Genetic studies of mitochondrial DNA in *Saccharomyces cerevisiae*

Soraia Joana Almeida Aires Esteves
Mestrado em Genética Forense
Departamento de Biologia
2018

Orientador
Pedro Soares, Assistant Professor/Researcher, Universidade do Minho

Coorientador
Ricardo Duarte, Post-Doc Researcher, Universidade do Minho

U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, _____/_____/_____

Agradecimentos

Em primeiro lugar, gostaria de expressar a minha sincera gratidão aos meus supervisores, Pedro Soares e Ricardo Duarte, por todo o apoio que permitiu que esta pesquisa fosse realizada. Ao Prof. Pedro Soares particularmente pela sua orientação, apoio e imenso conhecimento, por me orientar e por me ajudar a escrever esta tese.

Também gostaria de agradecer a todas as pessoas envolvidas, como Daniel Vieira, pela sua experiência em informática na simplificação de tarefas longas e demoradas, e Eduardo Conde-Sousa por sua ajuda na obtenção e conversão das sequências.

A todos os meus amigos, por estarem sempre lá para me animar, e por todos os momentos inesquecíveis e maravilhosos.

Finalmente, mas não menos importante, gostaria de agradecer à minha família por me encorajar e apoiar durante toda a minha vida.

Summary

Saccharomyces cerevisiae is a well-established model for genetic studies, more specifically mitochondria genetics, as a result of its functional similarity to the human DNA and the way it facilitates direct manipulation of the mitochondrial genome. Coincidentally, the first mitochondrial gene sequenced was from this species.

As a facultative anaerobe with a short generation time, it belongs in the *Saccharomyces* genus, along with other budding yeasts that are wildly used in industry and other human services. *S. cerevisiae*'s cell cycle alternates between sexual and asexual reproduction. During sexual reproduction, mitochondrial DNA is inherited from both parents with the occurrence of recombination.

The mitochondrial DNA of *S. cerevisiae* consists of long, linear molecules of heterogeneous size, about 85 kb, with around 35 genes and a low gene density, and extensive intergenic regions and a high A+T content. It is compressed into nucleoprotein structures called mitochondrial nucleoids (mt-nucleoids), which is a feature that allows for a better accommodation in the cell, as well as spatial regulation of gene expression and protection from DNA damage. As such, these characteristics of the genome make it more difficult to prepare, sequence and align the genetic material.

We collected sequences and their respective data from NCBI, submitting them to several tests of phylogeny, gene content, recombination, linkage, and genetic structure. Working with only the coding sections of the mitochondrial genome, we concluded that even under these conditions, there was still a prevalence of AT base pairs and that, even though the sequences were so small, there was a high level of diversity present in the form of SNP, with very high haplotypic diversity and a good discrimination between sequences. Recombination clearly plays an active and important role in the evolution and diversity of the species, providing high haplotypic diversity through new blends of existing variants. On the down side, recombination may be causing newly formed haplotypes to disappear, causing a lack of general patterns when studying the relationships between strains collected from different isolations sources and locations. It may also be responsible for large deletions and gene reversions that end up reflecting in the instability of the gene content.

Resumo

Saccharomyces cerevisiae é um modelo bem estabelecido de estudos genéticos, devido à sua similaridade funcional com o DNA humano e pela facilidade de manipulação direta do genoma mitocondrial. Coincidentemente, o primeiro gene mitocondrial sequenciado foi desta espécie.

Um anaeróbio facultativo com um tempo de geração curto, esta espécie pertence ao género *Saccharomyces*, juntamente com outras leveduras amplamente usadas na indústria e em outros serviços humanos. O ciclo celular de *S. cerevisiae* alterna entre reprodução sexual e reprodução assexual. Durante a reprodução sexual, o DNA mitocondrial é herdado de ambos os pais com a ocorrência de recombinação.

O mtDNA de *S. cerevisiae* consiste em moléculas lineares longas de tamanho heterogéneo, por volta de 85 kb, com cerca de 35 genes, baixa densidade genética, extensas regiões intergénicas e alto conteúdo de A + T. Encontra-se comprimido em estruturas nucleoproteicas chamadas nucleoides mitocondriais (mt-nucleoids), característica que permite uma melhor acomodação na célula, bem como a regulação espacial da expressão génica e proteção contra danos no DNA. Como tal, estas características do genoma tornam difícil a preparação, sequenciamento e alinhamento o material genético.

Da base de dados NCBI, foram recolhidas sequências, juntamente com os respetivos dados, submetendo-as a vários testes de filogenia, conteúdo genético, recombinação, associação entre genes e estrutura genética. Trabalhando apenas com as seções codificantes do genoma mitocondrial, concluímos que, mesmo nessas condições, ainda havia uma prevalência de pares de bases A e T e, embora as sequências fossem de reduzido tamanho, houve um alto nível de diversidade presente na forma de SNP, com diversidade haplotípica muito alta e uma boa discriminação entre sequências. A recombinação desempenha claramente um papel ativo e importante na evolução e diversidade da espécie, proporcionando alta variedade haplotípica através de novas combinações de variantes existentes. Pela negativa, a recombinação pode estar a fazer com que os haplótipos recém-formados desapareçam, causando uma falta de padrões gerais ao estudar as relações entre as linhagens provenientes de diferentes fontes e locais de isolamento. Pode também ser responsável por grandes deleções e reversões génicas que acabam por se refletir na instabilidade do conteúdo génico.

Key words

mtDNA, Saccharomyces, budding yeast, clustering, phylogeny, recombination

Abbreviations

NCBI - National Center for Biotechnology Information; SRA – Sequence read archive; LD – linkage disequilibrium; PCA – principal component analysis;

Index

1. Introduction

1.1. Saccharomyces cerevisiae

Saccharomyces cerevisiae - previously known by more than 80 synonyms, which include *Saccharomyces boulardii* and *Saccharomyces diastaticus* (Vaughan-Martini & Martini, 2011), - is a microbial agent important in human activities (Schacherer, Shapiro, Ruderfer, & Kruglyak, 2009). From dietary supplements to the brewing industries (Gianni Liti et al., 2009), such as baking and fermentation of wine, beer, and other alcoholic beverages (Gianni Liti et al., 2009; Schacherer et al., 2009; Wang, Liu, Liti, Wang, & Bai, 2012), *S. cerevisiae* has also established itself as one of the most important model organisms in genetics, genomics, and molecular biology. It is being sought out by several disciplines (Hittinger, 2013; Wolters, Chiu, & Fiumera, 2015), such as ecology, and population and evolutionary genetics (Wang et al., 2012).

Being a facultative anaerobe, with the capacity to obtain energy generated solely by fermentation, this species is also a significant model in the study of cellular and biochemical pathways responsible for the maintenance of respiratory activity (De Deken, 1966; Gancedo, 1998; J Piskur, Rozpedowska, Polakova, Merico, & Compagno, 2006).

S. cerevisiae was the first eukaryote to have its nuclear genome sequenced (Goffeau et al., 1996). The small 12 Mb genome is better annotated than that of any other eukaryote (Cherry et al., 2012), with two-thirds of the ~6000 identified genes of the species being characterized (Lipinski, Kaniak-golik, & Golik, 2010; Wang et al., 2012).

This species has a large genetic diversity and highly structured populations among wild isolates (Wolters et al., 2015). This yeast is one of the most thoroughly studied eukaryotic species, mainly due to the existence of a short generation time, the possibility to control the sexual cycle, and availability of a genome that is easy to manipulate (Johnston, 2000; Wang et al., 2012). Genetic analyses and phenotyping are fairly simple to obtain as well, and the engineering and testing of the effects of individual polymorphisms and their combinations on different genetic backgrounds (Schacherer et al., 2009), also the fact that researchers can exercise complete control over its genetics and environment (Ehrenreich et al., 2010; Storici, Durham, Gordenin, & Resnick, 2003) are all makers for a good model for complex analysis.

1.2. Budding yeast – the genus *Saccharomyces*

The term budding yeast is applied to organisms which, during mitotic growth, multiply by asymmetric division in a process named budding. At the starting point of each cell cycle, cells develop polarization and select a site for the bud to develop. That, in turn, grows until it reaches the size of the mother cell, at which point a septum is formed separating daughter cell from the mother cell. The process is followed by organelle partitioning by both active and directed transport of organelles between the two cells (Pruyne, Legesse-Miller, Gao, Dong, & Bretscher, 2004; Westermann, 2014).

The monophyletic fungi genus *Saccharomyces* originated approximately 10-20 million years ago (Kellis, Patterson, Endrizzi, Birren, & Lander, 2003; G. Liti, Barton, & Louis, 2006; Taylor & Berbee, 2006). It has seven species which have differences in gene content and architecture and protein sequences (Hittinger, 2013). Among those there is *S. paradoxus*, phylogenetically closer to *S. cerevisiae*, and *S. bayanus*, used to brew wine. It was in 1883 that Emil C. Hansen discovered the importance of yeast in brewing after having isolated and propagated a culture of the yeast (Boulton & Quain, 2001).

Yeasts are extensively utilized for industrial purposes, such as biomedical and bioenergy research (Hittinger, 2013), being fermentation a mean of developing biofuel technologies that produce ethanol as a renewable source of energy (Zhu, Sherlock, & Petrov, 2016). By studying these microorganisms in different ecosystems, their diversity is better explored, and new biotechnological properties can be found (Alvarenga, Carrara, Silva, & Oliveira, 2011; Amorim, Lopes, de Castro Oliveira, Buckeridge, & Goldman, 2011; Bravim, Palhano, Fernandes, & Fernandes, 2010; Silva, Batistote, & Cereda, 2013; Úbeda et al., 2016; Watanabe et al., 2010). As a result, the study of *Saccharomyces* ecology and diversity concentrates on natural and industrial fermentations and the habitats in which they occur, with emphasis on the most wanted features (Vaughan-Martini & Martini, 2011).

To obtain better suited yeasts for certain products, double and triple hybrid strains can be created, with various contributions from each of the species, depending on the needs of the project. One of the most famous industrial hybrids is the *Saccharomyces pastorianos*, which is adapted to the cold, and an alloploid of *S. cevisiae* and *S. eubayanus*. It is a very valuable asset to global brewing industry (Dunn & Sherlock, 2008; Libkind et al., 2011).

Saccharomyces life cycle is characterized by rare, but crucial, occurrences of sexual reproduction and outcrossing, and it's known that recombination of the nuclear genome occurs in nature at a high rate (Hittinger, 2013). The cells can have a haploid and a diploid phase, depending on the conditions that surround them. Under stress, diploid cells undertake sporulation followed by meiosis that produces four haploid spores (Herskowitz, 1988). Those

mother-daughter cells can then mate (haploselfing). As a consequence, sexual reproduction, outcrossing, and recombination are crucial to the survival of the species and genome renewal (Hittinger, 2013; Tsai, Bensasson, Burt, & Koufopanou, 2008), since recessive deleterious mutations have a chance at accumulating in populations during times of asexual reproduction (Masel & Lyttle, 2011; Mortimer, Romano, Suzzi, & Polsinelli, 1994).

In order to discover the evolutionary history of a species or genus, phenotypic differences must be understood on a genetic basis, by identifying polymorphisms among individuals of said group (Schacherer et al., 2009). It is possible nowadays to evaluate DNA sequence variation and genome evolution by the availability of sequence data (Dujon, 2006).

There have been studies of comparative mitochondrial genomic analysis between yeast species which have been used to reconstruct the evolutionary changes in genome organization and architecture (Wolters et al., 2015). This has the potential to illuminate the genetic and molecular mechanisms of evolution (Hittinger, 2013).

1.3. *Saccharomyces* mitochondrial genomes

Saccharomyces' mitochondrial genomes are highly diverse in structure and organization as a result of large rearrangements, accumulation of intergenic sequences, and of point mutations. On the course of their evolution, *Saccharomyces'* dynamic genome have suffered plentiful fluctuations. Therefore, among closely related yeasts, it's possible to deduce molecular events of evolution (Groth, Petersen, Piškur, & Pis, 2000).

Mitochondria move among cytoskeletal paths and fuse and divide often (Jakobs et al., 2003; Merz, Hammermeister, Altmann, Dürr, & Westermann, 2007; Westermann, 2010). The division of mitochondria between mother cell and bud involves a complex network of actions, with preservation zones ensuring the equal distribution of the organelles amid the cells (Fehrenbacher, Yang, Gay, Huckaba, & Pon, 2004; Simon, Karmon, & Pon, 1997; Yang, Palazzo, Swayne, & Pon, 1999). Some studies have revealed that when isolated from one another, yeasts accumulate distinct mutations (Groth et al., 2000).

Yeast species have very different genome sizes and gene order differs enough to be an indicator of a big sum of rearrangements through time (Jung, Friedrich, Reisser, Hou, & Schacherer, 2012), together with the existence of intergenic sequences, and polymorphism inside the coding regions points at the genome having suffered big modifications throughout its evolution (Groth et al., 2000). One of the reasons for those polymorphisms is the size and tenor of GC clusters that are significantly different within *Saccharomyces* yeasts (Groth et al., 2000).

The group of species of the *Saccharomyces* genus are divided in petite-positive and petite-negative yeasts (Barnett, 1992). Additionally, petite-positive yeasts are classified into sensu stricto (which includes *Saccharomyces cerevisie*) and sensu lato yeasts, which have mitochondrial genome of reduced size and significative differences in phylogeny and gene order. Sensu stricto yeasts also have very extensive intergenic sequences and are closely related (with similar genome size and gene order). Modern petite-positive *Saccharomyces* all share fermentation and respiration procedures, which brings the notion that all petite-positive yeasts must have suffered the same evolutionary pressures on the mitochondrial genes responsible for respiratory functions (Groth et al., 2000).

1.4. Geographical distribution and domestication – interaction with the human species

S. cerevisiae is a species that exists in phylogenetically distinct populations, both wild and domesticated, being oak trees (*Quercus*) an important habitat to *Saccharomyces*. This species is cosmopolitan (G. I. Naumov, Naumova, & Sniegowski, 1998; G. Naumov & Naumov, 1987), and present in environments that are not - or are only marginally - affected by human activity, like remote forest environments, throughout Russia, America, Asia, and all over Europe (Hittinger, 2013). It is usually isolated from substrates likely to offer suitable nutrients for the growth of this eukaryotic microbe (Hittinger, 2013; Wang et al., 2012), - like oak bark, exudates (sap), acorns, leaves, or nearby soil (Hittinger, 2013). Cross-breeding and the assembly of new combinations of pre-existing variations are thought to be made possible by human influence. That is supported by the population structure of the species which consists of a few well-defined, geographically isolated lineages and many different mosaics of these lineages (Gianni Liti et al., 2009).

Geography, environmental niches and the degree of human association are related to the complex pattern of differentiation that contains distinct identified lineages (Peter & Schacherer, 2015). As far as it is known, genetic differentiation indicates that population structure reflects ecological niches based on the source from which strains are isolated (Schacherer et al., 2009), rather than their geographical origins. Many closely related strains are found in widely separated locations (Gianni Liti et al., 2009). There have been studies that identified five well-defined, geographically isolated lineages Malaysian, West African, Sake, North American and ‘Wine/European’, as well as recombinant strains of these lineages (Wang et al., 2012).

