



Review/Revue

The rise of yeast population genomics

L'émergence de la génomique des populations de levures

Gianni Liti^a, Joseph Schacherer^{b,*}^a Centre for Genetics and Genomics, University of Nottingham, Nottingham NG7 2UH, UK^b Department of Genetics, Genomics and Microbiology, UMR 7156, University of Strasbourg/CNRS, 28, rue Goethe, 67083 Strasbourg cedex, France

ARTICLE INFO

Article history:

Received 23 November 2010

Accepted after revision 23 March 2011

Available online 1 July 2011

Keywords:

Polymorphisms
Population structure
Resequencing
Genomes
Yeasts

Mots clés :

Polymorphisme
Structure des populations
Reséquencage
Génomes
Levures

ABSTRACT

Genome sequences of multiple individuals are essential to determine the forces shaping sequence variation as well as to understand the relationship between genotype and phenotype. Because of their wide ecological, geographical and genetic diversity, yeast species represent an ideal model system for population genomics. Recently, there has been a renewed interest in characterizing the genetic diversity within yeast species such as *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*. Here, we review recent progress in the exploration of the intraspecific diversity using large collections of yeast isolates. These recent large-scale polymorphism surveys have increased our understanding of the population structures as well as the evolutionary history of the species. In addition, these resources represent a powerful framework for dissecting the relationship between genotype and phenotype.

© 2011 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

R É S U M É

Le séquençage systématique des génomes d'individus d'une même espèce est essentiel pour déterminer les processus évolutifs impliqués dans la variation de séquence mais également pour comprendre la relation existant entre génotypes et phénotypes. De par la diversité de leurs origines écologiques et géographiques, les levures sont des organismes modèles pour la génomique des populations. Récemment, il y a eu un regain d'intérêt pour la caractérisation de la diversité génétique intraspécifique chez les levures, comme *Saccharomyces cerevisiae* et *Saccharomyces paradoxus*. Dans cet article, nous faisons le point sur les études récentes ayant exploré la diversité des génomes de différentes collections de souches de levures. Ces études à grande échelle de la variation intraspécifique ont permis d'avoir une meilleure compréhension de la structure des populations mais également de l'histoire évolutive de ces espèces. De plus, les données générées représentent les fondations pour disséquer la relation existant entre génotypes et phénotypes.

© 2011 Académie des sciences. Publié par Elsevier Masson SAS. Tous droits réservés.

1. Introduction

Genomes vary in sequence as a result of evolution and hence polymorphism data allows us to elucidate the

evolutionary history of species. Moreover, genome-wide investigation of the patterns of polymorphism in a large sample of individuals is the first step to assess the relationship between genotype and phenotype within a population. Large-scale polymorphism surveys and analyses were first reported for a small number of species including *Drosophila melanogaster*, *Arabidopsis thaliana* and *Homo sapiens*.

* Corresponding author.

E-mail address: schacherer@unistra.fr (J. Schacherer).

For *A. thaliana*, the exploration of its genetic diversity began with a study in which 876 fragments (approximately 1% of the genome) across the genomes of 96 isolates, sampled worldwide, were compared by Sanger sequencing [1]. Then, to capture the genome-wide common sequence variation, a high-density array resequencing strategy was used on a subset of 20 accessions [2,3]. Major conclusions from these first studies were that most sequence variants are found worldwide although population structure and isolation by distance are evident. Linkage disequilibrium (LD) decays over about 10 kb. Hence, the average linkage disequilibrium in *A. thaliana* is not very different from that in humans, perhaps surprising, given the selfing nature of this organism. This result might reflect that outcrossing events are common and the effective population size is large. Today, a project with the goal of describing the whole-genome sequence variation in 1001 accessions of *A. thaliana* is underway (<http://1001genomes.org/>) [4,5]. The main motivation for the 1001 genomes project is to have a deeper insight into the genotype-phenotype relationship of this species by using a genome-wide association studies (GWAS) strategy.

In parallel, the release of the reference human genome sequence in 2001 provided the foundation for cataloguing human genetic variation [6]. A few years later, the International HapMap project released a catalogue of polymorphic sites containing approximately 3.1 million single nucleotide polymorphisms (SNPs) [7]. This resource allowed the mapping of thousands of genomic regions linked to disease susceptibility by genome-wide association studies. Today, an international collaboration aims to sequence 2500 human genomes from the five major population groups: Europe, East Asia, South Asia, West Africa and the Americas (<http://www.1000genomes.org/>). The pilot phase of this project was recently completed [8]. As for *A. thaliana*, the objective of this project is to identify large sets of functional polymorphisms that underlie phenotypic variation in multiple human populations.

