Genomic and Phenotypic Studies of Traits Affecting Cold Tolerance in Natural Yeast Species and Hybrids

A thesis submitted to

The University of Manchester for the degree of

Doctor of Philosophy in the Faculty of Biology, Medicine and Health

2018

Kobchai Duangrattanalert

School of Biological Sciences

- 20,526 words -

Table of Contents

Abstract

Over the past decades, scientific communities have put a significant amount of interest in genetic variation within and between *Saccharomyces* species and how it can be used for industries. In this thesis, the effect of inter and intraspecific variation on the phenotypes of natural yeast species and hybrids is explored.

To study intraspecific variation, fitness profiles of *S. cerevisiae* strain Σ1278b were constructed and compared with BY4743 from previous studies under nutrient starvation and temperature stress. It was found that genes involved in genome integrity maintenance and general metabolism are important for the growth of Σ1278b in complete and nutrient limited media, respectively. Moreover, the results also show that there is little fitness variation between the two yeast strains. Given that, less conserved genes are likely to produce background-specific phenotypes.

To study interspecific variation, experiments were carried out on independent lineages of a cross between *S. cerevisiae* (Sc) and *S. uvarum* (Su). These hybrids were constructed to resemble the natural hybrid *S. pastorianus* at the genome level. To examine the effect of chimeric protein complexes, hybrids that carry mitochondrial DNA (mtDNA) from *S. cerevisiae* were selected. It was found that combining *GCN1* from *S. cerevisiae* and *GCN20* from *S. uvarum* appeared to make the hybrids more robust under amino acid starvation and cold stress condition. In addition, it was found that variation in phenotypic traits could be caused by structural differences of the protein GCN1 and differential transcriptional regulation of the gene *GCN20*. Interestingly, this may explain biased retention of this protein complex in the natural hybrid *S. pastorianus*.

To study another aspect of interspecific variation, spot assays and RNA sequencing were carried out to determine the effect of mitotypes on the expression of nuclear genes in yeast hybrids. The results show that fitness is affected heavily by mitotypes as hybrids with mtDNA from *S. cerevisiae* is more favourable at warm and vice versa for mtDNA from *S. uvarum*. In addition it was found that mitotypes appear to have an influence on allelic expression bias in several biological pathways. This is true

especially in respiration as the expression of orthologs of this process from Sc and Su subgenomes was suppressed in complete medium at warm for Sc mtDNA and at cold for Su mtDNA, respectively.

And finally, as a side project, an open-source application, called gcatR, is developed to analyse growth curve data. This application is created in R language and contains several features including interactive data table, growth curve parameter estimation, time series clustering and classification, and fitting based on generalised additive models. What's more, it creates a set of diagrams (such as line graph, bar chart and heatmap) to represent the obtained results. More importantly, this application is designed to be user friendly and does not require users to have prior knowledge in any programming language. gcatR is provided free of charge and can be accessed at http://www.gcatr.manchester.ac.uk/ or downloaded to install on a local Windows device.

Declaration

No portion of the work referred to in the thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright statement

Acknowledgements

First and foremost, I would like to thank Professor Daniela Delneri for her guidance and excellent supervision throughout my PhD. More importantly, her patience and enthusiasm have immensely helped me with my study and research. I would also like to thank Dr. Jean-Marc Schwartz for his superb advice. I must also thank all members in Delneri's lab for their help with lab works and for making the last four years the best time of my life.

I would like to offer special thanks to my family - Mr Bavornbhoom Duangrattanalert, Mrs. Thitirat Soisamran and Miss Kanyarat Duangrattanalert. I would also like to express my appreciation to Mr. Yotsaporn Kaolim and other friends from Phrapathom Wittayalai school for their understanding and occasional support. Notably, my sincere gratitude goes to Evaldas Klumbys for being there when I needed and definitely for all the fun we have had for the past eight years.

My thanks are extended to Miss Somrudee Deepaisarn who helped me with printing, binding and submitting the revised version of my thesis.

Finally, I would like to acknowledge DPST (Development and Promotion of Science and Technology Talents Project) and the Royal Government of Thailand for funding my project.

Introduction

1.1 Yeast as a model organism

Saccharomyces cerevisiae has been domesticated and lived alongside humans for millennia [1]. It has a prominent historical association with various industries including baking, brewing, and winemaking. It has also been used extensively for modern scientific research and remarkably led to a better understanding of human biology and diseases. *S. cerevisiae* was the first eukaryote to be sequenced fully in 1996 [2]. Since then, a number of projects have been initiated to annotate and determine the functions of genes within the genome of this yeast. *Saccharomyces* Genome Deletion Project was led by an international team of researchers to produce an array of strains, each with a single gene deleted [3]. Other projects, however, attempted to unravel interactions between genes/proteins that occur in response both internal and external environments [4,5]

S. cerevisiae is an ideal model organism because of several aspects as follows. It is a small single-celled organism (diameter between 5 and 10 μm) with short generation time (doubling time around 1-2 hours at 30°C) [6]. Compared to other eukaryotes, it is simple to introduce mutations via homologous recombination allowing scientists to systematically alter genetic composition/structure. Despite a complex internal structures similar to plants and animals, *S. cerevisiae* has a much smaller proportion of non-coding DNA [7]. This effectively makes genetic and phenotypic studies less confounding. Research on this fungal species has a number of real world applications such as food production and drug development. Fundamentally, this makes *S. cerevisiae* an important economic driver.

1.2 Yeast cell cycle

Saccharomyces sensu stricto yeast has a generally interesting life cycle which allows cells to switch between two different states (see Figure 1) - haploids that reproduce via mitosis and diploids that reproduce via both mitosis and meiosis through a process known as sporulation. The haploids exist in two mating types,

MAT a and MAT α. Fusion of two haploids of the opposite mating type produces a diploid with phenotype MAT a and MAT α. At this stage, the diploid can proliferate via mitotic cell division or undergoes meiosis to produce tetrads or ascospores that contains four gametes (two MAT a and two MAT α). The meiotic cell division or sporulation occurs under harsh environmental conditions such as amino acid starvation. Homologous recombination also occurs at this stage. It shuffles yeast genetic materials, such that the daughter cells are slightly different from each other. The mitotic division of *S. cerevisiae* is called budding. It refers to asymmetric cell division that one of the daughter cells is smaller. Regardless of the difference in cell size, they inherit the same set of DNA from the mother cell. Like other species of fungi, *S. cerevisiae* can switch its mating type from MAT a and MAT α, and vice versa. This is known as homothallism. The mechanism of mating type switching.

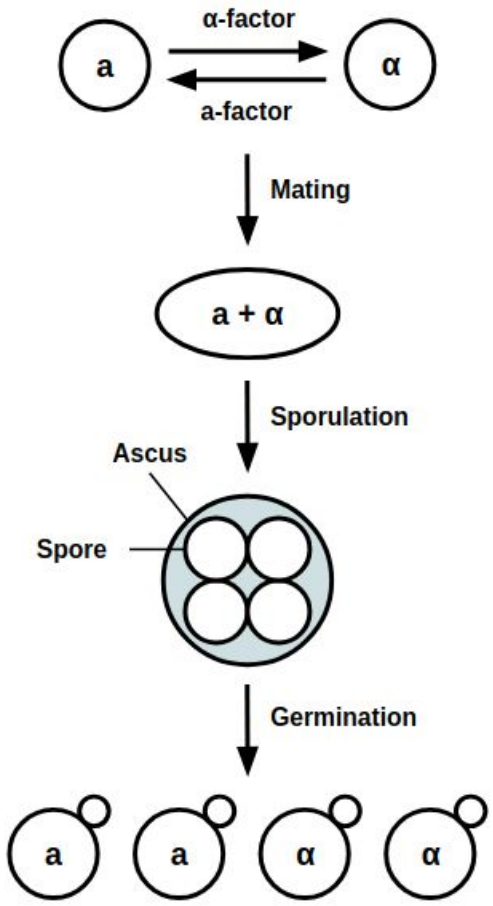


Figure 1. Schematic representation of yeast life cycle. Yeast can exist in both haploid and diploid states. As a haploid, it can mate with another haploid with opposite mating type and produce spores via meiosis. This diagram is an adaptation from Herskowitz I (1988) [8].

Mating type is determined by genetic profiles of the MAT locus and the two cryptic copies on chromosome III, called *HML a* and *HMR α* [9]. Note that these cryptic

genes are not transcribed due to interactions with specific silencing proteins. *MAT a* and *MAT α* are the determinants of mating type in *S. cerevisiae* and are located in the MAT locus. *MAT α* encodes for MAT α1 and MAT α2. MAT α1 is a co-activator that regulates alpha-specific genes, while MAT α2 forms a complex with MCM1 to repress a-specific genes in haploids. *MAT a* encodes for MAT a1 and MAT a2. MAT a1 functions with MAT α2 to repress haploid-specific genes in diploids, while the function of MAT a2 is currently unknown. In homothallic haploids, HO endonuclease is present and actively breaks the MAT locus. The locus is removed and repaired via recombination using *HML a* and *HMR α* as the donor for *MAT a* and *MAT α*, respectively. This results in the mating type switching and allows mating between yeasts of the same clonal population. 90% of the time this repair is carried out using the HML loci with the opposite mating type of the one present in the MAT locus hence inducing mating type switching.

1.3 Yeast natural history

It is believed that *S. cerevisiae* was domesticated as early as 3,150 BC [10]. Since then, it has been kept in domestication and used extensively in food industry and in research. Over the past centuries, the species has diversified into numerous strains, many of which are in association with human activities. Many strains are stored at the National Collection of Yeast Cultures (NCYC). NCYC has an online database (www.ncyc.co.uk) where the list of Saccharomyces and non-Saccharomyces yeast strains can be assessed. It is thought that *S. cerevisiae* is strictly domesticated, such that wild-type strains are simply the results of migration back to the wild. This idea was challenged in 2005 [11]. The authors presented a set of evidences. They argued that genetic variation of the wild-type strains is high, and the two lineages of brewing yeasts seem to be separated for thousands of years. According to these evidences, they gave the conclusion opposite to the above idea.

S. cerevisiae especially has diversified and evolved a number of distinct strains - each found in specific environments [12–14]. For example, S288C and BY strains are found strictly in laboratory. They are not flocculent and require only a minimal set of nutritional requirements. Σ1278b, however, has a diverged phenotype [15,16]. It is

able to exist in both unicellular and multicellular states. In fact, it is capable of growing filamentous structures and has relatively complex developmental process that facilitates fungal pathogenesis. This transition is formally known as dimorphic switch. This feature makes this strain a suitable model for studying fungal species with clinical importance. Other interesting strain is BC187. This strain is a derivative of the California wine barrel isolate and used rigorously in wine fermentation.

S. cerevisiae actually belong to a group of closely related yeast species called *Saccharomyces sensu stricto*. This group is composed of *S. cerevisiae*, *S. paradoxus*, *S. cariocanus*, *S. uvarum*, *S. mikatae*, *S. kudriavzevii*, *S. arboricola*, *S. eubayanus* and *S. jurei* [17–21]. The group also contains two natural hybrids - *S. pastorianus* [22] and *S. bayanus* [23] These yeasts have been documented in the wild. They are found either in isolation or in association with other yeast species. For example, *S. cerevisiae* was reported to co-exist with *S. paradoxus* in the woodland of North America [24]. It is controversial whether this *S. cerevisiae* is domesticated or wild. More *Saccharomyces* species have been isolated from natural sources. Some isolates are cold-adaptive and require low temperature for screening. For example, *S. uvarum* and *S. kudriavzevii* were obtained from bark samples. Further investigations reveal that they grow optimally at 26°C and 23°C, respectively [25]. In 2011, *S. eubayanus* which is a member of the *sensu stricto* group was discovered [26]. It is believed to be cold-adapted as it is found in the Northwestern region of Patagonia (a place where temperature is harshly cold throughout the year). And in 2017, *S. jurei* was found in Saint Auban, France, and proven to be a new species as determined by standard ITS and D1/D2 sequencing [21].

Reportedly, many co-existing yeast species appear to undergo hybridisation and form natural hybrids between species in the *sensu stricto* group. For example*, S. cerevisiae* is found to hybridise with *S. kudriavzevii* [27,28]. It is reported that the hybrid inherits thermal-adaptation traits from both parents as it is able to withstand a wider range of temperature. This allows the hybrid to competitively populate both warm and cold habitats and sometimes out-compete its parents. Another hybrid is *S. pastorianus*. It emerged around 15th–16th century from hybridisation events between *S. cerevisiae* and *S. uvarum*-like species which is now known to be *S.*

eubayanus [26,29,30]. More importantly, it is known for producing lager beer. The hybrid is subdivided into Saaz and Frohberg types (or group 1 and group 2, respectively) [31–33]. While Saaz-type strains (formerly known as *S. carlsbergensis*) are triploids carrying 2n *S. eubayanus* and 1n S. cerevisiae, Frohberg-type strains are tetraploids carrying equal portions (2n) of *S. cerevisiae* and *S. eubayanus*. In addition to variation in genetic compositions, they also exhibit phenotypes as group 1 *S. pastorianus* is more flocculent and less efficient at utilising sugars during fermentation [34–37].

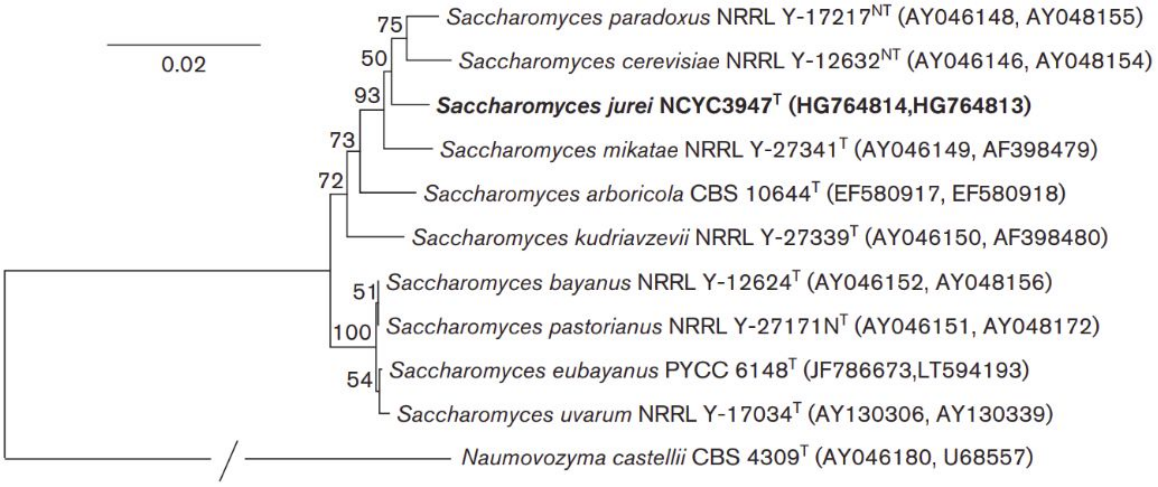


Figure 2. Neighbour-joining dendrogram of the *Saccharomyces sensu stricto* group. The diagram was constructed based on the combined sequences of the LSU D1/D2 and ITS regions (including 5.8S rRNA) of Saccharomyces jurei sp. nov. and its closest relatives. This diagram is taken from Naseeb S et al. (2017) [21].

1.4 Yeast genomes

Like to animals, *Saccharomyces* have two sets of genomes, nuclear genome and mitochondrial genome. Nuclear genome consists of 16 chromosomes ranging from 230 kb (chromosome I) to 1,531 kb (chromosome IV) in size. About 6,000 genes have been annotated for open reading frames (ORFs) [38]. While 70% of these ORFs accounts for protein-coding genes, the rest accounts for noncoding elements such as transfer RNA, ribosomal RNA, regulatory RNA, scaffold attachment regions,

telomeres and centromeres [39]. When compared to other eukaryotes, *S. cerevisiae* has a relatively compact and clean nuclear genome with a small number of introns [40]. This makes *S. cerevisiae* a suitable model organism for genomic and molecular studies. In addition, it gains more popularity in pharmaceutical research because the yeast shares multiple homologous genes with humans [41,42]. Mitochondria exist in multiple copies within a single yeast cell [43]. They are enclosed with double phospholipid bilayer and contain their own circular DNA with genes encoding machineries for aerobic respiration. Mitochondrial genome is smaller than the nuclear genome (about 85 kb) and contains seven to eight ORFs [44]. This significant reduction of the mitochondrial genome makes this organelle structurally and functionally dependent on proteins coded by genes in the nucleus.

1.5 Yeast mitochondria

The mitochondrion is an essential component of eukaryotic cells that evolved around two billion years ago from α-proteobacteria [45,46]. Since that time, this organelle has drastically changed and evolved as most genetic materials were lost or transferred rapidly to nucleus. What remains in the genome is a small number of protein encoding genes. For example, there are 13 genes in humans and 8 genes in budding yeasts [47]. In general, these genes are translated into the core subunits of respiratory chain complexes that are indispensable for growth under aerobic conditions. For instance, it has been shown that mutations of some mitochondrial genes are linked to age-related disorders such as Progeria, dementia and Alzheimer's disease [48–50]. Mitochondria are highly prone to accumulate deleterious mutations [51]. Without correction, these mutations will sequentially result in the formation of reactive oxygen species that aggressively damage cellular components and lead to more defects within both nuclear and mitochondrial genomes. This organelle is semi-autonomous and strictly dependent on hundreds of nuclear-encoded genes [52–55]. In yeast species, as many as 700 nuclear genes encoding for mitochondrial proteins are reported on *Saccharomyces* Genome Database. This suggests communication i.e. (anterograde and retrograde signaling

pathways) between the two compartments that relay biological information in response to stressful environments (see Figure 3).

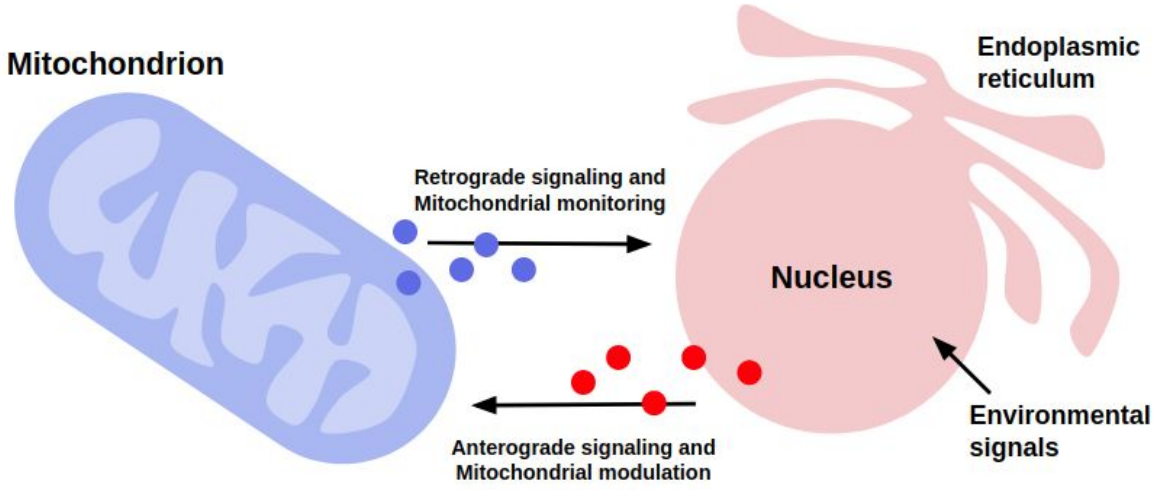


Figure 3. Schematic representation of communication between mitochondrion and nucleus. This interaction is composed of anterograde and retrograde signaling systems that help nucleus adjust and monitor mitochondrial functions in response to internal/external environments. This diagram is an adaptation from Kim HP (2014) [56].

Mitochondrial DNA (mtDNA) must be inherited vertically from one of the parents (see Figure 4) [44,57,58]. In yeast, two haploid parental cell fuse and give rise to a hybrid diploid cell that carries two parental genomes and initially both versions of mitochondria. However, despite fusion of most membrane-bound organelles, very rapidly only one single type of mitochondria is retained while the other is lost, within few generations, leaving the hybrid cell with two genomes but only one parental type of mitochondria. This phenomenon is known as uniparental mitochondrial inheritance. This process is particularly common in yeast species, especially in the *Saccharomyces sensu stricto* complex. This is the case as several members of this group frequently mates both in the wild and the industrial environments. *S. pastorianus* is one of the natural hybrids used extensively in the production of the lager beer. Based on several genomic studies, this hybrid is derived from

hybridisation events between *S. cerevisiae* and *S. eubayanus* with mitochondria passed down from the cold-adapted parent [57,59,60]

Several studies have demonstrated that mitochondrial inheritance is influenced by environmental conditions as well as the effect of mitochondrial DNA on fitness/transcription of nuclear-encoded genes in yeast hybrids [61,62]. For example, carbon source is known to influence the transmission of mitochondria [63]. Under fermentative conditions, cells are likely to lose mtDNA over time and start forming smaller petite colonies. Temperature tolerance is one of the key adaptations that determine evolution within the *Saccharomyces sensu stricto* group [64]. While *S . cerevisiae* grows optimally at the highest optimum at 32°C, *S. uvarum* grows optimally between 23 and 26°C. Given this information, temperature and possibly other environmental factors may play an important role in mitochondrial selection, and this choice subsequently affect the fitness of the organism.

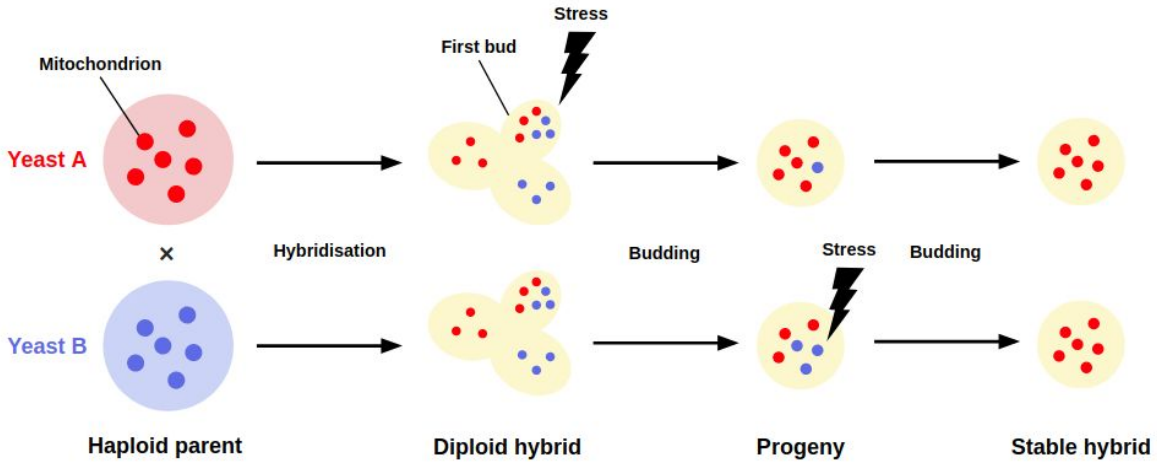


Figure 4. Possible mechanism to explain biased mtDNA transmission in yeast hybrids. After mating, parent A and parent B hybridise and become a diploid hybrid which carries both types of mtDNA. The hybrid is then selected based on the genotype of mtDNA which may play an important role during first budding (top) or budding of the subsequent stages (bottom). This diagram is an adaptation from Hsu YY et al. (2017) [62].

1.6 Cold-stress response

As yeasts inhabit a wide range of environments, they have to be able to withstand a number of stress factors including cold temperatures. According to literatures, yeasts employ many strategies to deal with different degrees of cold shock by adjusting composition of the lipid membrane and expressing specific types of proteins to modulate cellular functions.

1.6.1 Cell membrane remodelling

Cell membrane is an essential component of yeast cells. It encloses cellular contents and prevents them from mixing with external environment. In addition, it is involved in a variety of cellular activities, such as cell-cell adhesion, passive and facilitated diffusion, and cell signalling [65]. Plasma membrane is composed of lipid bilayer and other organic molecules, the composition of which determines structural and functional properties of the membrane as well as global behaviour of living cells.

Like other organisms, yeast plasma membrane contains phospholipids which are in association with sterols, proteins, and carbohydrates of different chain lengths. In *S. cerevisiae*, the membrane is enriched with sterols and certain species of phospholipids (i.e. phosphatidylcholine - PC and phosphatidylethanolamine - PE) [66,67]. Ergosterol is a type of sterols that is capable of maintaining rigidity of the membrane by reducing entropy of the phospholipids, leading to membrane condensation in liquid crystaline state. Yeast cell membrane also contains high proportions of saturated fatty acids (especially oleic acid and palmitoleic acid) which help maintain fluidity of the membrane by the effect of steric repulsion between acyl groups.

In addition to sterols and saturated acyl chains, lowering temperature also decreases membrane fluidity. This is detrimental to growth as it decreases expansion limit and causes cell rupture. To grow at low temperature, yeasts must find a mechanism to maintain flexibility of the membrane. One of the main mechanisms is changing composition of the membrane. *S. cerevisiae* which is a member of mesophilic yeasts responds to cold temperature by up-regulation of the expression and increasing the

activity of specific enzymes involved in fatty acid synthesis. This increases the ratio between palmitic acid and stearic acid as shortening of the acyl groups results in higher membrane flexibility. [68]. Palmitic acid can be unsaturated to produce palmitoleic acid. Interestingly, yeasts became more tolerant to low temperature as well as salt stress when oleate delta 12 unsaturases of sunflower (*Helianthus annuus*) were cloned into the genome of *S. cerevisiae* [69]. What's more, it has been demonstrated that increasing levels of sterols has fatal effects on cold-adapted yeasts. Belrhiti and colleagues showed that fitness of the cold-adapted *S. bayanus* is negatively correlated with the level of sterols in culture medium [70].

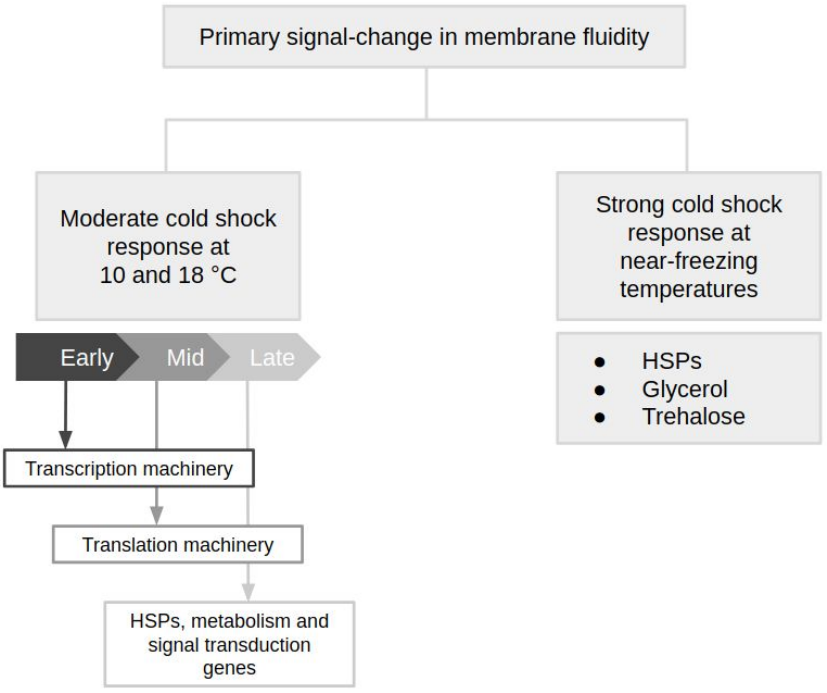


Figure 5. Hierarchical structure of cold-shock response in budding yeast. Cold-shock response is divided into two distinct modes - 1. Moderate cold-shock response which helps yeast maintain regular cell activities and 2. Strong cold-shock response which helps yeast survive the effect of desiccation and thawing at near-freezing temperatures. This diagram is an adaptation from et al. (2014) [71].

1.6.2 Cold-shock protein expression

In addition to changing composition of the cell membrane, yeasts have evolved a collection of cold-shock proteins (CSPs) to maintain regular cell activities and restore any damage caused by cold. Many CSPs are involved in the production of short and branched fatty acids to maintain membrane fluidity. Other are induced to chaperone activities of transcription and translation elements, to correct protein folding and to protect cells from being thawed by ice crystals. Figure 5 shows expression of various CSPs over the range of temperature and over the course of time after induction [71]. The responses between 10 and 18 °C focus on the maintenance of regular cell activities. Many CSPs of the moderate shock response are involved in transcription and translation.

Yeasts have several thermal sensors to monitor fluctuation of the surrounding environments. One of the sensors is plasma membrane. It was demonstrated that changes in physical states of the membrane trigger a cascade of signalling pathways that induces expression of several stress-response genes [72]. For example, SLN1, which is an osmotic sensor that regulates osmotic stress-response genes through a specific mitogen-activated protein kinase (MAPK) cascade, has been proposed that this protein could be involved cold-sensing mechanism [73–75]. NSR1 is a nucleolar protein required for ribosomal RNA processing [76–78]. It has been reported that yeasts with *NSR1* deleted are likely to become more sensitive to cold temperatures [79]. DBP2 is an ATP-dependent RNA helicase that, in association with other nucleoproteins, binds to transcripts of the DEAD-box protein family to facilitate transcription [80,81]. Similar to NSR1, this protein appears to be involved in cold-stress response [82]. Additional classes of CSPs are expressed in response to prolonged exposure to moderate temperature as well as to near-freezing temperature. These include heat-shock proteins (HSPs) and cryoprotectant proteins. HSPs (such as HSP12, HSP26, HSP42, and HSP104) are upregulated in response to different stressors such as temperature stress, oxidative stress and nutrient depletion [83–85]. They function both chaperones and components that maintain cell architecture.

As exposure to near-freezing temperatures can be devastating and may result in desiccation, thawing, and cell lysis, yeasts have to take extra measures to protect themselves by producing glycerol and trehalose [86–89]. These substances are capable of retaining water, resulting in slower water diffusion and water expansion. In *S. cerevisiae*, glycerol and trehalose are synthesised by glycerol phosphate dehydrogenase and trehalose phosphate synthase, respectively.

In wine-making industry, *S. cerevisiae* pre-dominates the fermentation process. It affects several aspects of wine quality such as aroma, taste, and color. Although pectin-free wine is more desirable due to higher clarity, the pectin of *S. cerevisiae* is sluggish at wine-fermentation temperatures (below 15°C for white wines) [90]. There was an attempt to find cold-active pectinases that belong to *Saccharomyces* species. A group of Indian researchers identified an isolate of *Saccharomyces* species with cold-active pectinases (15 - 21 U/ml at 5°C) [91]. However, they were not able to identify the isolates as well as the genes encoding the pectinolytic enzymes.

As mentioned previously, natural hybrids with cold-adapted yeasts are tolerant to cold temperature, suggesting the presence of cold-active enzymes in the parental strains. Phenotypic profiling has revealed that temperature has influence on the patterns of gene expression of different *Saccharomyces* species. Sugar and amino-acid metabolisms of the cold-adapted *Saccharomyces* and the hybrids are not retarded at low temperature [92]. This indicates that enzymes of certain metabolic pathways are adapted to work efficiently at low temperatures. The genes of these enzymes refer to cold-adaptive alleles. The next part of this review will introduce tools for systemic profiling to identify these alleles.

1.7 Experimental tool kits

1.7.1 Deletion collections

Study of the whole yeast genome became possible when the full genome of *S. cerevisiae* was published in 1996 [38]. The genome was annotated for ORFs, obtaining the first complete set of 5,885 protein-encoding genes. However, these genes were hypothetical. To identify functions of the genes, the international

consortium of yeast laboratories created a collection of yeast deletion strains. Over 90% of the genes were knocked out and replaced by the KanMX marker. The marker of each knock out is flanked by unique barcodes (*i.e.* UP-TAG and DOWN-TAG) as well as the universal sequences, which are useful for subsequent analyses using conventional technologies such as microarray and next generation sequencing. This invention is one of the milestones in the history of yeast research that helps us better understand the complexity of biology.