Several studies have suggested separate events of domestication/selection for *S. cerevisiae* (Gianni Liti et al., 2009; Wang et al., 2012). For example, the baker’s yeast has

been associated with human activity (Pretorius, 2000), leading to the idea that its use of fermentation led to its domestication (Gianni Liti et al., 2009). But this long association with humans is thought to have resulted in limited population differentiation of the species (Wang et al., 2012). Some say the population structure of *S. cerevisiae* provides support not only for the multiple domestication events theory, but also provides some insight into the origins of pathogenic strains (Schacherer et al., 2009). Since as a whole, the species is not domesticated, it is important to have studies that focuses on wild isolates, in order to clarify the population structure of the species (Wang et al., 2012). S. paradoxus, for example, is often found in cohabiting with *S. cerevisiae* though it seldom associates with human activity (Wang et al., 2012).

This knowledge is very important since human-associated strains gave rise to taxonomic controversies (Hittinger, 2013). Human activity gave a chance for population mingling and recombination (Gianni Liti et al., 2009; Magwene et al., 2011). This reasoning is why population genetics analysis provides important new understandings into ecological distribution, population structure and biogeography of the species. (Wang et al., 2012).

1.5. Clinical strains

S. cerevisiae strains from a large range of environments and even with human applications are capable of opportunistic colonization of human tissues (Schacherer et al., 2009), often found as an asymptomatic human gut commensal. But not always. It's also known to have low opportunistic virulence, having the worst cases happened with involvement of immunocompromised individuals, or with underlying medical conditions, or with a recent variety or combination of treatments (Zhu et al., 2016).

This species used in the fermentative production of bread, beer and wine is also involved in the treatment of antibiotic-related diarrhea and serves as a nutritional supplement commercialized as *S. boulardii* (Pérez-Torrado & Querol, 2016).

When analyzing strains isolated from clinical settings, they present low levels of virulence, though there have been strains used commonly, like in dietary supplements, that exhibit high virulence in infection models (Llopis et al., 2012, 2014; Pérez-Torrado et al., 2015). Hence the designation of opportunistic strain, since it only causes infection on elderly people, premature children, and overall patients suffering from immunosuppression (Pérez-Torrado & Querol, 2016).

These strains are said not to constitute a homogenous group, but are instead heterogenous mosaics that comprise alleles from several sub populations (Gianni Liti et al., 2009; Strope et al., 2015).

1.6. Mitochondrion and mitochondrial DNA

The mitochondrion is wildly called the “power-house” of the cell and is enclosed by two lipid membranes (Hsu & Chou, 2017; Saraste, 1999). The mitochondrion is a metabolic mediator, as it is a very important organelle fundamental for cell survival and several cellular functions. Among those is proliferation (Mitra, Wunder, Roysam, Lin, & Lippincott-Schwartz, 2009), the production of ATP, respiration, metabolite biosynthesis, ion homeostasis (Brookes, Yoon, Robotham, Anders, & Sheu, 2004), being key regulators of programmed cell death (Kroemer, Galluzzi, & Brenner, 2007), intervening in multiple signaling pathways (Jung et al., 2012; Müller, Lu, & Reichert, 2015; Solieri, 2010), mediate secondary massager signals to the nucleus (Al-Mehdi et al., 2012); and regulate aging (Basse, 2010; Zuin et al., 2008).

Every cell has numerous copies of the mitochondrial genome. It is indispensable to production of energy through oxidative phosphorylation, encoding several genes involved in the process that complement the action of those codified on the nucleus (Fritsch, Chabbert, Klaus, & Steinmetz, 2014). The number of molecules in each cell differ according to the species, tissue (Preuten et al., 2010), or culture conditions (Hori, Yoshida, Shibata, & Ling, 2009; Shay, Pierce, & Werbin, 1990), ranging between 50 and 200 copies per cell (Solieri, 2010).

This genome is a residue of an ancestral symbiont α-proteobacteria, having undertaken a significant size reduction, transferring genes to the nucleus (M. W. Gray, 2012; Wallace, 2007). Because mitochondria are such indispensable organelles and have a very small genome, they require nuclear-encoded proteins in order to express said genome (Turk, Das, Seibert, & Andrulis, 2013). That way, components encoded in both nuclear and mitochondrial genomes are necessary for mitochondrial biogenesis (Hsu & Chou, 2017).

This unique semi-autonomous organelle contains mitochondrial DNA and has its own ribosomes, which ensures that mitochondrial functions take place autonomously. It encodes a small subset of mitochondrial proteins, and also has an active role determining phenotypic diversity and fitness by interactions between mitochondrial and chromosomal genomes (Hsu & Chou, 2017; Solieri, 2010; Westermann, 2014). It’s growth is dependent of replication and expression of mitochondrial genes and nuclear-encoded proteins, and its replication and partitioning is not linked to the cell cycle (Hsu & Chou, 2017; Westermann, 2014). Most mitochondrial genomes are circular and double-stranded, forming linear head-to-tail molecules which are replicated by rolling circle mechanism (Bendich, 2010; Maleszka, Skelly, & Clark-Walker, 1991).

Mitochondrial DNA has a greater mutation rate than nuclear DNA (Jung et al., 2012). Because in most cases there is no recombination, its sequence variation is mostly due to the

accumulation of new mutations along radiating maternal lines (Torroni, Achilli, Macaulay, Richards, & Rgen Bandelt, 2006).

The majority of sexual eukaryotes have their mtDNA uniparentally inherited (Hsu & Chou, 2017), though that is not the case for budding yeasts.

Mitochondrial inheritance has been studied resorting to the budding yeast *S. cerevisiae*. The molecular machinery and cellular pathways, such as mitochondrial motility, tethering, fusion and fission, as well as mtDNA partitioning, affect the correct separation of mitochondria and mtDNA through cell division (Westermann, 2014).

This genome can be used in comparative genetics between closely related species in order to investigate chromosome architecture and evolution (Valach et al., 2011). There is a small amount of information related to mitochondrial genome variations within yeast species, although there is an increasing quantity of data available. Said data show the existence of substantial variations and a wide range of genome architectures among species (Jung et al., 2012).

1.7. Mitochondrial DNA of *S. cerevisiae* – history and morphology

It was in 1949 that mitochondrial inheritance was discovered in *Saccharomyces cerevisiae* (Ephrussi, 1949). The first mitochondrial gene sequenced was from this species (Françoise Foury, Roganti, Lecrenier, & Purnelle, 1998). As well as the first eukaryotic genome sequence, deriving from the strain S288C. Thousands of *S. cerevisiae* genomes have been sequenced since then to varying degrees of completions in the past few years (Engel et al., 2016).

The study of mitochondrial functions and cell biology in *S. cerevisiae* has advanced the understanding of mitochondrial genes (Wolters et al., 2015).

S. cerevisiae is, at this point, established as a reliable model for mitochondrial genetics, as a result of its functional similarity to the human DNA and the way it facilitates direct manipulation of the mitochondrial genome. The nuclear genome (comprising 16 chromosomes) of wild-type S288c was published in 1996 (Goffeau et al., 1996), and, a couple of years later in 1998, the first full mtDNA genome was sequenced from yeast (Françoise Foury et al., 1998).

The mitochondrial DNA of *S. cerevisiae* consists of long, linear molecules of heterogeneous size (Maleszka et al., 1991), about 85 kb, with a low gene density and extensive intergenic regions and a high A+T content (Françoise Foury et al., 1998; Miklos Zamaroczy & Bernardi, 1986). It is compressed into nucleoprotein structures called mitochondrial nucleoids (mt-nucleoids), which is a feature that allows for a better

accommodation in the cell, as well as spatial regulation of gene expression and protection from DNA damage (Bakkaiova et al., 2016).

S. cerevisiae has a mitochondrial genome that incorporates 35 genes (schematically represented in figure 2), most of them fundamental for oxidative phosphorylation (Turk et al., 2013). The coding portion contains a fundamental set of protein-coding genes for cytochrome *c* oxidase subunits I, II, and III (COX1, COX2, and COX3), ATP synthase subunits 6, 8, and 9 (ATP6, ATP8, and ATP9), apocyntochrome *b* (COB), a ribosomal protein (VAR1), genes for subunits 21S and 15S of ribosomal RNA, 24 tRNA (for translation), and the 9S RNA component of RNAse P (Françoise Foury et al., 1998; Groth et al., 2000; Hollingsworth & Martin, 1986). In sum, a group of mitochondrion-specific tRNA genes, two rRNA genes and a small set of protein-coding genes mainly involved in the electron transport system (Françoise Foury et al., 1998). It can also encompass up to 13 introns in three of its genes (COX1, COB, and in the 21S rRNA) (Lipinski et al., 2010). These genes' introns have been described in literature as such – COX1 holds seven introns, COB five, and 21S rRNA has one. Every intron is classified as group I or II, being mobile elements that suffer protein-assisted, autocatalytic RNA splicing (Lambowitz & Zimmerly, 2004; Lang, Laforest, & Burger, 2007). They can be distinguished according to their RNA secondary structures (Wolters et al., 2015).

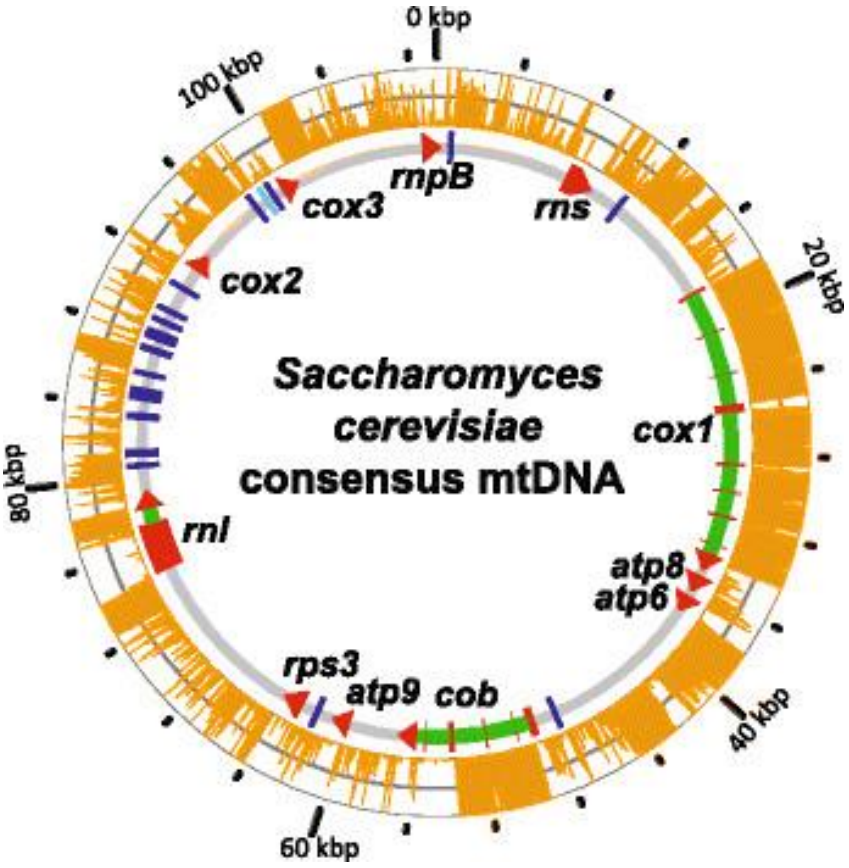

Figure 1 Consensus genome map of *Saccharomyces cerevisiae* mitochondrial DNA. Genes are indicated by red arrows, introns by green, and tRNA in blue. The light blue bars designate a sole tRNA encoded on the light strand. The orange

bars indicate the number of polymorphic sites within 100 bp windows (Wolters et al., 2015). The coding portion contains protein-coding genes for cytochrome *c* oxidase subunits I, II, and III (COX1, COX2, and COX3), ATP synthase subunits 6, 8, and 9(ATP6, ATP8, and ATP9), apocyntochrome *b* (COB), a ribosomal protein (VAR1), genes for subunits 21S and 15S of ribosomal RNA (LSU and SSU), 24 tRNA (for translation), and the 9S RNA component of RNAse P (RPM1) (Françoise Foury et al., 1998; Groth et al., 2000; Hollingsworth & Martin, 1986).

The intergenic regions comprise a big portion of the genome. They are characterized by long AT stretches and short GC clusters spread throughout (M de Zamaroczy & Bernardi, 1986; Weiller, Schueller, & Schweyen, 1989). They also have in their midst ori/rep/tra sequences. These are essential for the transmission of the mitochondrial DNA (Petersen et al., 2002; Wolters et al., 2015). GC clusters consist of several direct and indirect repeats (Groth et al., 2000), and are thought to influence mitochondrial DNA stability (Spírek, Soltésová, Horváth, Sláviková, & Sulo, 2002) and recombination (Dieckmann & Gandy, 1987) as they can also sometimes be inserted into coding genes, which ultimately results in alterations in the size of gene products (Butow, Perlman, & Grossman, 1985; Dujon, 1980; Hudspeth et al., 1984), which are important structural changes that can interfere with gene regulation (Hausner, 2003).

These characteristics of the genome make it more difficult to prepare, sequence and align sequences (Wolters et al., 2015).

A big portion of the yeast mitochondrial genome is arranged in linear concatemers, being more prominent in mother and nondividing cells, whereas growing buds show an arrangement in monomers (F. Ling, Hori, & Shibata, 2007; Feng Ling & Shibata, 2002; Solieri, 2010). As mentioned, mitochondrial DNA in *S. cerevisiae* is mainly kept in linear molecules that can vary a lot in length. Some studies infer that the molecules in circular configuration are templates for amplification. It supposedly occurs by a rolling circle mechanism that produces concatemers of linear arrays with numerous genome units (Westermann, 2014).

This species can survive purely by obtaining energy through fermentation, not needing mitochondrion to sustain itself (Jure Piskur, 1994). This is a rare feature among Eukariota, although the presence of mitochondria is essential to several metabolic functions (Lipinski et al., 2010). There are circumstances, though, in which this yeast is able to thrive without efficient mitochondria by fabricating respiratory-deficient petite mutants with the entirety or part of their mtDNA deleted (Fritsch et al., 2014). These mutants can appear because of mutations on mitochondria or even the lack of the organ altogether, or also as a result of mutations in nuclear-encoded genes responsible for oxidative phosphorylation (Bernardi, 1979; Goldring, Grossman, & Marmur, 1971; Heslot, Goffeau, & Louis, 1970).

1.8. Sexual reproduction and transmission of the genome

S. cerevisiae's mitochondrial system has a tendency to maintain a cell population of homogeneous copies of mtDNA, having associated, therefore, a process of homoplasmy, which means that all copies of mitochondrial DNA in the cell are alike (Lipinski et al., 2010). The species' frequency of genotypes is maintained no matter which strains are combined through mating. Diploid offspring establish homoplasmy after about 20 generations, through mitotic segregation, meaning that a low number of copies of mtDNA from the mother cells are transmitted to the new cells (Bonnefoy, Remacle, & Fox, 2007; Thrailkill & Birky, 1980).

During sexual reproduction in budding yeast there are two mating types: MATa and MATα. Studies hypothesize that the mtDNA from the MATα parent is eradicated shortly after the mating (Xu et al., 2000; Yan & Xu, 2003). The cells inherit 30-100 copies of mtDNA molecule (Miyakawa, Sando, Kawano, Nakamura, & Kuroiwa, 1987) from both parent cells.

