore recombination and ensure faithful chromosome segregation, thus promoting gamete viability (Rogers et al., 2018).

4, LOH and gamete viability

We exploited laboratory evolved living ancestor clones to test whether newly acquired LOH events can further increase gamete viability by ensuring correct segregation of the chromosomes carrying them. We dissected a median of 196 spores (total of 4984 spores) in the 8 mutation accumulation lines, as well as in the clones derived from very short-term clonal mitotic expansion (6 mitotic clones and 12 return-to-growth clones), which were selected for an additional LOH on chromosome II (**Fig. 2c**). The chromosome II of the living ancestor has 3 small LOH blocks, with one block located close to the centromere where recombination is suppressed (**Fig. 2a**). This highly heterozygous configuration with limited homozygous blocks resulted in 50% of the sequenced spores inheriting a non-recombined chromosome II (**Fig. 2a, Extended Data Fig. 5a**).

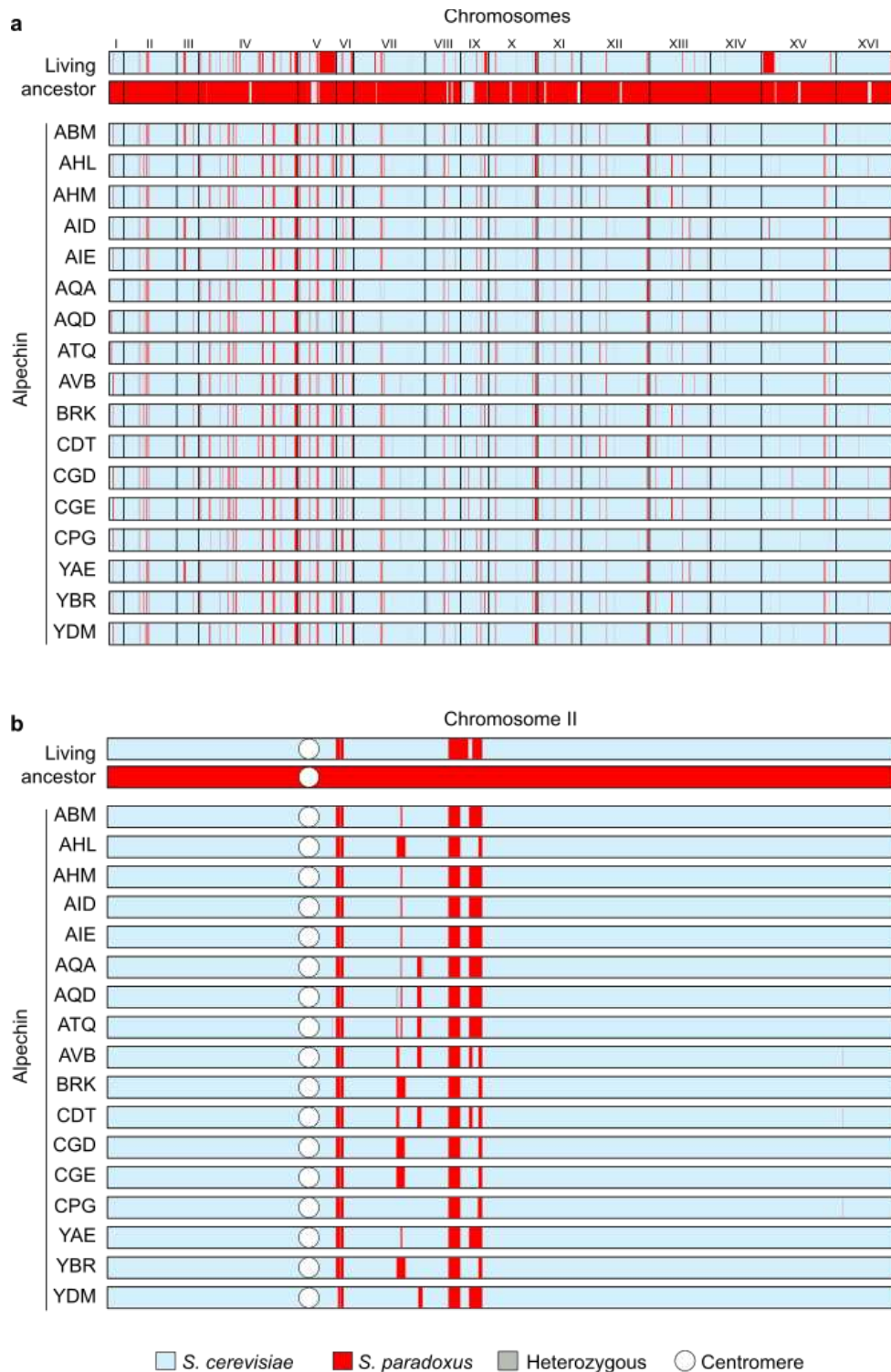
The spore viability of the six mitotic clones provides direct support for the causative role of LOH in the recovery of fertility. Specifically, all these mitotically evolved clones with an additional large LOH on chromosome II increased viability, with an average of 47% increase (i.e. from 3.93% to a median of 5.77%, **Extended Data Fig. 6c, Supplementary Table 4**). This increase in viability can largely be attributed to the additional LOH, which makes chromosome II homozygous for at least half of its length (**Fig. 2c**) and therefore provides a large region for recombination to occur. Since these clones were propagated only for a few cell divisions, we do not expect them to carry many additional, potentially confounding mutations. Consistently, genome analysis revealed only a single novel mutation in the six sequenced clones (**Supplementary table 10**).

The return-to-growth samples had a more marginal median increase (4.13%, **Extended Data Fig. 6c, Supplementary Table 4**) in spore viability, despite that they also should benefit from the additional chromosome II LOH and do not suffer from extensive mutations (7 mutations in the 12 clones, **Supplementary Table 10**). However, return-to-growth clones may suffer from other types of genomic alterations with negative effects on spore viability, such as chromosomal rearrangements triggered by the 150-200 DSBs induced by Spo11p. Indeed, examples of gross chromosomal rearrangements generated by break-induced replication mediated by Ty1 have been described in *S. cerevisiae* return-to-growth clones (Laureau et al., 2016; Mozzachiodi et al., 2020). Such type of gross chromosomal rearrangements would have highly detrimental effects on spore viability.

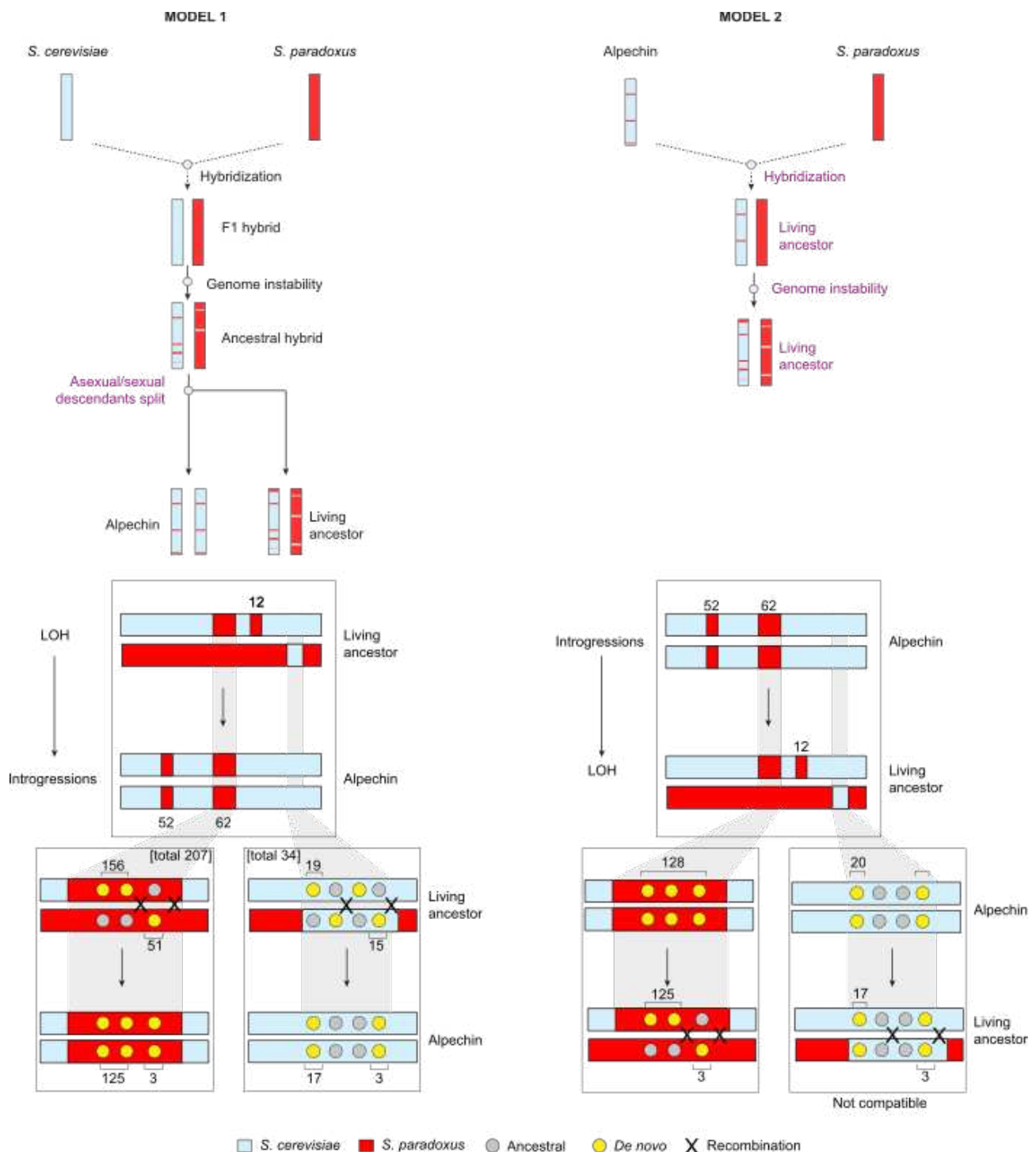
In contrast, gamete viability was equally likely to increase and decrease during the mutation accumulation line experiments, as recently reported (Charron et al., 2019). One line that underwent whole genome duplication recovered high fertility (43.5%, **Extended Data Fig. 6c, d and**

Supplementary Table 4), also as recently reported for a *S. cerevisiae*-*S. paradoxus* first-generation experimental hybrid (Charron et al., 2019). The tetraploidization was confirmed by FACS analysis and occurred between 40 and 60 single-cell bottlenecks of experimental evolution. This shows that tetraploidization is a powerful route to restore fertility (Greig, et al., 2002). Tetraploidization may have also played a role downstream the hybridization in the origin of Alpechins' introgressions, although it is still unknown whether tetraploids produce recombined diploid gametes. Two other mutation accumulation lines acquiring large LOH events in addition to those present in the living ancestor or going through chromosomal homogenization (loss and re-synthesis) increased their gamete viability to 6% (A887R33) and 10% (A887R31), respectively. Both lines acquired a large *de novo* LOH event on chromosome XIV (**Extended Data Fig. 6a**). In the living ancestor, chromosome XIV has only one single, small LOH block (**Extended Data Fig. 3**) and is therefore highly heterozygous and difficult to recombine, as it is evident from the genomes of meiotic gametes (**Extended Data Fig. 5a**). The heterozygosity of chromosome XIV likely serves as a barrier to gamete viability in the living ancestor, and the drastic increase in homozygosity due to the large, *de novo* LOH on this chromosome likely explains the increased fertility of the A887R31 and A887R33 lines. Acquiring additional LOH in chromosomes already rich in homozygous blocks appears to have no or a much lower beneficial effect on fertility, as expected given that these chromosomes already recombined more efficiently in the living ancestor. For example, novel LOH in chromosome VII (line A887R36) and chromosome IV (line A887R32) did not increase gamete viability. These results underscore that LOH have varying effects on fertility, depending on their size, location and distribution within and across chromosomes.

In contrast, spore viability in five samples did not change, or in some cases decreased, compared to the original living ancestor. This is fully expected considering that mutation accumulation lines evolved as diploid clones can accumulate recessive deleterious mutations within essential genes (19% of the genes, covering 13% of the genome), that are unmasked in haploid gametes. Consistent with this assumption, we detected 109 *de novo* mutations in the mutation accumulation lines, 23 of which affected essential genes (Supplementary Table 10). Each line carried at least one mutated essential gene. It is entirely plausible that the accumulation of detrimental *de novo* mutations counteracts the fertility benefits of additional LOH and explains the decreased viability in some mutation accumulation lines. We used the *de novo* mutations to estimate a base substitution rate of 2.31×10^{-10} /bp/gen, which is identical to what has been reported for experimental *S. cerevisiae*-*S. paradoxus* first-generation hybrids (Tattini et al., 2019).

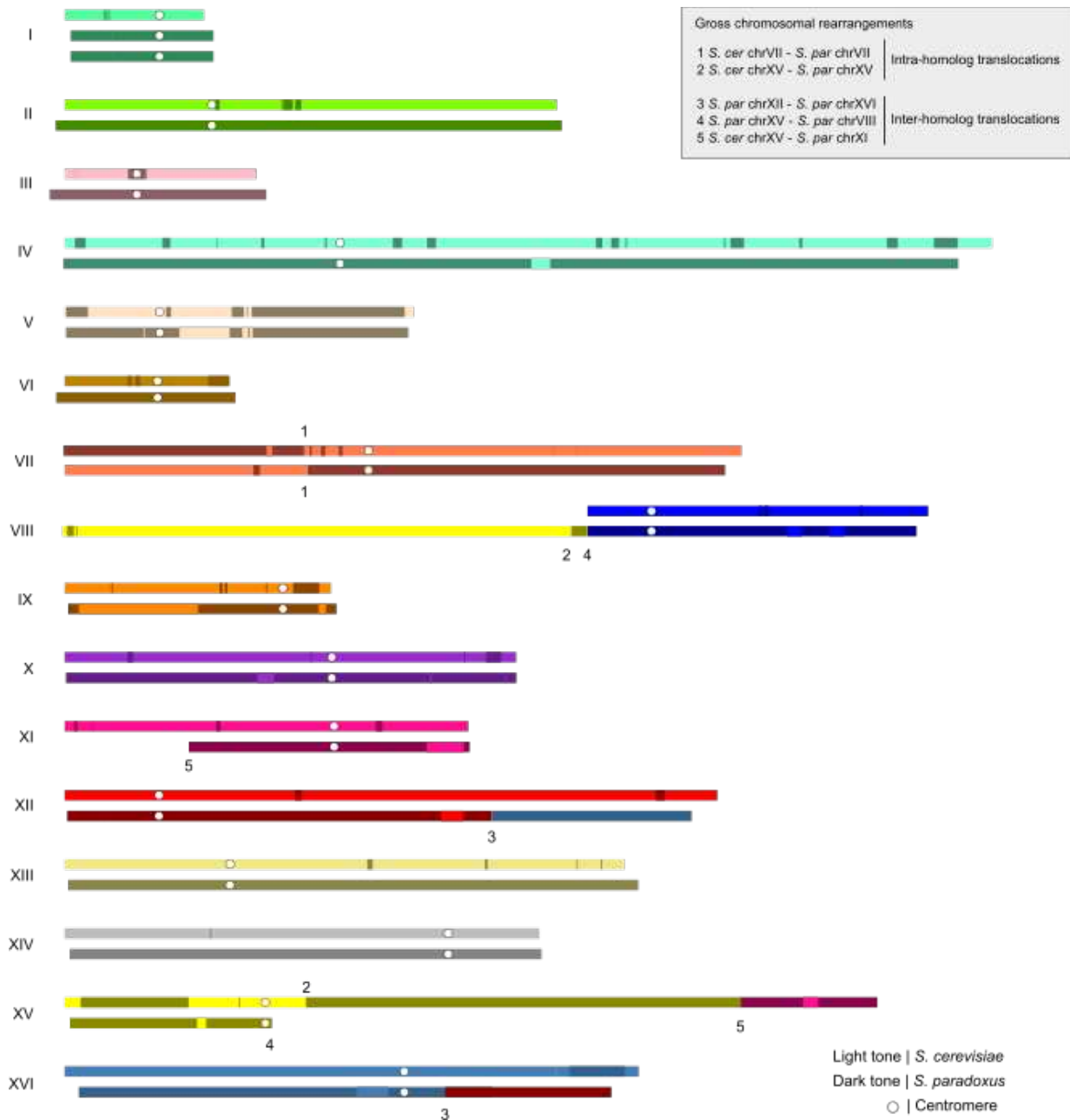


Extended Data Fig. 1 | Genome-wide distribution of LOH and introgressions. **a**, Genotype maps of the genome of the living ancestor (top) and the 17 Alpechins. The living ancestor's DNA is coloured to denote *S. cerevisiae* (blue) and *S. paradoxus* (red) DNA, respectively. The Alpechins' DNA is coloured to denote the *S. cerevisiae* background (blue) and the interspersed introgressions that are homozygous (red) or heterozygous (grey) for *S. paradoxus* DNA. There is extensive overlap between the positions of *S. paradoxus* LOH in the living ancestor and introgressions in the Alpechins. **b**, Zoom-in of chromosome II. Colours as in panel a. White circles indicate centromeres. Genomic coordinates are based on the *S. cerevisiae* DBVPG6765 genome assembly.

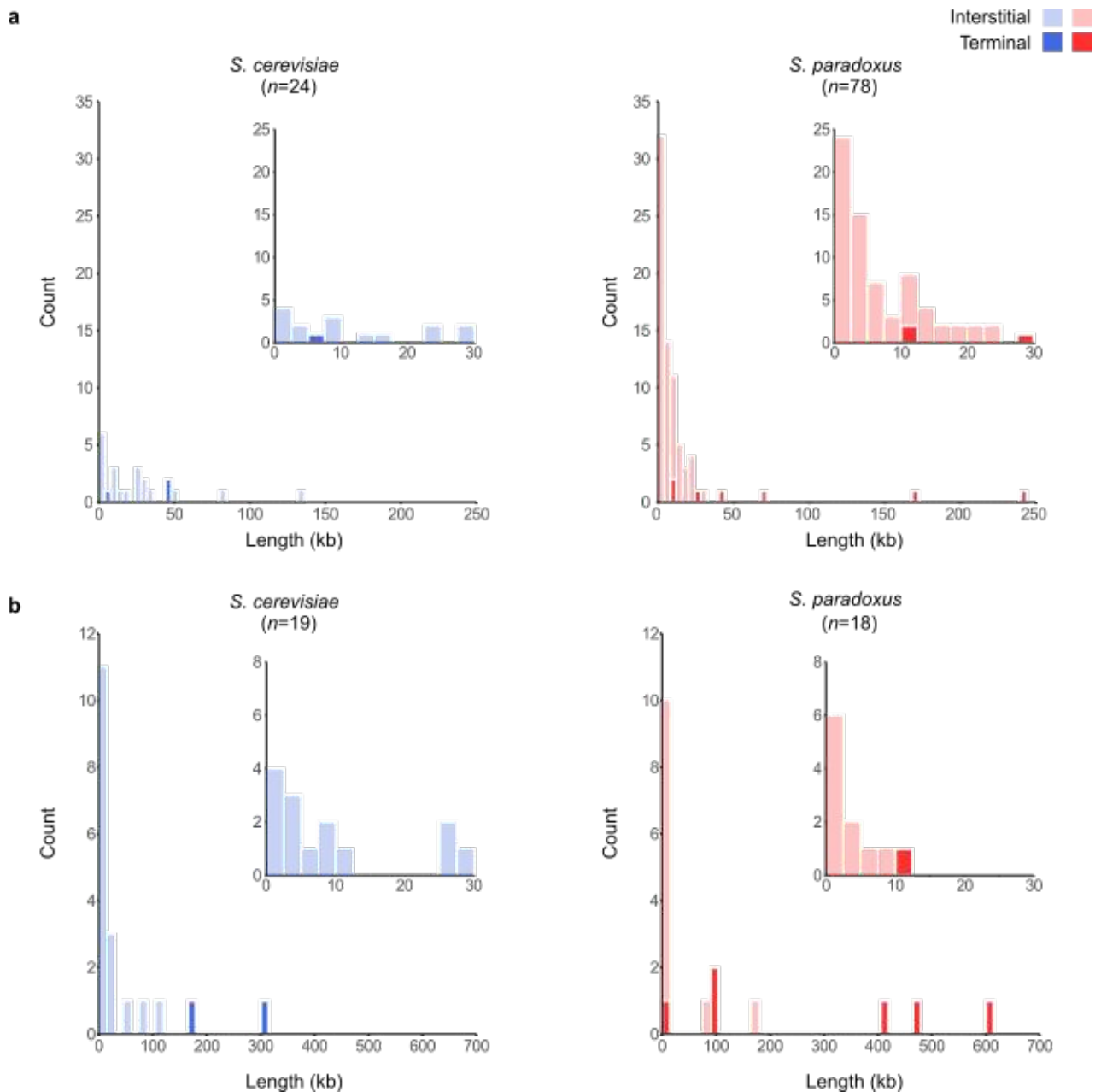


Extended Data Fig. 2 | Alternative evolutionary models of the living ancestor's and Alpechins' origin. The genome-wide overlap between LOH, introgressions and mutated sites within LOH regions with two copies of *S. cerevisiae* or *S. paradoxus* DNA allowed us to test the alternative models 1 (left) and 2 (right). See Supplementary Discussion 1 for details. Top boxes report the number of introgressions and LOH, partitioned into overlapping and non-overlapping events. Bottom boxes report the number of *de novo* point mutations in LOH blocks of the living ancestor. Left and right boxes show mutations in *S. paradoxus* and *S. cerevisiae* LOH blocks, respectively. Numbers around the top chromosome pair report the total number of *de novo* mutations, partitioned into those occurring on the *S. cerevisiae* and *S. paradoxus* haplotype, respectively. Numbers below

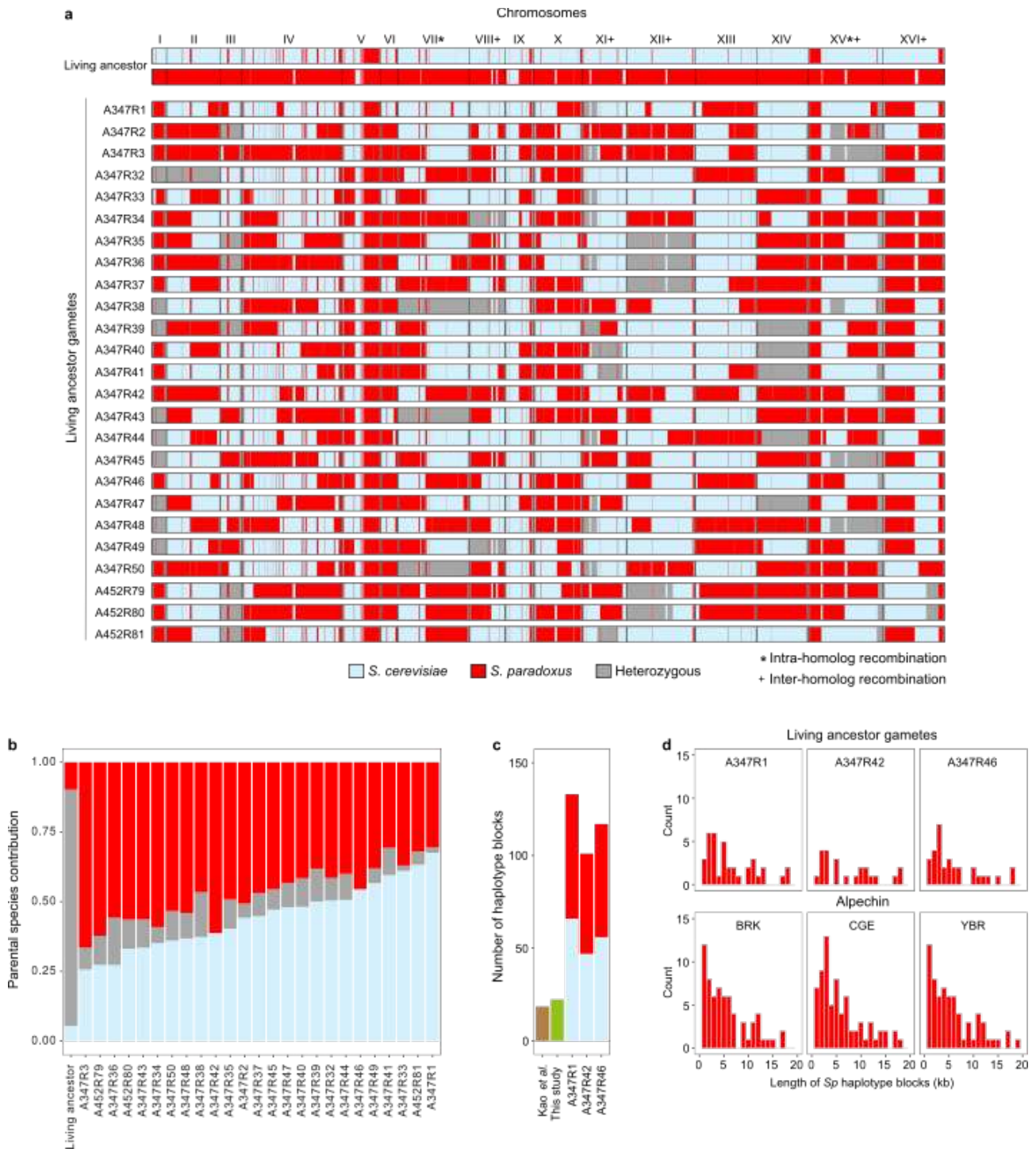
the bottom chromosome pair report *de novo* mutations that are shared between the living ancestor and the Alpechins, again partitioned into those occurring on the *S. cerevisiae* and *S. paradoxus* haplotype, respectively. Haplotype phasing data is available only for the living ancestor. Patterns of overlap between *S. paradoxus* LOH/introgressions and shared *de novo* mutations between the living ancestor's *S. paradoxus* LOH and the Alpechin strains are compatible with model 1. In contrast, the sequence of events in model 2 makes it incompatible with the observed data: the emergence of *S. cerevisiae* LOH and their *de novo* mutations after the hybridization between an Alpechin and a *S. paradoxus* implies that none of these *de novo* mutations should be present in any Alpechin strain. The presence of 20 mutations which are shared between the Alpechins and the living ancestor rules out model 2 in favour of model 1.



Extended Data Fig. 3 | Genome structure of the living ancestor. Each coloured bar represents a chromosome, with dark tones representing *S. paradoxus* DNA and light tones representing *S. cerevisiae* DNA. Dark blocks within *S. cerevisiae* chromosomes represent *S. paradoxus* LOH, while light blocks within *S. paradoxus* chromosomes represent *S. cerevisiae* LOH. White circles indicate centromeres and numbers (from 1 to 5) annotate the recombination breakpoints of the gross chromosomal rearrangements listed in the top-right box. Subgenome assignment for rearranged chromosomes is based on centromere flanking sequences. Genomic coordinates of *S. cerevisiae* and *S. paradoxus* chromosomes are based on the genome assemblies of DBVPG6765 and CBS432 strains, respectively. We cannot confidently say whether intrinsic (genetic) or extrinsic (environmentally induced) genome instability explain the astounding genome instability signature, nor we can say whether it emerged from a single event or accumulated gradually.

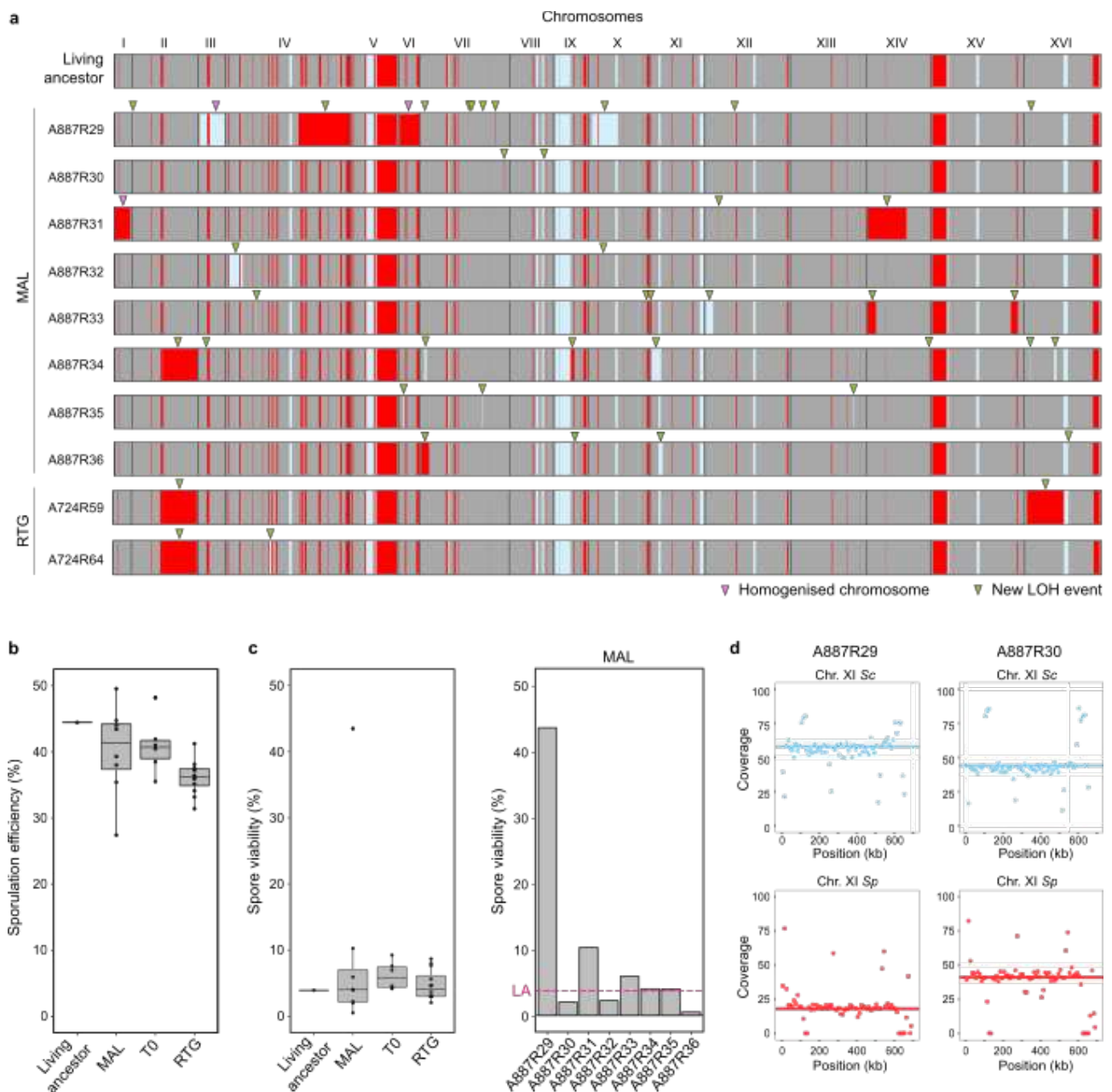


Extended Data Fig. 4 | LOH size distribution. a, Histograms of LOH sizes in the living ancestor. The left panel reports the size histogram of *S. cerevisiae* LOH events (two copies of *S. cerevisiae* DNA retained and loss of the *S. paradoxus* DNA copy). Bin width is 4 kb. Conversely, the right panel reports the size histogram of the *S. paradoxus* LOHs (two copies of *S. paradoxus* DNA retained and loss of the *S. cerevisiae* DNA copy). Dark and light colours respectively indicate terminal and interstitial LOH events. Insets show x-axis zoom-ins up to 30 kb. Bin width is 2.5 kb. **b,** Size histograms of new LOH events in the mutation accumulation lines. Bin width is 15 kb. Insets show x-axis zoom-ins up to 30 kb. Bin width is 2.5 kb. Panel order and colour codes are as in panel **a**.



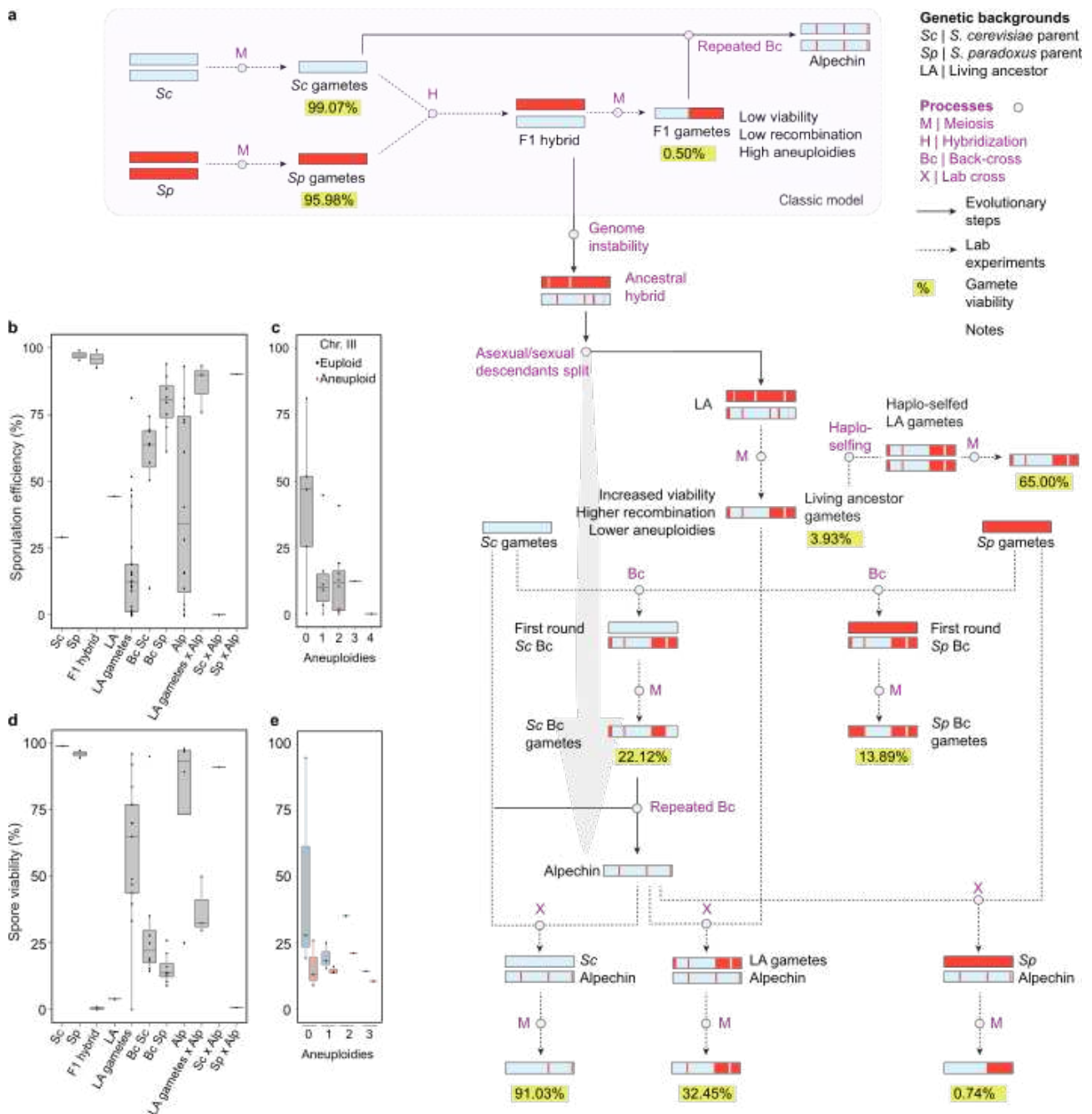
Extended Data Fig. 5 | Genome-wide genotype of the living ancestor gametes. **a**, We used the genome sequences of 25 living ancestor haploid gametes to infer their recombination landscape. Red and blue colours represent *S. paradoxus* and *S. cerevisiae* DNA, respectively. The upper part shows the living ancestor. Grey segments represent heterozygous regions for which both *S. paradoxus* and *S. cerevisiae* DNA is present and correspond to duplicated regions derived from segregation of rearranged chromosomes or aneuploidies. Aneuploidies were all gain-of-copies, with chromosome III particularly prone to duplicate (7/25 gametes), consistent with the presence of a single LOH block encompassing the centromere, where crossovers are rare (Pan et al., 2011). Chromosomes with intra-homolog and inter-homolog gross chromosomal rearrangements are

indicated with * and +, respectively. Genomic coordinates are based on the *S. cerevisiae* DBVPG6765 genome assembly. **b**, Percentage of the genome bearing *S. cerevisiae* or *S. paradoxus* DNA in the 25 living ancestor gametes. Colours as in panel **a**. **c**, The number of observed *S. cerevisiae* (blue) and *S. paradoxus* (red) haplotype blocks in three living ancestor gametes. These blocks derive from the combination of meiotic recombination events and pre-existing LOH. As comparison, we calculated the theoretical number of haplotype blocks expected from a single meiosis, considering only the number of crossovers, in an experimental *S. cerevisiae*-*S. paradoxus* F1 hybrid (data from Kao *et al.* (Kao et al., 2010), brown bar), and in the living ancestor if LOH were not present (green bar). We estimated the number of expected haplotypes (y) as $y=16(x+1)$, where x represents the number of crossovers per chromosome. **d**, Size distribution of *S. paradoxus* haplotype blocks in three living ancestor gametes (upper panels) and of introgression blocks in three Alpechins (lower panels).



Extended Data Fig. 6 | Living ancestor clonal evolution. **a**, Genome-wide genotype maps of the living ancestor and its diploid clones derived by mutation accumulation line protocol (MAL) through 120 single-cell bottlenecks and return-to-growth (RTG) clones derived by inducing and then aborting meiosis. Grey represents heterozygous DNA, red represents LOH blocks with two *S. paradoxus* DNA copies, while blue represents LOH blocks with two *S. cerevisiae* DNA copies. Green arrows indicate new LOH recombination events, purple arrows indicate chromosome homogenisation by loss and re-synthesis events. Genomic coordinates are based on the *S. cerevisiae* DBVPG6765 genome assembly. **b**, Sporulation efficiency (at day 3) and **c**, spore viability of the living ancestor, its mutation accumulation lines (MAL, $n=8$), its mitotically growing (T0, $n=6$) and return-to-growth (RTG, $n=12$) clones. Boxes: horizontal line: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR. The bar plot (right) shows spore viability of each single living ancestor's mutation accumulation line. One line (A887R29) has

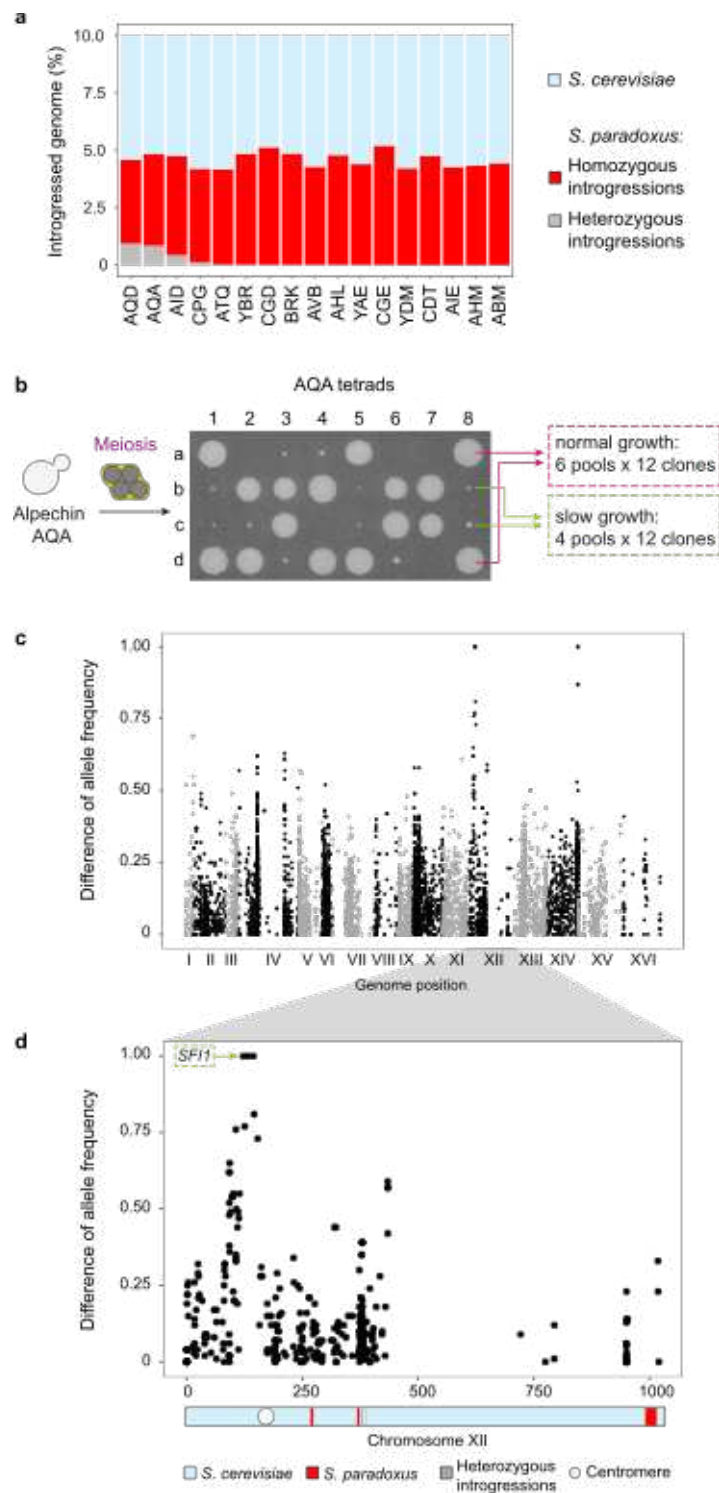
a huge increase in spore viability due to tetraploidization. The pink dotted line indicates the spore viability of the living ancestor. **d**, Coverage plots of chromosome XI of a tetraploid (A887R29) and a diploid (A887R30) mutation accumulation lines. The plots show a balanced chromosome number for the A887R30 diploid clone (1 copy of *S. cerevisiae* and 1 copy of *S. paradoxus*) and an unbalanced one for the A887R29 tetraploid clone (3 copies of *S. cerevisiae* and 1 copy of *S. paradoxus*). Dots represent mean coverage values in 10 kb non-overlapping windows. Lines represent whole-chromosome coverage medians.



Extended Data Fig. 7 | Reconstruction of the hybridization-to-introgression model. **a**, The grey-dotted box contains a representation of the classic model for the origin of introgressions. The discovery of the living ancestor rewrote part of this mechanistic process and enabled us to test the downstream stages. For each stage of the model, we indicate the measured median gamete viability (%). Plain lines point to naturally occurring stages, while dotted lines point to stages derived in the laboratory. The initial *S. cerevisiae*-*S. paradoxus* F1 hybrid derived from an outbreeding event and this has been shown to be rare (Ruderfer et al., 2006; Tsai et al., 2008), thus one or very few hybridizations initiated the process. Next, the F1 hybrid expanded clonally to a possibly large population size, depending on nutrients availability in its ecological niche. This mitotic clonal expansion was followed by the genome instability and LOH formation phase. It is hard to predict the rate of genome instability, since it might be influenced by both strain background and the

specific olive-oil related environment. However, genome instability of such magnitude might be rare and produce unviable combinations, likely reducing the ancestral hybrid population with LOH to one or few individuals. Next, the hybrid with LOH expanded mitotically, generating a population that kept evolving e.g. by accumulating additional LOH. Part of the population was maintained as asexual descendant, while part went through sexual reproduction and produced the Alpechin lineage. It is likely that the sporulation step was at the population level, therefore producing many spores. Next, one or more spores backcrossed to *S. cerevisiae*, which might occur rarely since it requires outbreeding. However, backcross of gametes from the ancestral hybrid was favoured by low gamete viability, which prevents intra-tetrad mating (that is the main way for spores to re-establish diploidy), while haplo-selfing was inhibited by fitness costs. We propose that this step bottlenecked the population to few individuals, consistent with our data showing that Alpechin strains share many of the LOH blocks. The living ancestor gametes were more fertile with *S. cerevisiae* than with *S. paradoxus*. We also crossed three living ancestor gametes to a haploid derivative of an Alpechin strain (CPG) and found fertility to be higher than in backcrosses with the most closely related *S. cerevisiae* lineage lacking introgressions. This is consistent with the Alpechin lineage being closer to the living ancestor. **b**, Sporulation efficiency after 3 days in sporulating conditions (central bar: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR) of hybridization-to-introgression stages. n =biologically independent samples. *S. cerevisiae* (*Sc*, $n=1$), *S. paradoxus* (*Sp*, $n=2$), experimental first-generation (F1) hybrid (F1 hybrid, $n=2$), living ancestor (LA, $n=1$), living ancestor gametes (LA gametes, $n=25$), backcross to *S. cerevisiae* (Bc *Sc*, $n=8$), backcross to *S. paradoxus* (Bc *Sp*, $n=8$), Alpechin (Alp, $n=16$), living ancestor gametes x Alpechin (LA gametes x Alp, $n=3$), *S. cerevisiae* x Alpechin (*Sc* x Alp, $n=1$), *S. paradoxus* x Alpechin (*Sp* x Alp, $n=1$). The *S. cerevisiae* x Alpechin cross (*Sc* x Alp) is a slow sporulator and requires additional time to sporulate, consistent with the presence of inefficient sporulation alleles in the Wine/European *S. cerevisiae* background (Gerke et al, 2017). **c**, Sporulation efficiency (at day 3) of living ancestor gametes grouped by number of aneuploidies (central bar: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR). n =biologically independent samples. 0 aneuploidies ($n=5$), 1 aneuploidy ($n=6$), 2 aneuploidies ($n=12$), 3 aneuploidies ($n=1$), 4 aneuploidies ($n=1$). Pink dots indicate spores with an additional copy of chromosome III, which remained haploid and thus do not sporulate. **d**, Spore viability (central bar: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR) of hybridization-to-introgression stages. n =biologically independent samples. *S. cerevisiae* (*Sc*, $n=1$), *S. paradoxus* (*Sp*, $n=2$), experimental first-generation (F1) hybrid (F1 hybrid, $n=2$), living ancestor (LA, $n=1$), living ancestor gametes (LA gametes, $n=13$), backcross to *S. cerevisiae* (Bc *Sc*, $n=8$), backcross to *S.*

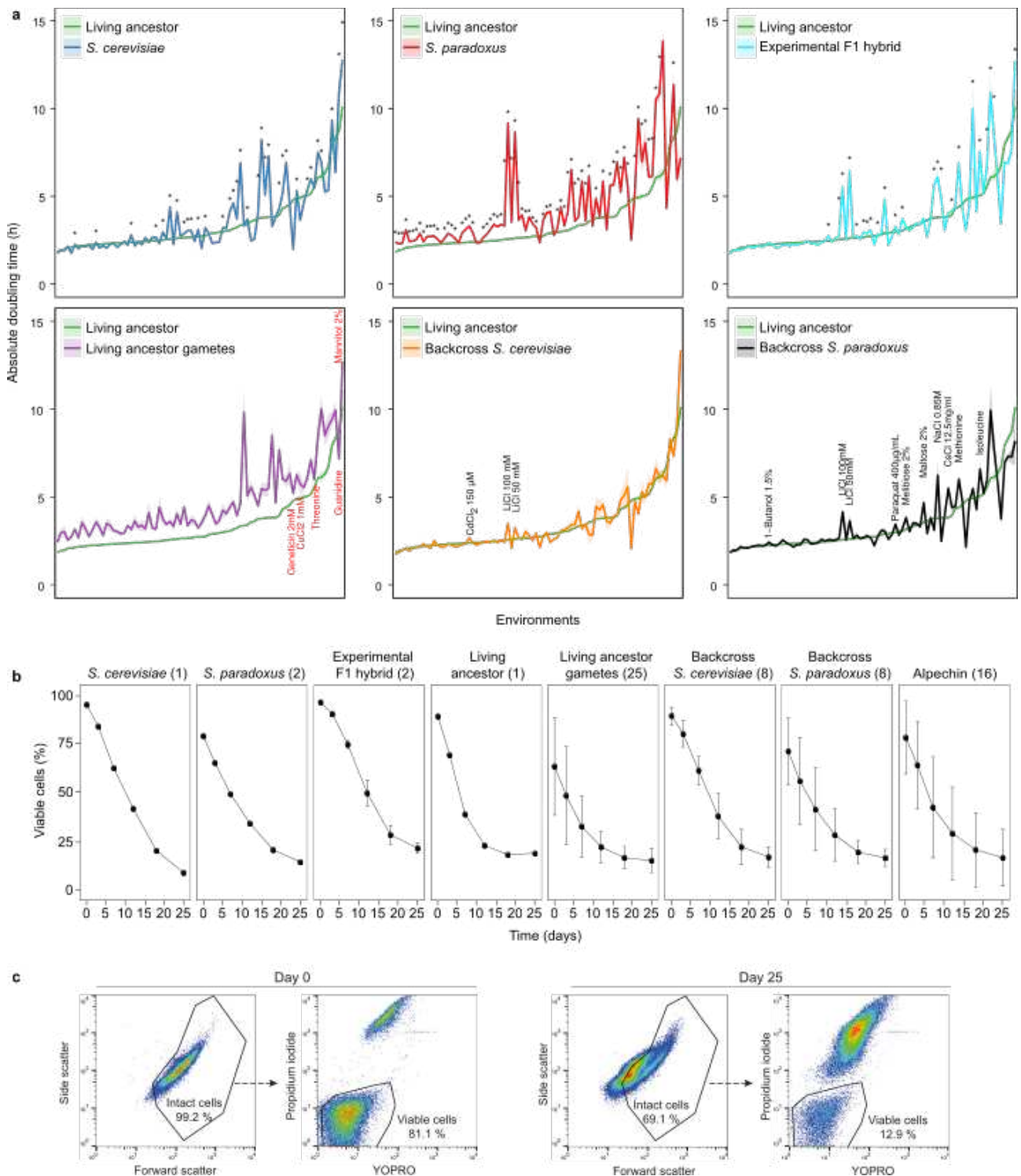
paradoxus (Bc *Sp*, $n=8$), Alpechin (Alp, $n=4$), living ancestor gametes x Alpechin (LA gametes x Alp, $n=3$), *S. cerevisiae* x Alpechin (*Sc* x Alp, $n=1$), *S. paradoxus* x Alpechin (*Sp* x Alp, $n=1$). The living ancestor has higher spore viability than the experimental *S. cerevisiae*-*S. paradoxus* F1 hybrid. The spore viability of the backcrosses to *S. cerevisiae* compares favourably to that of the *S. paradoxus* backcrosses. The Alpechins are highly fertile with a Wine/European *S. cerevisiae* (*Sc*) background without introgressions and moderately fertile with the living ancestor gametes, whereas they have strong reproductive barriers with *S. paradoxus* (*Sp*). **e**, Percentage of viable spores for haploid-stable living ancestor gametes backcrossed to *S. cerevisiae* (blue line boxes) or *S. paradoxus* (red line boxes), grouped by their number of aneuploidies. Boxes: horizontal line: median, upper/lower hinge: interquartile range (IQR), whiskers: max/min values. n =biologically independent samples. 0 aneuploidies ($n=6$), 1 aneuploidy ($n=6$), 2 aneuploidies ($n=2$), 3 aneuploidies ($n=2$).