Following decades, the project has been expanded. There are collections available in various backgrounds, i.e. BY4730, BY4739, BY4741, BY4742, and BY4743. Although they have led to numerous discoveries, there are certain drawbacks of using the strains strictly from laboratory background. For example, *ERC1* of the BY strains contains a frameshift mutation [93]. This mutation is specific to the laboratory strains and is not present in wild-type. What is more, the laboratory strains are not suitable for studying polymorphism because they have been subjected to extreme selection in laboratory environments, resulting in lineages of nearly identical genotypes. In contrary, wild-type strains are rich in genetic variation and provide the means to study polymorphism and discover new alleles.

Deletion collection of the strain Sigma 1278b was constructed in Charlie Boone's laboratory (http://www.utoronto.ca/boonelab/). This strain originates from wild-type background and is polymorphic compared to the strain S288C [15,16]. This provides an opportunity to explore phenotypic and genotypic variation with respect to evolutionary history of the strains. The collection was constructed by transferring the standard marker and barcodes from the S288c collection to the wild-type using the strategy shown in Figure 6. It is available in homozygous diploid and haploid of the MAT a and MAT α, as well as heterozygous diploid. The collection can be used for several purposes including genomic profiling to discover gene functions and new alleles.

.7.2 Competition experiments

Based on the concepts of metabolic control analysis, the effect of individual genes on growth and competitive fitness are small [94]. This suggests that the traditional approach of genomic profiling by experimenting on individual knockouts in isolation is

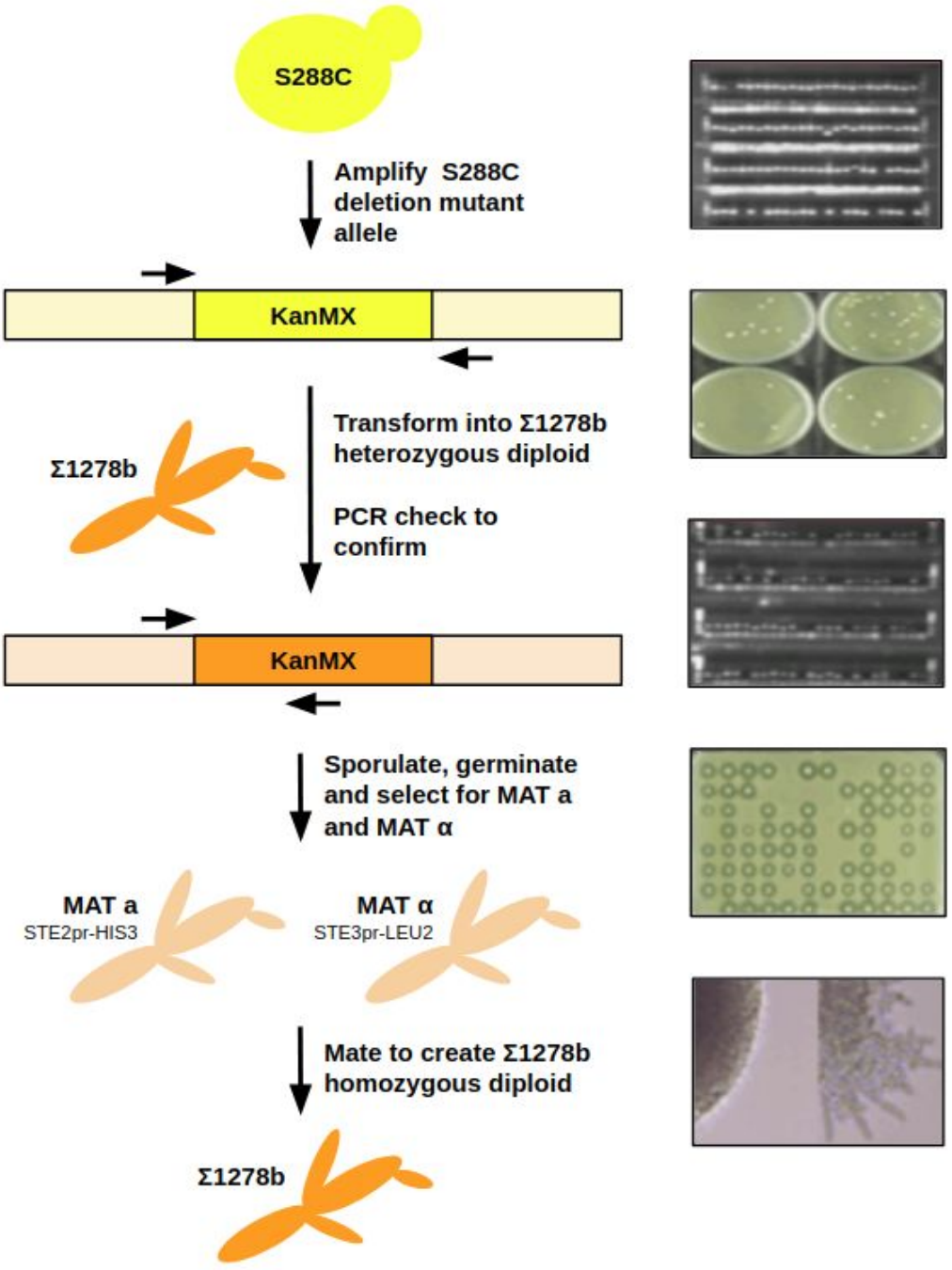


Figure 6. Construction of the Σ1278b deletion collection by transferring barcodes and selection markers from the original S288C collection. This diagram is an adaptation from http://www.utoronto.ca/boonelab/.

not effective. It may lead to mistreating the knockouts, which generate insignificant deviations from the non-mutated strain, as unrelated to a given phenotype.

In 1996, Ronald Davis' laboratory published a new methodology for genomic profiling, known as competition experiments [95]. Competition experiments have several advantages. The method reduces background variation and provides measurements in relative to other knockout strains. This results in higher sensitivity for discriminating two similar phenotypes. In addition, another advantage is cost-effective. Combining with microarray-based technology or next generation sequencing, measurements can be taken simultaneously from a pool of thousands of different knockout strains (see Figure 7) based on their unique barcodes (i.e. UP-TAG and DOWN-TAG). This makes the method less labour-intensive and cost-saving.The protocol of competition experiments were summarised and published in Yeast Systems Biology (Springer) [96]. It describes methodology and materials required for the experiments. This allows researchers to automate a genetic screening in a high-throughput fashion. Despite the drawback of possible cross interaction with the native proteins, this technology is still useful for a number of applications.

The method gains more attention from scientific community. It has been applied for research in different areas, such as drug targeting, stress adaptation, and general genomic and phenotypic profiling. For example, competition experiments were used to reveal genes involved in adaptation to UV stress and methyl methanesulfonate stress [97]. It was also used to reveal the functions of several non-coding RNA in *S. cerevisiae* [98]. All in all, competition experiment has a great potential for a wide range of applications including genomic profiling for cold-adaptive alleles in different strains of budding yeast.

1.7.3 Next-generation sequencing

DNA sequencing is a way to determine precise order of the four nucleotides in DNA molecules. Since 1970s, several methods have been developed to become more precise and more cost-effect, resulting in accelerated scientific discovery and research [99]. Next-generation sequencing (NGS), in particular, can be coupled with barcode analysis to identify deletant strains from cell cultures [96]. It can also be used to profile the quantities of mRNA transcripts to study transcriptomic response at high resolution [100–103]. In addition to these two, a variant of this technique, called chromatin immunoprecipitation sequencing or ChIP-seq, is used extensively to investigate interaction between DNA and protein molecules [104,105]

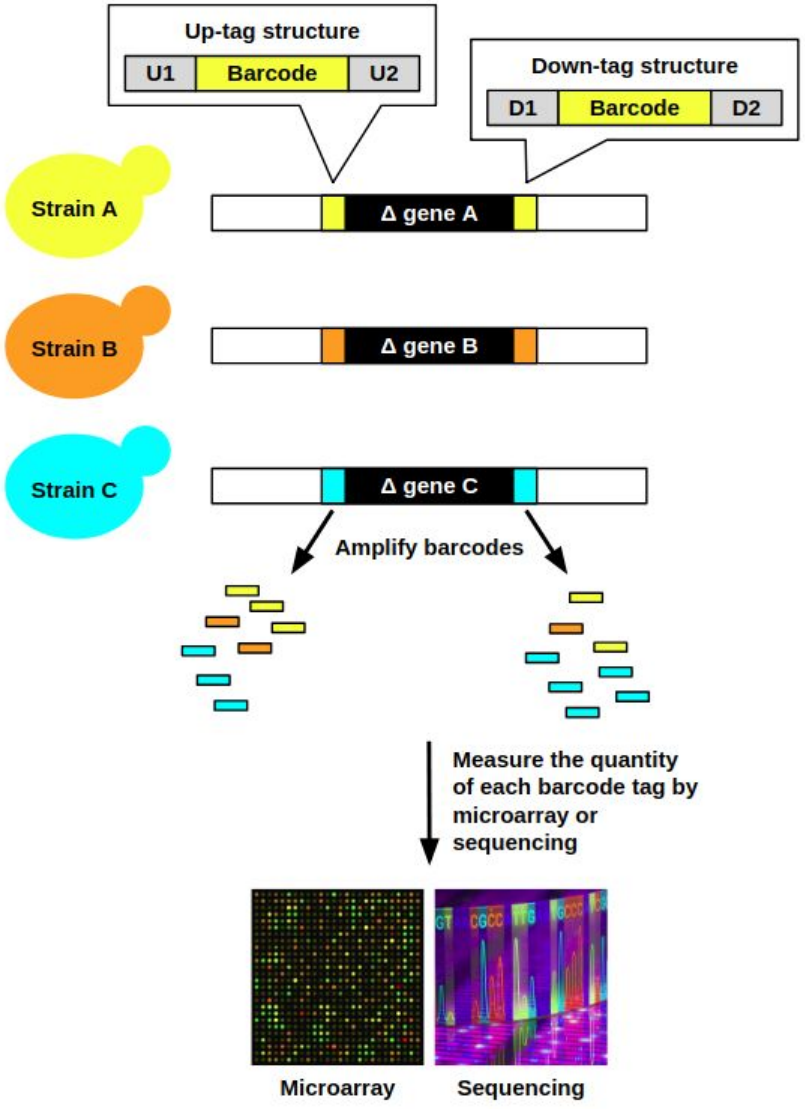


Figure 7. Schematic representation of barcode analysis of samples from competition experiments. Genomic DNA is extracted from cell cultures collected at different intervals during the competition. The extracts are then subjected to PCR amplification by universal primers to recover unique barcodes which can be measured by microarray or sequencing method. This diagram is an adaptation from Delneri D (2010) [106]

NGS has several advantages - e.g. cost-saving (about $40-50 per Gb), high accuracy (no non-specific hybridisation), and fast run time (27 hours to 11 days per run) [107–110]. However, it also has some drawbacks, such as operation cost (up to $654 K) and issues related with PCR amplification (preferential and uneven amplification). Many sequencing platforms are based on PCR amplification and bioimaging. These include Like to 454 Life Sciences, SOLiD, and Illumina/Solexa.

Illumina/Solexa sequencing relies on the formation of amplification clusters and the imaging of fluorescent dyes. Figure 8 shows the strategy of Illumina/Solexa sequencing: library preparation and solid-state bridge amplification [111]. In brief, genomic DNA is fragmented and ligated with adapter sequences, which allows hybridization with oligos on the surface of flow cell. Next, clusters are formed by solid-state bridge amplification to contain about 1,000 copies per cluster. The clusters are linearised and sequenced base by base using fluorescently labelled, reversibly terminated nucleotides. Images are taken after each round of synthesis, following by the removal of the fluorescent dyes. After the sequencing, the images are analysed to reconstruct the sequences using appropriate software.

RNA sequencing is a technique to reveal the presence and the quantity of RNA transcripts in a biological sample in the high-throughput manner. Unlike hybridization-based detection, it can be performed at high resolution to identify alternative splicing and other post-transcriptional modifications on RNA molecules [100–103]. RNA sequencing protocols are RNA samples are prepared and can be purified using a specific technique to enrich for mRNAs, microRNAs and lincRNAs etc. To construct a library, RNA is converted to cDNA by reverse transcriptase. Sequencing can be achieved by producing one read in a single-end sequencing reaction or by two ends separated by an unsequenced fragments in paired-end

reactions. After sequencing, reads can be mapped onto a reference genome to identify and quantify the levels of RNA transcripts. The results can be normalised and analysed by a variety of methods such as differential gene expression analysis, gene co-expression analysis, differential gene co-expression analysis, gene ontology enrichment analysis and pathway analysis [112–118]. This technology will continue to progress and accelerate how we study gene expression profiles more effectively.

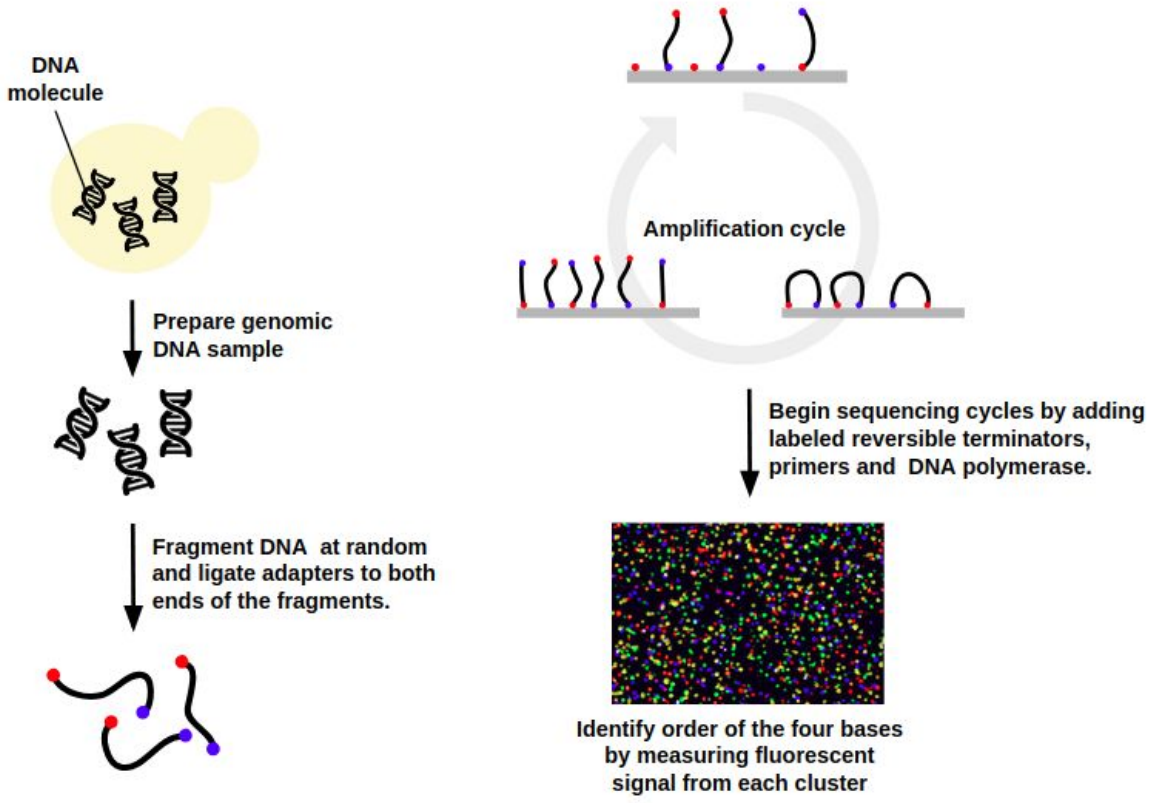


Figure 8. Schematic representation to show the principle of Illumina sequencing. In brief, genomic DNA is prepared from cell samples. It is then fragmented and ligated with adapters to mediate binding to the inside of the flow channel. Unlabeled nucleotides and enzymes are added to amplify individual DNA strands. When DNA clusters become dense enough, labeled reversible terminators, primers and DNA polymerase are added to initiate sequencing cycles. To identify precise order of the four bases, DNA clusters are excited to emit signals which can be captured by a sensitive camera.

1.8 Aims and objectives

This thesis covers a broad range of studies from intraspecific variation between different strains of *S. cerevisiae* to interspecific variation between *S. cerevisiae* and the cold-adapted *S. uvarum* and developing an analytic tool to help with large-scaled phenotypic screening.

To study intraspecific variation, knockout strains of the *S. cerevisiae* strain Σ1278b were obtained from Charlie Boone's lab and put in continuous cultures under six combinations of media (complete medium, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C). Samples were collected at different intervals during the competition. Barcodes were extracted and sequenced to measure the proportion of each strain type. The results were compared with datasets of the BY4743 strain from Delneri D et al. (2008) [119] and Paget CM et al. (2014) [120]. Several analyses were performed including differential fitness analysis to identify strains with significant changes in fitness, gene ontology enrichment analysis to identify relevant biological functions and generalised linear model testing to determine the effect of genomic backgrounds, media and temperatures.

To study interspecific variation between *Saccharomyces* yeasts, hybrids between *S. cerevisiae* and *S. uvarum* were created and subjected to a series of experiments. These two species were chosen because they evolve to fit in the two opposite ends of the temperature range - warm and cold, respectively. In addition, their hybrid appears to have similar genetic makeup to the natural hybrid *S. pastorianus*. This makes this hybrid a good model for studying the effect of hybridisation within the *sensu stricto* group.

According to Hewitt SK et al. (2014) [121], *S. pastorianus* carries a number of chimeric protein complexes. In this study, GCN1 and GCN20 were chosen because they form a binary complex with strong effects on mitotic fitness under various stress conditions [122–124]. To determine the effect of different allelic combinations, *GCN1* and *GCN20* were deleted to create four configurations of the protein complex (GCN1Sc-GCN20Sc, GCN1Sc-GCN20Su, GCN1Su-GCN20Su and GCN1Su-

GCN20Sc). Spot assays were performed to profile their phenotypes under twelve stress conditions. The conditions were selected to represent different types of stresses including temperature stress, amino acid starvation, acid stress, osmotic stress and oxidative stress. Quantitative real-time PCR and protein structure modeling were also carried out to explain of the results at molecular levels.

It is known that *S. pastrianus* carries only mtDNA originated from the cold-adapted parent [44]. This suggests the idea that mtDNA may play an important role in survival of the hybrid. To answer this question, *S. cerevisiae* x *S. uvarum* hybrids were grown under four combinations of media (complete medium to favour fermentation and glycerol medium to favour respiration) and temperatures (30°C to favour *S. cerevisiae* and 16°C to favour *S. uvarum*). RNA samples were collected at log phase and sequenced on the Illumina Hiseq platform. The results were analysed exhaustively to examine differential expression and differential co-expression of individual alleles.

And finally, during my PhD, I developed an analytic tool to analyse growth curve data. The programme was created in R using a number of pre-existing packages such as grofit, mgcv and Rtsne to run statistical analyses and to build an effective graphical representation of the results [125]. It was built into a Shiny platform so that it become user-friendly and suitable for users who have no experience in technical programming [126].

Phenotypic Profiling of the *S. cerevisiae* Strain Σ1278b

Kobchai Dungrattanalert and Daniela Delneri

Faculty of Biology, Medicine and Health,

University of Manchester, Manchester, United Kingdom, M1 7DN

Contributions

Kobchai Dungrattanalert:	Main investigator responsible for carrying out all experimental and computational works in this study.
Daniela Delneri:	Main supervisor.

Abstract

Saccharomyces cerevisiae in the Σ1278b background contains thousands of polymorphic nucleotides when compared to S288C and BY4743. Several studies have demonstrated that this variation has significant impact on the phenotypes of this strain. In this study, over 6,000 heterozygous mutants (each has a single gene deleted) from the Σ1278b collection were profiled under six combinations of media (complete medium, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C). Competition experiments were carried out and coupled with barcode analysis and sequencing on the Illumina Hiseq platform to measure the proportion of each strain type. Results were obtained and compared with datasets of the BY4743 strain from Delneri D et al. (2008) [119] and Paget CM et al. (2014) [120] to determine the impact of intraspecific variation between different strains of budding yeast. It was found that genes involved in transcriptional regulation, plasma membrane and mitochondria are significantly important for the growth of Σ1278b in nutrient limited media at low temperature. Genes involved in protein translation, however, are more crucial during growth in complete medium. Furthermore, fitness profiles of Σ1278b and BY4743 were compared to identify common/specific haploinsufficient and haploproficient genes. Additional analysis on sequence similarity between the two strains suggests that fitness differences may be explained by local intraspecific variation.

2.1 Background

Saccharomyces cerevisiae strain S288C was the first yeast strain to be sequenced fully in 1996 [38]. Since then, there has been a substantial increase in the number of complete genome sequences, leading to better understanding of genome evolution. Given that, we still do not much about yeasts in the wild. To address this problem, different aspects of yeast biology must be studied including ecological and geographical distributions [24,127], population structure [128–131] and different modes of reproduction cycle [132,133]. By doing so, we would understand how evolutionary pressure acts at the whole genome level and how it leads to successful adaptation to fit different environments.

Following the annotation of over 6,000 open reading frames (ORFs) in the complete yeast genome, a number of gene knockout collections have been created in various genetic backgrounds including S288C, BY4743 and Σ1278b. In general, they come in different formats - i.e. haploid MAT a, haploid MAT α, homozygous diploid and heterozygous diploid. The Σ1278b collection has been recently created by transferring barcodes and deletion markers from S288C. This is important as Σ1278b carries thousands of polymorphic nucleotides which give rise to different morphology and physiology when compared to S288C and BY4743 [15,16]

According to the theory of metabolic dominance, the effect of heterozygous loss-of-function alleles are often masked by the presence of genes/alleles with redundant functions [134]. This theory, however, has an exception as deletion of certain gene copies can lead to detrimentally abnormal phenotypes. This phenomenon is termed as haploinsufficiency (HI) [135]. However, the same process can also result in the gain of fitness under certain conditions. This phenomenon is termed as haploproficiency (HP), which is essentially the opposite of HI as deletion of one of the two copy of a given gene lead to noticeably improved phenotypes. It has been proposed that HI and HP can be explained on the basis of maintaining stoichiometry and balance of protein complexes within their corresponding biochemical reactions [136,137]

This study aimed to determine whether there is intraspecific variation of HI and HP between BY4743 and Σ1278b. As datasets of the BY strain have been created before by Delneri D et al. (2008) [119] and Paget CM et al. (2014) [120]. Hereby, only Σ1278b knockout strains were used for competition experiments and barcode analysis on the Illumina Hiseq platform. Over 6,000 strains of heterozygous diploid mutants (each with a single gene deleted) were pooled and competed under six combinations of media (YPD, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C). To quantitatively determine fitness profiles, barcodes within the genome of individual deletant strains were sequenced and analysed. Our findings revealed a number of HI and HP genes in the Sigma 1278b background that are responsible for starvation and cold stress responses at cold (16°C). Under rich YPD condition, genes involved in cytoplasmic protein translation are HP Σ1278b. Under carbon and nitrogen limited conditions, there were roughly equal number of HI and HP genes across different fermentation stages Interestingly, many of these genes are related to transcription regulation, endoplasmic reticulum and mitochondria. HI and HP Σ1278b profiles were then compared with those detected in the BY4743 background. The results showed that there is an overall overlap of HI and HP profiles between the two strains, however there are also a number of genes that display growth differences which may underpin the phenotypic dissimilarity between the two *S. cerevisiae* strains.

2.2 Materials and methods

2.2.1 Strains and culture conditions

Σ1278b heterozygous collection of over 6,000 knockout strains (each with a single gene deleted) was obtained from Charlie Boone's lab. The collection was constructed originally by transferring Deletions in the Sigma 1278b collection are transferring barcodes and deletion markers from S288C. All strains were maintained in YPD (2% (w/v) Bacto Peptone and 1% (w/v) Bacto yeast extract) with 2% (w/v) glucose) and stored at -80°C in cryoprotective medium (YPD with added 15% DMSO).

2.2.2 Competition experiments

Strains in the heterozygous collection were inoculated and grown overnight in YPD using 96-well plates at 30°C for 48 hours or until they reached stationary phase, followed by pooling 10 µl of individual cultures in fresh cryoprotective medium. Prior to competition experiments, 20 litres of YPD and F1-based media with limited nutrients were prepared and the chemostat system was set up as described in Deneri D (2011) [96]. In brief, pH probes were calibrated in standard buffers at pH 7 and subsequently at pH 4. After calibration, they were assembled with flask fermenters and autoclaved for sterilisation. Next, feeding lines of both medium and pH buffer were cleaned by 70% (v/v) ethanol and 1M KOH followed by rinsing with sterile water. Once the system was set, flask fermenters were filled with 100 ml of fresh media and inoculated with 500 µl of the knockout pool (the entire collection mixed together at equal volume at log phase). They were allowed to grow in batch for 24 hours before switching to continuous culture with dilution rate at 0.1/hour and at pH 4.5. It took another 48 hours for the culture to reach the steady state Then, 15-20 ml of sample cultures were collected after 3 (S1), 5 (S2) and 10 (S3) days. All the samples were taken within the 50 generations to avoid mutations. For each collection, OD600 of the samples were measured before storing at -20°C. All the experiment were carried out in duplicate.

2.2.3 Next-generation sequencing

Genomic DNA were extracted from the samples using the CTAB method [138]. The method is generally similar to the phenol-chloroform extraction with additional steps where cells are ruptured in the cationic detergent by sterilised glass beads. To monitor changes in population density, barcodes in the mutant strains were amplified using the same primers as the original BY collection. To avoid mutations, libraries were prepared using Phusion® High-Fidelity DNA polymerase and checked on 1% (w/v) agarose gel for quality control. The experiment generated 48 libraries (from six combinations of media (complete medium, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C) - each was run in duplicate and collected at batch, S1, S2 and S3) and four pool samples (controls) to

be sequenced on the Illumina Hiseq platform. The total of 5,764 strains were recovered after mapping by Bowtie [139,140]

2.2.4 Differential fitness analysis

After mapping, reads were counted and analysed by DESeq2 [141]. To identify significant changes in fitness during steady state (S1 → S3), the standard differential expression analysis based on the negative binomial distribution with size-factor/dispersion estimation for normalisation was run at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.01. Strains were classified as HI and HP when the values of fitness change were negative and positive, respectively.

2.2.5 Other statistical tests

Wilcoxon rank sum test with continuity correction was run to determine the effect of genomic backgrounds on the number of genes with significant fitness change. Analyses were performed using R programming language and run on RStudio version 31.0.153.

2.3 Results

2.3.1 Experimental strategy

In this study, strains in the Σ1278b heterozygous collection were profiled using competition experiments coupled with barcode analysis and sequencing on the Illumina Hiseq platform. Strains were cultured in six combinations of media (complete medium, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C). Samples were collected after 3 (S1), 5 (S2) and 10 (S3) days to avoid mutations as they are likely to emerge after 50 generations [142]. The total of 5,764 strains were recovered after mapping by Bowtie. Subsequently, the results were analysed by DESeq2 at the cutoff thresholds of absolute log2 fold change ≥ 2 and p-value ≤ 0.01. Strains were classified as HI and HP when the values of fitness change were negative and positive, respectively. Experimental strategy is summarised and shown in Figure 1.

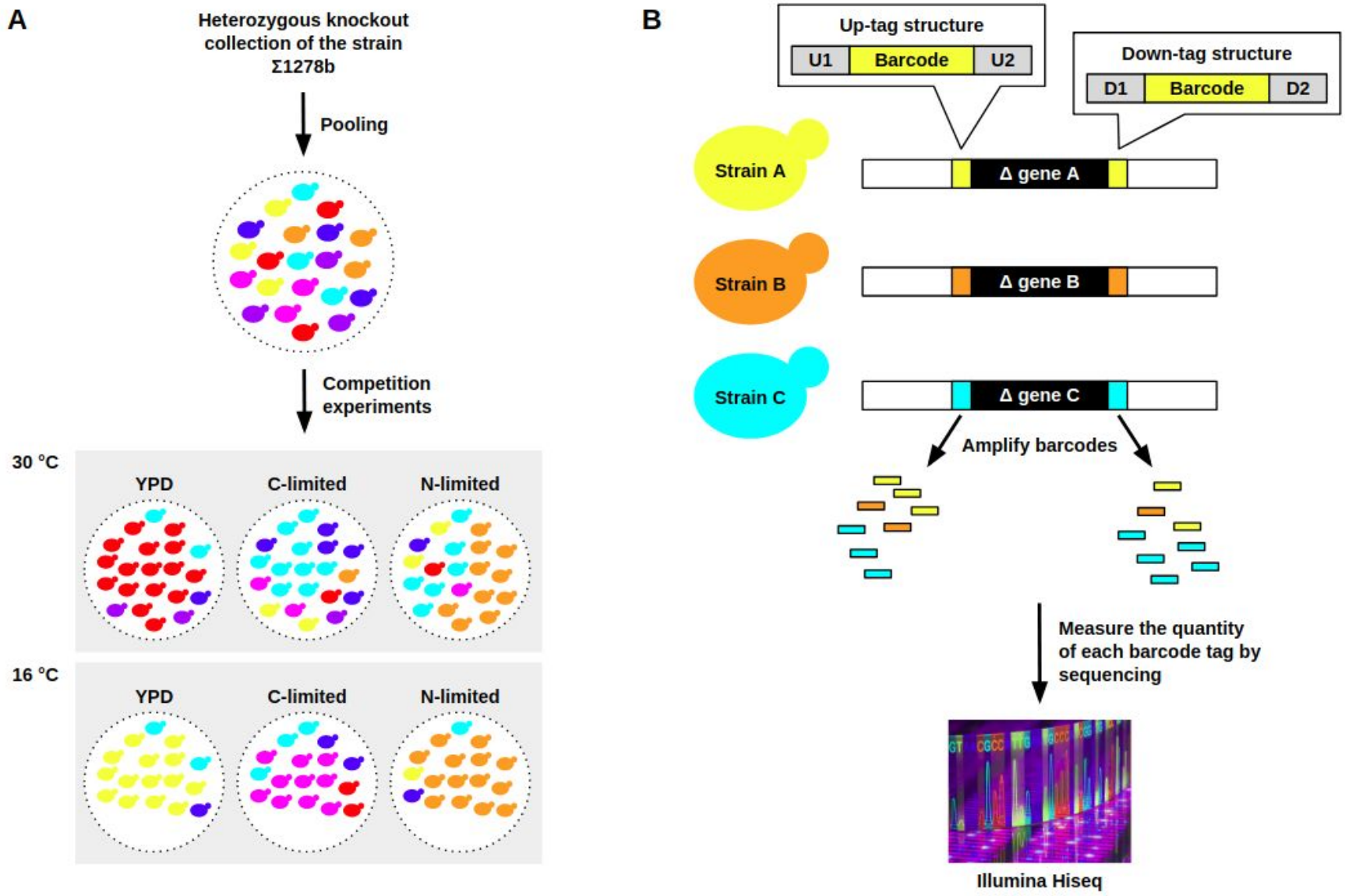

A
Heterozygous knockout collection of the strain Σ1278b
Pooling
Competition experiments
30 °C
YPD
C-limited
N-limited
16 °C
YPD
C-limited
N-limited
B
Up-tag structure
U1
Barcode
U2
Down-tag structure
D1
Barcode
D2
Strain A
Δ gene A
Strain B
Δ gene B
Strain C
Δ gene C
Amplify barcodes
Measure the quantity of each barcode tag by sequencing
Illumina Hiseq

Figure 1. Experimental strategy to obtain phenotypic profiles of the strain Σ1278b. (A) Strains in the knockout collection were pooled together and put into continuous cultures for competition experiments in six combinations of media (complete medium, carbon-limited medium and nitrogen-limited medium) and temperatures (16°C and 30°C). (B) It was followed by barcode analysis and sequencing on the Illumina Hiseq platform to identify and quantify the level of each strain type. The results were mapped and analysed by Bowtie and DESeq2, respectively.