The transmission of mitochondrial genome is affected by several genetic elements (Petersen et al., 2002). Even though the inheritance is biparental, the distribution of alleles is difficult to predict, because of different combinations between parental mtDNA, mitochondrial recombination and the loss of heteroplasmy (Westermann, 2014), as well as the easiness with which mobile elements of the mitochondrial genome are transferred laterally in populations (Goddard & Burt, 1999). All these aspects of the nature of the genome in study can explain possible divergence between nuclear population structures and those of mtDNA (Wolters et al., 2015).

Some studies suggest that population structure of nuclear genomes is generally followed by intraspecific diversity in mitochondrial genomes by studying the coding and intronic sequences, but there have been proof of the contrary, with replication not being tied to cell cycle (Lecrenier & Foury, 2000), which facilitates higher mutation rates for mitochondrial DNA (F. Foury, Hu, & Vanderstraeten, 2004).

Some populations seem to exhibit a recent expansion of mobile elements of intergenic sequences. Studies also reveal new introns never before found on the species as well as populations without introns thought to be conserved. Introns can contrast depending on the strain (Wolters et al., 2015).

The mitochondrial genome diversity seems to be limited to the population, which give information on recent evolutionary occurrences (Wolters et al., 2015).

1.9. *Saccharomyces cerevisiae* mtDNA recombination

Mitochondrial DNA recombination is a prevalent occurrence in plants (Arrieta-Montiel, Shedge, Davila, Christensen, & Mackenzie, 2009; Galtier, 2011), fungi (Dujon, Slonimski, & Weill, 1974; Fourie et al., 2013), protists (Michael W Gray, Burger, & Lang, 1999), and invertebrates (Ladoukakis & Zouros, 2001).

Budding yeasts have two phases, one haploid and one diploid in their life cycle that can both be transmitted asexually by mitotic division (Fisher, Buskirk, Vignogna, Marad, & Lang, 2018).

After sexual reproduction, *S. cerevisiae* mitochondrial DNA is inherited from both parents, allowing for a heteroplasmic state that lasts for a few generations and permits the occurrence of recombination. This occurrence has been described in yeast genomes since 1974 (Dujon et al., 1974). During the sexual cycle, the parental mitochondria fuse to form one organelle in the new zygote. Therefore, mitochondrial proteins and parental genomes mix in the same partition. The recombinant genome can then be found where both genomes interact, in the medial bud (Strausberg & Perlman, 1978; Zinn, Pohlman, Perlman, & Butow, 1987).

Some studies propose the existence of favored recombinogenic positions, so called hotspots. Most of these happen within intergenic regions, which could be related to the occurrence of repetitive and palindromic elements on those places, favoring their use as templates for recombination. Some happen within the tRNA gene cluster and rRNA regions. In the annotated genes, recombination hotspots have been found in the introns. As a result it is vastly assumed mitochondrial recombination occurs preferentially in non-protein-coding regions (Perez-Martinez, Broadley, & Fox, 2003; Zinn et al., 1987).

Recombination can occur through short repeats (Bernardi, 1979), or homologous pairing (Dujon et al., 1974), generating mutant molecules with a different gene order and/or intergenic deletions (Clark-Walker, 1989; J Piskur, 1988).

1.10. Evolution of the species

Several studies of the nuclear genome of *Saccharomyces cerevisiae* place the origin of the species in China, followed by a large number of separate domestication events, and the Taiwanese wild lineage is thought to be the most divergent population described until this day (Wang et al., 2012). Currently, *S. cerevisiae* can be found in various environments, being them human associated or wild, from fermented beverages to plants and insects.

Among the most important evolutionary events that influence the degree and variety of mutations and the proportion of adaptation are genome duplications. The biggest source in

variation throughout the evolution of *Saccharomyces cerevisiae* seems to be copy number variation. This structural duplication or deletion consists of portions of the genome that are repeated in various ways such as number and length (Peter et al., 2018).

Another big influencer of genome evolution is whole genome duplication, the synchronized replication of each chromosome. This alteration can be traced from ancient occurrences in recent genomes (Fisher et al., 2018). After whole genome duplication, nuclear genes acting on the mitochondria show greater evolution rates than genes with other functions, which supports the occurrence of petite mutants that don't need mitochondria to survive (Jung et al., 2012).

These kinds of evolutionary events are thought to intensify virulence and adaptation to stress in pathogenic fungi (Gerstein et al., 2015).

With the rise of genetic studies involving *S. cerevisiae*, there is a big number of published sequences, although the ratio from laboratory strains to natural isolates still tends a lot towards the domesticated (Peter et al., 2018).

1.11. Goals

With this work, we intend to better understand the evolution and origin of *S. cerevisae,* corroborating or not the notion that the species initially originated from China. We also aimed to gain clarification regarding the origin and phylogeny of the so-called clinical strains that seem to be a heterogenous group from different populations. Lastly, we meant to draw a clear picture of the relationship between the strains originated from different geographic locations and sources of isolation.

2. Methods

2.1. Collection of *S. cerevisiae* mitochondrial genomes

The data for this study was obtained on the online database with publicly available genetic data, the National Center for Biotechnology Information Search database (NCBI). The data was obtained in two formats: fasta sequences of the available complete mtDNA of *S. cerevisiae* deposited in the "nucleotide" section of NCBI (this data is completely annotated in terms of genes); and raw next generation sequencing files from genomic data of *S. cerevisiae* deposited in the Sequence Reads Archive of NCBI (SRA).

While the first type of data was directly obtainable from NCBI and usable in this study, the raw data required the alignment with a reference genome using bioinformatics' tools. For that task we requested the help of Post Doc Eduardo Conde-Sousa (CBMA). The procedure is described below.

The SRA toolkit (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft/) was used to programmatically download all *Saccharomyces cerevisiae* data housed within SRA. In order to extract the regions of interest and create a final fasta file containing all samples, a first alignment of each individual sample against the reference genome *Saccharomyces cerevisiae S288C* (GenBank: https://www.ncbi.nlm.nih.gov/genome/?term=saccharomyces+cerevisiae) was performed using the Burrows-Wheeler Aligner (BWA). Then samtools and bcftools were used to convert between file formats and extract target regions. Finally, mafft was used to align all samples against each other.

In parallel with the sequence retrieval, we also gathered information on the sequences. For some sequences, the "BioSample" section of NCBI contained data on the isolation source, geographical location and even collection date, in some cases. Most of them provided a small abstract with information on the work done with the sequences and we resorted to that and the published paper, when applicable, to gather data for the rest of the samples.

2.2. Analysis of the gene content of *S. cerevisiae* mitochondrial genome

Using the sequences directly extracted from the "nucleotide" section in NCBI, it was perfectly discernible that the complete mtDNA genomes would not facilitate an alignment using current multiple alignment software tools. This was caused by two main reasons: one is the

extremely high prevalence of AT base pairs throughout the mitochondrial DNA sequence of this species, as well as the high number of repeats and their unpredictability, and differences in size. About 85% of the reference sequence of mtDNA is T and A base pairs; apart from a common set of genes, different mtDNA genome in S. *cerevisiae* have different content. This is easy to access using the annotations provided by NCBI.

In order to compare the mtDNA between individuals, a common comparable portion of the mtDNA genome was compiled. This was done by excluding all the intergenic regions composed only by T and A and removing genes or pseudogenes that were not common between all genomes. We did some local alignments of certain portions to confirm annotations in NCBI. A total set of 5475 base pairs was obtained. We reached a final group of smaller, coding sequences that allowed us to analyze the large group of samples. All further analyses were focused on this portion of the sequence.

Following a quality step where sequences with more than 10% of unread positions were excluded, the remaining sequences were submitted to the software IMPUTOR (Jobin, Schurz, & Henn, 2018) to estimate/impute some of the remaining unread positions given the overall alignment and phylogenetic reconstructions of set of sequences in MEGA7 (Kumar, Stecher, Tamura, & Dudley, 2016).

We used DNAsp (DNA Sequence Polymorphism) (Librado & Rozas, 2005) and Arlequin (Excoffier, Laval, & Schneider, 2007) to obtain general statistics from the full dataset. The DNAsp software analyses nucleotide polymorphisms from DNA sequences and Arlequin has statistical tests that provide information on genetic and demographic features. For further uses, as the dataset was too large for visualization, we used a reduced dataset (below two thousand) that corresponded to sequences from which we had at least information on ecological source and geography.

2.3. Reconstruction of a detailed *S. cerevisiae* mitochondrial DNA phylogenetic tree

After the sequences were aligned against the S288C reference genome using the Bioedit tool (Hall, 1999), they were put through a number of processes in order to obtain a phylogenetic tree. First, we used mtDNA GeneSyn (Pavesi, Mauri, Iannelli, Gissi, & Pesole, 2004). GeneSyn is a software tool that allows automatic detection of conserved gene order from annotated genomes. We started with the tool “Calculate Polymorphisms”, saving it as a .txt file. Then, that file was transformed and exported using the tool “Export to Binary File” which was saved in a .rdf format.

Network (Bandelt, Forster, Sykes, & Richards, 1995) generates evolutionary trees and networks from genetic, linguistic, and other data. Network can then provide age estimates for any ancestor in the tree. Opening the .rdf file with the Network software, we could then see the polymorphic positions in every sequence and make adjustments such as the weight of each position, according to the needs of the analysis. After that we obtained an .out file, using the tool “Reduced Median”.

Network Publisher is an extended version of the free Network software's Draw subprogram, reading Network results .out.

Network was used to build an initial phylogeny, with mutations clearly displayed in the branches. A phylogenetic tree represents the evolutionary relationships among a set of organisms or groups of organisms, called taxa. The tips of the tree, or leaves, represent groups of descendent taxa and the nodes on the tree represent the common ancestors of those descendants. We aim to investigate the presence of homoplasy or recombinants in the dataset which are clear given the properties of network.

2.4. Recombination and linkage disequilibrium in *S. cerevisiae* mitochondrial DNA

In order to investigate recombination patterns across the *S. cerevisiae* mtDNA genome we used two approaches. For linkage analyses, only SNPs with frequencies higher than 10% were used.

The first approach consisted of using Haploview (Barrett, Fry, Maller, & Daly, 2005), a software tool that calculates basic LD statistics (D’ and r^2 – referring to coinheritance and allele frequency) from SNP files (PED) format. The software determines block of linkage disequilibrium (LD) across the full range of the sequence.

A second approach was the use of the PHASE software (Crawford et al., 2004). PHASE estimates the phase of the sequences, meaning that polymorphisms along a chromosome will be placed in the most probable haplotypic combination. The software also estimates recombination rates between polymorphisms. The software in this specific instance was only used with this purpose. For that, we combined the mtDNA in two but selected that the phase was already known. This way, recombination rates were calculated from the existing known data.

2.5. Genetic structure in *S. cerevisiae* mitochondrial DNA

As mtDNA of *S. cerevisiae* is a highly recombining system, a phylogenetic analysis is not fully appropriate. Nevertheless, we employed a neighbor-joining algorithm. Such algorithm is a distance-based algorithm and so it will define clusters based on general proximity between sequences and not assuming necessarily a common ancestor for the clades. We employed MEGA7 (Kumar et al., 2016) with a Maximum Composite Likelihood model and gamma-distributed rates (four). A bootstrap of 100 iterations was calculated.

Another method employed was principal component analysis (PCA). This is a mathematical approach that will attempt to describe the diversity in linear vectors. While nowadays the method is mostly employed in genome-wide data we will verify how the PCA vectors will group the diversity found in the mtDNA genomes. The software tool SmartPCA (Zhang, 1999) was be employed to do this.

A final approach is the application of individual ancestry estimation algorithms. These algorithms, also mainly used in genome-wide data, will categorize the diversity into a specific number of components (named K) defined by the user. Such components can represent ancestral population groups (although it is not necessarily true). We calculated K=2 to K=5 using sNMF software.

3. Results

3.1. Dataset

We obtained 184 *S. cerevisiae* mitochondrial DNA genomes from the “nucleotide” section of NCBI and a total of 12357 genomes from the raw data deposited in SRA (Sequence Read Archive). The 184 complete genomes from NCBI were used to define a comparable dataset between all genomes by detecting a common set of genes present in all mtDNA genomes. Figure 2 represents the regions and set of genes used in the following analyses.

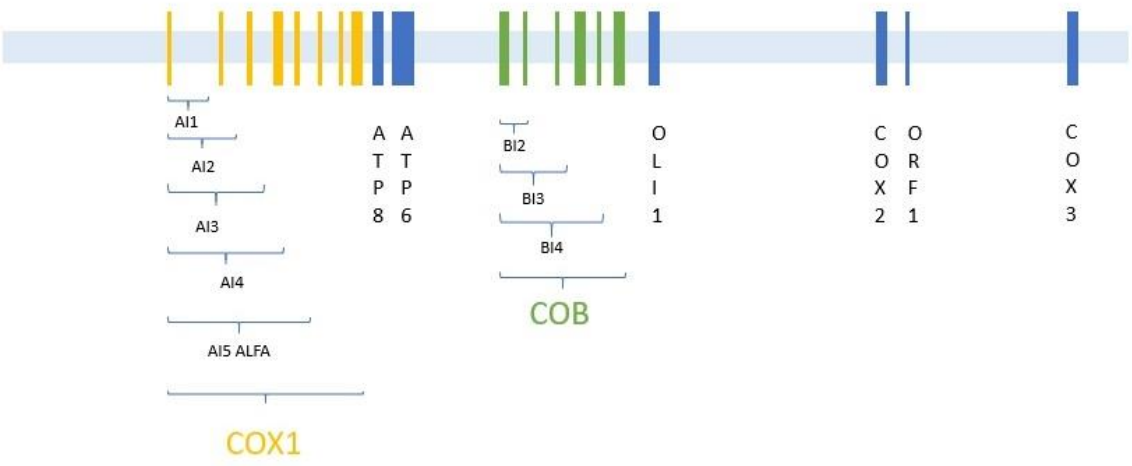

Figure 2 Representation of the coding portions highlighted from the genome of S. cerevisiae (85kb), containing the main genes of the species present across the different complete mtDNA genomes.

Following the cut of the different genomes, and an exclusion of sequences with an excess of unread portions, we obtained a final dataset of 6647 genomes with 5475 base pairs that was further submitted to IMPUTOR to obtain an estimate of as many of these missing positions in as many sequences possible.

In order to characterize our dataset. we attained general statistics using DNAsp and Arlequin. These statistics are summarized in Table 1.

Table 1 Summary statistics for the final dataset of 6647 *S. cerevisiae* mtDNA genomes

Number of sequences	**6647**
Size	5475
Number of polymorphic sites	526
Number of haplotypes	1265
Number of observed transitions	273
Number of observed transversions	291
C compositions (%)	13.09
T compositions (%)	41.33
A compositions (%)	30.95
G compositions (%)	14.63
Gene diversity	0.9665 +/- 0.0032
Mean number of pairwise differences	38.632869 +/- 16.797110
Nucleotide diversity (average over loci)	0.007354 +/- 0.003535
Tajima's D p-value	0.10600
Fu's FS p-value	0.59400

3.2. Recombination

Because of the nature of the mitochondrial DNA of *S. cerevisiae*, with a high recombination rate and high concentration of AT, all attempts at building a full network representative of the diversity of the species were futile. The level of homoplasy was enormous. Attempts to simplify the phylogeny, by weighting the characters based on previous runs, only highlights the presence of recombination represented by square central figures with continuous variants in the molecule from one side of the square and variants from other section in the other.