Extended Data Fig. 8 | Mapping the slow growth phenotype in the Alpechin AQA strain. a, Percentage of introgressed DNA in the Alpechin strains ranked for amount of heterozygous introgressions (grey, ordered first in the stack bar). Most introgressed material is in homozygous state (red, ordered second). Three strains, including AQA, have a considerable fraction of heterozygous introgressions. The rest of the genome up to 100% is *S. cerevisiae* background (blue, ordered third) and the y-axis is cut at 10%. **b,** Tetrad dissection of the Alpechin AQA strain showing a 2:2 colony size segregation pattern (two slow-growing and two normal-growing colonies in each meiosis), consistent with a single heterozygous variant underlying the growth defect. **c,** Each dot

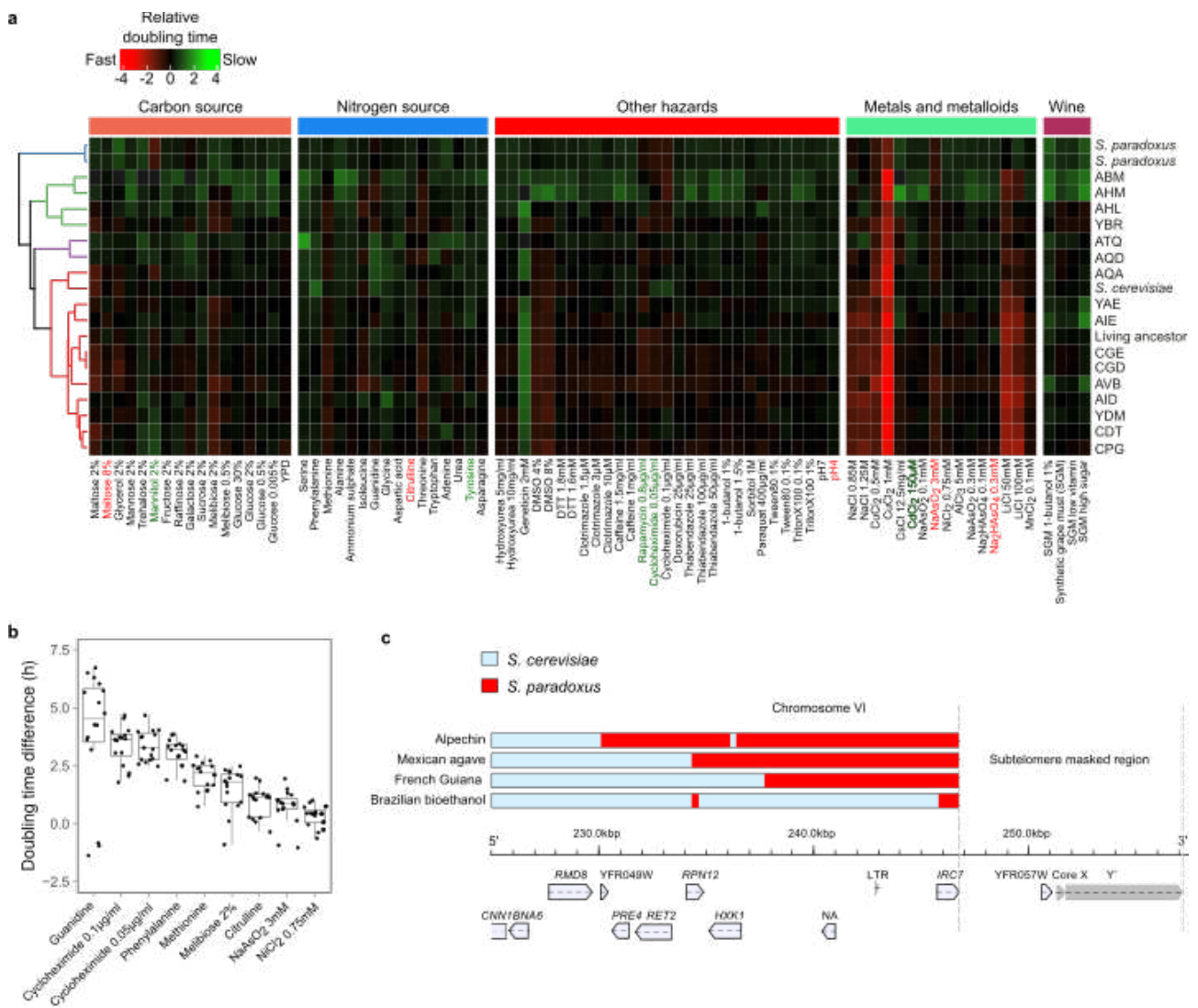
represents the allele frequency (AF) difference of reference and non-reference alleles across the genome between pools of big and small spores. Two regions show fixation of the opposite alleles in the big and small pools and chromosome XII is compatible with the assumption that the deleterious allele associated with the small phenotype has the lowest frequency in the *S. cerevisiae* population.

d, Chromosome XII AF zoom and map of homozygous (red) and heterozygous (grey) introgressions on a *S. cerevisiae* genome backbone (blue) of the Alpechin strain AQA. The growth defect maps to the *SFI1* gene, which is far from the closest introgression, indicating that the slow growth phenotype is not driven by a deleterious effect of heterozygous introgressions masked by the *S. cerevisiae* allele. The white circle indicates the centromere. Genomic coordinates are based on the *S. cerevisiae* DBVPG6765 genome assembly.



Extended Data Fig. 9 | The competitive fitness of the living ancestor. **a**, Mitotic growth (absolute doubling time, h) of the living ancestor ($n=40$) as compared to the closest extant relatives of its *S. cerevisiae* DBVPG6765 (upper left, $n=8$) and *S. paradoxus* N17 (upper middle, $n=16$) parents and a diploid experimental F1 hybrid between the two (upper right, $n=16$), the living ancestor's diploid gametes (lower left, $n=200$) and backcrosses of its haploid-stable gametes to *S. cerevisiae* (lower middle, $n=64$) and *S. paradoxus* parents (lower right, $n=64$). The number of observations (n) results from the combination of the strains in each group and the number of replicates performed in the experiment. Significant differences enumerated in the main text are indicated with asterisks or name

(Welsh test, FDR $q=0.05$) and presented in Supplementary Table 7. For the living ancestor gametes, we instead indicate (red text) the rare environments where no significant difference was found. **b**, The percentage of surviving cells (average \pm SD) for each step of the hybridization-to-introgression process during a 25-days chronological lifespan experiment. n =biologically independent samples. *S. cerevisiae* ($n=1$), *S. paradoxus* ($n=2$), experimental first-generation (F1) hybrid ($n=2$), living ancestor ($n=1$), living ancestor gametes ($n=25$), backcross to *S. cerevisiae* ($n=8$), backcross to *S. paradoxus* ($n=8$), Alpechin ($n=16$). **c**, Gating strategy used for the estimation of the percentage of surviving cells shown in panel **b**.



Extended Data Fig. 10 | The adaptive value of introgressions. **a**, Heatmap of the relative doubling time (mean of $n=40$ replicates for the living ancestor and $n=8$ replicates for all the other strains) of the living ancestor, its modern Alpechin descendants and the closest living relatives of its *S. cerevisiae* and *S. paradoxus* parents. Strains were grouped (left) based on doubling time similarity using hierarchical clustering (Pearson's r , complete linkage) and branch colours highlight four major clusters. Environments in which the living ancestor grows significantly (Welsh test, FDR $q=0.05$) faster (green) or slower (red) than the Alpechins are coloured. **b**, Difference in relative doubling time (central bar: median, upper/lower hinge: IQR, whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR) between Alpechins ($n=128$) and *S. cerevisiae* ($n=8$) in the environments which favour Alpechins' growth (Welsh test, FDR $q=0.05$). The number of observations (n) results from the combination of the strains in each group and the number of replicates performed in the experiment. Each dot represents the average among the 8 replicates of each of the 16 Alpechins. **c**, Genotype maps of the Alpechin, Mexican agave, French Guiana and Brazilian bioethanol clades in a shared introgressed region containing the *IRC7* (*YFR055W*) gene. This gene was previously reported to be introgressed in strains belonging to the Wine/European

clade (Roncoroni et al., 2011). The *IRC7* gene present in the Alpechins has a European *S. paradoxus* ancestry, while Mexican Agave, Brazilian Bioethanol and French Guiana lineages inherited the *S. paradoxus* American allele, underlying an independent introgression and maintenance of *IRC7* that supports adaptive introgression. Blue and red colours indicate *S. cerevisiae* and *S. paradoxus* regions, respectively. Subtelomeric regions are masked genome wide because they lack synteny between *S. cerevisiae* and *S. paradoxus* and cannot be aligned for markers genotyping, although inspection of this specific VI-R subtelomeric region in both *S. cerevisiae* and *S. paradoxus* genomes indicates that the telomeric repeats are very close to *IRC7*²⁴. The bottom panel shows genetic features annotation of this region from the *S. cerevisiae* Wine/European strain DBVPG6765.

Methods

Natural and experimental yeast strains

The strains used in this work are listed in supplementary table 1. The living ancestor, CBS7002, was isolated from olive oil wastewater (Alpechin, in Spanish) by J. Santa Maria in the 1970s and obtained from the CBS yeast culture collection (<http://www.wi.knaw.nl/Collections/>). This strain was initially proposed as a type strain of the new species *S. hispalensis*, which is now regarded as synonymous to *S. cerevisiae*. The Alpechins were also mostly isolated from the same substrate and the same geographic location as the living ancestor. However, two Alpechins were isolated from Californian olive fermentations inoculated with European olives, one was isolated from the olive fruit fly and four were isolated from humans (as fungal infections or from faeces). Genome sequences of the Alpechins used in this study were previously reported (Peter et al., 2018), whereas new samples associated with this work were sequenced with Illumina paired-end technology at the NGS platform of Institut Curie according to the manufacturer's standard protocols.

We reconstructed the missing evolutionary stages linking the living ancestor and the Alpechins and leading to genome introgressions. First, we generated homozygous diploid *S. cerevisiae* (DBVPG6765, Wine/European subpopulation) and *S. paradoxus* (N17, European subpopulation) strains. Second, we mated a haploid *S. cerevisiae* DBVPG6765 and a haploid *S. paradoxus* N17 to generate an experimental *S. cerevisiae*-*S. paradoxus* F1 hybrid with subgenomes closely resembling those of the ancestral hybrid before the genome shock. Next, we induced meiosis in the living ancestor by a five-days incubation in potassium acetate (KAc) medium (Liti et al., 2017) and isolated 25 viable meiotic spores by tetrad dissection. The living ancestor is homothallic and therefore able to switch mating type during vegetative growth. 18 of these spores were auto-diploidized by mating-type switching and haplo-selfing to form homozygotic diploids, while 7 spores remained haploid due to the presence of chromosome III aneuploidies. The ploidy was

further confirmed by flow cytometry analysis, as previously described (Peter et al., 2018). Chromosome III contains the *MAT* locus, which controls the mating type. When both *MAT* alleles are present due to aneuploidies, spores are unable to switch mating type and haplo-selfing is inhibited (Haber, 2012).

We deleted the *HO* mating-type switch gene in the living ancestor and generated eight additional haploid-stable gametes (*ho::NAT*, either *ura3::KanMX* or *ura3::HygMX*). These spores were backcrossed to *S. paradoxus* N17 and to *S. cerevisiae* DBVPG6765 respectively to produce ($n=16$) diploid hybrids that were either $\sim 75\%$ *S. paradoxus* or $\sim 75\%$ *S. cerevisiae*.

We deleted the *HO* gene also in one diploid Alpechin (CPG) and generated a haploid-stable gamete (*MATa*, *ho::HygMX*, *ura3::KanMX*). This gamete was crossed with *S. paradoxus* N17, *S. cerevisiae* DBVPG6765 and with three haploid-stable living ancestor gametes.

Mutation accumulation lines

We evolved the living ancestor as 8 parallel mitotic mutation accumulation lines on YPD solid medium and passed them through a single-cell bottleneck every ~ 48 hours (~ 18.5 generations) at 30°C , for a total of 120 bottlenecks (~ 2220 generations). At each bottleneck, a random colony, having expanded from a single cell, was re-streaked to isolate single colony-forming units. To avoid selection bias, we picked the closest colony to the centre of each cultivation plate at each streak, regardless of its size. The ploidy of the evolved lines was measured by flow cytometry, as previously described (Peter et al., 2018).

Rates of *de novo* LOH formation and return to growth (RTG) experiments

Rates of *de novo* LOH formation were measured for the living ancestor and the experimental *S. cerevisiae*-*S. paradoxus* first generation hybrid using a fluctuation assay, as previously described (Vázquez-García et al., 2017). In diploid strains, both *URA3* gene copies were deleted from their native locus on chromosome V and one *URA3* copy was inserted back into the *LYS2* locus on chromosome II, thereby replacing one of the *LYS2* copies. Relative rates of *de novo* LOH formation were measured on 5-FOA plates as loss of this single *URA3* gene copy, reflecting the replacement of the chromosome region carrying the *URA3* allele with the homologous region from the other chromosome II copy, which carried the *LYS2* allele. These rates were measured in populations of both mitotically-growing cells (18 hours in YPD) as well as cells in which meiosis had been initiated (6 hours in KAc) and then aborted by returning the cells to mitotic growth (18 hours in YPD), as previously described (Mozzachiodi et al., 2020; Tattini et al., 2019). We performed 5 biological replicates (1 technical replicate for each) for each strain background and condition. We sequenced the genome of 6 mitotically growing clones and 12 return-to-growth clones derived from

the living ancestor cells that had been selected to have lost the *URA3* allele. We confirmed the presence of a large *de novo* LOH on chromosome II in all analysed samples and checked the presence of additional *de novo* LOH events, as described in the paragraph “Mapping LOH and introgressions’ boundaries”.

Estimation of sporulation efficiency and spore viability

We measured sporulation efficiency and spore viability as previously described (De Chiara et al., 2020), for all the stages of the hybridization-to-introgression process as well as for the genome-sequenced clones of the living ancestor derived by asexual growth (6 clones), return-to-growth (12 clones) and mutation accumulation protocol (8 clones) (**Supplementary Tables 4 and 5**). Sporulation was induced by incubation in liquid KAc medium and its efficiency was measured by counting ~200 cells at days 1 and 3 for the *S. cerevisiae* and *S. paradoxus* parents, the experimental F1 hybrid, the backcrosses and the crosses of the Alpechin strain with *S. cerevisiae* and *S. paradoxus*. All the other samples were measured at multiple time points ($t=0, 1, 2, 3, 4, 9$ days). Spores were then isolated by tetrad dissection and their viability was scored as the percentage of spores forming visible colonies after six days of growth on YPD at 30°C.

Bulk segregant analysis of Alpechin AQA growth defect

We induced sporulation of the Alpechin strain AQA (OS1357) in liquid KAc medium. Spores were isolated using a dissecting microscope and incubated in YPD medium for four days at 30°C. We pooled spores based on their colony size, forming ten pools. Six pools contained only big colonies while four contained only small colonies. Each pool was composed of 12 spores, except one (A504R10) that contained 11 spores. DNA was extracted from each pool and sequenced with Illumina technology (mean coverage of 55X). The markers used in the bulk segregant analysis were extracted from the Alpechin strain AQA by mapping onto the *S. cerevisiae* DBVPG6765 assembly, using BWA (v0.7.12). Post-processing steps, including sorting and duplicates removal were performed by SAMtools (v1.2) and picard tools (v2.8.0). Variant calling was performed by freebayes (v0.9.5). We extracted the markers assuming that the 2:2 segregation pattern in the spores was caused by a heterozygous SNV for which one of the two alleles was deleterious and recessive, slowing the growth of haploid spores but not that of the Alpechin diploid parent. Therefore, we extracted only the SNVs whose allelic fraction (AF) was equal to 0.5 in the Alpechin AQA parent. Afterwards, we mapped Illumina reads of the pools on the *S. cerevisiae* DBVPG6765 assembly and performed the subsequent steps as described above. Then, we estimated the AF of the heterozygous markers in all the pools and excluded the ones which were not compatible with the following assumptions: i) All the pools must have a homozygous genotype for the marker. ii) Big and small

pools must have a different allele for the marker. iii) The deleterious allele associated with the small phenotype must be the one with the lowest frequency in the *S. cerevisiae* population. Only one marker was compatible with these criteria and was therefore deemed to explain the variation in the spore's growth.

Estimating population size doubling time during clonal reproduction

We measured population size doubling time during clonal reproduction for all the hybridization-to-introgression stages. One Alpechin strain (BRK) was excluded from the analysis because of contamination. All yeast populations were stored at $-80\text{ }^{\circ}\text{C}$ in 20% glycerol and were cultivated at $30\text{ }^{\circ}\text{C}$ in temperature and humidity-controlled cabinets. Yeast strains were revived from frozen 96-well stocks by robotic transfer (Singer RoToR; long pins) of a random sub-sample of each thawed population ($\sim 50,000$ cells) to a Singer PlusPlate in 1536 array format on solid Synthetic Complete (SDC) medium composed of 0.14% Yeast Nitrogen Base (CYN2210, ForMedium), 0.50% $(\text{NH}_4)_2\text{SO}_4$, 0.077% Complete Supplement Mixture (CSM; DCS0019, ForMedium), 2.0% (w/w) glucose, pH buffered to 5.80 with 1.0% (w/v) succinic acid and 0.6% (w/v) NaOH. In every fourth position, fixed spatial controls (genotype: YPS128, *MATa/MAT α*) were introduced to account for spatial variation across plates in the subsequent experimental stage. Controls were similarly sub-sampled from a separate 96-well plate and introduced to the pre-culture array (Singer RoTor; long pins). Populations were pre-cultivated for 72 hours at $30\text{ }^{\circ}\text{C}$. For pre-cultures of nitrogen-limited environments, the background medium was modified to avoid nitrogen storing and later growth on stored nitrogen: CSM was replaced by 20 mg/L uracil (not converted into usable nitrogen metabolites) and $(\text{NH}_4)_2\text{SO}_4$ was reduced to growth-limiting concentrations (30 mg N/L). We cast all solid plates 24 hours prior to use, on a levelled surface, by adding 50 mL of medium in the same upper right corner of the plate. We removed excess liquid by drying plates in a laminar air-flow in a sterile environment. Pre-cultured populations were mitotically expanded until stationary phase (2 million cells; 72 h), were again subsampled ($\sim 50,000$ cells; short pins) and transferred to experimental plates, containing the medium of interest (**Supplementary Table 7**). Synthetic grape must (SGM) was prepared as previously described (Beltran et al., 2004). We tracked population size expansion using the Scan-o-matic system (v1.5.7, <https://github.com/Scan-o-Matic/scanomatic>). Plates were maintained undisturbed and without lids for the duration of the experiment (72 h) in high-quality desktop scanners (Epson Perfection V800 PHOTO scanners, Epson Corporation, UK) standing inside dark, humid and thermostatic cabinets with intense air circulation. Images were analyzed and phenotypes were extracted and normalized against the fourth position controls using Scan-o-Matic (Zackrisson et al., 2016). We extracted the normalized, relative population size doubling time for each experiment, D_r , by subtracting the control value for that position, $D_r =$

$\log_2(D) - \log_2(D_{control,local})$. D_r is reported as the relative doubling time. We also report absolute doubling times, while maintaining the normalization, as $D_{normabs} = 2^{D^{norm}} \text{Mean}(D_{control,local})$. We performed 8 biological replicates per sample, except for the living ancestor for which we performed 40 biological replicates.

Estimation of chronological life span

Strains were grown overnight in liquid SDC medium, diluted 100x in 200 μL of fresh SDC in a 96-well plate and incubated at 30 °C for 25 days. We started taking samples after the net growth in each cell population had ceased, 72 hours after incubation. At each time point, 5 μL of cells were transferred in 100 μL of staining solution (Phosphate-buffered saline + 3 μM propidium iodide + 200 nM YO-PRO-1) in a 96-well plate and incubated 10 minutes in the dark at 30 °C. Cell viability was measured by high-throughput flow cytometry on a FACS-Calibur using the HTS module, as previously described (Barré et al., 2020). Cell debris and heavily damaged old necrotic cells were first removed based on forward scatter and side scatter at day 0. Then, cells were excited with the 488 nm laser and fluorescence was read with FL-1 and FL-3 filters, corresponding to YO-PRO-1 and propidium iodide fluorescence, respectively. Non-fluorescent cells were considered viable whereas fluorescent ones were considered dead. The final viability was calculated as the product of "Intact cells" and "Viable cells" (**Extended Data Fig. 9c**). We measured each sample once. Flow cytometry data were subsequently analysed using FlowJo (v10.3).

Sequencing and *de novo* assembly of the living ancestor's genome

The complex genome structure of the living ancestor makes full assembly challenging. Specifically, large blocks of LOH with near-identical sequence are interspersed in the highly diverged parental subgenomes and they are often merged by the assembler. We combined multiple sequencing and experimental approaches to overcome this challenge.

First, we deep-sequenced the living ancestor with both short (~200X, Illumina) and long (~146X, PacBio) read technology. DNA was extracted and sequenced for 2 x 100 (paired-end) cycles using the Illumina HiSeq2000. *De novo* assemblies from short reads were obtained using ABySS (v1.9.0) (Simpson et al., 2009). DNA was extracted and sequenced by PacBio technology at the Lausanne genomic technologies facility (Lausanne, Switzerland) and raw reads were processed by standard SMRT Analysis pipeline (v2.3.0) to obtain filtered subreads. The assembly was processed with the LRSDAY (v1.0.0) pipeline (Yue & Liti, 2018). We performed *de novo* assembly using Canu (v1.5) (Koren et al., 2017) by setting the overall (diploid) genome size to 24 Mb. We then aligned the raw PacBio reads to our Canu assembly by palign (distributed with GenomicConsensus v2.0.0) and polished the assembly by Quiver (distributed with GenomicConsensus v2.0.0) (Chin et al., 2013),

correcting sequencing errors. We further mapped the Illumina reads to the polished genome assembly for another round of error-correction, using Pilon (v1.22) (Walker et al., 2014). Scaffolds were manually visualised and inspected as dot-plots using the program MUMMERplot included in the suite MUMmer (v3.1) (Kurtz et al., 2004).

We additionally used the genome sequence of 25 living ancestor gametes to phase the living ancestor's subgenomes and to determine its karyotype structure. We analysed the genomes of the living ancestor's gamete and used the coverage to infer their segregation and recombination patterns. Most (65%, excluding aneuploidies, chromosome VII and chromosome XV) of the chromosomes had non-recombinant and uniform coverage profiles, that reflected the living ancestor's genome configuration. Several structural variants were evident using both the PacBio assembly and the segregation profiles and these were validated through manual inspection of the corresponding scaffolds.

Phylogenetic analysis of the living ancestor's subgenomes

We performed *de novo* genome assembly based on deep (~200X) Illumina sequencing data for the living ancestor and 32 representative strains sampled from the 1002 Yeast Genome Project⁵. The genome assembly was performed by ABySS (v1.9.0) (Simpson et al., 2009) with the option '-k 64'. All assembled scaffolds shorter than 1 kb were removed. The 33 assemblies were subsequently annotated using LRSDAY (Yue et al., 2017; Yue & Liti, 2018). For the living ancestor, we further partitioned its annotated genes into two subsets: the *S. cerevisiae*-like genes and the *S. paradoxus*-like genes, based on the tblastx (v2.2.31+) search against our curated yeast proteome database containing genes from 7 *S. cerevisiae* and 5 *S. paradoxus* representative strains (Yue et al., 2017). Each gene was classified as *S. cerevisiae*-like or *S. paradoxus*-like if its top hit (cutoff: sequence identity \geq 90% and match coverage \geq 70%) matched with *S. cerevisiae* or *S. paradoxus*. We identified a total of 5489 *S. cerevisiae*-like and 5699 *S. paradoxus*-like genes and treated these two sets separately as two independent subgenomes. We refer to these as "Living ancestor Sc" and "Living ancestor Sp" respectively in our downstream gene orthology and phylogenetic analysis. We employed proteinortho (v5.15) (Lechner et al., 2011) to identify orthologous gene groups shared among Living ancestor Sc and Living ancestor Sp, the other 32 representative *S. cerevisiae* strains and the 18 strains (7 *S. cerevisiae* + 5 *S. paradoxus* + 6 outgroup species from the *Saccharomyces* genus) that we analysed before (Yue et al., 2017). Based on this analysis, 2258 one-to-one orthologous groups shared among all these 52 strains (including Living ancestor Sc and Living ancestor Sp) were identified and used in the downstream phylogenetic analysis. The phylogenetic analysis was performed as previously described (Yue et al., 2017). The six *Saccharomyces* outgroup species are not displayed in the phylogenetic tree showed in Fig. 1a.

Mapping LOH and introgressions' boundaries

We annotated LOH and introgression regions based on the allelic fraction at a defined set of single nucleotide markers derived from the alignment of DBVPG6765 (Wine/European *S. cerevisiae*) and CBS432 (European *S. paradoxus*) genome assemblies. We generated a list of 1059399 reliable markers between the two assemblies using the software nucmer included in the suite MUMmer (v3.1) (Kurtz et al., 2004), with the option “-C” to exclude markers with ambiguous position assignments. These markers reside within the core chromosome regions where near-perfect synteny is observed between the *S. cerevisiae* and *S. paradoxus* genomes. Because synteny is lost within subtelomeres (Yue et al., 2017), these regions were excluded from our analysis. The over 1M markers distributed over 11210867 bp of aligned genome gives a very dense map of polymorphic markers at an average distance of ~10 bp. We mapped Illumina reads on the *S. cerevisiae* DBVPG6765 assembly using BWA (v0.7.12). Read mapping runs were stored in the Sequence Alignment/Map (SAM) format and its binary form (BAM) and post-processing steps were performed by picard tools (v2.8.0). Sorting, indexing and per-position coverage calculation were performed using SAMtools (v1.2).

We used the software freebayes (v0.9.5) to call markers from the BAM files. Results were stored in Variant Call Format (VCF) files. We examined the fraction of the two possible alleles at every marker and used in-house Perl scripts to assign a genotype with the following distinguishing criteria: i) If the allelic fraction is equal to 0.5 the marker is in heterozygosity, ii) If the allelic fraction equals 0 or 1 the marker is in homozygosity for the *S. cerevisiae* or the *S. paradoxus* subgenome, respectively. Markers with a coverage <20 were considered unassigned and ignored in the subsequent analyses. For both the Alpechins (except YAE and YDM) and the original non-laboratory evolved living ancestor, we increased this threshold to <50 given their higher genome coverage.

In the case of the living ancestor and its derived clones, we annotated regions containing at least 10 consecutive markers in homozygosity (interspaced by less than 10 markers with a different state) as LOH. In the case of the Alpechins, we used the same parameters and annotated regions of homozygosity for the *S. paradoxus* allele as homozygous introgressions, and regions of heterozygosity as heterozygous introgressions. As additional evidence, we tested the reliability of LOH and introgression predictions by manually inspecting recombinant reads spanning their boundaries using IGV (v2.3.68). LOH and introgressions for which we could not find recombinant reads were excluded.

The same process was used to map introgression boundaries in strains belonging to three clades previously described⁵: Brazilian bioethanol, Mexican agave and French Guiana. For each of the

introgressed clades (including Alpechin), we separately summed up the introgressions in their isolates in order to build a pan-introgression reference collection. In case multiple isolates within the same clade had different boundaries for an overlapping introgression, the largest event was kept. Overall, we genotyped 104435 markers in LOH in the living ancestor (63375 in *S. paradoxus* LOH and 41060 in *S. cerevisiae* LOH). These markers were distributed along 1704137 bp (1092514 for *S. paradoxus* LOH and 611623 for *S. cerevisiae* LOH), giving a total of 15.2% of genome in LOH (9.7% for *S. paradoxus* and 5.5% for *S. cerevisiae*). We genotyped 77955 introgressed markers in the Alpechin clade, distributed along 902805 bp (with a range of 467265 – 583627 bp in individual strains), giving a total of 8% of introgressed genome (ranging from 4% to 5% in individual strains). Subtelomeres were excluded from the genome size estimations.

We intersected the genomic coordinates of *S. paradoxus* LOH in the living ancestor with the genomic coordinates of introgressions in the Alpechins, using the program `intersectBed` included in the suite `bedtools` (v2.17). We retrieved the genomic positions in which *S. paradoxus* LOH and introgressions overlapped, for a total of 534736 overlapping bp (containing 43872 markers). We calculated the expected random number of overlapping markers (8221) as the total number of markers (1059399) multiplied by the percentage of genome in *S. paradoxus* LOH in the living ancestor (9.7%) and the percentage of introgressed genome in the Alpechin clade (8%).

Variant calling and SNVs analysis

Illumina reads were mapped to a concatenated reference assembly, which included both the *S. cerevisiae* DBVPG6765 and the *S. paradoxus* CBS432 assemblies (Yue et al., 2017). Competitive mapping against the two highly diverged species assemblies was performed by BWA (v0.7.12). Read mapping runs were stored in the Sequence Alignment/Map (SAM) format and its binary form (BAM). Sorting, indexing and per-position coverage calculation were performed using SAMtools (v1.2). Post-processing steps were performed by picard tools (v2.8.0). Single nucleotide variants (SNVs) were called from read mappings using freebayes (v0.9.5). We filtered SNVs to keep only the ones within LOH regions with two copies of the *S. paradoxus* subgenome. Subsequent filtering was performed based on the assumption that *de novo* mutations occurring after the LOH appearance should be in heterozygous state in the living ancestor's genome (AF=0.5). The allele present on the *S. paradoxus* reference genome was annotated as ancestral, whereas the alternative one was annotated as a *de novo* mutation. We confirmed that all the predicted *de novo* mutations were absent from the European and Far East Asian *S. paradoxus* populations (MAF=0) (Bergstrom et al., 2014). We verified the presence of these mutations in 25 gametes derived from the living ancestor and retained only those segregating. We assigned the alleles to the two possible haplotypes (hosted on the *S. cerevisiae* or the *S. paradoxus* subgenome of the living ancestor) according to the species that

contributed to the gametes' genomes for those specific flanking genomic positions. A total of 207 sites passed all these criteria. This set was used to investigate the relationship between the living ancestor and the Alpechin strains. We annotated the alleles present at these positions in the Alpechins' introgressions and determined the number of *de novo* mutations retained and the haplotype that was most likely inherited. We applied the same approach to the heterozygous SNVs within *S. cerevisiae* LOH regions and used the results to determine which model best explains the living ancestor-Alpechin origin (**Extended Data Fig. 2 and Supplementary Discussion 1**).

Sequencing and genomic analysis of the living ancestor gametes

The living ancestor was sporulated in KAc medium (Liti et al., 2017). 25 spores were isolated using a dissecting microscope, incubated in YPD medium at 30°C and scored for viability after six days (**Supplementary Table 4**). We sequenced the living ancestor gametes with Illumina technology (mean coverage of 100X). Mapping and *de novo* assembly of their genomes were performed as described in the paragraph “Variant calling and SNVs analysis”. The sequencing depth of coverage was extracted using SAMtools (v1.2). Coverage analyses and statistical tests were performed using R (v3.5.3). Chromosomes with values close to 100X mean coverage for both species (200X total) across their entire length were considered as aneuploidies. Recombination sites were manually mapped as described in the paragraph “Mapping LOH and introgressions' boundaries”.

Estimation of the living ancestor's base substitution rate and phenotypic impact

Read mapping and variant calling of the living ancestor before propagation (generation zero, G_0) and the mutation accumulation lines were performed as described in the paragraph “Variant calling and SNVs analysis”. We filtered the VCF files to keep only SNVs with quality higher than 100 and coverage higher than 20, and we excluded SNVs in subtelomeric regions (Yue et al., 2017).

We subtracted the SNVs present in the living ancestor G_0 from its derived mutation accumulation lines using the *vcf-isec* program included in the suite VCFtools (v0.1.16). Next, we searched for shared SNVs present in at least two mutation accumulation lines and removed them from the VCF files, based on the assumption that identical *de novo* mutations are highly unlikely to arise independently and derived from incorrect mapping or poor coverage. All the remaining SNVs were manually inspected using IGV (v2.3.68) and they were used to estimate the living ancestor's base substitution rate (μ) as $\mu = (n/bp)/g$, where n is the number of *de novo* mutations present in each line, bp is the (diploid) genome size of the living ancestor minus subtelomeric regions (22418023 bp) and g is the number of generations performed in the mutation accumulation experiment (2220). The median of base substitution rates among the eight lines was taken as the living ancestor's base substitution rate.

Finally, *de novo* mutations were annotated using the *S. cerevisiae* DBVPG6765 and the *S. paradoxus* CBS432 genome annotations (Yue et al., 2017) and their effect was predicted by Variant Effect Predictor (v87) (McLaren et al., 2016) with the “--custom” parameter for custom reference genomes.

Molecular dating of genome instability and Alpechin emergence

The set of 207 SNVs described in the paragraph “Variant calling and SNVs analysis” was further refined for molecular dating analyses. We used the *S. paradoxus* CBS432 genome annotation (Yue et al., 2017) to extract SNVs sites that are either at the third codon position of coding regions or in intergenic regions. We retained 148 SNVs to estimate the timing since the LOH began in the living ancestor and the split between the living ancestor and the Alpechin lineage. We first used a strict molecular clock model and estimated the number of generations (g) as $g = k_s / (2\mu)$ (Fay & Benavides, 2005). k_s is the substitution rate (SNVs in third position and intergenic regions/length of third positions plus intergenic regions in *S. paradoxus* LOH overlapping with introgressions) and μ is the per-nucleotide per-generation base substitution rate estimated through the analysis of the living ancestor’s mutation accumulation lines (2.31×10^{-10}). Given that 148 substitutions were detected in the 312644 bp of the living ancestor’s *S. paradoxus* LOH that overlapped with the Alpechins’ introgressions, we estimated 1024316 generations since the genome shock has begun. This corresponds to a range of 2806 – 10243 years if we assume 100 – 365 clonal generations per year (Liti et al., 2006; Rolland & Dujon, 2011). Among the 148 SNVs, 90 have been inherited by at least one Alpechin isolate. We cannot confidently say whether the remaining 58 substitutions have been lost due to chromosome segregation or because they correspond to recent mutations that occurred after the split between the living ancestor and the Alpechin lineage. The 90 SNVs shared between the living ancestor and the Alpechin isolates resulted in a maximum age of 622895 generations occurring between the genome shock and the split. We therefore estimated that the Alpechins’ birth occurred at the most 401421 generations ago, corresponding to a range of 1100 – 4014 years. This estimation assumes a similar generation time in nature in both branches after their split.

Next, we accounted for variation in the rate of the molecular clock using a random, local molecular clock model. First, we introduced mutations in the *S. paradoxus* CBS432 genome assembly in the positions corresponding to the 148 sites used in the previous molecular dating analysis, using the *vcf-consensus* program included in the suite VCFtools (v0.1.16). This allowed us to recreate living ancestor-like and Alpechin-like genome sequences. For the living ancestor-like sequence, we introduced mutations at all the 148 sites. For the Alpechin-like sequence, we introduced mutations at the 90 sites that retained mutations in the Alpechin lineage. We then extracted the sequences corresponding to the 312644 bp of the living ancestor's *S. paradoxus* LOH that overlapped with the

Alpechins' introgressions from the *S. paradoxus* CBS432 genome assembly, from the living ancestor-like sequence and from the Alpechin-like sequence. We also extracted the sequences corresponding to the living ancestor's *S. paradoxus* LOH from the *S. cerevisiae* DBVPG6765 genome assembly.

We obtained four fasta sequences (DBVPG6765, CBS432, living ancestor and Alpechin) and aligned them using FSA (v1.15.9) with the "--refinement --exonerate --anchored" options (Bradley et al., 2009). The "--anchored" option of FSA further depends on MUMmer (v3.23) (Kurtz et al., 2004) and exonerate (v2.2.0) (Slater & Birney, 2005) to assist the alignment. The aligned sequences were trimmed by trimAl (v1.4) (Capella-Gutiérrez et al., 2009) with the option "-gt 1.0" to remove gap positions. Finally, the filtered alignment was used to conduct a molecular dating analysis with BEAST2 (v2.6.2) (Bouckaert et al., 2019) with the following parameter settings:

- Substitution model: HKY (Hasegawa et al., 1985), with 4 Gamma category count and empirical base frequencies.
- Clock model: random local clock (Suchard & Drummond, 2010).
- Tree prior: calibrated Yule model (Heled & Drummond, 2010).
- Molecular clock rate prior: Gamma (alpha=0.001, beta=1000).
- Speciation rate prior: Gamma (alpha=2.0, beta=2.0).
- Molecular clock rate prior: Gamma (alpha=2.0, beta=2.0).
- Calibration node: *S. cerevisiae* - *S. paradoxus* divergence time with a log normal distribution of $M=4.9$, $S=0.11$, which translates into a median divergence time of 4.87 million years ago (mya) with 95% central 95% probability range covering 4.06 mya - 5.84 mya (based on previous estimate (Shen et al., 2018)).
- MCMC chain length: 100000000 with 1000 pre burn-in.
- MCMC chain logging frequency: every 1000 steps.

Two independent runs were performed and the convergence of these two MCMC runs was assessed by Tracer (v1.7.1) (<http://tree.bio.ed.ac.uk/software/tracer/>). TreeAnnotator (distributed with BEAST2) was used to generate the consensus tree with 10% burn-in. The resulting consensus tree was visualized with FigTree (v1.4.4) (<http://tree.bio.ed.ac.uk/software/figtree/>).

GO term analysis and annotation of LOH and introgressed genes

Introgressed genes were annotated using the CBS432 and DBVPG6765 genome annotation (**Supplementary Table 9**) (Yue et al., 2017). Standard GO term analysis was performed with the GO Term Finder tool available at *Saccharomyces* Genome Database (SGD). Significant GO terms were extracted by the algorithm implemented in the tool, with a False Discovery Rate (FDR) corrected α threshold of 0.05 (**Supplementary Table 8**).

Acknowledgements

We thank Anders Bergström, Gilles Fischer, Bertrand Llorente, Simone Mozzachiodi and Lorenzo Tattini for critical reading of the manuscript. This work was supported by Agence Nationale de la Recherche (ANR-11-LABX-0028-01, ANR-13-BSV6-0006-01, ANR-15-IDEX-01, ANR-16-CE12-0019 and ANR-18-CE12-0004) and the Swedish Research Council (2014-6547, 2014-4605 and 2018-03638). We thank the Institut Curie NGS platform for generating the sequence data. J.S. is a fellow of the University of Strasbourg Institute for Advanced Study and a member of the Institut Universitaire de France.

Competing interest

The authors declare no competing interests.

Data availability

The genome sequences generated in this study are available at Sequence Read Archive (SRA), NCBI under the accession codes BioProject ID PRJNA594913, Biosample ID SAMN13540515 - SAMN13540586. Re-analysed sequencing data are available as part of the project ERP014555 in the Sequence Read Archive (SRA). The phenotype data are available within the supplementary information files. The source data underlying Figs. 1b, 2a-d, 4a-d and Extended Data Fig. 1a-b, 2, 4a-b, 5a-d, 6a-d, 7b-e, 8a, c, d, 9a-b, 10a-c are provided in the Source Data File. All the strains generated in this work are freely available upon request.

Code and tables availability

Custom Bash and Perl scripts used in this study and supplementary tables are available at <https://github.com/mdangiolo89/A-yeast-living-ancestor-reveals-the-origin-of-genomic-introgressions>.

References

- Almeida, P., Gonçalves, C., Teixeira, S., Libkind, D., Bontrager, M., Masneuf-Pomarède, I., Albertin, W., Durrens, P., Sherman, D. J., Marullo, P., Todd Hittinger, C., Gonçalves, P., & Sampaio, J. P. (2014). A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature Communications*, 5(May). <https://doi.org/10.1038/ncomms5044>
- Arnold, B. J., Lahner, B., DaCosta, J. M., Weisman, C. M., Hollister, J. D., Salt, D. E., Bomblies, K., & Yant, L. (2016). Borrowed alleles and convergence in serpentine adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 113(29), 8320–8325. <https://doi.org/10.1073/pnas.1600405113>
- Barbosa, R., Almeida, P., Safar, S. V. B., Santos, R. O., Morais, P. B., Nielly-Thibault, L., Leducq, J.-B., Landry, C. R., Gonçalves, P., Rosa, C. A., & Sampaio, J. P. (2016). Evidence of natural hybridization in Brazilian wild lineages of *Saccharomyces cerevisiae*. *Genome Biology and Evolution*, evv263. <https://doi.org/10.1093/gbe/evv263>
- Barré, B. P., Hallin, J., Yue, J. X., Persson, K., Mikhalev, E., Irizar, A., Holt, S., Thompson, D., Molin, M., Warringer, J., & Liti, G. (2020). Intragenic repeat expansion in the cell wall protein gene HPF1 controls yeast chronological aging. *Genome Research*, 30(5), 697–710. <https://doi.org/10.1101/gr.253351.119>
- Beltran, G., Novo, M., Rozès, N., Mas, A., & Guillamón, J. M. (2004). Nitrogen catabolite repression in *Saccharomyces cerevisiae* during wine fermentations. *FEMS Yeast Research*, 4(6), 625–632. <https://doi.org/10.1016/j.femsyr.2003.12.004>
- Bergstrom, A., Simpson, J. T., Salinas, F., Barr??, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, 31(4), 872–888. <https://doi.org/10.1093/molbev/msu037>
- Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., Du Plessis, L., Poppinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., ... Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, 15(4), 1–28. <https://doi.org/10.1371/journal.pcbi.1006650>
- Bradley, R. K., Roberts, A., Smoot, M., Juvekar, S., Do, J., Dewey, C., Holmes, I., & Pachter, L. (2009). Fast statistical alignment. *PLoS Computational Biology*, 5(5). <https://doi.org/10.1371/journal.pcbi.1000392>
- Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15), 1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>
- Charron, G., Marsit S., Henault M., Martin H., Landry, C. R. (2019). Spontaneous whole-genome duplication restores fertility in interspecific hybrids. *Nature Communications*, 2019.