2.3.2 Fitness profiling of the strain Σ1278b

Σ1278b mutants were profiled using continuous culture experiments under nutritional and temperature stress conditions. If a gene is critical for controlling fitness through a specific pathway, this very gene is expected to impair growth upon reducing its dosage from 2 to 1 [96,106]. For that reason, this study should be able to identify genes involved in biologically important pathways at cold.

As shown in Table 1, carbon-limited medium and nitrogen-limited medium had the highest number of genes with significant fitness changes during Pool-Batch and during S1-S3, respectively. Also, it is clear that more genes were affected during steady stages than the earlier phase of the fermentation. Chi-squared test was run on these numbers and confirmed that media and fermentation stages are strong contributors for the growth of this yeast strain (Chi-squared = 85.54, df = 2, p-value < 0.001). Regarding the proportion of HI and HP genes in each condition, it was found that genes with significant fitness change during Pool-Batch were more haploinsufficient in most conditions except for nitrogen-limited medium (see Figure 2). In contrary, genes were more haploproficient in later stages of the fermentation.

Gene ontology mapping was carried out to identify relevant biological functions of the significant genes (see Table 1-6). The results showed that it was impossible to map around 20% of the genes to the Gene ontology database. In nitrogen-limited medium, many processes were affected during the early stage of the fermentation.

	Pool-Batch	S1-S3
YPD	90	162
C-limited	107	84
N-limited	68	309

Table 1. The numbers of genes that showed significant fitness change during Pool-Batch or S1-S3 in the three tested media. The values were derived from the standard differential expression analysis with cutoff thresholds at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.01.

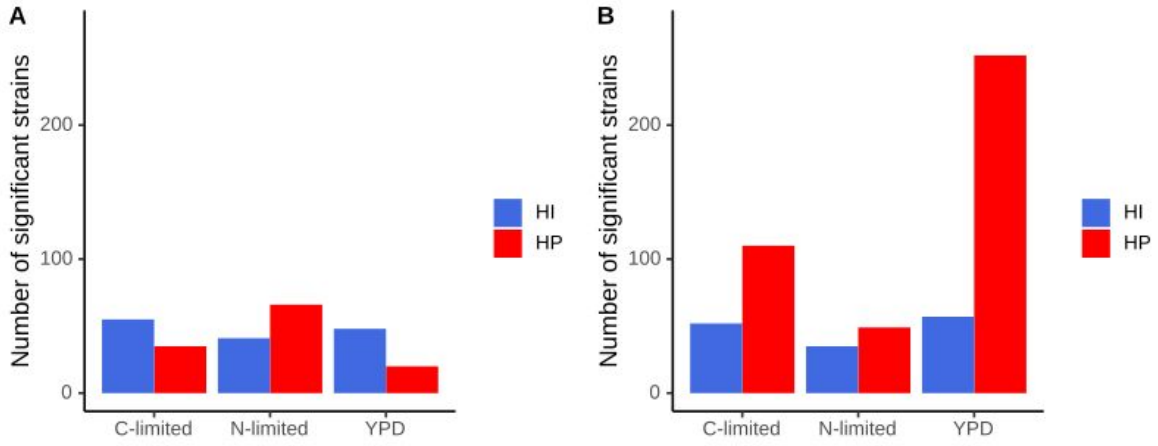


Figure 2. The numbers of HI and HP genes during Pool-Batch or S1-S3 in the three tested media. While genes with log2 fold change is less than -2 were classified as haploinsufficient, those with log2 fold change larger than 2 were classified haploproficient.

These processes include carbohydrate metabolism, protein targeting and proteolysis. While genes that are associated with the first process (such as *GAL2*, *MSS11*, *OPI10* and *UGP1*) were haploinsufficient, genes that are related to the last two pathways (such as *COS1*, *GGA1*, *PIM1* and *PEX10*) turned out to be haploproficient. Annotations of carbon-limited medium and nitrogen-limited medium are similar during steady stage. Genes related to cellular response to DNA damage appeared to be underrepresented. This means that genes such as *RAD5* and *RAD54* are required for growing at cold temperature. Note that these two genes

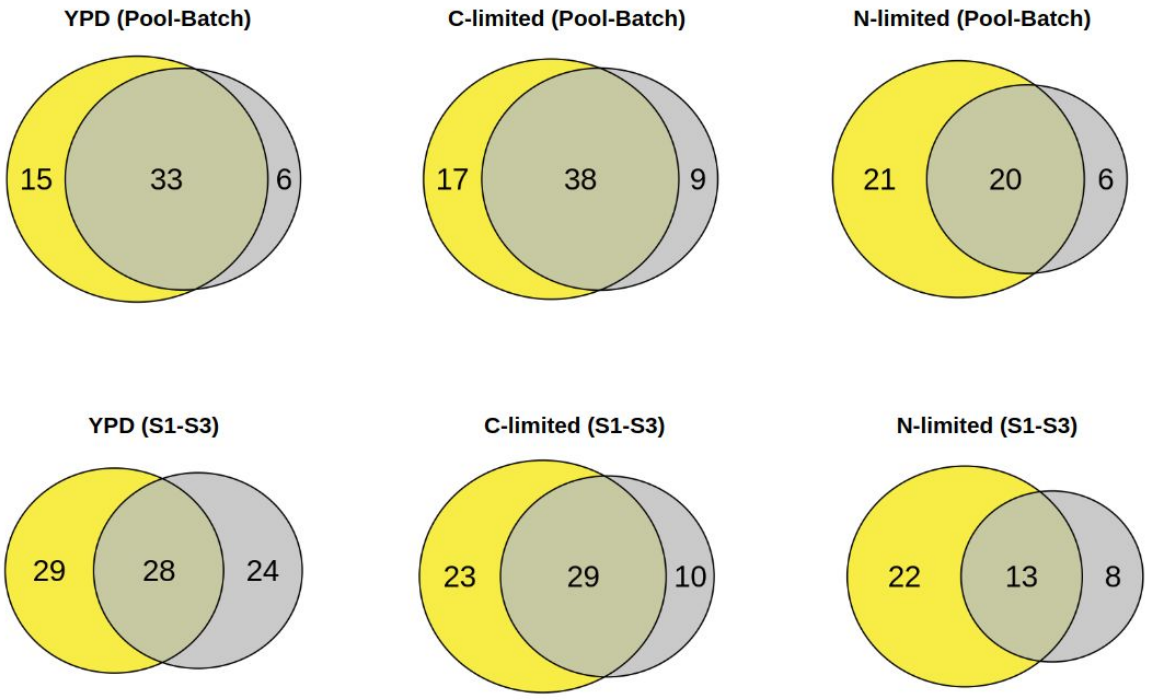


Figure 3. Venn diagrams to show the overlap of haploinsufficient genes between Σ1278b (yellow) and BY4743 (gray). Gene lists were obtained from the standard differential expression analysis with cutoff thresholds at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.01.

belong to the SWI2/SNF2 family and are parts of major DNA repair pathways in yeast [143–145]

2.3.3 intraspecific variation between Σ1278b and BY4743

Venn diagrams were constructed to examine intraspecific variation between Σ1278b and BY4743 (see Figure 3-4). Overall, a large degree of overlapping between the two strains were observed. In most cases, Σ1278b had marginally more significant genes than the BY. and roughly, the overlaps tend to account more than 50% of the gene sets. However, there is an exception as BY4743 had about 20 more genes that were overrepresented in the complete medium during steady stage than the Sigma. In this condition, both BY4743 and Σ1278b-specific genes are associated with cytoplasmic translation and structural organisation within yeast cells. All in all, the

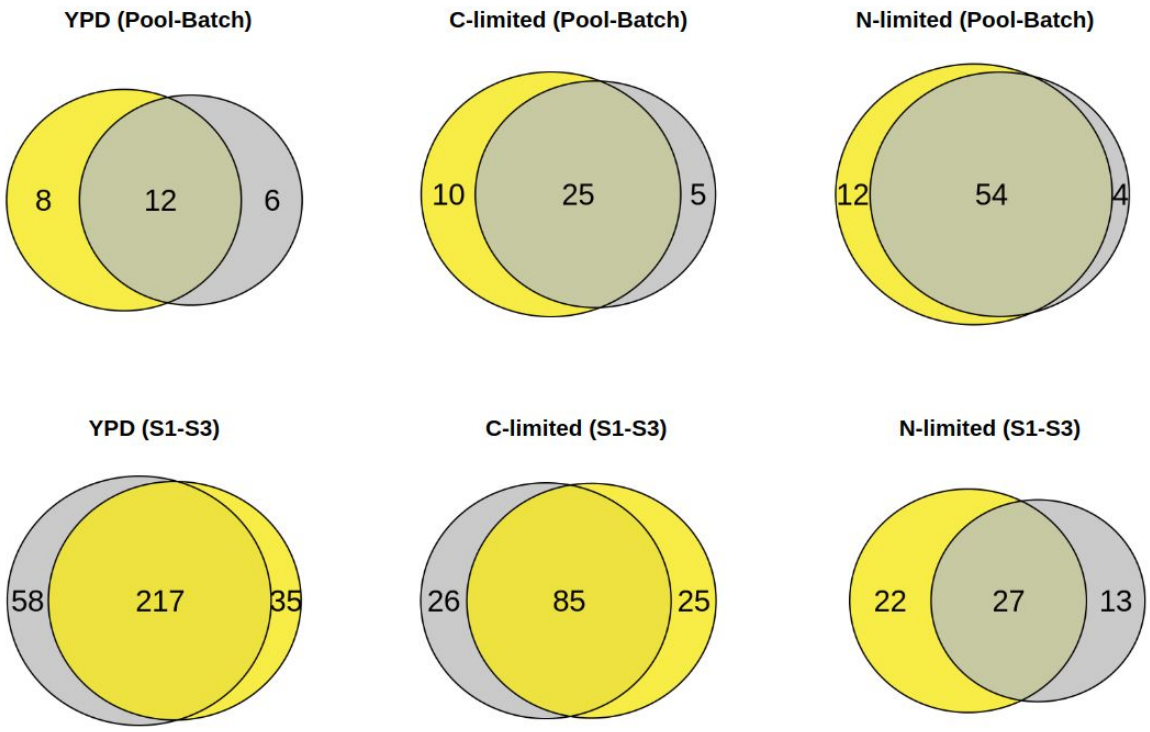


Figure 4. Venn diagrams to show the overlap of haploproficient genes between Σ1278b (yellow) and BY4743 (gray). Gene lists were obtained from the standard differential expression analysis with cutoff thresholds at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.01.

results show that, in general, there is only marginal variation of overall growth under the tested conditions.

2.3.4 Relationship between genomic conservation and fitness changes

As described in the previous section, based on the competition experiment data, Σ1278b and BY4743 are phenotypically very similar. Given that, a small subset of genes are apparently background-specific. It is possible that this could be results of polymorphism within the genomes of these yeast strains. To formally test this hypothesis, sequence dissimilarity of the background-specific genes and the genes that fitness changes were common between the BY and the Sigma were computed. In this study, Damerau-Levenshtein distance was calculated from optimal alignment of DNA sequences [146,147]. Note that, HI and HP were pooled together for

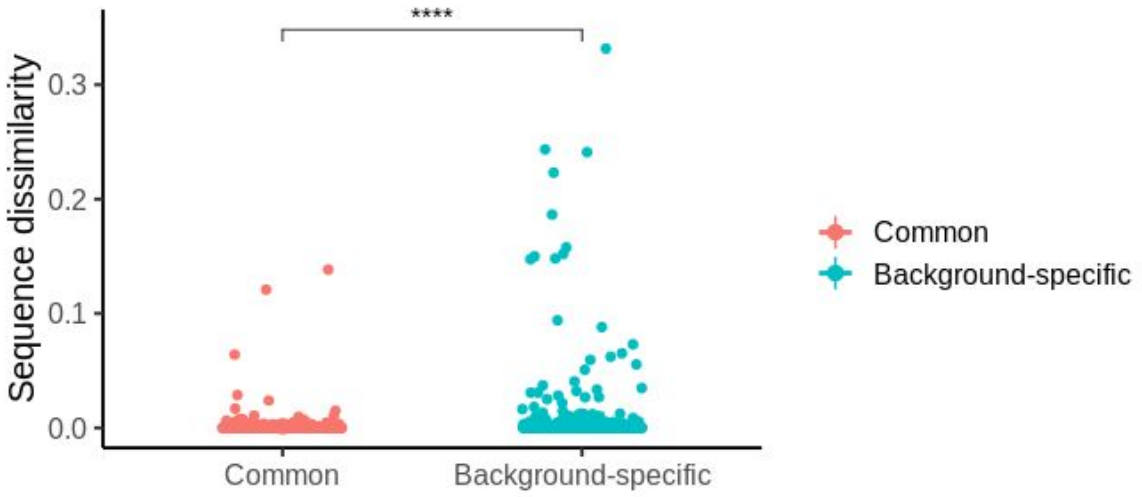


Figure 5. An error plot to show distributions of DNA sequence dissimilarity between Σ1278b and BY4743 of genes that had significant fitness change common or specific to the strain types. Wilcoxon rank sum test with continuity correction was performed to calculate statistical significance between the two.

simplicity. It was found that background-specific genes have slightly higher sequence distance (w = 64736, p-value < 0.001) as shown an error plot in Figure 5. Genes with high distance scores in the background-specific group are *STU1*, *STB2*, *FMP40* and *BSC5*. while the first three genes were significant in the Sigma background, the last gene was significant only in the BY.

2.4 Discussion

In this study, experiments were carried out to obtain fitness profiles of the Σ1278b and to compare intraspecific variation between this strain and BY4743. It was clear that media and fermentation stages have strong influences over the growth of these two strains and probably of *S. cerevisiae* in general [148–150]. In particular, more significant genes were observed during steady stage. Especially under starvation conditions, it was found that genes are involved mostly in DNA repair. Similar events have also been observed in other studies [151–153]. *RAD5* and *RAD54* stands out from the rest as they are a pair of homologous genes that are parts of major DNA

repair mechanisms [143–145]. The results show that remove these genes are likely to impair growth in the Sigma strain under nutritional stress conditions.

As mentioned previously that there is only little phenotypic variation between Σ1278b and BY4743. Even though there were slightly more significant genes in the Sigma than in the BY but this could happen as Σ1278b may not be accustomed to to grow in the laboratory setting optimised for BY strains. It is known that different strains of *S. cerevisiae* tend to have different nutrient requirements [154–156]. The only case that more BY4743 were observed was in the complete medium during S1-S3. A number of HP genes in this condition are involved in chromatin organisation. Michaela C et al. (2013) have demonstrated that haploproficiency are related to the maintenance of genomic integrity [142].

Additionally, relationship between genomic conservation and fitness changes was analysed. It is interesting to discover that DNA sequence dissimilarity has a significant impact on yeast growth. It is known that Σ1278b contains thousands of polymorphic nucleotides that give rise to different morphology and physiology when compared to S288C and BY4743 [15,16]. It was observed that *STU1* which play a key role in microtubule organisation [157,158]. This agrees with the finding as Σ1278b would require this gene for filamentous growth. Based on the results, intraspecific variation may alter protein functions and lead differences in some phenotypes. It has been reported that rewiring is happening in S. cerevisiae at the metabolic level [159–161]. This evidence supports and helps explain the results of this study.

2.5 Conclusion

Competition experiments coupled with barcode analysis and sequencing on the Illumina Hiseq platform were carried out to identify genes with relevant functions in response to nutrient starvation and temperature stress in the Σ1278b background. It was found that genes involved in general metabolism are crucially important for the growth of this strain in nutrient limited media at low temperature. Genes involved in maintaining genomic integrity, however, are more important for growth in complete

medium during steady stage. By comparing data of Σ1278b and BY4743, it was clear that there is little fitness variation between the two strains. Given that, less conserved genes are likely to produce background-specific phenotypes. To validate the results, monoculture experiments and other phenotypic assays could be performed on the mutants in the heterozygous collection of the Sigma and the BY. This should overcome the limitation of high throughput screening as it may suffer from false discoveries.

2.6 Supplementary materials

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	8 out of 48 genes, 16.7%	FYV1, HUA2, YBR116C, YLR123C, YLR317W, YNL114C, YNL235C, YTP1
response to chemical	5 out of 48 genes, 10.4%	FYV5, ROG3, SCP160, UBA4, YKL075C
signaling	5 out of 48 genes, 10.4%	FYV5, LEM3, ROG3, SCP160, SOK1
conjugation	4 out of 48 genes, 8.3%	FYV5, PRM2, ROG3, SCP160
protein complex biogenesis	4 out of 48 genes, 8.3%	ATP12, BIM1, HSP60, VMA21

GO-Slim term	**Cluster frequency**	**Genes annotated to the term**
biological process unknown	5 out of 20 genes, 25%	ECM15, YHR033W, YKL091C, YKL123W, YML089C
organelle assembly	3 out of 20 genes, 15%	KIP3, PKC1, TIF5
cytoskeleton organization	2 out of 20 genes, 10%	KIP3, PKC1
mitotic cell cycle	2 out of 20 genes, 10%	KIP3, PKC1
ion transport	2 out of 20 genes, 10%	PMA1, VPS27

Table S5. Functional annotation of genes with significant fitness change in YPD during Pool-Batch. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	15 out of 57 genes, 26.3%	KEL3, MRX9, RTC4, YBR051W, YBR099C, YFR054C, YGL036W, YGL138C, YGR018C, YGR045C, YHR033W, YJR056C, YNR061C, YOL085C, YPL034W
mitotic cell cycle	5 out of 57 genes, 8.8%	BRN1, BUD3, MCM1, SGS1, SWI5
organelle fission	5 out of 57 genes, 8.8%	BRN1, DNM1, FIS1, SGS1, SWI5
mitochondrion organization	4 out of 57 genes, 7.0%	DNM1, FIS1, NGR1, TRZ1
DNA replication	4 out of 57 genes, 7.0%	BRN1, MCM1, PSF1, SGS1

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	56 out of 252 genes, 22.2%	AIM9, CMS1, CTL1, FYV1, KRE28, NAB6, NOP13, PAU13, PIN2, PRM8, RTR2, SKG3, TDA11, TFB6, UBP11, YBL062W, YBL065W, YBR113W, YBR184W, YBR277C, YCR049C, YDL242W, YDR090C, YDR094W, YDR220C, YDR413C, YDR417C, YDR445C, YGR066C, YGR114C, YGR115C, YHL017W, YHL041W, YHR125W, YIL086C, YIR035C, YJL107C, YJL118W, YKL123W, YLR400W, YLR402W, YLR407W, YML018C, YMR119W-A, YMR130W, YNL140C, YOL019W, YOL118C, YOL131W, YOR024W, YPK9, YPL067C, YPL162C, YPR012W, YPR039W, YPR136C
cytoplasmic translation	30 out of 252 genes, 11.9%	RBG1, RPL15B, RPL27A, RPL32, RPL34A, RPL5, RPS0B, RPS11B, RPS13, RPS14B, RPS15, RPS16A, RPS17A, RPS18B, RPS19A, RPS19B, RPS20, RPS21A, RPS24A, RPS27B, RPS29B, RPS3, RPS30A, RPS4B, RPS5, RPS6A, RPS6B, RPS7B, SRO9, TIF34
rRNA processing	25 out of 252 genes, 9.9%	EMG1, NOC4, NOP12, POP6, RIO1, RPS0B, RPS11B, RPS13, RPS14B, RPS16A, RPS18B, RPS20, RPS21A, RPS24A, RPS27B, RPS6A, RPS6B, RRP46, SNU66, SPB1, SPT5, UTP15, UTP18, UTP30, UTP9
ribosomal small subunit biogenesis	23 out of 252 genes, 9.1%	DRS2, EMG1, NOC4, RIO1, RPS0B, RPS11B, RPS13, RPS14B, RPS16A, RPS17A, RPS18B, RPS19A, RPS19B, RPS20, RPS21A, RPS24A, RPS27B, RPS6A, RPS6B, UTP15, UTP18, UTP30, UTP9
chromatin organization	18 out of 252 genes, 7.1%	APC9, BDF2, CDC6, DOT6, EAF6, FOB1, FPR4, GCR1, HDA2, HDA3, HEL2, IOC3, LGE1, MGA2, RIF1, SGF11, SPN1, SPT6

Table S6. Functional annotation of genes with significant fitness change in YPD during S1-S3. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	20 out of 55 genes, 36.4%	BSC2, DAS2, FMP40, NVJ2, PAU2, PMU1, RMD1, TDA11, YDL041W, YDL071C, YEL028W, YGL074C, YIR016W, YKL077W, YKR073C, YLR123C, YLR434C, YOL036W, YOR345C, YPL197C
nuclear transport	5 out of 55 genes, 9.1%	CLA4, MTR4, THP2, YRA1, ZUO1
lipid transport	4 out of 55 genes, 7.3%	CLA4, DRS2, LAM5, MDM34
response to chemical	3 out of 55 genes, 5.5%	CLA4, DRS2, GLR1
regulation of organelle organization	3 out of 55 genes, 5.5%	CDC13, CLA4, SLY1

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	10 out of 35 genes, 28.6%	STB2, YGR168C, YIL077C, YJL132W, YLR124W, YNL319W, YPL035C, YPL039W, YPR089W, YPR136C
cellular response to DNA damage stimulus	4 out of 35 genes, 11.4%	BDF1, LCD1, RAD28, RTT107
transcription from RNA polymerase II promoter	4 out of 35 genes, 11.4%	CST6, NDT80, RPH1, SPP41
regulation of cell cycle	3 out of 35 genes, 8.6%	HOS2, LCD1, RTT107
response to chemical	3 out of 35 genes, 8.6%	CST6, HPA3, YAR1

Table S1. Functional annotation of genes with significant fitness change in carbon-limited medium during Pool-Batch. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	10 out of 52 genes, 19.2%	ARI1, CPR2, EFM6, EMC10, UIP4, YDL177C, YDR042C, YIL163C, YKL036C, YLR400W
cellular response to DNA damage stimulus	5 out of 52 genes, 9.6%	ESA1, RAD54, RTT107, TOF1, YAF9
DNA repair	5 out of 52 genes, 9.6%	ESA1, RAD54, RTT107, TOF1, YAF9
transcription from RNA polymerase II promoter	5 out of 52 genes, 9.6%	ELP6, ESA1, MFT1, PGD1, TFA2
regulation of cell cycle	4 out of 52 genes, 7.7%	ESA1, RTT107, SPO16, TOF1

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	19 out of 110 genes, 17.3%	DLT1, HGH1, SET4, YBL010C, YBR285W, YCR016W, YDR431W, YER134C, YHL045W, YHR049C-A, YIR020C, YJL064W, YJL118W, YJL119C, YKR073C, YMR196W, YOR169C, YPR126C, YPR130C
transmembrane transport	10 out of 110 genes, 9.1%	ATP7, BAP3, CCC2, FCY21, PEX12, PEX2, SEC65, SSA1, YME1, YOL163W
response to chemical	9 out of 110 genes, 8.2%	AGA2, GPA1, IMP2', PIP2, SLT2, TRX1, UGA2, URM1, YBP1
transcription from RNA polymerase II promoter	9 out of 110 genes, 8.2%	AFT1, ARP7, BMH1, CAF130, CTK1, HHF1, IMP2', PIP2, RPB4
protein modification by small protein conjugation/removal	9 out of 110 genes, 8.2%	BMH1, MPE1, PCI8, PEX12, PEX2, PSH1, RPL40B, SSA1, URM1

Table S2. Functional annotation of genes with significant fitness change in carbon-limited medium during S1-S3. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	9 out of 41 genes, 22%	NVJ2, YBR232C, YBR259W, YCR102W-A, YKL030W, YMR057C, YOR029W, YOR041C, YOR345C
cellular response to DNA damage stimulus	4 out of 41 genes, 9.8%	ARP8, CDC9, EAF5, THP1
carbohydrate metabolic process	4 out of 41 genes, 9.8%	GAL2, MSS11, OPI10, UGP1
DNA repair	4 out of 41 genes, 9.8%	ARP8, CDC9, EAF5, THP1
protein complex biogenesis	4 out of 41 genes, 9.8%	CYC3, HSP60, SEC6, USO1

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	13 out of 66 genes, 19.7%	FAP1, YBR027C, YDR413C, YFR054C, YGL081W, YGR127W, YHR049C-A, YIL077C, YLR049C, YLR225C, YMR290W-A, YNL179C, YPR177C
protein targeting	7 out of 66 genes, 10.6%	COS1, GGA1, PEX10, SNX4, SPC3, SSS1, YRB1
proteolysis involved in cellular protein catabolic process	6 out of 66 genes, 9.1%	COS1, GGA1, GID7, PIM1, PNG1, YRB1
response to chemical	5 out of 66 genes, 7.6%	BOL2, CAT8, MCR1, PNG1, YAR1
transmembrane transport	5 out of 66 genes, 7.6%	ATP1, HXT3, PEX10, SSS1, UGA4

Table S3. Functional annotation of genes with significant fitness change in nitrogen-limited medium during Pool-Batch. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

GO-Slim term	Cluster frequency	Genes annotated to the term
transcription from RNA polymerase II promoter	5 out of 35 genes, 14.3%	CRT10, PCF11, PDR1, PHO4, TAF12
protein complex biogenesis	5 out of 35 genes, 14.3%	PMA1, RAD5, SCO1, SLA1, STU1
cellular response to DNA damage stimulus	4 out of 35 genes, 11.4%	DNA2, EAF6, RAD5, SAC3
response to chemical	4 out of 35 genes, 11.4%	GPR1, HUB1, PDR1, YGL039W
DNA repair	4 out of 35 genes, 11.4%	DNA2, EAF6, RAD5, SAC3

GO-Slim term	Cluster frequency	Genes annotated to the term
biological process unknown	8 out of 49 genes, 16.3%	ECM9, SSP120, YBR206W, YEL023C, YGL260W, YKL077W, YPL257W, YPR012W
transcription from RNA polymerase II promoter	7 out of 49 genes, 14.3%	ARO80, DIG2, ELP4, HIR2, INO4, RSF1, THP1
mitotic cell cycle	6 out of 49 genes, 12.2%	BNI4, BUD23, CHL1, INN1, PDS5, THP1
nuclear transport	6 out of 49 genes, 12.2%	ASM4, BRL1, BUD23, MSG5, NSP1, THP1
nucleobase-containing compound transport	5 out of 49 genes, 10.2%	ASM4, BRL1, NRT1, NSP1, THP1

Table S4. Functional annotation of genes with significant fitness change in nitrogen-limited medium during S1-S3. The mapping was performed using GO Slim Mapper on Saccharomyces Genome Database. Top 5 were selected and shown in the above table. There are two lists of annotations - top and bottom tables are designated for the haploinsufficient genes and the haploproficient genes, respectively.

Impacts of Chimeric GCN1 - GCN20 Complexes in Yeast Hybrids

Kobchai Dungrattanalert, Sarah K Hewitt and Daniela Delneri

Faculty of Biology, Medicine and Health,

University of Manchester, Manchester, United Kingdom, M1 7DN

Contributions

Kobchai Dungrattanalert:	Main investigator responsible for RNA sequencing experiment and transcriptome analysis.
Sarah K Hewitt	Main investigator responsible for strain construction and fitness profiling.
Daniela Delneri:	Main supervisor.

Abstract

Saccharomyces sensu stricto is a group of closely related yeast species with tendency to undergo hybridisation. *S. pastorianus*, which is used actively in lager industry, is derived from *S. cerevisiae* (Sc) and *S. uvarum* (Su)-like species which is now known to be *S. eubayanus* [26,29,30]. A prior study by Sarah K Hewitt SK et al. (2014) [121] suggests that this hybrid has undergone a series of chromosomal rearrangement including duplication, translocation, inversion and selective gene loss. The latter gave rise to chimeric protein complexes that could be associated with specific phenotypic traits in lager yeast. In this study, we investigated the impact of different allelic combinations of the protein complex GCN1-GCN20 on growth performance of the *ScxSu* hybrids. It seems that the subunit GCN1 is critical for yeast growth as it directly interacts with the kinase of translational initiation factor eIF2 and promotes expression of many stress-response genes. As expected, the natural configuration GCN1 Sc-GCN20 Su was the fittest. It grew better than other strains in nearly all testing conditions. We found that this phenotypic advantage is contributed mainly by Sc allele of the gene *GCN1*. Our expression data showed that *GCN1* was expressed differentially with respect to the type of alleles. *GCN20* from S. uvarum however had a more stable expression and may indicate higher selective pressure on this gene allele. This study has revealed the benefits of chimeric protein complexes and how they may evolutionarily arise within yeast hybrids.

3.1 Background

Saccharomyces sensu stricto is a group of closely related yeast species with tendency to undergo hybridisation. *S. pastorianus*, for instance, is an infamous example of natural interspecific hybrids [22]. It is known commonly as lager yeast as it is used extensively for lager production. Among other hybrids, mystery around the origin of this species is largely solved by decades of experiments. Based on several genomic studies, *S. pastorianus* was formed by a recent hybridisation event (15th–16th century) between *S. cerevisiae* (Sc) and *S. uvarum* (Su)-like species which is now known to be *S. eubayanus* [26,29,30]. S. eubayanus is a new species discovered Patagonia that may have been incidentally imported into Europe. However, this theory is largely controversial as there may be an alternative source of this yeast species from within the European continent.

S. pastorianus can be subdivided into two distinct groups based on genomic composition, namely Saaz and Frohberg types (or group 1 and group 2, respectively) [31–33]. While Saaz-type strains (formerly known as *S. carlsbergensis*) are triploids carrying 2n *S. eubayanus* and 1n S. cerevisiae but with 3.5 Mb of DNA missing from the *S. cerevisiae* portion, Frohberg-type strains (such as Weihenstephan strain WS34/70) are tetraploids carrying 2n *S. eubayanus* and 2n *S. cerevisiae* with insignificant losses of DNA from both parents. Despite different genomic contents, Saaz and Frohberg essentially shared the same origin as demonstrated by genetic rearrangements that are common in both lineages. In addition to variation in genetic compositions, they also have distinctly diverged phenotypes as group 1 *S. pastorianus* is more flocculent and less efficient at utilising sugars during fermentation [34–37]. It has been suggested that this variation is caused fundamentally by genomic rearrangements.

After interspecific hybridisation, genomes are subject to a variety of genomic rearrangements including translocation, inversion, duplication and selective gene loss/gain [121,162–164]. Some studies suggest that this process is driven primarily by external environments as a number of stresses have been demonstrated to promote this phenomenon and result in significant adaptations such as copper and

sulphite resistance in wine yeasts [161,165–169]. At larger scales, genomic rearrangement is believed to contribute to speciation as structural changes (including deletion, insertion, duplication, inversion and translocation) are likely to alter genome organization of individuals and result in accumulation of relevant mutations within species and subpopulation [170,171]. Note that there are four main mechanisms of speciation - 1. Allopatric which refers to speciation by geographic isolation, 2. Peripatric which refers to allopatric speciation of a small population with rare mutations, 3. Parapatric which refers to speciation by behavioural or temporal separation and 4. Sympatric which refers to speciation by exploring new niches or by rearranging genomic contents [172–175]. Selective gene loss in particular is the process by which chimeric protein complexes are formed. Since most biological functions are carried out by protein complexes, interspecific assembly of these proteins are likely to have significant impacts on yeast phenotypes. Sarah K Hewitt SK et al. (2014) sequenced three strains of *S. pastorianus* and identify a list of chimeric protein complexes that are retained in both group 1 and group 2, and those that are more specific to one of the two types of lager yeast [121]. In this study, we decided to focus on GCN1-GCN20 complex which is a protein complex that is kept in the Sc-Su configuration exclusively within group 1.