As made clear by the representation in Figure 3, the recombination pattern is evident throughout the samples gathered.

As a result, to properly analyze the effect of recombination on diversity patterns, we explored linkage disequilibrium using the D’ (figure 4A) and r^2 (figure 4B) statistics. As somewhat expected, SNPs located in the same exon/gene are in linkage, however every single block never extends across two genes, suggesting a permanent break of linkage always occurring between genes.

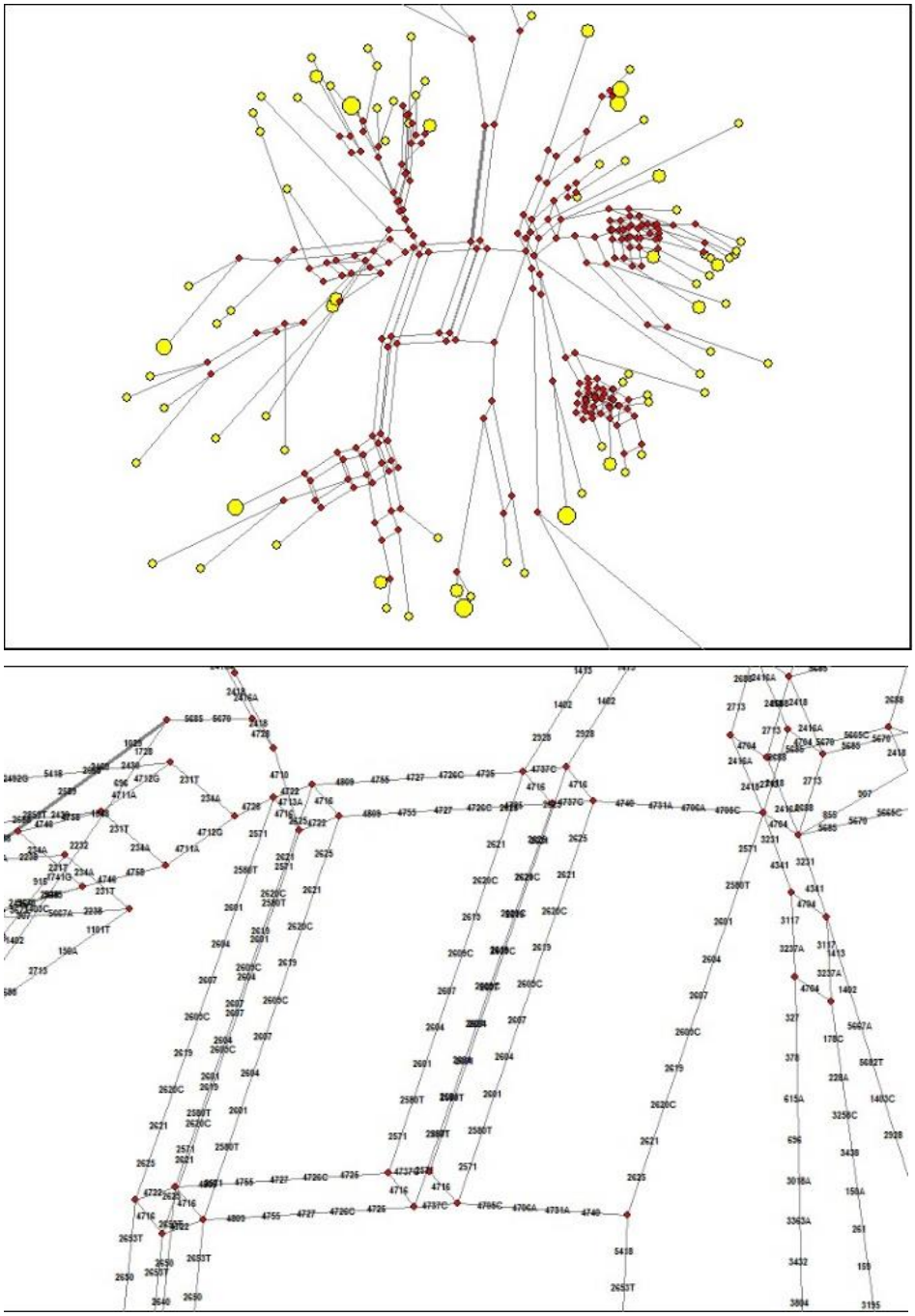

Figure 3 One of the networks obtained with 146 sequences of Saccharomyces cerevisiae. A. general pattern; B. detail of the central section.

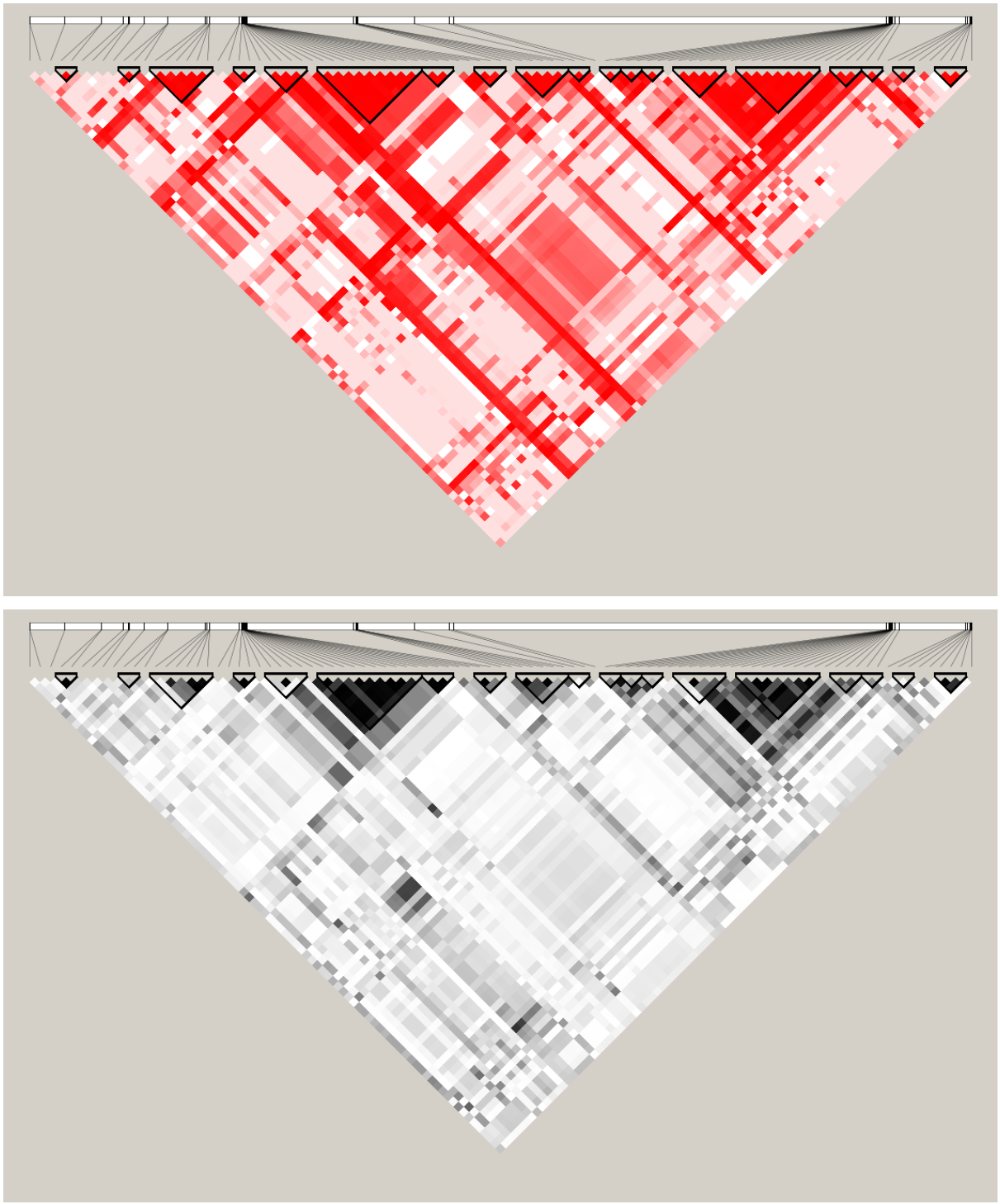

Figure 4 Patterns of linkage disequilibrium (LD) across the mtDNA genome of S. cerevisiae displayed using D' (A.) and r^2 (B). Blocks of Linkage Disequilibrium are highlighted.

PHASE estimates a frequency of recombination across the mtDNA genome (figure 5). Although the method is too sensitive for typing errors, it seems evident that recombination in general is prevalent across the entirety of the mtDNA genome.

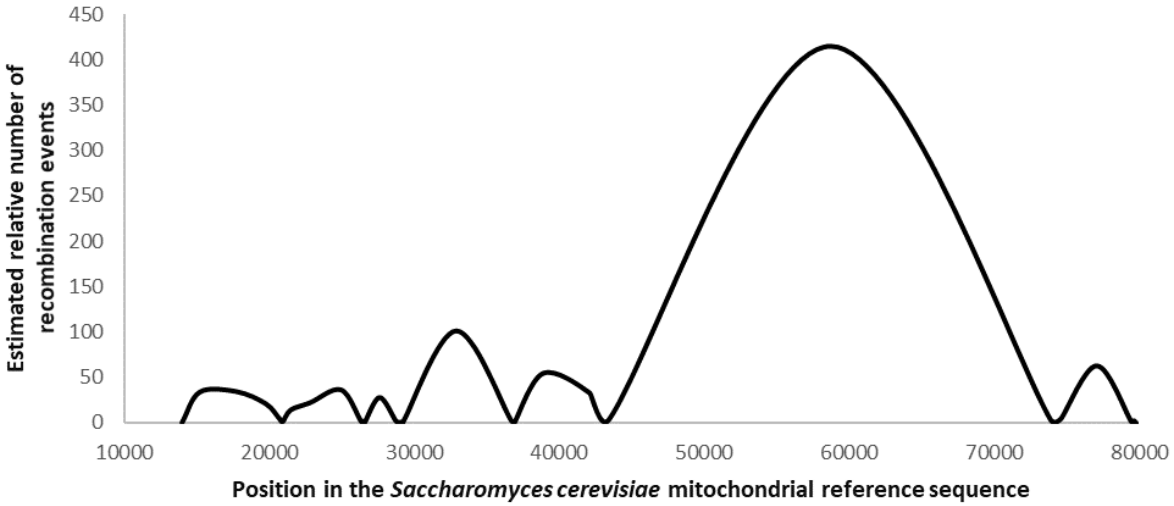

Figure 5 Estimated relative amount of recombination in the *S. cerevisiae* mitochondrial genome in relation to an estimated background recombination rate

3.3. Genetic structure

We investigated the genetic structure of the mtDNA genome of *S. cerevisiae* using three methodologies, all based on general patterns (clustering and genetic distances) that were more fit for recombining systems: neighbor-joining trees, PCA and sNMF, an individual ancestry estimator. As the number of sequences/organisms was too large to display in most analysis, we opted to reduce the sample group to only those which information on source and/or geography was known leading to a dataset just below 2000 samples.

Figure 6 displays the neighbor-joining tree with the geography highlighted while in figure 7 the same tree is highlighting the source (environment) of the samples.

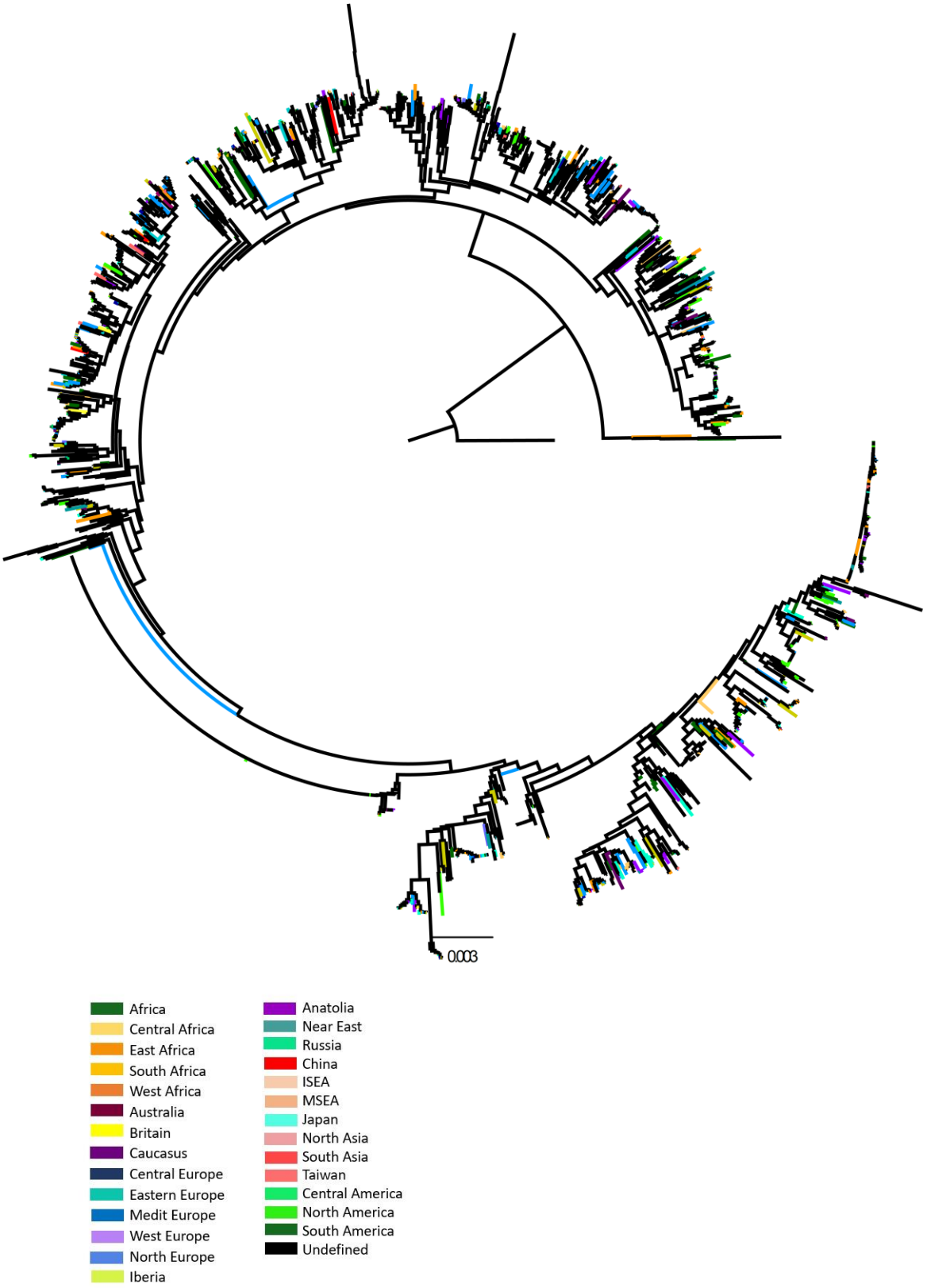

Figure 6 Neighbor-joining tree of 1985 *S. cerevisiae* mtDNA genomes. Colors in the branches correspond to geographic locations.