<http://dx.doi.org/10.1038/s41467-019-12041-8>

- Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., Clum, A., Copeland, A., Huddleston, J., Eichler, E. E., Turner, S. W., & Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods*, *10*(6), 563–569. <https://doi.org/10.1038/nmeth.2474>
- Chumki, S. A., Dunn, M. K., Coates, T. F., Mishler, J. D., Younkin, E. M., & Casper, A. M. (2016). Remarkably long-tract gene conversion induced by fragile site instability in *Saccharomyces cerevisiae*. *Genetics*, *204*(1), 115–128. <https://doi.org/10.1534/genetics.116.191205>
- De Chiara, M., Barré, B., Persson, K., Chioma, A. O., Irizar, A., Schacherer, J., Warringer, J., & Liti, G. (2020). *Domestication reprogrammed the budding yeast life cycle*. 1–28. <https://doi.org/10.1101/2020.02.08.939314>
- Dunn, B., Paulish, T., Stanbery, A., Piotrowski, J., Koniges, G., Kroll, E., Louis, E. J., Liti, G., Sherlock, G., & Rosenzweig, F. (2013). Recurrent Rearrangement during Adaptive Evolution in an Interspecific Yeast Hybrid Suggests a Model for Rapid Introgression. *PLoS Genetics*, *9*(3). <https://doi.org/10.1371/journal.pgen.1003366>
- Dutta, A., Lin, G., Pankajam, A. V., Chakraborty, P., Bhat, N., Steinmetz, L. M., & Nishant, K. T. (2017). Genome dynamics of hybrid *Saccharomyces cerevisiae* during vegetative and meiotic divisions. *G3: Genes, Genomes, Genetics*, *7*(11), 3669–3679. <https://doi.org/10.1534/g3.117.1135>
- Edelman, N. B., Frandsen, P. B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R. B., García-accinelli, G., Belleghem, S. M. Van, & Patterson, N. (2019). Butterfly Radiation. *Science*, *599*(November), 594–599.
- Enard, D., & Petrov, D. A. (2018). Evidence that RNA Viruses Drove Adaptive Introgression between Neanderthals and Modern Humans. *Cell*, *175*(2), 360-371.e13. <https://doi.org/10.1016/j.cell.2018.08.034>
- Fay, J. C., & Benavides, J. A. (2005). Evidence for domesticated and wild populations of *saccharomyces cerevisiae*. *PLoS Genetics*, *1*(1), 0066–0071. <https://doi.org/10.1371/journal.pgen.0010005>
- Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G., & Louis, E. J. (2000). Chromosomal evolution in *Saccharomyces*. *Nature*, *405*(6785), 451–454. <https://doi.org/10.1038/35013058>
- Gallone, B., Steensels, J., Mertens, S., Dzialo, M. C., Gordon, J. L., Wauters, R., Theßeling, F. A., Bellinazzo, F., Saels, V., Herrera-Malaver, B., Prahl, T., White, C., Hutzler, M., Meußdoerffer, F., Malcorps, P., Souffriau, B., Daenen, L., Baele, G., Maere, S., & Verstrepen, K. J. (2019). Interspecific hybridization facilitates niche adaptation in beer yeast. *Nature Ecology and Evolution*, *3*(11), 1562–1575. <https://doi.org/10.1038/s41559-019-0997-9>
- Gerstein, A. C., Kuzmin, A., & Otto, S. P. (2014). Loss-of-heterozygosity facilitates passage through haldane’s sieve for *saccharomyces cerevisiae* undergoing adaptation. *Nature Communications*, *5*(May), 1–9. <https://doi.org/10.1038/ncomms4819>
- Greig, D. (2009). Reproductive isolation in *Saccharomyces*. *Heredity*, *102*(1), 39–44.

<https://doi.org/10.1038/hdy.2008.73>

- Greig, Duncan, Borts, R. H., Louis, E. J., & Travisano, M. (2002). Epistasis and hybrid sterility in *Saccharomyces*. *Proceedings. Biological Sciences / The Royal Society*, 269(1496), 1167–1171. <https://doi.org/10.1098/rspb.2002.1989>
- Greig, Duncan, Louis, E. J., Borts, R. H., & Travisano, M. (2002). Hybrid speciation in experimental populations of yeast. *Science*, 298(5599), 1773–1775. <https://doi.org/10.1126/science.1076374>
- Haber, J. E. (2012). Mating-type genes and MAT switching in *Saccharomyces cerevisiae*. *Genetics*, 191(1), 33–64. <https://doi.org/10.1534/genetics.111.134577>
- Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, 105(S1), 795–809. <https://doi.org/10.1093/jhered/esu033>
- Hasegawa, M., Kishino, H., & Yano, T. aki. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22(2), 160–174. <https://doi.org/10.1007/BF02101694>
- Heled, J., & Drummond, A. J. (2010). Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*, 27(3), 570–580. <https://doi.org/10.1093/molbev/msp274>
- Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, 512(7513), 194–197. <https://doi.org/10.1038/nature13408>
- Hunter, N., Chambers, S. R., Louis, E. J., & Borts, R. H. (1996). The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. TL - 15. *The EMBO Journal*, 15 VN-r(7), 1726–1733. /Users/emilyclare/Documents/ReadCube Media/EMBO J 1996 Hunter N.pdf%5Cnhttp://europepmc.org/abstract/MED/8612597
- Gerke J., Lorenz K., Cohen B. (2009). Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. *Science*. 323(5913): 498–501
- James, T. Y., Michelotti, L. A., Glasco, A. D., Clemons, R. A., Powers, R. A., James, E. S., Rabern Simmons, D., Bai, F., & Ge, S. (2019). Adaptation by loss of heterozygosity in *Saccharomyces cerevisiae* clones under divergent selection. *Genetics*, 213(2), 665–683. <https://doi.org/10.1534/genetics.119.302411>
- Kaniewski, D., Van Campo, E., Boiy, T., Terral, J. F., Khadari, B., & Besnard, G. (2012). Primary domestication and early uses of the emblematic olive tree: Palaeobotanical, historical and molecular evidence from the Middle East. *Biological Reviews*, 87(4), 885–899. <https://doi.org/10.1111/j.1469-185X.2012.00229.x>
- Kao, K. C., Schwartz, K., & Sherlock, G. (2010). A genome-wide analysis reveals no nuclear Dobzhansky-Muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genetics*, 6(7), 1–12. <https://doi.org/10.1371/journal.pgen.1001038>

- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, 5(2). <https://doi.org/10.1186/gb-2004-5-2-r12>
- Langdon, Q. K., Peris, D., Baker, E. P., Opulente, D. A., Bond, U., Gonçalves, P., Sampaio, J. P., & Libkind, D. (2020). *and Domesticated Yeasts*. 2019(11), 1576–1586. <https://doi.org/10.1038/s41559-019-0998-8>. Fermentation
- Laureau, R., Loeillet, S., Salinas, F., Bergström, A., Legoix-Né, P., Liti, G., & Nicolas, A. (2016). Extensive Recombination of a Yeast Diploid Hybrid through Meiotic Reversion. *PLOS Genetics*, 12(2), e1005781. <https://doi.org/10.1371/journal.pgen.1005781>
- Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F., & Prohaska, S. J. (2011). Proteinortho : Detection of (Co -) orthologs in large-scale analysis. *BMC Bioinformatics*. 12, Article number:124.
- Lee, H. Y., Chou, J. Y., Cheong, L., Chang, N. H., Yang, S. Y., & Leu, J. Y. (2008). Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. *Cell*, 135(6), 1065–1073. <https://doi.org/10.1016/j.cell.2008.10.047>
- Liti, G., Barton, D. B. H., & Louis, E. J. (2006). Sequence diversity, reproductive isolation and species concepts in saccharomyces. *Genetics*, 174(2), 839–850. <https://doi.org/10.1534/genetics.106.062166>
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. a, Davey, R. P., Roberts, I. N., Burt, A., Tsai, I. J., Bergman, C. M., Bensasson, D., Michael, J. T., Kelly, O., Oudenaarden, A. Van, Barton, D. B. H., Bailes, E., Alex, N., Ba, N., Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*. 458(7236), 337–341. <https://doi.org/10.1038/nature07743>.
- Liti, G., Warringer, J., & Blomberg, A. (2017). Isolation and laboratory domestication of natural yeast strain. *Cold Spring Harbor Protocols*, 2017(8), 626–630. <https://doi.org/10.1101/pdb.prot089052>
- Magwene, P. M., Kayikçi, Ö., Granek, J. A., Reininga, J. M., Scholl, Z., & Murray, D. (2011). Outcrossing, mitotic recombination, and life-history trade-offs shape genome evolution in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 108(5), 1987–1992. <https://doi.org/10.1073/pnas.1012544108>
- Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology and Evolution*, 20(5), 229–237. <https://doi.org/10.1016/j.tree.2005.02.010>
- Marsit, S., Leducq, J. B., Durand, É., Marchant, A., Filteau, M., & Landry, C. R. (2017). Evolutionary biology through the lens of budding yeast comparative genomics. *Nature Reviews Genetics*, 18(10), 581–598. <https://doi.org/10.1038/nrg.2017.49>
- Martin, S. H., & Jiggins, C. D. (2017). Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development*, 47, 69–74. <https://doi.org/10.1016/j.gde.2017.08.007>
- McClure, A. W., Jacobs, K. C., Zyla, T. R., & Lew, D. J. (2018). Mating in wild yeast: Delayed interest in sex after spore germination. *Molecular Biology of the Cell*, 29(26), 3119–3127.

<https://doi.org/10.1091/mbc.E18-08-0528>

- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 1–14. <https://doi.org/10.1186/s13059-016-0974-4>
- Morales, L., & Dujon, B. (2012). Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and Molecular Biology Reviews : MMBR*, *76*(4), 721–739. <https://doi.org/10.1128/MMBR.00022-12>
- Mozzachiodi, S., Tattini, L., Llored, A., Irizar, A., Škofljanc, N., Angiolo, D., Chiara, M. De, Barré, B. P., Yue, J., Lutazi, A., Loillet, S., Laureau, R., Marsit, S., Stenberg, S., Persson, K., Legras, J., & Dequin, S., Warringer J., Nicolas A., Liti G. Aborting meiosis overcomes hybrid sterility.
- Bozdag O., Ono J., Denton J., Karakoc E., Hunter N., Leu J., Greig D. Engineering recombination between diverged yeast species reveals genetic incompatibilities. BioArxiv. doi: <https://doi.org/10.1101/755165>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. *Nature*. 556(7701):339-344.
- Pontes, A., Čadež, N., Gonçalves, P., & Sampaio, J. P. (2019). A Quasi-Domesticated Relic Hybrid Population of *Saccharomyces cerevisiae* × *S. paradoxus* Adapted to Olive Brine. *Frontiers in Genetics*, *10*(MAY), 1–14. <https://doi.org/10.3389/fgene.2019.00449>
- Prüfer, K., Racimo, F., Patterson, N., Jay, F., Sankararaman, S., Sawyer, S., Heinze, A., Renaud, G., Sudmant, P. H., De Filippo, C., Li, H., Mallick, S., Dannemann, M., Fu, Q., Kircher, M., Kuhlwilm, M., Lachmann, M., Meyer, M., Ongyerth, M., ... Pääbo, S. (2014). The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature*, *505*(7481), 43–49. <https://doi.org/10.1038/nature12886>
- Rogers, D. W., McConnell, E., Ono, J., & Greig, D. (2018). Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast. *PLoS Biology*, *16*(11), 1–17. <https://doi.org/10.1371/journal.pbio.2005066>
- Rolland, T., & Dujon, B. (2011). Yeasty clocks: Dating genomic changes in yeasts. *Comptes Rendus - Biologies*, *334*(8–9), 620–628. <https://doi.org/10.1016/j.crvi.2011.05.010>
- Roncoroni, M., Santiago, M., Hooks, D. O., Moroney, S., Harsch, M. J., Lee, S. A., Richards, K. D., Nicolau, L., & Gardner, R. C. (2011). The yeast IRC7 gene encodes a β -lyase responsible for production of the varietal thiol 4-mercapto-4-methylpentan-2-one in wine. *Food Microbiology*, *28*(5), 926–935. <https://doi.org/10.1016/j.fm.2011.01.002>
- Ruderfer, D. M., Pratt, S. C., Seidel, H. S., & Kruglyak, L. (2006). Population genomic analysis of outcrossing and recombination in yeast. *Nature Genetics*, *38*(9), 1077–1081. <https://doi.org/10.1038/ng1859>
- Schacherer, J. (2015). Chromosomal onset of reproductive isolation in *Saccharomyces cerevisiae*. *Current Biology*, *24*(10), 1153–1159. <https://doi.org/10.1016/j.cub.2014.03.063>

- Sharp, N. P., Sandell, L., James, C. G., & Otto, S. P. (2018). The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(22), E5046–E5055. <https://doi.org/10.1073/pnas.1801040115>
- Shen, X. X., Opulente, D. A., Kominek, J., Zhou, X., Steenwyk, J. L., Buh, K. V., Haase, M. A. B., Wisecaver, J. H., Wang, M., Doering, D. T., Boudouris, J. T., Schneider, R. M., Langdon, Q. K., Ohkuma, M., Endoh, R., Takashima, M., Manabe, R., Ichiroh, Čadež, N., Libkind, D., Rokas, A. (2018). Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell*, *175*(6), 1533–1545.e20. <https://doi.org/10.1016/j.cell.2018.10.023>
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, *19*(6), 1117–1123. <https://doi.org/10.1101/gr.089532.108>
- Slater, G. S. C., & Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics*, *6*, 1–11. <https://doi.org/10.1186/1471-2105-6-31>
- Slon, V., Mafessoni, F., Vernot, B., Filippo, C. De, Grote, S., Viola, B., Hajdinjak, M., Peyrégne, S., Nagel, S., Brown, S., Kelso, J., Meyer, M., Prüfer, K., & Pääbo, S. (2019). The genome of the offspring of a Neandertal mother and a Denisovan father. *Nature*. *561*(7721), 113–116. <https://doi.org/10.1038/s41586-018-0455-x>.
- Smukowski Heil, C. S., DeSevo, C. G., Pai, D. A., Tucker, C. M., Hoang, M. L., & Dunham, M. J. (2017). Loss of Heterozygosity Drives Adaptation in Hybrid Yeast. *Molecular Biology and Evolution*, *34*(7), 1596–1612. <https://doi.org/10.1093/molbev/msx098>
- Suchard, M. A., & Drummond, A. J. (2010). Bayesian random local clocks, or one rate to rule them all. *BMC Biology*, *8*(1), 114. <http://www.biomedcentral.com/1741-7007/8/114>
- Sun, Y., Corcoran, P., Menkis, A., Whittle, C. A., Andersson, S. G. E., & Johannesson, H. (2012). Large-scale introgression shapes the evolution of the mating-type chromosomes of the filamentous ascomycete *Neurospora tetrasperma*. *PLoS Genetics*, *8*(7). <https://doi.org/10.1371/journal.pgen.1002820>
- Symington, L. S., Rothstein, R., & Lisby, M. (2014). Mechanisms and Regulation of Mitotic Recombination in *Saccharomyces cerevisiae*. *Genetics*. *198*(3): 795–835.
- Tattini, L., Tellini, N., Mozzachiodi, S., D’Angiolo, M., Loeillet, S., Nicolas, A., & Liti, G. (2019). Accurate Tracking of the Mutational Landscape of Diploid Hybrid Genomes. *Molecular Biology and Evolution*, *36*(12), 2861–2877. <https://doi.org/10.1093/molbev/msz177>
- Tsai, I. J., Bensasson, D., Burt, A., & Koufopanou, V. (2008). Population genomics of the wild yeast *Saccharomyces paradoxus*: Quantifying the life cycle. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(12), 4957–4962. <https://doi.org/10.1073/pnas.0707314105>
- Vázquez-García, I., Salinas, F., Li, J., Fischer, A., Barré, B., Hallin, J., Bergström, A., Alonso-Perez, E., Warringer, J., Mustonen, V., & Liti, G. (2017). Clonal Heterogeneity Influences the Fate of New Adaptive Mutations. *Cell Reports*, *21*(3), 732–744.

<https://doi.org/10.1016/j.celrep.2017.09.046>

- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C. A., Zeng, Q., Wortman, J., Young, S. K., & Earl, A. M. (2014). Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*, 9(11). <https://doi.org/10.1371/journal.pone.0112963>
- Wolf, A. B., & Akey, J. M. (2018). Outstanding questions in the study of archaic hominin admixture. *PLoS Genetics*, 14(5), 1–14. <https://doi.org/10.1371/journal.pgen.1007349>
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Yue, J. X., & Liti, G. (2018). Long-read sequencing data analysis for yeasts. *Nature Protocols*, 13(6), 1213–1231. <https://doi.org/10.1038/nprot.2018.025>
- Zackrisson, M., Hallin, J., Ottosson, L. G., Dahl, P., Fernandez-Parada, E., Ländström, E., Fernandez-Ricaud, L., Kaferle, P., Skyman, A., Stenberg, S., Omholt, S., Petrovic, U., Warringer, J., & Blomberg, A. (2016). Scan-o-matic: High-resolution microbial phenomics at a massive scale. *G3: Genes, Genomes, Genetics*, 6(9), 3003–3014. <https://doi.org/10.1534/g3.116.032342>

Chapter 5

Telomeres are shorter in wild *S. cerevisiae* populations than in domesticated ones

Jia Xing Yue and me designed and implemented the pipeline to estimate telomere length; Matteo De Chiara and me performed and analysed GWAS; Benjamin Barré performed and analysed mitochondrial phenotyping; Marie-Josèphe Giraud-Panis and me performed Southern blots; Eric Gilson and Gianni Liti conceived and supervised the project; Eric Gilson, Gianni Liti and me wrote the paper.

Abstract

Telomeres are ribonucleoproteins that cap chromosome-ends. Their length is maintained and strictly controlled by telomerase. Telomere length has been measured in several laboratory strains of the budding yeast *Saccharomyces cerevisiae*, but little is known about its variation in wild isolates. Moreover, the quantitative contribution of telomere length to the organismal fitness remains largely unexplored due to the lack of an adequate number of strains and the laboriousness of the current measuring techniques. The recent publication of the genomes and phenomes of more than 1000 yeast strains enabled us to explore telomere length variation on an unprecedented scale. Here, we developed a fast and cost-effective bioinformatic pipeline to estimate the average telomere length of these strains from whole-genome-sequences. Our results revealed natural telomere length variation among the isolates, which can be explained by genetic factors and ecological origin. Notably, telomere length was shorter in wild isolates than in isolates from domestic environments. Genetic association studies of wild isolates revealed an enrichment for loss-of-function mutations in genes known to regulate telomere length and showed that the return of domesticated yeasts to a wild habitat leads to telomere shortening. Moreover, telomere length variation is associated with mitochondrial metabolism in wild but not in domesticated strains. Overall, these findings suggest that short telomeres contribute to the fitness of yeasts living in wild environments and that this constraint is lost upon domestication.

Introduction

Telomeres are ribonucleoprotein structures situated at the ends of chromosomes (Blackburn & Gall, 1978). They comprise tandem repeats of a DNA sequence whose motif and length vary among species (Fulnečková et al., 2013; Monaghan, 2010). They maintain chromosome structure by distinguishing natural chromosome-ends from accidental double-stranded breaks (DSBs) and by contributing to chromosome segregation during cell division. Telomere DNA length (TL) is determined by the opposite actions of elongation and degradation pathways: elongation is ensured by telomerase, a reverse transcriptase that uses RNA as a template (Greider & Blackburn, 1985, 1987), or by alternative lengthening of telomeres (ALT) recombination pathway (Lundblad & Blackburn, 1993; Teng & Zakian, 1999; Teng et al., 2000); degradation is linked to the semi-conservative replication of DNA extremities coupled to specific nuclease activities (Olovnikov, 1973; Watson, 1972). In mammalian somatic cells, telomerase is not or only weakly expressed and telomeres shorten at each cell division. By contrast, telomerase is expressed constitutively in the budding yeast *Saccharomyces cerevisiae* allowing telomeric DNA to be maintained at an equilibrium length (Wellinger & Zakian, 2012). The mean TL differs among species. Human telomeres can range between 5 and 15 kb while mice have extremely long telomeres (even >100 kb) (Monaghan, 2010). In comparison, yeast telomeres are relatively short at 300±75 bp (Wellinger & Zakian, 2012). Telomere length is a complex trait: in budding yeast it is modulated by more than 400 genes, named the Telomere Length Maintenance (TLM) genes, and by various environmental factors, including heat, caffeine and ethanol (Kupiec, 2014; Romano et al., 2013).

Telomere length also varies among individuals of the same species, including yeast, nematode, humans, rodents, bats and birds (Cook et al., 2016; Foley et al., 2018; Hansen et al., 2016; Heidinger et al., 2012; Liti et al., 2009; Zijlmans et al., 1997). In humans, telomere length variation can have profound consequences on fitness and has been associated with different pathological conditions: short telomeres are correlated with a higher risk of developing cardiovascular and degenerative diseases and a higher risk of mortality, while long telomeres increase the risk of cancer through their capacity to increase the proliferative potential of cells, favoring their immortalization (Aviv & Shay, 2018). Telomere length variation among different human ethnic groups has recently been associated to selection to attenuate the risk of melanoma in individuals with light skin pigmentation (Hansen et al., 2016; Mangino et al., 2015). In birds, telomere length has been used to predict key life-history traits like growth, reproduction and lifespan. Within a given species, individuals with shorter mean telomere lengths have a lower life expectancy (Hausmann et al., 2005; Heidinger et al., 2012). However, among species, the rate of telomere attrition might be a better predictor of lifespan than the mean telomere length (Whittemore et al., 2019). It was proposed that telomere attrition is an adaptive strategy based on life-history regulation and

environmental adaptation (Young, 2018).

In yeast telomerase is constitutively expressed and it guarantees the homeostasis of telomere length over time (Wellinger & Zakian, 2012). Most of the knowledge on their telomere length regulation comes from studies of a few laboratory-derived strains. Of note, examples of extreme telomere length variation among different subpopulations of both *S. cerevisiae* and its closest wild relative *S. paradoxus* have been reported (Liti et al., 2009). This raises several questions: does selection act on telomere length regulation and does telomere length affect organismal fitness and influence its adaptation to the environment? The lack of an adequate number of strains, especially the ones isolated from wild habitats, has so far prevented these questions from being addressed adequately.

The recent publication of the genomes and phenomes of more than 1000 yeast (Peter et al., 2018) strains enabled us to explore telomere length variation at an unprecedented scale. Here, we report the first comprehensive analysis of telomere length variation in yeast strains isolated from a wide range of ecological niches and geographic areas, including domesticated and wild environments. For this purpose, we developed a fast and cost-effective bioinformatic pipeline to estimate the average telomere length from whole-genome sequences. We described this variation across 26 well-defined lineages and identified the genetic variants involved. We correlated TL with a series of phenotypes including growth using different carbon and nitrogen sources, mitochondrial metabolism, sporulation capacity and chronological lifespan (CLS). The results revealed different telomere length regulation patterns in wild and domestic yeasts that correlate with their mitochondrial activities.

Results

Telomere length plasticity within a large collection of *Saccharomyces cerevisiae* isolates

We estimated the TL of 918 strains out of the 1011 *Saccharomyces cerevisiae* strains using a bioinformatic pipeline designed to infer TL from *S. cerevisiae* next-generation sequencing data (**Supplementary Discussion 1 and Extended Data Figs. 1, 2**). TL was normally distributed with a median of 236±96 bp (**Extended Data Fig. 3A**), a value that is close to that of the S288C laboratory strain. Some isolates have either very short (e.g. 41 bp) or very long (e.g. 1210 bp) telomeres (**Extended Data Fig. 3A and Supplementary Table 1**). TL varied significantly among 26 previously described clades and three mosaic lineages (two-tailed Kruskal-Wallis test, $p=9.154e^{-07}$) (Peter et al., 2018). TL was longer in the Alpechin and Mexican agave clades, where it exceeded 300 bp, and it was shorter in Malaysian and Ecuadorean clades, where it remained below 200 bp (**Fig. 1 and Supplementary Table 2**). The Wine/European TL had a distribution comparable to that of the entire collection, ranging from 45 to 750 bp (median=238±101 bp), consistent with the largest clade sample size (~1/3 of the whole collection) (**Extended Data Fig. 3B**). This lineage

contains four well-defined subclades (Semi-wild, Clinical/Y' amplification, Clinical/*S. boulardii*, Georgian) (Peter et al., 2018). TL varied significantly among the subclades (two-tailed Kruskal-Wallis test, $p=3.394e-08$). It was highest in the Clinical/Y' amplification subclade where it exceeded 600 bp, and lowest in the Clinical/*S. boulardii* subclade, where it remained below 200 bp (**Extended Data Fig. 3C and Supplementary Table 2**). Clinical/Y' amplification strains are known to have aberrant TL and amplification of interstitial telomeric sequences (ITS) and Y' elements (Bergstrom et al., 2014). Their TL was confirmed by teloblot (**Extended Data Fig. 2B**). The other subclades and remaining Wine/European isolates had a TL comparable to the median of the collection.

We examined whether ploidy and aneuploidy had an effect on TL but we found no variation in TL among strains with variable ploidy and aneuploidy (two-tailed Kruskal-Wallis test, $p=0.15$ and $p=0.76$, respectively) (**Extended Data Figs. 3D, E**).

We also measured the ITS content across the 918 yeasts collection and found it to be highly variable but positively correlated with TL (Pearson's $R=0.43$, $p<2.2e-16$) and Y' content (Pearson's $R=0.81$, $p<2.2e-16$) (**Supplementary Discussion 2 and Extended Data Figs. 4, 5**). The marked association between ITS and Y' can simply be explained by the presence of short stretches of telomeric repeats between Y' copies. The correlations between TL and Y'/ITS were not expected. Overall, our analyses showed that TL exhibits inter- and intra-clade variation.

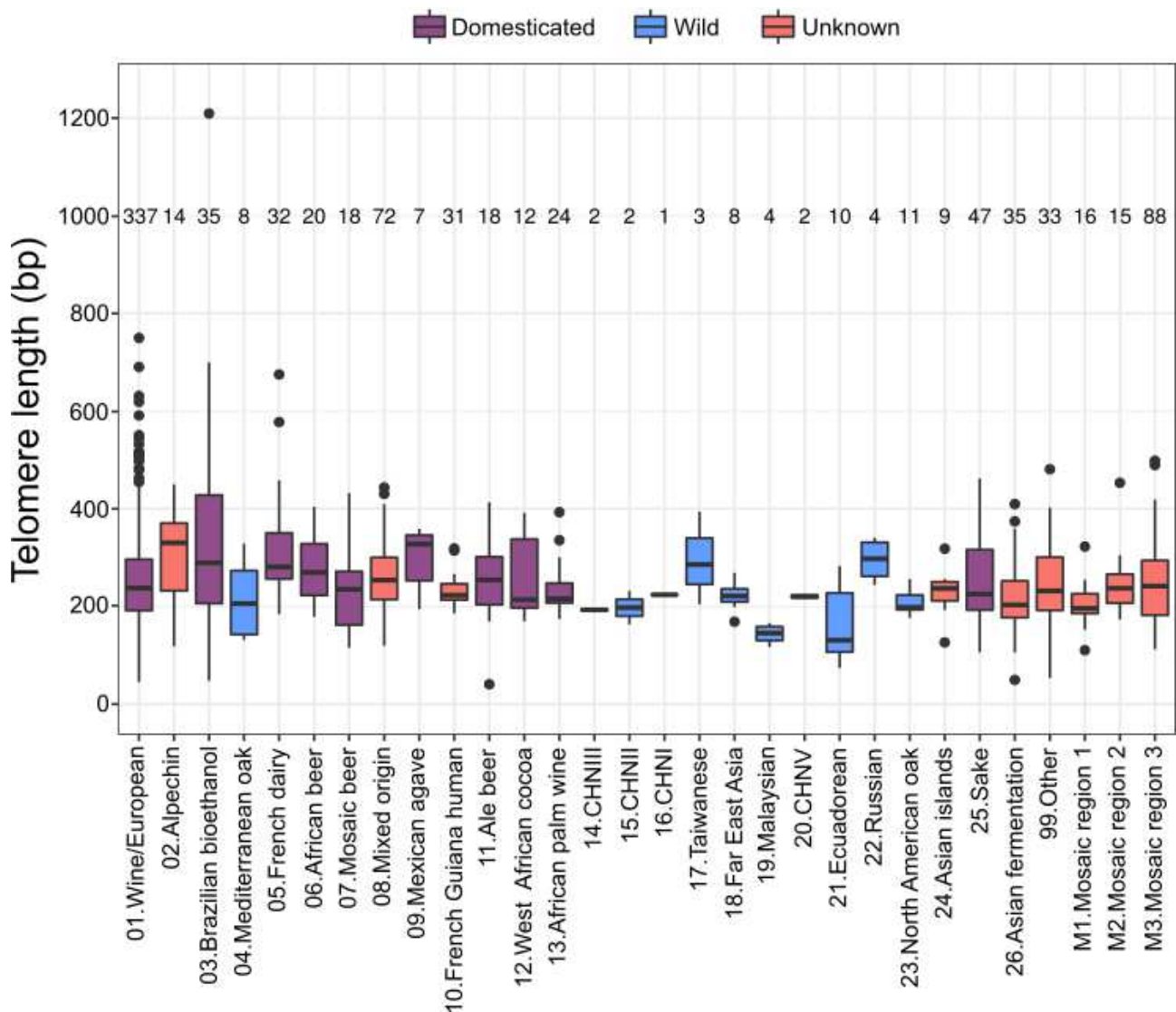


Figure 1 - Telomere length of the phylogenetic lineages described in the 1002 Yeast Genome Project (Peter et al., 2018). Colours represent the clade classification (domesticated, wild or unknown), which is the same as in (Peter et al., 2018): clades are assigned as wild or domesticated based on the environmental origin of the isolates dominating (>66%) in the clade. Clades were labelled “unknown” if the dominance criteria was not reached. The order of clades on the x-axis is the same as in (Peter et al., 2018). In the box plots, horizontal lines denote the median, top and bottom hinges denote the IQR, whiskers denote maximum and minimum values within upper and lower hinges $\pm 1.5 \times$ IQR. Numbers on top represent the number of isolates in each clade.

Domesticated and wild yeast isolates exhibit different TLs

Since domestication is an important lifestyle trait in *Saccharomyces cerevisiae* (De Chiara et al., 2020), we investigated whether it was associated with TL. Isolates were classified as wild ($n=55$) or domesticated ($n=550$) based on the dominant origin (>66%) in their clade, according to (Peter et al., 2018). Domesticated isolates (244 ± 108 bp) had significantly (two-tailed Wilcoxon test, $p=9.296e-$

⁰⁵) longer telomeres than wild ones (213 \pm 66 bp), with a median difference of \sim 30 bp and a significant TL variance (two-tailed Fligner-Killeen test, $p=0.02$) between the two groups (**Fig. 2A**). Consistently, TL was longer in domesticated clades than in wild ones (**Fig. 1**).

Within the domesticated group, most isolates originated from a domesticated environments and were called “anthropic”, while others were isolated in nature and were called “feral”. The genome ancestries of feral strains resembled those of isolates from domesticated clades, meaning that they had been domesticated and then they returned to natural habitats. We compared TL between these groups and found that feral isolates had a shorter median TL than anthropic ones (domesticated-anthropic=256 \pm 105 bp, domesticated-feral=227 \pm 80 bp, two-tailed Wilcoxon test, $p=0.001379$) (**Fig. 2B**). This confirms that the long TL pattern is associated with domestication and indicates that the reintroduction in a natural environment results in TL shortening.

Since the strains belonging to domesticated clades were isolated mostly in Europe and Africa, while the wild ones were derived mostly from America and Asia, the association of long TL variation with domesticated strains might reflect an effect of isolation location. However, we found no TL variation (two-tailed Kruskal-Wallis test, $p=0.11$) across continents, confirming that the domestication process, and not the geographic origin, drives TL variation (**Extended Data Fig. 6A**).

To further determine whether domestication results in a long TL pattern, we compared the TL of *S. cerevisiae* to that of *S. paradoxus*, its closest phylogenetic relative which was never domesticated. We estimated the TL of *S. cerevisiae* ($n=33$) and *S. paradoxus* ($n=24$) strains from the *Saccharomyces* Genome Resequencing Project (SGRP) collection using already available Sanger sequencing data (Liti et al., 2009). *S. cerevisiae* telomeres were longer than their *S. paradoxus* counterparts with a median difference of \sim 75 bp (*S. cerevisiae*=256 \pm 93 bp, *S. paradoxus*=181 \pm 91 bp, two-tailed Wilcoxon test, $p=3.96e^{-12}$). SGRP *S. cerevisiae* telomeres had lengths comparable to those of domesticated *S. cerevisiae* strains, while *S. paradoxus* telomeres had lengths comparable to *S. cerevisiae* wild strains (**Fig. 2C and Extended Data Fig. 6B**). Moreover, *S. paradoxus* TL is very different in its three continental subpopulations, with short telomeres in European isolates and very long ones in North American isolates (Liti et al., 2009). Therefore, the high variance in the *S. paradoxus* group indicates that genetic drift in isolated subpopulations can also result in large TL variation.

We conclude that *S. cerevisiae* strains exhibit different TL patterns according to their lifestyles: long for domesticated and short for wild strains. Impressively, the domesticated pattern is reversible as exemplified in the feral isolates, showing that TL can change according to the lifestyle conditions.

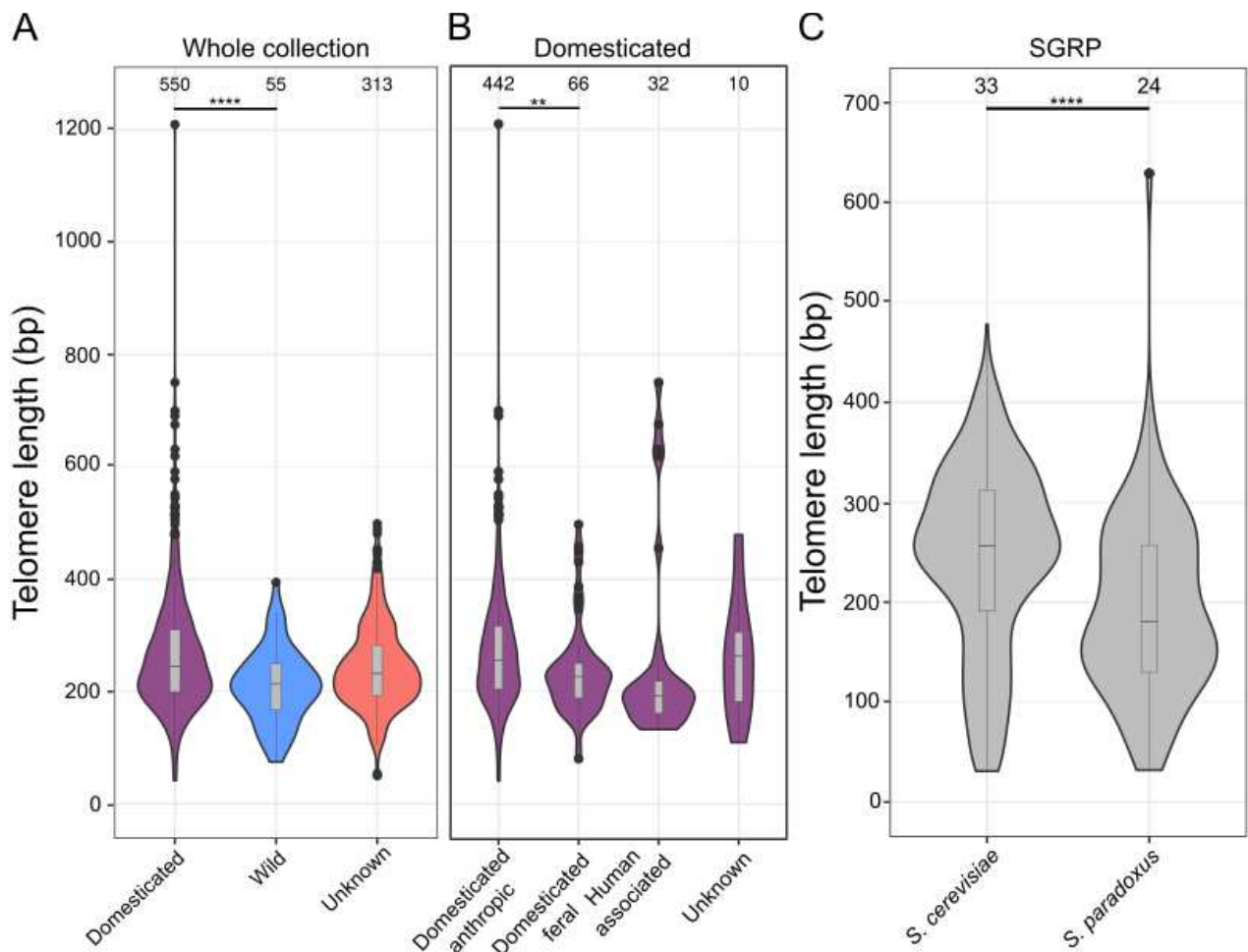


Figure 2 – **A**, Telomere length of domesticated vs wild isolates. Isolates classification as in Fig. 1. Box plots are as in Fig. 1. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **B**, Telomere length of domesticated isolates, divided in groups based on their environment of isolation. Domesticated-anthropic are strains isolated from fermentation processes; domesticated-feral are strains isolated from natural environments; human-associated comprises strains isolated from the human body; unknown comprises strains whose environment of isolation is unknown. Box plots are as in Fig. 1. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **C**, Telomere length of *S. cerevisiae* vs *S. paradoxus* strains from the *Saccharomyces* genome resequencing project. Box plots are as in Fig. 1. Numbers on top represent the number of isolates in each group. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Novel genetic variants and loss of function of TLM genes explain TL variation

To search for associations between common (minor allele frequency (MAF) > 5%) single nucleotide variants (SNVs), copy number variants (CNV) and TL, we conducted a genome-wide association study (GWAS) across 555 euploid diploid strains. We identified 88 QTL/Ns (false

discovery rate (FDR) corrected $\alpha=0.05$), divided into 62 SNVs and 26 CNVs, including 4 mitochondrial CNVs. Among the SNVs, 41 were intragenic while 21 were in non-coding regions. Of the intragenic ones, 17 were missense while 24 were synonymous (**Fig. 3A and Supplementary Table 4**) Overall, GWAS hits showed enrichment in genes involved in RNA metabolism and intracellular trafficking (**Supplementary Table 4**).

Interestingly, four QTL/Ns are TLM genes (*HTL1*, *YHL012W*, *YKU70*, *CSR2*), but overall there was not a significant enrichment in TLM genes among the QTL/Ns. Despite not being classified as TLM genes, 25 of the remaining QTL/Ns encode proteins that belong to multi-protein complexes or pathways involved in telomere length regulation, and three are in the immediate upstream/downstream regions of TLM genes (*YLR230W*, *RAD50* and *SRB8/AHC2*).

Among the rest, one SNV was inside a gene (*PPS1*) located close to *RIF1*, while a group of seven variants (multiple *CDC42* and *TOP3* hits) was located near *EST1*. The presence of intergenic and synonymous hits suggests that these variants act by regulating the expression of neighboring TLM genes.

For the four mitochondrial and 20 out of the 22 nuclear CNVs, a shorter TL was associated with a higher copy number, whereas the two remaining CNVs (*YCR102C* and *YFR054C*) showed the opposite trend (see **Fig. 3B** for an example of mitochondrial CNV **and Extended Data Fig. 7** for the nuclear CNV). In almost all the CNVs (except *YCR102C*, *YFR054C* and mitochondrial CNVs) TL variation is driven by the Clinical/*S. boulardii* subclade and the Ecuadorean clade, which contain many isolates carrying multiple gene copies (**Extended Data Fig. 7**). In most of the SNVs (57/62), the allele with the lowest frequency in the population was associated with longer telomeres, but no major clades emerged as the main drivers of TL variation (**Extended Data Figs. 8, 9, 10**).

We tested whether some QTL/Ns could explain the TL difference between domesticated and wild isolates. In a subset of 16 SNVs, the frequency of the minor allele was higher than expected in domesticated clades, but lower in wild ones. In 13 cases, the TL phenotype conferred by the minor allele was in accordance with the TL phenotype of the clades which showed a high frequency of that allele. Among these 13 cases, 12 had a minor allele conferring longer telomeres and prevalent in domesticated clades, and one (*AHA1_YDR215C*) had a minor allele fixed in the wild clades and conferring shorter telomeres. We show these 13 examples in (**Fig. 3C and Supplementary Table 4**). The 12 SNVs associated with long telomeres in domestic clades are enriched in genes involved in exonucleolytic activity and intracellular trafficking (**Supplementary Table 4**). We tested whether these 13 SNVs could also explain the TL difference between anthropic and feral isolates but the results were not statistically significant.

Rare variants can evade GWAS detection, so we traced their effect by pinpointing likely loss of function (LOF) variants in TLM genes ($n=383$). LOF variants were enriched in wild strains (two-

tailed X^2 test, $p=7e^{-9}$), but this enrichment was more pronounced in genes whose deletion causes telomere shortening ($p=4.3e^{-15}$) rather than lengthening ($p=0.02$), in accordance with the trend of wild isolates to have shorter telomeres. We then tested whether the LOF variants could also explain the TL difference between anthropic and feral isolates but the results were not significant. Of note, the Malaysian clade carries multiple chromosomal rearrangements (Yue et al., 2017) and short telomeres, which might be due to fixed LOF variants in *SIR4*, whose deletion causes telomere shortening and genome instability, or to *HUR1*, a gene involved in non-homologous end-joining repair. Other clades carried LOF variants in TLM genes, which could contribute to their TL phenotype. (**Supplementary Table 5**). Among the clades with the longest telomeres, the clinical/Y' amplification, Alpechin and Brazilian Bioethanol did not carry any LOF. However, the *RIF2* gene of all the Alpechin strains is introgressed from *S. paradoxus* and the protein carries 13 aminoacid substitutions affecting its stability and the interaction with its partners, providing an explanation for the long phenotype of this clade. The long phenotype of the French Dairy clade was associated to 3 LOF variants (*HUR1*, *MND2*, *GUP2*), including 1 in the gene *GUP2*, whose deletion causes telomere lengthening. The African beer clade presented LOF variants in both telomere lengthening (*YBR284W*) and shortening (*MET7*) genes. Among the strains with the shortest telomeres, the clinical/*S. boulardii* subclade carried LOFs in *AZR1* and *WHI2*, whose deletion leads to telomere shortening. On the contrary, a third LOF in the gene *RPS30B* causes telomere lengthening. A LOF in *RNR1*, encoding a subunit of the ribonucleotide-diphosphate reductase, is fixed in the North American clade. The inactivation of this gene causes telomere shortening due to reduced availability of dNTPs and is in accordance with the phenotype of the clade. A LOF in *URN1*, causing telomere shortening, was detected in the Ecuadorean strains.

In conclusion, we detected a set of genetic variants and LOF mutations in TLM genes that contribute to the TL phenotypes observed in our *S. cerevisiae* collection, both at the single clade level and between domesticated and wild strains. Moreover, the enrichment of LOF variants in wild strains suggests that telomere length is under selection in natural environments.

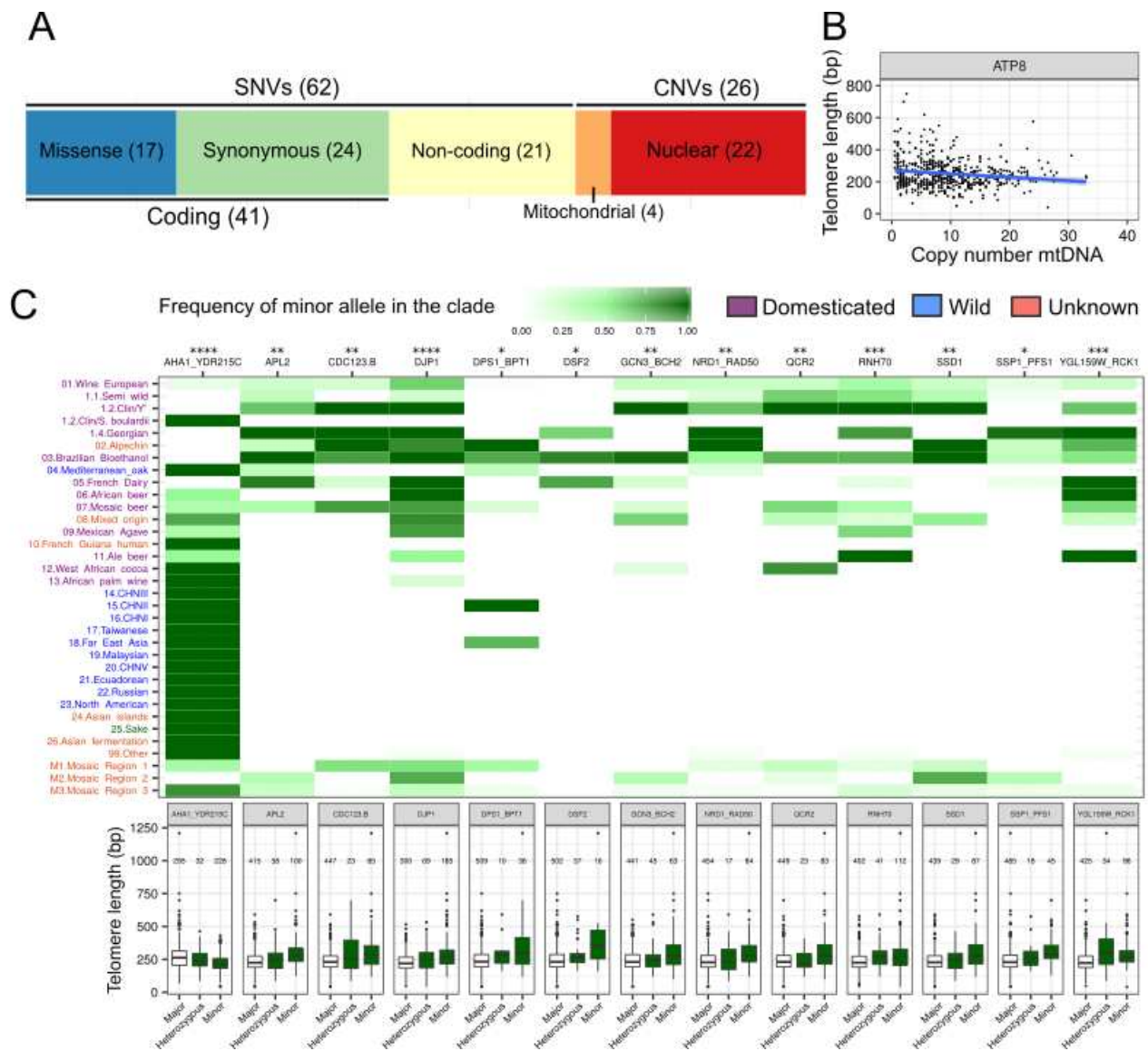


Figure 3 – **A**, Map representing the number of QTLs and QTNs associated to telomere length variation divided by functional group. SNVs: single nucleotide variants; CNVs: copy number variants. Numbers in brackets denote the number of QTLs/QTNs contained in that group. **B**, Example of mitochondrial CNV for the gene *ATP8*, which is associated to telomere length variation. Each point represents a single isolate and the line represents a linear regression function. Isolates with longer telomeres carry less copy numbers of *ATP8*. The other 3 mitochondrial CNVs show the same behaviour. **C**, The upper plot represents a heatmap of the minor allele frequency (MAF) of 13 QTNs, which show a significant enrichment in minor alleles in either domesticated or wild clades relative to the expected values. Expected values were estimated based on the absolute frequency of each minor allele in the whole collection and the proportion of domesticated/wild isolates. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. Squares represent the MAF in each clade, and the intensity of the green tone is directly proportional to the MAF in the clade. The order of clades on the y-axis is the same as in (Peter et al., 2018) and the clade classification is as in Fig. 1.