In *S. cerevisiae*, amino acid starvation sets off the expression of over 40 genes that are related to biosynthesis of several amino acids [176,177]. This process requires active translation of GCN4 which is a global activator of genes essential for general control in yeast cells [178–180]. In response to activation by the GCN complex, GCN4 is expressed at much higher levels followed by the transcription of other genes within a few hours after initiation. During the initiation process, uncharged tRNA is delivered from GCN1 to GCN2 and sequentially to 40S ribosomal subunits. GCN1 is a large globular protein and with homology to the N-terminal HEAT repeat domain of fungal translation elongation factor 3 (eEF3) [181]. Essentially, it is a positive regulator of GCN2 as yeast strains with *GCN1* deleted is unable activate GCN2 during stresses [182,183]. In addition to GCN2, Gcn1 forms a complex with GCN20 but such interaction is necessarily required for translation activation [122–124]. At the molecular level, GCN20 binds to GCN1 via short region on the N

terminus (1–189). This interaction is mediated exclusively by G1444 in the eEF3-like region (1330–1617) of GCN1. GCN20 also has an active site that intrinsically binds to ribosome as it can be detected by polysome cosedimentation assays. To form a complex with GCN1 and ribosome, GCN20 requires the presence of ATP molecules that bind preferentially to the C-terminus. But some experiments have suggested that this ATP binding domain is dispensable and not absolutely required under most conditions.

In this study, we aimed to determine the effect of chimeric protein complex in yeast hybrids. More specifically GCN1-GCN20 was chosen and tested using a variety of techniques including serial dilution spotting, real-time PCR, transcription factor binding site prediction and 3D structure modelling. Significantly, we found that retention of *GCN1* Sc and *GCN20* Su within the hybrid is probably driven by structural variation and transcriptional regulation rewiring, respectively.

3.2 Materials and methods

3.2.1 Strains and culture conditions

S. cerevisiae strain BY4741 and *S. uvarum* strain NCYC2669 were used in this study as parental strains and obtained originally from Thermo Scientific UK and from the National Collection of Yeast Cultures, respectively. In addition to background mutations in the parent BY4741, Delneri's lab has created a variant that carries a KanMX cassette at the neutral *AAD3* locus to facilitate selection in the presence of kanamycin. ScxSu hybrids were created by crossing the parents and sequentially selected for individuals that carry mitochondria from the Su parent by restriction fragment length polymorphism based on the restriction pattern of the mitochondrial genes *COX2* and *COX3*. Deletions of specific GCN1 and GCN20 alleles were introduced via homologous recombination to replace the original gene copies with hphMX and natMX selection markers, respectively. This process was carried out previously by Dr. Sarah K Hewitt (a former PhD student in Delneri's lab). Three biological replicates of each strain were made and maintained in either YPD (2% (w/v) Bacto Peptone, 1% (w/v) Bacto yeast extract and 2% (w/v) glucose).

Experiments were carried out in SD-based medium (6.7 g/L yeast nitrogen base without amino acids and 2% (w/v) glucose). All cultures were incubated with shaking at 200 rpm until they reach OD600 between 0.0 and 0.6 (around mid log phase).

3.2.2 Serial-dilution spot assays

To determine the effect of chimeric protein complexes, three biological replicates of ScxSu carrying different copies of *GCN1* and *GCN20* and the hybrid parents were selected and used for sensitivity screening. In brief, cells were grown overnight in YPD at 30 ºC with shaking at 200 rpm and washed twice in distilled water. They were suspended to make cell solutions at OD600 around 0.5 and diluted five times with the dilution factor of 1/10. It was followed by spotted under twelve combinations of chemical treatments (2% w/v glucose, 0.02% w/v glucose, 5% v/v ethanol, 5% v/v dimethyl sulfoxide (DMSO), 60mM acetic acid and 8μM methionine sulfoximine (MSX)) and temperatures (10°C and 30°C) for 48 hours or until growth became visually stable. Images were taken and processed by adaptive filters using a 10x10 pixel moving window with offset at 0.16 from the averaged intensity value. Spot size and intensity were quantified using the R package EBImage [184]. As shown in the equation below, size (S) and intensity (I) of each spot (i) were multiplied and normalised by its corresponding dilution factor (D). The values were then added together to represent fitness (F).

$$F = \sum \left(\frac{\left(S_i \times I_i \right)}{D_i} \right)$$

3.2.3 Quantitative real-time PCR

Expression levels of Sc and Su alleles of the genes *GCN1* and *GCN20* in Sc, Su and ScxSu background were quantified using SYBR™ Green master mix (Thermo Fisher Scientific) and StepOnePlus™ real-time PCR system. Briefly, cells were grown overnight in YPD at 30 ºC with shaking at 200 rpm and washed twice in distilled water, followed by inoculation in 10 mL of fresh SD medium with 2% w/v glucose. Under the same condition, cultures were incubated to reach OD600 between 0.5 and

0.6 (mid log phase). Next, total RNA samples were extracted using Isolate II RNA Mini Kit (Bioline) and converted to cDNA using GoScript™ reverse transcription system (Promega). To account for differences in overall expression levels, *ACT1* Sc and *ACT1* Su were used as references. Primers were designed by Primer3Plus [185] to target specific alleles and amplify around 100-200 bp during the PCR reaction (see Table S1). Ct values were estimated and converted into relative gene expression.

3.2.4 Transcription factor prediction

To predict transcription factors and their binding sites, 250bp upstream regions were obtained from Saccharomyces Genome Database and screened for DNA motifs available on Yeastract [186,187]. This database is an online repository of approximately 163,000 regulatory interactions between transcription factors and genes in budding yeast.

3.2.5 3D protein structure prediction

To predict 3D structures of GCN1 and GCN20, protein sequences were obtained from Saccharomyces Genome Database and evaluated using RaptorX (http://raptorx.uchicago.edu/) [188,189]. This online tool provides a powerful and reliable platform for template-based tertiary structure prediction. To examine the obtained structures, PyMOL version 1.74 was used to align protein structures and graphically compare multiple versions of the GCN proteins [190]

3.2.6 Other statistical tests

Analysis of variance (ANOVA) was run to determine the effect of allele types on the phenotype and expression levels of *GCN1* and *GCN20*. Analyses were performed using R programming language and run on RStudio version 31.0.153.

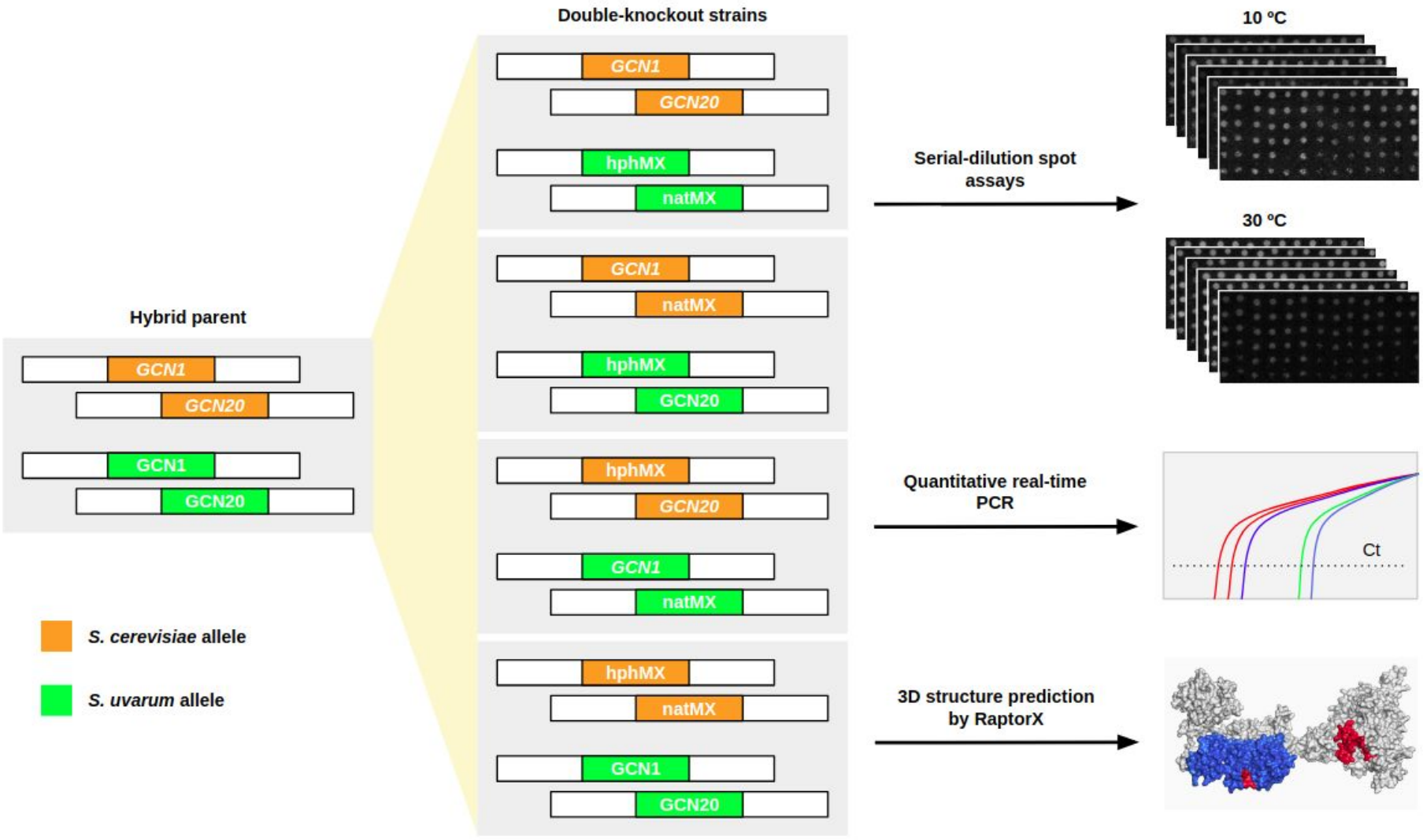
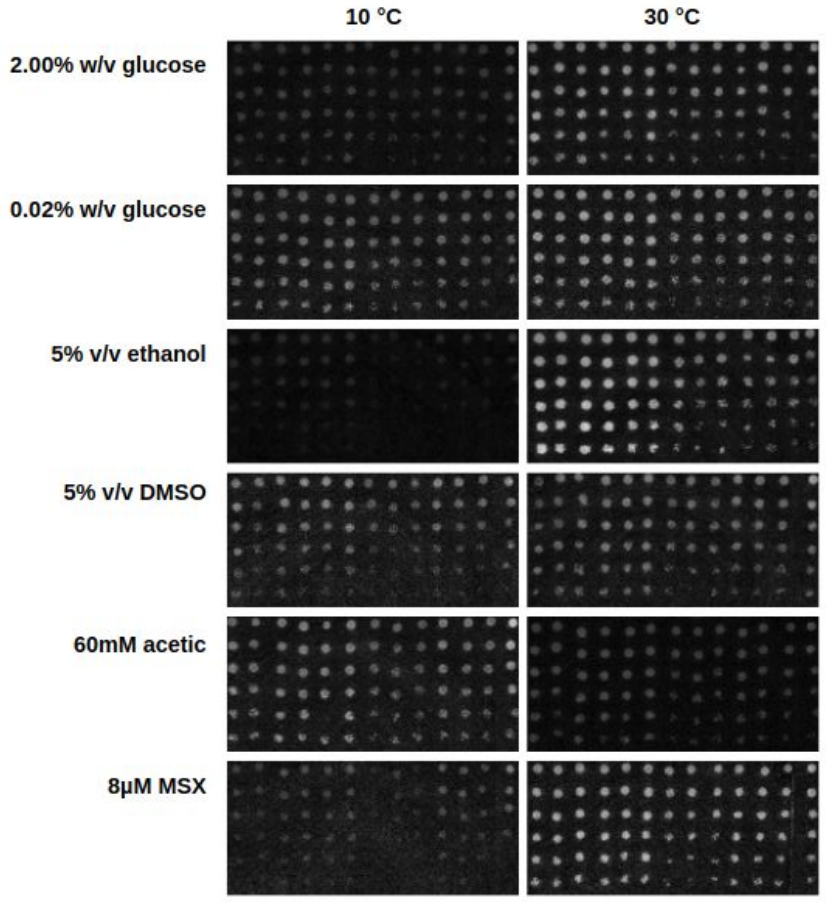

Hybrid parent
GCN1
GCN20
GCN1
GCN20
S. cerevisiae allele
S. uvarum allele
Double-knockout strains
GCN1
GCN20
hphMX
natMX
GCN1
natMX
hphMX
GCN20
hphMX
GCN20
GCN1
natMX
hphMX
natMX
GCN1
GCN20
Serial-dilution spot assays
10 °C
30 °C
Quantitative real-time PCR
Ct
3D structure prediction by RaptorX

Figure 1. Experimental strategy to examine the effect of different chimeric protein complexes in yeast hybrids. Different combinations of GCN1 and GCN20 were created by double deletion on the ScxSu hybrid carrying Su mitochondria. The resulting double-knockout strains were then characterised by serial-dilution spot assays under twelve different conditions, quantitative real-time PCR to measure the levels of gene expression and 3D structure prediction by RaptorX to build complete/partial structures of the protein GCN1 and GCN20 of both allele types.

3.3 Results

3.3.1 Experimental strategy

To examine the effect of chimeric protein complexes in yeast hybrids, three biological replicates of the ScxSu hybrid were created and forced to carry specific combinations of GCN1 and GCN20 by Dr. Sarah K Hewitt (a former PhD student in Delneri's lab). This protein complex was chosen primarily because of its retention within the genome of group 1 *S. pastorianus* [121]. The knockout strains were characterised by serial-dilution spot assays under twelve combinations of chemical treatments (2% w/v glucose, 0.02% w/v glucose, 5% v/v ethanol, 5% v/v DMSO, 60mM acetic acid and 8μM MSX) and temperatures (10°C and 30°C). Images were taken and processed by adaptive filters using a 10x10 pixel moving window with offset at 0.16 from the averaged intensity value. Spot size and intensity were quantified and used to calculate fitness scores (see Materials and methods). Quantitative real-time PCR and 3D structure prediction were also carried out to explain phenotypic variation amongst different configurations of the protein complex GCN1-GCN20. Experimental strategy is summarised and shown in Figure 1.


10 °C
30 °C
2.00% w/v glucose
0.02% w/v glucose
5% v/v ethanol
5% v/v DMSO
60mM acetic
8µM MSX

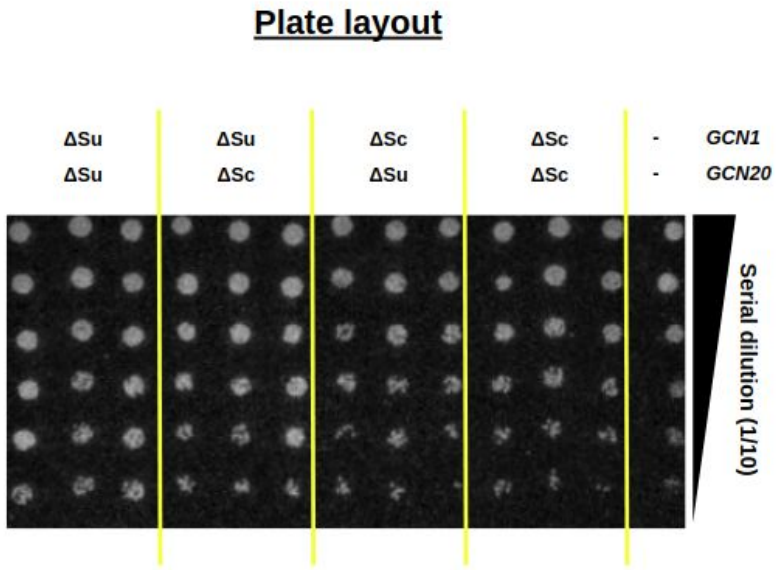


Plate layout
ΔSu ΔSu
ΔSu ΔSc
ΔSc ΔSu
ΔSc ΔSc
- GCN1
- GCN20
Serial dilution (1/10)

Figure 2. Phenotypes of different GCN1-GCN20 complexes in the ScxSu hybrid background. Cells were cultured overnight in YPD at 30°C before exposure to twelve combinations of chemical treatments and temperatures for 48 hours or until no more growth observed. (left) Images were taken (raw without processing). (right) Cells were diluted and spotted on the solid media in this layout.

3.3.2 Phenotypes of different GCN1-GCN20 complexes

If chimeric GCN1-GCN20 has an effect on the phenotype of yeast hybrids, it is expected that deletion of *GCN1* and *GCN20* alleles will affect growth and resistance towards certain stressors. Figure 2 shows the results of serial-dilution spot assays. It is clear that configuration of the protein complex indeed has significant impacts on stress responses. Differences were seen more clearly at 10°C but not much at 30°C.

It appears that *GCN1* deletion had stronger influence on mitotic fitness in stressful environments. In all twelve conditions, strains with *GCN1* Sc grew drastically better than *GCN1* Su. As for *GCN20*, Sc allele of this gene seemed to correlate with growth defects especially under intense amino acid and nucleoside starvation. This starvation was induced by MSX which is a potent inhibitor of glutamine synthetase [191–193]. Interestingly, the presence of *GCN20* Su appeared to greatly enhance phenotypes and make the ScxSu hybrids perform better at cold. To further exploit the data, images of the spot assays were processed to compute fitness scores from spot size and intensity. The obtained values were normalised and analysed by ANOVA (see Table S2). The results show that phenotype of the hybrids is indeed determined by allele types of both *GCN1* and *GCN20* (Df = 1; F-value 5.059; P-value = 0.025). This implies that the two GCN protein may intact differentially to each other in the hybrid background. The data were simplified and shown as a heatmap in Figure 3. The figure is undoubtedly a better graphical representation of this dataset as it shows contrast between strains with similar growth profiles. It is clear that certain configurations of the protein complex GCN1-GCN20 provide higher competitive advantages under stress conditions. In particular, *GCN1* Sc and *GCN20*

Su which are retained in the natural hybrid *S. pastorianus* were able to fit into a wider range of environments.

3.3.3 Expression of *GCN1* and *GCN20* in yeast hybrids

Fitness assays have revealed the effect of chimeric protein complexes on yeast phenotypes. In addition to these findings, quantitative real-time PCR was carried out to explain how different types of *GCN1* and *GCN20* give rise to different fitness

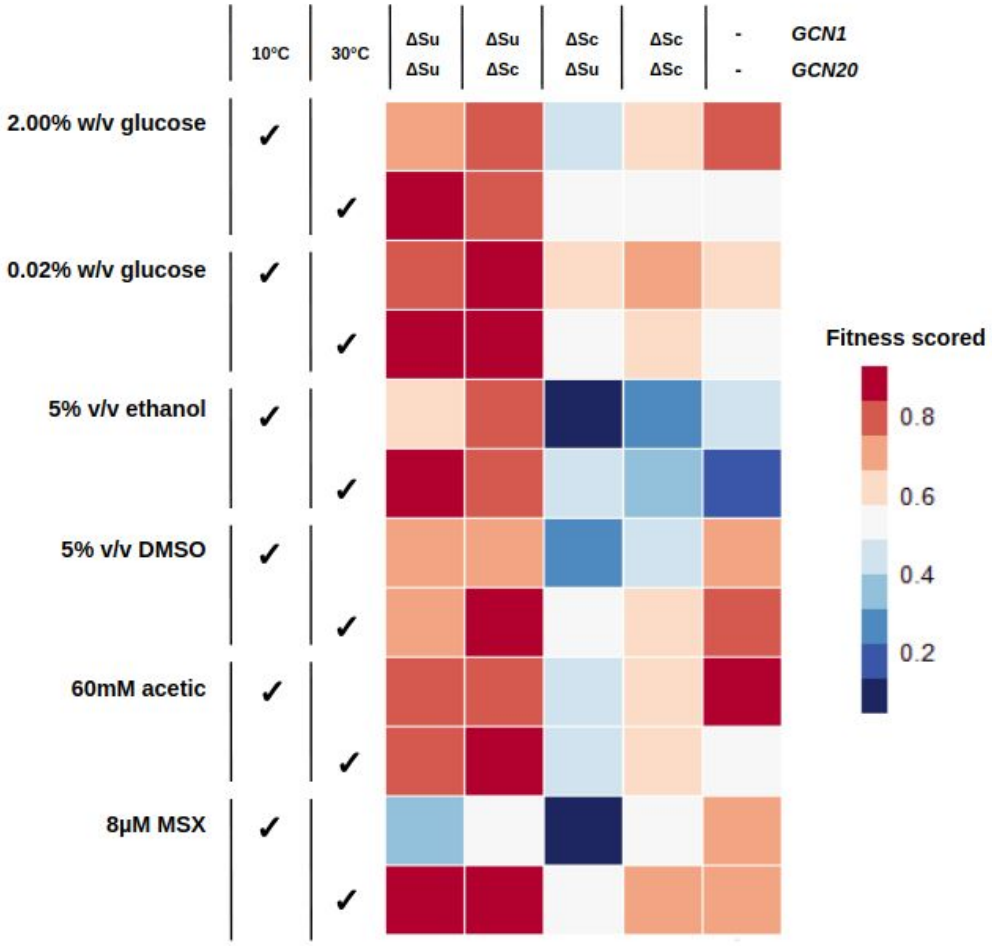


Figure 3. Heatmap showing phenotypic variation amongst different configuration of the GCN1-GCN20 complexes in the hybrid background. Images of the serial-dilution spot assays were taken and analysed as mentioned in the Materials and methods. Spot size and intensity were quantified and converted into fitness scores. The scores were normalised for fair comparison across different conditions.

profiles. For fair comparison, the experiment was done in triplicates in the hybrid parents providing a system where both copies of *GCN1* and *GCN20* coexist and every configuration of the protein complex can potentially be formed depending on expression of the protein subunits. For simplicity, the hybrids were grown in 2.00% w/v glucose at 10°C and at 30°C. Note that, in this study, *ACT1* of both Sc and Su were used as references to account for differences in overall gene expression. Data were collected and analysed by ANOVA (see Table S3-4). Also, the results were plotted and shown in Figure 4. Allele types and temperatures appeared to strongly affect the levels of gene expression as both *GCN1* Sc and *GCN20* Sc were upregulated at 30°C but downregulated at 10°C, and vice versa for those derived from the Su subgenome. This is true for most cases except for *GCN20* Su as it was expressed constantly at high levels at both warm and cold. This may explain the retention of *GCN20* Su as it is definitely more advantageous to retain this copy of the gene especially at low temperatures.

To explain this phenomenon, additional analysis was carried out on the 250bp upstream regions of both *GCN20* Sc and *GCN20* Su. Transcription factors for both *GCN20* were predicted by motif matching on Yeastract. Table 1 shows that *GCN20* Su is enriched with a larger number of transcription factors binding sites when compared to the Sc allele. While 3 transcription factors are specific to the Sc allele, 8 other transcription factors are found to be specific to the Su counterpart. 7 transcription factors (AZF1, FKH1, FKH2, GSM1, HAA1, HSF1 and STB1) are, however, common to both alleles and could be involved in general regulation of the

***GCN20* Sc**	***GCN20* Su**	**Transcription factors**
✓		ASH1, MOT1 and RLM1
	✓	HAC1, HAP2, HAP3, HAP4, IME1 RGT1, XBP1 and YAP1
✓	✓	AZF1, FKH1, FKH2, GSM1, HAA1 HSF1 and STB1

Table 1. Potential transcriptions factors that may bind to the upstream regions of *GCN*20 Sc and *GCN*20 Su. The prediction was based on 250 bp upstream of the coding regions and all available binding sites in the Yeastract repository.

gene. This indicates that *GCN20* Su may have a tighter regulation such that its expression could be adjusted in response to internal and external environments. This set of data, however, cannot explain the retention of *GCN1* Sc in the natural hybrid as it is transcriptionally unfavourable at lower temperatures.

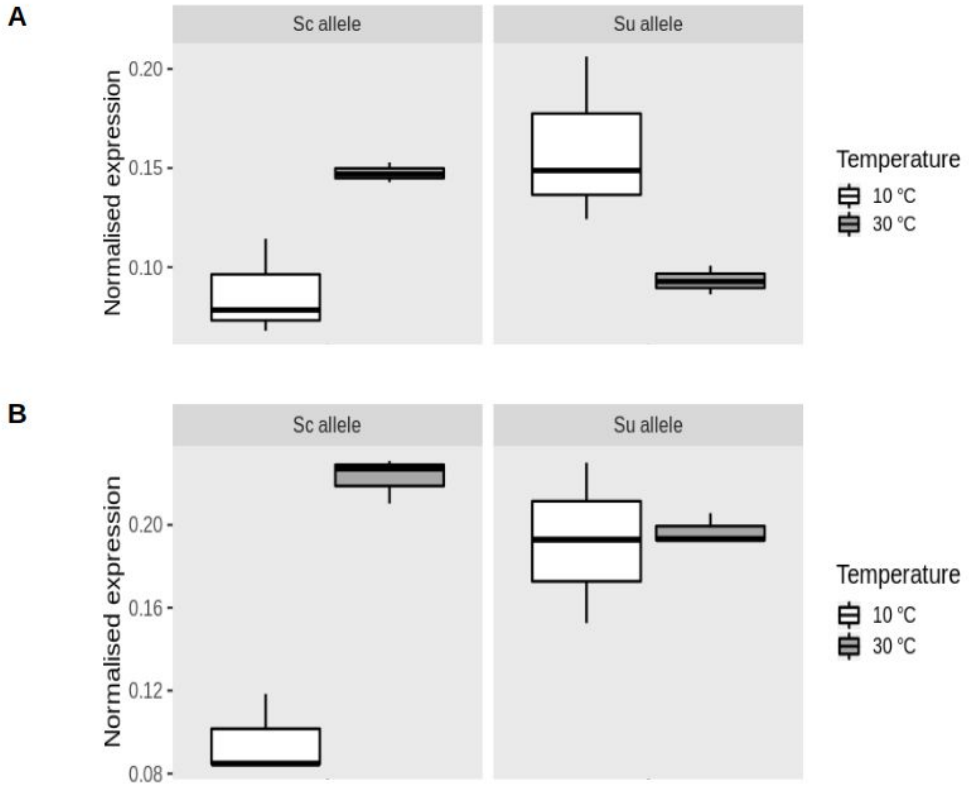


Figure 4. Expression levels of *GCN1* and *GCN20* of individual Sc and Su alleles in the ScxSu hybrids. Cells were cultured and grown in SD medium at 10°C and 30°C to reach OD600 between 0.5 and 0.6 (mid log phase) before converted to cDNA and quantified using quantitative real-time PCR. Ct values were estimated and normalised by *ACT1* levels of their respective sub-genomes.

3.3.4 Three-dimensional structure of the protein GCN1

Based on the fitness data, GCN1 Sc is required for stress responses but it is unclear how exactly this particularly copy of GCN1 is associated with such phenotypes. One possibility is that functional differences of GCN1 Sc and GCN1 Su are caused by structural variation. To test this hypothesis, 3D structures of both GCN1 proteins were predicted using RaptorX - an online tool that provides a powerful and reliable platform for template-based tertiary structure prediction.

The sequences of both GCN1 Sc and GCN1 Su were retrieved from Saccharomyces Genome Database. Both proteins contain the total of 2,672 amino acid residues and several key features including M1 (754-796) and M7 (1,458-1,465) for binding to ribosome, eEF3-like domain (1,330-1,617) for binding to GCN20 and RWDB domain (2,052-2,428) for binding to GCN2 [181]. Since GCN2 binding domain has been modelled by Ramachandra R et al. (2017) [194], it was a better choice to figure out structures of the remaining three domains. 1,381 residues (672-2,052) between M1 and M7 of GCN1 Sc and GCN1 Su were obtained and used for structure modelling on RaptorX. The results show that GCN1 in general has two distinct domains based on the best aligned templates (see Table S5). As shown in Figure 5, domain 1 (1,316-2,062) contain binding sites for both ribosome and GCN20. Despite minor variation of protein sequences between the two alleles, domain 1 of both Sc and Su are structurally similar and may not cause of phenotypic differences observed in previous experiments. Based on the results from RaptorX, domain 1 is structurally similar to the Elongation Factor 3 (template: 2iw3A; p-value < 0.05). Unlike domain 1, domain 2 (672-1,315) was more difficult to model as RaptorX required more templates to fit this section of the protein. More importantly, domain 2 of GCN1 Sc and GCN1 Su were built using different sets of templates due to distinct variation at the structure levels.

3.3.5 Three-dimensional structure of the protein GCN20

Based on the expression data, *GCN20* Su appears to be transcriptionally more robust than *GCN20* Sc. Given that, it is still a chance that the observed differences in

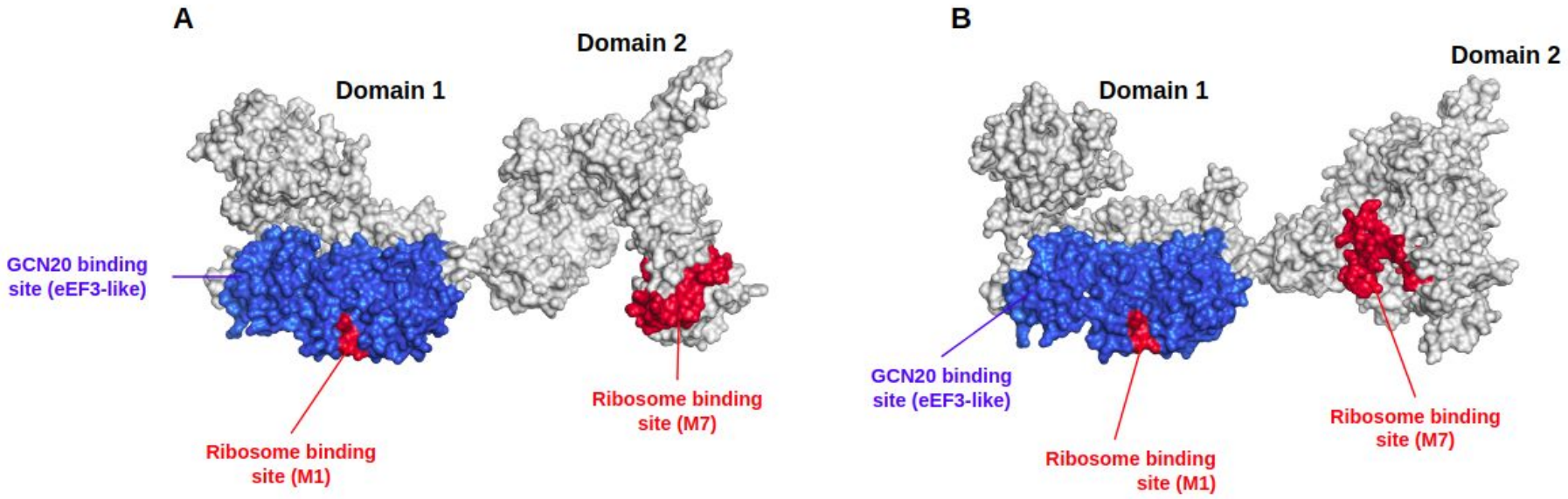

A
Domain 2
Domain 1
GCN20 binding
site (eEF3-like)
Ribosome binding
site (M7)
Ribosome binding
site (M1)
B
Domain 2
Domain 1
GCN20 binding
site (eEF3-like)
Ribosome binding
site (M7)
Ribosome binding
site (M1)

Figure 5. Predicted structures of GCN1 Sc and GCN1 Su, respectively. To make the prediction, amino acid sequences (672-2,052) were retrieved from Saccharomyces Genome Database and submitted to RaptorX for threading against multiple homologous templates. Templates with the best model quality were then assigned to specific sections of the proteins sequences.

phenotype of the two GCN20 proteins could be caused by structural variation. To test this hypothesis, 3D structures of both GCN20 proteins were predicted using RaptorX.