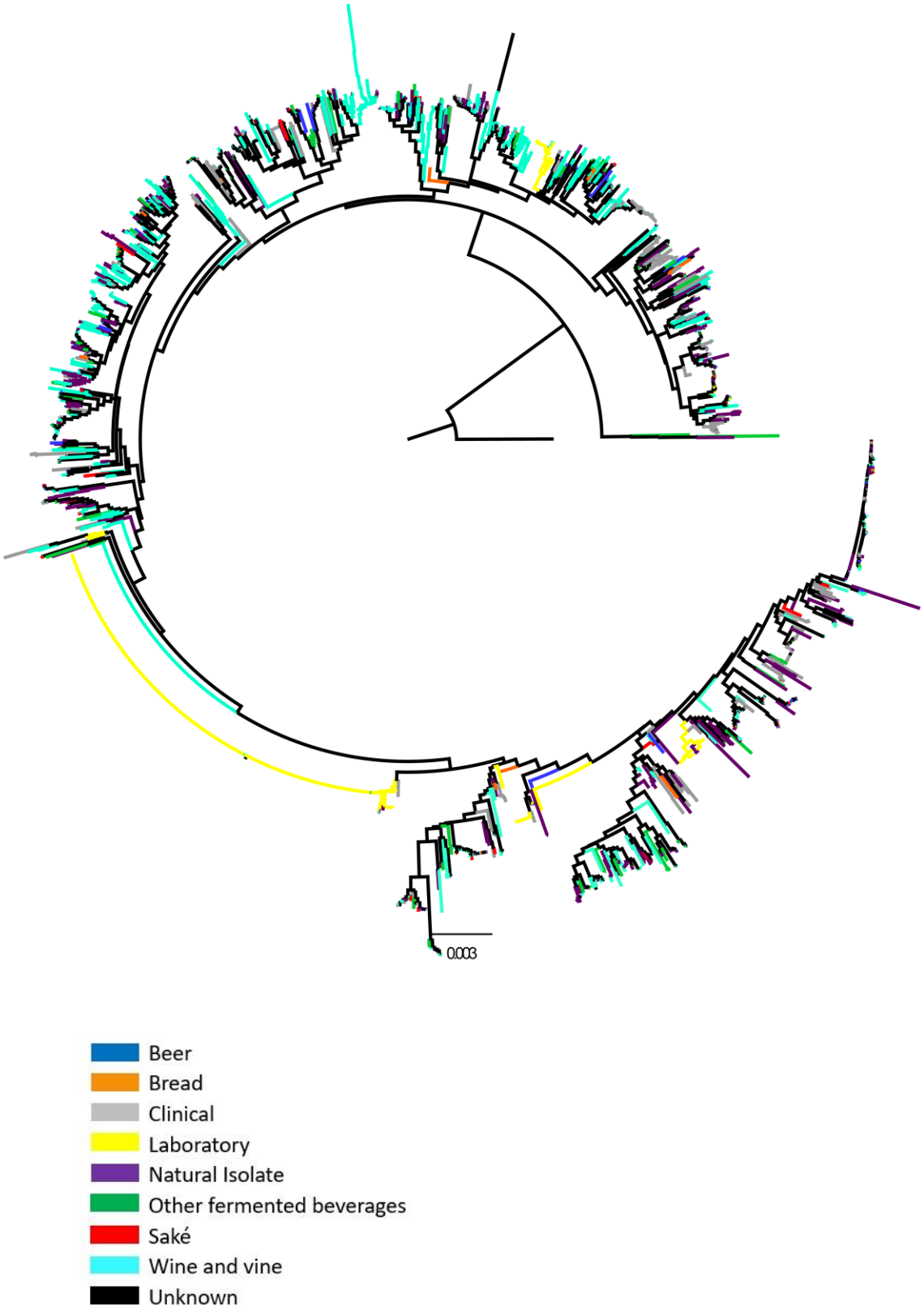

Figure 7 Neighbor-joining tree of 1896 Sc mtDNA genomes. Colors in the branches correspond to source (environment) locations.

The five major PCAs, explaining in total 55% of the diversity are displayed, comparing pairs of principal vectors in Figures 7 (pinpointing their geography) and 8 (highlighting their type of source).

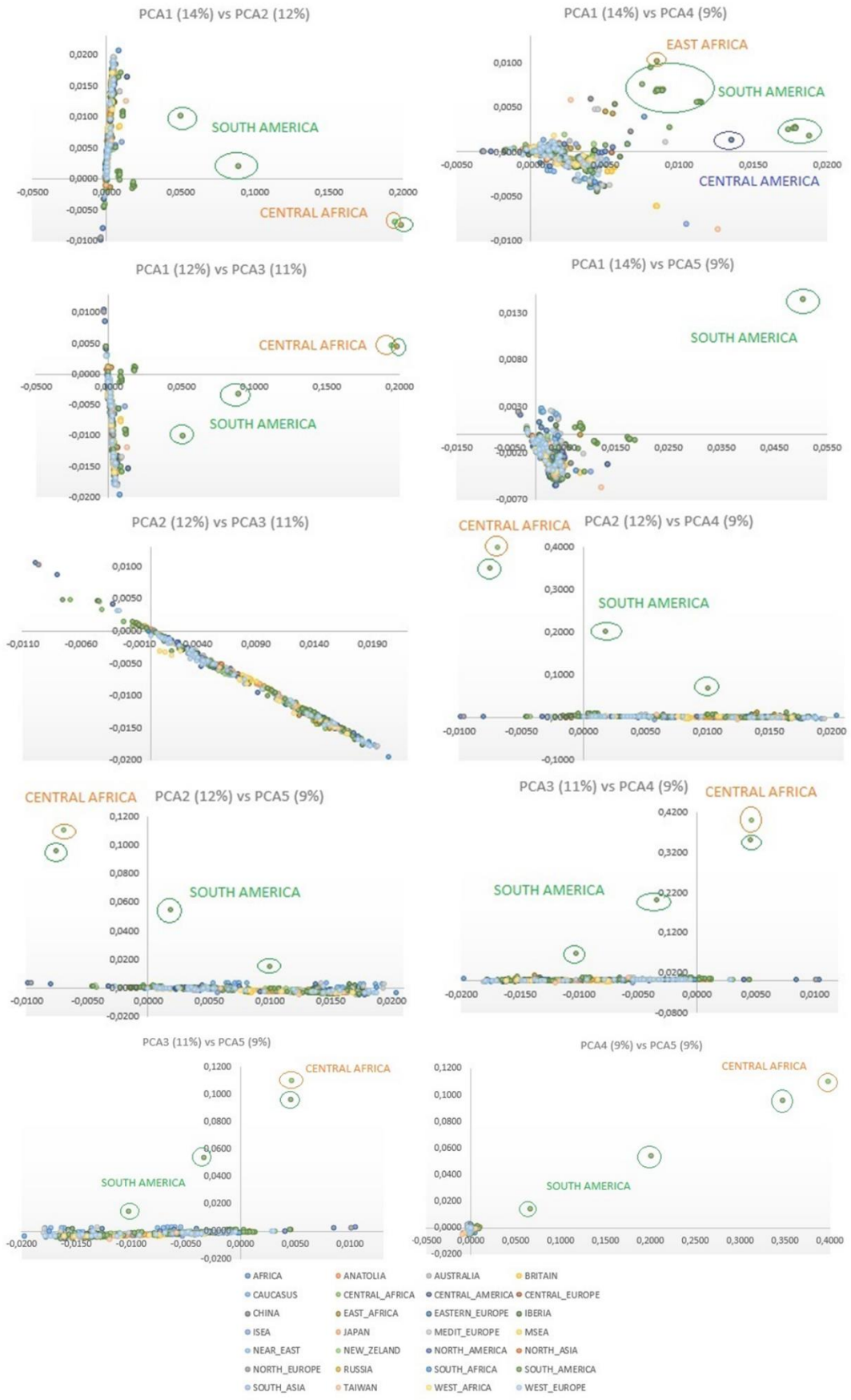

Figure 8 Comparison of the first five principal component in the PCA of the mtDNA genomes of *S. cerevisiae*, indicating their geographic location.

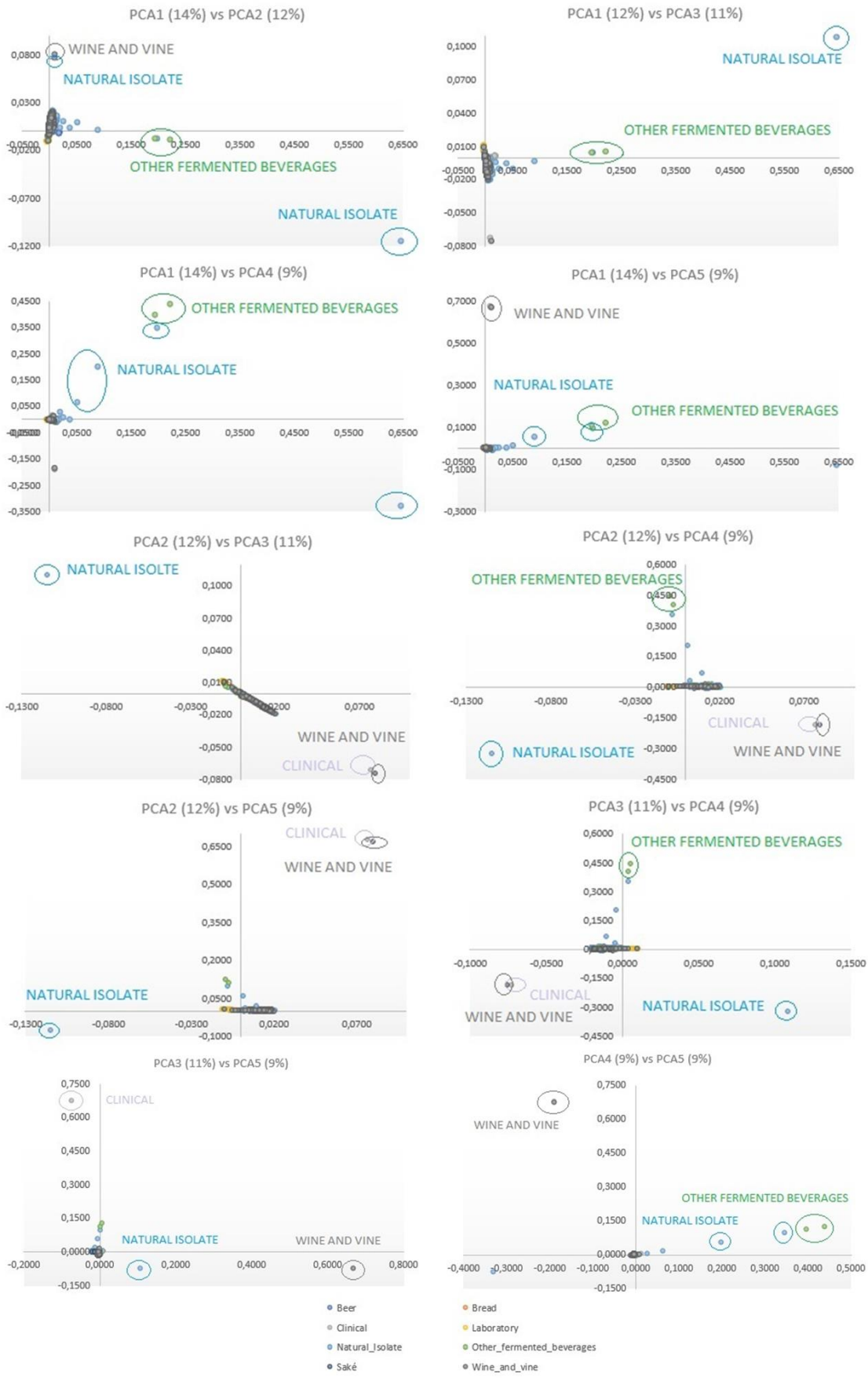

Figure 9 Comparison of the first five principal component in the PCA of the mtDNA genomes of *S. cerevisiae*, indicating their type of source.

Finally, Figures 9 and 10 display the results of sNMF, considering between 2 and 5 components, separating samples by geography and source respectively.

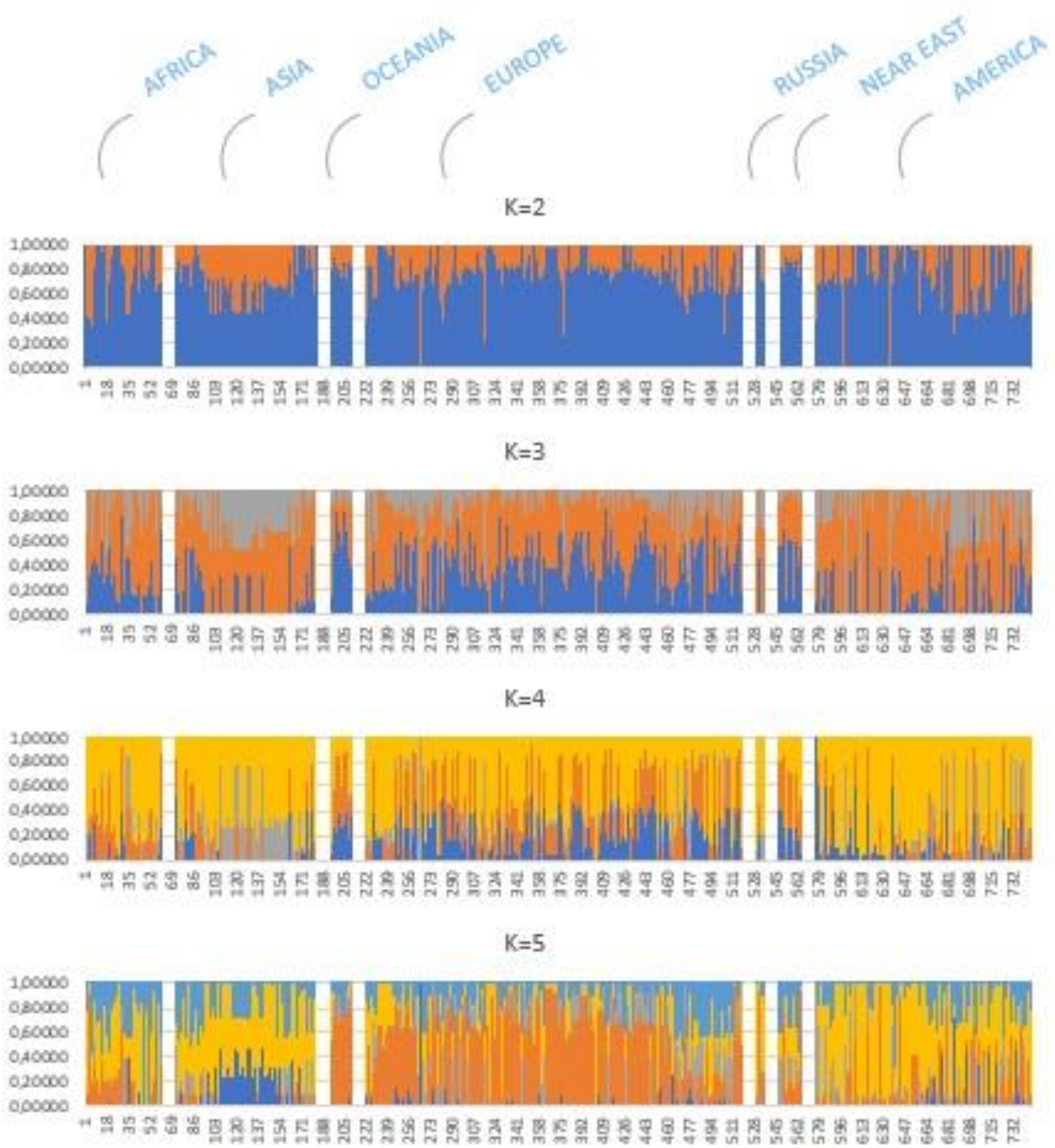

Figure 10 Individual ancestry estimates of *S. cerevisiae* mtDNA genomes using the sNMF algorithm. Splits correspond to geographic locations.