The name of the QTN is based on its host gene. QTNs with two gene names are found in intergenic regions and their name comprises the upstream and downstream genes. The bottom plot represents the telomere length of isolates divided by their genotype for the 13 QTNs. Major: homozygous for the major allele; heterozygous: heterozygous genotype; minor: homozygous for the minor allele. Numbers on top represent the number of isolates in each group. The isolates whose genotype was unknown are not displayed. QTN names are as in the upper plot. Box plots are as in Fig. 1.

Mitochondrial metabolism is associated with TL variation in wild isolates

We investigated the ecophysiological significance of the different TLs of the yeast collection by testing their association with 155 phenotypes related to growth in different carbon and nitrogen sources, in the presence of chemical compounds and environmental challenges, reproductive capacity and CLS under normal and stressful conditions. No significant correlation was found between TL and sporulation-related phenotypes, while a few associations emerged with growth in alternative carbon and nitrogen sources, and are indicated by * in (Fig. 4A) (Supplementary Table 6).

Notably, the copy number of mitochondrial DNA was negatively associated with TL, in accordance with the results of the GWAS (Figs. 3B and 4A). Moreover, there were positive associations of TL with mitochondrial volume and activity in both YPEG and YPD (Spearman's $R=0.16$ and $p=0.0003$ for mitochondrial volume in YPEG, $R=0.11$ and $p=0.03$ for mitochondrial activity in YPEG, $R=0.12$ and $p=0.02$ for mitochondrial volume in YPD, $R=0.11$ and $p=0.03$ for mitochondrial activity in YPD, FDR=5%) (Figure 4A). Since domesticated and wild yeast isolates have different TLs, we checked their associations with the phenotype dataset separately. The correlation coefficients of all mitochondrial phenotypes were higher when considering the wild isolates alone (Spearman's $R=0.23$ for mitochondrial volume in YPEG, $R=0.22$ for mitochondrial activity in YPEG, $R=0.22$ for mitochondrial volume in YPD, $R=0.34$ for mitochondrial activity in YPD), although p -values failed to reach statistical significance in this setting (Figs. 4A, B and Supplementary Table 6). For Pearson's correlation, the coefficients were higher and TL and mitochondrial metabolism co-vary in wild strains according to a linear model (Supplementary Table 6). When considering domesticated strains, we observed the opposite trend, with correlation coefficients close to 0, meaning that wild strains are the main drivers of the association between TL and mitochondria.

In line with these results, wild isolates had lower mitochondrial activity and volume in both YPD and YPEG as compared to domesticated ones (two-tailed Wilcoxon test, $p=5.584e^{-7}$ and $p=2.928e^{-9}$ for mitochondrial volume and activity in YPD, respectively; $p=1.057e^{-14}$ and $p<2.2e^{-16}$ for mitochondrial volume and activity in YPEG, respectively; Fig. 4C only YPD in Fig4C), suggesting that domestication induced a higher metabolic rate. As expected, the differences were more

pronounced in YPEG medium, where respiration is stimulated and mitochondria are required to grow. We then examined the mitochondrial metabolism of feral strains and found they had a lower mitochondrial volume than anthropic ones (two-tailed Wilcoxon test, $p=0.0008$ in YPD and $p=0.0008$ in YPEG). Mitochondrial activity was lower in feral strains than in anthropic ones but not significantly (two-tailed Wilcoxon test, $p=0.22$ in YPD and $p=0.30$ in YPEG; **Fig. 4D**). These results suggest that mitochondrial metabolism, like TL, can adapt to the reintroduction in a wild environment. Overall, these findings indicate that mitochondrial metabolism is markedly associated with TL in wild isolates while no such correlation is observed in domesticated isolates.

Interestingly, the wild strain TLs were correlated with chronological lifespan (CLS) (Spearman's $R=0.48$, $p=0.03$, $R=0.47$, $p=0.04$, for CLS at 7 and 21 days of incubation in SDC, respectively) (**Fig. 4A and Supplementary Table 6**).

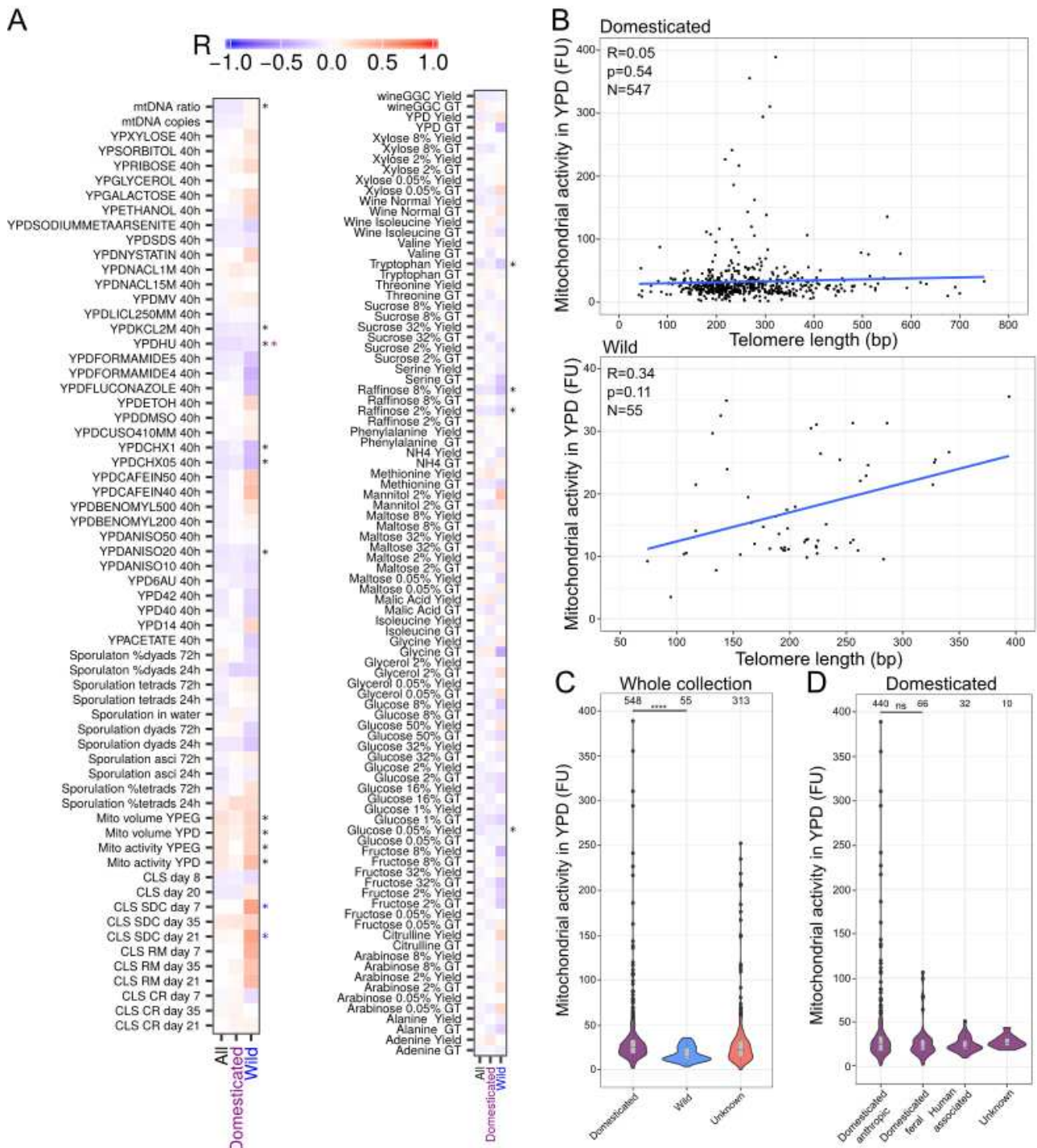


Figure 4 – **A**, Heatmap of the Spearman's correlation coefficient between telomere length and other 155 available phenotypes. Blue tones represent negative coefficients while red tones represent positive coefficients. Asterisks indicate phenotypes whose association with telomere length is statistically significant, and their colour represents the group for which the association is significant (all, domesticated or wild). Classifications and relative colours are as in Fig. 2A. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **B**, Example of association between mitochondrial activity in YPD and telomere length. The upper plot shows domesticated isolates ($n=547$) and the

bottom one shows wild ($n=55$) isolates. Each point represents a single isolate and the line represents a linear regression function. Three isolates with extreme values are not displayed in the upper plot. The association is significant when considering the whole collection ($n=918$), while it is not when considering domesticated or wild isolates separately. However, correlation coefficients are higher when considering the wild isolates, while they are close to 0 when considering domesticated isolates. In general, isolates with longer telomeres have a higher mitochondrial activity. The other 3 mitochondrial phenotypes (mitochondrial activity and volume in YPD and YPEG) show the same behaviour, while the copy number of mitochondrial DNA is negatively correlated with telomere length. **C**, Mitochondrial activity in YPD of domesticated vs wild isolates. Isolates classification as in Fig. 1. Box plots are as in Fig. 1. Numbers on top represent the number of isolates in each group. Two isolates with extreme values are not displayed. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. **D**, Mitochondrial activity in YPD of domesticated isolates, divided in groups as in Fig. 2B. Box plots are as in Fig. 1. Numbers on top represent the number of isolates in each group. Two isolates with extreme values are not displayed. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

Discussion

By estimating the telomeric DNA length (TL) of nearly 1000 *Saccharomyces cerevisiae* isolates collected worldwide and associated with an extensive genomic and phenotyping dataset (Peter et al., 2018), this work generates a unique corpus of data allowing studies of the role of telomeres in yeast evolution and fitness. To our knowledge, no similar integrated approach has been used to study telomere functions on such a high-resolution, ecological scale.

We observed a wide range of normally distributed TL variation (median of 236±96 bp) that is close to that previously measured in the classical S288C laboratory strain. This variation cannot be simply explained by ploidy and geographic characteristics and is associated with several genetic variants. Polygenic adaptation contributing to telomere length setting in natural populations is in accordance with a previous QTL analysis performed with a smaller yeast population (Liti et al., 2009). Both studies uncovered sequence variants in the two genes encoding the Ku heterodimer, a key regulator of telomere length and subtelomeric silencing (Laroche et al., 1998). Remarkably, TL was shorter in the wild strains than in the domesticated isolates. Moreover, rare loss of function mutations in 28 genes expected to shorten TL appear to be enriched in the wild strains but not in the domesticated ones.

These results suggest that possessing short telomeres in a wild habitat confers a selective advantage for yeast. Since a wild habitat forces the yeast to cope with both resource scarcity and fluctuating environments, one can postulate that short telomeres could help survival and fitness upon exposure

to various stresses. This might result from the transcriptional outcome of short telomeres (Ye et al., 2014). In yeast, the expression of subtelomeric genes can succumb to heterochromatinization initiated at telomeres by the recruitment and subsequent subtelomeric spreading of the Sir2-Sir3-Sir4 complex (Ottaviani et al., 2007). Thus, the expression of subtelomeric genes depends in part on TL, which modulates heterochromatin establishment and stabilization. Importantly, the transcriptomic signature in response to various stressors (e.g. heat and osmotic shock, starvation, membrane deformation, genotoxic and mitochondrial DNA loss) is enriched in subtelomeric genes (Ai et al., 2002; Smith et al., 2011; Stone & Pillus, 1996). This transcriptional response can contribute directly to adaptation to a wild habitat since several subtelomeric genes are involved in carbon source and nutrient utilization (Pryde & Louis, 1997), cell wall and stress responses (Ai et al., 2002; Robyr et al., 2002; Wyrick et al., 1999). For instance, subtelomeric derepression occurs in response to glucose starvation via a global decrease in histones level (Bruhn et al., 2020) and the derepression of subtelomeric genes encoding cell wall components contributes to adaptation of yeasts to the membrane deforming molecule chlorpromazine (Ai et al., 2002). In addition to modulating subtelomeric silencing, stress can lead to the release of transcription factors from telomeres allowing them to bind and regulate their targets genome-wide (Buck & Lieb, 2009; Maillet et al., 1996; Martin et al., 1999; Platt et al., 2013). Thus, the adaptation to stressful environments via reorganization of the chromosome-end chromatin might drive telomere shortening in wild isolates. In agreement with this hypothesis, several loss-of-function mutations found at high frequency in wild clades affected genes directly involved in subtelomeric silencing and gene expression such as Sir4, Hst1, the closest paralog of Sir2 (Kang et al., 2017), a set of genes involved in non nonsense-mediated mRNA decay (*NMD2*, *RPL13B* and *RPS16B*) (Lew et al., 1998) and the chromatin remodeler *SNF5*.

Among a wide range of phenotypic traits tested for TL association, mitochondria show the best correlations with TL: positive correlations with mitochondrial content and activity and negative ones with mitochondrial DNA copy number. These results suggest that the short telomeres in wild isolates are linked to mitochondrial homeostasis and redox metabolism. In fact, a wealth of diverse data in yeast and mammals indicate that telomeres and mitochondria are intimately interconnected. First, telomere changes can control mitochondrial integrity. In telomerase negative yeast cells, telomere shortening leads to an increased number of mitochondria (Nautiyal et al., 2002). In mammals, telomere shortening activates p53, which in turn represses *PGC1- α* and *- β* gene expression leading to mitochondrial dysfunction (Sahin et al., 2011). Most mammalian telomeric proteins have mitochondrial functions (Ahmed et al., 2008; Chen et al., 2013; Kim et al., 2017). Interestingly, subtelomeric silencing can be linked to mitochondrial activity since a loss of the telomeric protein TRF2 leads to the repression of the subtelomeric *SIRT3* gene leading to

mitochondrial dysfunction (Robin et al., 2020). Conversely, mitochondrial dysfunction can alter telomere structure (Qian et al., 2019). Due to its high guanine content, telomeric DNA is sensitive to the oxidative stress that results from mitochondrial respiration, leading to the accumulation of 8-oxo-guanine, which disrupts the binding of telomere protective factors (Opresko et al., 2005) and antioxidant proteins that are specifically associated with telomeres (Aeby et al., 2016), prevents extension by telomerase and leads to telomere shortening (Passos et al., 2007). Finally, we propose that a wild habitat imposes a trade-off between telomere and mitochondrial functions to cope with nutrient limitations and environment disturbances. On one hand, long telomeres could be associated to enhanced mitochondrial respiration and reactive oxygen species production. This might lead to a Sir3-dependent increase in chronological lifespan (CLS) through an enhanced subtelomeric silencing (Schroeder et al., 2012), thereby explaining the positive correlation we observed between TL and CLS. On the other hand, short telomeres could lead to mitochondrial dysfunction, and the high mitochondrial copy number associated with short telomeres might constitute a possible compensatory mechanism (Kotiadis et al., 2014).

By contrast, the TLs of domesticated isolates are longer than those of wild isolates. A simple explanation is that the absence of the constraints of a wild environment relieves the selective pressure to maintain short telomeres. This relaxation regarding telomere length regulation is also reflected by the heterogeneity of TL distribution in domesticated isolates, some measuring more than 1 kb and harboring more Y' subtelomeric elements interspersed with short stretches of telomeric sequences (ITS). The existence of natural yeasts carrying ultralong telomeres is in accordance with previous studies showing that artificially elongated telomeres do not affect yeast fitness (Harari et al., 2017; Harari & Kupiec, 2018). Y' amplification mediated by homologous recombination is observed as a response to telomerase ablation in laboratory strains. Therefore, it is tempting to speculate that an absence of selection on telomere length in domestic strains may allow the transient loss of telomerase expression giving rise to recombination relaxation in the subtelomeres. The strains that are classified as feral, i.e. considered as having returned to the wild after a period of domestication, have shorter TLs like the wild isolates, substantiating the notion that short telomeres are rapidly selected in yeasts living in wild environments. The limited number of feral strains precluded an optimal analysis to decipher whether this return to short telomeres is driven by genetic selection.

Overall, our analyses revealed links between telomere length and mitochondrial functions in a large population of yeast isolates collected around the world. This association is observed in wild but not in domesticated isolates uncovering the key role of lifestyle in shaping telomere-mitochondria relationships. These results give an evolutionary meaning to the plethora of mechanisms connecting telomeres and mitochondria that were described in laboratory strains and whose ecophysiological

significance is still elusive. They also give credence to the notion that short telomeres can be beneficial. If telomere shortening in response to stress generally works through its adverse effects on premature aging, it is a biological process widely conserved throughout evolution and consequently expected to have a key role in fitness. This was evoked in human genetics to explain the shorter telomeres in European populations versus African ones as conferring a better resistance to melanoma (Hansen et al., 2016). Our results provide evidence that short telomeres are selected during evolution as a trade-off among genome maintenance, mitochondrial metabolism and stress transcriptional responses to increase fitness in habitats with limited resource and fluctuating environment.

Supplementary discussion

1) Bioinformatic estimation of telomeric DNA content in *Saccharomyces cerevisiae*

An estimate of telomeric DNA length present at the ends of chromosomes (abbreviated TL) can be inferred from NGS data by counting the number of reads containing a sufficiently high number of telomeric DNA repeated sequences. This approach was implemented for human genomes containing regular telomeric DNA repeats ((T2AG3)_n) (Ding et al., 2014; Hakobyan et al., 2016; Lee et al., 2017; Nersisyan & Arakelyan, 2015). However, it does not account for the presence of interstitial telomeric sequences (ITS), which might confound the estimation of telomere length. Moreover, some organisms, such as *Saccharomyces cerevisiae*, use irregular telomeric DNA repeated sequences ((TG1-3)_n), making the use of human genomes-oriented pipelines impossible. In order to exploit the huge amount of NGS data available for the budding yeast to perform telomere biology studies, we developed an original pipeline to estimate TL and ITS content from whole genome sequencing (see **methods**).

First, we examined different criteria to distinguish telomeric/ITS reads from the others, and we started by deriving the frequency distribution of stretches of telomeric repeats of different length in the complete genome assemblies of 7 *S. cerevisiae* and 5 *S. paradoxus* strains already published (Yue et al., 2017). In all the cases, the distribution was skewed towards small values, with an abundance of very short telomeric repeat stretches, mostly ranging between 2 and 20 bp of length. There was a sudden drop in frequency for stretches longer than 20 bp (**Extended Data Fig. 1A**). Very long stretches were exclusively present in subtelomeres (ITS) or telomeres. Based on this observation, we tested gradually increasing thresholds, ranging from 20 to 50 bp of telomeric stretch, and checked which one gave the best performance. We simultaneously replaced all the native telomeres of the 12 genome assemblies with 14 synthetic telomeres of known length, ranging from 18 to 651 bp, building a total of 168 synthetic genomes. The use of 12 genomes with variable ITS content allowed us to test how much the genomic background and the difference in ITS content

impact the TL measurement. We simulated Illumina reads from these genomes and applied our pipeline to these datasets. The estimates of ITS content and TL correlated positively with the real values across the range of parameters tested, with the best correlation at the threshold of 40 bp ($R=0.9638$, $p<2.2e-16$ for ITS content, $R=0.9821$, $p<2.2e-16$ for TL) (**Extended Data Fig. 1B and Supplementary Table 9**).

We further validated the efficiency of our pipeline in the 12 strains by using already published Pacbio assemblies, Sanger sequencing data derived from the *Saccharomyces* genome resequencing project (SGRP) and the gold-standard technique of Southern blotting (Teloblot) (Bergstrom et al., 2014; Liti et al., 2009; Yue et al., 2017). First, we used custom Perl scripts to identify Sanger reads containing full-length telomeres and estimated TL by averaging the TLs of those reads. In parallel, we estimated TL by applying our pipeline to Illumina sequencing data derived from the same strains. We correlated results and found significant positive correlations for ITS content and TL (Pearson's $R=0.87$ and $P=0.0002$, $R=0.64$ and $P=0.02$, respectively) (**Extended Data Fig. 1C, D and Supplementary Table 10**). In addition, we measured TL of 5 *S. cerevisiae* strains (DBVPG6044, DBVPG6765, Y12, YPS128, UWOPS034614) and 2 *S. paradoxus* strains (CBS432 and N44) by Teloblot and confirmed a positive correlation between all the measurements (Pearson's $R=0.99$ and $P=0.0001$ between the Sanger sequencing and the Teloblot, $R=0.79$ and $P=0.05$ between the Teloblot and the pipeline) (**Extended Data Fig. 2A and Supplementary Table 10**). One strain, N44, did not work due to the complete absence of Y' elements in its genome and was not included in the statistical analysis.

We additionally measured the performance of the pipeline on the 918 *S. cerevisiae* collection. We chose one representative strain for 17 clades and measured TL by Teloblot. We correlated these results with the pipeline measurements and found an overall positive correlation (Pearson's $R=0.87$, $P=2.497e-05$) (**Extended Data Fig. 2B and Supplementary Table 10**). Two strains (25.Sake and 08.Mixed origin) did not give any terminal restriction fragment (TRF) and were not included in the statistical analysis. One strain from the Clinical/Y' amplification subclade had extremely long telomeres, in accordance with previous reports showing that this strain has atypical TL and amplification of ITS and Y' elements (Bergstrom et al., 2014). Taken together, these results show that the TL and ITS estimations of our pipeline are reliable and can be used for further analyses.

The recent publication of the complete genome sequencing of the *Saccharomyces cerevisiae* knock-out collection and its use to identify genes involved in the maintenance of genome stability, including TLM ones, offered the opportunity to test our pipeline against an already validated dataset (Puddu et al., 2020). Puddu *et al.* applied a bioinformatic approach which is similar to the one used in our study, and identified, among others, 12 new TLM genes validated by Teloblot. We applied our pipeline to the same dataset, adding knock-outs for known telomere-associated proteins, like

RIF1, *RIF2*, *TEL1* and the subunits of the telomerase complex (*EST1*, *EST2*, *EST3*). In all the cases our pipeline detected a significant difference in TL respect to the wild-type samples and the results were in line with what previously described in Puddu *et al.* (Pearson's $R=0.87$, $p=5.168e^{-16}$), confirming the reliability of our approach (**Extended Data Fig. 2C, D and Supplementary Table 8**). The method used by Puddu *et al.* is efficient in detecting TL changes in strains sharing the same genetic background and basal ITS content. However, other collections, as the 918 yeasts one, are constituted by strains derived from very different ecological niches and might vary in their ITS content, possibly leading to a bias in TL estimation. In this setting, our additional module classifying ITS and telomeric reads allows to greatly improve results and to distinguish between elongated/shortened telomeres or type I survivors carrying ITS/Y' amplifications. In the case of *EST* genes, we confirmed that their long TL derives from amplification of ITS (**Extended Data Fig. 2B**). Our pipeline is freely available through github and constitutes a valid tool to estimate TL and ITS content in a fast and cost-effective way.

2) Telomere length and ITS content are closely associated

We measured ITS content across the 918 yeasts collection and found it to show a hyperbolic distribution, with the majority of the strains containing less than 500 bp of ITS per cell. The overall distribution ranged from 0 to 92890 bp per cell (median=1295+-5676 bp) (**Extended Data Fig. 4A and Supplementary Table 1**). We subsequently estimated ITS content in the phylogenetic clades and found it to vary significantly among them (two-tailed Kruskal-Wallis test, $p<2.2e^{-16}$). It was highest in Mexican agave and Brazilian bioethanol, where the median exceeded 6000 bp per cell, while it was lowest in Mediterranean oak, Ecuadorean and CHNV clades, where ITS were almost absent (**Extended Data Fig. 4B and Supplementary Table 7**). ITS distribution in the Wine/European clade was similar to the whole collection and ITS content varied among the Wine/European subclades (two-tailed Kruskal-Wallis test, $p=8.95e^{-14}$). It followed the same trend of TL, being highest in the Clinical/Y' amplification subclade where the median exceeded 50000 bp, and lowest in the Clinical/*S. boulardii* subclade, where it remained below 200 bp (**Extended Data Fig. 4C, D and Supplementary Table 7**).

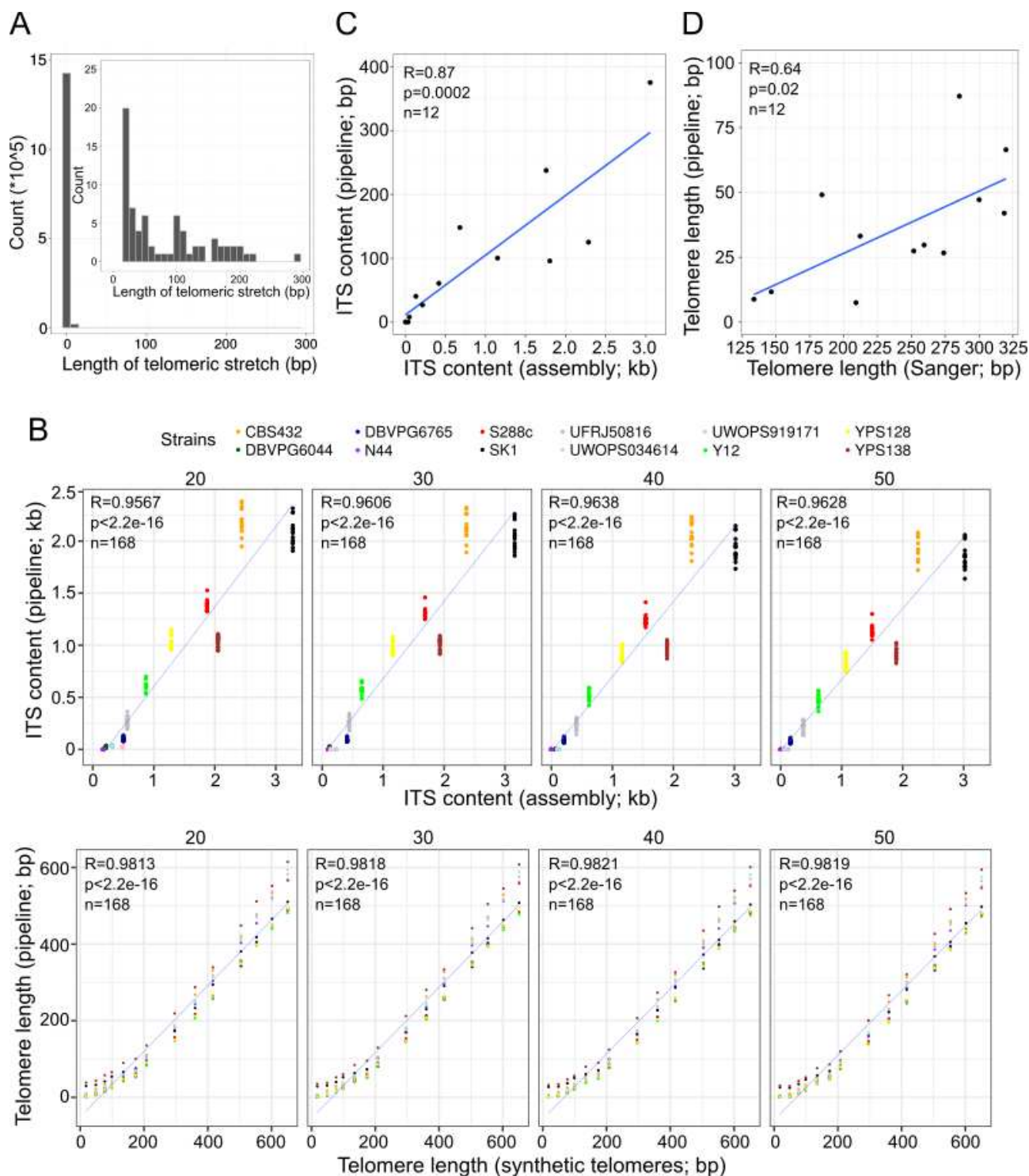
The similar behaviour of TL and ITS content suggests that they might be closely associated. We measured the correlation coefficient between the two variables and found a significant positive correlation (Pearson's $R=0.43$, $p<2.2e^{-16}$) (**Extended Data Fig. 5A**). In yeast, subtelomeric ITS are usually accompanied by Y' elements. We estimated the copy number of Y' elements in the collection and confirmed a significant positive correlation with ITS content (Pearson's $R=0.81$, $p<2.2e^{-16}$) (**Extended Data Fig. 5B and Supplementary Table 1**). By consequence, TL was also positively correlated with the copy number of Y' elements (Pearson's $R=0.39$, $p<2.2e^{-16}$) (**Extended Data Fig.**

5C). This result suggests that either ITS might be a *cis*-acting factor influencing the homeostasis of TL or viceversa, or that both TL and ITS content are under the control of the same determinants. To test these hypotheses, we measured ITS content in knock-outs strains of known TLM genes (Puddu et al., 2020) and detected a significant difference in ITS content respect to the wild-type samples. In most of the cases, the direction of the variation in ITS content was the same as for TL, although some opposite cases were also observed, confirming that ITS and telomeres are positively correlated ($R=0.63$, $p=1.156e^{-6}$) and are under the control of the same genetic determinants (**Extended Data Fig. 5D and Supplementary Table 8**).

We then checked how ITS behaved in domesticated and wild yeasts. ITS content was lower in wild yeasts than in domesticated ones (two-tailed Wilcoxon test, $p=1.348e^{-5}$) (**Extended Data Fig. 5E**). This finding revealed a possible role of domestication in reshaping not only telomeres', but also ITS' dynamics. Among the wild clades, the Taiwanese, Russian and North-American behaved like domesticated ones, having higher ITS content (**Extended Data Fig. 4B**). We then investigated whether domestication played a role in the association between TL and ITS content. However, TL and ITS content were always correlated to each other even if we considered only domesticated (Pearson's $R=0.48$, $p<2.2e^{-16}$) or only wild isolates (Pearson's $R=0.36$, $p=0.007$), meaning that this association is mediated by other factors.

The domesticated isolates had wider variance than the wild ones and reached both the lower and upper extremes of the distribution. However, no variation in ITS content was detected between domesticated-anthropic and domesticated-feral strains (two-tailed Wilcoxon test, $p=0.49$) (**Extended Data Fig. 5F**), suggesting that, despite the reintroduction in a wild environment has an immediate effect on telomeres, the effects on ITS need a longer time to manifest.

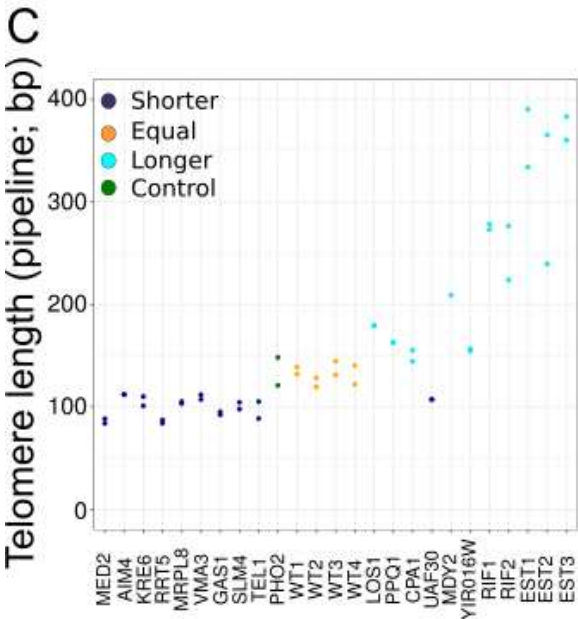
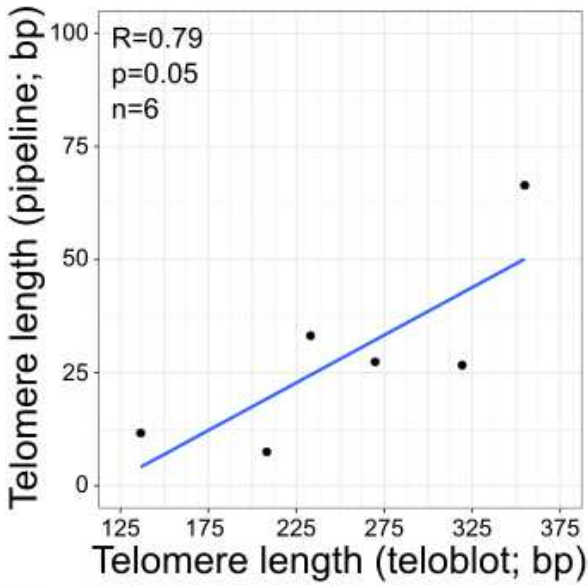
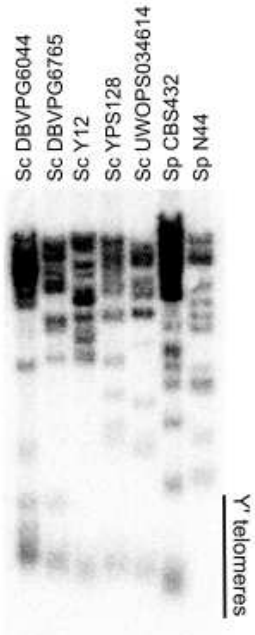
In conclusion, we show that telomeres and ITS biology are closely associated and interdependent, and that telomeres and ITS are under the control of the same genetic factors.



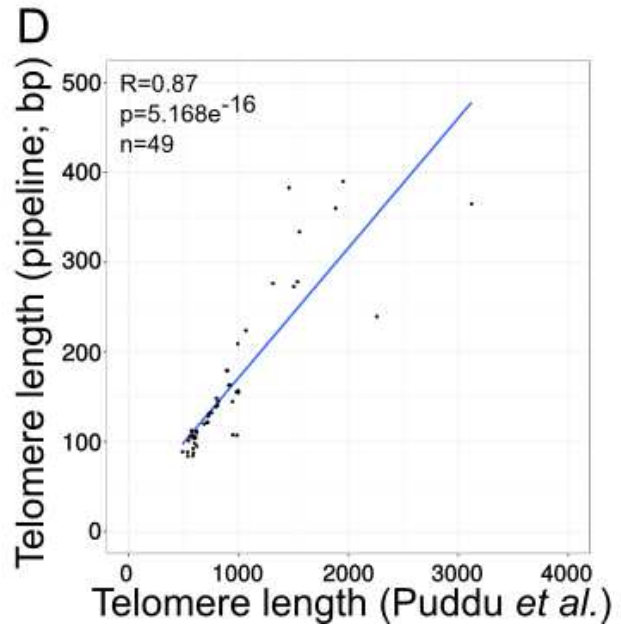
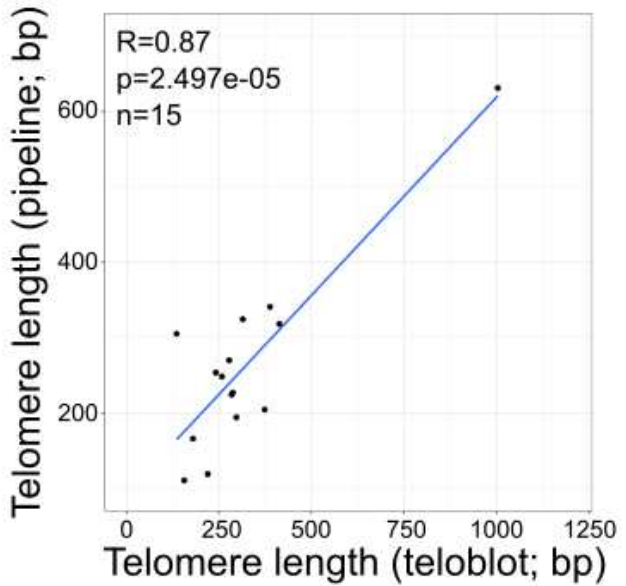
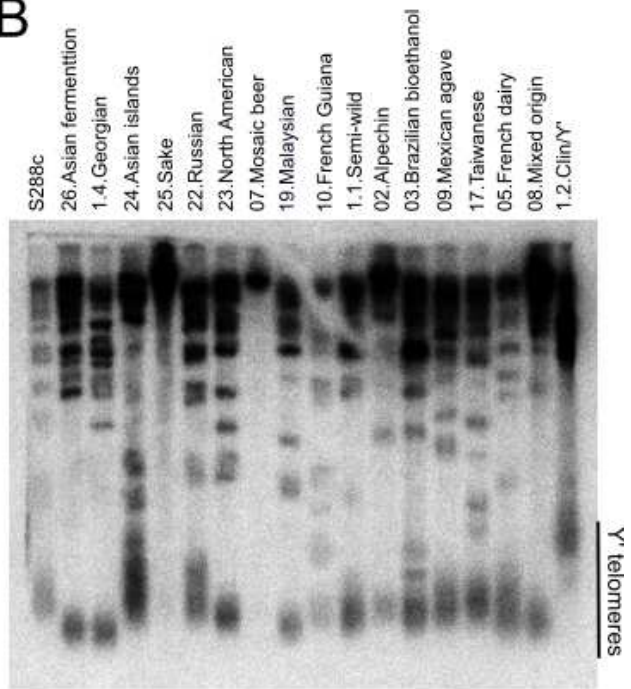
Extended Data Figure 1 – **A**, Frequency distribution of stretches of telomeric-like repeats in the genome assembly of the *S. paradoxus* European strain CBS432. Bin width is 10 bp. The inset shows a magnified view of the y axis up to 25 counts. Bin width is 10 bp. The first bin (0-10 bp) is not shown in the inset. **B**, The upper panels show ITS content in 168 simulated sequencing runs based on 12 already published *S. cerevisiae* and *S. paradoxus* genome assemblies, modified to carry synthetic telomeres of increasing length (not shown). ITS content is measured by the pipeline (y -axis) and estimated directly from the genome assemblies (x -axis). Colours denote the 12 assemblies.

Each point represents a single simulated NGS run and the line represents a linear regression function. Side-by-side panels represent correlations obtained using increasing thresholds of telomeric repeat stretches (20 to 50 bp) to detect telomeric reads. The bottom panels show telomere length in the same 168 simulated sequencing runs, estimated by the pipeline (y-axis) and compared to the known length of the synthetic telomeres (x-axis). Colours are as in the upper panels. **C**, ITS content in the 12 genome assemblies, measured by the pipeline (y-axis) and estimated by the genome assemblies (x-axis). Each point represents a single isolate and the line represents a linear regression function. **D**, Telomere length in the 12 genome assemblies, measured by the pipeline (y-axis) and estimated from Sanger sequencing reads (x-axis). Each point represents a single isolate and the line represents a linear regression function.

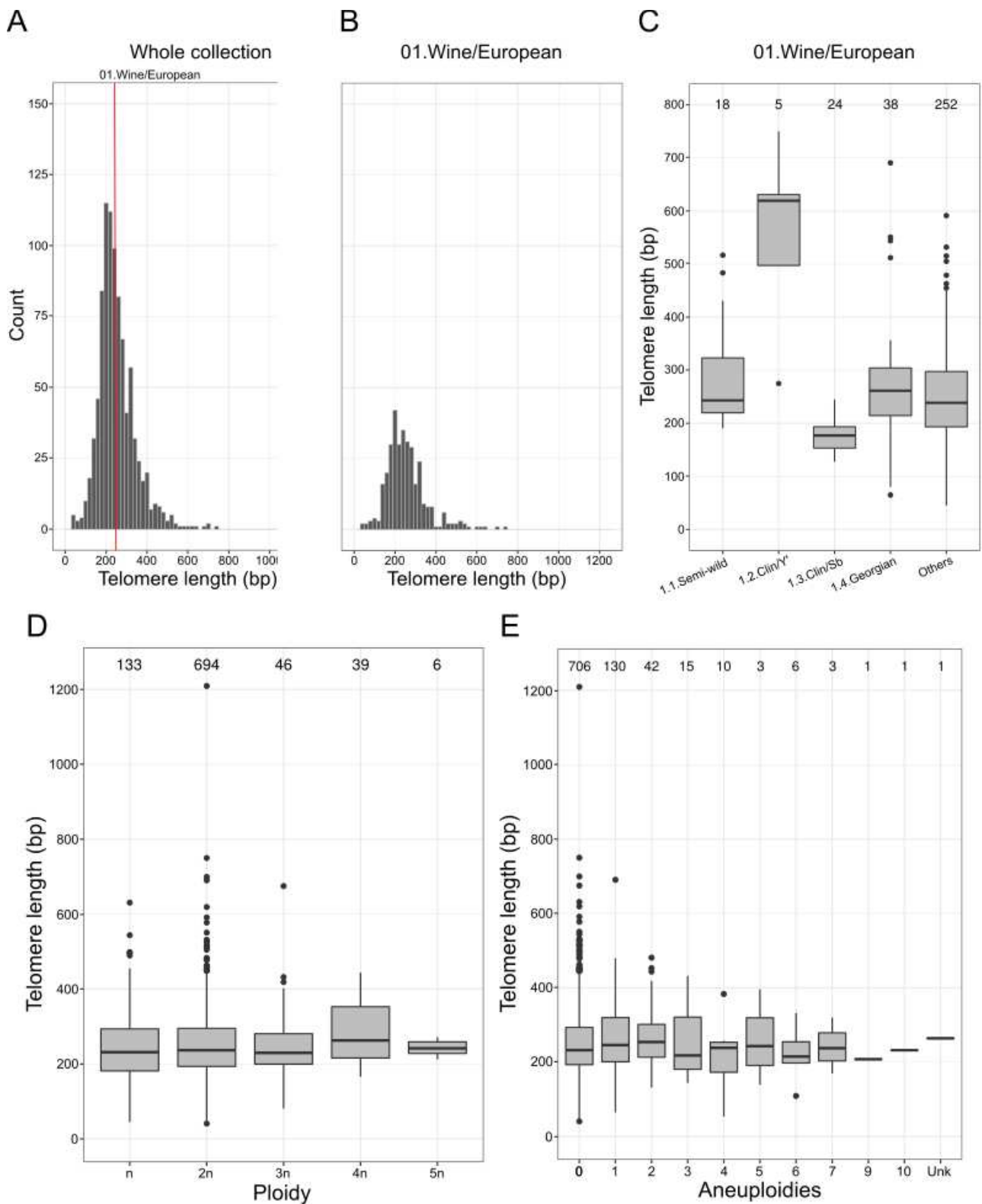
A



B

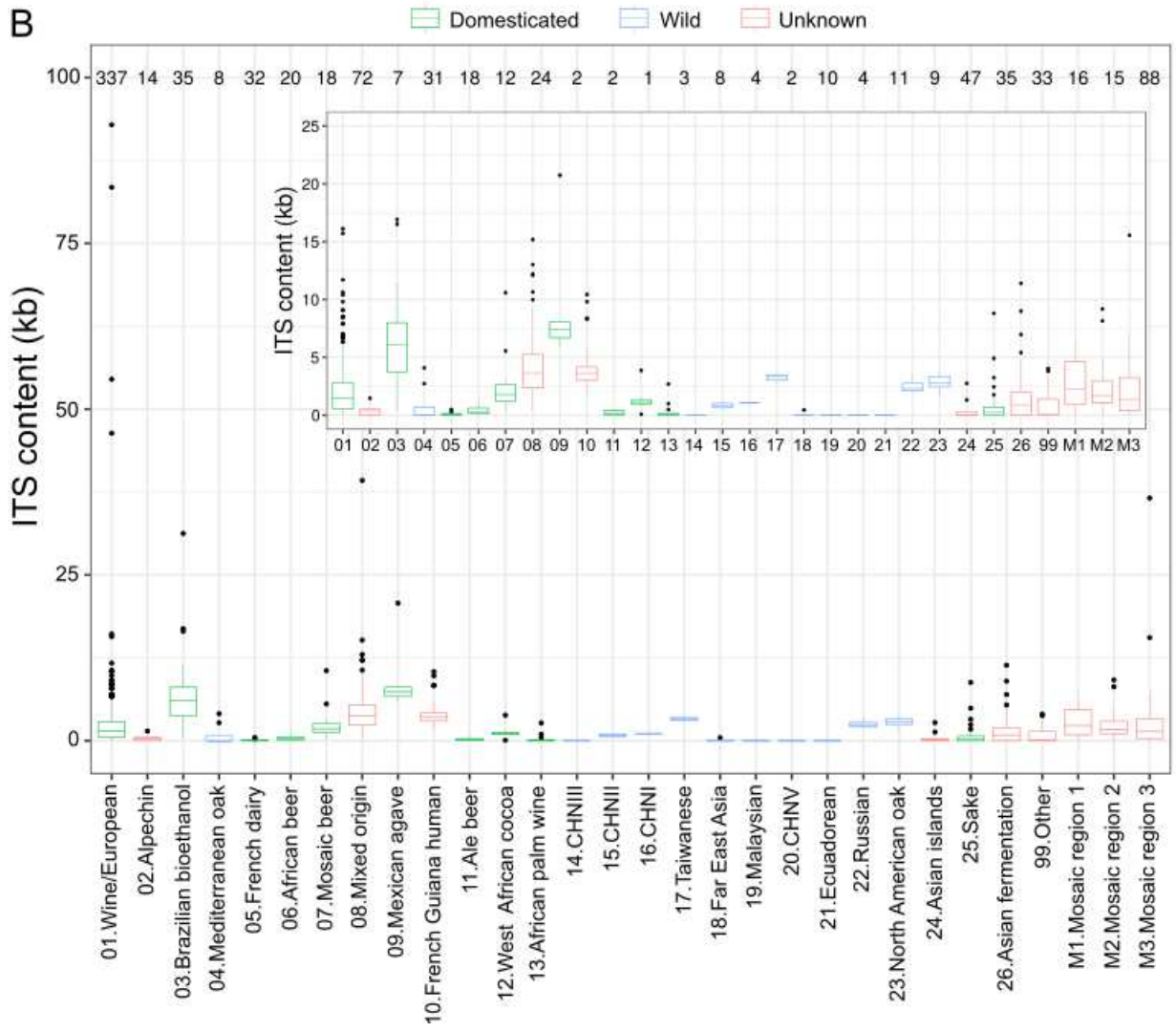
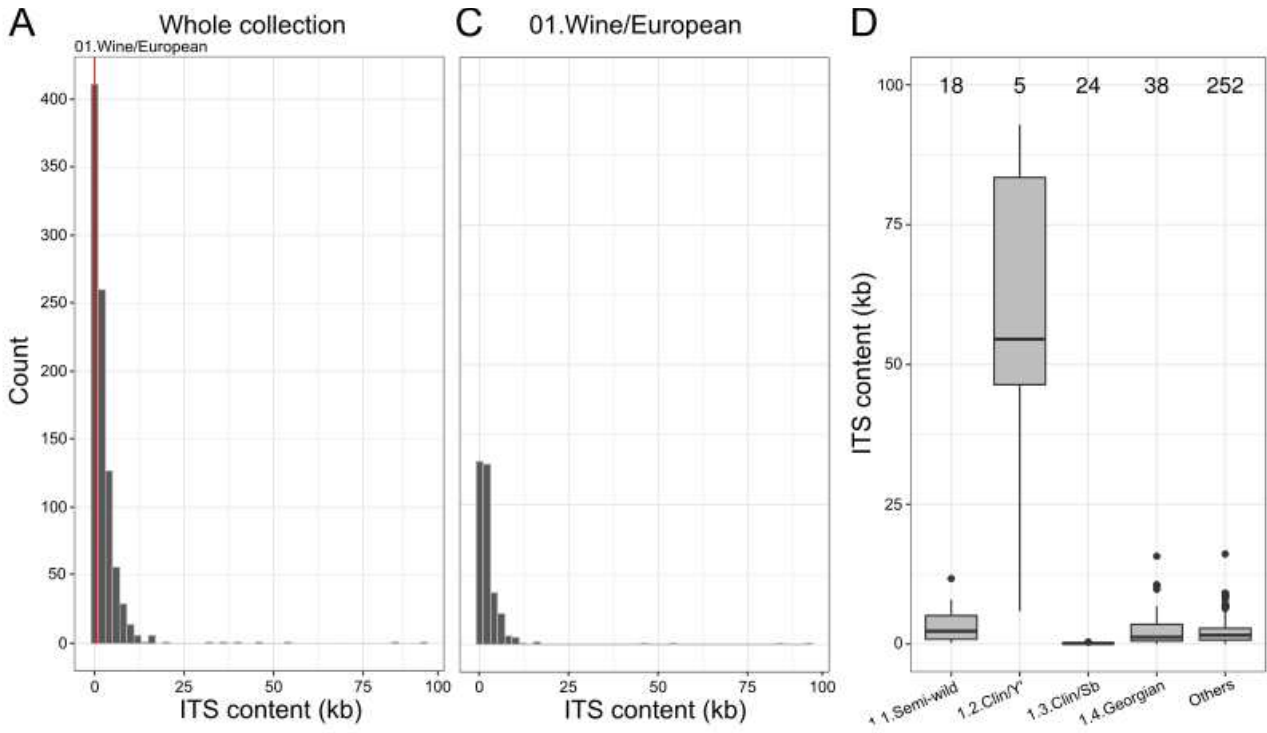


Extended Data Figure 2 – A, The upper panel shows a telomeric restriction fragment (TRF) blot for a subset of the 12 strains of Extended Data Fig. 1D. The black line denotes TRFs resulting from the digestion of a Y' element. No TRF was detected for the strain N44 because it does not carry any Y' elements. The bottom panel shows telomere length in the same subset of strains, estimated by the pipeline (y-axis) and compared to the value resulting from the blot (x-axis). Each point represents a single isolate and the line represents a linear regression function. **B**, The upper panel shows a TRF blot for a subset of 17 representative strains of the major clades of the 918 yeasts collection. The black line denotes TRFs resulting from the digestion of a Y' element. No TRF was detected for the representatives of the Sake and Mosaic beer clades, as they carry a limited amount of Y' elements. The bottom panel shows telomere length in the same set of strains, estimated by the pipeline (y-axis) and compared to the value resulting from the blot (x-axis). Each point represents a single isolate and the line represents a linear regression function. **C**, Telomere length of knock-out strains for known telomere length maintenance genes described in (Puddu et al., 2020) (x-axis), estimated by the pipeline (y-axis). Colours denote either wild-type strains (yellow) or knock-outs for genes which shorten (dark blue) or lengthen (light blue) telomeres. One gene is not involved in telomere maintenance and is shown as control (green). **D**, Telomere length in the same strains of panel C, measured by the pipeline (y-axis) and compared to the estimations of Puddu et al. 2019 (x-axis). The line represents a linear regression function.