The sequences of both GCN20 Sc and GCN20 Su were retrieved from Saccharomyces Genome Database. They are small globular proteins that contain 193 amino acid residues. Unlike GCN1, GCN20 is not well studied and there is not much information about the structure of this protein. The most up-to-date model of GCN20 is P43535 deposited on the www.proteinmodelportal.org [195]. The results show that GCN20 is composed two domains based on the best aligned templates (see Table S6). As shown in Figure 6, domain 1 (644-1,381) of GCN20 Sc and GCN20 Su were modelled after the structure of a translation factor EttA (template: 3j5sD; p-value < 0.05). The second domain (1-643), however, was more challenging to model as alignment scores were critically low such that the solutions gave out by RaptorX may not be reliable. The only match that is common in both GCN20 is the highly conserved PWI tri-peptide (PWI) domain (template: 1mp1A; p-value < 0.05). Additionally, the prediction suggest that a portion around 35 residues of this region could be highly disordered and may act functionally as an inter-modular linker (see Figure S1). All in all, there is no evidence to support that fitness differences between GCN20 Sc and GCN20 Su are caused by variation at the structure level.

3.4 Discussion

In this study, a series of experiments were carried out to examine the effect of chimeric protein complexes in yeast hybrids. The experiments demonstrated that *GCN1* Sc and *GCN20* Su are required to enhance stress responses in the hybrid

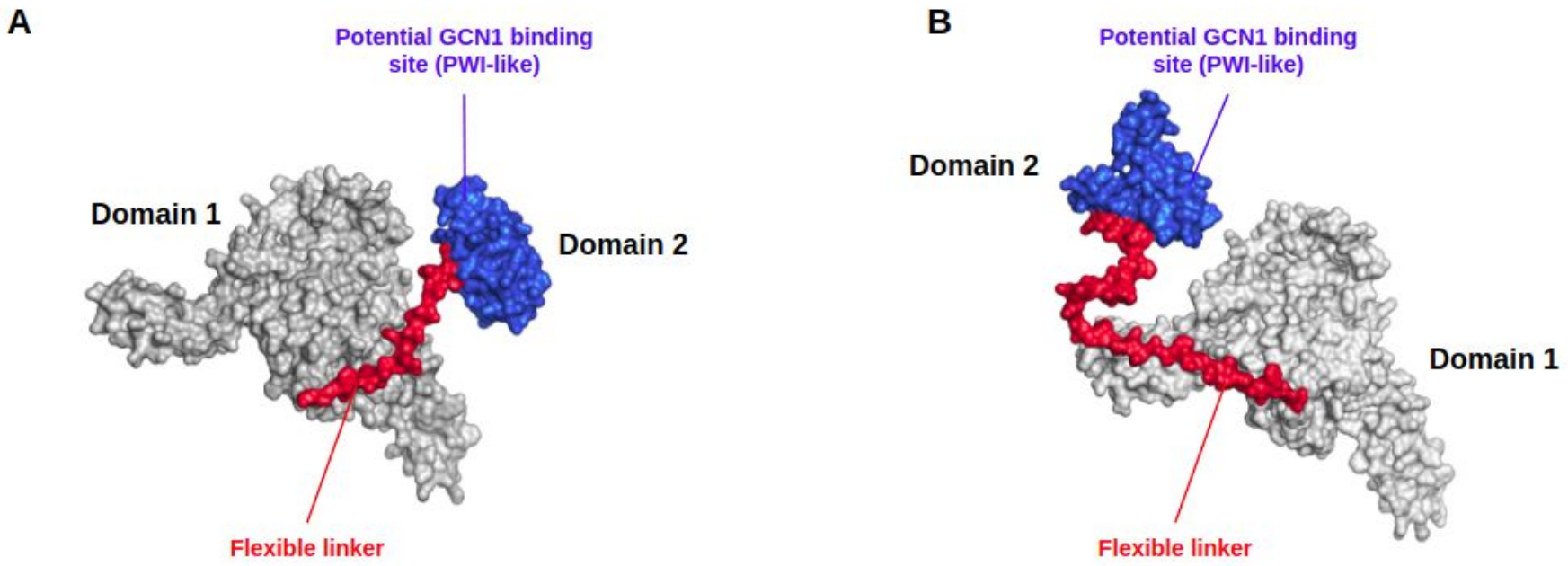

A
Potential GCN1 binding site (PWI-like)
Domain 1
Domain 2
Flexible linker
B
Potential GCN1 binding site (PWI-like)
Domain 2
Domain 1
Flexible linker

Figure 6. Predicted structures of GCN20 Sc and GCN20 Su, respectively. To make the prediction, amino acid sequences were retrieved from Saccharomyces Genome Database and submitted to RaptorX for threading against multiple homologous templates. Templates with the best model quality were then assigned to specific sections of the proteins sequences.

background. More importantly, this allele combination appeared to take an edge and outperform the others in most conditions. Interesting ones are 5% v/v ethanol and 8μM MSX at 10°C which induce strong nutritional starvation and moderate cold shock. This result agrees with previous works that used classical genetics to work out the phenotypes of *GCN1* [196–198] and *GCN20* [199–201]. However, none have considered the influence of allele types. This is important as it may explain why *GCN1* Sc-*GCN20* Su is conserved only in the group 1 *S. pastorianus*. Unlike group 2, group 1 hybrid has lost a large portion of the *S. cerevisiae* genome but retains nearly all genomic parts of the *S. eubayanus* parent [31–33]. This makes the hybrid more tolerant to low temperatures and more suitable for cold fermentation [36,37,202,203]. It is likely that retaining GCN1 from *S. cerevisiae* may help the group 1 hybrid with nutrient utilisation during fermentation and GCN20 from the cold adapted parent may act as a modulator that fine-tune the function of GCN1 under suboptimal conditions.

Gene expression studies shows that *GCN1* Sc and *GCN20* Sc are transcribed more actively at warm 30°C. This complies with the nature of budding yeast that grows optimally at around this temperature [25,204]. However, unexpectedly, *GCN20* Su is expressed constantly at high levels at both warm and cold. This would give a competitive advantage to the hybrid as it would be more resilient to a broader range of environments. Interestingly, upstream region of GCN20 Su is enriched with transcription factor binding sites. It is known that noncoding regions evolve more rapidly than coding regions, which can result in the loss or gain of binding elements. Changes in transcription factor binding can lead to different gene expression patterns even between closely related species [205,206]. More specifically, it has

been demonstrated that transcription factor binding sites are diverged across related species within the *sensu stricto* group [207,208]. Based on breakpoint analysis by Hewitt SK et al. (2017) [121], there is no major breakpoint near *GCN20* of the group 1 *S. pastorianus*. This implies that GCN20 expression of lager yeast could be similar to that of our synthetic hybrid in this study.

Three-dimensional structures of GCN1 and GCN20 were predicted by RaptorX to examine structural variation and how it may affect phenotype of the yeast hybrids. Based on the prediction results, C-terminus of GCN1 is composed of domain 1 (1,316-2,062) and domain 2 (672-1,315). Domain 1 of both GCN1 Sc and GCN1 Su were fit to the same template (template: 2iw3A; p-value < 0.05) at exceptionally high model quality. This indicates that the prediction of this GCN1 segment is generally reliable. This also implies that domain 1 of both GCN1 proteins are structurally and probably functionally identical. On the other hand, domain 2 of this protein was more difficult to model as multiple templates were required to align with this section of the protein. As domain 2 of GCN1 carries a binding site for ribosome, this could mean that GCN1 Sc and GCN1 Su may interact differentially with ribosomes. It is known that group 1 *S. pastorianus* relies exclusively on *S. eubayanus* ribosomal DNA [31,209,210]. This suggests that while GCN1 Sc could be able to bind to both ribosomes from both parents, GCN1 Su has evolved to bind specifically to only ribosome from *S. uvarum*. Similar to GCN1, GCN20 is also composed of two domains. The first domain is modelled with high prediction scores after the structure of EttA which is a regulatory translation factor in the ATP-binding cassette F (ABC-F) protein family [211,212]. It is possible that this domain may play a crucial role in fine-tuning the function of GCN1 [181]. In contrast to the first domain, the second domain on the N-terminus is poorly characterised as this region is highly disorder. However, it has a weak similarity to the highly conserved PWI tri-peptide (PWI) domain. PWI is a short domain known for interaction with a variety of polynucleotides [213–216]. Given that, PWI module can bind preferentially to proteins. And in this case domain 2 may interact with GCN20 binding site in GCN1. To summarise, there is no evidence to support that fitness differences between the two GCN20 are caused by structural variation.

3.5 Conclusion

In this study, a series of experiments were used to demonstrate the effect of chimeric protein complexes in yeast hybrids. The results show that chimeric GCN1-GCN20 have striking impacts on a wide range of phenotypes including amino acid starvation and cold stress response. For example, combining *GCN1* from *S. cerevisiae* and *GCN20* from *S. uvarum* appeared to make the hybrids more robust under stress conditions. This allelic combination is the one that is retained within group 1 *S. pastorianus*. In addition, it was found that variation in phenotypic traits could be caused by structural differences of the protein GCN1 and differential transcriptional regulation of the gene *GCN20*. Interestingly, this may explain biased retention of this protein complex in the natural hybrid *S. pastorianus*. Finally, different types of experiments such as domain swapping and site directed mutagenesis could be performed to support the finding of this study.

3.6 Supplementary materials

Name	Size (bp)	Sequence
GCN1Sc_fwd	20	GTCGCTGCATTTAAGCTTCC
GCN1Sc_rev	20	TTCACCCACGACACGAATTA
GCN1Su_fwd	20	TGGTTCCAAACACACCAAGA
GCN1Su_rev	20	TTGGCAACAATTGCGATAAA
GCN20Sc_fwd	20	ACCGGTGATACGTCCAAGAG
GCN20Sc_rev	20	CTCTTGGCGACTTTCTTTGC
GCN20Su_fwd	20	TCAACGGTGACATCGAACAT
GCN20Su_rev	20	CCTTTTCGTCGTCGATCAAT
ACT1Sc_fwd	20	TGTCACCAACTGGGACGATA
ACT1Sc_rev	20	GGCTTGGATGGAAACGTAGA
ACT1Su_fwd	20	CACGGTATCGTGACCAACTG
ACT1Su_rev	20	GGCTTGGATGGAAACGTAGA

Table S1. Primers for quantitative real-time PCR. They were designed by Primer3Plus based the sequences of *GCN1*, *GCN20* and *ACT1* of both Sc and Su alleles from Saccharomyces Genome Database.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Treatment	5	29902.932	5980.586	30.279	< 0.001	***
Temperature	1	117001.006	117001.006	592.363	< 0.001	***
GCN1	2	58947.794	29473.897	149.223	< 0.001	***
GCN20	1	3368.436	3368.436	17.054	< 0.001	***
Treatment:Temperature	5	162490.339	32498.068	164.534	< 0.001	***
Treatment:GCN1	10	17434.64	1743.464	8.827	< 0.001	***
Temperature:GCN1	2	14185.878	7092.939	35.911	< 0.001	***
Treatment:GCN20	5	4133.838	826.768	4.186	< 0.001	***
Temperature:GCN20	1	1127.34	1127.34	5.708	0.017	*
GCN1:GCN20	1	999.168	999.168	5.059	0.025	*
Treatment:Temperature:GCN1	10	20207.724	2020.772	10.231	< 0.001	***
Treatment:Temperature:GCN20	5	896.327	179.265	0.908	0.475	
Treatment:GCN1:GCN20	5	635.419	127.084	0.643	0.667	
Temperature:GCN1:GCN20	1	14.36	14.36	0.073	0.788	
Treatment:Temperature:GCN1:GCN20	5	654.44	130.888	0.663	0.652	
Residuals	876	173023.661	197.516	-	-	

Table S2. ANOVA result of the spot assay data. The analysis was performed to include interaction terms to examine relationship between treatments, temperature sand allele types of the two *GCN* genes.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Temperature	1	0	0	0.042	0.843	
Allele.type	1	0	0	0.424	0.533	
Temperature:Allele.type	1	0.012	0.012	19.84	0.002	**
Residuals	8	0.005	0.001	-	-	

Table S3. ANOVA result of the quantitative real-time PCR data of the gene *GCN1*. The analysis was performed to include interaction terms to examine relationship between temperatures and allele types.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Temperature	1	0.013	0.013	25.41	0.001	**
Allele.type	1	0.004	0.004	7.158	0.028	**
Temperature:Allele.type	1	0.011	0.011	21.607	0.002	**
Residuals	8	0.004	0.001	-	-	

Table S4. ANOVA result of the quantitative real-time PCR data of the gene *GCN20*. The analysis was performed to include interaction terms to examine relationship between temperatures and allele types.

Name	Domain	Rank	Template	Score	uGDT/GDT	uSeqID/SeqID	P-value
GCN1 Sc	1	1	2iw3A	445	275/37	113/15	8.50E-07
GCN1 Sc	2	1	2qmrA	382	202/31	78/12	6.60E-07
GCN1 Sc	2	2	5tqcA	384	200/31	63/10	6.20E-07
GCN1 Sc	2	3	2ie3A	370	198/31	65/10	1.10E-06
GCN1 Sc	2	4	2bkuB	382	198/31	61/9	6.50E-07
GCN1 Sc	2	5	4a0cA	376	193/30	44/7	8.50E-07
GCN1 Su	1	1	2iw3A	414	185/31	86/14	5.90E-08
GCN1 Su	2	1	2bkuB	352	124/22	59/11	4.40E-07
GCN1 Su	2	2	4xriA	347	124/22	64/12	5.40E-07
GCN1 Su	2	3	2ie3A	348	115/22	52/10	5.20E-07
GCN1 Su	2	4	2qmrA	359	116/21	58/11	3.20E-07
GCN1 Su	2	5	4a0cA	352	108/21	46/9	4.30E-07

Table S5. Protein structure prediction results of GCN1 Sc and GCN1 Su. score refers to the quality of sequence alignment. uGDT/GDT is the ratio between unnormalised and normalised global distance test score to measure model quality. uSeqID/SeqID is the ratio between the number of identical residues in the alignment and the same number normalised by the protein/domain sequence length. This ratio can indicate whether the predicted model has a correct fold. P-value refers to the relative model quality.

Name	Domain	Rank	Template	Score	uGDT/GDT	uSeqID/SeqID	P-value
GCN20 Sc	1	1	3j5sD	462	342/61	131/23	1.30E-10
GCN20 Sc	2	1	1mp1A	47	39/45	8/9	3.40E-02
GCN20 Sc	2	2	2qy9A	45	40/46	5/6	4.40E-02
GCN20 Sc	2	3	2qnaA	47	39/45	8/9	3.30E-02
GCN20 Sc	2	4	2yhsA	46	37/43	5/6	3.70E-02
GCN20 Sc	2	5	2bkuB	47	36/41	9/10	3.50E-02
GCN20 Su	1	1	3j5sD	463	340/61	134/24	8.20E-11
GCN20 Su	2	1	1mp1A	46	38/35	9/8	4.40E-07
GCN20 Su	2	2	4ydeA	43	38/35	6/6	5.40E-07
GCN20 Su	2	3	4qmhA	45	38/35	9/8	5.20E-07
GCN20 Su	2	4	3on4A	46	36/33	11/10	3.20E-07
GCN20 Su	2	5	3crjA	44	35/32	10/9	4.30E-07

Table S6. Protein structure prediction results of GCN20 Sc and GCN20 Su. score refers to the quality of sequence alignment. uGDT/GDT is the ratio between unnormalised and normalised global distance test score to measure model quality. uSeqID/SeqID is the ratio between the number of identical residues in the alignment and the same number normalised by the protein/domain sequence length. This ratio can indicate whether the predicted model has a correct fold. P-value refers to the relative model quality.

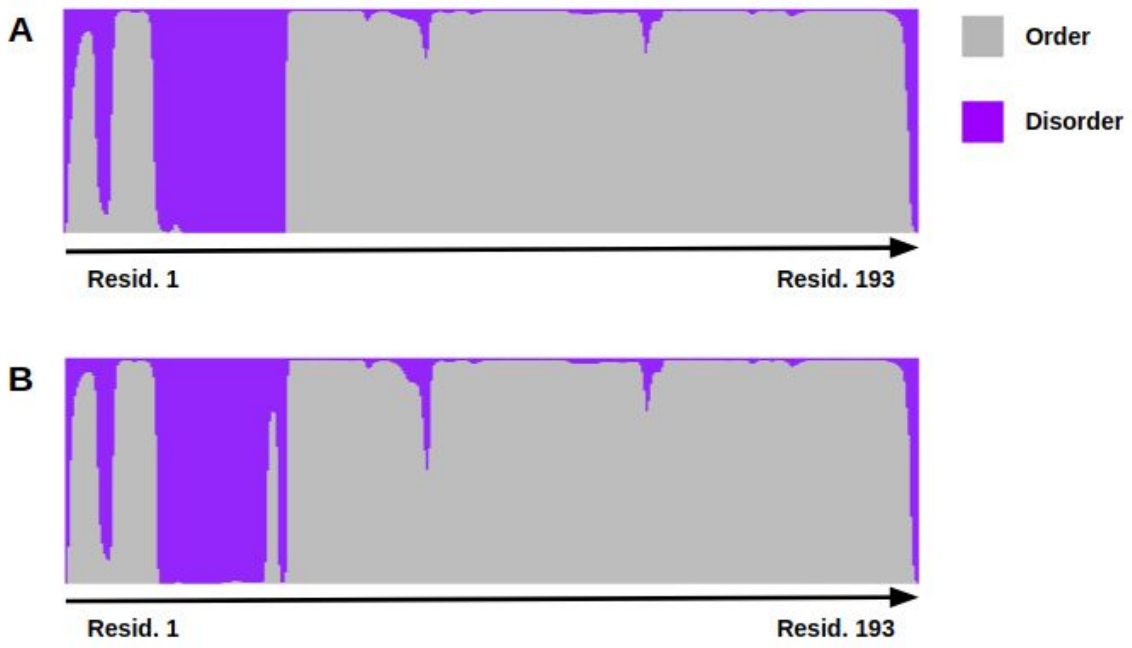


Figure S1. Predicted order/disorder regions of GCN20 Sc and GCN20 Su. RaptorX employs a powerful deep convolutional neural fields to predict protein structure properties. For a residue to be labelled as disordered, it must be in a segment that contains more than three residues with missing atomic coordinates in the X-ray structure.

Impacts of mitotypes on global gene expression in yeast hybrids

Kobchai Dungrattanalert, Sarah K Hewitt and Daniela Delneri

Faculty of Biology, Medicine and Health,

University of Manchester, Manchester, United Kingdom, M1 7DN

Contributions

Kobchai Dungrattanalert:	Main investigator responsible for RNA sequencing experiment and transcriptome analysis.
Sarah K Hewitt	Main investigator responsible for constructing hybrid stains and profiling their phenotypes.
Daniela Delneri:	Main supervisor.

Abstract

In eukaryotes, mitochondria take important roles in aerobic respiration, apoptosis and cellular metabolism of several metabolites such as fatty acids, amino acids and heme. Mitochondrial DNA is inherited since it cannot be synthesised anew. In yeast hybrids, it has been observed that mitochondrial inheritance is uniparental that only the mitochondrial DNA of one parent is maintained. This is perhaps to avoid incompatibility which reduces overall fitness. The mitochondria is composed mainly of nuclear-encoded proteins, and there are only 8 mitochondrial genes. Here the impact of different types of mitochondria on the transcriptome was investigated in yeast hybrids carrying different types of mitochondria. S. cerevisiae strain BY4741 and S. uvarum strain NCYC2669 were crossed and those with a single mitotype (either S. cerevisiae or S. uvarum mitochondria) were selected. The hybrids were cultured in rich YPD medium or YP medium with glycerol and at two different temperatures (28°C and 16°C). Their total RNAs were extracted for transcriptome analysis by sequencing on Illumina platform. The results show that mitotype has indeed significant impacts on several aspects of the ScxSu hybrids depending on the types of mitochondria. In addition to effects on mitochondrial retention and mitotic fitness, mitotype appear to cause allelic expression bias in various biological pathways. This means that hybrid cells work preferentially with one set of parental alleles or the other according to different functions. This work has shed light on the mechanism that explains why favouring one parental mtDNA over the other may help an organism adapt to different environments and how this phenomenon may drive nuclear retention of specific alleles in yeast hybrids.

4.1 Background

The mitochondrion is a membrane-bound organelle in eukaryotic cells that evolved around two billion years ago from α-proteobacteria [45,46]. Since that time, this organelle has lost or transferred a large portion of genetic materials to nucleus. In yeast, what remains in the genome is a set of 8 protein encoding genes [47]. In general, these genes are translated into the core subunits of respiratory chain complexes that are indispensable for growth under aerobic conditions. For instance, it has been shown that mutations of some mitochondrial genes are linked to age-related disorders such as Progeria, dementia and Alzheimer's disease [48–50]. Mitochondria are likely to accumulate deleterious mutations [51]. Without correction, these mutations will lead to the formation of reactive oxygen species that aggressively damage cellular components and result in more defects within both nuclear and mitochondrial genomes. The function of this organelle relies on hundreds of nuclear-encoded genes [52–55]. In yeast species, 700 nuclear genes are found to encode for mitochondrial proteins are reported on *Saccharomyces* Genome Database. This suggests communication i.e. (anterograde and retrograde signaling pathways) between the two compartments that relay biological information in response to stressful environments.

Mitochondrial DNA (mtDNA) must be inherited vertically from one of the parents [44,57,58]. In yeast, two haploid parental cell fuse and give rise to a hybrid diploid cell that carries two parental genomes and initially both versions of mtDNA. However, despite fusion of most membrane-bound organelles, very rapidly only one single type of mitochondria is retained while the other is lost and cells become homoplasmic. This phenomenon is known as uniparental mitochondrial inheritance. This process is particularly common in yeast species, especially in the *Saccharomyces sensu stricto* complex. This is the case as several members of this group frequently mates both in the wild and the industrial environments. *S. pastorianus* is one of the natural hybrids used extensively in the production of the lager beer [22]. Based on several genomic studies, this hybrid is derived from hybridisation events between *S. cerevisiae* (Sc) and *S. uvarum* (Su)-like species

which is now known to be *S. eubayanus* with mitochondria passed down from the cold-adapted parent [26,29,30]

Several studies have demonstrated that mitochondrial inheritance is influenced by environmental conditions as well as the effect of mitochondrial DNA on fitness/transcription of nuclear-encoded genes in yeast hybrids [62–64]. For example, carbon source is known to influence the transmission of mitochondria [217–219]. Temperature tolerance is one of the key adaptations that determine evolution within the *Saccharomyces sensu stricto* group [220]. While *S . cerevisiae* grows optimally at the highest optimum at 32°C, *S. uvarum* grows optimally between 24 and 27°C. Given this information, temperature and possibly other environmental factors may play an important role in mitochondrial selection, and this choice subsequently affect the fitness of the organism.

In this study, the effect of mitotype on fitness and global gene expression in yeast hybrids was investigated. It was found that mtDNA is associated with fitness differences and nuclear allelic expression in specific biological pathways depending on carbon source and temperature. This work explains why favouring one parental mtDNA over the other may help an organism adapt to different environments and how this phenomenon may drive nuclear retention of specific nuclear alleles in yeast hybrids.

4.2 Materials and methods

4.2.1 Strains and culture conditions

S. cerevisiae strain BY4741 and *S. uvarum* strain NCYC2669 were used in this study as parental strains and obtained originally from Thermo Scientific UK and from the National Collection of Yeast Cultures, respectively. In addition to background mutations in the parent BY4741, Delneri lab has created a variant that carries a KanMX cassette at the neutral *AAD3* locus to facilitate selection in the presence of kanamycin. Sc x Su hybrids were successfully created by Dr. Sarah Katherine Hewitt by crossing the parents and sequentially selected for individuals that carry a single type of mitochondria by restriction fragment length polymorphism based on the

restriction pattern of the mitochondrial genes *COX2* and *COX3*. Identity of the mitochondria were checked again by Sanger sequencing using primers that bind to unique regions within the Sc and Su mitochondrial genomes (see Table S1). Primers were designed by Primer3Plus [185] to amplify around 400-500 bp. Hybrid that carries mtDNA from *S. cerevisiae* is named HMtSc, and hybrid that carries mtDNA from *S. uvarum* is named HMtSu. Both parents and the hybrids were maintained and cultured in either YP media (2% (w/v) Bacto Peptone and 1% (w/v) Bacto yeast extract) with 2% (w/v) glucose or 2% (v/v) glycerol with additional 2% (w/v) agar for solid media. 2% (w/v) agar was added to solidify the YP broth. All cultures were incubated with shaking at 200 rpm until they reach OD600 between 0.0 and 0.6 (around mid log phase).

4.2.2 Serial-dilution spot assays

To determine the effect of mtDNA, three biological replicates of ScxSu carrying different mitochondrial DNA were created and used for fitness screen. In brief, cells were grown overnight in YPD at 30 ºC with shaking at 200 rpm and washed twice in distilled water. They were suspended to make cell solutions at OD600 around 0.5 and diluted five times with the dilution factor of 1/10. It was followed by spotted under four combinations of medium (complete medium (YPD) and YP + 2% v/v glycerol) and temperatures (28°C, 16°C, 10°C and 4°C) for 24 hours or until growth became visually stable. At the end, images were taken to capture the state of cell growth.

4.2.3 Sample preparation and RNA sequencing

First, cells were grown overnight and inoculated in 50ml of fresh YPD and YP + glycerol at 28°C and 16°C with shaking at 200rpm to reach mid log phase (OD600 between 0.5 and 0.6). At this stage, cells replicate at high rates and contain significantly large amounts of RNA and other biochemical materials. RNA samples were isolated using Trizol (Ambion, Life Technologies, USA) by following instructions provided by the manufacturer. In summary, cells were collected by snap freeze in liquid nitrogen, ground to fine powder and suspended in Trizol to break open cell wall. This step was followed by centrifugation at 12,000 x g for 10 minutes to discard

cell debris. Supernatants were collected and combined with chloroform (200µl to every 1ml volume) to obtain RNA in the upper aqueous phase after centrifugation at 12,000 x g for 5 minutes. The upper phase was transferred and mixed with an equal volume of isopropanol to precipitate RNA by centrifugation at 12,000 x g for 10 minutes after 10-minute incubation at room temperature. To further improve quality of the RNA, pellets were collected and washed twice in 1ml (DEPC)-treated 70% ethanol and air dried before mixing with 50% (v/v) lithium chloride (Ambion, Life Technologies, USA) and stored overnight at -20°C. RNA was then isolated by additional wash in 70% ethanol before adding 50µl DEPC-H2O to dissolve the samples. Integrity of the RNA samples was checked on 1.5% agarose gel. Those with sharp 18C and 28S without DNA contamination or signs of degradation were proceeded to the next step. The quantity and quality of RNA was determined using a Nanodrop 1000 spectrophotometer (Thermo Scientific, USA). The extracted samples were sequenced on an Illumina Hiseq 2500 (Illumina, USA) platform in the Genomic Technologies Core Facility at The University of Manchester, UK. The experiment was carried out in triplicate for both parents and the hybrids HMtSc and HMtSu.

4.2.4 Read assembly and annotation

S. cerevisiae (UCSC SacCer3) and *S. uvarum* (Scannell DR et al. 2011) [221] were combined to create a reference genome. Reads were mapped and counted using STAR mapper and HTseq, respectively [222,223]. To avoid misidentification of individual alleles, reads that were common between the two sub-genomes were excluded. The total of 5,339 genes were recovered across the tested strains and conditions.

4.2.5 Differential expression and co-expression analysis

After mapping, reads were counted and analysed by DESeq2 [141]. To identify significant changes in expression of the orthologues, the standard differential expression analysis based on the negative binomial distribution with size-factor/dispersion estimation for normalisation was run at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.05. Differential co-expression analyses was analysed

based on Pearson correlation of individual alleles at the cutoff value of r ≥ 0.75 [224]. The network was constructed and clusters of the alleles were identified to show those similar expression patterns across the four testing conditions and the two hybrids.

4.2.6 Functional enrichment analysis

To identify relevant biological functions, enrichment analysis was performed using the R package KEGGprofile. This tool maps gene lists to available pathways in KEGG database [225]. Annotation terms with number of genes ≥ 5 and adjusted p-value ≤ 0.05 were selected.

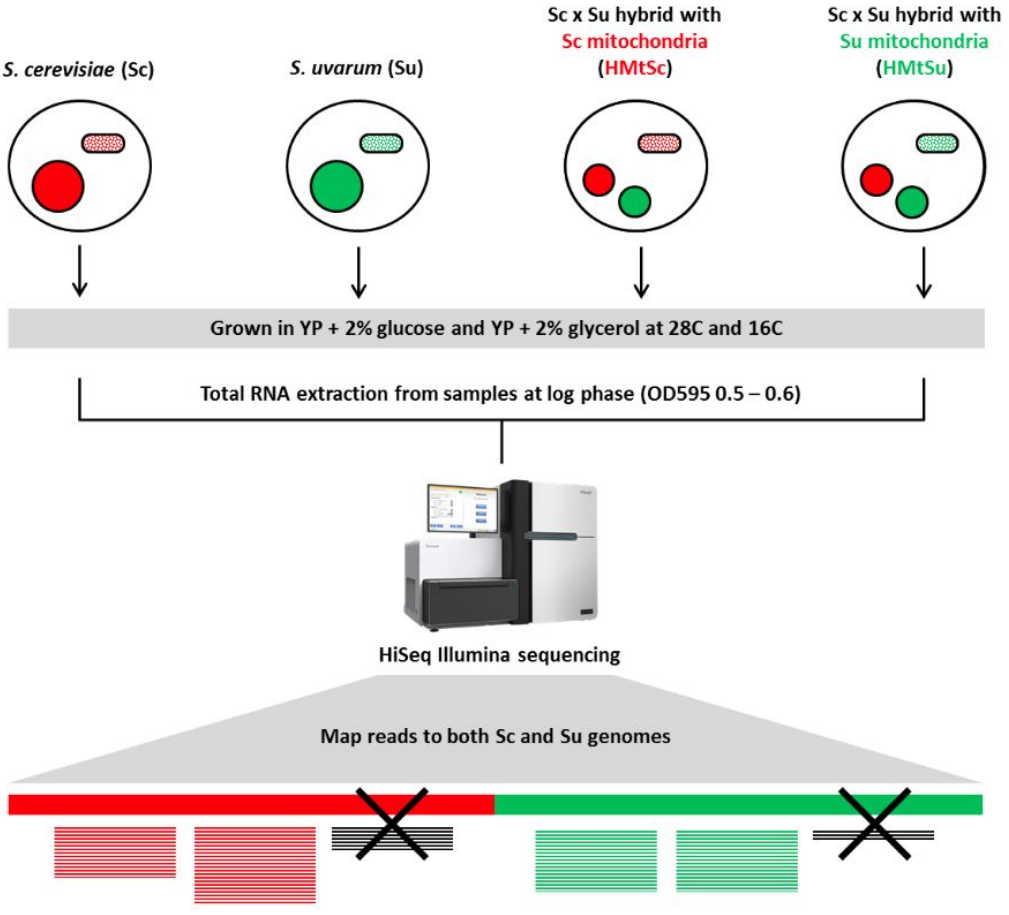


Figure 1. Schematic diagram of sample preparation and sequencing strategy. Cells were grown in YPD and YP - glycerol at 28°C and 16°C to reach mid-log phase (OD600 between 0.50 and 0.60). RNA samples were extracted and sequenced on the Illumina platform.

4.3 Results

4.3.1 Experimental strategy

S. cerevisiae strain BY4741 and *S. uvarum* strain NCYC2669 were crossed to create six independent lineages of yeast hybrids. Those with a single type of mitochondrial DNA were selected and analysed using serial-dilution spot assays and RNA sequencing on the Illumina HiSeq platform. Spot assays were carried out in eight combinations of medium (YPD and YP + glycerol) and temperatures (28°C, 16°C, 10°C and 4°C) for 24 hours or until growth became visually stable. These hybrids, along with the parental strains, were grown in the same media but only at 28°C and 16°C. To analyse gene expression, a strategy was designed to deal with both sub-genomes in the hybrid strains. In short, *S. cerevisiae* (UCSC SacCer3) and *S. uvarum* (Scannell DR et al. 2011) were combined to create a reference genome. Reads were mapped and counted using STAR mapper and HTseq, respectively. To avoid misidentification of individual alleles, reads that were common between the two sub-genomes were excluded (see Figure 1). Reads from the parental strains were also counted using the same workflow to ensure that subsequent analyses were not influenced the absence of certain reads.