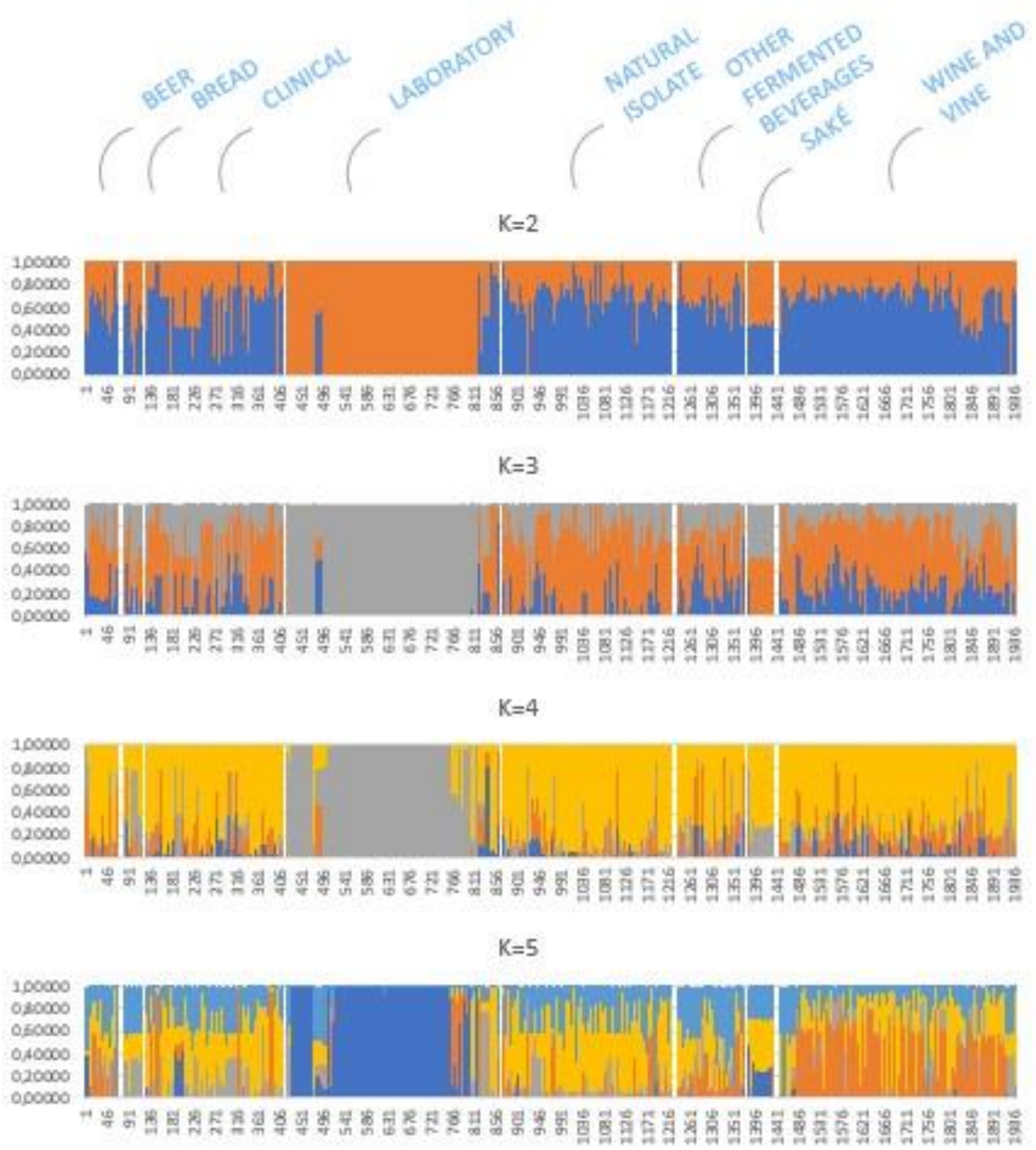

Figure 11 Individual ancestry estimates of S. cerevisiae mtDNA genomes using the sNMF algorithm. Splits correspond to source of the sample.

Geography

Analyzing the three methods, it seems evident there is very little geographic clustering as far as *S. cerevisiae* mtDNA is concerned. The neighbor-joining tree samples from different locations are extremely interleaved and intertwined between different geographies.

Again, that point is evident in the PCA. While a few samples are placed more distantly, in PC1 to PC5, it is difficult to discern any major trend. These misplaced outsider points are mostly from South America (but also Africa) but they are exceptions and could even be low quality samples. In general, all samples are displayed within a continuous trend (gradients)

established by the pairs of principal components. However, the general picture is that these samples do not display any type of geographic clustering and are placed, independently of geography, along these gradients.

The situation is very similar for sNMF. For K=2 no pattern is visible. For K=3 there seems to exist some preponderance of the blue component in that analysis in Europe and Africa, while Asia shows higher frequency of the grey component. K=4 patterns are less discernable and at K=5, again an orange component seems more frequent in Europe (and associated Russia and Near East). A darker blue component is basically only displayed in Asia. Nevertheless, the patterns are very mild and the geographic clustering of *S. cerevisiae* mtDNA is basically inexistent.

Source

In terms of source, the clustering is actually stronger. In the case of the neighbor-joining tree, laboratory strains are mostly placed within a single branch, while wine and vine strains are located across a set of well-defined clades. Nevertheless, all the other categories seem spread across the tree, namely the clinical strains.

In terms of PCA, the results are as displayed for the geographic locations. While there are some samples outside the major group of sequences, from “other fermented beverages” and “natural source”, most of the samples fall within a gradient between the different pairs of principal components without displaying any specific grouping.

In terms of sNMF, we can see at K=2 that the laboratory strains display a clearly different pattern from other samples that is maintained for higher Ks, having a nearly specific component in K=5. For K=3 to K=5 the samples from “sake” seem to display different proportions of the different components, though it is of minor notoriety. At K=5, the “wine and vine” lineages display a high frequency of the orange component, suggesting some deeper level of clustering as displayed by some major branches in the neighbor-joining tree.

4. Discussion

Right at the beginning of this project, it was clear that the alignment and manipulation of the sequences gathered was extremely difficult to perform given even the most sophisticated alignment tools. The great number of AT base pairs, the extension of the intergenic sequences, as well as the fluctuating size of the genome and its gene content, caused difficulties in analyzing the mitochondrial DNA. To date, the number of studies focusing on this molecule is very scarce in literature. Even studies directed at genomics tend to completely exclude this molecule in the analysis. For that reason, we reduced the length of the sequences, by identifying the more prevalent/common coding genes among the 184 genomes from the “nucleotide” segment of NCBI, eventually sizing them to 5475 base pairs. By performing this task, we developed an analysis frame for the study of *S. cerevisiae* mitochondrial genome in the future.

After the establishment of an equivalent mtDNA sequence, we obtained basic statistics for the total set of individuals. One stand-out point on a molecular level is that, although we restricted our analysis to the coding sequences, the percentage of T and A is very high (72.28% of the total nucleotides) comparable with the percentage above 90% in the full DNA molecule. One question that must be addressed for this smaller sequence is whether this segment will provide enough diversity to render a good level of discrimination between sequences. The results show that the diversity is quite high, with over 500 SNPs detected. Considering the recombination effect, these 500 SNPs contribute to a very high haplotypic diversity and a good discrimination between sequences. There were also no signs of apparent selective pressure on the evolution of the *S. cerevisiae* mitochondrial genome, which is a preferable feature in genetic systems when cataloguing diversity.

Contrary to most mitochondrial DNA molecules of eukaryotic individuals, *S. cerevisiae* and other yeast’s mitochondrial genomes demonstrate undoubtedly the occurrence of recombination at high rates throughout its extension. Recombination patterns were observed very clearly through the Network analysis, linkage disequilibrium (LD) analysis, and estimation of phase recombination rates, confirming the presence of recombination throughout the molecule. It is of our opinion that recombination plays a very important role in the evolution of the genome, not only instigating very high haplotypic diversity through new combinations of existing variants, but also by prompting a probable fast disruption of newly formed haplotypes leading to lack of general patterns when considering the source of isolation and geographic clustering as we observed. High recombination throughout the molecule can also be responsible for the instability in the gene content across different genomes, as cases of large

deletion and gene conversions are not uncommon in high recombination regions of the human genome.

Linkage disequilibrium (LD) is useful to provide information on how a population is structured by identifying the association between alleles. It allows us to understand the recombination patterns in the molecule and also to direct the following analyses. Higher LD only occurs within genes or even just exons of the genes. Haplotypic blocks occur in intragenic regions only, and they are further separated by intronic or intragenic regions. This result is supported by the high number of recombination events estimated by PHASE. It is likely that the high concentration of AT base pairs throughout these regions, as well as high number of repeats, is a major biochemical feature that is underlying the high recombination rate.

Considering the steep recombination rate, and the lack of haploblocks, it is clear that examinations based on the establishment of lineages and phylogenetic analyses are inadequate. We based our examination on the establishment of clusters that could agglomerate common diversity, highlighting geographic structuring or common source. The neighbor-joining tree, although a phylogenetic method, is based on genetic distances between sequences. PCA and sNMF, an individual ancestry estimator, also aim to establish patterns between sequences. The lack of LD through most of the molecule makes these analyses feasible.

The neighbor-joining tree shows us that this species' strains do not display a pattern of distribution, either considering location or source of isolation as the variable under scrutiny. The only slight clustering of notice is of the laboratory strains that generally form a single branch, and wine and vine strains appear in some well-defined clades, but not an individual one or even a limited number. Furthermore in PCA, location-wise, most data formed single clusters near the axis of the graphics with a small number of outliers corresponding to South America and Central Africa samples. These single clusters are extended throughout one on the vectors, but it does not establish any type of geographic trend. Similarly, that same pattern was observed for the source of isolation with outliers prevalently of the natural source and other fermented beverages.

The sNMF study doesn't show a preponderant geographic clustering, supporting the results from the neighbor-joining and the PCA. The only instance where some hints of clustering are visible is related to the source, mainly related to laboratory strains and a hypothetical wine and vine component, and even these are only possible when submitting the algorithm to a higher number of components. Nevertheless, these specific clusters at five components are more preponderant in these groups but not all the samples have it. One important point is that even though a previous large study suggested that Asia was the origin point of *S. cerevisiae* we did not obtain any patterns that reiterate that, neither as outliers in Asia or Asia representing a region of higher diversity of the yeast. In a similar fashion as

obtained with nuclear genomes, clinical strains did not show any type of clustering. This seems to suggest that the acquisition of virulence was probably the result of multiple events.

As we gathered the mitochondrial DNA from the SRA section of the NCBI database, we also obtained the nuclear DNA of those same individuals. As a result, we will be able to do further perform studies similar to this one, but focusing on the remaining, which will allow us to corroborate/contrast the results of this study. We are particularly interested in inspecting the geographical patterns of the nuclear genome, considering the apparent lack of pattern in the mitochondrial genome.

5. Conclusion - Forensic genetics context

Forensic genetics are best known as the investigation of human DNA to help in criminal cases, even though there are several more areas in which forensic methods excel at. For instance, the study of nonhuman genetic material provides insight into situations such as animal attacks, trafficking of illegal species, bioterrorism, or even fraudulent food composition.

The study of species like the fungi *S. cerevisiae* can be applied to some of the areas mentioned before. In the context of forensic genetics, it can provide useful insight into product labeling as well fraudulent food composition since yeast species are wildly used in the production of alcoholic beverages, bread, among others.

6. References

Al-Mehdi, A.-B., Pastukh, V. M., Swiger, B. M., Reed, D. J., Patel, M. R., Bardwell, G. C., … Gillespie, M. N. (2012). Perinuclear Mitochondrial Clustering Creates an Oxidant-Rich Nuclear Domain Required for Hypoxia-Induced Transcription. *Science Signaling*, *5*(231), ra47-ra47. https://doi.org/10.1126/scisignal.2002712

Alvarenga, R., Carrara, A., Silva, C., & Oliveira, E. (2011). Potential application of Saccharomyces cerevisiae strains for the fermentation of banana pulp. *African Journal of Biotechnology*, *10*(18), 3608–3615. Retrieved from https://www.ajol.info/index.php/ajb/article/view/93438

Amorim, H. V., Lopes, M. L., de Castro Oliveira, J. V., Buckeridge, M. S., & Goldman, G. H. (2011). Scientific challenges of bioethanol production in Brazil. *Applied Microbiology and Biotechnology*, *91*(5), 1267–1275. https://doi.org/10.1007/s00253-011-3437-6

Arrieta-Montiel, M. P., Shedge, V., Davila, J., Christensen, A. C., & Mackenzie, S. A. (2009). Diversity of the Arabidopsis Mitochondrial Genome Occurs via Nuclear-Controlled Recombination Activity. *Genetics*, *183*(4), 1261–1268. https://doi.org/10.1534/genetics.109.108514

Bakkaiova, J., Marini, V., Willcox, S., Nosek, J., Griffith, J. D., Krejci, L., & Tomaska, L. (2016). Yeast mitochondrial HMG proteins: DNA-binding properties of the most evolutionarily divergent component of mitochondrial nucleoids. *Bioscience Reports*, *36*(1), e00288–e00288. https://doi.org/10.1042/BSR20150275

Bandelt, H. J., Forster, P., Sykes, B. C., & Richards, M. B. (1995). Mitochondrial portraits of human populations using median networks. *Genetics*, *141*(2), 743–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8647407

Barnett, J. A. (1992). The taxonomy of the genusSaccharomyces meyenex reess: A short review for non-taxonomists. *Yeast*, *8*(1), 1–23. https://doi.org/10.1002/yea.320080102

Barrett, J. C., Fry, B., Maller, J., & Daly, M. J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*, *21*(2), 263–265. https://doi.org/10.1093/bioinformatics/bth457

Basse, C. W. (2010). Mitochondrial inheritance in fungi. *Current Opinion in Microbiology*, *13*(6), 712–719. https://doi.org/10.1016/j.mib.2010.09.003

Bendich, A. J. (2010). The End of the Circle for Yeast Mitochondrial DNA. *Molecular Cell*, *39*(6), 831–832. https://doi.org/10.1016/j.molcel.2010.09.005

Bernardi, G. (1979). The petite mutation in yeast. *Trends in Biochemical Sciences*, *4*(9), 197–201. https://doi.org/10.1016/0968-0004(79)90079-3