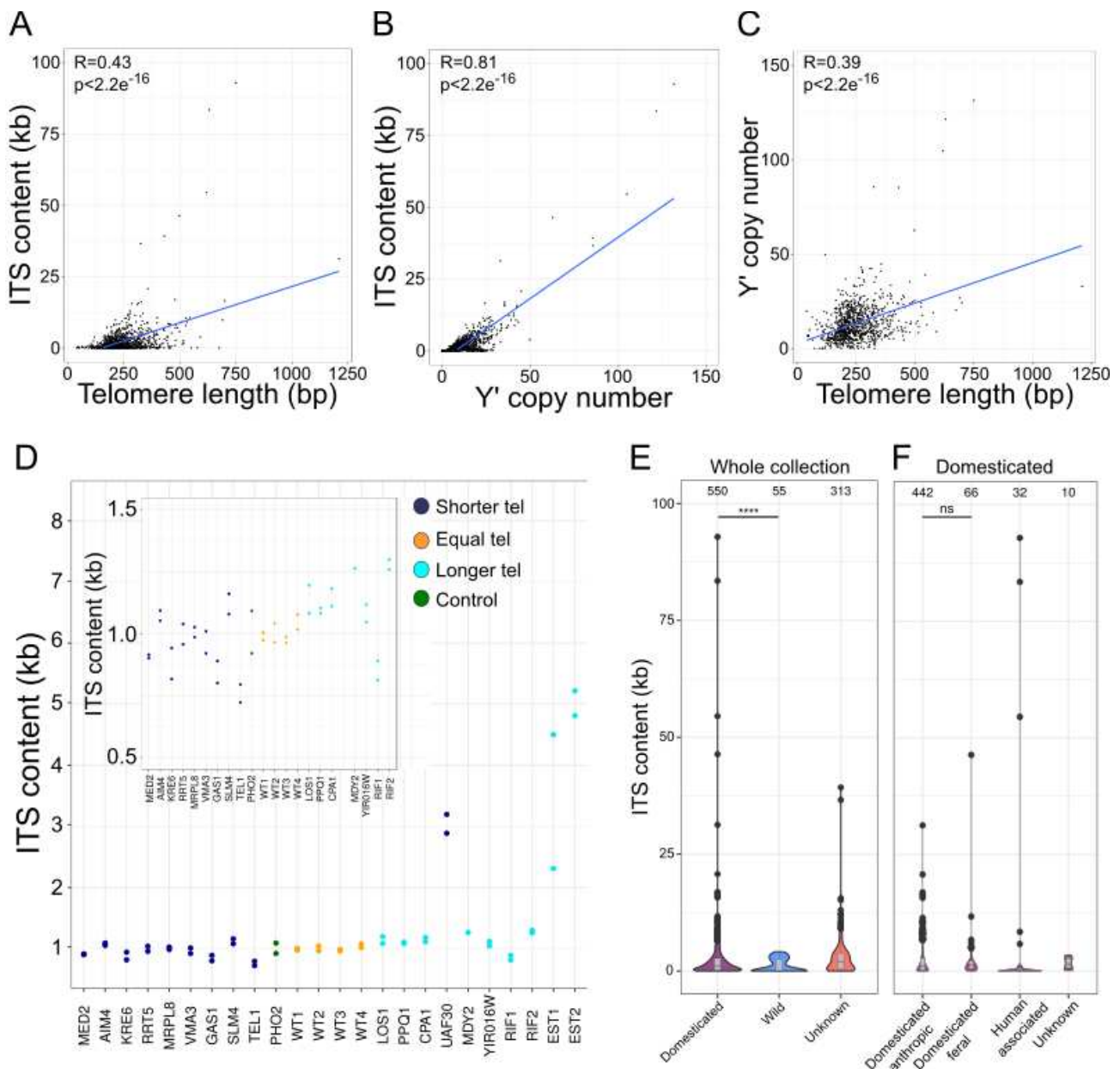


Extended Data Figure 3 – **A**, Frequency distribution of telomere length in the 918 yeasts collection. Bin width is 20 bp. The red line indicates the median of the Wine/european clade. **B**, Frequency distribution of telomere length in the isolates of the Wine/European clade. Bin width is 20 bp. **C**, Telomere length of the subclades of the Wine/European lineage, described in the 1002 Yeast Genome Project (Peter et al., 2018). Boxplots as in Fig. 1. Numbers on top represent the

number of isolates in each clade. **D**, Telomere length of isolates grouped by their ploidy. Box plots are as in Fig. 1. Numbers on top represent the number of isolates in each group. **E**, Telomere length of isolates grouped by their number of aneuploidies. Box plots as in Fig. 1. Numbers on top represent the number of isolates in each group.

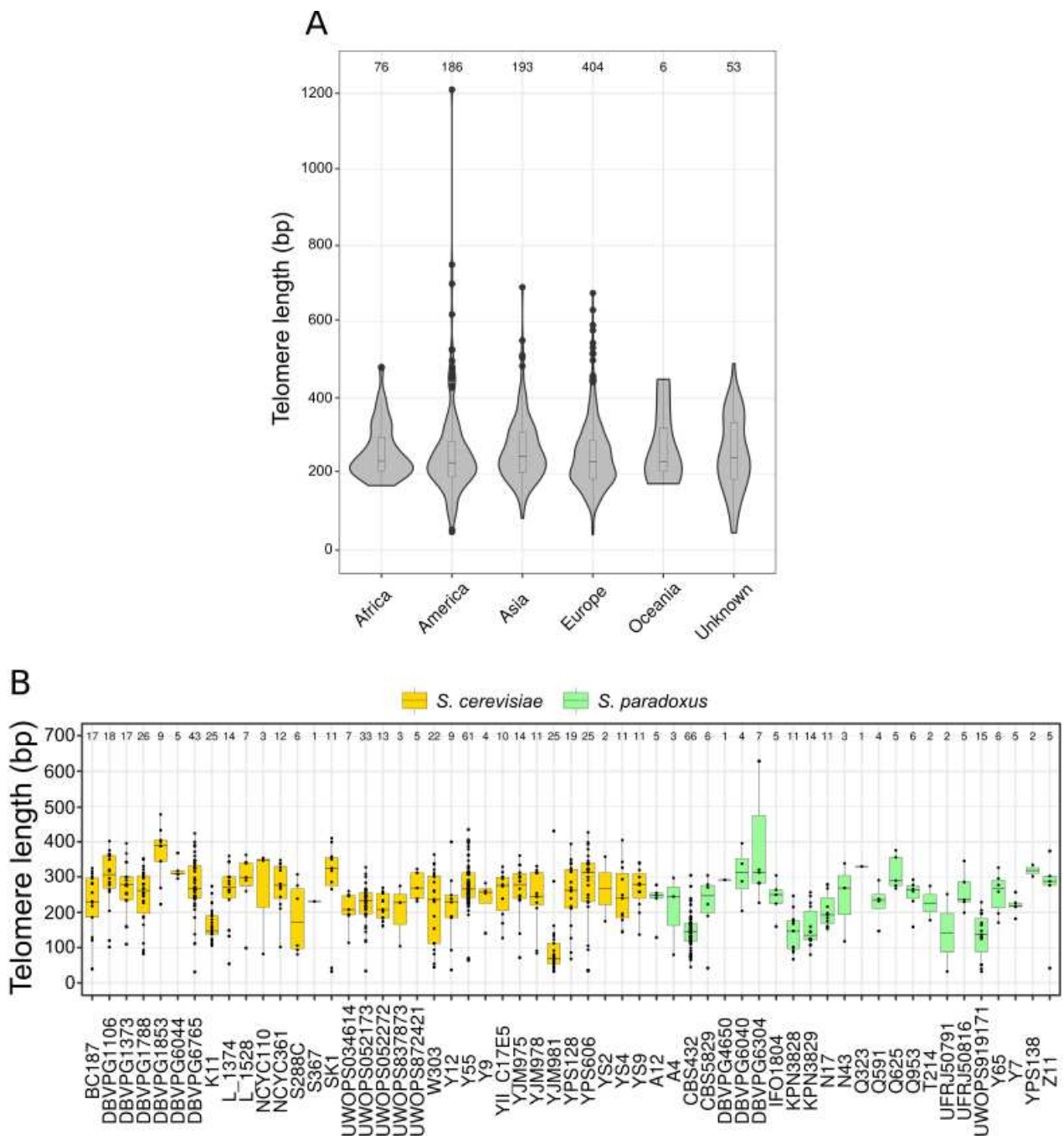


Extended Data Fig. 4 – A, Frequency distribution of ITS content in the 918 yeasts collection. Bin width is 2 kb. The red line indicates the median of the Wine/european clade. **B**, ITS content of the phylogenetic lineages described in the 1002 Yeast Genome Project (Peter et al., 2018). Colours represent the clade classification (domesticated, wild or unknown), which is the same as in Fig. 1. The order of clades on the x-axis is the same as in Fig. 1. Boxplots are as in Fig. 1. Numbers on top represent the number of isolates in each clade. The inset shows a magnified view of the y-axis up to 25 kb. **C**, Frequency distribution of ITS content in the isolates of the Wine/European clade. Bin width is 2 kb. **D**, ITS content of the subclades of the Wine/European lineage, described in the 1002 Yeast Genome Project (Peter et al., 2018). Boxplots are as in Fig. 1. Numbers on top represent the number of isolates in each clade.

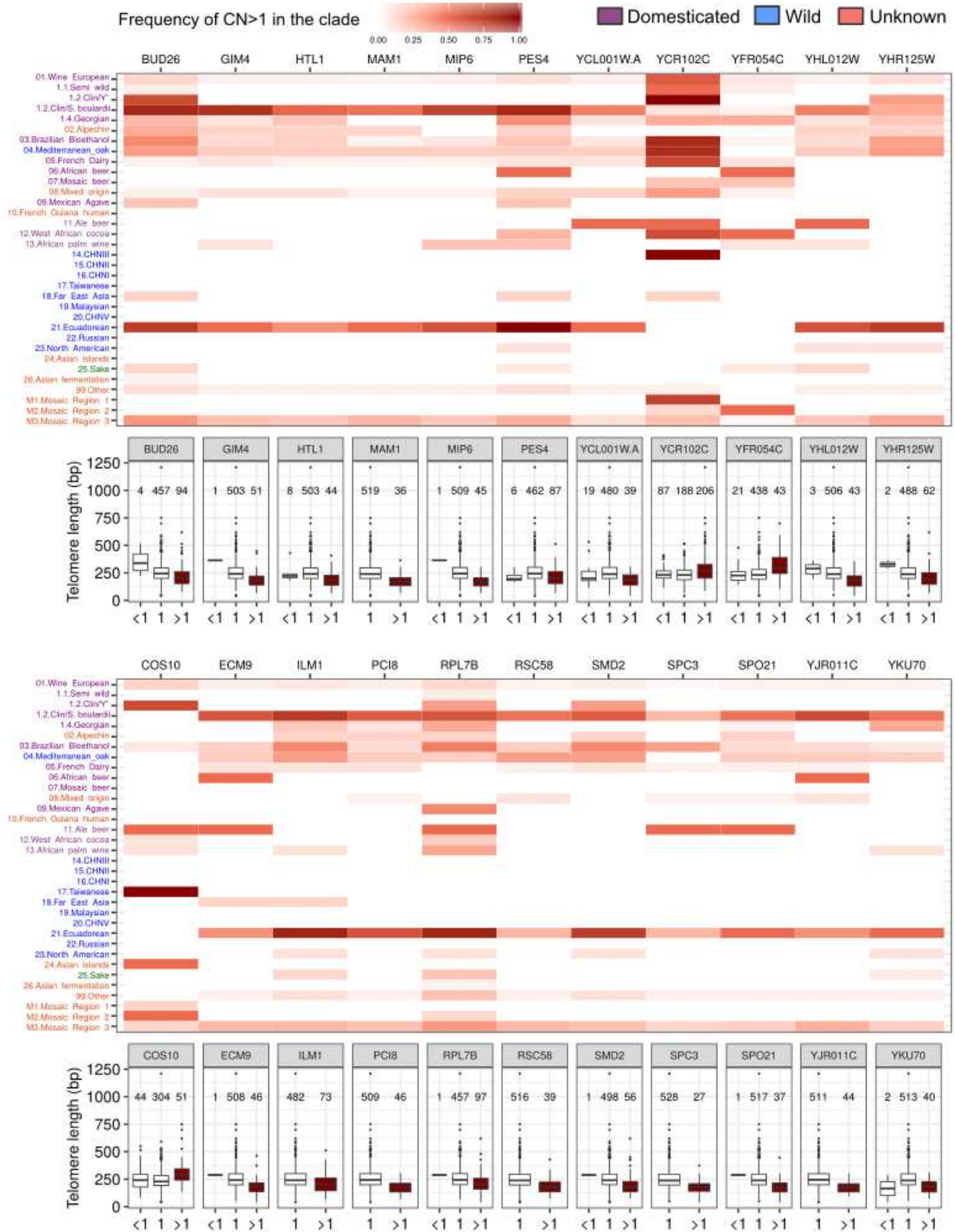


Extended Data Fig. 5 – **A**, Correlation between telomere length and ITS content in the 918 yeasts collection. Each point represents a single isolate and the line represents a linear regression function. **B**, Correlation between ITS content and Y' copy number in the 918 yeasts collection. Each point represents a single isolate and the line represents a linear regression function. **C**, Correlation between telomere length and Y' copy number in the 918 yeasts collection. Each point represents a single isolate and the line represents a linear regression function. **D**, ITS content of knock-out strains for known telomere length maintenance genes described in (Puddu et al., 2020) (x-axis), estimated by the pipeline (y-axis). Colours are as in Extended Data Fig. 2C. The inset shows a magnified view of the y-axis up to 1.5 kb. The knock-outs for *UAF30*, *EST1*, *EST2* and *EST3* are not shown in the inset. **E**, ITS content of domesticated vs wild isolates. Isolates classification as in Fig. 1. Box plots are as in Fig. 1. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$. **F**, ITS

content of domesticated isolates, divided in groups as in Fig. 2B. Box plots are as in Fig. 1.
* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.



Extended Data Fig. 6 – A, Telomere length of isolates grouped by their continent of isolation. Box plots are as in Fig. 1. Numbers on top represent the number of isolates in each continent. **B**, Telomere length of single isolates of the *Saccharomyces* genome resequencing project (Liti, Carter, et al., 2009). Box plots are as in Fig. 1. Numbers on top represent the number of telomeric reads used to estimate the telomere length in each isolate. Colours represent *S. cerevisiae* and *S. paradoxus* isolates.



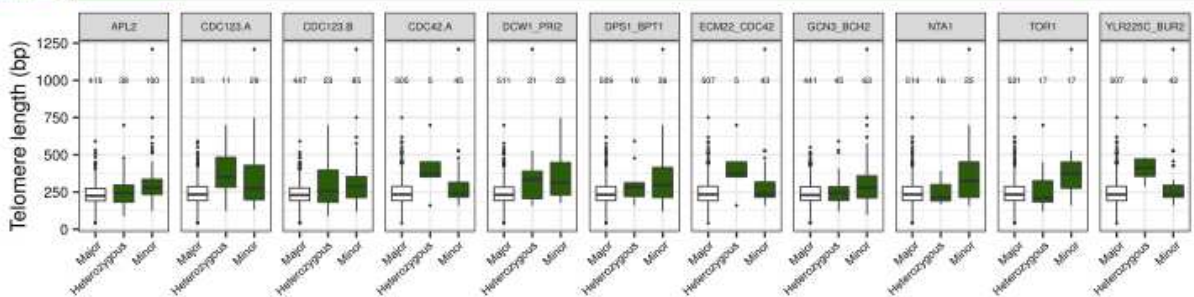
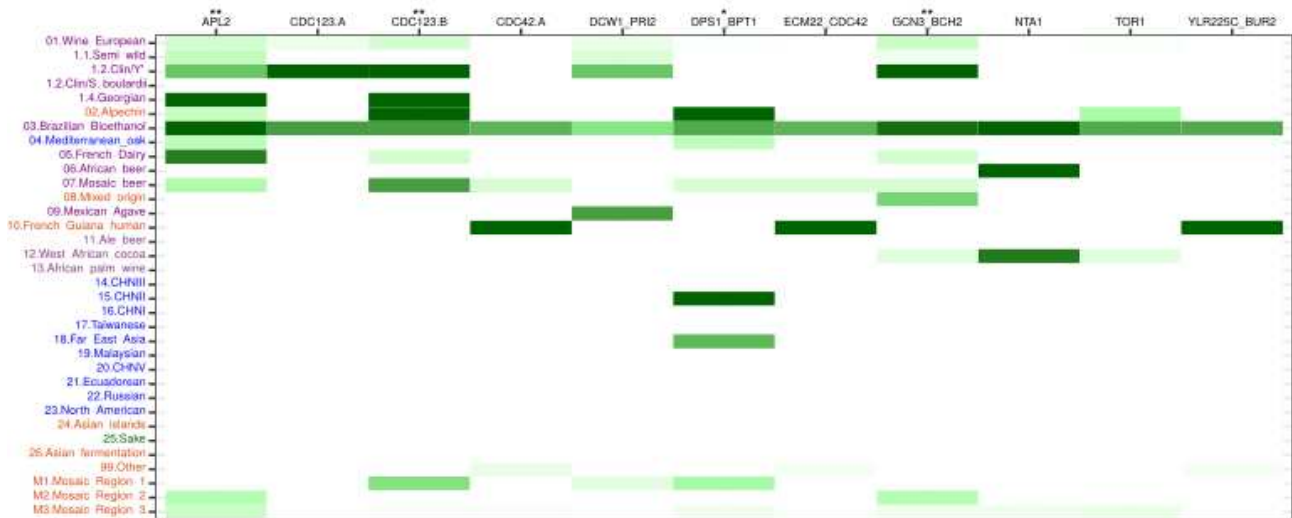
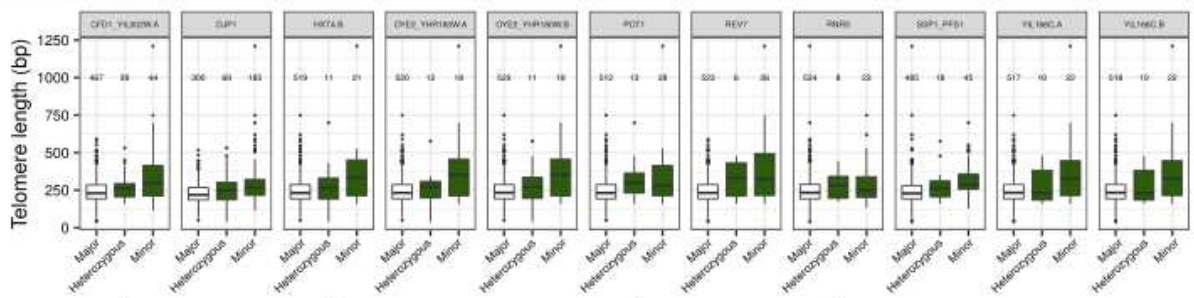
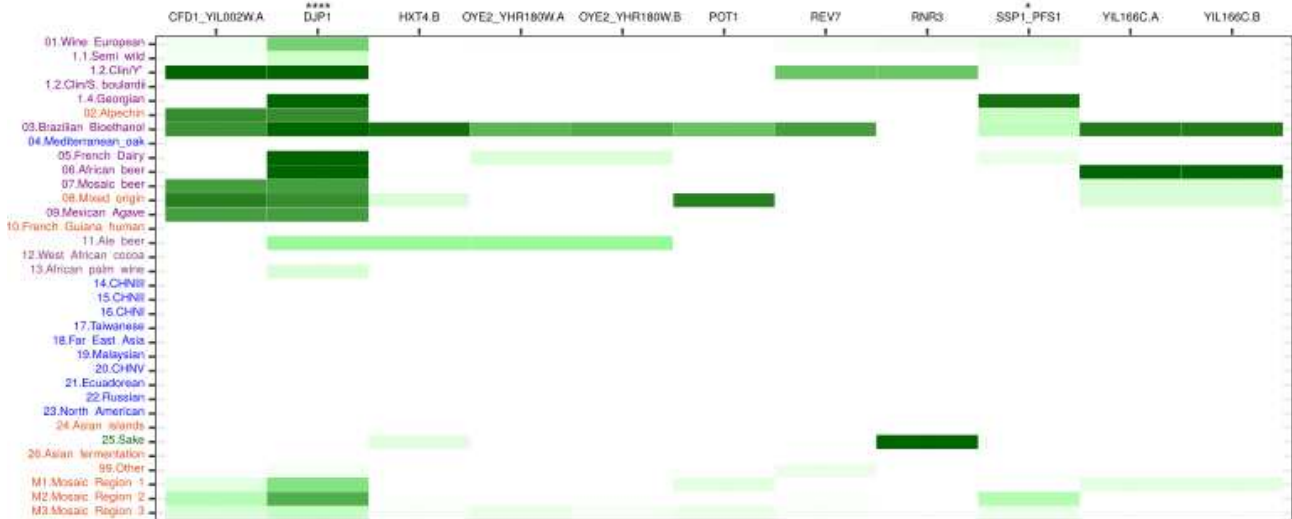
Extended Data Fig. 7 – The upper plot represents a heatmap of the gene copy number (CN) of the 22 QTLs for CNV. Squares represent the fraction of isolates carrying more than 1 copy of the gene in each clade, and the intensity of the red tone is directly proportional to the fraction in the clade. The order of clades on the y-axis is the same as in Fig. 3C and the clade classification is as in Fig. 1.

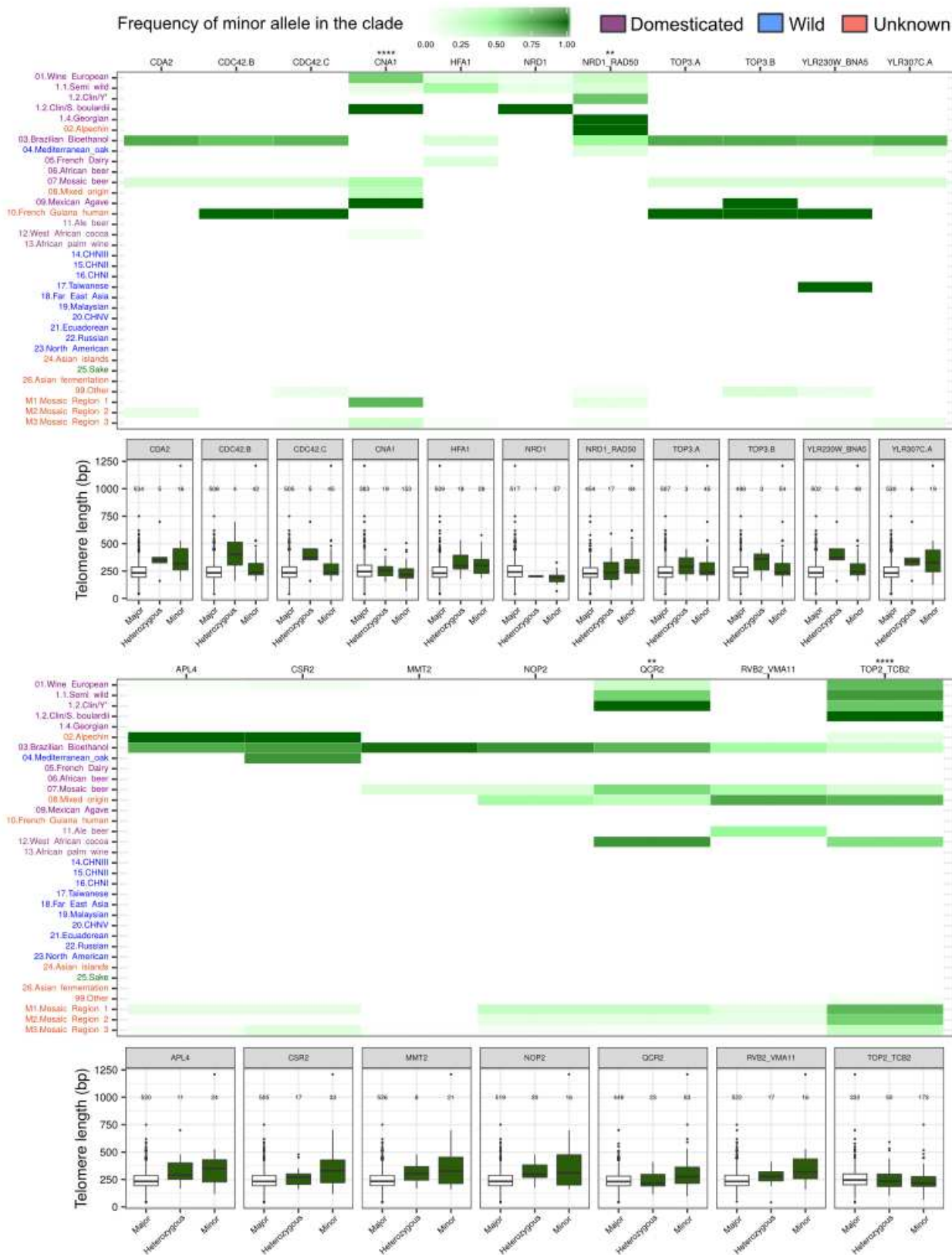
The bottom plot represents the telomere length of isolates divided by their CN for the 22 QTLs. <1: less than 1 copy of the gene; 1: 1 copy of the gene; >1: more than 1 copy of the gene. Numbers on top represent the number of isolates in each group. The isolates whose gene CN was unknown are not displayed. Box plots are as in Fig. 1.



Frequency of minor allele in the clade Domesticated Wild Unknown

0.00 0.25 0.50 0.75 1.00





Extended Data Figs. 8, 9, 10 – The upper plot represents a heatmap of the minor allele frequency (MAF) of QTNs. Squares represent the MAF in each clade, and the intensity of the green tone is directly proportional to the MAF in the clade. The order of clades on the y-axis is the same as in Fig. 3C and the clade classification is as in Fig. 1. The name of the QTN is the same as in Fig. 3C.

The bottom plot represents the telomere length of isolates divided by their genotype, and genotypes are as in Fig. 3C. Numbers on top represent the number of isolates in each group. The isolates whose genotype was unknown are not displayed. QTN names are as in the upper plot. Box plots are as in Fig. 1. Asterisks indicate QTNs which show a significant enrichment in minor alleles in either domesticated or wild clades relative to the expected values (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$).

Methods

Statistics and data reproducibility

We estimated TL in 918 already sequenced *S. cerevisiae* strains (Peter et al., 2018). Strains classifications, including phylogenetic lineages and domesticated/wild assignments, were maintained as described in (De Chiara, Barré, et al., 2020). For most of the analyses performed in this study we used the whole collection ($n=918$). A subset of euploid diploid strains ($n=555$) was used for the genome-wide association study in order to avoid the confounding effect of ploidy variation and aneuploidy. The genome-wide association study was performed as described in (De Chiara, Barré, et al., 2020). Genotype data, LOF variants and the set of phenotypes used in the correlation analysis were already available from (Peter et al., 2018), except for mitochondrial activity and volume, which were measured as previously described (De Chiara, Friedrich, et al., 2020). Standard GO term analysis was performed with the GO Term Finder tool available at Saccharomyces Genome Database (SGD). Significant GO terms were extracted by the algorithm implemented in the tool, with a False Discovery Rate (FDR) corrected α threshold of 0.1 (**Supplementary Table 4**). Mitochondrial hits were not included in the GO term analysis. LOF mutations were annotated as private to one or few clades and listed in Supplementary Table 5 if their frequency in the clade was higher than 0.6 but their frequency in the whole collection was lower than 0.2.

Statistical analyses were performed using R (version 3.5.3). Two-group comparisons were performed using the function *wilcox.test*, performing a 2-tailed Mann-Whitney test. Multiple-group comparisons were performed using the function *kruskal.test*, performing a non-parametric ANOVA test. Subsequent pairwise analyses were performed using the function *dunnTest* included in the package FSA. When multiple tests were performed on the same dataset, we applied a FDR correction for multiple hypothesis testing with a threshold of $\alpha=0.05$. Correlation analyses were performed using the function *cor.test* and applying both Pearson's and Spearman's models. P-values were subsequently corrected using both FDR and Bonferroni correction for multiple hypothesis testing.

Overview of the pipeline strategy to estimate TL and ITS content

For each strain, we screened all Illumina paired-end reads to extract the ones coming from telomeres and ITS. We classified reads as telomeric or ITS-derived if they contained a stretch of telomeric repeats (either C{1,3}A or TG{1,3}) longer or equal to 40 bp. The telomeric repeats must immediately follow one another with a limited number of gaps allowed. The list of telomeric/ITS reads is stored in an output file.

Subsequently, we performed mapping with BWA (version 0.7.12) (Li & Durbin, 2009) on a modified SGD reference genome in which we masked all repetitive sequences, including telomeres, ITS and Y' elements. The list of repetitive sequences in the SGD genome was generated and masked using RepeatMasker (<http://www.repeatmasker.org/>). A representative, long version Y' element was appended to this reference genome as an additional chromosome entry. The representative Y' element was chosen to be the one from TEL09L, as previously described (Yue et al., 2017). Read mapping runs were stored in the Sequence Alignment/Map (SAM) format and its binary form (BAM). Sorting, indexing and per-position coverage calculation were performed using SAMtools (version 1.2) (Li et al., 2009).

We annotated the mapping position of all our telomeric/ITS reads and their pairs and classified reads as ITS-derived based on the following assumptions: i) if a telomeric/ITS read contains a stretch of TG{1,3} repeats and its pair maps on a non-subtelomeric part of the genome, this read must come from a ITS; ii) if a telomeric/ITS read contains a stretch of TG{1,3} repeats and its pair maps on the representative Y' element, this read must come from a ITS.

We classified reads as non-Y'-associated ITS-derived if they respected the first criterion. We classified reads as Y'-associated ITS-derived if they respected the second criterion. All the remaining reads were classified as telomeric. We computed the average Y'-associated ITS content (YIC), non-Y'-associated ITS content (NYIC) and telomere length (TL) as:

$$YIL = \frac{(\sum^N S) * 2}{C} \quad NYIL = \frac{(\sum^N S) * 2}{C} \quad TL = \frac{\sum^N S}{C} - YIL - NYIL$$

where N =number of reads assigned to each category (Y' and non-Y' associated ITS-derived and telomeric), S =length of the telomeric repeats stretch contained in reads assigned to each category, and C =coverage in regions with a GC content of 50-80%, which is similar to what is found in yeast telomeric repeats. For simplicity, we define YIC in the main text as “ITS content”. To estimate the copy number of Y' elements, we calculated the average coverage on the representative TEL09L Y' element and divided it by the average coverage in regions with a GC content of 50-80%.

We applied our pipeline to a dataset of 7 *S. cerevisiae* and 5 *S. paradoxus* strains sequenced with both Illumina, Sanger and Pacbio technology and compared the estimation of our pipeline to the actual TL and ITS content values. TL and ITS content values were previously determined as described in the paragraphs “ITS content estimation in 12 yeast genome assemblies” and “TL estimation in the SGRP collection”. We observed a rate of underestimation of 4-fold for TL values, and a rate of 10-fold for ITS content values, resulting from our pipeline. We adjusted this bias in the analysis of the 918 yeasts collection by increasing their TL of 4-fold and ITS content of 10-fold. Values reported in Supplementary Table 1 are already adjusted.

ITS content and TL estimation in 12 yeast genome assemblies

We estimated TL and ITS content in 7 *S. cerevisiae* and 5 *S. paradoxus* strains sequenced with Pacbio technology and whose complete genome assemblies were already available (Yue et al., 2017). We estimated the frequency and distribution of stretches of telomeric repeats in the 12 genome assemblies by using custom Perl scripts. We merged the repeats which were interspaced by only 1 bp into a unique telomeric repeat stretch by using the software mergeBed included in the suite bedtools (version 2.17.0) (Quinlan & Hall, 2010) with the option “-d 1”. Then, we filtered the list to keep only the stretches longer than 10 bp. The stretches located at chromosome-ends were annotated as telomeres and were not kept for further analysis due to the unreliability of assembling softwares to correctly assemble chromosome-ends. The interstitial telomeric repeats stretches were annotated as Y'-associated ITS if they immediately preceded an annotated Y' element, while the others were discarded. The ITS content of a strain was determined as the sum of the length of its ITS annotations.

TL of these 12 yeast strains was determined by analysing their already available Sanger-sequencing reads, included in the SGRP project, as described in the paragraph “TL estimation in the SGRP collection”. Southern blots of these strains and of the representatives of the clades in the 918 yeasts collection were performed as previously described (Harari et al., 2017).

TL estimation in the SGRP collection

We estimated the TL of SGRP strains by using custom Perl scripts. First, we filtered the Sanger-sequencing reads to retain only the telomeric ones. We classified reads as telomeric if they contained a stretch of telomeric repeats (either C_{1,3}A or TG_{1,3}) longer or equal to 40 bp and located at the extremities of the read. The telomeric repeats must immediately follow one another with a limited number of gaps allowed. Then, TL was manually estimated from each telomeric read by manually counting the telomeric repeats. Each read was considered as a single telomere. The TL of a strain was calculated as the average of the TLs of its single telomeres.

Simulations of sequencing runs from genomes carrying synthetic telomeres

To estimate the efficiency of our pipeline, we applied it to artificially constructed datasets, consisting of simulated sequencing runs from genomes carrying synthetic telomeres of known length. We masked the native telomeres of the 12 previously described genome assemblies by using the program `maskfasta` contained in the suite `bedtools` (version 2.17.0) (Quinlan & Hall, 2010). Subsequently, we replaced the masked native telomeres with 14 synthetic telomeres of known length, consisting of yeast telomeric $TG\{1,3\}$ or $C\{1,3\}A$ repeats, and creating a total of 168 (12 x 14) synthetic genomes. The lengths of the synthetic telomeres had the following values: 18, 49, 77, 99, 135, 174, 208, 296, 360, 416, 504, 553, 602, 651 bp. ITS were not modified. Illumina reads were then simulated from the 168 synthetic genomes using the software `dwgsim` (<https://github.com/nh13/DWGSIM>) with the following parameters: type of read=paired-end, read length=100 bp, sequencing error rate=0%, insert size=300 bp, coverage=30.

References

- Aeby, E., Ahmed, W., Redon, S., Simanis, V., & Lingner, J. (2016). Peroxiredoxin 1 Protects Telomeres from Oxidative Damage and Preserves Telomeric DNA for Extension by Telomerase. *Cell Reports*, 17(12), 3107–3114. <https://doi.org/10.1016/j.celrep.2016.11.071>
- Ahmed, S., Passos, J. F., Birket, M. J., Beckmann, T., Brings, S., Peters, H., Birch-Machin, M. A., von Zglinicki, T., & Saretzki, G. (2008). Telomerase does not counteract telomere shortening but protects mitochondrial function under oxidative stress. *Journal of Cell Science*, 121(7), 1046–1053. <https://doi.org/10.1242/jcs.019372>
- Ai, W., Bertram, P. G., Tsang, C. K., Chan, T.-F., & Zheng, X. F. S. (2002). Regulation of Subtelomeric Silencing during Stress Response that Sir3 alone is sufficient to spread silencing in the absence of other silencing factors. *Molecular Cell*. Volume 10. Issue 6. Pages 1295-1305.
- Aviv, A., & Shay, J. W. (2018). Reflections on telomere dynamics and ageing-related diseases in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 373(1741). <https://doi.org/10.1098/rstb.2016.0436>
- Bergstrom, A., Simpson, J. T., Salinas, F., Barr??, B., Parts, L., Zia, A., Nguyen Ba, A. N., Moses, A. M., Louis, E. J., Mustonen, V., Warringer, J., Durbin, R., & Liti, G. (2014). A high-definition view of functional genetic variation from natural yeast genomes. *Molecular Biology and Evolution*, 31(4), 872–888. <https://doi.org/10.1093/molbev/msu037>
- Blackburn, E. H., & Gall, J. G. (1978). A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in Tetrahymena. *Journal of Molecular Biology*, 120(1), 33–53. [https://doi.org/10.1016/0022-2836\(78\)90294-2](https://doi.org/10.1016/0022-2836(78)90294-2)
- Bruhn, C., Ajazi, A., Ferrari, E., Lanz, M. C., Batrin, R., Choudhary, R., Walvekar, A., Laxman, S., Longhese, M. P., Fabre, E., Smolka, M. B., & Foiani, M. (2020). Subtelomeric Silencing. *Nature Communications*. <http://dx.doi.org/10.1038/s41467-020-17961-4>
- Buck, M. J., & Lieb, J. D. (2006). A chromatin-mediated mechanism for specification of conditional transcription factor targets. *Nat Genet*. 38(12), 1446–1451. <https://doi.org/10.1038/ng1917.A>
- Chen, L., Zhang, Y., Zhang, Q., Li, H., Luo, Z., & Fang, H. (2013). Mitochondrial Localization of Telomeric Protein TIN2 Links Telomere Regulation to Metabolic Control. *Mol Cell*. 47(6), 839–850.
- Cook, D. E., Zdraljevic, S., Tanny, R. E., Seo, B., Riccardi, D. D., Noble, L. M., Rockman, M. V., Alkema, M. J., Braendle, C., Kammenga, J. E., Wang, J., Kruglyak, L., Félix, M. A., Lee, J., & Andersen, E. C. (2016). The genetic basis of natural variation in *Caenorhabditis elegans* telomere length. *Genetics*. <https://doi.org/10.1534/genetics.116.191148>
- De Chiara, M., Barré, B., Persson, K., Chioma, A. O., Irizar, A., Schacherer, J., Warringer, J., & Liti, G. (2020). Domestication reprogrammed the budding yeast life cycle. *Bioarxiv*. <https://doi.org/10.1101/2020.02.08.939314>.
- De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., & Liti, G. (2020).

- Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biology*, 18(1), 1–15. <https://doi.org/10.1186/s12915-020-00786-4>
- Ding, Z., Mangino, M., Aviv, A., Spector, T., & Durbin, R. (2014). Estimating telomere length from whole genome sequence data. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gku181>
- Foley, N. M., Hughes, G. M., Huang, Z., Clarke, M., Jebb, D., Whelan, C. V., Petit, E. J., Touzalin, F., Farcy, O., Jones, G., Ransome, R. D., Kacprzyk, J., O'Connell, M. J., Kerth, G., Rebelo, H., Rodrigues, L., Puechmaille, S. J., & Teeling, E. C. (2018). Growing old, yet staying young: The role of telomeres in bats' exceptional longevity. *Science Advances*, 4(2), 1–12. <https://doi.org/10.1126/sciadv.aao0926>
- Fulnečková, J., Ševčíková, T., Fajkus, J., Lukešová, A., Lukeš, M., Vlček, Č., Lang, B. F., Kim, E., Eliáš, M., & Sýkorová, E. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution*, 5(3), 468–483. <https://doi.org/10.1093/gbe/evt019>
- Greider, C. W., & Blackburn, E. H. (1985). Identification of a specific telomere terminal transferase activity in tetrahymena extracts. *Cell*, 43(2 PART 1), 405–413. [https://doi.org/10.1016/0092-8674\(85\)90170-9](https://doi.org/10.1016/0092-8674(85)90170-9)
- Greider, C. W., & Blackburn, E. H. (1987). The telomere terminal transferase of tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, 51(6), 887–898. [https://doi.org/10.1016/0092-8674\(87\)90576-9](https://doi.org/10.1016/0092-8674(87)90576-9)
- Hakobyan, A., Nersisyan, L., & Arakelyan, A. (2016). Quantitative trait association study for mean telomere length in the South Asian genomes. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btw027>
- Hansen, M. E. B., Hunt, S. C., Stone, R. C., Horvath, K., Herbig, U., Ranciaro, A., Hirbo, J., Beggs, W., Reiner, A. P., Wilson, J. G., Kimura, M., Vivo, I. De, Chen, M. M., Kark, J. D., Levy, D., Nyambo, T., Tishkoff, S. A., & Aviv, A. (2016). Shorter telomere length in Europeans than in Africans due to polygenetic adaptation. *Human Molecular Genetics*, 25(11), 2324–2330. <https://doi.org/10.1093/hmg/ddw070>
- Harari, Y., & Kupiec, M. (2018). Do long telomeres affect cellular fitness? In *Current Genetics*. <https://doi.org/10.1007/s00294-017-0746-z>
- Harari, Y., Zadok-Laviel, S., & Kupiec, M. (2017). Long Telomeres Do Not Affect Cellular Fitness in Yeast. *mBio*. <https://doi.org/10.1128/mBio.01314-17>.
- Hausmann, M. F., Winkler, D. W., & Vleck, C. M. (2005). Longer telomeres associated with higher survival in birds. *Biology Letters*, 1(2), 212–214. <https://doi.org/10.1098/rsbl.2005.0301>
- Heidinger, B. J., Blount, J. D., Boner, W., Griffiths, K., Metcalfe, N. B., & Monaghan, P. (2012). Telomere length in early life predicts lifespan. *Proceedings of the National Academy of Sciences of the United States of America*, 109(5), 1743–1748. <https://doi.org/10.1073/pnas.1113306109>
- Kang, W. K., Devare, M., & Kim, J. Y. (2017). HST1 increases replicative lifespan of a sir2Δ mutant in the absence of PDE2 in *Saccharomyces cerevisiae*. *Journal of Microbiology*, 55(2),

123–129. <https://doi.org/10.1007/s12275-017-6535-z>

- Kim, H., Li, F., He, Q., Deng, T., Xu, J., Jin, F., Coarfa, C., Putluri, N., Liu, D., & Songyang, Z. (2017). Systematic analysis of human telomeric dysfunction using inducible telosome/shelterin CRISPR/Cas9 knockout cells. *Cell Discovery*, 3. <https://doi.org/10.1038/celldisc.2017.34>
- Kotiadis, V. N., Duchon, M. R., & Osellame, L. D. (2014). Mitochondrial quality control and communications with the nucleus are important in maintaining mitochondrial function and cell health. *Biochimica et Biophysica Acta - General Subjects*, 1840(4), 1254–1265. <https://doi.org/10.1016/j.bbagen.2013.10.041>
- Kupiec, M. (2014). Biology of telomeres: Lessons from budding yeast. In *FEMS Microbiology Reviews*. <https://doi.org/10.1111/1574-6976.12054>
- Laroche, T., Martin, S. G., Gotta, M., Gorham, H. C., Pryde, F. E., Louis, E. J., & Gasser, S. M. (1998). Mutation of yeast Ku genes disrupts the subnuclear organization of telomeres. *Current Biology*, 8(11), 653–657. [https://doi.org/10.1016/s0960-9822\(98\)70252-0](https://doi.org/10.1016/s0960-9822(98)70252-0)
- Lee, M., Napier, C. E., Yang, S. F., Arthur, J. W., Reddel, R. R., & Pickett, H. A. (2017). Comparative analysis of whole genome sequencing-based telomere length measurement techniques. *Methods*. <https://doi.org/10.1016/j.ymeth.2016.08.008>
- Lew, J. E., Enomoto, S., & Berman, J. (1998). Telomere Length Regulation and Telomeric Chromatin Require the Nonsense-Mediated mRNA Decay Pathway. *Molecular and Cellular Biology*, 18(10), 6121–6130. <https://doi.org/10.1128/mcb.18.10.6121>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liti, G., Carter, D. M., Moses, A. M., Warringer, J., Parts, L., James, S. a, Davey, R. P., Roberts, I. N., Burt, A., Tsai, I. J., Bergman, C. M., Bensasson, D., Michael, J. T., Kelly, O., Oudenaarden, A. Van, Barton, D. B. H., Bailes, E., Alex, N., Ba, N., ... Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*. 458(7236), 337–341. <https://doi.org/10.1038/nature07743>.
- Liti, G., Haricharan, S., Cubillos, F. A., Tierney, A. L., Sharp, S., Bertuch, A. A., Parts, L., Bailes, E., & Louis, E. J. (2009). Segregating YKU80 and TLC1 alleles underlying natural variation in telomere properties in wild yeast. *PLoS Genetics*, 5(9). <https://doi.org/10.1371/journal.pgen.1000659>
- Lundblad, V., & Blackburn, E. H. (1993). An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell*, 73(2), 347–360. [https://doi.org/10.1016/0092-8674\(93\)90234-H](https://doi.org/10.1016/0092-8674(93)90234-H)
- Maillet, L., Boscheron, C., Gotta, M., Marcand, S., Gilson, E., & Gasser, S. M. (1996). Evidence for silencing compartments within the yeast nucleus: A role for telomere proximity and Sir protein concentration in silencer-mediated repression. *Genes and Development*, 10(14), 1796–1811. <https://doi.org/10.1101/gad.10.14.1796>

- Mangino, M., Christiansen, L., Stone, R., Hunt, S. C., Horvath, K., Eisenberg, D. T. A., Kimura, M., Petersen, I., Kark, J. D., Herbig, U., Reiner, A. P., Benetos, A., Codd, V., Nyholt, D. R., Sinnreich, R., Christensen, K., Nassar, H., Hwang, S. J., Levy, D., ... Aviv, A. (2015). DCAF4, a novel gene associated with leucocyte telomere length. *Journal of Medical Genetics*, 52(3), 157–162. <https://doi.org/10.1136/jmedgenet-2014-102681>
- Martin, S. G., Laroche, T., Suka, N., Grunstein, M., & Gasser, S. M. (1999). Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell*, 97(5), 621–633. [https://doi.org/10.1016/S0092-8674\(00\)80773-4](https://doi.org/10.1016/S0092-8674(00)80773-4)
- Monaghan, P. (2010). Telomeres and life histories: The long and the short of it. In *Annals of the New York Academy of Sciences* (Vol. 1206, pp. 130–142). Blackwell Publishing Inc. <https://doi.org/10.1111/j.1749-6632.2010.05705.x>
- Nautiyal, S., DeRisi, J. L., & Blackburn, E. H. (2002). The genome-wide expression response to telomerase deletion in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(14), 9316–9321. <https://doi.org/10.1073/pnas.142162499>
- Nersisyan, L., & Arakelyan, A. (2015). Computel: Computation of mean telomere length from whole-genome next-generation sequencing data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0125201>
- Olovnikov, A. M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of Theoretical Biology*, 41(1), 181–190. [https://doi.org/10.1016/0022-5193\(73\)90198-7](https://doi.org/10.1016/0022-5193(73)90198-7)
- Opresko, P. L., Fan, J., Danzy, S., Wilson, D. M., & Bohr, V. A. (2005). Oxidative damage in telomeric DNA disrupts recognition by TRF1 and TRF2. *Nucleic Acids Research*, 33(4), 1230–1239. <https://doi.org/10.1093/nar/gki273>
- Ottaviani, A., Gilson, E., & Magdinier, F. (2007). Telomeric position effect: From the yeast paradigm to human pathologies? *Biochimie*. 90(1):93-107.
- Passos, J. F., Saretzki, G., & Von Zglinicki, T. (2007). DNA damage in telomeres and mitochondria during cellular senescence: Is there a connection? *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkm893>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates Species-wide genetic and phenotypic diversity. *Nature*. 556(7701):339-344.
- Platt, J. M., Ryzkin, P., Wanat, J. J., Donahue, G., Ricketts, M. D., Barrett, S. P., Waters, H. J., Song, S., Chavez, A., Abdallah, K. O., Master, S. R., Wang, L. S., & Johnson, F. B. (2013). Rap1 relocalization contributes to the chromatin-mediated gene expression profile and pace of cell senescence. *Genes and Development*, 27(12), 1406–1420. <https://doi.org/10.1101/gad.218776.113>

- Pryde., L. (2015). *Saccharomyces cerevisiae*. *The Dictionary of Genomics, Transcriptomics and Proteomics*, 0, 1–1. <https://doi.org/10.1002/9783527678679.dg11446>
- Puddu, F., Herzog, M., Selivanova, A., Wang, S., & Zhu, J. (2020). Genome architecture and stability in the *S. cerevisiae* knockout collection. *Nature*, 573(7774), 416–420. <https://doi.org/10.1038/s41586-019-1549-9>.Genome
- Qian, Kumar, Roginskaya, Fouquerel, Opresko, P. L., Shiva, S., Watkins, S. C., Kolodieznyi, D., Bruchez, M. P., & Van Houten, B. (2019). Chemoptogenetic damage to mitochondria causes rapid telomere dysfunction. *Proceedings of the National Academy of Sciences of the United States of America*, 116(37), 18435–18444. <https://doi.org/10.1073/pnas.1910574116>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Robin, J. D., Jacome Burbano, M. S., Peng, H., Croce, O., Thomas, J. L., Laberthonniere, C., Renault, V., Lototska, L., Pousse, M., Tessier, F., Bauwens, S., Leong, W., Sacconi, S., Schaeffer, L., Magdinier, F., Ye, J., & Gilson, E. (2020). Mitochondrial function in skeletal myofibers is controlled by a TRF2-SIRT3 axis over lifetime. *Aging Cell*, 19(3), 1–16. <https://doi.org/10.1111/acel.13097>
- Robyr, D., Suka, Y., Xenarios, I., Kurdistani, S. K., Wang, A., Suka, N., Grunstein, M., & Hall, B. (2002). Microarray Deacetylation Maps Determine Genome-Wide Functions for Yeast Histone Deacetylases. *Cell*, 109(4), 437–446.
- Romano, G. H., Harari, Y., Yehuda, T., Podhorzer, A., Rubinstein, L., Shamir, R., Gottlieb, A., Silberberg, Y., Pe'er, D., Ruppin, E., Sharan, R., & Kupiec, M. (2013). Environmental Stresses Disrupt Telomere Length Homeostasis. *PLoS Genetics*. <https://doi.org/10.1371/journal.pgen.1003721>
- Sahin, E., Colla, S., Liesa, M., Moslehi, J., Müller, F. L., Guo, M., Cooper, M., Kotton, D., Fabian, A. J., Walkey, C., Maser, R. S., Tonon, G., Foerster, F., Xiong, R., Wang, Y. A., Shukla, S. A., Jaskelioff, M., Martin, E. S., Heffernan, T. P., ... DePinho, R. A. (2011). Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature*, 470(7334), 359–365. <https://doi.org/10.1038/nature09787>
- Smith, J. J., Miller, L. R., Kreisberg, R., Vazquez, L., Wan, Y., & Aitchison, J. D. (2011). Environment-responsive transcription factors bind subtelomeric elements and regulate gene silencing. *Molecular Systems Biology*, 7(455). <https://doi.org/10.1038/msb.2010.110>
- Stone, E. M., & Pillus, L. (1996). Activation of an MAP kinase cascade leads to Sir3p hyperphosphorylation and strengthens transcriptional silencing. *Journal of Cell Biology*, 135(3), 571–583. <https://doi.org/10.1083/jcb.135.3.571>
- Teng, S.-C., & Zakian, V. A. (1999). Telomere-Telomere Recombination Is an Efficient Bypass Pathway for Telomere Maintenance in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology*, 19(12), 8083–8093. <https://doi.org/10.1128/mcb.19.12.8083>
- Teng, S. C., Chang, J., McCowan, B., & Zakian, V. A. (2000). Telomerase-independent lengthening of yeast telomeres occurs by an abrupt Rad50p-dependent, Rif-inhibited recombinational

process. *Molecular Cell*, 6(4), 947–952. [https://doi.org/10.1016/S1097-2765\(05\)00094-8](https://doi.org/10.1016/S1097-2765(05)00094-8)

Watson. (1972). *Nature New Biology*, 238, 37–38.

Wellinger, R. J., & Zakian, V. A. (2012). Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: Beginning to end. In *Genetics*. <https://doi.org/10.1534/genetics.111.137851>

Whittemore, K., Vera, E., Martínez-Nevado, E., Sanpera, C., & Blasco, M. A. (2019). Telomere shortening rate predicts species life span. *Proceedings of the National Academy of Sciences of the United States of America*, 116(30), 15122–15127. <https://doi.org/10.1073/pnas.1902452116>

Wyrick, J. J., Holstege, F. C. P., Jennings, E. G., Causton, H. C., Shore, D., Grustein, M., Lander, E. S., & Young, R. A. (1999). Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature*, 402(6760), 418–421. <https://doi.org/10.1038/46567>

Ye, J., Renault, V. M., Jamet, K., & Gilson, E. (2014). Transcriptional outcome of telomere signalling. *Nature Reviews Genetics*, 15(7), 491–503. <https://doi.org/10.1038/nrg3743>

Young, A. J. (2018). The role of telomeres in the mechanisms and evolution of life-history trade-offs and ageing. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 373, Issue 1741). Royal Society Publishing. <https://doi.org/10.1098/rstb.2016.0452>

Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>

Zijlmans, J. M. J. M., Martens, U. M., Poon, S. S. S., Raap, A. K., Tanke, H. J., Ward, R. K., & Lansooorp, P. M. (1997). Telomeres in the mouse have large inter-chromosomal variations in the number of t2ag3 repeats. *Proceedings of the National Academy of Sciences of the United States of America*, 94(14), 7423–7428. <https://doi.org/10.1073/pnas.94.14.7423>

Chapter 6

Inactivation of the DNA damage response rescues organismal fitness in telomere-humanized yeast

Eric Gilson, Gianni Liti and me designed the experiments; Benjamin Barré and me performed and analysed the experiments; Eric Gilson and Gianni Liti conceived and supervised the project; Eric Gilson, Gianni Liti and me wrote the paper.

Abstract

Telomeres are ribonucleoproteins which cap chromosome-ends and impede the activation of DNA-damage-response (DDR). Their length is maintained by the reverse transcriptase telomerase using a RNA template. Telomeric sequences are conserved across the tree of life and consist of G-rich repeated units, but variation in length and sequence is present among distant taxa. Each species is likely fine-tuned for its own telomere properties, however their quantitative contribution to the organismal fitness remains largely unexplored. In the budding yeast *S. cerevisiae*, the telomerase RNA template is encoded by *TLC1* and carries degenerated TG₁₋₃ repeats. Previous studies have shown that a 16-bp editing of *TLC1* enables to globally reconfigure telomeres by generating yeasts with newly synthesised exact T₂AG₃ repeats, which show an intrinsic telomere dysfunction and a chronic activation of the DDR. Since T₂AG₃ corresponds to the telomeric repeat found in humans (as well as many other organisms), these telomere-engineered yeasts were dubbed humanized yeasts. In this work, we evolved multiple lines of humanized and wild-type yeasts to characterize the effect of telomere variation on fitness. We used a strategy that sequentially combined two experimental evolution paradigms: first, we evolved cells through mutation accumulation lines (MALs) to minimize selection. This enables to investigate the cellular and genomics effects during the fitness decay driven by telomere dysfunction. Next, we submitted MALs to adaptive evolution by multiple serial transfers (STs) of large population sizes, to map mutations that counteract the fitness decline. During MALs, humanized yeasts gradually slowed their growth and shortened chronological lifespan. Whole-genome-sequencing revealed that they had increased mutation rate and genome instability, with recurrent aneuploidies on chromosome XVI. We observed recurrent mutations in DDR genes, including *TEL1* and *MEC1*. The *TEL1* line underwent multiple chromosomal rearrangements, subtelomeric amplifications and recovered growth fitness. After multiple STs, most humanized lines recovered a wild-type fitness, with independent occurrence of mutations in the MRX complex, a key effector of DDR. Overall, our results show that humanized

telomeres increase mutation rate and cause a severe fitness decline which is rescued by the inactivation of DDR.

Introduction

Telomeres are ribonucleoproteins that protect the end of linear chromosomes of eukaryotic organisms (Greider & Blackburn, 1987). They prevent chromosome-ends to be mistakenly recognised as DNA double-strand breaks and repaired generating end-to-end chromosome fusions (McClintock, 1941). Their additional function is to solve the so-called “end replication problem” (Olovnikov, 1973; Watson, 1972). Telomeres are composed of tandem repeats of a DNA sequence that act as a buffer, preventing internal genes to be deleted as a result of chromosome end resection. Telomere length is maintained via the specialised enzyme telomerase. This enzyme is a reverse transcriptase that synthesizes new telomeric repeats using a RNA moiety as template (Greider & Blackburn, 1989).

Telomeric repeat units can vary across the tree of life. The vertebrate-like TTAGGG (T₂AG₃) repeat is the most widespread, being present in the majority of Metazoa, but also in some classes of protists, fungi and plants, and is thought to be the ancestral telomeric repeat from which all the others evolved (Fulnečková et al., 2013). Variations of this motif constitute the prevalent telomeric repeat in other taxa, like TTTAGGG in plants (Peska & Garcia, 2020), TTAGG in insects (Kuznetsova, 2020), TTAGGC in nematodes (Zetka & Müller, 1996) and degenerate motifs in ascomycetous yeasts (Teixeira & Gilson, 2005). Despite the great degree of diversity, all these motifs have some common characteristics, like being short, tandemly-repeated and G-rich, and the principles of telomere biology are conserved even among distantly related eukaryotes (Kupiec, 2014). The molecular mechanisms of telomere maintenance have been thoroughly investigated in the last 30 years (Chan & Blackburn, 2002; De Lange, 2005; Greider & Blackburn, 1989; Lundblad & Szostak, 1989; Marcand et al., 1997). However, we still lack a mechanistic understanding of the evolutionary processes that led to the astonishing variety of telomeric sequences in eukaryotes, and how such diversity affects fitness.

Human telomeres are constituted by the canonical T₂AG₃ repeat, whereas telomeres of the budding yeast *Saccharomyces cerevisiae* are constituted by the degenerate TG₁₋₃ repeat (Wellinger & Zakian, 2012). Yeast telomeres that have the same sequence as human telomeres can be generated by using a “humanized” version of *TLC1*, the gene encoding the RNA template of telomerase, in which the template region is altered to encode T₂AG₃ repeats, instead of the native TG₁₋₃ (Bah et al., 2004, 2011; Henning et al., 1998). The humanization of yeast telomeres is accompanied by a switch in telomere-binding proteins, in which RAP1p, originally bound to native TG₁₋₃ repeats, is replaced by TBF1p (Alexander & Zakian, 2003; Brevet et al., 2003; Ribaud et al., 2012), and induces the activation of DNA damage response (DDR) pathways (Di Domenico et al., 2009; Di Domenico et al., 2013). Humanized yeast can be used as a model to investigate the mechanisms of telomere evolution and how organisms adapt to new telomeric sequences.

In this study, we characterized the evolutionary response of *S. cerevisiae* to telomere variation at the phenotypic and genomic level. First, we evolved cells through mutation accumulation lines (MALs) to minimize selection and investigate the cellular and genomic effects of telomere dysfunction. Unlike other experimental designs in which a component of the telomerase holoenzyme is deleted and cells can only survive a few generations, this approach allowed us to characterize the long-term effects of a mild but chronic telomere dysfunction. During MALs, humanized yeasts showed a progressive fitness decay, accompanied by higher rates of mutations and genome instability. Then, we submitted MALs to adaptive evolution by multiple serial transfers (STs) of large population sizes, to map mutations that counteract the fitness decline. After multiple STs, most humanized lines recovered a wild-type fitness, with independent occurrence of mutations in the DDR pathway. Overall, our results reveal that telomere biology is characterized by a high degree of evolutionary plasticity, and indicate a possible route of telomere evolution through the inactivation of DDR.

Results

Humanized yeasts show fitness decay upon MAL

Fitness is defined as the capacity of an organism to thrive in a given environment (Di Gregorio et al., 2016), and results from the contribution of both its genotype and transcriptional profile. Although there is not a univocal estimation, fitness can be quantified by using a combination of various parameters, like growth rate, metabolic rate and lifespan.

Telomeric repeat units are highly variable across the tree of life, and each species is likely fine-tuned for its own telomere properties. However, the quantitative contribution of telomere sequence variation to the organismal fitness has remained largely unexplored. We used the humanized yeast model system to address this question, and we took advantage of humanized and WT strains already generated in previous studies (Bah et al., 2004, 2011). From a single humanized and WT ancestors, we evolved respectively 16 parallel haploid lines of humanized yeasts and 11 lines of wild-type yeasts by using a mutation accumulation lines (MALs) protocol aimed to minimise selective pressure. MALs were passed through a single-cell bottleneck (scb) every ~72 h (~20 generations) for a total of 100 scb, and samples were collected and frozen every 20 scb (**see methods**).

Previous studies showed that humanized yeasts progressively accumulate G2/M cell cycle delay due to the chronic activation of the Rad53-dependent DNA damage checkpoint pathway (Di Domenico et al., 2009). We confirmed these results by observing a progressive decrease in colony size in humanized lines (data not shown), and by comparing growth performance in humanized and wild-type yeasts at the beginning and at the end of the MAL experiment. Humanized and WT yeasts had the same growth rate at 2 scb, but humanized grew slower than WT at 100 scb (two-tailed Wilcoxon test, $p=1.933e-15$), as reflected by their higher population doubling time, while WT maintained the

same growth rate during the MAL experiment (**Figure 1a**). We observed the same tendency with the length of the lag phase. Both humanized and WT had the same lag duration at 2 scb, but humanized had a longer lag phase than WT at 100 scb (two-tailed Wilcoxon test, $p=1.382e-06$) (**Figure 1b**). Interestingly, both humanized and WT show a higher variance at 100 scb, probably due to the accumulation of mutations during MAL.

Next, we asked whether telomere humanization had an effect on chronological lifespan (CLS, e.g. the ability of cells to survive in non-dividing conditions). We measured CLS at regular time points during a 15 days experiment and compared the percentage of surviving cells in humanized and WT lines at each timepoint. Overall, humanized yeasts had a lower survival rate than WT at all timepoints, and the difference with the WT became more pronounced as we approached the end of the two weeks (two-tailed Wilcoxon test, $p=2.373e-10$ at 15 days) (**Figure 1c**).

Then, we explored the pace at which this fitness decay occurred by sampling multiple timepoints. We estimated growth performance and CLS in 4 representative lines of humanized yeasts (5 for CLS) and 3 representative lines of WT yeasts. These experiments revealed that humanized and WT yeasts had similar growth performance (represented by lag duration and population doubling time) at the beginning of MALs. The performance of WT lines did not change during the MAL experiment, while it progressively worsened for humanized yeasts (**Extended Data Fig. 1a,b**). We observed the same phenomenon for CLS, as WT lines maintained a constant survival rate during MAL, while humanized lines progressively decreased their survival rate (**Extended Data Fig. 1c**). Surprisingly, humanized lines seem to partially recover their fitness starting from 60 scb. After this timepoint, their growth performance and survival rate improve, even though they do not match those of WT yeasts even at 100 scb.

Overall, our experiments confirmed that telomere humanization causes a severe fitness decline, which has a very rapid onset, but this is partially rescued at a later stage.

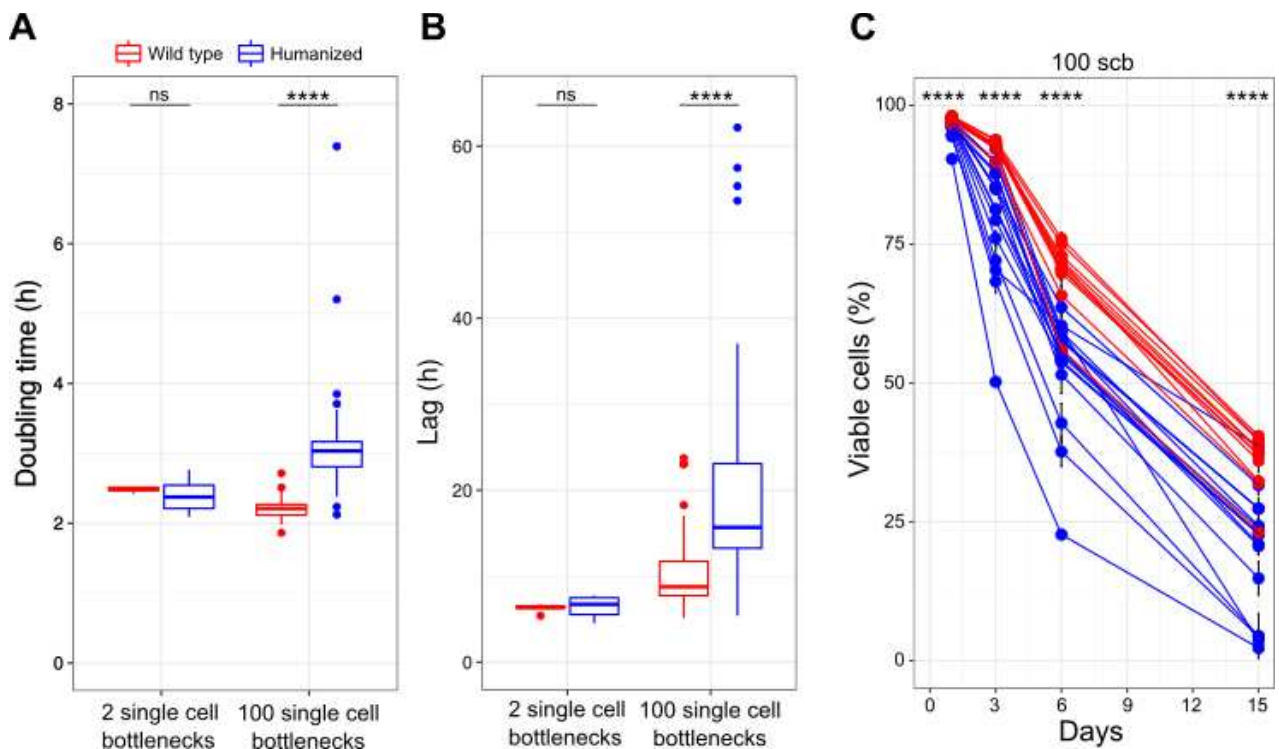


Figure 1 – Fitness decay. **a**, Population doubling time of WT and humanized lines at 2 ($n=2$ and $n=4$, respectively) and 100 scb ($n=11$ and $n=16$, respectively). Boxes: horizontal line: median, upper/lower hinge: interquartile range (IQR), whiskers: largest/smallest value within upper/lower hinge $\pm 1.5 \times$ IQR. ns: non-significant, ****: $p < 0.0001$. The experiment was performed in triplicate. **b**, Lag phase duration of WT and humanized lines at 2 ($n=2$ and $n=4$, respectively) and 100 scb ($n=11$ and $n=16$, respectively). Boxplots as in Fig. 1a. ns: non-significant, ****: $p < 0.0001$. The experiment was performed in triplicate. **c**, Percentage of surviving cells (average \pm SD) for WT and humanized lines at 100 scb during a 15-days chronological lifespan experiment ($n=11$ and $n=16$, respectively). ns: non-significant, ****: $p < 0.0001$. The experiment was performed in triplicate.

Humanized yeasts have higher oxidative stress than WT yeasts

Telomere and mitochondrial biology are closely connected and telomere dysfunction is known to affect mitochondrial metabolism through the transcriptional dysregulation of mitochondrial biogenesis pathways (Finkel, 2011; Robin et al., 2020; Sahin et al., 2011). We investigated whether telomere humanization had an effect on mitochondrial content and activity through flow cytometry. At 100 scb, humanized yeasts had lower mitochondrial activity than WT (two-tailed Wilcoxon test, $p=1.807 \times 10^{-7}$), but higher mitochondrial content (two-tailed Wilcoxon test, $p < 2.2 \times 10^{-16}$) (Figure 2a,b). These apparent contrasting results might be explained by the fact that replication of dysfunctional mitochondria can result in an increase of mitochondrial content but not of their

activity. We asked whether the increase of mitochondrial content was accompanied also by an increase in copy number of mitochondrial genome. We whole-genome sequenced all the 16 humanized and 11 WT lines and estimated the mtDNA copy number by analysing coverage along three mitochondrially-encoded genes: *COX2*, *COX3* and *ATP6* (De Chiara et al., 2020). While starting from similar mtDNA copy number at 2 scb (average of 2 mtDNA CN), humanized yeasts increased their mtDNA copies during MALs to a final average of ~3 copies, while WT did not, resulting in a significant difference (two-tailed Wilcoxon test, $p=0.0001789$ at 100 scb) (**Extended Data Fig. 2a**).

Abnormalities in mtDNA can manifest in yeast through a characteristic small colony size, due to a block in the aerobic respiratory chain pathway, and these yeast mutants are called “petites” (Chen & Clark-Walker, 1999). We tested whether telomere humanization and its consequent disruption of the mitochondrial equilibrium had an effect in the propensity to generate “petite” colonies. Indeed humanized yeasts generated a high proportion of petite colonies when plated on YPG medium (~30% on average). This effect was already present at the beginning of the MALs and did not change during the experiment. This phenotype was linked to telomere humanization as demonstrated by the fact that WT yeasts had a negligible percentage of “petites” (close to 0%) (**Extended Data Fig. 2b**).

Mitochondrial dysfunction is usually characterized by the increase of reactive oxygen species derived from the malfunctioning of the respiratory chain, which have been suggested to have a relevant role in the onset of aging in both proliferating and non-proliferating tissues (Murphy, 2009). We measured the amount of superoxide anion ($O_2^{\bullet-}$), the proximal mitochondrial ROS, in humanized and WT lines, by flow cytometry. We found a clear difference in the amount of ROS between the two genetic backgrounds, with humanized yeasts having higher ROS than WT (two-tailed Wilcoxon test, $p=8.554e-05$) (**Figure 2c**). Some humanized lines exhibited a particularly high amount of ROS, while one WT had a discordant behaviour respect to the rest and had a ROS amount comparable to that of the highest ROS-containing humanized lines. We next characterized the pace at which mitochondrial dysfunction occurred and measured ROS at different timepoints during MALs in representatives of humanized and WT yeasts. This analysis revealed that ROS accumulation started soon after the beginning of MALs and kept increasing until the end in humanized yeasts, while ROS content did not change in WT yeasts (**Extended Data Fig. 2c**). Differently from other fitness data presented in the previous paragraph, the mitochondrial fitness does not seem to be rescued at late stages of the MAL experiment, as reflected by the fact that ROS do not decrease after that point but seem to have reached a plateau.

Overall, this experiments confirmed the close connection between telomere, DDR and mitochondrial biology already postulated in previous literature, and showed that telomere

humanization causes mitochondrial dysfunction. The malfunctioning of the respiratory chain induces a rapid increase in the amount of ROS and in the tendency to form “petite” colonies, which is not rescued by the increase in mitochondrial DNA, and actively contribute to the deterioration of metabolic fitness in humanized yeasts.

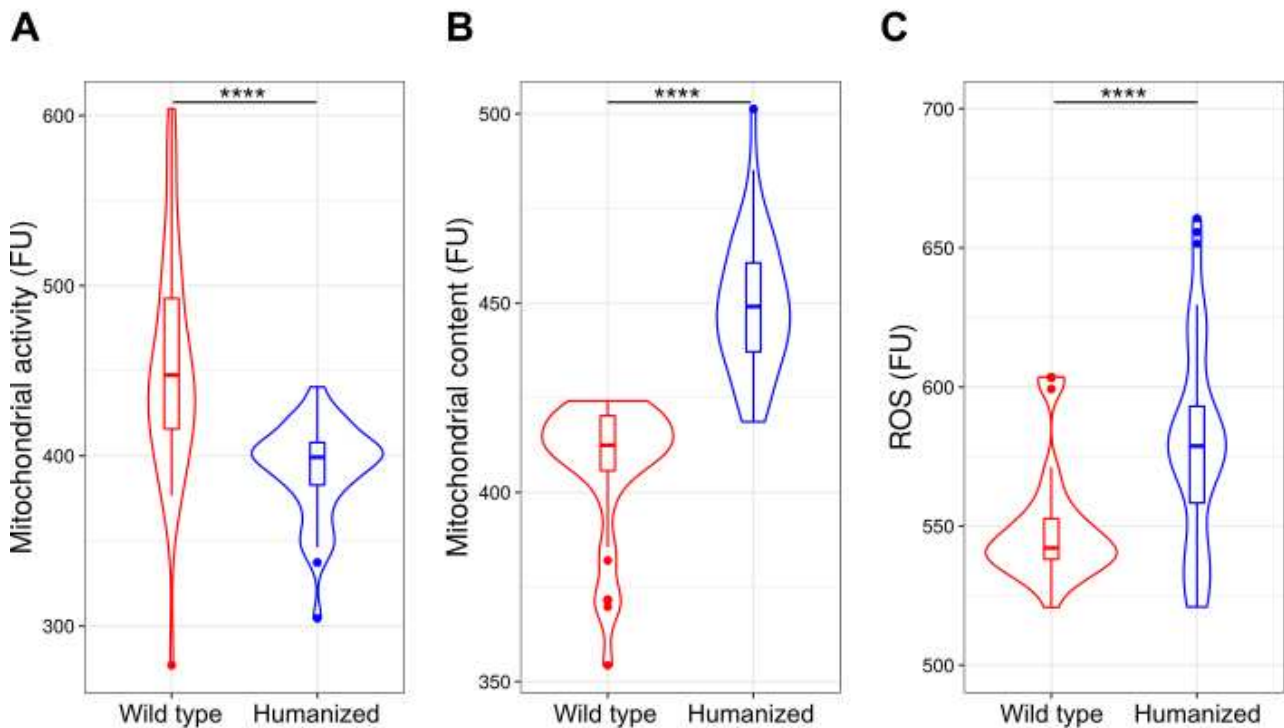


Figure 2 – Oxidative stress. **a**, Mitochondrial activity of WT and humanized lines at 100 scb ($n=11$ and $n=16$, respectively), measured in fluorescence units (FU). Boxplots as in Fig. 1a. ns=non-significant, ****: $p<0.0001$. The experiment was performed in triplicate. **b**, Mitochondrial content of WT and humanized lines at 100 scb ($n=11$ and $n=16$, respectively), measured in fluorescence units (FU). Boxplots as in Fig. 1a. ns=non-significant, ****: $p<0.0001$. The experiment was performed in triplicate. **c**, Content of reactive oxygen species (ROS) of WT and humanized lines at 100 scb ($n=11$ and $n=16$, respectively), measured in fluorescence units (FU). Boxplots as in Fig. 1a. ns=non-significant, ****: $p<0.0001$. The experiment was performed in triplicate.

Telomere humanization induces genome instability

After revealing the effect of telomere humanization on cellular fitness, we sought to understand what was the impact of telomere humanization on the genomic landscape. We performed whole-genome sequencing of all the 16 humanized and 11 WT lines at the end of the MAL protocol (100 scb) and of the two ancestral strains (one humanized and one WT) from which all the others derived. First, we estimated telomere length and repeat composition by using a previously described bioinformatic pipeline (chapter 5). We measured the length of the telomeric portion containing

native TG_{1-3} repeats and found that WT yeasts have longer TG_{1-3} portions than humanized yeasts. We next estimated the length of the T_2AG_3 portion and, as expected, WT yeasts did not contain any T_2AG_3 repeats at both 2 and 100 scb. On the contrary, humanized yeasts accumulated T_2AG_3 repeats during the MAL experiment starting from completely native TG_{1-3} telomeres at 2 scb.

We next sum-up the TG_{1-3} and the T_2AG_3 components to determine the total length of telomeres and found that, overall, WT yeasts had longer telomeres than humanized ones, in accordance with previous literature (Bah et al., 2004). This indicates that, despite containing an active telomerase, humanized yeasts cannot replenish telomeres at the same pace to which they lose them (**Figure 3a and Extended Data Fig.3a**).

Subsequently, we explored the structure of subtelomeres by using the same pipeline and estimated the content of interstitial telomeric sequences (ITS) and the copy number (CN) of Y' elements. Our analysis evidenced an increase in ITS content in humanized, but not in WT yeasts, during MAL. This was accompanied by an increase in CN of Y' elements, and, globally, the two variables correlated well ($R=0.92$, $P=9.192e-13$) (**Extended Data Fig.3b-d**). Two humanized lines carried a very high amount of ITS and Y', a characteristic signature of type I survivors, which are strains that survive telomerase inactivation thanks to recombination-mediated telomere lengthening mechanisms. Interestingly, we discovered that humanized yeasts also carried ITS with T_2AG_3 repeats, which were not present either in WT yeasts or in humanized ones at 2 scb. We manually analysed sequencing reads containing segments of these ITS and found them to be in three possible conformations: i) T_2AG_3 -only ITS which were constituted entirely by T_2AG_3 repeats, ii) mixed ITS with an internal segment constituted by TG_{1-3} repeats and a telomere-proximal segment constituted by T_2AG_3 repeats, iii) sandwich ITS containing an internal and a telomere-proximal segments with TG_{1-3} repeats and a central part constituted by T_2AG_3 repeats (**Extended Data Fig.4a**). The presence of these ITS structures indicates that recombinational mechanisms are active in humanized yeasts and contribute to telomere-subtelomere dynamics. These results are in agreement with the Y' amplification previously reported in humanized yeasts (Di Domenico et al., 2013).

Haploid yeasts are known to spontaneously diploidize during MALs (Sharp et al., 2018). We checked for the occurrence of ploidy changes in our lines and we found that all but one lines remained haploid. The only line that became diploid was a humanized one and the ploidy change occurred at the very end of the MAL experiment, between 80 and 100 scb. Diploidization likely occurred through mitotic defects as the *MAT* locus remained homozygous (*MATa/MATa*).

We next analysed coverage profiles along chromosomes to determine whether CNV and aneuploidies occurred during MAL. We discovered 17 aneuploidies in our lines, which were all gain of one copy. Among these, only one was found in a WT line, whereas all the others involved humanized lines, which were thus more subject to this type of genome instability (two-tailed

Wilcoxon test, $p=0.0001$). The majority of aneuploidies involved chromosome XVI, which had been duplicated in 12 out of the 16 humanized lines (**Figure 3b**). We tested whether an additional copy of chromosome XVI might have an adaptive value for humanized yeasts and checked if this aneuploidy was associated with any of the previously characterized phenotypes. We found that XVI-aneuploid humanized lines had shorter telomeres than XVI-euploid humanized ones. We further characterized this difference by comparing the length of the TG₁₋₃ and the T₂AG₃ portions of telomeres in XVI-euploid and XVI-aneuploid lines, and found that the difference was driven by the T₂AG₃ segment, while the TG₁₋₃ one was similar between the two groups (**Figure 3c**).

Finally, we determined the karyotype structure of humanized and WT lines by pulsed-field gel electrophoresis (PFGE). Humanized and WT ancestors at 2 scb, as well as all the WT lines at 100 scb, had a normal karyotype, identical to the one of the S288C control. On the contrary, humanized lines at 100 scb have different band patterns which are compatible with both chromosomal rearrangements and subtelomeric amplifications similar to the ones observed in type I survivors. Interestingly, most of the band shifts involve the chromosome XVI band, which has a higher molecular weight than the control (**Extended Data Fig.4b**).

In conclusion, these analysis confirmed that humanized lines have mixed telomeres composed of both TG₁₋₃ and T₂AG₃ repeats. Humanized telomeres trigger genome instability which manifests in a high aneuploidy rate, subtelomeric amplifications and structural rearrangements. Surprisingly, chromosome XVI is a preferential target of genome instability and might have an adaptive role in the survival of humanized yeasts.

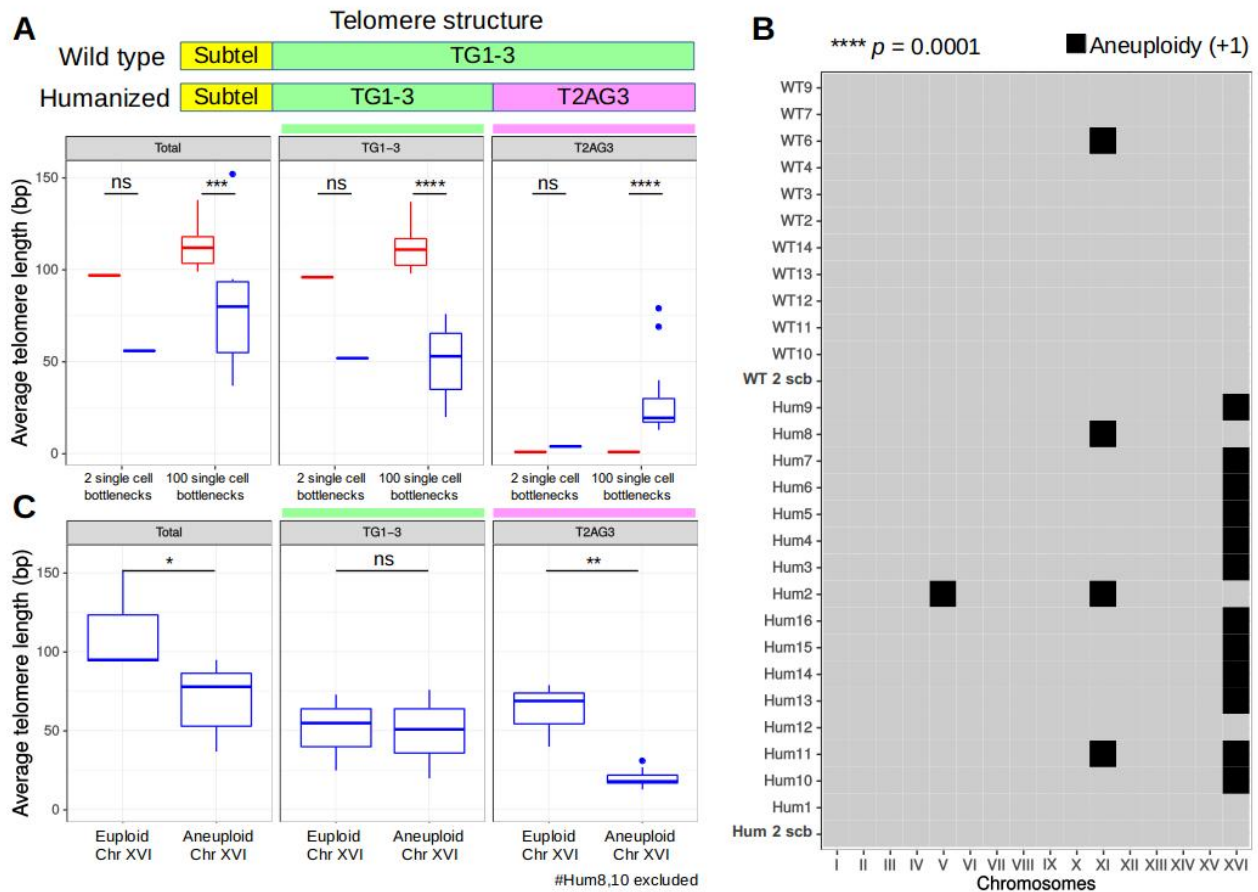


Figure 3 – Genome instability in humanized yeasts – a, Schematic illustration of telomere structure in wild type and humanized yeasts. Telomere length estimation in WT and humanized yeasts at 2 ($n=2$ and $n=4$, respectively) and 100 scb ($n=11$ and $n=14$, respectively). Telomere length estimations are divided in three categories: TG1-3, T2AG3, total (TG1-3 + T2AG3). Boxplots as in Fig. 1a. ns=non-significant, ***: $p<0.001$, ****: $p<0.0001$. Humanized lines 8 and 10 have been excluded from this analysis due to their high ITS content. **b**, The plot represents the state of each chromosome in the MALs. Grey squares represent euploid chromosomes which maintained their original copy number (1), while black squares represent chromosomes which underwent aneuploidy. In this case, all aneuploidies are gain of copies (+1). Chromosome XVI highlights as a preferential site for aneuploidy. **c**, Telomere length estimation in humanized yeasts divided in two groups on the basis of the state of their chromosome XVI: euploid (one copy) or aneuploid (two copies), at the end of the MAL protocol (100 scb) ($n=3$ and $n=11$, respectively). Telomere length estimations are divided in three categories as in figure 3a. Boxplots as in Fig. 1a. ns=non-significant, *: $p<0.05$, **: $p<0.01$. Humanized lines 8 (XVI-euploid) and 10 (XVI-aneuploid) have been excluded from this analysis due to their high ITS content.

Humanized yeasts have higher mutation rate than WT

We whole-genome sequenced the humanized and WT lines at 100 scb plus the two ancestors at 2 scb in order to characterize possible effects of telomere humanization on the mutational spectrum. We compared the humanized and WT lines with their respective ancestors to infer the new mutational events occurred during the MAL experiment. In total, we identified 250 point mutations in the humanized lines, and 87 in the WT. Among the 250 found in humanized lines, 122 were located in subtelomeres and 128 were located in the core parts of the chromosomes. For the WT lines, 15 mutations were found in subtelomeres while 72 occurred in the core genome. Given that subtelomeres constitute only a minor fraction of the genome (1310254 bp vs 10761072 bp in subtelomeres and core genome, respectively), humanized yeasts have an enrichment in mutations occurring there (binomial test, $p < 0.0001$), while WT yeasts have a subtelomere/core mutation distribution similar to what is expected by chance (binomial test, $p = \text{ns}$). We observed the same tendency for insertions/deletions of less than 100 bp, which were also enriched at subtelomeric sites in humanized (binomial test, $p < 0.0001$), but not in WT yeasts (binomial test, $p = \text{ns}$) (**Extended Data Fig. 5a**).

Then, we used these results to estimate base substitution and indels rates. To account for differences in doubling time between humanized and WT yeasts, we measured the number of cells per colony throughout the MAL experiment in representatives of humanized and WT yeasts, and used these values to calculate mutation rates per generation. We estimated an initial growth rate of ~ 20 generations/bottleneck (~ 72 h) for both humanized and WT. While WT maintained a constant growth rate during MAL, humanized progressively slowed their growth and performed only ~ 14 generations/bottleneck at the end of MAL (**Extended Data Fig. 5b**). This accounted for a total of 2140 generations performed by WT lines in the MAL experiment, and a total of 1640 generations performed by humanized lines. Using these values, we estimated a genome-wide point mutation rate of $7.9e^{-10}$ for humanized yeasts, and $3.0e^{-10}$ for WT. Subdividing the mutations in those occurring in subtelomeres and in the core genome did not change the result, with humanized yeasts having higher mutation rate than WT, but evidenced a propensity for humanized yeasts to mutate their subtelomeric sites, where the mutation rate was 10-fold higher than in the core genome ($3.55e^{-9}$ vs $4.53e^{-10}$, respectively) (**Figure 4a**). We did the same analysis for insertions/deletions and mirrored the previous results, with humanized yeasts having a higher indel rate than WT at both core and subtelomeric sites (**Extended Data Fig. 5c**).

Given that humanized yeasts have a high amount of ROS and these are known to damage the mitochondrial DNA, we tested whether there was a difference in mtDNA mutation rate (85779 bp) between the two groups. Surprisingly, we did not find any significant difference in point mutation rate and indel rate between humanized and WT lines (**Extended Data Fig. 5d,e**). We further

investigated whether the higher mutation rate of humanized yeasts was also characterized by a specific mutational signature. However, we did not identify any enrichment in transitions vs transversions or any specific base substitution (data not shown).

We tested whether selection during MAL might have affected point mutation rates. Assuming that coding and non-coding regions are equally likely to mutate, point mutations were not more/less likely to occur in genes or to be nonsynonymous than expected in WT lines, indicating an absence of selective pressure in MAL. However, point mutations were found in coding regions less often than expected (observed fraction: 0.59, expected fraction: 0.75; binomial test: $P < 0.0001$) in humanized yeasts, indicating an effect of selection in shaping the distribution of mutations in this group (**Extended Data Fig. 6a,b**). To have a better insight into the selection outcome, we checked if the mutations in humanized and WT lines were enriched for specific gene categories. A GO term analysis revealed that WT lines were enriched in genes encoding ion binding proteins ($p=0.03$), while humanized lines were enriched in genes involved in the ceramid biosynthesis process (0.1) and encoding cell periphery proteins ($p=0.06$). Interestingly, humanized yeasts had also an enrichment in effectors of the DNA damage sensing pathway (*TEL1* and *MEC1*; $p=0.09$). We further characterized the phenotypic effects of mutations, with a particular attention on the *TEL1/MEC1* mutations. Multiple humanized and WT mutations were predicted to have important phenotypic effects, either by affecting the stability of the protein or its interaction with other partners. The *TEL1* mutation was an aminoacid substitution located in the FAT domain of the protein and predicted to affect its function. An alignment of the yeast *TEL1* protein with its human ortholog ATM revealed that a deletion of the corresponding ATM aminoacid (together with the upstream two) was already found in ataxia telangiectasia patients (Wright et al., 1996), in a sporadic case of T-cell prolymphocytic leukemia (Vorechovsky et al., 1997) and in a breast cancer patient with a family history of multiple malignancies (Vořechovský et al., 1996). These findings are a strong indication that our mutation might be a loss-of-function mutation inactivating the DNA damage response (**Extended Data Fig. 6c**). Further characterization of the humanized line carrying this mutation revealed that it underwent considerable amplification of ITS and Y' elements and was the only line undergoing diploidization (**Extended Data Fig. 6d**). We Sanger-sequenced this line at all the timepoints and determined that the *TEL1* mutation occurred between 60 and 80 scb, and could therefore have triggered the diploidization. We performed long-read sequencing and genome assembly of this line and found that it underwent chromosomal rearrangements involving multiple chromosomes (**Extended Data Fig. 6e**).

We compared the timing of the *TEL1* mutation occurrence with the growth rate profile of the humanized line carrying it, and found them to be remarkably compatible. The partial recovery of growth performance occurs after the appearance of the *TEL1* mutation (**Figure 4b**). Together with

previous evidence, it strongly indicates that the inactivation of the DDR can rescue fitness in humanized yeasts.

Overall, these findings reveal that telomere humanization increases mutation rate, and it can have an impact not only on the surrounding genomic regions, but also on more distant ones, as it is demonstrated by the increase in mutation rate in core parts of the genome. Despite MAL should be selectively neutral, we found signals of selection on humanized lines, which accumulated mutations in specific cellular pathways. In particular, LOF mutations in the DDR pathway seem to have a beneficial effect on humanized yeasts and induce a partial fitness recovery.