4.3.2 Phenotypes of hybrids carrying different mtDNA

To understand the influence of environmental conditions on overall fitness, hybrids and the parents were analysed by serial-dilution spot assays. It was found that hybrids with Sc mtDNA grew better than hybrids with Su mtDNA at warm (28°C), whilst it turned out to be vice versa at cold (4°C) especially in glycerol medium (see Figure 2). It is clear that growth of different hybrid strains on a respiratory carbon source was generally more striking. The results also show that mtDNA may not affect the fitness of both hybrids, but the parents, at 10°C and 16°C. This indicates that nuclear genome has stronger effects than mitochondrial genome at these temperatures. All in all, there is a relationship between genetic background and overall fitness under the tested conditions.

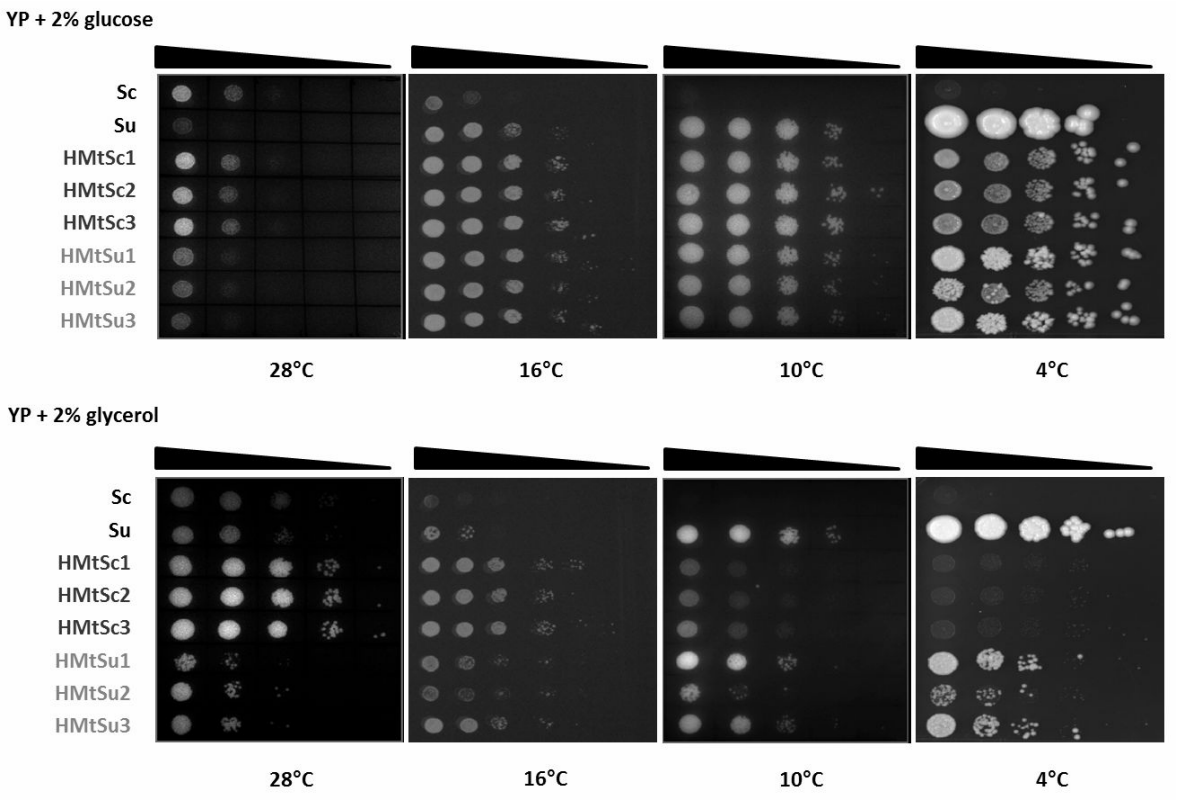


Figure 2. Serial-dilution spot assays of the hybrids carrying different types of mitochondria on fermentative YPD or respiratory YP + 2% v/v glycerol at a range of temperatures from 4°C to 28°C.

4.3.3 Plasticity in allele-specific expression in yeast hybrids

5,339 genes which is around 90% of the total yeast genome were counted with reliable and clear orthologous relationship between the two parents across the tested strains and conditions. The reads in each sample ranges between 2.8 millions and 22.7 millions. To validate this approach, principle component analysis (PCA) was carried out based on the expression data. Figure 3 demonstrated that media and allele type appeared to affect differential gene expression as shown by separation on the first component and the second component respectively. Temperatures appeared to be the weakest contributor after genetic background as it is nearly impossible to distinguish them on the PCA map.

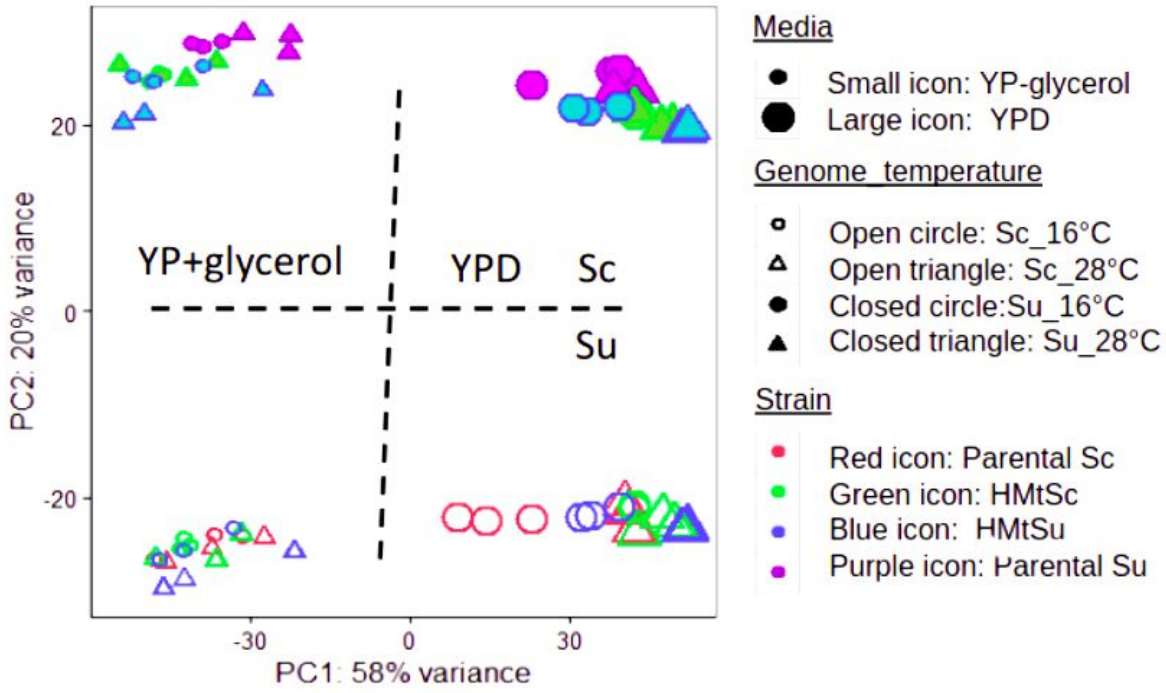


Figure 3. PCA projection of the expression data from RNAseq of the ScxSu hybrids and their parents grown in fresh YPD and YP + glycerol at 28°C and 16°C. Reads were then mapped and counted using STAR mapper and HTseq. Counts were normalised by DESeq2 before passing then to PCA analysis.

To determine the effect of genetic background (i. e. hybrids vs homozygous parent) on differential gene expression, expression levels of *S. cerevisiae* or *S. uvarum* alleles in the hybrids were compared against expression profiles in the respective parents across the four tested growth conditions (see Figure 4). It was found that differentially expressed genes between parental strains and the hybrids were influenced greatly by mitotypes and growth conditions. For example, differential expression between the parent Sc and alleles in the Sc subgenome of the hybrid HMtSc was overall greater than at cold. However, these expression changes at cold are mitigated in the presence of Su mtDNA. Nonetheless, higher levels of differential

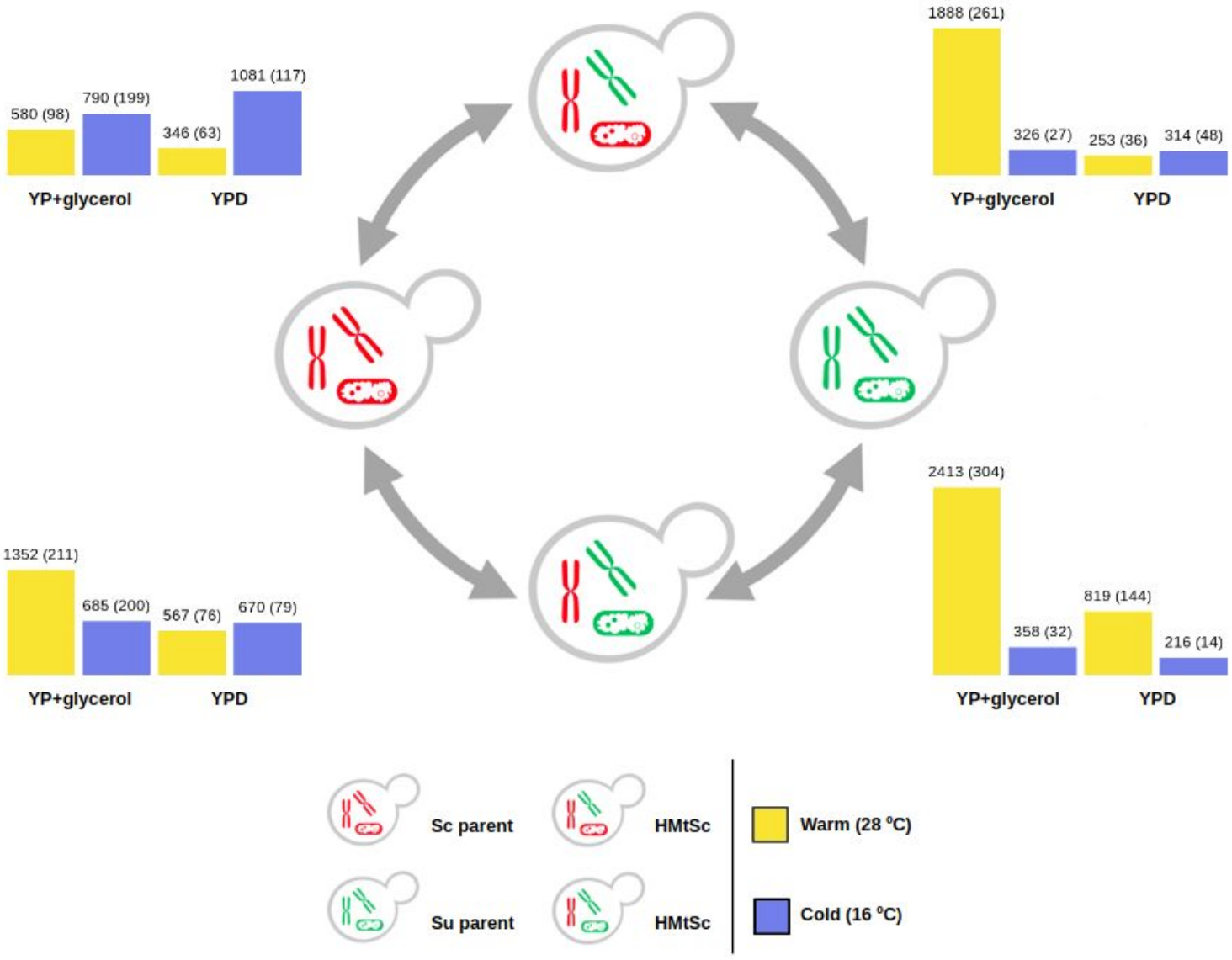

580 (98)
790 (199)
346 (63)
1081 (117)
YP+glycerol
YPD
1888 (261)
326 (27)
253 (36)
314 (48)
YP+glycerol
YPD
1352 (211)
685 (200)
567 (76)
670 (79)
YP+glycerol
YPD
2413 (304)
358 (32)
819 (144)
216 (14)
YP+glycerol
YPD
Sc parent
Su parent
HMtSc
HMtSc
Warm (28 °C)
Cold (16 °C)

Figure 4. Differential gene expression between yeast hybrids and the parents. Expression data were analysed using DESeq2 and cutoffs were set at absolute log2 fold change ≥ 2.00 and p-value ≤ 0.05 The analyses were done repeatedly for the four culture conditions and for each pair of yeast hybrids and the parents. Mitochondrial-associated genes that are shown in parentheses were derived from Saccharomyces Genome Database.

gene expression were observed when compared Su parent with Su alleles in the hybrid HMtSc under respiratory condition at warm. However, these changes were aggravated in the presence of *S. cerevisiae* mitochondria. What's more, the results show that expression of certain alleles is more affected at warm temperature in the presence of mitochondria from *S. uvarum*. In addition, the study revealed that Sc subgenome was upset by cold temperature as more than 20% of the differentially expressed genes are associated with mitochondria regardless of the type of mtDNA (Exact binomial test: p-value < 0.001; Table S2).

4.3.4 Gene co-expression in yeast hybrids

To further exploit the data, a co-expression network was constructed based on Pearson correlation coefficients of the absolute log2 fold change values. Note that, the values were derived from orthologs that showed significant differential expression in the hybrids and in the four combinations of medium (YPD and YP + glycerol) and temperatures (28°C and 16°C). Correlation was cut-off at 0.94 to keep ortholog pairs with strong connection and to not break the network structure into several isolated communities (see Figure S1). Clustering by random walktrap [226] were performed to identify alleles with similar expression profiles. As shown in Figure 5, the network has 1,915 nodes and 3,0627 edges with its topology very close to scale-free.

As shown in Figure 5, the network is composed of six main clusters (each contains around 100-300 genes). Interestingly, these clusters are not made of the same proportion of Sc and Su alleles (Chi-squared = 907.87, df = 6, p-value < 0.001; see Figure S2). While cluster 1, 2 and 5 are clearly biased towards the Sc subgenome, cluster 3, 4 and 6 apparently contain more Su. As for cluster 0, it has about the same

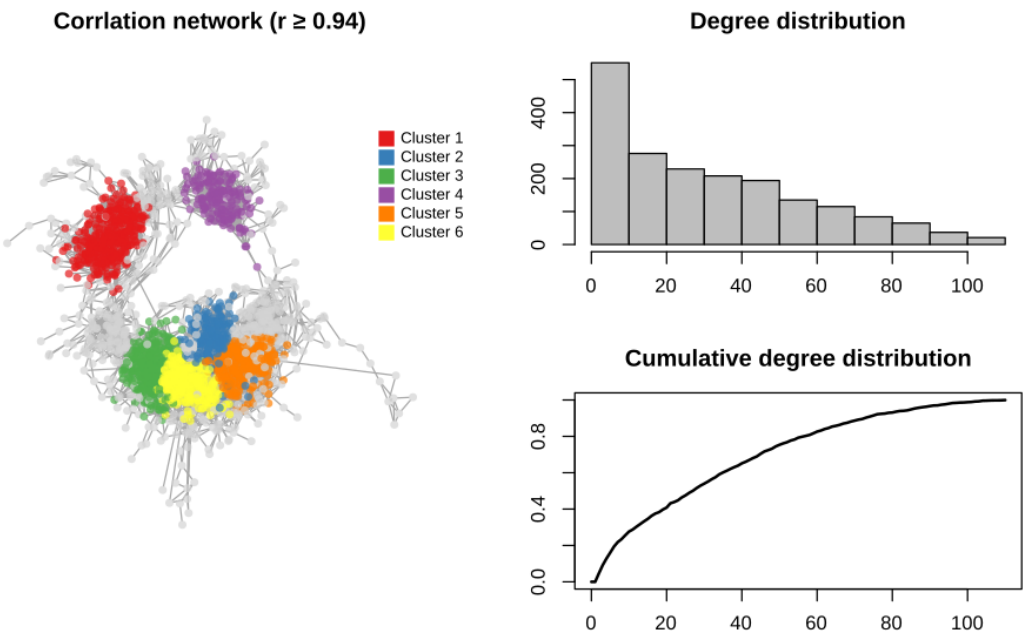


Figure 5. Co-expression network of orthologs with significant differential expression in the hybrids and in the four combinations of medium (YPD and YP + glycerol) and temperatures (28°C and 16°C). It was constructed based on Pearson correlation coefficients of the absolute log2 fold change values with the cutoff threshold of 0.94. Clustering by random walktrap were then performed to identify alleles with similar expression profiles. Degree values were computed and shown in top right and bottom right panels.

numbers of Sc and Su. This cluster is generally a collection of orthologs with ambiguous expression patterns and may not be biologically important.

To associate each cluster with relevant biological functions, enrichment analysis was performed to map gene lists to available pathways in KEGG database [225]. The results were put together and shown in Table 1. Clusters 1 and 4 are enriched with genes for ribosome biogenesis. In addition, cluster 1 also contains orthologs for and other pathways related to amino acid and nucleic acid metabolism. Based on expression profiles in Figure S3, orthologs in cluster 1 which are mostly Sc were upregulated in complete medium at 28°C in both HMtSc and HMtSu. The expression pattern is, however, different in cluster 4 as orthologs in this cluster were downregulated in glycerol medium at 16°C in the same genetic backgrounds. As for

Cluster	Pathway	P-value	
Cluster 1	Ribosome biogenesis in eukaryotes	< 0.001	***
Cluster 1	Metabolic pathways	< 0.001	***
Cluster 1	Purine metabolism	< 0.001	***
Cluster 1	Arginine and proline metabolism	0.003	**
Cluster 1	Pyrimidine metabolism	< 0.001	***
Cluster 1	Lysine biosynthesis	< 0.001	***
Cluster 1	Lysine degradation	< 0.001	***
Cluster 1	RNA polymerase	< 0.001	***
Cluster 1	Histidine metabolism	< 0.001	***
Cluster 3	Oxidative phosphorylation	< 0.001	***
Cluster 3	Metabolic pathways	< 0.001	***
Cluster 4	Ribosome biogenesis in eukaryotes	< 0.001	***
Cluster 4	Cell cycle - yeast	< 0.001	***
Cluster 4	Meiosis - yeast	< 0.001	***

Cluster	Pathway	P-value	
Cluster 4	RNA transport	0.003	**
Cluster 4	MAPK signaling pathway - yeast	0.003	**
Cluster 4	DNA replication	< 0.001	***
Cluster 5	Oxidative phosphorylation	< 0.001	***
Cluster 5	Metabolic pathways	< 0.001	***
Cluster 5	Biosynthesis of secondary metabolites	< 0.001	***
Cluster 5	Peroxisome	< 0.001	***
Cluster 5	Citrate cycle (TCA cycle)	< 0.001	***
Cluster 6	Oxidative phosphorylation	0.003	**
Cluster 6	Metabolic pathways	< 0.001	***
Cluster 6	Glycolysis / Gluconeogenesis	0.002	**
Cluster 6	Biosynthesis of secondary metabolites	< 0.001	***
Cluster 6	Citrate cycle (TCA cycle)	< 0.001	***

Table 1. Functional annotation of cluster 1-6. The analysis was performed by mapping gene lists to available pathways in KEGG database using the R packages KEGGprofile. Annotation terms with number of genes ≥ 5 and adjusted p-value ≤ 0.05 were selected.

cluster 5 and 6, they are enriched with genes for glycolysis, TCA cycle and electron transport chain. This means that these clusters could be involved in aerobic respiration. While cluster 5 which contains more than 80% Sc was upregulated in glycerol medium at 28°C, orthologs in cluster 6 which contains about 90% Su were expressed strongly in the same medium but at cold temperature. Even though cluster 5 and 6 are involved in cellular respiration, the expression of some orthologs in these clusters was suppressed in complete medium at 28°C for HMtSc and at 16°C for HMtSu. It could be that these are the conditions where hybrids with Sc mitotype and hybrids with Su mitotype switch to fermentation, respectively.

4.4 Discussion

To explain the effect of mitotype, *S. cerevisiae* and *S. uvarum* were crossed to create hybrids that carry a single type of mtDNA. Based on the fitness data, it is clear that mitotype has a strong effect on the fitness of yeast hybrids especially under respiratory conditions. This generally agrees with other studies that ScxSu with mtDNA from *S. cerevisiae* are likely to do better under certain stress conditions [227]. Other studies have suggested the role of this mitochondria in cell growth, sugar uptake and drug tolerance during fermentation [228,229]. All in all, it is reasonable to say that it is advantageous to retain mtDNA from the Sc parent to grow more effectively in respiratory conditions at warm. On the other hand, non-*S. cerevisiae* mitochondrial DNA have tendency to ferment better in industrial environments at lower temperatures. This is true especially for the natural hybrid *S. pastorianus* which carries mtDNA from the cold adapted parent *S. eubayanus* [22,27]. As shown in Figure 2, hybrids with Su mtDNA grew at much faster rates than

hybrids with Sc mtDNA. Given that, hybridisation may also play an important role as both HMtSc and HMtSu had higher fitness than the parents at 10°C and 16°C.

RNA sequencing was carried out to examine the impact of mitotype on the transcription of individual orthologs in yeast hybrids. It was shown that Sc subgenome of the hybrids was more differentially expressed at warm when compared to the parents, and vice versa for the Su counterpart. What interesting is that both Sc and Su alleles displayed a high level of expression plasticity, implying that allelic expression could be controlled by different trans-acting elements from each subgenome in the hybrid backgrounds [230–232]. This means that the expression can be regulated by environmental factors. As shown in the PCA map, media appeared to be the strongest contributor followed by subgenomes, temperatures and mitotypes. Alternatively, there is a possibility that gene expression differences between hybrids and the parents were caused by differences in gene dosage from the Sc or Su subgenome.

The co-expression network was constructed to study interaction between orthologs in the hybrid backgrounds. Based on degree distribution, the network is very close to scale-free. This network topology occurs widely in natural systems such as gene regulatory network, metabolic network and protein-protein interaction network [233–237]. This suggests that the network is organised into dense clusters with a small number of hubs or central controllers that may control the expression of the orthologs. In total, there are six main clusters in the co-expression map - each has a unique set of properties including the proportion of Sc and Su alleles, expression profile across the two hybrids and the four tested conditions and finally the list of associated biological functions. As mentioned previously that cluster 1 and 4 may function in ribosome genesis, orthologs in these clusters were expressed nearly identically in both hybrids. It is possible that this could be the result of hybridisation. A shift in transcriptional regulation can occur widely in plants and animals [238–241]. Likewise, the expression of respiration machineries in cluster 5 and 6 were consistent in both hybrids at least in the glycerol medium. However, some orthologs in these two clusters behaved rather differently in YPD. As shown in Figure S3, respiration seemed to be suppressed in complete medium at 28°C in the HMtSc

background and at 16°C in the HMtSu background. Together with the expression of orthologs for amino acid metabolism in cluster 1, the evidence points out that mitochondria in the hybrids could be functionally compromised such that cells had to inevitably switch to fermentation [242–244]. Potentially, this could be explained by genetic variation of cis-regulatory elements and mitochondrial genome of the two related yeast species [245,246]. To summarise, assume that this regulatory variation is essential for optimal growth within their specific environments, combining the two genomes may increase a repertoire of functional abilities towards a wider range of stresses.

4.5 Conclusion

In this study, the effect of mitotypes was demonstrated on several aspects of yeast hybrids. Evidently, mitotic fitness is affected heavily by the types of mitochondrial DNA. The results show that mtDNA from *S. cerevisiae* is more favourable at warm and vice versa for mtDNA from *S. uvarum*. To explain this phenomenon, RNA sequencing was performed to inspect global expression of nuclear genes in different genetic backgrounds. Specifically, differential expression analysis and co-expression analysis were run in an attempt to explain relationship between fitness and expression patterns. Most importantly, mitotypes appear to have a direct impact on allelic expression bias in several biological pathways. For example, the expression of orthologs for respiration from Sc and Su subgenomes was suppressed in complete medium at 28°C for HMtSc and at 16°C for HMtSu, respectively. Potentially, this could be explained by genetic variation of cis-regulatory elements and mitochondrial genome of the two related yeast species. Finally, more experiments could be performed to validate the findings of the study. For instance, one could do a high throughput screening for mitochondrial and nuclear proteins to see whether it is consistent with the results here.

4.6 Supplementary materials

Name	Size (bp)	Sequence
ATP6_fwd	22	GCTTAAAGGACAAATTGGAGGT
ATP6_rev	20	AGCCCAGACATATCCCTGAA
COX1_fwd	20	GGTATGGCAGGAACAGCAAT
COX1_rec	20	TTTTTCACACTGCCTGTGCT
COX2_fwd	20	TCAGGATTCAGCAACACCAA
COX2_rec	20	CATCAGCAGCTGTTACAACGA
COX3_fwd	20	TCATGCCTTAATCGCAGGTA
COX3_rec	20	CCCGTTAGGAGTTTTGAAGG
COB_fwd	20	ATTATCCACGGGACCAATGA
COB_rev	20	TTGATGATGGTTGTGGTGAA

Table S1. Primers for Sanger sequencing. The primers were designed by Primer3Plus based the sequences of *ATP6, COX1, COX2, COX3* and *COB*.

Medium	Temperature	Mitotype	Subgenome	Number of differentially expressed genes	Proportion of mitochondrial-associated genes (%)	P-value	
YPD	28 °C	Sc	Sc	346	18.21	< 0.001	***
YPD	28 °C	Su	Sc	567	13.4	0.033	*
YP + glycerol	28 °C	Sc	Sc	580	16.9	< 0.001	***
YP + glycerol	28 °C	Su	Sc	1352	15.61	< 0.001	***
YPD	16 °C	Sc	Sc	1081	10.82	0.016	*
YPD	16 °C	Su	Sc	670	11.79	0.046	*
YP + glycerol	16 °C	Sc	Sc	790	25.19	< 0.001	***
YP + glycerol	16 °C	Su	Sc	685	29.2	< 0.001	***
YPD	28 °C	Sc	Su	253	14.23	0.044	*
YPD	28 °C	Su	Su	819	17.58	< 0.001	***
YP + glycerol	28 °C	Sc	Su	1888	13.82	< 0.001	***
YP + glycerol	28 °C	Su	Su	2413	12.6	0.019	*
YPD	16 °C	Sc	Su	314	15.29	0.016	*
YPD	16 °C	Su	Su	216	6.48	< 0.001	***
YP + glycerol	16 °C	Sc	Su	326	8.28	< 0.001	***
YP + glycerol	16 °C	Su	Su	358	8.94	0.011	*

Table S2. Exact binomial tests for the proportion between mitochondrial and non-mitochondrial genes with significant differential expression in yeast hybrids.

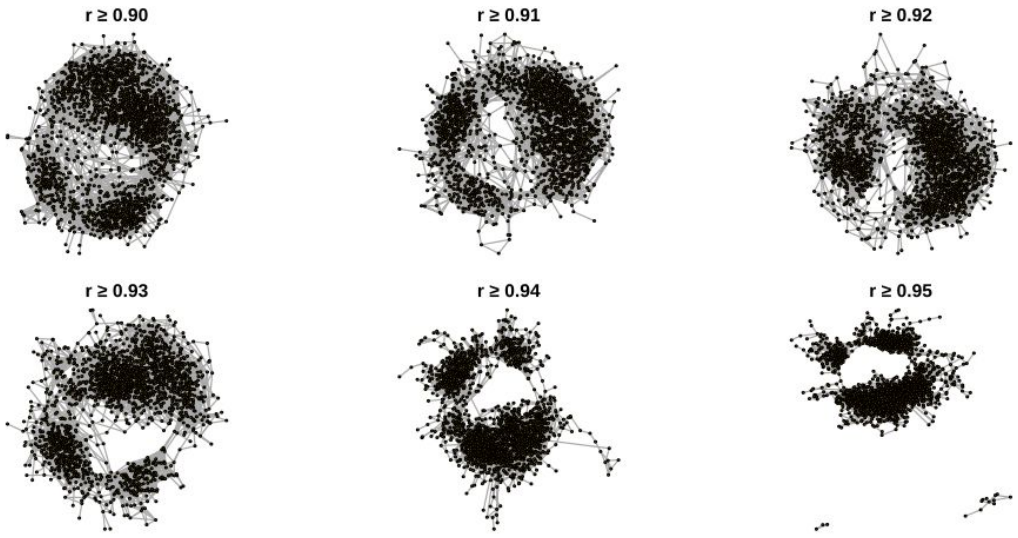


Figure S1. Co-expression networks of orthologs with significant differential expression in the hybrids and in the four tested conditions. It was constructed based on Pearson correlation coefficients of the absolute log2 fold change values with the cutoff threshold between 0.90 and 0.95.

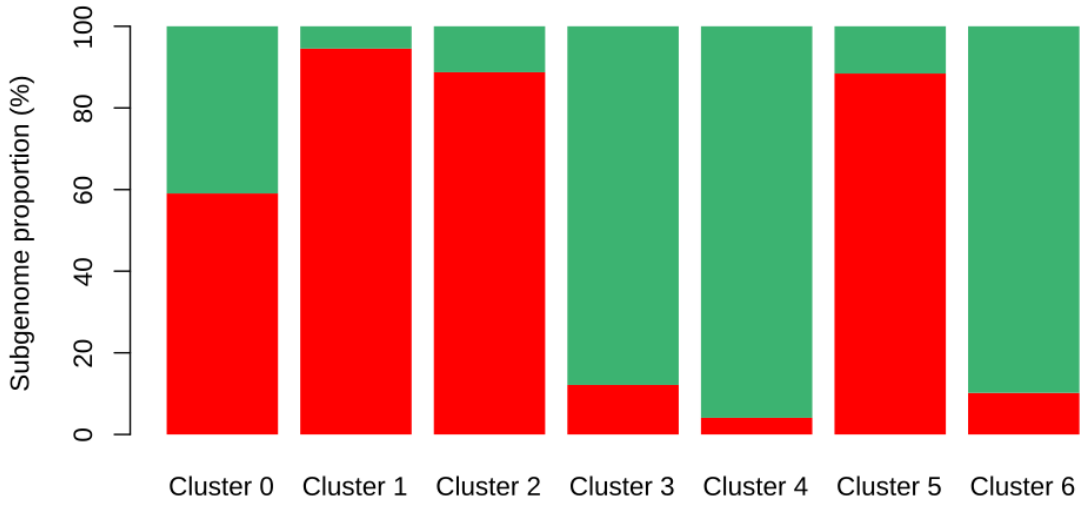


Figure S2. Proportions of Sc and Su alleles in cluster 1-6. This set of numbers was derived from the counts of allele types in the clusters of the co-expression network that was created at the cutoff threshold of 0.94. Red color represent Sc alleles and blue color represent Su alleles.