Bonnefoy, N., Remacle, C., & Fox, T. D. (2007). Genetic Transformation of Saccharomyces cerevisiae and Chlamydomonas reinhardtii Mitochondria. In *Methods in cell biology* (Vol. 80, pp. 525–548). https://doi.org/10.1016/S0091-679X(06)80026-9

Boulton, C., & Quain, D. (2001). *Brewing yeast and fermentation*. Blackwell Science. Retrieved from https://www.wiley.com/en-us/Brewing+Yeast+and+Fermentation-p-9781405152686

Bravim, F., Palhano, F. L., Fernandes, A. A. R., & Fernandes, P. M. B. (2010). Biotechnological properties of distillery and laboratory yeasts in response to industrial stresses. *Journal of Industrial Microbiology & Biotechnology*, *37*(10), 1071–1079. https://doi.org/10.1007/s10295-010-0755-0

Brookes, P. S., Yoon, Y., Robotham, J. L., Anders, M. W., & Sheu, S.-S. (2004). Calcium, ATP, and ROS: a mitochondrial love-hate triangle. *American Journal of Physiology-Cell Physiology*, *287*(4), C817–C833. https://doi.org/10.1152/ajpcell.00139.2004

Butow, R. A., Perlman, P. S., & Grossman, L. I. (1985). The unusual varl gene of yeast mitochondrial DNA. *Science (New York, N.Y.)*, *228*(4707), 1496–501. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2990030

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., … Wong, E. D. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, *40*(D1), D700–D705. https://doi.org/10.1093/nar/gkr1029

Clark-Walker, G. D. (1989). In vivo rearrangement of mitochondrial DNA in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, *86*(22), 8847–51. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2682661

Crawford, D. C., Bhangale, T., Li, N., Hellenthal, G., Rieder, M. J., Nickerson, D. A., & Stephens, M. (2004). Characterizing fine-scale variation in human recombination Evidence for substantial fine-scale variation in recombination rates across the human genome. *NATURE GENETICS*, *36*. https://doi.org/10.1038/ng1376

De Deken, R. H. (1966). *The Crabtree Effect: A Regulatory System in Yeast*. *J. gen. Microbiol* (Vol. 44). Retrieved from www.microbiologyresearch.org

de Zamaroczy, M., & Bernardi, G. (1986). The primary structure of the mitochondrial genome of Saccharomyces cerevisiae--a review. *Gene*, *47*(2–3), 155–77. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3549452

Dieckmann, C. L., & Gandy, B. (1987). Preferential recombination between GC clusters in yeast mitochondrial DNA. *The EMBO Journal*, *6*(13), 4197–203. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3327690

Dujon, B. (1980). Sequence of the intron and flanking exons of the mitochondrial 21S rRNA

gene of yeast strains having different alleles at the omega and rib-1 loci. *Cell*, *20*(1), 185–97. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6156002

Dujon, B. (2006). Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in Genetics*, *22*(7), 375–387. https://doi.org/10.1016/j.tig.2006.05.007

Dujon, B., Slonimski, P. P., & Weill, L. (1974). Mitochondrial genetics. IX. A model for recombination and segregation of mitochondrial genomes in Saccharomyces cerevisiae. *Genetics*, *78*(1), 415–437. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4613610

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

Ehrenreich, I. M., Torabi, N., Jia, Y., Kent, J., Martis, S., Shapiro, J. A., … Kruglyak, L. (2010). Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature*, *464*(7291), 1039–1042. https://doi.org/10.1038/nature08923

Elson, J. L., Turnbull, D. M., & Howell, N. (2004). Comparative Genomics and the Evolution of Human Mitochondrial DNA: Assessing the Effects of Selection. *The American Journal of Human Genetics*, *74*(2), 229–238. https://doi.org/10.1086/381505

Engel, S. R., Weng, S., Binkley, G., Paskov, K., Song, G., & Cherry, J. M. (2016). From one to many: Expanding the Saccharomyces cerevisiae reference genome panel. *Database*, *2016*(2), 1–5. https://doi.org/10.1093/database/baw020

Ephrussi, B. (1949). Action de l'acriflavine sur les levures. *Unités Biologiques Douées de Continuité Génétiqué*.

Excoffier, L., Laval, G., & Schneider, S. (2007). Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, *1*, 47–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19325852

Fehrenbacher, K. L., Yang, H.-C., Gay, A. C., Huckaba, T. M., & Pon, L. A. (2004). Live Cell Imaging of Mitochondrial Movement along Actin Cables in Budding Yeast. *Current Biology*, *14*(22), 1996–2004. https://doi.org/10.1016/j.cub.2004.11.004

Fisher, K. J., Buskirk, S. W., Vignogna, R. C., Marad, D. A., & Lang, G. I. (2018). Adaptive genome duplication affects patterns of molecular evolution in Saccharomyces cerevisiae. *PLoS Genetics*, *14*(5), 1–22. https://doi.org/10.1371/journal.pgen.1007396

Fourie, G., van der Merwe, N. A., Wingfield, B. D., Bogale, M., Tudzynski, B., Wingfield, M. J., & Steenkamp, E. T. (2013). Evidence for inter-specific recombination among the mitochondrial genomes of Fusarium species in the Gibberella fujikuroi complex. *BMC Genomics*, *14*, 605. https://doi.org/10.1186/1471-2164-14-605

Foury, F., Hu, J., & Vanderstraeten, S. (2004). Mitochondrial DNA mutators. *Cellular and Molecular Life Sciences*, *61*(22), 2799–2811. https://doi.org/10.1007/s00018-004-4220-

y

Foury, F., Roganti, T., Lecrenier, N., & Purnelle, B. (1998). The complete sequence of the mitochondrial genome of Saccharomyces cerevisiae è ne. *FEBS Letters*, *440*(3), 325–331. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9872396

Fritsch, E. S., Chabbert, C. D., Klaus, B., & Steinmetz, L. M. (2014). A genome-wide map of mitochondrial DNA recombination in yeast. *Genetics*, *198*(2), 755–771. https://doi.org/10.1534/genetics.114.166637

Galtier, N. (2011). The intriguing evolutionary dynamics of plant mitochondrial DNA. *BMC Biology*, *9*(1), 61. https://doi.org/10.1186/1741-7007-9-61

Gancedo, J. M. (1998). Yeast carbon catabolite repression. *Microbiology and Molecular Biology Reviews : MMBR*, *62*(2), 334–61. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9618445

Gerstein, A. C., Fu, M. S., Mukaremera, L., Li, Z., Ormerod, K. L., Fraser, J. A., … Nielsen, K. (2015). Polyploid Titan Cells Produce Haploid and Aneuploid Progeny To Promote Stress Adaptation. *MBio*, *6*(5), e01340-15. https://doi.org/10.1128/mBio.01340-15

Goddard, M. R., & Burt, A. (1999). Recurrent invasion and extinction of a selfish gene. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(24), 13880–5. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10570167

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., … Oliver, S. G. (1996). Life with 6000 Genes. *Science*, *274*(5287), 546, 563–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8849441

Goldring, E. S., Grossman, L. I., & Marmur, J. (1971). Petite mutation in yeast. II. Isolation of mutants containing mitochondrial deoxyribonucleic acid of reduced size. *Journal of Bacteriology*, *107*(1), 377–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/5563875

Gray, M. W. (2012). Mitochondrial Evolution. *Cold Spring Harbor Perspectives in Biology*, *4*(9), a011403–a011403. https://doi.org/10.1101/cshperspect.a011403

Gray, M. W., Burger, G., & Lang, B. F. (1999). Mitochondrial Evolution, *283*(March), 1476–1482.

Groth, C., Petersen, R. F., Piškur, J., & Pis, J. (2000). Diversity in organization and the origin of gene orders in the mitochondrial DNA molecules of the genus Saccharomyces. *Molecular Biology and Evolution*, *17*(12), 1833–1841. https://doi.org/10.1093/oxfordjournals.molbev.a026284

Hall, T. A. (1999). BioEdit: a user-frienly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Oxford University Press*. Retrieved from http://brownlab.mbio.ncsu.edu/JWB/papers/1999Hall1.pdf

Hausner, G. (2003). Fungal mitochondrial genomes, plasmids and introns. *Applied Mycology*

and Biotechnology, *3*(C), 101–131. https://doi.org/10.1016/S1874-5334(03)80009-6

Herskowitz, I. (1988). Life cycle of the budding yeast Saccharomyces cerevisiae. *Microbiological Reviews*, *52*(4), 536–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3070323

Heslot, H., Goffeau, A., & Louis, C. (1970). Respiratory metabolism of a "petite negative"yeast Schizosaccharomyces pombe 972h-. *Journal of Bacteriology*, *104*(1), 473–81. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4394400

Hittinger, C. T. (2013). 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

Ho, S. Y. W., Lanfear, R., Bromham, L., Phillips, M. J., Soubrier, J., Rodrigo, A. G., & Cooper, A. (2011). Time-dependent rates of molecular evolution. *Molecular Ecology*, *20*(15), 3087–3101. https://doi.org/10.1111/j.1365-294X.2011.05178.x

Ho, S. Y. W., Shapiro, B., Phillips, M. J., Cooper, A., & Drummond, A. J. (2007). Evidence for Time Dependency of Molecular Rate Estimates. *Systematic Biology*, *56*(3), 515–522. https://doi.org/10.1080/10635150701435401

Hollingsworth, M. J., & Martin, N. C. (1986). RNase P activity in the mitochondria of Saccharomyces cerevisiae depends on both mitochondrion and nucleus-encoded components. *Molecular and Cellular Biology*, *6*(4), 1058–64. https://doi.org/10.1128/MCB.6.4.1058

Hori, A., Yoshida, M., Shibata, T., & Ling, F. (2009). Reactive oxygen species regulate DNA copy number in isolated yeast mitochondria by triggering recombination-mediated replication. *Nucleic Acids Research*, *37*(3), 749–761. https://doi.org/10.1093/nar/gkn993

Hsu, Y. Y., & Chou, J. Y. (2017). Environmental Factors Can Influence Mitochondrial Inheritance in the Saccharomyces Yeast Hybrids. *PLoS ONE*, *12*(1), 1–16. https://doi.org/10.1371/journal.pone.0169953

Hudspeth, M. E., Vincent, R. D., Perlman, P. S., Shumard, D. S., Treisman, L. O., & Grossman, L. I. (1984). Expandable var1 gene of yeast mitochondrial DNA: in-frame insertions can explain the strain-specific protein size polymorphisms. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(10), 3148–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6328501

Jakobs, S., Martini, N., Schauss, A. C., Egner, A., Westermann, B., & Hell, S. W. (2003). Spatial and temporal dynamics of budding yeast mitochondria lacking the division component Fis1p. *Journal of Cell Science*, *116*(10), 2005–2014. https://doi.org/10.1242/jcs.00423

Jobin, M., Schurz, H., & Henn, B. M. (2018). IMPUTOR: Phylogenetically Aware Software for Imputation of Errors in Next-Generation Sequencing. *Genome Biology and Evolution*, *10*(5), 1248–1254. https://doi.org/10.1093/gbe/evy088

Johnston, M. (2000). The yeast genome : on the road to the Golden Age, 617–623. Retrieved from https://www.ncbi.nlm.nih.gov/labs/articles/11088011/

Jung, P. P., Friedrich, A., Reisser, C., Hou, J., & Schacherer, J. (2012). Mitochondrial Genome Evolution in a Single Protoploid Yeast Species. *G3: Genes|Genomes|Genetics*, *2*(9), 1103–1111. https://doi.org/10.1534/g3.112.003152

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

Kroemer, G., Galluzzi, L., & Brenner, C. (2007). Mitochondrial Membrane Permeabilization in Cell Death. *Physiological Reviews*, *87*(1), 99–163. https://doi.org/10.1152/physrev.00013.2006

Kumar, S., Stecher, G., Tamura, K., & Dudley, J. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets Downloaded from. *Mol. Biol. Evol*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Ladoukakis, E. D., & Zouros, E. (2001). Direct Evidence for Homologous Recombination in Mussel (Mytilus galloprovincialis) Mitochondrial DNA. *Molecular Biology and Evolution*, *18*(7), 1168–1175. https://doi.org/10.1093/oxfordjournals.molbev.a003904

Lambowitz, A. M., & Zimmerly, S. (2004). Mobile Group II Introns. *Annual Review of Genetics*, *38*(1), 1–35. https://doi.org/10.1146/annurev.genet.38.072902.091600

Lang, B. F., Laforest, M.-J., & Burger, G. (2007). Mitochondrial introns: a critical view. *Trends in Genetics*, *23*(3), 119–125. https://doi.org/10.1016/j.tig.2007.01.006

Lecrenier, N., & Foury, F. (2000). New features of mitochondrial DNA replication system in yeast and man. *Gene*, *246*(1–2), 37–48. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10767525

Libkind, D., Hittinger, C. T., Valerio, E., Goncalves, C., Dover, J., Johnston, M., … 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*, *108*(35), 14539–14544. https://doi.org/10.1073/pnas.1105430108

Librado, P., & Rozas, J. (2005). *DnaSP v5: A software for comprehensive analysis of DNA poly-morphism data*. Retrieved from http://www.ub.edu/dnasp

Ling, F., Hori, A., & Shibata, T. (2007). DNA Recombination-Initiation Plays a Role in the Extremely Biased Inheritance of Yeast [rho-] Mitochondrial DNA That Contains the Replication Origin ori5. *Molecular and Cellular Biology*, *27*(3), 1133–1145. https://doi.org/10.1128/MCB.00770-06

Ling, F., & Shibata, T. (2002). Recombination-dependent mtDNA partitioning: in vivo role of Mhr1p to promote pairing of homologous DNA. *The EMBO Journal*, *21*(17), 4730–40. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12198175

Lipinski, K. A., Kaniak-golik, A., & Golik, P. (2010). Maintenance and expression of the S. Cerevisiae mitochondrial genome-From genetics to evolution and systems biology. *BBA - Bioenergetics*, *1797*(6–7), 1086–1098. https://doi.org/10.1016/j.bbabio.2009.12.019

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., … Louis, E. J. (2009). Population genomics of domestic and wild yeasts. *Nature*, *458*(7236), 337–341. https://doi.org/10.1038/nature07743

Llopis, S., Hernández-Haro, C., Monteoliva, L., Querol, A., Molina, M., & Fernández-Espinar, M. T. (2014). Pathogenic Potential of Saccharomyces Strains Isolated from Dietary Supplements. *PLoS ONE*, *9*(5), e98094. https://doi.org/10.1371/journal.pone.0098094

Llopis, S., Querol, A., Heyken, A., Hube, B., Jespersen, L., Fernández-Espinar, M., & Pérez-Torrado, R. (2012). Transcriptomics in human blood incubation reveals the importance of oxidative stress response in Saccharomyces cerevisiae clinical strains. *BMC Genomics*, *13*(1), 419. https://doi.org/10.1186/1471-2164-13-419

Magwene, P. M., Kayıkçı, 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*, *108*(5), 1987–1992. https://doi.org/10.1073/pnas.1012544108

Maleszka, R., Skelly, P. J., & Clark-Walker, G. D. (1991). Rolling circle replication of DNA in yeast mitochondria. *The EMBO Journal*, *10*(12), 3923–3929. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1935911

Masel, J., & Lyttle, D. N. (2011). The consequences of rare sexual reproduction by means of selfing in an otherwise clonally reproducing species. *Theoretical Population Biology*, *80*(4), 317–22. https://doi.org/10.1016/j.tpb.2011.08.004