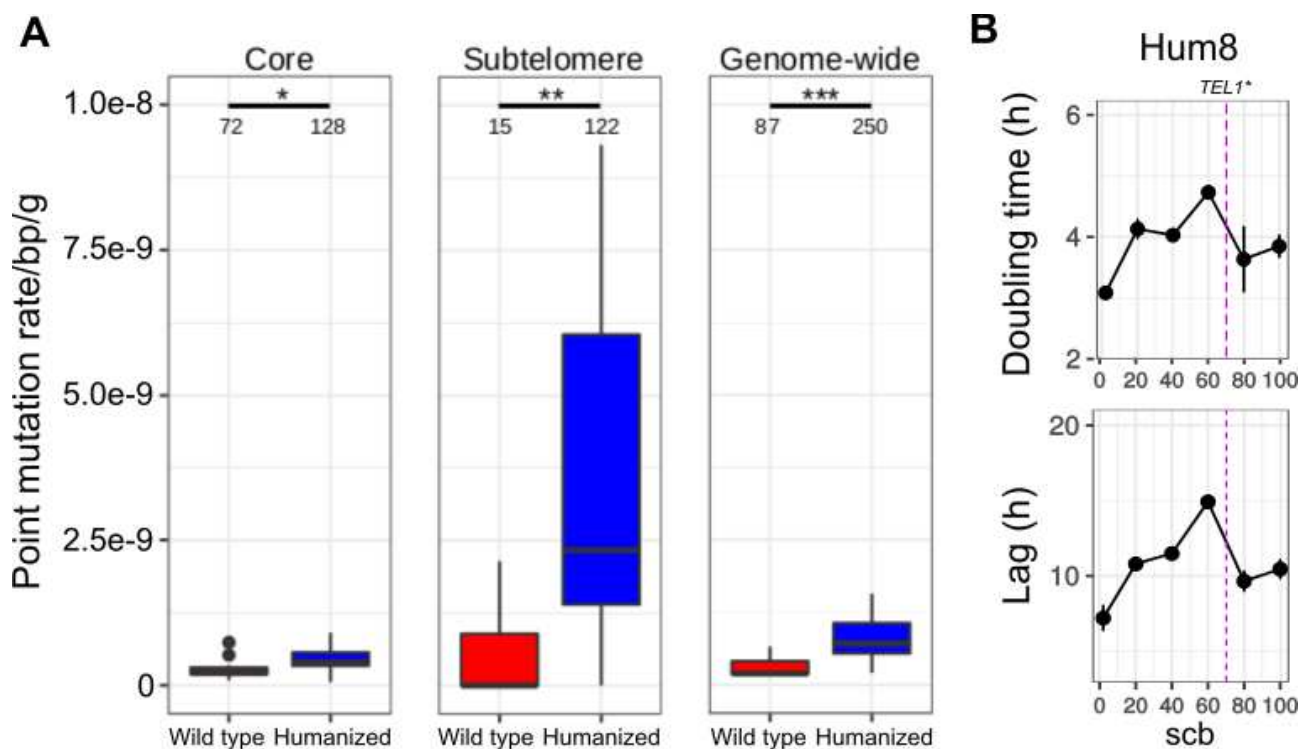


Figure 4 – Effect of telomere humanization on mutation rate – a, Point mutation rate per basepair per generation in WT and humanized yeasts ($n=11$ and $n=16$, respectively), subdivided in core genome (10761072 bp), subtelomeres (1310254 bp) and genome-wide (core + subtelomere, 12071326 bp). Numbers on top indicate the number of mutations occurred in each group. Boxplots as in Fig. 1a. ns=non-significant, *: $p<0.05$, **: $p<0.01$, ***: $p<0.001$. b, Growth performance (population doubling time and lag duration) of the humanized line carrying a mutation in *TEL1* (Hum8), over the course of the MAL experiment. The pink dashed line indicates the time interval of occurrence of the *TEL1* mutation and it is compatible with the recovery of fitness observed in this line. The experiment was performed in triplicate.

Adaptation to telomere sequence variation passes through DDR inactivation

After characterizing the effects of telomere humanization on cellular fitness and genomic landscape, we carried out an adaptive evolution experiment to study whether humanized yeasts were able to recover their fitness and adapt to telomere sequence variation. We evolved 5 parallel populations of humanized yeasts and 2 WT resulting from the MAL protocol for 31 serial transfers (ST) in rich liquid media. Among the humanized lines, we included the one carrying a mutation in *TEL1*. In this experimental setting, spontaneous mutations that increase cellular fitness will be positively selected and spread within the population. We measured the growth performance before and after the ST protocol to check if there was fitness recovery. The spotting assay, population doubling time and lag duration values evidenced a clear fitness recovery in 3 of the humanized lines, which reached the same level of the ancestors at the beginning of the MAL protocol. By contrast, the 2 WT and the 2 other humanized lines did not recover fitness during the ST protocol, in line with the “diminishing return” principle (Wei & Zhang, 2019). The *TEL1* line did not recover any fitness, as it already started with a fitness comparable to the WT (**Figure 5**).

We whole-genome sequenced the 7 lines at the end of the ST protocol and compared them with their ancestors at the beginning of ST/end of MAL to characterize the genomic changes responsible for the fitness recovery. We did not find any changes in telomere length, ITS content and Y' CN respect to the beginning of the ST. WT lines kept having longer telomeres than humanized ones, even after the ST protocol.

We analysed the genome in search of mutations that occurred during ST and found only a limited number of them, in line with our expectations. Interestingly, the three humanized lines which completely recovered their fitness carried mutations in components of the MRX complex (*MRE11:K477I*, *MRE11:Q447P*, *RAD50:L171P*), another effector of the DDR acting upstream of *TEL1/ATM* (**Figure 5**) (Gobbini et al., 2016). These mutations were fixed in the population, underlining their strong selection. They caused an aminoacid substitution in a conserved region of the resulting proteins and were predicted to have an important impact. None of the other mutations occurred in the ST protocol reached fixation or was predicted to have an impact on the protein, further supporting that these 3 mutations were under strong selection and accounted for the fitness recovery in their respective lines.

In conclusion, the adaptive evolution by ST experiment confirmed what was already evident from the MAL experiment: humanized yeasts need to inactivate the DNA damage checkpoint in order to survive in the long term and adapt to the variation in the telomeric sequence. Other genomic modifications might participate to the evolutionary adaptation, but further work is needed to dissect their contribution and will be carried out during 2021.

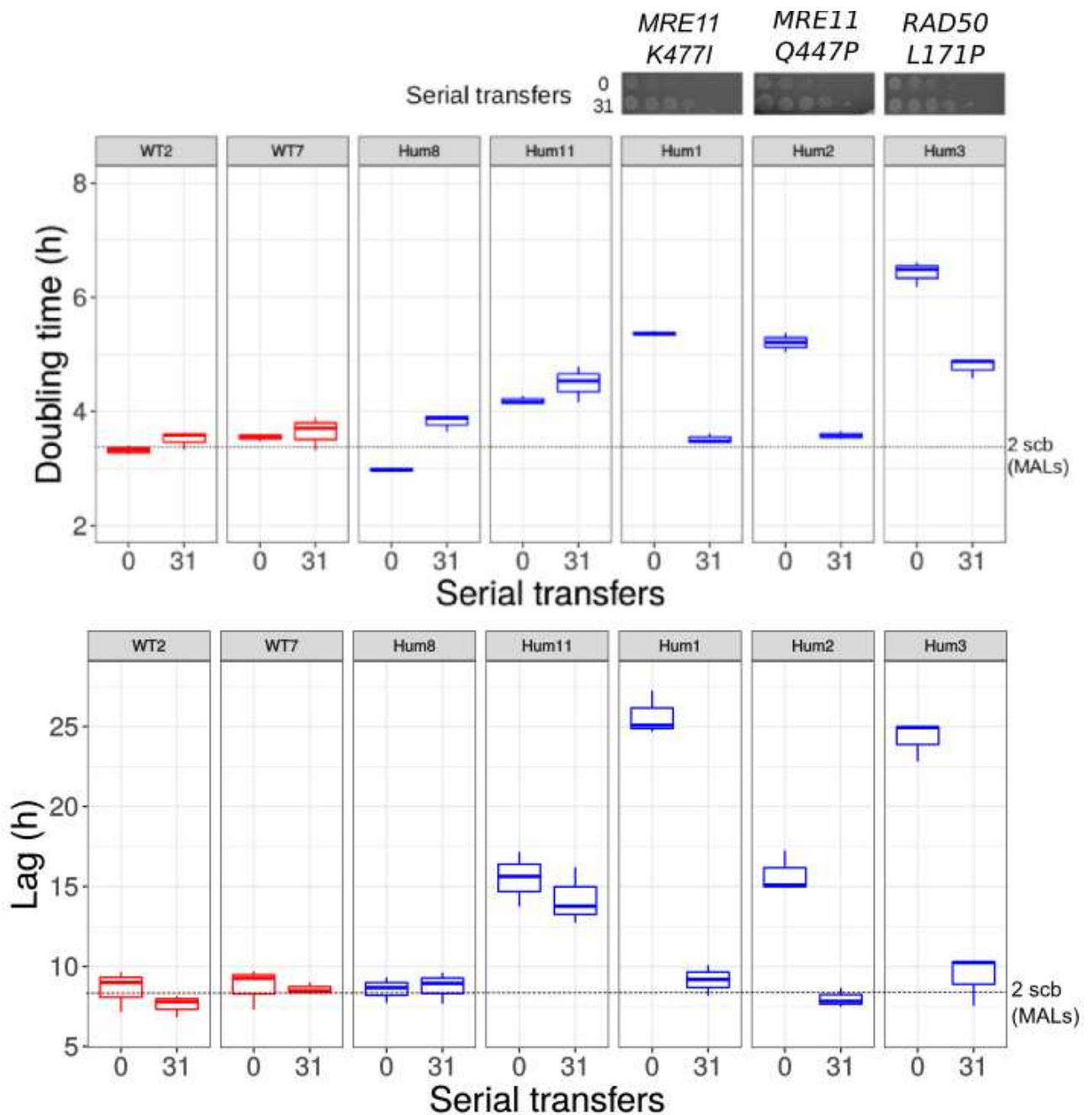
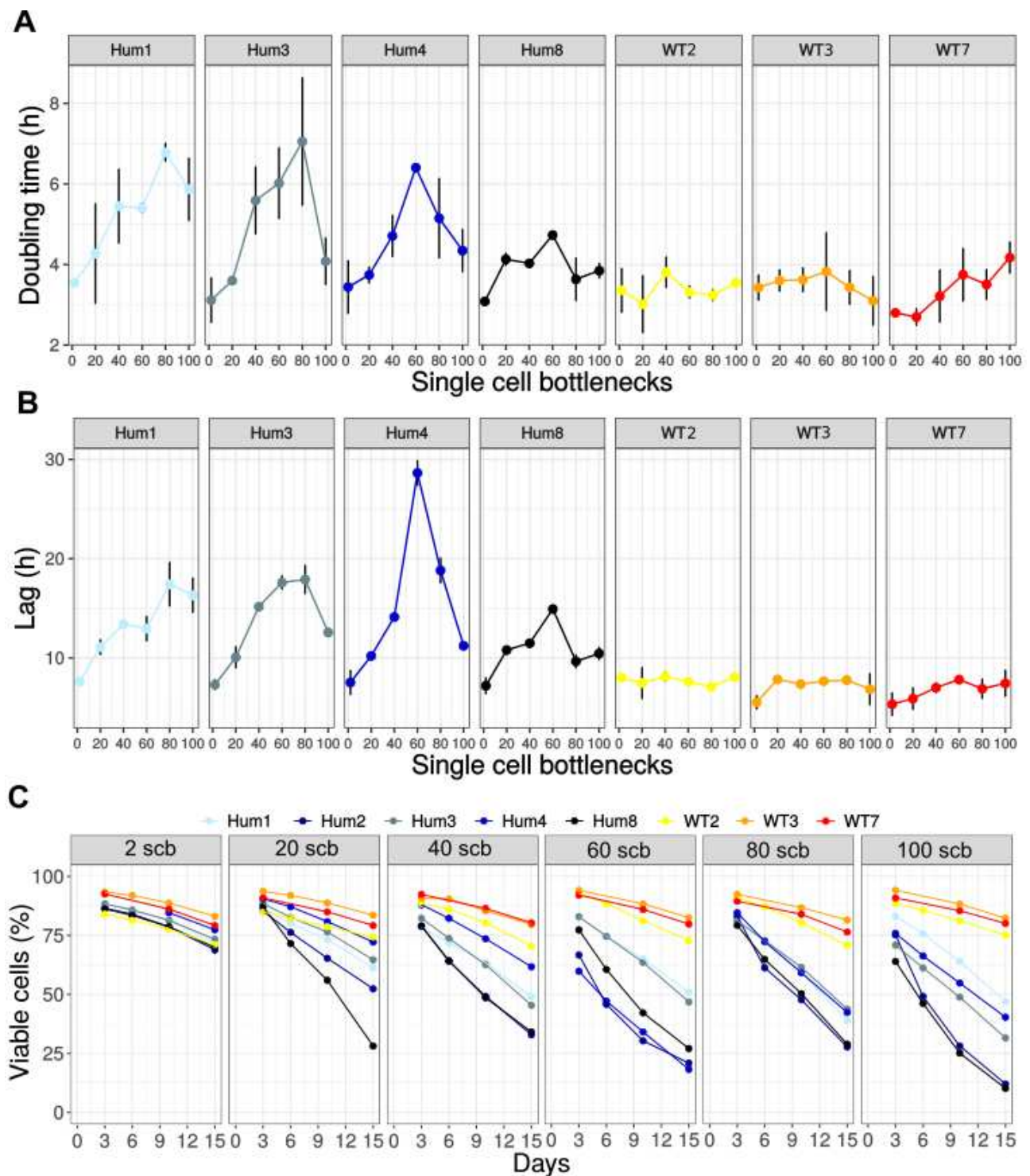
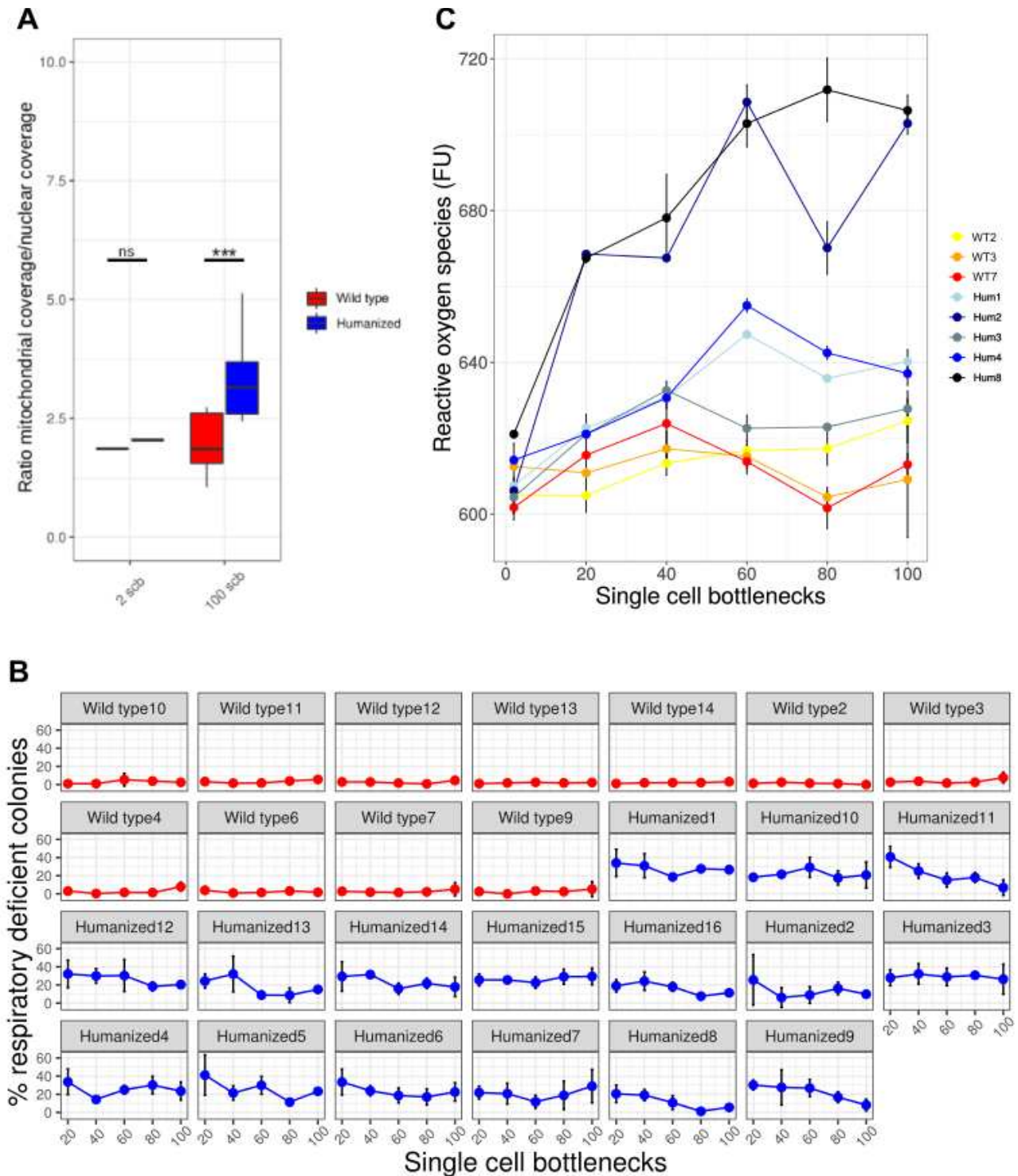


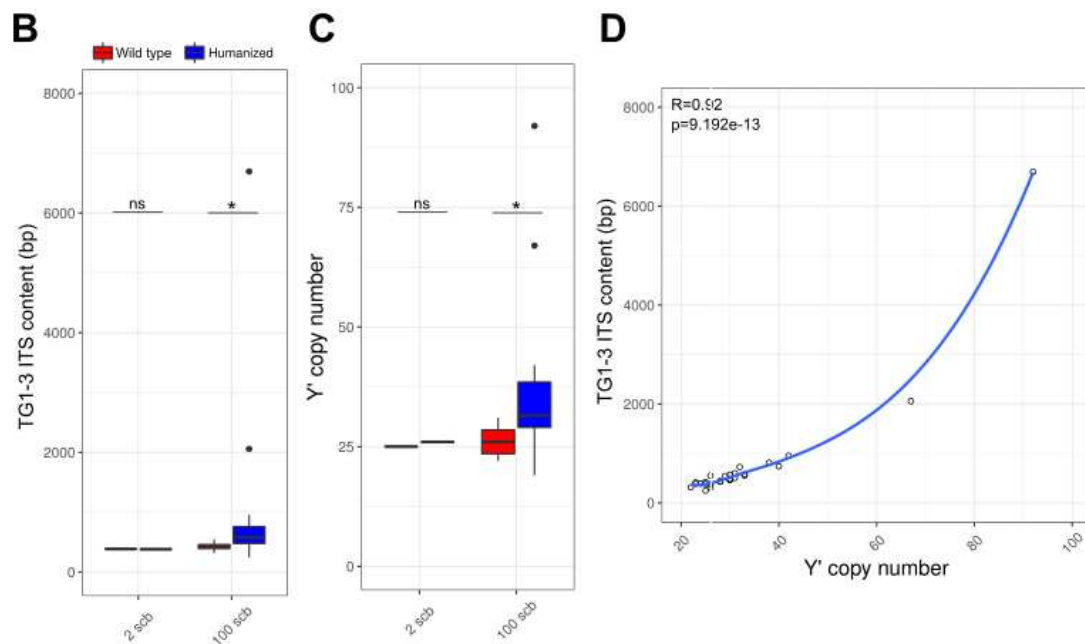
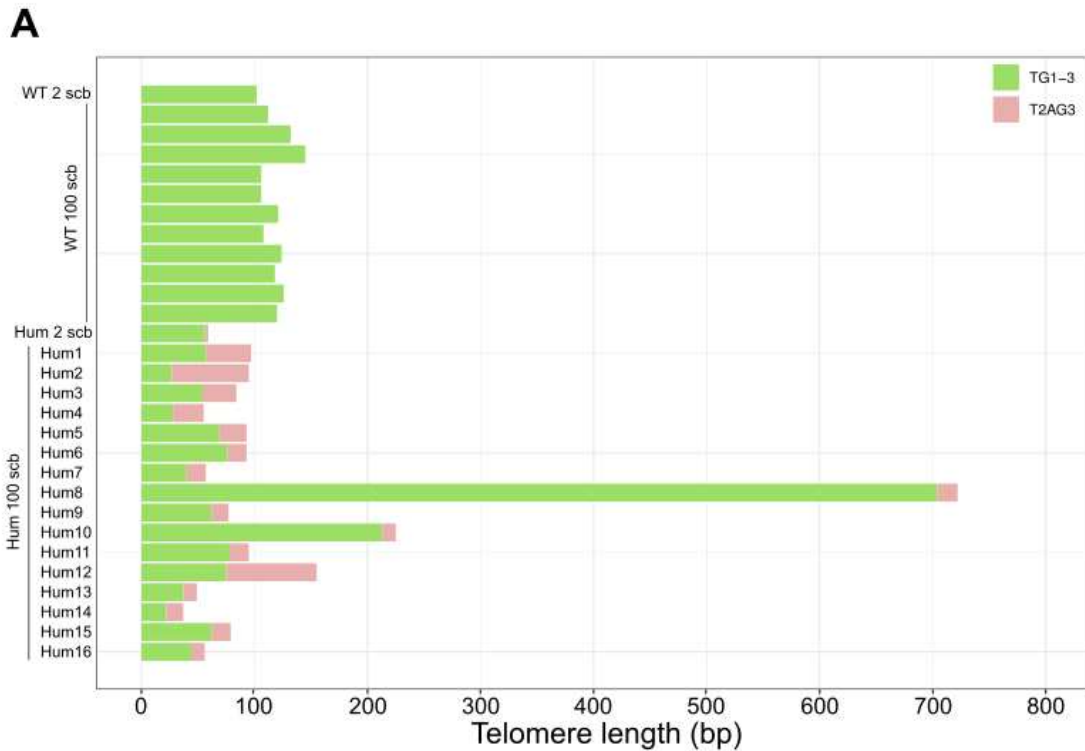
Figure 5 – Adaptive effect of DDR mutations. Population doubling time (upper panel) and lag phase duration (bottom panel) of WT and humanized lines before (0) and after (31) the ST protocol. Boxes as in Figure 1a. The experiment was performed in triplicate. The grey dashed line represents the doubling time and lag duration at the beginning of the MAL protocol. Data denote a complete recovery of fitness in 3 out of the 5 humanized lines, which reach the same values of WT. Pictures on top of the plot represent spotting assays of the 3 humanized lines that showed a fitness recovery. Upper and bottom series respectively represent lines before and after the ST protocol, and each spot represents an increasing dilution factor of 10 from left to right. Annotations on top of the spotting assays indicate the gene experiencing the mutation (either *MRE11* or *RAD50*) and its impact on the protein.



Extended Data Figure 1 – a, Population doubling time (average \pm SD) of 3 representative WT and 4 representative humanized lines at regular intervals during the MAL experiment. The experiment was performed in triplicate. **b**, Lag phase duration (average \pm SD) of 3 representative WT and 4 representative humanized lines at regular intervals during the MAL experiment. The experiment was performed in triplicate. **c**, Percentage of surviving cells (average \pm SD) for 3 representative WT and 5 representative humanized lines at regular intervals during the MAL experiment during a 15-days chronological lifespan experiment. The experiment was performed in triplicate. All the panels highlight a partial recovery of fitness starting from 60 scb.

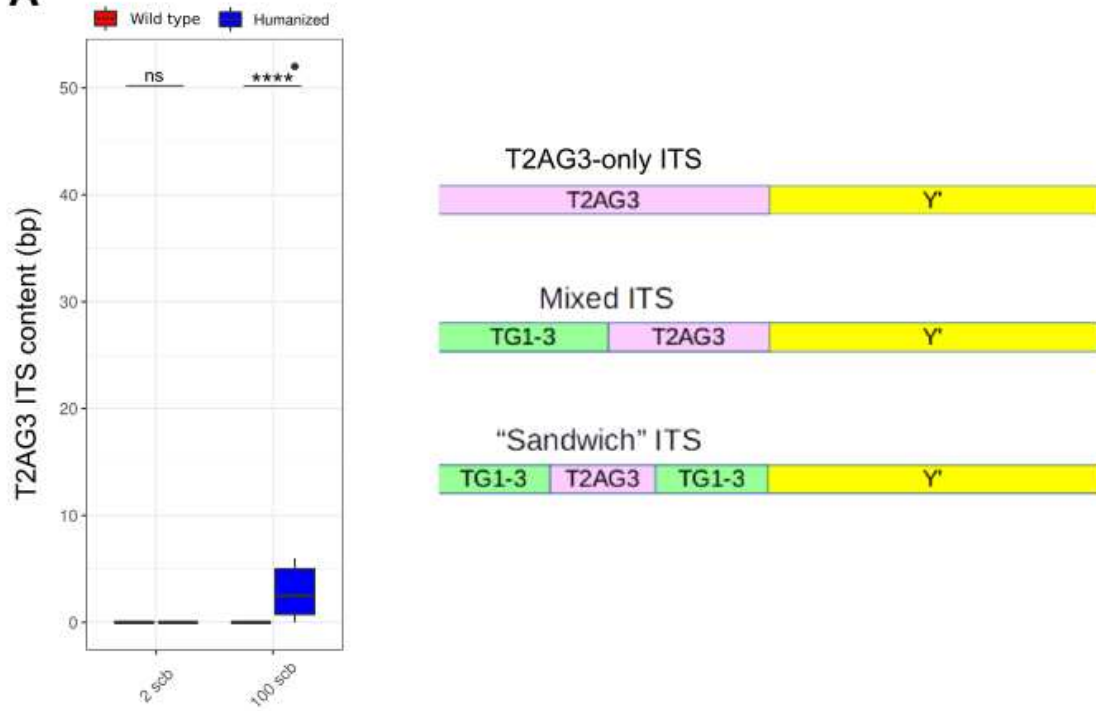
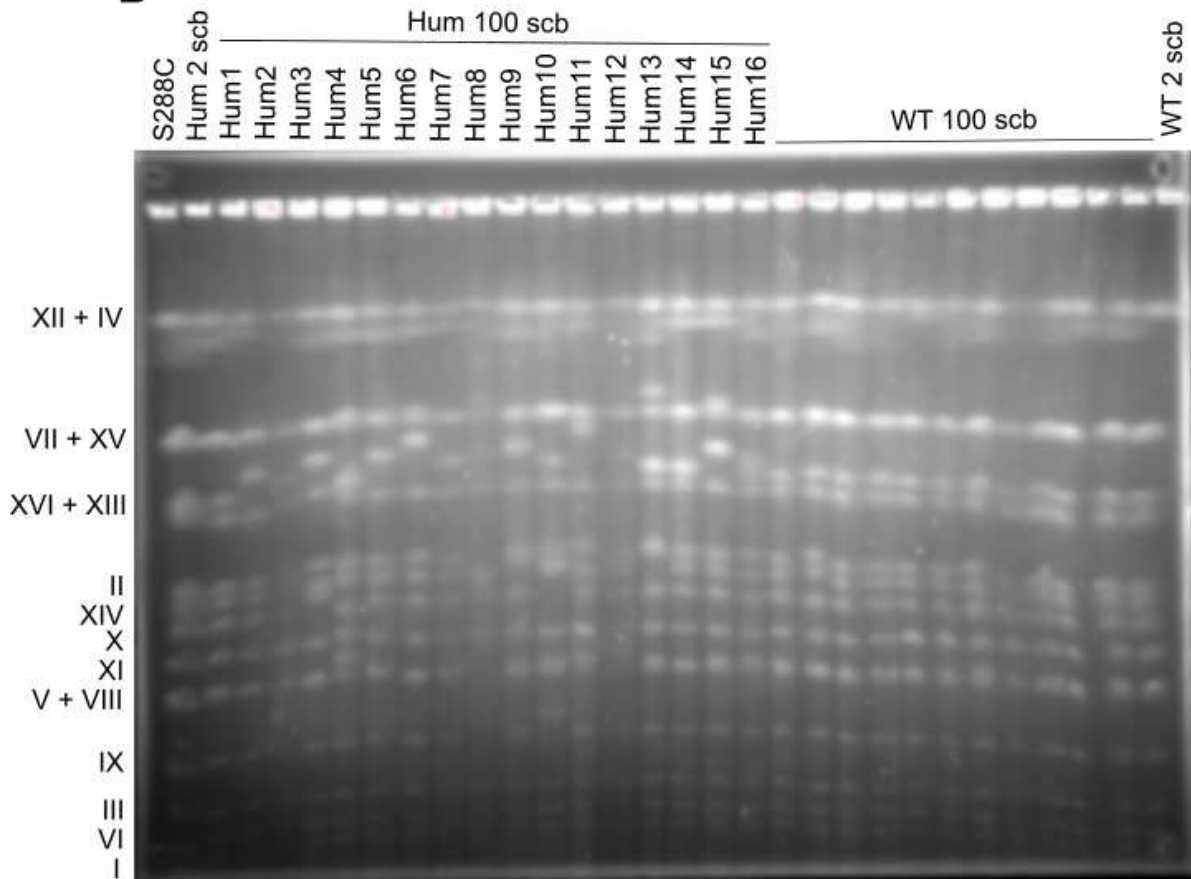


Extended Data Figure 2 – **a**, Estimation of the copy number of mtDNA in WT and humanized lines at the beginning (2 scb) ($n=1$ and $n=1$, respectively) and at the end (100 scb) ($n=11$ and $n=16$, respectively) of the MAL protocol. Boxplots as in Figure 1a. ns: non-significant; ***: $p < 0.001$. **b**, Percentage of respiratory-deficient (*petites*) colonies in YPG medium for all the WT and humanized lines across regular intervals during the MAL protocol. **c**, Content of reactive oxygen species (average \pm SD) of 3 representative WT and 5 representative humanized lines at regular intervals during the MAL experiment, measured in fluorescence units (FU). The experiment was performed in triplicate.



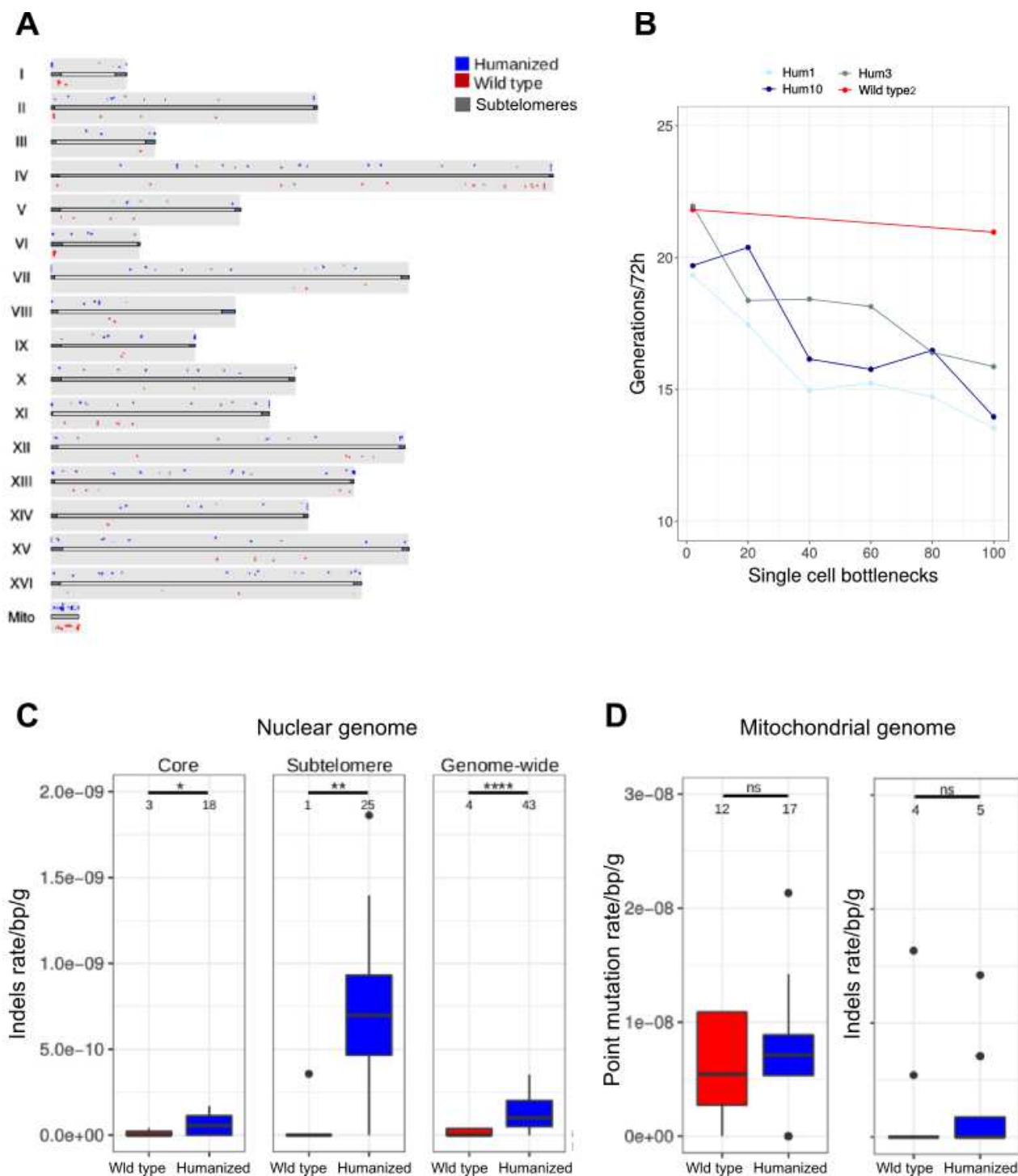
Extended Data Figure 3 – Chromosome-end structure in WT and humanized yeasts - a, Telomere length estimation in each WT and humanized line at the beginning (2 scb, 2 representatives) and at the end (100 scb, all the lines) of the MAL protocol. Telomere length estimations are divided in two categories: TG1-3 (green) and T2AG3 (pink). The order of telomeric repeats reflects telomere structure. **b,** Estimation of ITS content (TG1-3 repeats only) in WT and humanized lines at the beginning (2 scb) ($n=1$ and $n=1$, respectively) and at the end (100 scb) ($n=11$

and $n=16$, respectively) of the MAL protocol. Boxplots as in Figure 1a. ns: non-significant; *: $p<0.05$. **c**, Estimation of Y' copy number in WT and humanized lines at the beginning (2 scb) ($n=1$ and $n=1$, respectively) and at the end (100 scb) ($n=11$ and $n=16$, respectively) of the MAL protocol. Boxplots as in Figure 1a. ns: non-significant; *: $p<0.05$. **d**, Correlation between TG₁₋₃ ITS content and Y' copy number across all the sequenced WT and humanized lines at the beginning (2 scb; $n=1$ and $n=1$, respectively) and at the end (100 scb; $n=11$ and $n=16$, respectively) of the MAL protocol ($n=29$, Pearson's $R=0.92$, $P=0.192e^{-13}$). The blue line represents an interpolation of the data.

A**B**

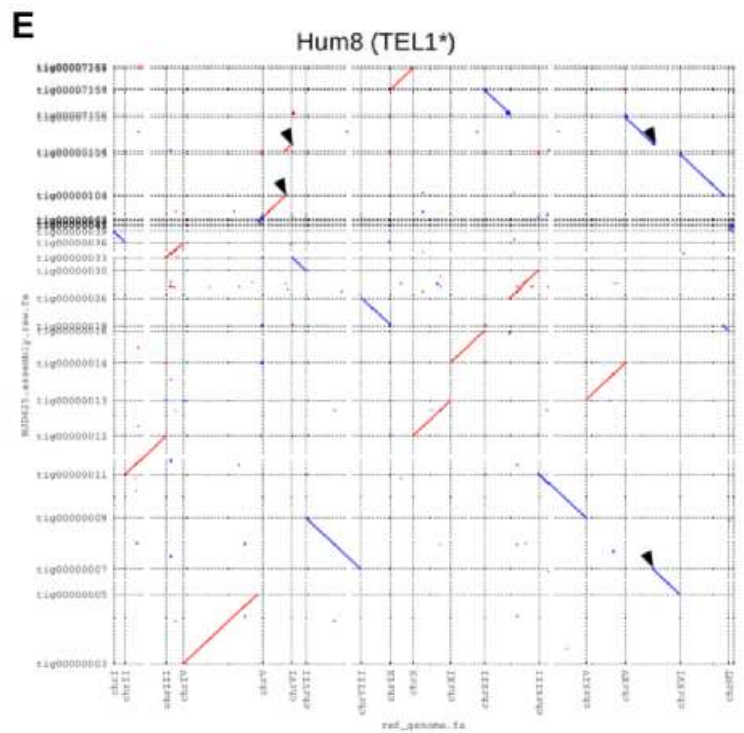
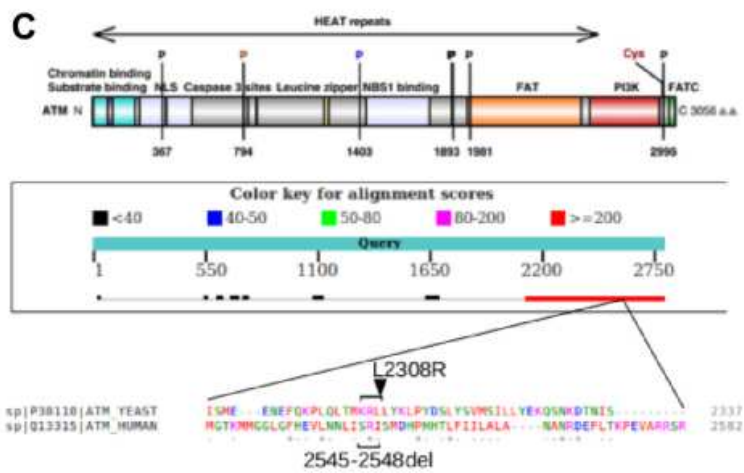
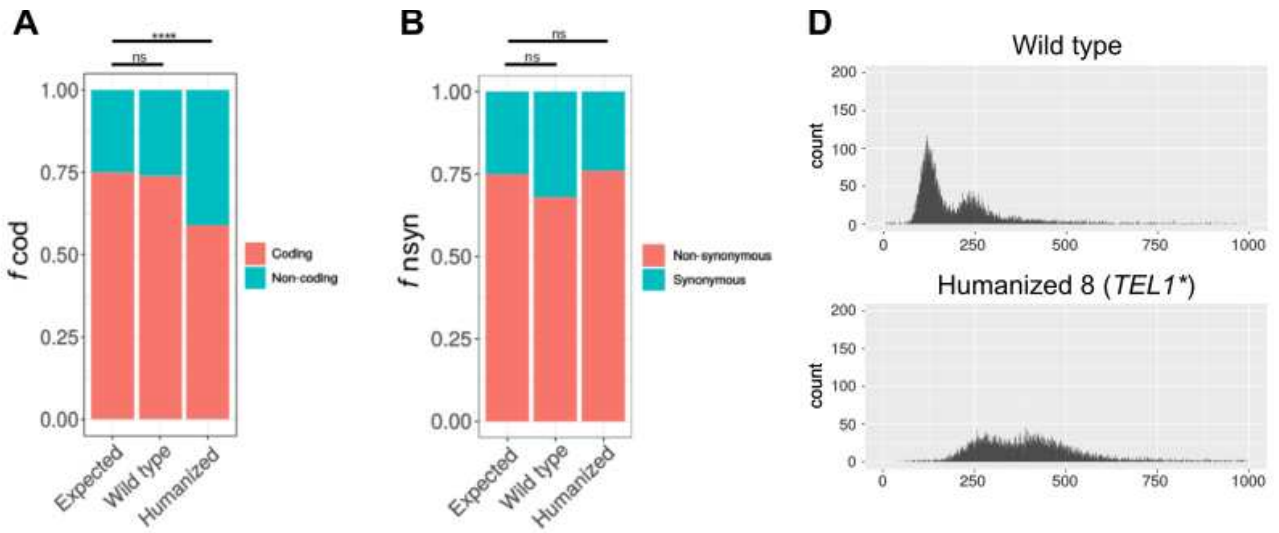
Extended Data Figure 4 – Telomere humanization and genome instability – a, Estimation of ITS content (T2AG3 repeats only) in WT and humanized lines at the beginning (2 scb) ($n=1$ and $n=1$, respectively) and at the end (100 scb) ($n=11$ and $n=16$, respectively) of the MAL protocol (left panel). Boxplots as in Figure 1a. ns: non-significant; *****: $p<0.0001$. Schematic illustration of the three types of structure of T2AG3-containing ITS, retrieved from the manual inspection of Illumina reads containing T2AG3 ITS repeats. The humanized outlier at 100 scb contains T2AG3-only ITS.

b, CHEF gel of WT and humanized lines at the beginning (2 scb) and at the end (100 scb) of the MAL protocol. Lanes contain, from left to right: 1) a S288C laboratory control strain, 2) a humanized line at 2 scb, 3-18) humanized lines at 100 scb, 19-29) WT lines at 100 scb, 30) a WT line at 2 scb. Karyotypes of the S288C control, WT 2 scb, Hum 2 scb and WT 100 scb are identical, while the humanized lines at 100 scb present multiple band shifts. Most of the shifts involve the band containing chromosome XVI, which increased its molecular weight in the majority of the humanized lines. The humanized line 8, carrying the *TEL1* mutation, shows an extensive pattern variation respect to the rest, compatible with multiple chromosomal rearrangements and chromosomal amplifications.



Extended Data Figure 5 – Telomere humanization and mutation rates – **a**, Genome-wide distribution of point mutations and insertions/deletions in all the humanized (upper line, blue spots) and WT (bottom line, red spots) yeasts. Genome coordinates are based on the *S. cerevisiae* SGD genome assembly. Dark grey areas at the chromosome tips represent subtelomeres and their boundaries are taken from (Yue et al., 2017). The presence of multiple spots on the same vertical line denotes a high density of mutations in a restricted area, and it is especially frequent in subtelomeres (for example chr VII L). **b**, Estimation of the number of generations performed at every single-cell bottleneck (~72 hours) across the entire MAL experiment, in one representative

WT line and 3 representative humanized lines. The number of generations performed is constant in WT lines, while it progressively decreases in humanized lines. **c**, Insertion/deletion rate per basepair per generation in WT and humanized yeasts ($n=11$ and $n=16$, respectively), subdivided in core genome (10761072 bp), subtelomeres (1310254 bp) and genome-wide (core + subtelomere, 12071326 bp). Numbers on top indicate the number of mutations occurred in each group. Only events shorter than 100 bp are considered for this analysis. Boxplots as in Fig. 1a. ns=non-significant, *: $p<0.05$, **: $p<0.01$, ****: $p<0.0001$. **d**, Point mutation rate and insertion/deletion rate per basepair per generation in WT and humanized yeasts ($n=11$ and $n=16$, respectively) across the mitochondrial genome (85779 bp). Numbers on top indicate the number of mutations occurred in each group. Boxplots as in Fig. 1a. ns=non-significant.



Extended Data Figure 6 – Characterization of the humanized line 8 (*TEL1)** – **a**, Fraction of point mutations occurring in coding vs non-coding regions in WT and humanized lines ($n=11$ and $n=16$, respectively), as compared to what is expected assuming a random distribution of mutation events (0.75 in coding regions). ns=non-significant, *****: $p<0.0001$. While WT do not deviate from the expected distribution, humanized yeasts carry less mutations than expected in coding regions, indicating the presence of selective pressure during the MAL experiment. **b**, Fraction of synonymous vs non-synonymous point mutations in WT and humanized lines ($n=11$ and $n=16$, respectively), as compared to what is expected assuming a random occurrence of mutation events (0.75 non-synonymous). ns=non-significant. Both WT and humanized do not deviate from the expected distribution. **c**, Schematic representation of the *TEL1* mutation in humanized line 8. The mutation involves an aminoacid substitution, from leucine (L) to arginine (R), at position 2308 of the protein. The substitution is predicted to have important effects on the protein, as it is located in a very conserved site, the FAT domain. The alignment of the complete Tel1 protein with its human ortholog ATM shows that the C-terminal part of the protein, where the mutation is located, is very conserved from yeast to humans. A deletion of 3 aminoacids surrounding our site in the human ortholog has been associated to ataxia telangiectasia (the main disease resulting from ATM dysfunction) and other severe human pathologies, supporting that our *TEL1* mutation might cause the loss of function of the Tel1 protein. **d**, Histogram showing the number of cells in the G1 (left peak) or G2 (right peak) phase of the cell cycle. The upper panel shows a representative plot for a WT yeast at 100 scb, while the bottom panel shows the same plot for humanized line 8 at 100 scb, carrying the *TEL1* mutation. The peaks are shifted to the right in the bottom panel, indicating an increased propidium-iodide fluorescence compatible with a doubled DNA content in this line. **e**, Comparison of the *S. cerevisiae* reference genome (x-axis) and our humanized line 8 assembly (y-axis). Sequence homology signals are indicated in red (forward match) or blue (reverse match). Black arrows indicate breakpoints of chromosomal rearrangements in the humanized line 8 genome assembly.

Methods

Yeast strains used in this study

The two initial yeast strains (one for humanized and one for wild type lines) to which the MAL protocol was applied derive from (Bah et al., 2004) and their genotype is: RWY12 (*Mata*, *ura3-52*, *lys 2-801*, *ade2-101*, *trp1-Δ1*, *his3-Δ200*, *leu2-Δ1*, *tlc1Δ::LEU2*, VR-ADE2-T). The replacement of the *TLC1* gene, encoding for the telomerase RNA template, ensures that these yeasts cannot use their native RNA to elongate their telomeres. In addition to these genetic modifications, the two initial strains contain two plasmids: *paz1* and *pTLC1TRP*. The *paz1* plasmid contains the wild type version of the *TLC1* gene and the selection marker *URA3*, conferring the ability to synthesize uracil and thus to grow on media lacking this amino acid. The *pTLC1TRP* plasmid contains the *TLC1* gene and the selection marker *TRP*, conferring the ability to synthesize tryptophan and thus to grow on media lacking this amino acid. The *TLC1* gene present in the *pTLC1TRP* plasmid is different in the two initial strains: in the initial strain used to derive WT lines, the *pTLC1* gene is the native yeast one; in the initial strain used to derive humanized lines, the *pTLC1* gene is almost identical to the native yeast one, except for a modification in the part used as template for telomere elongation (5'---CACCACACCCACACAC---3'), which has been replaced by a humanized version (5'---CUAACCCU---3'). Overall, both initial strains have the same genotype and contain the same two plasmids (*paz1* and *pTLC1TRP*), but the strain used to derive humanized lines contains a humanized *TLC1* gene in its *pTLC1TRP* plasmid. The presence of the *paz1* plasmid carrying a native *TLC1* gene in both initial strains ensures that a functional telomerase activity was maintained during the construction of the engineered strain.

Mutation accumulation lines

Prior to the beginning of the MAL protocol, we ensured to get rid of the *paz1* plasmid in both initial strains, in order to have telomerase rely only on the *TLC1* copy present in the *pTLC1TRP* plasmid, which is different in the two strains. To do this, we plated the two initial strains on media containing 5-FOA, a chemical which kills cells with a functional *URA3* gene, and we retrieved surviving colonies which had lost the *URA3*-containing *paz1* plasmid. At this point, we started our MAL protocol, and since two streaking-for-single were performed during the *paz1* exclusion procedure, we dated this timepoint as 2 single-cell bottlenecks.

We derived 16 parallel lines from the humanized yeast initial strain, and 16 lines from the WT initial strain. 5 of the WT lines became “petites” at different timepoints during the MAL protocol and were not characterized further in the study. MALs were propagated from each parental background on Leucin/Tryptophan dropout solid medium. The use of a leucine-deprived medium ensures to maintain only cells which kept the auxotrophic *LEU2* marker, conferring the ability to

synthesize leucine and thus to grow on media lacking this amino acid, inserted in the native *TLC1* locus in the nuclear genome. At the same time, the use of a tryptophan-deprived medium ensures to maintain only cells which kept their pTLC1TRP plasmid. Cells were passed through a single-cell bottleneck every ~72 h (~20 generations) at 30 °C, for a total of 100 bottlenecks. Samples were collected and frozen every 20 bottlenecks. At each single-cell bottleneck, a random colony was streaked to isolate the next single colony. To avoid any involuntary selection, at each streak, the closest colony to the center of the plate was picked, independently of its size.

To determine the number of generations passed after 72 h, three colonies for the humanized background and one colony for the WT background were independently resuspended in 100 µl of sterile water and serially diluted. Twenty microliters of each dilution were plated on solid YPD medium and grown for ~72 h at 30 °C. The number of colonies was manually counted in the plate with a suitable dilution and the number of generations (G) was estimated according to $G = \log_2(n \times d)$, where n is the number of cells counted on the plate and d is the corresponding dilution factor. This procedure was performed at all timepoints during MAL for the representatives of the humanized background, while it was performed only at the initial (2 scb) and final (100 scb) for the representative of the WT background.