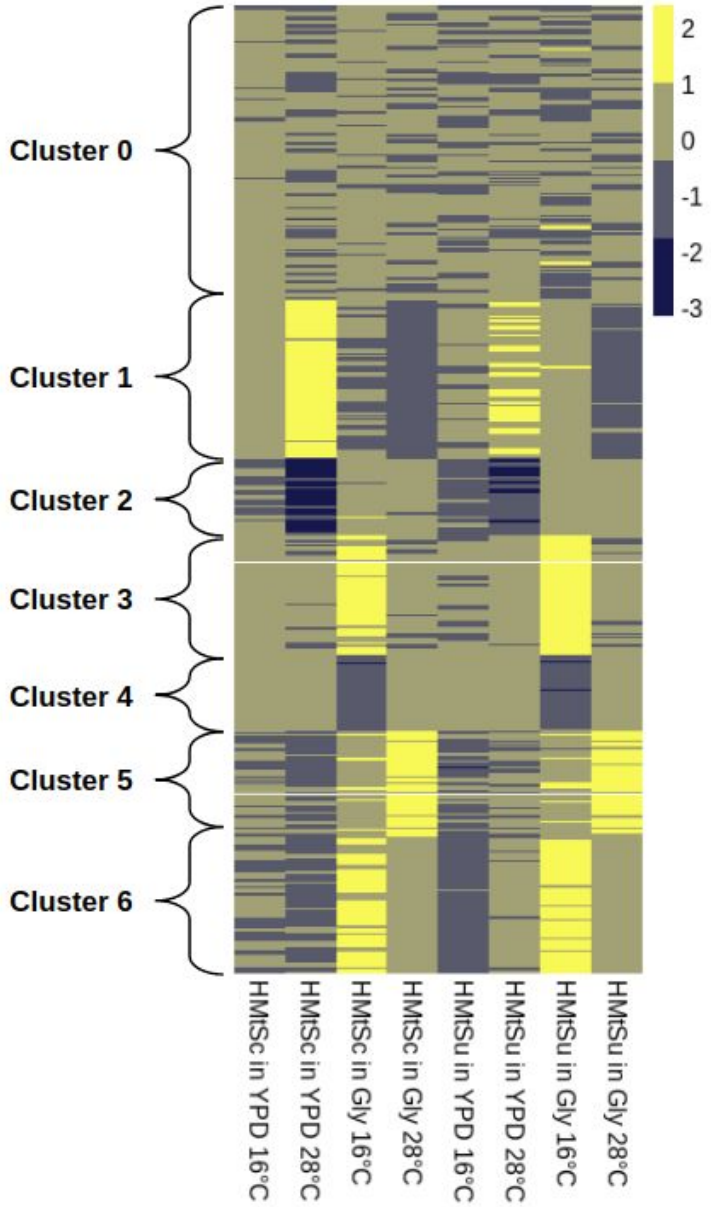


Figure S3. Expression profiles of cluster 1-6. This set of numbers was derived from absolute log2 fold change values (hybrid vs, parent) of individual orthologs in the clusters of the co-expression network that was created at the cutoff threshold of 0.94. Yellow and navy blue represent high and low relative expression levels, respectively.

gcatR: An all-round tool for growth curve data analysis

Kobchai Dungrattanalert and Daniela Delneri

Faculty of Biology, Medicine and Health,

University of Manchester, Manchester, United Kingdom, M1 7DN

Contributions

Kobchai Dungrattanalert:	Main developer of the software
Daniela Delneri:	Main supervisor.

Abstract

A shiny application gcatR (**g**rowth **c**urve **a**nalysis **t**ool based on **R**) was developed with a graphical user interface for interactive analysis of growth curves and fitness data. This quantitative user-friendly tool is able to analyse different aspects of phenotypic data including a complete analysis workflow routine. It is accessible by scientists who have no computational knowledge and programming skills, and it is implemented with several key features containing interactive data table, growth curve parameter estimation, time series clustering and classification, and fitting based on generalised additive models. gcatR allows users to conveniently analyse and visualise phenotypic data and the relevant figures can be created and customised to report the results. The application and source code are available online at http://www.gcatr.manchester.ac.uk/.

5.1 Background

Quantitative phenotyping and growth curve assays are essential tools for biological studies, since they allow researchers to explore and comprehensively understand complex relationship between environments and cell growth. For examples, growth curves can be used to assess the effects of pH, temperatures and other environmental factors on the cell fitness [247–250]. In response to demand for high throughput experiments, automation has been developed and has become the crucial part of modern research, to advance knowledge at the system level by simultaneously testing hundreds and thousands of samples all at the same time [251]

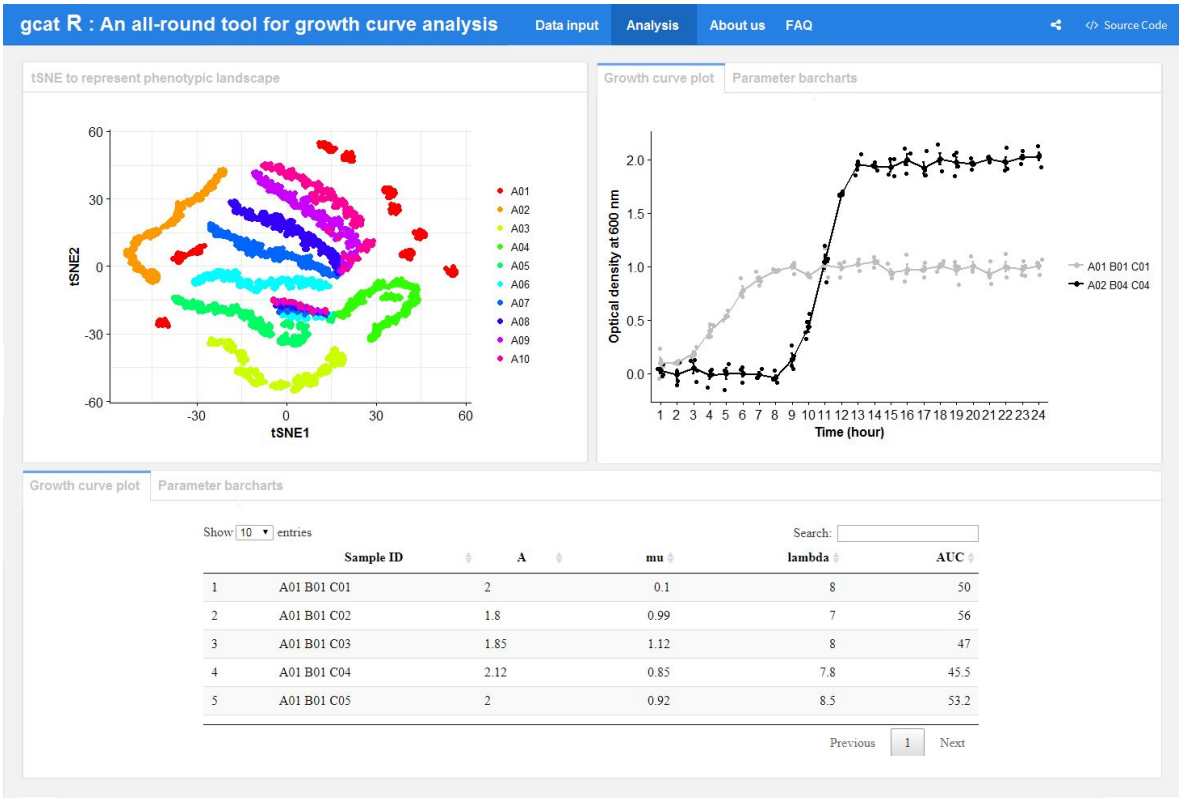


Figure 1. Screenshot of the Result panel. The panel is organised into three different sections: 1. tSNE plot to represent phenotypic landscape of the growth curve data, 2. Growth curve plot and parameter barcharts of the selected samples and 3. Interactive table showing growth curve parameters of all samples in the dataset.

Despite the benefits of these biological assays, very few applications have been developed to systematically analyze cell growth. *grofit*, for example, is an add-on application that must be run on an R working environment [252]. It is built to provide multiple sophisticated workflows to fit growth curve data analysis and estimate useful parameters - maximum growth rate, lag phase, maximum biomass and area under curve. Even though there are other softwares such as the R package growthcurves [253], these tools are not suitable for most end-users as they require extensive knowledge in programming. Nonetheless, *GATHODE* is an excellent example of softwares with a more user-friendly interface [254]. More importantly, it can be downloaded to run locally without installing additional programs or dealing with complicated settings. Therefore there is a need to develop a quantitative tool with simple interface to analyses different aspects of growth curve data in the high throughput manner.

5.2 Design and features

We decided to developed an open-source application called *gcatR* (growth curve analysis tool based on R). This application is capable of analysing and producing a set of fully customized figures to report the results. It requires users to upload a single CSV file that contains growth curve data and their corresponding experimental design. We implement *gcatR* with a number of useful features including interactive data table, growth curve parameter estimation, time series clustering and classification, and fitting based on generalised additive models. *gcatR* enables users to create an interactive database using the R package DT. To estimate growth curve parameters, *gcFitSpline* from the R package *grofit* is integrated into our workflow to analyse fit growth curve data that may not comply with any known growth curve model [252]. We also implement t-distributed stochastic neighbor embedding to compress growth curve data in order to compare individual samples based on their spatial locations [255]. This technique is coupled by clustering to infer possible phenotypes in the test subjects [256]. What is more, we combine generalized additive model (GAM) from the package *mgcv* to fit time series data and estimate statistical significance of the tested variables [257]. For example, it can be used in

an experimental setup where researchers aim to determine the effect of certain environmental factors such as salinity, osmoticity and temperature. And finally, *gcatR* has features to generate attractive figures to visualise growth curves using line plot, growth curve parameters and scatter plot/heatmap using bar chart and clustering result using scatter plot. Please see Figure 1 for a screenshot of gcatR.

We built *gcatR* using R programming language on the Shiny platform to be a well established technology for engaging users who may not have basic computation skills [126]. Effectively, it provides a number of widget controls that allows users to fine tune their findings or customize the look of figures generated. This application can be deployed in two ways: first, it can either be downloaded and run locally on an R console; and second, it can be accessed online when the application is hosted on a public server. The second approach has become increasingly more popular as the cost of renting a server is getting cheaper and it does not require much technical/programming skills.

Despite the advantages of cloud hosting, demand for an on-premise application has been growing in several industries. This is particularly due to concern for privacy and security of certain sensitive documents, in fact, despite strict hosting policies, the risk of data being hacked is still a major worry for the majority of companies.. Inno Setup provides such solution as it helps with compiling R codes and conversion to a desktop application that can be run locally on a personal computer without independent installation of the R program.

5.3 Conclusion

gcatR is created entirely in the R language with several key features including interactive data table, growth curve parameter estimation, time series clustering and classification, and fitting based on generalised additive models. gcatR is available online at http://www.gcatr.manchester.ac.uk/ or it can be downloaded from the same link for installation on a local Windows device free of charge.

Conclusions

In chapter 1, general idea of yeast biology was presented together with relevant tools for experimentation. A number of topics was introduced including life cycle, natural history, mitochondria and cold-stress response. After reviewing hundreds of scientific literatures, it is clear that the community has put a significant amount of interest in genetic variation within and between *Saccharomyces* species and how it can be exploited for industrial use. Take this into account, this thesis aimed to explore the effect of inter and intraspecific variation on the phenotypes of natural yeast species and hybrids. In this chapter, the content of this thesis is summarised and the impact of the findings is discussed together with directions of future work.

Chapter 2 features the results of competition experiments that were coupled with barcode analysis and sequencing on the Illumina Hiseq platform to identify genes with relevant functions in response to nutrient starvation and temperature stress in *S. cerevisiae* strain Σ1278b. In short, genes involved in genome integrity maintenance and general metabolism appear to be crucially important for the growth of Σ1278b in complete and nutrient limited media, respectively. Additional analysis was carried out and revealed that there is little fitness variation between this strain and BY4743. Given that, less conserved genes are likely to produce background-specific phenotypes. To validate the results, monoculture experiments and other phenotypic assays could be performed on the mutants in the heterozygous collection of the Sigma and the BY. This should overcome the limitation of high throughput screening as it may suffer from false discoveries.

Chapter 3 features the effect of chimeric protein complexes in yeast hybrids. In short, the results show that different combinations of GCN1-GCN20 have significant impacts on a wide range of phenotypes including amino acid starvation and cold stress response. For example, combining *GCN1* from *S. cerevisiae* and *GCN20* from *S. uvarum* appeared to make the hybrids more robust under stress conditions. This allelic combination is the one that is retained within group 1 *S. pastorianus*. In addition, it was found that variation in phenotypic traits could be caused by structural differences of the protein GCN1 and differential transcriptional regulation of the gene

GCN20. Interestingly, this may explain biased retention of this protein complex in the natural hybrid. Finally, different types of experiments such as domain swapping and site directed mutagenesis could be performed to support the finding of this study.

Chapter 4 features the effect of mitotypes on several aspects of yeast hybrids. In short, the results show that mtDNA from *S. cerevisiae* grew better at warm and vice versa for mtDNA from *S. uvarum*. To explain this phenomenon, RNA sequencing was performed to inspect global expression of nuclear genes in different genetic backgrounds. And, differential expression analysis and co-expression analysis were run in an attempt to explain relationship between fitness and expression patterns. In this study, it was proven that mitotypes have a direct impact on allelic expression bias in several biological pathways. For example, the expression of orthologs for respiration from Sc and Su subgenomes is likely to be suppressed in complete medium at 28°C for HMtSc and at 16°C for HMtSu, respectively. Potentially, this could be explained by genetic variation of cis-regulatory elements and mitochondrial genome of the two related yeast species. Finally, more experiments could be performed to validate the findings of the study. For instance, one could do a high throughput screening for mitochondrial and nuclear proteins to see whether it is consistent with the results here.

And finally, chapter 5 introduces an open-source application - gcatR. This application is created in R language and contains several features including interactive data table, growth curve parameter estimation, time series clustering and classification, and model fitting based on generalised additive models. What's more, it creates a set of diagrams (such as line graph, bar chart and heatmap) to represent the obtained results. More importantly, this application is designed to be user friendly and does not require users to have prior knowledge in any programming language. gcatR is provided free of charge and can be accessed at http://www.gcatr.manchester.ac.uk/ or downloaded to install on a local Windows device.

This thesis as a whole has demonstrated that genetic variation within and between *Saccharomyces* species has considerable effects on different aspects of yeast

phenotypes. This, in general, is a great example of adaptive differentiation. Systemic comparison between different strains of *S. cerevisiae* has provided an instance of single-nucleotide polymorphism as a number of less conserved genes was identified and linked to specific characteristics of the strain Σ1278b. Similar mutations have also been observed in other strains of budding yeast and many are associated with unique enological and technological properties [258,259].

In addition, adaptive traits can arise from hybridisation. For example, *S. pastorianus* which is a cross between *S. cerevisiae* and *S. eubayanus* that carries mitochondrial DNA from the cold adapted parent has been reported to display a set of phenotypes suitable for cold fermentation [34–37]. As reported in the previous section, chimeric protein complexes which emerge from specific gene loss after hybridisation appear to influence overall fitness of yeast hybrids under a broad range of environments through structural differences and differential transcriptional regulation. Alternatively, novel traits can be the result of biased mitochondrial inheritance. It was shown that transcription of a given subgenome is concerted in specific pathways depending on the type of mitochondrial DNA. This means that cells are likely to work preferentially with one set of parental alleles or the other according to different cellular functions.

This is great but the question is what could we do with this knowledge? In recent years, there is a trend of integrating biology with computer science and using the state of the art machine learning to extract patterns or to make a precise prediction of biological systems. Many research groups have tried and achieved impressive results [260–263]. In this case, one could build a predictive model from genotype and phenotype data to capture the complexity of yeast systems and use it to improve/develop a strain for biotechnological and pharmaceutical purposes.

References

1. Ayto J, editor. Vienna bread. An A-Z of Food and Drink. Oxford University Press; 2002.
2. Engel SR, Dietrich FS, Fisk DG, Binkley G, Balakrishnan R, Costanzo MC, et al. The reference genome sequence of *Saccharomyces cerevisiae:* then and now. G3. 2014;4: 389–398.
3. Giaever G, Nislow C. The yeast deletion collection: a decade of functional genomics. Genetics. 2014;197: 451–465.
4. Scarcelli JJ, Viggiano S, Hodge CA, Heath CV, Amberg DC, Cole CN. Synthetic genetic array analysis in *Saccharomyces cerevisiae* provides evidence for an interaction between RAT8/DBP5 and genes encoding P-body components. Genetics. 2008;179: 1945–1955.
5. Sheff M a., Thorn KS. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae.* Yeast. 2004;21: 661–70.
6. Boekhout T, Robert V. Yeasts in Food. Woodhead Publishing Limited; 2003.
7. McCutcheon JP. Computational identification of non-coding RNAs in *Saccharomyces cerevisiae* by comparative genomics. Nucleic Acids Res. 2004;32.
8. Herskowitz I. Life cycle of the budding yeast *Saccharomyces cerevisiae.* Microbiol Rev. 1988;52: 536–553.
9. Klar AJ, Fogel S, Radin DN. Switching of a mating-type a mutant allele in budding yeast *Saccharomyces cerevisiae*. Genetics. 1979;92: 759–776.
10. Cavalieri D, McGovern PE, Hartl DL, Mortimer R, Polsinelli M. Evidence for *S. cerevisiae* fermentation in ancient wine. J Mol Evol. 2003;57 Suppl 1: S226–S232.
11. Fay JC, Benavides JA. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae.* PLoS Genet. 2005;1: 66–71.
12. Peter J, De Chiara M, Friedrich A, Yue J-X, Pflieger D, Bergström A, et al. Genome evolution across 1,011 *Saccharomyces cerevisiae* isolates. Nature. 2018;556: 339–344.

13. Legras J-L, Merdinoglu D, Cornuet J-M, Karst F. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. Mol Ecol. 2007;16: 2091–2102.
14. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458: 337–341.
15. Winzeler EA, Castillo-Davis CI, Oshiro G, Liang D, Richards DR, Zhou Y, et al. Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics. 2003;163: 79–89.
16. Chin BL, Ryan O, Lewitter F, Boone C, Fink GR. Genetic variation in *Saccharomyces cerevisiae*: circuit diversification in a signal transduction network. Genetics. 2012;192: 1523–1532.
17. Vaughan Martini A, Martini A. Three newly delimited species of *Saccharomyces sensu stricto.* Antonie van Leeuwenhoek. 1987;53.
18. Wang S-A, Bai F-Y. *Saccharomyces arboricolus* sp. nov., a yeast species from tree bark. Int J Syst Evol Microbiol. 2008;58: 510–514.
19. Naumov GI, James SA, Naumova ES, Louis EJ, Roberts IN. Three new species in the *Saccharomyces sensu stricto* complex: *Saccharomyces cariocanus, Saccharomyces kudriavzevii* and *Saccharomyces mikatae.* Int J Syst Evol Microbiol. 2000;50 Pt 5: 1931–1942.
20. Naumov GI, Naumova ES, Louis EJ. Two new genetically isolated populations of the *Saccharomyces sensu stricto* complex from Japan. J Gen Appl Microbiol. 1995;41.
21. Naseeb S, James SA, Alsammar H, Michaels CJ, Gini B, Nueno-Palop C, et al. *Saccharomyces jurei* sp. nov., isolation and genetic identification of a novel yeast species from Quercus robur. Int J Syst Evol Microbiol. 2017;67: 2046–2052.
22. Masneuf I, Hansen J, Groth C, Piskur J, Dubourdieu D. New hybrids between *Saccharomyces sensu stricto* yeast species found among wine and cider production strains. Appl Environ Microbiol. 1998;64: 3887–3892.
23. Nguyen H-V, Legras J-L, Neuvéglise C, Gaillardin C. Deciphering the hybridisation history leading to the Lager lineage based on the mosaic genomes

of Saccharomyces bayanus strains NBRC1948 and CBS380. PloS one. 2011;6: e25821.

24. Sniegowski PD, Dombrowski PG, Fingerman E. *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. FEMS yeast Res. 2002;1: 299–306.
25. Salvadó Z, Arroyo-López FN, Guillamón JM, Salazar G, Querol A, Barrio E. Temperature adaptation markedly determines evolution within the genus *Saccharomyces*. Appl Environ Microbiol. 2011;77: 2292–2302.
26. Libkind D, Hittinger CT, Valério E, Gonçalves C, Dover J, Johnston M, et al. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc Natl Acad Sci United States Am. 2011;108: 14539–14544.
27. González SS, Barrio E, Gafner J, Querol A. Natural hybrids from *Saccharomyces cerevisiae, Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. FEMS yeast Res. 2006;6: 1221–1234.
28. Erny C, Raoult P, Alais A, Butterlin G, Delobel P, Matei-Radoi F, et al. Ecological success of a group of *Saccharomyces cerevisiae/Saccharomyces kudriavzevii* hybrids in the northern european wine-making environment. Appl Environ Microbiol. 2012;78: 3256–3265.
29. Baker E, Wang B, Bellora N, Peris D, Hulfachor AB, Koshalek JA, et al. The Genome Sequence of Saccharomyces eubayanus and the Domestication of Lager-Brewing Yeasts. Mol Biol Evol. 2015;32: 2818–2831.
30. Peris D, Langdon QK, Moriarty RV, Sylvester K, Bontrager M, Charron G, et al. Complex Ancestries of Lager-Brewing Hybrids Were Shaped by Standing Variation in the Wild Yeast *Saccharomyces eubayanus*. PLoS Genet. 2016;12: e1006155.
31. Dunn B, Sherlock G. Reconstruction of the genome origins and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. Genome Res. 2008;18: 1610–1623.
32. Gasch A. Faculty of 1000 evaluation for Reconstruction of the genome origins

and evolution of the hybrid lager yeast *Saccharomyces pastorianus*. F1000 - Post-publication peer review of the biomedical literature. 2008.

33. Monerawela C, Bond U. The hybrid genomes of *Saccharomyces pastorianus*: A current perspective. Yeast. 2018;35: 39–50.
34. Gibson BR, Storgårds E, Krogerus K, Vidgren V. Comparative physiology and fermentation performance of Saaz and Frohberg lager yeast strains and the parental species Saccharomyces eubayanus. Yeast. 2013;30: 255–266.
35. Ekberg J, Rautio J, Mattinen L, Vidgren V, Londesborough J, Gibson BR. Adaptive evolution of the lager brewing yeast *Saccharomyces pastorianus* for improved growth under hyperosmotic conditions and its influence on fermentation performance. FEMS yeast Res. 2013;13: 335–349.
36. Adamczyk J, Deregowska A, Skoneczny M, Skoneczna A, Natkanska U, Kwiatkowska A, et al. Copy number variations of genes involved in stress responses reflect the redox state and DNA damage in brewing yeasts. Cell Stress & Chaperon-. 2016;21: 849–864.
37. Hull RM, Cruz C, Jack CV, Houseley J. Environmental change drives accelerated adaptation through stimulated copy number variation. PLoS Biol. 2017;15: e2001333.
38. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. Sci. 1996;274: 546, 563–546, 567.
39. Non-coding DNA – a brief review. J Appl Biol & Biotechnol. 2017;
40. Hooks KB, Delneri D, Griffiths-Jones S. Intron evolution in *Saccharomycetaceae*. Genome Biol Evol. 2014;6: 2543–2556.
41. Sturgeon CM, Kemmer D, Anderson HJ, Roberge M. Yeast as a tool to uncover the cellular targets of drugs. Biotechnol J. 2006;1: 289–298.
42. Tugendreich S, Bassett DE, McKusick VA, Boguski MS, Hieter P. Genes conserved in yeast and humans. Hum Mol Genet. 1994;3 Spec No: 1509–1517.
43. Crawford N. Book Review: Mitochondria Structure and Function. Ann Clin Biochem Int J Biochem Lab Med. 1971;8.
44. Sharer JD. Mitochondrial genetics. Curr Protoc Hum Genet. 2011;Chapter 19: Unit19.0.

45. Martin WF, Müller M, editors. Origin of Mitochondria and Hydrogenosomes. Springer Berlin Heidelberg; 2007.
46. Williamson D. The curious history of yeast mitochondrial DNA. Nat Rev Genet. 2002;3: 475–481.
47. Sager R. Mitochondrial Genetics of Yeast. Cytoplasmic Genes and Organelles. Elsevier; 1972. p. 105–144p.
48. Payne BAI, Chinnery PF. Mitochondrial dysfunction in aging: Much progress but many unresolved questions. Biochim et Biophys Acta. 2015;1847: 1347–1353.
49. Balaban RS, Nemoto S, Finkel T. Mitochondria, Oxidants, and Aging. Cell. 2005;120: 483–495.
50. Linnane AW. Mitochondria and aging: the universality of bioenergetic disease. Aging. 1992;4: 267–271.
51. Neiman M, Taylor DR. The causes of mutation accumulation in mitochondrial genomes. Proceedings Biol Sci. 2009;276: 1201–1209.
52. Quirós PM, Mottis A, Auwerx J. Mitonuclear communication in homeostasis and stress. Nat Rev Mol cell Biol. 2016;17: 213–226.
53. Koch L. Genetic variation: Nuclear and mitochondrial genome interplay. Nat Rev Genet. 2016;17: 502.
54. Poyton RO, McEwen JE. Crosstalk between nuclear and mitochondrial genomes. Annu Rev Biochem. 1996;65: 563–607.
55. Prokisch H, Scharfe C, Camp DG, Xiao W, David L, Andreoli C, et al. Integrative Analysis of the Mitochondrial Proteome in Yeast. PLoS Biol. 2004;: 160.
56. Kim HP. Lightening up Light Therapy: Activation of Retrograde Signaling Pathway by Photobiomodulation. Biomol & Ther. 2014;22: 491–496.
57. Azpiroz R, Butow RA. Mitochondrial inheritance in yeast. Methods Enzymol. 1995;260: 453–465.
58. Birky CW, Demko CA, Perlman PS, Strausberg R. Uniparental inheritance of mitochondrial genes in yeast: dependence on input bias of mitochondrial DNA and preliminary investigations of the mechanism. Genetics. 1978;89: 615–651.
59. Hoekstra RF. Evolutionary origin and consequences of uniparental mitochondrial inheritance. Hum Reprod. 2000;15 Suppl 2: 102–111.

60. Piškur J. Inheritance of the Yeast Mitochondrial Genome. Plasmid. 1994;31.
61. Rainieri S, Kodama Y, Nakao Y, Pulvirenti A, Giudici P. The inheritance of mtDNA in lager brewing strains. FEMS yeast Res. 2008;8: 586–596.
62. Hsu Y-Y, Chou J-Y. Environmental Factors Can Influence Mitochondrial Inheritance in the *Saccharomyces* Yeast Hybrids. PloS one. 2017;12: e0169953.
63. Breton S, Stewart DT. Atypical mitochondrial inheritance patterns in eukaryotes. Genome. 2015;58: 423–431.
64. Yan Z, Sun S, Shahid M, Xu J. Environment factors can influence mitochondrial inheritance in the fungus *Cryptococcus neoformans*. Fungal Genet Biol : FG & B. 2007;44: 315–322.
65. Lombard J. Once upon a time the cell membranes: 175 years of cell boundary research. Biol Direct. 2014;9: 32.
66. van der Rest ME, Kamminga AH, Nakano A, Anraku Y, Poolman B, Konings WN. The plasma membrane of Saccharomyces cerevisiae: structure, function, and biogenesis. Microbiol Rev. 1995;59: 304–322.
67. Santos E, Leal F, Sentandreu R. The plasma membrane of *Saccharomyces cerevisiae*. Molecular structure and asymmetry. Biochim et Biophys Acta. 1982;685: 329–339.
68. Rubio-Texeira M, Van Zeebroeck G, Voordeckers K, Thevelein JM. *Saccharomyces cerevisiae* plasma membrane nutrient sensors and their role in PKA signaling. FEMS yeast Res. 2010;10: 134–149.
69. Walker G. Yeast Physiology and Biotechnology. John Wiley & Sons, Inc. NY. John Wiley & Sons, Inc. NY; 1998.
70. Belrhiti A. Membrane Lipid and Sterol Distribution in *Saccharomyces bayanus*. Process Biochem. 1995;30.
71. Buzzini P, Margesin R. Cold-Adapted Yeasts: A Lesson from the Cold and a Challenge for the XXI Century. Cold-adapted Yeasts. Springer Berlin Heidelberg; 2014. pp. 3–22.
72. Loertscher J, Larson LL, Matson CK, Parrish ML, Felthauser A, Sturm A, et al. Endoplasmic reticulum-associated degradation is required for cold adaptation and regulation of sterol biosynthesis in the yeast *Saccharomyces cerevisiae*.

Eukaryot cell. 2006;5: 712–722.

73. O'Rourke SM, Herskowitz I, O'Shea EK. Yeast go the whole HOG for the hyperosmotic response. Trends Genet : TIG. 2002;18: 405–412.
74. Reiser V, Raitt DC, Saito H. Yeast osmosensor Sln1 and plant cytokinin receptor Cre1 respond to changes in turgor pressure. J cell Biol. 2003;161: 1035–1040.
75. Hayashi M, Maeda T. Activation of the HOG pathway upon cold stress in Saccharomyces cerevisiae. J Biochem. 2006;139: 797–803.
76. Lee WC, Xue ZX, Mélèse T. The NSR1 gene encodes a protein that specifically binds nuclear localization sequences and has two RNA recognition motifs. J cell Biol. 1991;113: 1–12.
77. Yan C, Mélèse T. Multiple regions of NSR1 are sufficient for accumulation of a fusion protein within the nucleolus. J cell Biol. 1993;123: 1081–1091.
78. Lee WC, Zabetakis D, Mélèse T. NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. Mol Cell Biol. 1992;12: 3865–3871.
79. Wu K, Wu P, Aris JP. Nucleolar protein Nop12p participates in synthesis of 25S rRNA in Saccharomyces cerevisiae. Nucleic acids Res. 2001;29: 2938–2949.
80. Barta I, Iggo R. Autoregulation of expression of the yeast Dbp2p "DEAD-box" protein is mediated by sequences in the conserved DBP2 intron. EMBO J. 1995;14: 3800–3808.
81. Defining the Role of DEAD-box Proteins in RNA Biology♦: The DEAD-box RNA Helicase Dbp2 Connects RNA Quality Control with Repression of Aberrant Transcription. J Biol Chem. 2012;287.
82. Cloutier SC, Ma WK, Nguyen LT, Tran EJ. The DEAD-box RNA helicase Dbp2 connects RNA quality control with repression of aberrant transcription. J Biol Chem. 2012;287: 26155–26166.
83. Richter K, Haslbeck M, Buchner J. The heat shock response: life on the verge of death. Mol cell. 2010;40: 253–266.
84. Whitley D, Goldberg SP, Jordan WD. Heat shock proteins: a review of the molecular chaperones. J Vasc Surg. 1999;29: 748–751.
85. Lindquist S, Craig EA. The heat-shock proteins. Annu Rev Genet. 1988;22:

631–677.

86. Aguilera J, Randez-Gil F, Prieto JA. Cold response in S*accharomyces cerevisiae*: new functions for old mechanisms. FEMS Microbiol Rev. 2007;31: 327–341.
87. Panadero J, Pallotti C, Rodriguez-Vargas S, Randez-Gil F, Prieto JA. A Downshift in Temperature Activates the High Osmolarity Glycerol (HOG) Pathway, Which Determines Freeze Tolerance in *Saccharomyces cerevisiae*. J Biol Chem. 2005;281.
88. Duskova M, Borovikova D, Herynkova P, Rapoport A, Sychrova H. The role of glycerol transporters in yeast cells in various physiological and stress conditions. FEMS Microbiol Lett. 2015;362: 1–8.
89. Kitichantaropas Y, Boonchird C, Sugiyama M, Kaneko Y, Harashima S, Auesukaree C. Cellular mechanisms contributing to multiple stress tolerance in *Saccharomyces cerevisiae* strains with potential use in high-temperature ethanol fermentation. AMB Express. 2016;6: 107.
90. Merín MG, Mendoza LM, Farías ME, Morata de Ambrosini VI. Isolation and selection of yeasts from wine grape ecosystem secreting cold-active pectinolytic activity. Int J Food Microbiol. 2011;147: 144–148.
91. Naga Padma P, Anuradha K, Reddy G. Pectinolytic yeast isolates for cold-active polygalacturonase production. Innov Food Sci & Emerg Technol. 2011;12.
92. Ortiz-Tovar G, Pérez-Torrado R, Adam AC, Barrio E, Querol A. A comparison of the performance of natural hybrids *Saccharomyces cerevisiae* × *Saccharomyces kudriavzevii* at low temperatures reveals the crucial role of their S. kudriavzevii genomic contribution. Int J Food Microbiol. 2018;274: 12–19.
93. Fehrmann S, Bottin-Duplus H, Leonidou A, Mollereau E, Barthelaix A, Wei W, et al. Natural sequence variants of yeast environmental sensors confer cell-to-cell expression variability. Mol Syst Biol. 2013;9: 695.
94. Moreno-Sánchez R, Saavedra E, Rodríguez-Enríquez S, Olín-Sandoval V. Metabolic control analysis: a tool for designing strategies to manipulate metabolic pathways. J Biomed & Biotechnol. 2008;2008: 597913.
95. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW. Quantitative

phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet. 1996;14: 450–456.