Merz, S., Hammermeister, M., Altmann, K., Dürr, M., & Westermann, B. (2007). Molecular machinery of mitochondrial dynamics in yeast. *Biological Chemistry*, *388*(9), 917–926. https://doi.org/10.1515/BC.2007.110

Mitra, K., Wunder, C., Roysam, B., Lin, G., & Lippincott-Schwartz, J. (2009). A hyperfused mitochondrial state achieved at G1-S regulates cyclin E buildup and entry into S phase. *Proceedings of the National Academy of Sciences*, *106*(29), 11960–11965. https://doi.org/10.1073/pnas.0904875106

Miyakawa, I., Sando, N., Kawano, S., Nakamura, S., & Kuroiwa, T. (1987). Isolation of morphologically intact mitochondrial nucleoids from the yeast, Saccharomyces cerevisiae. *Journal of Cell Science*, *88 (Pt 4)*, 431–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3332668

Mortimer, R. K., Romano, P., Suzzi, G., & Polsinelli, M. (1994). Genome renewal: A new phenomenon revealed from a genetic study of 43 strains ofSaccharomyces cerevisiae derived from natural fermentation of grape musts. *Yeast*, *10*(12), 1543–1552. https://doi.org/10.1002/yea.320101203

Müller, M., Lu, K., & Reichert, A. S. (2015). Mitophagy and mitochondrial dynamics in Saccharomyces cerevisiae. *Biochimica et Biophysica Acta - Molecular Cell Research*, *1853*(10), 2766–2774. https://doi.org/10.1016/j.bbamcr.2015.02.024

Naumov, G. I., Naumova, E. S., & Sniegowski, P. D. (1998). Saccharomyces paradoxus and Saccharomyces cerevisiae are associated with exudates of North American oaks. *Canadian Journal of Microbiology*, *44*(11), 1045–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10029999

Naumov, G., & Naumov, G. I. (1987, January 1). Genetic basis for classification and identification of the ascomycetous yeasts. Retrieved from https://www.scienceopen.com/document?vid=bc4950df-35a7-4cfb-b45a-574331505987

Nei, M. (1971). Extinction time of deleterious mutant genes in large populations. *Theoretical Population Biology*, *2*(4), 419–425. https://doi.org/10.1016/0040-5809(71)90030-X

Pavesi, G., Mauri, G., Iannelli, F., Gissi, C., & Pesole, G. (2004). GeneSyn: a tool for detecting conserved gene order across genomes. *Bioinformatics*, *20*(9), 1472–1474. https://doi.org/10.1093/bioinformatics/bth102

Perez-Martinez, X., Broadley, S. A., & Fox, T. D. (2003). Mss51p promotes mitochondrial Cox1p synthesis and interacts with newly synthesized Cox1p. *The EMBO Journal*, *22*(21), 5951–5961. https://doi.org/10.1093/emboj/cdg566

Pérez-Torrado, R., Llopis, S., Perrone, B., Gómez-Pastor, R., Hube, B., & Querol, A. (2015). Comparative Genomic Analysis Reveals a Critical Role of De Novo Nucleotide Biosynthesis for Saccharomyces cerevisiae Virulence. *PLOS ONE*, *10*(3), e0122382. https://doi.org/10.1371/journal.pone.0122382

Pérez-Torrado, R., & Querol, A. (2016). Opportunistic strains of Saccharomyces cerevisiae: A potential risk sold in food products. *Frontiers in Microbiology*, *6*(JAN), 1–5. https://doi.org/10.3389/fmicb.2015.01522

Peter, J., Chiara, M. De, Friedrich, A., Yue, J., Pflieger, D., Bergström, A., … Liti, G. (2018). Saccharomyces cerevisiae isolates Saccharomyces cerevisiae isolates. *Nature*, *556*, 339–344. https://doi.org/10.1038/s41586-018-0030-5

Peter, J., & Schacherer, J. (2015). Population genomics of yeasts : towards a comprehensive view across a broad evolutionary scale, (January), 73–81. https://doi.org/10.1002/yea

Petersen, R. F., Langkjær, R. B., Hvidtfeldt, J., Gartner, J., Palmen, W., Ussery, D. W., … Piškur, J. (2002). Inheritance and organisation of the mitochondrial genome differ

between two Saccharomyces yeasts. *Journal of Molecular Biology*, *2836*(02), 627–636. https://doi.org/10.1016/S0022-2836(02)00037-2

Phillips, M. J. (2009). Branch-length estimation bias misleads molecular dating for a vertebrate mitochondrial phylogeny. *Gene*, *441*(1), 132–140. https://doi.org/10.1016/j.gene.2008.08.017

Piskur, J. (1988). Transmission of yeast mitochondrial loci to progeny is reduced when nearby intergenic regions containing ori sequences are deleted. *Molecular & General Genetics : MGG*, *214*(3), 425–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3063946

Piskur, J. (1994). Inheritance of the Yeast Mitochondrial Genome.

Piskur, J., Rozpedowska, E., Polakova, S., Merico, A., & Compagno, C. (2006). How did Saccharomyces evolve to become a good brewer? *Trends in Genetics*, *22*(4), 183–186. https://doi.org/10.1016/j.tig.2006.02.002

Pretorius, I. S. (2000). Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast*, *16*(8), 675–729. https://doi.org/10.1002/1097-0061(20000615)16:8<675::AID-YEA585>3.0.CO;2-B

Preuten, T., Cincu, E., Fuchs, J., Zoschke, R., Liere, K., & Börner, T. (2010). Fewer genes than organelles: extremely low and variable gene copy numbers in mitochondria of somatic plant cells. *The Plant Journal*, *64*(6), 948–959. https://doi.org/10.1111/j.1365-313X.2010.04389.x

Pruyne, D., Legesse-Miller, A., Gao, L., Dong, Y., & Bretscher, A. (2004). MECHANISMS OF POLARIZED GROWTH AND ORGANELLE SEGREGATION IN YEAST. *Annual Review of Cell and Developmental Biology*, *20*(1), 559–591. https://doi.org/10.1146/annurev.cellbio.20.010403.103108

Ruiz-Pesini, E., & Wallace, D. C. (2006). Evidence for adaptive selection acting on the tRNA and rRNA genes of human mitochondrial DNA. *Human Mutation*, *27*(11), 1072–1081. https://doi.org/10.1002/humu.20378

Saraste, M. (1999). Oxidative phosphorylation at the fin de siècle. *Science (New York, N.Y.)*, *283*(5407), 1488–93. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10066163

Schacherer, J., Shapiro, J. A., Ruderfer, D. M., & Kruglyak, L. (2009). Comprehensive polymorphism survey elucidates population structure of Saccharomyces cerevisiae. *Nature*, *458*(7236), 342–345. https://doi.org/10.1038/nature07670

Shay, J. W., Pierce, D. J., & Werbin, H. (1990). Mitochondrial DNA copy number is proportional to total cell DNA under a variety of growth conditions. *The Journal of Biological Chemistry*, *265*(25), 14802–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2394698

Silva, R. O. da, Batistote, M., & Cereda, M. P. (2013). Alcoholic fermentation by the wild

yeasts under thermal, osmotic and ethanol stress. *Brazilian Archives of Biology and Technology*, *56*(2), 161–169. https://doi.org/10.1590/S1516-89132013000200001

Simon, V. R., Karmon, S. L., & Pon, L. A. (1997). Mitochondrial inheritance: Cell cycle and actin cable dependence of polarized mitochondrial movements inSaccharomyces cerevisiae. *Cell Motility and the Cytoskeleton*, *37*(3), 199–210. https://doi.org/10.1002/(SICI)1097-0169(1997)37:3<199::AID-CM2>3.0.CO;2-2

Soares, P., Abrantes, D., Rito, T., Thomson, N., Radivojac, P., Li, B., … Pereira, L. (2013). Evaluating Purifying Selection in the Mitochondrial DNA of Various Mammalian Species. *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0058993

Soares, P., Ermini, L., Thomson, N., Mormina, M., Rito, T., Röhl, A., … Richards, M. B. (2009). Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *American Journal of Human Genetics*, *84*(6), 740–759. https://doi.org/10.1016/j.ajhg.2009.05.001

Solieri, L. (2010). Mitochondrial inheritance in budding yeasts: Towards an integrated understanding. *Trends in Microbiology*, *18*(11), 521–530. https://doi.org/10.1016/j.tim.2010.08.001

Spírek, M., Soltésová, A., Horváth, A., Sláviková, E., & Sulo, P. (2002). GC clusters and the stability of mitochondrial genomes of Saccharomyces cerevisiae and related yeats. *Folia Microbiologica*, *47*(3), 263–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12094735

Storici, F., Durham, C. L., Gordenin, D. A., & Resnick, M. A. (2003). Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proceedings of the National Academy of Sciences*, *100*(25), 14994–14999. https://doi.org/10.1073/pnas.2036296100

Strausberg, R. L., & Perlman, P. S. (1978). The effect of zygotic bud position on the transmission of mitochondrial genes in Saccharomyces cerevisiae. *Molecular & General Genetics : MGG*, *163*(2), 131–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/355844

Strope, P. K., Skelly, D. A., Kozmin, S. G., Mahadevan, G., Stone, E. A., Magwene, P. M., … 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*, *25*(5), 762–74. https://doi.org/10.1101/gr.185538.114

Taylor, J. W., & Berbee, M. L. (2006). Dating divergences in the Fungal Tree of Life: review and new analyses. *Mycologia*, *98*(6), 838–49. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17486961

Thrailkill, K. M., & Birky, C. W. (1980). Intracellular population genetics: evidence for random

drift of mitochondrial allele frequencies in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *Genetics*, *96*(1), 237–62. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7009322

Torroni, A., Achilli, A., Macaulay, V., Richards, M., & Rgen Bandelt, H.-J. (2006). Harvesting the fruit of the human mtDNA tree. https://doi.org/10.1016/j.tig.2006.04.001

Tsai, I. J., Bensasson, D., Burt, A., & Koufopanou, V. (2008). *Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle POPULATION BIOLOGY*. *PNAS March* (Vol. 25). Retrieved from http://www.pnas.org/content/pnas/105/12/4957.full.pdf

Turk, E. M., Das, V., Seibert, R. D., & Andrulis, E. D. (2013). The Mitochondrial RNA Landscape of Saccharomyces cerevisiae. *PLoS ONE*, *8*(10), 1–21. https://doi.org/10.1371/journal.pone.0078105

Úbeda, J. F., Chacoń-Ocaña, M., Díaz-Hellín, P., Ramírez-Pérez, H., Briones, A., Patricia, D., … Briones, A. (2016). Genetic and phenotypic characterization of Saccharomyces spp. Strains isolated in distillery plants. *FEMS Yeast Research*, *16*(December 2015), 1–6. https://doi.org/10.1093/femsyr/fow035

Valach, M., Farkas, Z., Fricova, D., Kovac, J., Brejova, B., Vinar, T., … Nosek, J. (2011). Evolution of linear chromosomes and multipartite genomes in yeast mitochondria. *Nucleic Acids Research*, *39*(10), 4202–4219. https://doi.org/10.1093/nar/gkq1345

Vaughan-Martini, A., & Martini, A. (2011). Saccharomyces Meyen ex Reess (1870). *The Yeasts*, 733–746. https://doi.org/10.1016/B978-0-444-52149-1.00061-6

Wallace, D. C. (2007). Why Do We Still Have a Maternally Inherited Mitochondrial DNA? Insights from Evolutionary Medicine. *Annual Review of Biochemistry*, *76*(1), 781–821. https://doi.org/10.1146/annurev.biochem.76.081205.150955

Wang, Q.-M. M., Liu, W.-Q. Q., Liti, G., Wang, S.-A. A., & Bai, F.-Y. Y. (2012). Surprisingly diverged populations of Saccharomyces cerevisiae in natural environments remote from human activity. *Molecular Ecology*, *21*(22), 5404–5417. https://doi.org/10.1111/j.1365-294X.2012.05732.x

Watanabe, T., Srichuwong, S., Arakane, M., Tamiya, S., Yoshinaga, M., Watanabe, I., … Nakamura, T. (2010). Selection of stress-tolerant yeasts for simultaneous saccharification and fermentation (SSF) of very high gravity (VHG) potato mash to ethanol. *Bioresource Technology*, *101*(24), 9710–9714. https://doi.org/10.1016/j.biortech.2010.07.079

Weiller, G., Schueller, C. M. E., & Schweyen, R. J. (1989). Putative target sites for mobile G+C rich clusters in yeast mitochondrial DNA: Single elements and tandem arrays. *MGG Molecular & General Genetics*, *218*(2), 272–283. https://doi.org/10.1007/BF00331278

Westermann, B. (2010). Mitochondrial dynamics in model organisms: What yeasts, worms and flies have taught us about fusion and fission of mitochondria. *Seminars in Cell & Developmental Biology*, *21*(6), 542–549. https://doi.org/10.1016/j.semcdb.2009.12.003

Westermann, B. (2014). Mitochondrial inheritance in yeast. *BBA - Bioenergetics*, *1837*(7), 1039–1046. https://doi.org/10.1016/j.bbabio.2013.10.005

Wolters, J. F., Chiu, K., & Fiumera, H. L. (2015). Population structure of mitochondrial genomes in Saccharomyces cerevisiae. *BMC Genomics*, *16*(1), 1–13. https://doi.org/10.1186/s12864-015-1664-4

Xu, J., Ali, R. Y., Gregory, D. A., Amick, D., Lambert, S. E., Yoell, H. J., … Mitchell, T. G. (2000). Uniparental Mitochondrial Transmission in Sexual Crosses in Cryptococcus neoformans. *Current Microbiology*, *40*(4), 269–273. https://doi.org/10.1007/s002849910053

Yan, Z., & Xu, J. (2003). Mitochondria are inherited from the MATa parent in crosses of the basidiomycete fungus Cryptococcus neoformans. *Genetics*, *163*(4), 1315–25. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12702677

Yang, H. C., Palazzo, A., Swayne, T. C., & Pon, L. A. (1999). A retention mechanism for distribution of mitochondria during cell division in budding yeast. *Current Biology : CB*, *9*(19), 1111–4. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10531006

Zamaroczy, M., & Bernardi, G. (1986). The primary structure of the mitochondrial genome of Saccharomyces cerevisiae - a review.

Zhang, Y. (1999). Smart PCA, 1351–1356.

Zhu, Y. O., Sherlock, G., & Petrov, D. A. (2016). Whole Genome Analysis of 132 Clinical Saccharomyces cerevisiae Strains Reveals Extensive Ploidy Variation, *6*(August), 2421–2434. https://doi.org/10.1534/g3.116.029397

Zinn, A. R., Pohlman, J. K., Perlman, P. S., & Butow, R. A. (1987). Kinetic and segregational analysis of mitochondrial DNA recombination in yeast. *Plasmid*, *17*(3), 248–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3306735

Zuin, A., Gabrielli, N., Calvo, I. A., García-Santamarina, S., Hoe, K.-L., Kim, D. U., … Hidalgo, E. (2008). Mitochondrial Dysfunction Increases Oxidative Stress and Decreases Chronological Life Span in Fission Yeast. *PLoS ONE*, *3*(7), e2842. https://doi.org/10.1371/journal.pone.0002842