Estimation of the growth performance

Strains have been incubated two nights in 200 µL in liquid Leu/Trp dropout medium in a non-shaking incubator at 30 °C. The day after, strains were mixed and 2 µl of the mixed culture was resuspended in 200 µL of fresh Leu/Trp dropout liquid medium in 96-well plate. The plate was incubated in a TECAN plate reader machine (Tecan, Infinite F200 Pro) at 30 °C for 72 hours and optical density (600 nm) was measured every 15 minutes during the experiment. Final results were analysed using the software PRECOG (Fernandez-Ricaud et al., 2016).

Chronological lifespan estimation

Strains were grown two nights in liquid Leu/Trp dropout medium, diluted 100x in 200 µL of fresh Leu/Trp dropout in a 96-well plate and incubated at 30 °C for 15 days. We measured viability at regular timepoints during the experiment. At each time point, 5 µL of cells were transferred in 100 µL of staining solution (Phosphate-buffered saline + 3 µM propidium iodide + 200 nM YO-PRO-1) in a 96-well plate and incubated 10 minutes in the dark at 30 °C. Cell viability was measured by high-throughput flow cytometry on a FACS-Calibur using the HTS module, as previously described (Barré et al., 2020). Cells were excited with the 488 nm laser and fluorescence was read with FL-1 and FL-3 filters, corresponding to YO-PRO-1 and propidium iodide fluorescence, respectively. Non-fluorescent cells were considered viable whereas fluorescent ones were considered dead. We

measured each sample three times. Flow cytometry data were subsequently analysed using the R package *flowCore* (Hahne et al., 2009).

Ploidy estimation

Ploidy was estimated for strains at 2 and 100 scb using a propidium iodide (PI) staining assay. For the humanized line 8, carrying the *TEL1* mutation, ploidy was estimated at all timepoints during the MAL protocol. Cells were incubated in 100 μ L of Leu/Trp dropout liquid media in 96-well plates and left two nights at 30 °C. After that, 3 μ L were taken and resuspended in 100 μ L of cold 70% ethanol for 3 h at 4 °C, washed twice with PBS, and resuspended in 100 μ L of staining solution (15 μ M PI, 100 μ g/ml RNase A, 0.1% v/v Triton-X, in PBS) and finally incubated for 3 h at 37 °C in the dark. Samples were analysed on a FACS-Calibur flow cytometer using the HTS module for processing 96-well plates. Cells were excited at 488 nm and fluorescence was collected with a FL2-A filter. The distributions of both FL2-A and FSC-H values have been processed to find the two main density peaks, which correspond to the two cell populations in G1 and G2 phase of the cell cycle, as described in (Peter et al., 2018).

Mitochondrial phenotyping

Cells were incubated in 100 μ L of Leu/Trp dropout liquid media in 96-well plates and left two nights at 30 °C. After that, 10 μ L of cells were transferred into 190 μ L of a HEPES 10 mM-glucose 2% solution. Samples were washed twice with HEPES-glucose and finally resuspended in 100 μ L of mitochondrial staining solution (HEPES-glucose + 100 nM MitoTracker Green (Molecular Probes) + 20 nM MitoTracker Deep Red (Molecular Probes)) in a 96-well plate and incubated 30 min in the dark at room temperature. MitoTracker Green is known to accumulate passively into the mitochondria proportionally to their content, whereas MitoTracker Deep Red is known to accumulate into mitochondria proportionally to their membrane potential. The samples were analysed on a FACSCalibur using the HTS module. The relative mitochondrial volume was estimated by reading fluorescence in the FL-1 channel, whereas the relative mitochondrial activity was estimated by reading fluorescence in the FL-4 channel, as described in (De Chiara et al., 2020).

ROS estimation

ROS content was measured using the fluorescent molecule dihydroethidium (DHE). DHE can permeate cellular membranes and be oxidized by $O_2^{\bullet-}$, producing fluorescent derivatives. Cells were incubated in 100 μ L of Leu/Trp dropout liquid media in 96-well plates and left two nights at 30 °C. After that, 10 μ L of cells were transferred into 190 μ L of a HEPES 10 mM-glucose 2% solution. Samples were washed twice with HEPES-glucose and finally resuspended in 100 μ L of

DHE staining solution (HEPES-glucose + 800 μ M DHE) in a 96-well plate and incubated 30 min in the dark at room temperature. After that, the staining solution was removed and replaced by 100 μ L HEPES-glucose. Samples were analysed on a FACSCalibur using the HTS module. The relative ROS content was estimated by reading fluorescence in the FL-2 channel.

Estimation of the tendency to form “petites”

Yeast strains are incubated two nights in 200 μ L of Leu/Trp dropout liquid medium at 30 °C. After that, serial dilutions are prepared and plated on YPG solid medium (0.1% glucose, 3% glycerol, 1% yeast extract, 2% peptone, 2% agar). Plates are then incubated at 30 °C for 7 days and colonies are counted from the most suitable plate. The proportion of petites colonies is estimated as the number of petites colonies vs the total number of colonies in the plate. Some of the putative “petites” colonies are then re-streaked on YPEG media (3% glycerol, 1% yeast extract, 2% peptone, 2% agar) to confirm their phenotype.

Sequencing

After 100 single-cell bottlenecks, cells were inoculated in 5-ml liquid Leu/Trp dropout cultures and grown overnight at 30 °C in a shaking incubator. DNA was extracted using “Yeast Masterpure” kit (Epicentre, USA) following the manufacturer’s instructions. Illumina paired-end libraries (2 \times 150 bp) were prepared according to manufacturer’s standard protocols and sequenced with an HiSeq 2500 instrument, at the NGS platform of Institut Curie. For the MAL experiment, two initial strains at 2 scb (one with the humanized background and one with the wild-type background) and all the strains at 100 scb were sequenced. For the ST experiment, all the 5 humanized and 2 WT lines at 31 ST were sequenced.

Telomere length estimation

Telomere length, ITS content and copy number of Y' elements were determined using the bioinformatic pipeline described in chapter 5 of this thesis. Since all samples in this study derived from the same strain background (RWY12), we reasoned that the background noise would not be so high and thresholds for repeat detection were decreased accordingly. We used a threshold of 20 bp of telomeric stretch length to estimate telomere length and ITS content with the TG₁₋₃ repeat, and a threshold of 12 bp (corresponding to 2 humanized repeats) with the T₂AG₃ repeat. Differently from what described in chapter 5, we did not observe any underestimation in our results and consequently did not apply any mathematical adjustments.

Genomic analyses of MALs

We mapped Illumina reads resulting from MALs on the *S. cerevisiae* SGD genome assembly using BWA (v0.7.12) (Li & Durbin, 2009). Read mapping runs were stored in the Sequence Alignment/Map (SAM) format and its binary form (BAM) and post-processing steps were performed by picard tools (v2.8.0). Sorting, indexing and per-position coverage calculation were performed using SAMtools (v1.2) (Li et al., 2009). Chromosomes with doubled coverage values (~120) respect to the genome-wide median (~60) across their entire length were considered as aneuploidies. The copy number of mitochondrial DNA was estimated as described in (De Chiara et al., 2020).

We used the software freebayes (v0.9.5) to call markers from the BAM files. Results were stored in Variant Call Format (VCF) files. We filtered the VCF files to keep only SNVs/indels with quality higher than 20 and coverage higher than 10. We subtracted the SNVs/indels present in the sequenced initial humanized strain at 2 scb from its derived mutation accumulation lines, and the SNVs/indels present in the sequenced initial WT strain at 2 scb from its derived lines, using the vcf-isec program included in the suite VCFtools (v0.1.16). Next, we searched for shared SNVs present in at least two mutation accumulation lines and removed them from the VCF files, based on the assumption that identical de novo mutations are highly unlikely to arise independently and derived from incorrect mapping or poor coverage. All the remaining SNVs/indels were manually inspected using IGV (v2.3.68) (Thorvaldsdóttir et al., 2013) and they were used to estimate point mutation and insertion/deletion rates (μ) as $\mu=(n/bp)/g$, where n is the number of de novo mutations present in each line, bp is the (haploid) genome size of the *S. cerevisiae* SGD genome assembly and g is the number of generations performed in the mutation accumulation experiment (2140 for WT lines and 1640 for humanized lines). de novo mutations were annotated using the *S. cerevisiae* SGD genome assembly implemented in the online software Variant Effect Predictor (VEP) (McLaren et al., 2016). Further annotation of the functional impact of mutations was carried out using mutfunc (Wagih et al., 2018).

GO term analysis

Standard GO term analysis was performed separately on the mutations extracted from WT and humanized lines with the GO Term Finder tool available at *Saccharomyces* Genome Database (SGD). Significant GO terms were extracted by the algorithm implemented in the tool, with a False Discovery Rate (FDR) corrected α threshold of 0.1.

Experimental validation of variants

Two de novo SNVs from independent humanized lines were validated by Sanger sequencing. One variant was located inside the gene *MEC1* and the other one was inside the gene *TEL1*. A pair of primers (upstream and downstream) was designed for each SNV using Unipro UGENE (Okonechnikov et al., 2012), and used to perform Polymerase Chain Reaction (PCR). We checked the presence of the mutations at all timepoints during the MAL protocol. PCR products were sequenced by Eurofins Genomics. The presence and the genotype of the variants were checked by visual inspection of the electropherograms. We confirmed the presence of both variants and discovered that the mutations in *TEL1* occurred between 60 and 80 scb, while the mutation in *MEC1* occurred between 40 and 60 scb. The two mutations occurred in separate humanized lines.

Long reads sequencing and structural variant analysis

Yeast cells from humanized line 8 were grown for two nights in liquid Leu/Trp dropout media at 30 °C. Genomic DNA was extracted using “Yeast Masterpure” kit (Epicentre, USA) according to the manufacturer's instructions. The MINION (Oxford Nanopore) sequencing library was prepared using the SQK-LSK109 sequencing kit according to the manufacturer's protocol. The library was loaded onto a FLO-MIN106 flow cell and sequencing was run for 72 hours. Long read basecalling and scaffolding were performed using the pipeline LRSDAY (v 1.6) (Yue & Liti, 2018) and the dotplot was generated using mummerplot (Marçais et al., 2018).

Serial transfers protocol

5 humanized lines and 2 WT lines derived from the MAL protocol were then re-evolved through a serial transfer protocol. Strains were inoculated in tubes containing 5 ml of Leu/Trp dropout liquid medium and incubated for two nights at 30 °C in a shaking incubator. Transfers were performed every 2 days by diluting 50 µl of the previous culture into 5 ml of fresh Leu/Trp dropout medium (1:100 dilution). We performed a total of 31 serial transfers. At the end of the protocol, DNA was extracted for the samples and sequenced as described in the paragraph “Mutation accumulation lines”.

Genomic analyses of STs

We mapped Illumina reads resulting from STs and extracted variants as described in the paragraph “Genomic analyses of MALs”. We filtered the VCF files to keep only SNVs/indels with quality higher than 20 and coverage higher than 5. We subtracted the SNVs/indels present in the corresponding mutation accumulation line at 100 scb from which each ST line was derived, using

the vcf-isec program included in the suite VCFtools (v0.1.16). Next, we searched for shared SNVs present in at least two STs and removed them from the VCF files, based on the assumption that identical de novo mutations are highly unlikely to arise independently and derived from incorrect mapping or poor coverage. All the remaining SNVs/indels were manually inspected using IGV (v2.3.68) (Thorvaldsdóttir et al., 2013). de novo mutations were annotated using the *S. cerevisiae* SGD genome assembly implemented in the online software Variant Effect Predictor (VEP) (McLaren et al., 2016). Further annotation of the functional impact of mutations was carried out using mutfunc (Wagih et al., 2018). Telomere length, ITS content and Y' copy number measurements were performed with the same methods and parameters described in the paragraph “Telomere length estimation”.

Spotting assay

Cells were incubated in Leu/Trp dropout liquid media for two nights at 30 °C. After that, they were serially diluted in water. For each 10-fold dilution, 5 µL were spotted either on Leu/Trp dropout agar plates. Plates were then incubated 72 h at 30 °C, and images were acquired with a Epson Perfection V330 scanner.

References

- Alexander, M. K., & Zakian, V. A. (2003). Rap1p telomere association is not required for mitotic stability of a C₃TA₂ telomere in yeast. *EMBO Journal*, 22(7): 1688–1696.
- Bah, A., Bachand, F., Clair, É., Autexier, C., & Wellinger, R. J. (2004). Humanized telomeres and an attempt to express a functional human telomerase in yeast. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkh511>
- Bah, A., Gilson, E., & Wellinger, R. J. (2011). Telomerase is required to protect chromosomes with vertebrate-type T2AG3' ends in *Saccharomyces cerevisiae*. *Journal of Biological Chemistry*. <https://doi.org/10.1074/jbc.M111.220186>
- Barré, B. P., Hallin, J., Yue, J. X., Persson, K., Mikhalev, E., Irizar, A., Holt, S., Thompson, D., Molin, M., Warringer, J., & Liti, G. (2020). Intragenic repeat expansion in the cell wall protein gene HPF1 controls yeast chronological aging. *Genome Research*, 30(5), 697–710. <https://doi.org/10.1101/gr.253351.119>
- Brevet V., Berthiau A., Civitelli L., Donini P., Schramke V., Geli V., Ascenzioni F., Gilson E. (2003). The number of vertebrate repeats can be regulated at yeast telomeres by Rap1-independent mechanisms. *EMBO Journal*. 22(7): 1697–1706.
- Chan, S. W. L., & Blackburn, E. H. (2002). New ways not to make ends meet: Telomerase, DNA damage proteins and heterochromatin. *Oncogene*, 21(4 REV. ISS. 1), 553–563. <https://doi.org/10.1038/sj/onc/1205082>
- Chen, X. J., & Clark-Walker, G. D. (1999). The petite mutation in yeasts: 50 years on. *International Review of Cytology*, 194(February 2000), 197–238. [https://doi.org/10.1016/s0074-7696\(08\)62397-9](https://doi.org/10.1016/s0074-7696(08)62397-9)
- De Chiara, M., Friedrich, A., Barré, B., Breitenbach, M., Schacherer, J., & Liti, G. (2020). Discordant evolution of mitochondrial and nuclear yeast genomes at population level. *BMC Biology*, 18(1), 1–15. <https://doi.org/10.1186/s12915-020-00786-4>
- De Lange, T. (2005). Shelterin: The protein complex that shapes and safeguards human telomeres. *Genes and Development*, 19(18), 2100–2110. <https://doi.org/10.1101/gad.1346005>
- Di Domenico, E. G., Auriche, C., Viscardi, V., Longhese, M. P., Gilson, E., & Ascenzioni, F. (2009). The Mec1p and Tel1p checkpoint kinases allow humanized yeast to tolerate chronic telomere dysfunctions by suppressing telomere fusions. *DNA Repair*. <https://doi.org/10.1016/j.dnarep.2008.10.005>
- Di Domenico, E. G., Mattarocci, S., Cimino-Reale, G., Parisi, P., Cifani, N., D'Ambrosio, E., Zakian, V. A., & Ascenzioni, F. (2013). Tel1 and Rad51 are involved in the maintenance of telomeres with capping deficiency. *Nucleic Acids Research*, 41(13), 6490–6500. <https://doi.org/10.1093/nar/gkt365>
- Di Gregorio, A., Bowling, S., & Rodriguez, T. A. (2016). Cell Competition and Its Role in the Regulation of Cell Fitness from Development to Cancer. *Developmental Cell*, 38(6), 621–634. <https://doi.org/10.1016/j.devcel.2016.08.012>

- Fernandez-Ricaud, L., Kourtchenko, O., Zackrisson, M., Warringer, J., & Blomberg, A. (2016). PRECOG: A tool for automated extraction and visualization of fitness components in microbial growth phenomics. *BMC Bioinformatics*, *17*(1), 1–15. <https://doi.org/10.1186/s12859-016-1134-2>
- Finkel. (2011). Telomeres and Mitochondrial Function. *Circ Res*. 108(8): 903–904.
- Fulnečková, J., Ševčíková, T., Fajkus, J., Lukešová, A., Lukeš, M., Vlček, Č., Lang, B. F., Kim, E., Eliáš, M., & Sýkorová, E. (2013). A broad phylogenetic survey unveils the diversity and evolution of telomeres in eukaryotes. *Genome Biology and Evolution*, *5*(3), 468–483. <https://doi.org/10.1093/gbe/evt019>
- Gobbini, E., Cassani, C., Villa, M., Bonetti, D., & Longhese, M. P. (2016). Functions and regulation of the MRX complex at DNA double-strand breaks. In *Microbial Cell* (Vol. 3, Issue 8, pp. 329–337)
- Greider, C. W., & Blackburn, E. H. (1987). The telomere terminal transferase of tetrahymena is a ribonucleoprotein enzyme with two kinds of primer specificity. *Cell*, *51*(6), 887–898. [https://doi.org/10.1016/0092-8674\(87\)90576-9](https://doi.org/10.1016/0092-8674(87)90576-9)
- Greider, C. W., & Blackburn, E. H. (1989). A telomeric sequence in the RNA of Tetrahymena telomerase required for telomere repeat synthesis. *Nature*, *337*(6205), 331–337. <https://doi.org/10.1038/337331a0>
- Hahne, F., LeMeur, N., Brinkman, R. R., Ellis, B., Haaland, P., Sarkar, D., Spidlen, J., Strain, E., & Gentleman, R. (2009). flowCore: A Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics*, *10*. <https://doi.org/10.1186/1471-2105-10-106>
- Henning, K. A., Moskowitz, N., Ashlock, M. A., & Liu, P. P. (1998). Humanizing the yeast telomerase template. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(10), 5667–5671. <https://doi.org/10.1073/pnas.95.10.5667>
- Kupiec, M. (2014). Biology of telomeres: Lessons from budding yeast. In *FEMS Microbiology Reviews*. <https://doi.org/10.1111/1574-6976.12054>
- Kuznetsova. (2020). Telomere structure in insects: A review. *Journal of Zoological Systematics and Evolutionary Research*, *58*(1), 127–158. <https://doi.org/10.1111/jzs.12332>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lundblad, V., & Szostak, J. W. (1989). A mutant with a defect in telomere elongation leads to senescence in yeast. *Cell*, *57*(4), 633–643. [https://doi.org/10.1016/0092-8674\(89\)90132-3](https://doi.org/10.1016/0092-8674(89)90132-3)
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLoS Computational Biology*, *14*(1), 1–14. <https://doi.org/10.1371/journal.pcbi.1005944>

- Marcand, S., Gilson, E., & Shore, D. (1997). A protein-counting mechanism for telomere length regulation in yeast. *Science*, 275(5302), 986–990. <https://doi.org/10.1126/science.275.5302.986>
- McClintock, B. (1941). The Stability of Broken Ends of Chromosomes in *Zea Mays*. *Genetics*, 26(2), 234–282.
- McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, 17(1), 1–14. <https://doi.org/10.1186/s13059-016-0974-4>
- Murphy, M. P. (2009). How mitochondria produce reactive oxygen species. *Biochemical Journal*, 417(1), 1–13. <https://doi.org/10.1042/BJ20081386>
- Okonechnikov, K., Golosova, O., Fursov, M., Varlamov, A., Vaskin, Y., Efremov, I., German Grehov, O. G., Kandrov, D., Rasputin, K., Syabro, M., & Tleukenov, T. (2012). Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics*, 28(8), 1166–1167. <https://doi.org/10.1093/bioinformatics/bts091>
- Olovnikov, A. M. (1973). A theory of marginotomy. The incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon. *Journal of Theoretical Biology*, 41(1), 181–190. [https://doi.org/10.1016/0022-5193\(73\)90198-7](https://doi.org/10.1016/0022-5193(73)90198-7)
- Peska, V., & Garcia, S. (2020). Origin, Diversity, and Evolution of Telomere Sequences in Plants. *Frontiers in Plant Science*, 11(February), 1–9. <https://doi.org/10.3389/fpls.2020.00117>
- Peter, J., De Chiara, M., Friedrich, A., Yue, J.-X., Pflieger, D., Bergström, A., Sigwalt, A., Barre, B., Freel, K., Llored, A., Cruaud, C., Labadie, K., Aury, J.-M., Istace, B., Lebrigand, K., Barbry, P., Engelen, S., Lemainque, A., Wincker, P., ... Schacherer, J. (2018). Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates Species-wide genetic and phenotypic diversity. *Nature*. 556(7701):339-344.
- Ribaud, V., Ribeyre, C., Damay, P., & Shore, D. (2012). DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. *EMBO Journal*, 31(1), 138–149. <https://doi.org/10.1038/emboj.2011.349>
- Robin, J. D., Jacome Burbano, M. S., Peng, H., Croce, O., Thomas, J. L., Laberthonniere, C., Renault, V., Lototska, L., Pousse, M., Tessier, F., Bauwens, S., Leong, W., Sacconi, S., Schaeffer, L., Magdinier, F., Ye, J., & Gilson, E. (2020). Mitochondrial function in skeletal myofibers is controlled by a TRF2-SIRT3 axis over lifetime. *Aging Cell*, 19(3), 1–16. <https://doi.org/10.1111/accel.13097>
- Sahin, E., Colla, S., Liesa, M., Moslehi, J., Müller, F. L., Guo, M., Cooper, M., Kotton, D., Fabian, A. J., Walkey, C., Maser, R. S., Tonon, G., Foerster, F., Xiong, R., Wang, Y. A., Shukla, S. A., Jaskelioff, M., Martin, E. S., Heffernan, T. P., ... DePinho, R. A. (2011). Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature*, 470(7334), 359–365. <https://doi.org/10.1038/nature09787>
- Sharp, N. P., Sandell, L., James, C. G., & Otto, S. P. (2018). The genome-wide rate and spectrum of spontaneous mutations differ between haploid and diploid yeast. *Proceedings of the National*

Academy of Sciences of the United States of America, 115(22), E5046–E5055.
<https://doi.org/10.1073/pnas.1801040115>

- Teixeira, M. T., & Gilson, E. (2005). Telomere maintenance, function and evolution: The yeast paradigm. *Chromosome Research*, 13(5), 535–548. <https://doi.org/10.1007/s10577-005-0999-0>
- Thorvaldsdóttir, H., Robinson, J. T., & Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2), 178–192. <https://doi.org/10.1093/bib/bbs017>
- Vořechovský, I., Luo, L., Lindblom, A., Negrini, M., Webster, A. D. B., Croce, C. M., & Hammarström, L. (1996). ATM mutations in cancer families. *Cancer Research*, 56(18), 4130–4133.
- Vorechovsky, I., Luol, R. I. L., Poroni, L., Martin, J. S., Cetcvsky, D., Amlot, P. L., Yaxlef, J. C., Hammarstrorn, L., David, A. B., & Yuille, M. A. R. (1997). Clustering of missense mutations in the ataxia-telangiectasia gene in a sporadic T-cell leukaemia. *Nature Genetics*, 17(september), 96–99.
- Wagih, O., Busby, B., Galardini, M., Memon, D., Typas, A., & Beltrao, P. (2018). Comprehensive variant effect predictions of single nucleotide variants in model organisms. *Mol Syst Biol*. 14(12): e8430.
- Watson. (1972). *Nature New Biology*, 238, 37–38.
- Wei, X., & Zhang, J. (2019). Patterns and Mechanisms of Diminishing Returns from Beneficial Mutations. *Molecular Biology and Evolution*, 36(5), 1008–1021. <https://doi.org/10.1093/molbev/msz035>
- Wellinger, R. J., & Zakian, V. A. (2012). Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: Beginning to end. In *Genetics*. <https://doi.org/10.1534/genetics.111.137851>
- Wright, J., Teraoka, S., Onengut, S., Tolun, A., Gatti, R. A., Ochs, H. D., & Concannon, P. (1996). A high frequency of distinct ATM gene mutations in ataxia-telangiectasia. *American Journal of Human Genetics*, 59(4), 839–846.
- Yue, J. X., Li, J., Aigrain, L., Hallin, J., Persson, K., Oliver, K., Bergström, A., Coupland, P., Warringer, J., Lagomarsino, M. C., Fischer, G., Durbin, R., & Liti, G. (2017). Contrasting evolutionary genome dynamics between domesticated and wild yeasts. *Nature Genetics*, 49(6), 913–924. <https://doi.org/10.1038/ng.3847>
- Yue, J. X., & Liti, G. (2018). Long-read sequencing data analysis for yeasts. *Nature Protocols*, 13(6), 1213–1231. <https://doi.org/10.1038/nprot.2018.025>
- Zetka, M. C., & Müller, F. (1996). Telomeres in nematodes. *Seminars in Cell and Developmental Biology*, 7(1), 59–64. <https://doi.org/10.1006/scdb.1996.0009>

Conclusions and perspectives

In this PhD work that lasted over four years, I used the budding yeast *Saccharomyces cerevisiae* as model organism to investigate important aspects of genome evolution: the origin of interspecies introgressions and telomere evolution.

Introgressions and hybrid sterility

Introgressions are thought to initiate from ancient hybridization events followed by repeated backcrossing to one parental population/species. They are characterized by allelic replacement, meaning that the genetic fragment which is transferred from one species/population to another will not insert in a random part of the recipient genome, but it will replace its homologous sequence. This implies that recombination, and, therefore, sex, must be involved in the process. The presence of Neanderthal and Denisovan introgressions into the modern human genome has been successfully explained by such hybridization-backcross model, given the low sequence divergence between the two populations. We often find gene introgressions also between species whose DNA are so different that sex fails to produce offspring. The *Saccharomyces sensu stricto* complex constitutes one of such examples, as sequence divergence among its species can be pretty high and hybrids between them usually fail to generate viable offspring. How could these yeast DNA introgressions occur when the hybrid is sterile? Our living ancestor described in chapter 4 offered us the answer: the occurrence of scattered blocks of sequence identity in a genome that otherwise is completely heterozygous creates preferential sites for recombination, overcoming in this way the major reproductive barrier and giving origin to the current population that now inhabits the wastewaters of olive oil production (Alpechin).

The *S. paradoxus* introgressions present in the *S. cerevisiae* Alpechin strains are enriched in genes coding for cell membrane and cell wall proteins, suggesting a possible role in the adaptation of yeasts to live in the harsh environmental conditions of olive oil wastewater (OWW). OWW is in fact characterized by an acidic pH, a high salt concentration causing osmotic stress, and by the presence of phenolic compounds and a complex microbial community. Alpechin strains have better growth performance than other strains in this environment, but the genetic bases underlying this adaptation are not yet completely understood. Future directions of this project will involve the characterization of the contribution of introgressed genes in the adaptation of Alpechin yeasts to this environment. I found that three genes have been independently introgressed and retained in four *S. cerevisiae* lineages: Alpechin, Brazilian bioethanol, French Guyana and Mexican agave. Among these, the gene *IRC7* constitutes an interesting candidate as it has been shown to be introgressed not

only in these 4 lineages, but also in some clinical *S. cerevisiae* isolates, while the majority of Wine/European strains carry a deleted copy.

Another interesting aspect of this work is that we characterize, for the first time, the presence of an ancient hybrid and its derivative strains in the same ecological niche. The coinhabitation of these two populations might reflect an adaptation to utilize different resources in the same ecological niche. In our phenotyping of asexual fitness in over 80 environments, we evidence the presence of environments in which either the Alpechin strains perform better than the living ancestor or viceversa. It would be interesting to better characterize the fitness of the Alpechins and the living ancestor in OWW, discover if and which resources are differentially utilized by the two, and unravel the genes underlying this behaviour.

Telomere evolution

Telomeres are variable among taxa. Species typically have very different average telomere lengths, and even among the same species, telomere length can vary among individuals and populations. Telomeres are not only different in terms of length, but also in terms of sequence, and a great diversity of telomeric DNA motifs was reported among organisms. Despite we have a wide knowledge about telomere diversity, we still lack a mechanistic understanding of the evolutionary processes that led to the variety of telomeric lengths and sequences in eukaryotes, and how such diversity affects fitness.

I approached this scientific problem by two different angles: in one project, I characterized the diversity of telomere lengths in the widest collection of *S. cerevisiae* isolates currently available, and correlated these data with phenotypes already available for this collection. Despite the presence of numerous wild strains, the majority of the collection was nevertheless constituted by domesticated-derived isolates. I believe this study constitutes a first step towards the understanding of yeast telomere biology in the wild, but more sampling in wild environments is needed in order to decipher the biological significance of telomere length in different ecological settings.

The analysis of such dataset answered many important questions in yeast telomere biology. We now know that aneuploidy and ploidy do not affect telomere length, along with a long list of other phenotypes. Interestingly, the comparison of telomere length between domesticated and wild isolates revealed an unexpected difference in telomere length, with wild isolates carrying shorter telomeres than domesticated ones. We still lack a complete understanding of the environmental pressures at the origin of this difference, but an intriguing explanation might rely on the impact of telomere length on the transcriptomic profile of strains. Shorter telomeres might decrease the transcriptional repression present at subtelomeric sites and allow stress response genes to be

expressed. Moreover, a shorter telomeric tract needs less telomere-binding proteins, which are in turn free to bind other genetic targets in the genome and modulate their expression. The lack of transcriptomic and proteomic data for the 1002 yeast collection unfortunately prevented us from testing these hypotheses. Hopefully, the publication of such data in the future will allow us to answer these important scientific questions.

The need to measure telomere length in such a wide collection (over 900 isolates) constituted a huge technical obstacle for me, due to the great laboriousness and time-consuming nature of Teloblots. This was at the same time a source of great frustration and a scientific challenge, as it motivated and challenged me to develop an alternative technique to measure telomere length in a quicker and less expensive way, that resulted in an additional achievement of my PhD work: a bioinformatic pipeline to measure telomere length and ITS content in yeast. The pipeline will be made freely available upon publication of chapter 5 and will constitute a bioinformatic resource for other researchers in the field. I believe the huge amount of data (telomere length, ITS content and Y' copy number) that I generated in this project will also constitute a resource for other colleagues in the telomere research field. ITS content is especially relevant in this sense, because it is still unknown whether ITS have other functional roles in addition to constituting an alternative mechanism of telomere maintenance.

In my second project, I shifted my attention to telomere sequence diversity and investigated the molecular mechanisms at the basis of telomere sequence evolution. To do this, I chose a model organism called “telomere-humanized yeast”, which is essentially a *S. cerevisiae* yeast carrying human telomeric repeats. The choice of this experimental model constitutes a sort of reverse evolution, as the telomeric repeats present in humans are thought to be the ancestral ones from which all the other variants evolved. I submitted my humanized yeasts to a MAL protocol, in order to characterize the long term effect of telomere humanization, and then submitted MAL to adaptive evolution, in order to test if and how humanized yeasts could recover a normal fitness. All the data that I recovered from these experiments pointed into one direction: the inactivation of the DNA-damage sensing pathway is a necessary step in order to recover fitness and being able to survive with a non-native telomeric sequence. Despite its beneficial role in telomere evolution, the inactivation of the DDR pathway can have very detrimental effects. Cells experience thousands of DNA lesions every day, as a consequence of chemical and environmental agents, and the DDR constitutes the first sensor of DNA damage. If the pathway is inactivated, these lesions will not be detected and repaired, leading to the onset of chromosomal aberrations. Mutations in key components of the DDR pathway often constitute driver mutations in the development of cancer. Our humanized line 8, carrying a putative loss-of-function mutation in the gene *TEL1*, constitutes a

perfect example of this tradeoff. On one hand, the occurrence of the mutation is followed by a complete recovery of growth performance. On the other hand, a careful look at the genome revealed that this line underwent multiple chromosomal rearrangements, extensive subtelomeric amplifications and even a whole-genome duplication. The additional evolution of this line through a ST protocol resulted in the accumulation of even more chromosomal aberrations, mainly constituted by segmental duplications of various chromosomes (data not shown).

Telomere dysfunction and cancer are intimately connected and often observed in concurrence. Their connection passes through the occurrence of bridge-fusion-breakage cycles upon telomere dysfunction, which result in chromosomal rearrangements that can trigger the onset of cancer. Our experiments revealed that the connection between telomeres and cancer might also pass through an increase in point mutation rate upon telomere dysfunction. If a loss-of-function mutation occurs in a key oncosuppressor gene and is positively selected because it confers some growth advantage, the cell line carrying this mutation will experience genomic instability similar to what is observed in cancer cells.

On a more evolutionary scale, the onset of such mutations and their consequent genomic instability might also favor an early process of speciation, thanks to the reproductive isolation conferred by different genomic structures.

However, given the important and detrimental consequences of DDR inactivation on the cell, it is unlikely that such inactivation is maintained on a permanent basis. A more likely scenario of telomere sequence evolution might involve the transient inactivation of the DDR by a similar type of mutation, followed by the occurrence of a compensatory mutation somewhere else, able to restore a complete DNA damage sensing activity.

The availability of high-throughput genetic engineering techniques will allow this and other questions to be answered in the near future. As an example, we currently do not know anything about the evolutionary consequences that other telomeric repeats might have in yeast: an interesting research avenue could be to construct not only telomere-“humanized” yeasts, but also telomere-“insectized”, “vegetized” and other yeasts carrying the spectrum of telomeric repeats observed in nature, to characterize the phenotypic consequences of such telomeric repeats and how yeasts adapt to that challenge.

Annex: other projects

Aborting meiosis overcomes hybrid sterility

Simone Mozzachiodi, Lorenzo Tattini, Agnes Llored, Agurtzane Irizar, Neža Škofljanc, Melania D'Angiolo, Matteo De Chiara, Benjamin Barré, Jia-Xing Yue, Angela Lutazi, Sophie Loeillet, Raphaëlle Laureau, Souhir Marsit, Simon Stenberg, Benoit Albaud, Karl Persson, Jean-Luc Legras, Sylvie Dequin, Jonas Warringer, Alain Nicolas, Gianni Liti. Submitted.

Authors contributions

A.N. and G.L. conceived the project; S.Mo., L.T., J.W., A.N., G.L., designed the experiments; S.Mo., L.T., A.L., A.I., B.B., N.Š., M.D.C., J-X. Y., M.D., A.L., S.L., S.S., R.L., S.Ma., S.S., K.P., performed and analysed the experiments; B.A., performed the sequencing, J.L.L., S.D., J.W., A.N., G.L., contributed with resources and reagents; S.Mo., J.W., A.N., G.L. supervised the project; G.L. coordinated the project; S.Mo. and G.L. wrote the paper with input from L.T., A.N., J.W.

Abstract

Hybrids between species or diverged lineages contain fundamentally novel genetic combinations, but an impaired meiosis often make them evolutionary dead ends. Here, we explored recombination induced by an aborted meiosis, namely return-to-growth (RTG), across a panel of 20 yeast diploid backgrounds with different genome structures and levels of sterility. Genome analyses of 284 RTG-evolved and control clones revealed extensive regions of loss-of-heterozygosity in sterile hybrids with either a defective meiosis or a heavily rearranged karyotype, whereas RTG recombination was reduced by high levels of sequence divergence between parental subgenomes. The RTG recombination preferentially occurs in regions with local sequence homology and in meiotic recombination hotspots. The loss-of-heterozygosity had a profound impact on both sexual and asexual fitness and enabled genetic mapping of phenotypic differences in sterile lineages where linkage or association analyses fail. We propose that RTG gives sterile hybrids access to a natural route for genome recombination and adaptation.

Accurate Tracking of the Mutational Landscape of Diploid Hybrid Genomes

Lorenzo Tattini, Nicolò Tellini, Simone Mozzachiodi, [Melania D'Angiolo](#), Sophie Loeillet, Alain Nicolas, Gianni Liti (2019). *Molecular Biology and Evolution*.

Author Contributions

L.T. designed and implemented computational methods, performed simulations, analyzed data, and wrote the manuscript. N.T. analyzed data and performed simulations. S.M. tested computational methods and performed experimental validations. [M.D.](#) performed mutation accumulation experiments. S.L. conducted experiments. A.N. revised the manuscript. G.L. coordinated and designed the study and wrote the manuscript.

Abstract

Mutations, recombinations, and genome duplications may promote genetic diversity and trigger evolutionary processes. However, quantifying these events in diploid hybrid genomes is challenging. Here, we present an integrated experimental and computational workflow to accurately track the mutational landscape of yeast diploid hybrids (MuLoYDH) in terms of single-nucleotide variants, small insertions/deletions, copy-number variants, aneuploidies, and loss-of-heterozygosity. Pairs of haploid *Saccharomyces* parents were combined to generate ancestor hybrids with phased genomes and varying levels of heterozygosity. These diploids were evolved under different laboratory protocols, in particular mutation accumulation experiments. Variant simulations enabled the efficient integration of competitive and standard mapping of short reads, depending on local levels of heterozygosity. Experimental validations proved the high accuracy and resolution of our computational approach. Finally, applying MuLoYDH to four different diploids revealed striking genetic background effects. Homozygous *Saccharomyces cerevisiae* showed a ~4-fold higher mutation rate compared with its closely related species *S. paradoxus*. Intraspecies hybrids unveiled that a substantial fraction of the genome (~250 bp per generation) was shaped by loss-of-heterozygosity, a process strongly inhibited in interspecies hybrids by high levels of sequence divergence between homologous chromosomes. In contrast, interspecies hybrids exhibited higher single-nucleotide mutation rates compared with intraspecies hybrids. MuLoYDH provided an unprecedented quantitative insight into the evolutionary processes that mold diploid yeast genomes and can be generalized to other genetic systems.

Simultaneous occurrence of covert infections with small RNA viruses in the lepidopteran *Spodoptera exigua*

Agata K. Jakubowska, Melania D'Angiolo, Rosa M. González-Martínez, Anabel Millàn Leiva, Arkaitz Carballo, Rosa Murillo, Primitivo Caballero, Salvador Herrero (2014). Journal of invertebrate pathology.

Author contributions

I sequenced and analysed the genome of three novel RNA viruses infecting *Spodoptera exigua* and realized the graphical abstract and figure 1 of the paper.

Abstract

Viral covert infections in invertebrates have been traditionally attributed to sublethal infections that were not able to establish an acute infection. Recent studies are revealing that, although true for some viruses, other viruses may follow the strategy of establishing covert or persistent infections without producing the death of the host. Recently, and due to the revolution in the sequencing technologies, a large number of viruses causing covert infections in all type of hosts have been identified.

The beet armyworm, *Spodoptera exigua* (Lepidoptera: Noctuidae) is a worldwide pest that causes significant losses to agricultural and ornamental plant industries. In a previous project we used NGS to obtain a comprehensive transcriptome of the larval stage, revealing the presence of an important number of unigenes belonging to novel RNA viruses, most of them from the order *Picornavirales*. In order to characterize *S. exigua* viral complex, in this work we have completed the genomic sequences of two picorna-like viruses, and compared them to a SeIV1, a member of *Iflaviridae* previously described by our group. We performed additional studies to determine virus morphology, horizontal transmission, tissue and life stage distribution and abundance in the hosts. We discuss the role of virus persistent infections on insect populations.

Curriculum Vitae

D'Angiolo Melania Jennifer

email: melania.dangiolo@unice.fr

WORK EXPERIENCE

- September 2016 to date **IRCAN: Institute for Research on Cancer and Aging Nice, France:** PhD fellow via Labex Signallife International PhD programme. Teams: Population genomics and complex traits (CNRS UMR 7284 - INSERM U1081), Telomeres, senescence and cancer (IRCAN). Thesis directors: Dr Gianni Liti, Pr Eric Gilson.
- September 2015 - June 2016 **IRCAN: Institute for Research on Cancer and Aging Nice, France:** Internship for Master's Thesis via Erasmus Mobility Placement Programme. Team: Population genomics and complex traits (CNRS UMR 7284 - INSERM U1081). Thesis director: Dr Gianni Liti.
- January - September 2012 **Department of Genetics at the University of Valencia, Spain:** Internship for Bachelor's Thesis via Erasmus Mobility Placement Programme. Team: Biological control of pests (CBP). Thesis director: Dr Salvador Herrero.

ACADEMIC EDUCATION

- 2016-2021 (as planned) **Doctor of Philosophy in Life Sciences and Health.** Université Cote d'Azur, France.
- 2013-2016 **Master of Science in Molecular and Cellular Biology.** University of Pisa, Italy.
- 2008-2013 **Bachelor of Science in Molecular Biology.** University of Pisa, Italy.

AWARDS

- May 2019 Awarded a fellowship for a 4th PhD year by Fondation pour la Recherche Médicale (FRM).
- July 2017 Awarded a travel grant to attend the EMBL course "Mass spectrometry and RNAseq".
- June 2016 Selected for a PhD fellowship in the Labex Signallife International PhD programme.
- May 2016 Awarded a travel grant to attend the congress "BITS2016".
- April 2015 Awarded an Erasmus Placement Scholarship granted by the European Commission.
- November 2011 Awarded an Erasmus Placement Scholarship granted by the European Commission.

LANGUAGE SKILLS

Italian (First Language), English (C1), French (C1), Spanish (C1), German (A1), Korean (notions).

JOB-RELATED SKILLS

- Molecular biology techniques Electrophoresis, PCR, cloning, DNA/RNA extraction and purification, sequencing library preparation, q-PCR, phenotyping, flow cytometry, manipulation of radioactive probes, southern blot, Oxford Nanopore sequencing.
- Whole genome analysis in next generation sequencing datasets Short and long reads mapping and assembly, coverage and CNV analysis, SNP detection, genome annotation, linkage analysis and genome-wide association studies, algorithmics and programming, phylogenetics.

Operating systems Windows, Linux/Unix.
Programming languages Perl (proficient), R (intermediate), Bash (basic).

EDITORIAL WORK

Reviewer for the peer-reviewed journal "Genome Biology and Evolution".

PUBLICATIONS

Publications in peer-reviewed journals

D'Angiolo M, De Chiara M, Yue JX, Irizar A, Stenberg S, Persson K, Llored A, Barré B, Schacherer J, Marangoni R, Gilson E, Warringer J, Liti G. A yeast living ancestor reveals the origin of genomic introgressions. *Nature* (2020).

D'Angiolo M. et al. Telomeres are shorter in wild *S. cerevisiae* populations than in domesticated ones. (In preparation).

D'Angiolo M. et al. Inactivation of the DNA damage response rescues organismal fitness in telomere-humanized yeast. (In preparation).

Mozzachiodi S, Tattini L, Llored A, Irizar A, Skofljanc N, **D'Angiolo M**, De Chiara M, Barré B, Yue JX, Lutazi A, Loeillet S, Laureau R, Marsit S, Stenberg S, Baulande S, Persson K, Legras JL, Dequin S, Warringer J, Nicolas A, Liti G. Aborting meiosis overcomes hybrid sterility. (Submitted).

Tattini L, Tellini N, Mozzachiodi S, **D'Angiolo M**, Loeillet S, Nicolas A, Liti G. Accurate tracking of the mutational landscape of diploid hybrid genomes reveals genetic background effects. *Molecular Biology and Evolution* (2019).

Jakubowska AK, **D'Angiolo M**, González-Martínez RM, Millán-Leiva A, Carballo A, Murillo R, Caballero P, Herrero S. Simultaneous occurrence of covert infections with small RNA viruses in the lepidopteran *Spodoptera exigua*. *Journal of Invertebrate Pathology* (2014).

Conference presentations

D'Angiolo M, De Chiara M, Yue JX, Irizar A, Stenberg S, Persson K, Llored A, Barré B, Schacherer J, Marangoni R, Gilson E, Warringer J, Liti G. A yeast living ancestor reveals the origin of genomic introgressions. EMBL Conference: Molecular mechanisms in Evolution and Ecology. Heidelberg (2020). Selected talk.

D'Angiolo M, Barré B, Gilson E, Liti G. Inactivation of the DNA damage response rescues fitness in telomere-humanized yeast. 4th Labex Signalife meeting. Nice (2020). Selected talk.

D'Angiolo M, De Chiara M, Yue JX, Irizar A, Stenberg S, Persson K, Llored A, Barré B, Schacherer J, Marangoni R, Gilson E, Warringer J, Liti G. A yeast living ancestor reveals the origin of genomic introgressions. Nice Genomics Winter School. Nice (2019). Invited talk.

D'Angiolo M, Barré B, Gilson E, Liti G. Inactivation of the DNA damage response rescues fitness in telomere-humanized yeast. EMBO workshop: Comparative genomics of eukaryotic microbes – Genomes in flux and flux between genomes. Sant Feliu de Guixols (2019). Selected talk.

D'Angiolo M, De Chiara M, Yue JX, Irizar A, Stenberg S, Persson K, Llored A, Barré B, Schacherer J, Marangoni R, Gilson E, Warringer J, Liti G. A yeast living ancestor reveals the origin of genomic introgressions. Journées de l'Ecole Doctorale de Nice. Nice (2019).

Jakubowska AK, **D'Angiolo M**, Millán-Leiva A, Herrero S. Simultaneous and persistent infections of picornaviruses in the lepidopteran *Spodoptera exigua*. Proceedings of the XII National Congress of Virology. Burgos (2013).

Posters

D'Angiolo M, Barré B, Gilson E, Liti G. Inactivation of the DNA damage response rescues fitness in telomere-humanized yeast. Aquatic research models to study regeneration and aging workshop. Nice (2019).

D'Angiolo M, De Chiara M, Yue JX, Irizar A, Llored A, Stenberg S, Barré B, Loeillet S, Marangoni R, Gilson E, Nicolas A, Warringer J, Liti G. Genome instability initiates introgressions in budding yeast. Journées de l'Ecole Doctorale de Nice. Nice (2018).

Nicolas A, Loeillet S, Llored A, Marsit S, Laureau R, Tattini L, **D'Angiolo M**, Bedrat A, Serero A, Legras JL, Mozzachiodi S, Baulande S, Dequin S, Liti G. Recombination of yeast hybrids through meiotic reversion. EMBO Conference on Meiosis. Hvar (2017).

Laureau R, Loeillet S, **D'Angiolo M**, Llored A, De Chiara M, Yue JX, Barré B, Legoix-Né P, Liti G, Nicolas A. Recombination of hybrid yeast strains upon reversion of meiotic progression. EMBO Conference: Experimental Approaches to Evolution and Ecology using Yeast and Other Model Systems. Heidelberg (2016).

D'Angiolo M, De Chiara M, Yue JX, Llored A, Irizar A, Lutazi A, Barré B, Marangoni R, Nicolas A, Liti G. A dominant role of mitotic recombination in shaping interspecies genome introgression. 13th Annual Meeting of the Bioinformatics Italian Society. Salerno (2016).

Jakubowska AK, **D'Angiolo M**, Millan-Leiva A, Carballo A, Murillo R, Caballero P, Herrero S. Simultaneous covert infections with three different RNA viruses in the lepidopteran *Spodoptera exigua*. 47th International congress on invertebrate pathology and microbial control. Mainz (2014).

Jakubowska AK, **D'Angiolo M**, Millan-Leiva A, Murillo R, Caballero P, Herrero S. Infecciones persistentes y simultáneas con 3 picomavirus en poblaciones de *Spodoptera exigua*. 8th National Congress of Applied Entomology. Matarò (2013).

Nice, 30.11.2020