96. Delneri D. Competition experiments coupled with high-throughput analyses for functional genomics studies in yeast. Methods Mol Biol. 2011;759: 271–282.
97. Hanway D, Chin JK, Xia G, Oshiro G, Winzeler EA, Romesberg FE. Previously uncharacterized genes in the UV- and MMS-induced DNA damage response in yeast. Proc Natl Acad Sci United States Am. 2002;99: 10605–10610.
98. Parker S, Fraczek MG, Wu J, Shamsah S, Manousaki A, Dungrattanalert K, et al. Large-scale profiling of noncoding RNA function in yeast. PLoS Genet. 2018;14: e1007253.
99. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. Nature. 2017;550: 345–353.
100. Corney DC. RNA-seq Using Next Generation Sequencing. Mater Methods. 2013;3.
101. Next-generation sequencing in transcriptomics: RNA-seq experiments. Molecular Data Analysis Using R. John Wiley & Sons, Inc.; 2017. pp. 145–166.
102. Corney DC. RNA-seq Using Next Generation Sequencing. Mater Methods. 2012;2.
103. CS P. Transcriptome Analysis on RNA-seq Data. J Next Gener Seq & Appl. 2015;01.
104. Furey TS. ChIP–seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. Nat Rev Genet. 2012;13.
105. Park PJ. ChIP-seq: advantages and challenges of a maturing technology. Nat Rev Genet. 2009;10: 669–80.
106. Delneri D. Barcode technology in yeast: application to pharmacogenomics. FEMS yeast Res. 2010;10: 1083–1089.
107. Levy SE, Myers RM. Advancements in Next-Generation Sequencing. Annu Rev Genomic- Hum Genet. 2016;17: 95–115.
108. Next-next generation sequencing (next2generation sequencing, third generation sequencing). The Dictionary of Genomics, Transcriptomics and

Proteomics. Wiley-VCH Verlag GmbH & Co. KGaA; 2015. pp. 1–1.

109. von Bubnoff A. Next-Generation Sequencing: The Race Is On. Cell. Cell Press; 2008;132: 721–723.

110. Meng Q, Yu J. Next Generation DNA Sequencing Technologies and Applications. Next Generation Sequencing and Whole Genome Selection in Aquaculture: Liu/Next Generation Sequencing and Whole Genome Selection in Aquaculture. Wiley-Blackwell; 2010. pp. 35–56.

111. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomic-. 2012;13: 341.

112. Costa-Silva J, Domingues D, Lopes FM. RNA-Seq differential expression analysis: An extended review and a software tool. PloS one. 2017;12: e0190152.

113. Gene set enrichment analysis of RNA seq data integrating differential expression and splicing.pdf.

114. Grant GR, Farkas MH, Pizarro AD, Lahens NF, Schug J, Brunk BP, et al. Comparative analysis of RNA-Seq alignment algorithms and the RNA-Seq unified mapper (RUM). Bioinforma. 2011;27: 2518–28.

115. WANG X, WANG X-W, WANG L-K, FENG Z-X, ZHANG X-G. A Review on The Processing and Analysis of Next-generation RNA-seq Data*: A Review on The Processing and Analysis of Next-generation RNA-seq Data. Prog Biochem Biophys. 2010;37.

116. Gillis J, Pavlidis P. A methodology for the analysis of differential coexpression across the human lifespan. BMC Bioinforma. 2009;10: 306.

117. Yang E-W, Girke T, Jiang T. Differential gene expression analysis using coexpression and RNA-Seq data. Bioinformatics. 2013;29.

118. Draghici S, Khatri P, Tarca AL, Amin K, Done A, Voichita C, et al. A systems biology approach for pathway level analysis A systems biology approach for pathway level analysis. Genome Res. 2007;17: 1537–1545.

119. Delneri D, Hoyle DC, Gkargkas K, Cross EJM, Rash B, Zeef L, et al. Identification and characterization of high-flux-control genes of yeast through

competition analyses in continuous cultures. Nat Genet. 2008;40: 113–117.

120. Paget CM, Schwartz J-M, Delneri D. Environmental systems biology of cold-tolerant phenotype in *Saccharomyces* species adapted to grow at different temperatures. Mol Ecol. 2014;23: 5241–5257.

121. Hewitt SK, Donaldson IJ, Lovell SC, Delneri D. Sequencing and characterisation of rearrangements in three *S. pastorianus* strains reveals the presence of chimeric genes and gives evidence of breakpoint reuse. PloS one. 2014;9: e92203.

122. Marton MJ, Vazquez de Aldana CR, Qiu H, Chakraburtty K, Hinnebusch AG. Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIF2alpha kinase GCN2. Mol Cell Biol. 1997;17: 4474–4489.

123. Vazquez de Aldana CR, Marton MJ, Hinnebusch AG. GCN20, a novel ATP binding cassette protein, and GCN1 reside in a complex that mediates activation of the eIF-2 alpha kinase GCN2 in amino acid-starved cells. EMBO J. 1995;14: 3184–3199.

124. Garcia-Barrio M. Association of GCN1-GCN20 regulatory complex with the N-terminus of eIF2alpha kinase GCN2 is required for GCN2 activation. EMBO J. 2000;19.

125. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna; 2008.

126. Beeley C. Web Application Development with R Using Shiny. Packt Publishing; 2013.

127. Sampaio JP, Gonçalves P. Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus.* Appl Environ Microbiol. 2008;74: 2144–2152.

128. Aa E, Townsend JP, Adams RI, Nielsen KM, Taylor JW. Population structure and gene evolution in *Saccharomyces cerevisiae*. FEMS yeast Res. 2006;6: 702–715.

129. Koufopanou V, Hughes J, Bell G, Burt A. The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus.*

Philos Trans R Soc London Ser B, Biol Sci. 2006;361: 1941–1946.

130. Kuehne HA, Murphy HA, Francis CA, Sniegowski PD. Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations. Curr Biol : CB. 2007;17: 407–411.

131. Liti G, Barton DBH, Louis EJ. Sequence diversity, reproductive isolation and species concepts in Saccharomyces. Genetics. 2006;174: 839–850.

132. Ruderfer DM, Pratt SC, Seidel HS, Kruglyak L. Population genomic analysis of outcrossing and recombination in yeast. Nat Genet. 2006;38: 1077–1081.

133. Tsai IJ, Bensasson D, Burt A, Koufopanou V. Population genomics of the wild yeast Saccharomyces paradoxus: Quantifying the life cycle. Proc Natl Acad Sci United States Am. 2008;105: 4957–4962.

134. Deutschbauer AM, Jaramillo DF, Proctor M, Kumm J, Hillenmeyer ME, Davis RW, et al. Mechanisms of haploinsufficiency revealed by genome-wide profiling in yeast. Genetics. 2005;169: 1915–25.

135. Cleary MA. Haploinsufficiency. Encyclopedia of Genetics. Elsevier; 2001. p. 911.

136. Veitia RA. Gene Dosage Balance in Cellular Pathways: Implications for Dominance and Gene Duplicability. Genetics. 2004;168.

137. Veitia RA. Gene dosage balance: deletions, duplications and dominance. Trends Genet : TIG. 2005;21: 33–35.

138. Zhang YJ, Zhang S, Liu XZ, Wen HA, Wang M. A simple method of genomic DNA extraction suitable for analysis of bulk fungal strains. Lett Appl Microbiol. 2010;51: 114–118.

139. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma. 2010;Chapter 11: Unit 11.7.

140. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2.

141. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550.

142. de Clare M, Oliver SG. Copy-number variation of cancer-gene orthologs is sufficient to induce cancer-like symptoms in *Saccharomyces cerevisiae*. BMC Biol. 2013;11: 24.

143. Glassner BJ, Mortimer RK. Synergistic interactions between RAD5, RAD16 and RAD54, three partially homologous yeast DNA repair genes each in a different repair pathway. Radiat Res. 1994;139: 24–33.

144. Johnston LH, Johnson AL. The DNA repair genes RAD54 and UNG1 are cell cycle regulated in budding yeast but MCB promoter elements have no essential role in the DNA damage response. Nucleic acids Res. 1995;23: 2147–2152.

145. Widlak P, Pietrowska M, Lanuszewska J. The role of chromatin proteins in DNA damage recognition and repair Mini-review. Histochem Cell Biol. 2005;125.

146. Boytsov L. Indexing methods for approximate dictionary searching: Comparative analysis. J Exp Algorithmics. 2011;16.

147. Navarro G. A guided tour to approximate string matching. ACM Comput Surv. 2001;33.

148. Broach JR. Nutritional control of growth and development in yeast. Genetics. 2012;192: 73–105.

149. Hanscho M, Ruckerbauer DE, Chauhan N, Hofbauer HF, Krahulec S, Nidetzky B, et al. Nutritional requirements of the BY series of *Saccharomyces cerevisiae* strains for optimum growth. FEMS yeast Res. 2012;12: 796–808.

150. Mohammed HH, Brown TL, Beck BH, Yildirim-Aksoy M, Eljack RM, Peatman E. The effects of dietary inclusion of a *Saccharomyces cerevisiae* fermentation product in a commercial catfish ration on growth, immune readiness, and columnaris disease susceptibility. J Appl Aquac. 2018; 1–17.

151. Gangloff S, Arcangioli B. DNA repair and mutations during quiescence in yeast. FEMS yeast Res. 2017;17.

152. Sweet DH, Jang YK, Sancar GB. Role of UME6 in transcriptional regulation of a DNA repair gene in *Saccharomyces cerevisiae*. Mol Cell Biol. 1997;17: 6223–6235.

153. Fu Y, Pastushok L, Xiao W. DNA damage-induced gene expression in Saccharomyces cerevisiae. FEMS Microbiol Rev. 2008;32: 908–926.

154. Gutiérrez A, Beltran G, Warringer J, Guillamón JM. Genetic basis of variations in nitrogen source utilization in four wine commercial yeast strains. PloS one. 2013;8: e67166.

155. Gutiérrez A, Chiva R, Sancho M, Beltran G, Arroyo-López FN, Guillamon JM. Nitrogen requirements of commercial wine yeast strains during fermentation of a synthetic grape must. Food Microbiol. 2012;31: 25–32.

156. Godard P, Urrestarazu A, Vissers S, Kontos K, Bontempi G, van Helden J, et al. Effect of 21 different nitrogen sources on global gene expression in the yeast *Saccharomyces cerevisiae*. Mol Cell Biol. 2007;27: 3065–3086.

157. Pasqualone D, Huffaker TC. STU1, a suppressor of a beta-tubulin mutation, encodes a novel and essential component of the yeast mitotic spindle. J cell Biol. 1994;127: 1973–1984.

158. Yin H, You L, Pasqualone D, Kopski KM, Huffaker TC. Stu1p is physically associated with beta-tubulin and is required for structural integrity of the mitotic spindle. Mol Biol cell. 2002;13: 1881–1892.

159. Qin J, Zhou YJ, Krivoruchko A, Huang M, Liu L, Khoomrung S, et al. Modular pathway rewiring of Saccharomyces cerevisiae enables high-level production of L-ornithine. Nat Commun. 2015;6: 8224.

160. Sandai D, Yin Z, Selway L, Stead D, Walker J, Leach MD, et al. The evolutionary rewiring of ubiquitination targets has reprogrammed the regulation of carbon assimilation in the pathogenic yeast Candida albicans. mBio. 2012;3.

161. Marsit S, Dequin S. Diversity and adaptive evolution of *Saccharomyces* wine yeast: a review. FEMS yeast Res. 2015;15.

162. Sáez-Vásquez J, Gadal O. Genome organization and function: a view from yeast and Arabidopsis. Mol Plant. 2010;3: 678–690.

163. Dunn B, Paulish T, Stanbery A, Piotrowski J, Koniges G, Kroll E, et al. Recurrent rearrangement during adaptive evolution in an interspecific yeast hybrid suggests a model for rapid introgression. PLoS Genet. 2013;9: e1003366.

164. El-Mabrouk N, Sankoff D. Hybridization and Genome Rearrangement. Combinatorial Pattern Matching. Springer Berlin Heidelberg; 1999. pp. 78–87.

165. Sipiczki M. Diversity, variability and fast adaptive evolution of the wine yeast (*Saccharomyces cerevisiae*) genome—a review. Ann Microbiol. 2011;61.

166. Fay JC, McCullough HL, Sniegowski PD, Eisen MB. Population genetic variation in gene expression is associated with phenotypic variation in

Saccharomyces cerevisiae. Genome Biol. 2004;5: R26.

167. Warringer J, Zörgö E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, et al. Trait variation in yeast is defined by population history. PLoS Genet. 2011;7: e1002111.

168. Park H, Bakalinsky AT. SSU1 mediates sulphite efflux in *Saccharomyces cerevisiae*. Yeast. 2000;16: 881–888.

169. Nardi T, Corich V, Giacomini A, Blondin B. A sulphite-inducible form of the sulphite efflux gene SSU1 in a *Saccharomyces cerevisiae* wine yeast. Microbiol. 2010;156: 1686–1696.

170. Feulner PGD, De-Kayne R. Genome evolution, structural rearrangements and speciation. J Evol Biol. 2017;30: 1488–1490.

171. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci United States Am. 2013;110: E1743–E1751.

172. Mechanisms of speciation. Proceedings from the International Meeting on Mechanisms of Speciation. May 4-8, 1981, Rome, Italy. Prog Clin Biol Res. 1982;96: 1–546.

173. Dieckmann U, Doebeli M, Metz JAJ, Tautz D, editors. Ecological Mechanisms of Speciation. Adaptive Speciation. Cambridge University Press; 2004. pp. 169–172.

174. Pianka ER. Mechanisms of Speciation. Proceedings from the International Meeting on Mechanisms of Speciation, Sponsored by the Accademia Nazionale dei Lincei, May 4-8, 1981, Rome, Italy. Progress in Clinical and Biological Research, Volume 96.Claudio Barigozzi. Q Rev Biol. 1983;58.

175. Powell JR. Genetic and nongenetic mechanisms of speciation. Prog Clin Biol Res. 1982;96: 67–74.

176. Bae NS, Seberg AP, Carroll LP, Swanson MJ. Identification of Genes in that Are Haploinsufficient for Overcoming Amino Acid Starvation. G3. 2017;7: 1061–1084.

177. Ljungdahl PO, Daignan-Fornier B. Regulation of amino acid, nucleotide, and phosphate metabolism in *Saccharomyces cerevisiae*. Genetics. 2012;190: 885–929.

178. Hinnebusch AG. Translational regulation of GCN4 and the general amino acid control of yeast. Annu Rev Microbiol. Annual Reviews Inc; 2005;59: 407–50.

179. Tripathi G. Gcn4 co-ordinates morphogenetic and metabolic responses to amino acid starvation in *Candida albicans*. EMBO J. 2002;21.

180. Albrecht G, Mösch HU, Hoffmann B, Reusser U, Braus GH. Monitoring the Gcn4 protein-mediated response in the yeast *Saccharomyces cerevisiae*. J Biol Chem. 1998;273: 12696–12702.

181. Castilho BA, Shanmugam R, Silva RC, Ramesh R, Himme BM, Sattlegger E. Keeping the eIF2 alpha kinase Gcn2 in check. Biochim et Biophys Acta. 2014;1843: 1948–1968.

182. Sattlegger E, Hinnebusch AG. Polyribosome binding by GCN1 is required for full activation of eukaryotic translation initiation factor 2alpha kinase GCN2 during amino acid starvation. J Biol Chem. 2005;280: 16514–16521.

183. Sattlegger E. Separate domains in GCN1 for binding protein kinase GCN2 and ribosomes are required for GCN2 activation in amino acid-starved cells. EMBO J. 2000;19.

184. Pau G, Fuchs F, Sklyar O, Boutros M, Huber W. EBImage--an R package for image processing with applications to cellular phenotypes. Bioinformatics. 2010;26.

185. Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. Primer3Plus, an enhanced web interface to Primer3. Nucleic acids Res. 2007;35: W71–W74.

186. Teixeira MC. The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res. 2006;34.

187. Teixeira MC, Monteiro PT, Guerreiro JF, Gonçalves JP, Mira NP, Dos Santos SC, et al. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. Nucleic acids Res. 2014;42: D161–D166.

188. Källberg M, Margaryan G, Wang S, Ma J, Xu J. RaptorX server: a resource for template-based protein structure modeling. Methods Mol Biol. 2014;1137:

17–27.

189. Källberg M, Wang H, Wang S, Peng J, Wang Z, Lu H, et al. Template-based protein structure modeling using the RaptorX web server. Nat Protoc. 2012;7: 1511–1522.

190. DeLano WL. The PyMOL Molecular Graphics System. DeLano Scientific LLC, Palo Alto, CA, USA.; 2009.

191. Tate JJ. Methionine Sulfoximine Treatment and Carbon Starvation Elicit Snf1-independent Phosphorylation of the Transcription Activator Gln3 in Saccharomyces cerevisiae. J Biol Chem. 2005;280.

192. Crespo JL, Powers T, Fowler B, Hall MN. The TOR-controlled transcription activators GLN3, RTG1, and RTG3 are regulated in response to intracellular levels of glutamine. Proc Natl Acad Sci United States Am. 2002;99: 6784–6789.

193. Ecker N, Mor A, Journo D, Abeliovich H. Induction of autophagic flux by amino acid deprivation is distinct from nitrogen starvation-induced macroautophagy. Autophagy. Taylor & Francis; 2010;6: 879–890.

194. Rakesh R, Krishnan R, Sattlegger E, Srinivasan N. Recognition of a structural domain (RWDBD) in Gcn1 proteins that interacts with the RWD domain containing proteins. Biol Direct. 2017;12: 12.

195. Arnold K, Kiefer F, Kopp J, Battey JND, Podvinec M, Westbrook JD, et al. The Protein Model Portal. J Struct Funct Genomic-. 2009;10: 1–8.

196. Breslow DK, Cameron DM, Collins SR, Schuldiner M, Stewart-Ornstein J, Newman HW, et al. A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat Methods. 2008;5.

197. VanderSluis B, Hess DC, Pesyna C, Krumholz EW, Syed T, Szappanos B, et al. Broad metabolic sensitivity profiling of a prototrophic yeast deletion collection. Genome Biol. 2014;15: R64.

198. Messenguy F, Scherens B. Induction of “General Control” and thermotolerance in cdc mutants of *Saccharomyces cerevisiae*. Mol & Gen Genet : MGG. 1990;224: 257–263.

199. Qian W, Ma D, Xiao C, Wang Z, Zhang J. The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. Cell reports. 2012;2:

1399–1410.

200. Pir P, Gutteridge A, Wu J, Rash B, Kell DB, Zhang N, et al. The genetic control of growth rate: a systems biology study in yeast. BMC Syst Biol. 2012;6: 4.

201. Davey HM, Cross EJM, Davey CL, Gkargkas K, Delneri D, Hoyle DC, et al. Genome-wide analysis of longevity in nutrient-deprived *Saccharomyces cerevisiae* reveals importance of recycling in maintaining cell viability. Environ Microbiol. 2012;14: 1249–1260.

202. Walther A, Hesselbart A, Wendland J. Genome sequence of *Saccharomyces carlsbergensis*, the world's first pure culture lager yeast. G3. 2014;4: 783–793.

203. Olesen K, Felding T, Gjermansen C, Hansen J. The dynamics of the *Saccharomyces carlsbergensis* brewing yeast transcriptome during a production-scale lager beer fermentation. FEMS yeast Res. 2002;2: 563–573.

204. Walsh RM, Martin PA. Growth of *Saccharomyces cerevisiae* and *Saccharomyces uvarum* in a temperature gradient incubator. J Inst Brew. 1977;83.

205. Kasowski M, Grubert F, Heffelfinger C, Hariharan M, Asabere A, Waszak SM. Variation in transcription factor binding among humans. New Biotechnol. 2010;27.

206. Zheng W, Zhao H, Mancera E, Steinmetz LM, Snyder M. Genetic analysis of variation in transcription factor binding in yeast. Nature. 2010;464: 1187–1191.

207. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, et al. Divergence of Transcription Factor Binding Sites Across Related Yeast Species. Science. 2007;317: 815–819.

208. Almeida P, Barbosa R, Bensasson D, Gonçalves P, Sampaio JP. Adaptive divergence in wine yeasts and their wild relatives suggests a prominent role for introgressions and rapid evolution at noncoding sites. Mol Ecol. 2017;26: 2167–2182.

209. van den Broek M, Bolat I, Nijkamp JF, Ramos E, Luttik MAH, Koopman F, et al. Chromosomal Copy Number Variation in *Saccharomyces pastorianus* Is Evidence for Extensive Genome Dynamics in Industrial Lager Brewing Strains.

Appl Environ Microbiol. 2015;81: 6253–6267.

210. Chen P, Dong J, Yin H, Bao X, Chen L, He Y, et al. Genome comparison and evolutionary analysis of different industrial lager yeasts (*Saccharomyces pastorianus*): Comparative analysis of genomic characteristics of lager yeasts. J Inst Brew. 2016;122.

211. Boël G, Smith PC, Ning W, Englander MT, Chen B, Hashem Y, et al. The ABC-F protein EttA gates ribosome entry into the translation elongation cycle. Nat Struct & Mol Biol. 2014;21: 143–151.

212. Tinton S a., Schepens B, Bruynooghe Y, Beyaert R, Cornelis S. Regulation of the cell-cycle-dependent internal ribosome entry site of the PITSLRE protein kinase: roles of Unr (upstream of N-ras) protein and phosphorylated translation initiation factor eIF-2alpha. Biochem J. 2005;385: 155–63.

213. Szymczyna BR, Bowman J, McCracken S, Pineda-Lucena A, Lu Y, Cox B, et al. Structure and function of the PWI motif: a novel nucleic acid-binding domain that facilitates pre-mRNA processing. Genes & Dev. 2003;17: 461–475.

214. Blencowe BJ, Ouzounis CA. The PWI motif: a new protein domain in splicing factors. Trends Biochem Sci. 1999;24: 179–180.

215. Zhan X, Qian B, Cao F, Wu W, Yang L, Guan Q, et al. An Arabidopsis PWI and RRM motif-containing protein is critical for pre-mRNA splicing and ABA responses. Nat Commun. 2015;6: 8139.

216. Absmeier E, Rosenberger L, Santos KF, Becke C, Wahl MC. PWI-like domain of *Chaetomium thermophilum* Brr2. 2015.

217. Goldthwaite CD, Cryer DR, Marmur J. Effect of carbon source on the replication and transmission of yeast mitochondrial genomes. Mol & Gen Genet : MGG. 1974;133: 87–104.

218. Williamson DH. The effect of environmental and genetic factors on the replication of mitochondrial DNA in yeast. Symp Soc Exp Biol. 1970;24: 247–276.

219. Birky CW. Effects of glucose repression of the transmission and recombination of mitochondrial genes in yeast (*Saccharomyces cerevisiae*). Genetics. 1975;80: 695–709.

220. Alonso-Del-Real J, Lairón-Peris M, Barrio E, Querol A. Effect of Temperature on the Prevalence of Non Species against a Wine Strain in Wine Fermentation: Competition, Physiological Fitness, and Influence in Final Wine Composition. Front Microbiol. 2017;8: 150.

221. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, et al. The Awesome Power of Yeast Evolutionary Genetics: New Genome Sequences and Strain Resources for the Saccharomyces sensu stricto Genus. G3. 2011;1: 11–25.

222. Dobin A, Davis C a., Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinforma. 2013;29: 15–21.

223. Anders S, Pyl PT, Huber W. HTSeq - A Python framework to work with high-throughput sequencing data. bioRxiv. Cold Spring Harbor Labs Journals; 2014; 0–5.

224. Lui TWH, Tsui NBY, Chan LWC, Wong CSC, Siu PMF, Yung BYM. DECODE: an integrated differential co-expression and differential expression analysis of gene expression data. BMC Bioinforma. 2015;16: 182.

225. Kanehisa M. The KEGG Database. pp. 91–103.

226. Pons P, Latapy M. Computing Communities in Large Networks Using Random Walks. J Graph Algorithms Appl. 2006;10.

227. Picazo C, Gamero-Sandemetrio E, Orozco H, Albertin W, Marullo P, Matallana E, et al. Mitochondria inheritance is a key factor for tolerance to dehydration in wine yeast production. Lett Appl Microbiol. 2015;60: 217–222.

228. Motomura S, Horie K, Kitagaki H. Mitochondrial activity of sake brewery yeast affects malic and succinic acid production during alcoholic fermentation: Mitochondrial activity of sake brewery yeast. J Inst Brew. 2012;118.

229. O'Connor-Cox ESC, Lodolo EJ, Axcell BC. Mitochondrial relevance to yeast fermentation performance. J Inst Brew. 1996;102.

230. de Meaux J, Pop A, Mitchell-Olds T. Cis-regulatory evolution of chalcone-synthase expression in the genus *Arabidopsis*. Genetics. 2006;174: 2181–2202.

231. Guo M, Yang S, Rupe M, Hu B, Bickel DR, Arthur L, et al. Genome-wide

allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. Plant Mol Biol. 2008;66: 551–563.

232. He F, Zhang X, Hu J, Turck F, Dong X, Goebel U, et al. Genome-wide analysis of cis-regulatory divergence between species in the *Arabidopsis* genus. Mol Biol Evol. 2012;29: 3385–3395.

233. Jeong H, Tombort B, Albert R, Oltvai LN, Barabasi AL. The large-scale organization of metabolic networks. The Structure and Dynamics of Networks. Princeton University Press;

234. Keskin O, Gursoy A, Ma B, Nussinov R. Principles of protein-protein interactions: what are the preferred ways for proteins to interact? Chem Rev. ACS Publications; 2008;108: 1225–1244.

235. Vallabhajosyula RR, Chakravarti D, Lutfeali S, Ray A, Raval A. Identifying hubs in protein interaction networks. PloS one. 2009;4: e5344.

236. Nicolau M, Schoenauer M. On the evolution of scale-free topologies with a gene regulatory network model. Bio Syst. 2009;98: 137–148.

237. Nicolau M, Schoenauer M. Evolving scale-free topologies using a Gene Regulatory Network model. 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence). IEEE; pp. 3747–3754.

238. Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. Changes to gene expression associated with hybrid speciation in plants: further insights from transcriptomic studies in Senecio. Philos Trans R Soc London Ser B, Biol Sci. 2008;363: 3055–3069.

239. Hegarty MJ, Barker GL, Brennan AC, Edwards KJ, Abbott RJ, Hiscock SJ. Extreme changes to gene expression associated with homoploid hybrid speciation. Mol Ecol. 2009;18: 877–889.

240. Hegarty MJ, Barker GL, Wilson ID, Abbott RJ, Edwards KJ, Hiscock SJ. Transcriptome shock after interspecific hybridization in senecio is ameliorated by genome duplication. Curr Biol : CB. 2006;16: 1652–1659.

241. Zhang M, Liu X-K, Fan W, Yan D-F, Zhong N-S, Gao J-Y, et al. Transcriptome analysis reveals hybridization-induced genome shock in an interspecific F hybrid

from Camellia. Genome. 2018;61: 477–485.

242. Iurina NP, Odintsova MS. Mitochondrial signaling: retrograde regulation in yeast *Saccharomyces cerevisiae*. Genetika. 2008;44: 1445–1452.

243. Jazwinski SM, Kriete A. The yeast retrograde response as a model of intracellular signaling of mitochondrial dysfunction. Front Physiol. 2012;3: 139.

244. Ballinger SW. Beyond retrograde and anterograde signaling: Mitochondrial - nuclear interactions as a means for evolutionary adaptation and contemporary disease susceptibility. Free Radic Biol Med. 2012;53.

245. Lin Z, Li W-H. The evolution of aerobic fermentation in *Schizosaccharomyces pombe* was associated with regulatory reprogramming but not nucleosome reorganization. Mol Biol Evol. 2011;28: 1407–1413.

246. Lin Z, Wang T-Y, Tsai B-S, Wu F-T, Yu F-J, Tseng Y-J, et al. Identifying cis-regulatory changes involved in the evolution of aerobic fermentation in yeasts. Genome Biol Evol. 2013;5: 1065–1078.

247. Lianou A, Koutsoumanis KP. Effect of the growth environment on the strain variability of Salmonella enterica kinetic behavior. Food Microbiol. 2011;28: 828–837.

248. Knight KP, McKellar RC. Influence of cinnamon and clove essential oils on the D- and z-values of Escherichia coli O157:H7 in apple cider. J Food Prot. 2007;70: 2089–2094.

249. Medina A, Lambert RJW, Magan N. Rapid throughput analysis of filamentous fungal growth using turbidimetric measurements with the Bioscreen C: a tool for screening antifungal compounds. Fungal Biol. 2012;116: 161–169.

250. Cheeseman KE, Denyer SP, Hosein IK, Williams GJ, Maillard J-Y. Evaluation of the bactericidal efficacy of three different alcohol hand rubs against clinical isolates of Staphylococcus aureus using an ex vivo carrier test. J Hosp Infect. 2011;77: 21–24.

251. Toussaint M, Conconi A. High-throughput and sensitive assay to measure yeast cell growth: a bench protocol for testing genotoxic agents. Nat Protoc. 1: 1922–1928.

252. Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, Kschischo M. grofit:

Fitting Biological Growth Curves withR. J Stat Softw. 2010;33.

253. Savitsky TD, Paddock SM. Bayesian Semi- and Non-parametric Models for Longitudinal Data with Multiple Membership Effects in R. J Stat Softw. 2014;57: 1–35.

254. Jung PP, Christian N, Kay DP, Skupin A, Linster CL. Protocols and programs for high-throughput growth and aging phenotyping in yeast. PloS one. 2015;10: e0119807.

255. van der Maaten LJP, Hinton GE. Visualizing High-Dimensional Data Using t-SNE. J Mach Learn Res. 2008;9: 2579–2605.

256. Wang YM, Chew WC. Accelerating the iterative inverse scattering algorithms by using the fast recursive aggregate T-matrix algorithm. Radio Sci. 1992;27.

257. Generalized Additive Models: An Introduction with R. Chapman and Hall/CRC; 2006.

258. Jubany S, Tomasco I, Ponce de León I, Medina K, Carrau F, Arrambide N, et al. Toward a global database for the molecular typing of *Saccharomyces cerevisiae* strains. FEMS yeast Res. 2008;8: 472–484.

259. Fry J. Faculty of 1000 evaluation for Resurrecting ancestral alcohol dehydrogenases from yeast. F1000 - Post-publication peer review of the biomedical literature. 2005.

260. Lakhal L, Rhazi L, Bonhoure J, Aussenac T. Predicting rheological behavior of wheat dough based on machine learning and front-face fluorescence spectroscopy on wheat flour. CFW Plex. 2013;

261. Okser S, Pahikkala T, Airola A, Salakoski T, Ripatti S, Aittokallio T. Regularized machine learning in the genetic prediction of complex traits. PLoS Genet. 2014;10: e1004754.

262. Borgwardt K. Predicting phenotypic effects of gene perturbations in C. elegans using an integrated network model. BioEssays : news Rev Mol Cell Dev Biol. 2008;30: 707–710.

263. Guzzetta G, Jurman G, Furlanello C. A machine learning pipeline for quantitative phenotype prediction from genotype data. BMC Bioinforma. 2010;11 Suppl 8: S3.