The complete genome sequence of *S. cerevisiae* was a milestone in the field of genomics in the 1990s [9,10]. The *S. cerevisiae* laboratory strain S288c became the pioneer eukaryotic genome. Ever since, the hemiascomycetous yeasts (the subphylum of fungi that includes *S. cerevisiae*) have been used as model organisms for evolutionary and comparative genomics. Because of the structure of their genome (small and compact), these organisms represent a powerful model for comparative genomics and studies of genome evolution [11]. The availability of genome sequence data represents an unprecedented opportunity to evaluate DNA sequence variation and genome evolution in a phylum spanning a broad evolutionary range. The wealth of these data on interspecific sequence differences stands in contrast to our limited knowledge of variation within yeast species. Nevertheless, significant progress has been made over the last few years to characterize the polymorphic variation within yeast species.

Here, we review recent studies that explored the genetic diversity within large collections of yeast isolates of the same species. Yeast population genomics to date have mainly been focused on two species: *S. cerevisiae* and *S. paradoxus*. These comprehensive genome analyses have

increased our understanding of the population structures as well as the evolutionary history of the species. We first begin with a discussion of the two complementary approaches that focused on either individual or species genomic variation. We then discuss how these polymorphism resources lay the foundation for dissecting the relationship between genotype and phenotype.

2. From comparative genomics to population genomics

By comparing genome sequences at different phylogenetic distances, diverse questions can be addressed. At very large evolutionary distances, broad insights about the gene content as well as evolution of the genome structures can be gleaned. At moderate phylogenetic distances, sequence comparisons allow a deeper insight into the functional parts of the genome and help our understanding of the relationship between genotype and phenotype. Together, comparative and population genomics provide a comprehensive and complementary view of genome sequence variation. Because of their wide ecological, geographical, clinical and industrial distribution and their genome size, yeasts are ideal model organisms to understand genome evolution in different species as well as in natural populations.

Since 2000, yeasts (or more precisely hemiascomycetes) have been at the forefront of comparative genomics [12]. Today, more than 40 hemiascomycetes species are either completely or partially sequenced [11]. Among these sequenced organisms, there are species evolutionary distant from *S. cerevisiae* but also closely related species such as those belonging to the *Saccharomyces sensu stricto* complex. Because of the variable phylogenetic distance between these organisms, yeast genome comparisons were very fruitful. As an example, the comparison of the *Saccharomyces sensu stricto* species led to the mapping of conserved regulatory motifs along the *Saccharomyces* genome [13–15]. Sequencing of species such as *Kluyveromyces waltii* (recently renamed *Lachancea waltii*) and *Ashbya gossypii* provided clear proof of a major event in the history of the hemiascomycetous subphylum: the whole genome duplication event in the lineage leading to *S. cerevisiae* [16,17]. Finally, comparisons of more distant species represented a powerful model for studies of genome evolution [18,19].

Nevertheless, a single genome sequence is not representative of a species. It is now known that *S. cerevisiae* strains can vary in Ty element [20,21], gene content [22] and similar findings have been found for other species, including humans. Of particular importance is the observation that some genes with fundamental effects on life-history traits are not even present or functional in the S288c reference strain [23]. In most of the comparative genomic studies, it is, therefore, sometimes difficult to know if the observed genomic variation is either strain or species specific.

Now, new molecular methods allow the characterization of genetic variation at the whole-genome level between and within a species. The generation of high-density arrays suitable for whole-genome variation detection was the first major technological breakthrough

[2,24]. Today, next-generation sequencing technology provides a novel opportunity for collecting genome-scale sequence data [8]. In the near future, this current technological revolution in sequencing will allow the exploration of large numbers of new species as well as to have a better insight into natural populations. All these strategies lay the foundation for the evolving field of population genomics.

3. *Saccharomyces cerevisiae* population genomics

Previous studies have aimed for a global view of variation across the *S. cerevisiae* genome through a variety of different means. Nucleic acid polymorphism among *S. cerevisiae* isolates has been documented using different molecular markers [25,26], identifying single features polymorphisms (SFPs) [27], by multilocus sequence analysis [28–31] or using selected segregating sites [32]. While these strategies offer some insight into allelic variation between *S. cerevisiae* strains, they provide polymorphism information for only a very small fraction of the genome, limiting our ability to make inference about the general structure of a population. Hence, the next step was to obtain a precise measure of the diversity within *S. cerevisiae* species and to compare whole genomes of a large population at the nucleotide level.

A first exploration of the yeast whole-genome variation at the nucleotide level focused on seven commonly used laboratory *S. cerevisiae* strains: A364A, W303, FL100, CEN.PK, Σ 1278b, SK1 and BY4716 [33]. Genomic polymorphism maps were generated using high-density arrays. A major conclusion of this study was that all the studied laboratory strains except SK1 are derivatives of the S288c reference strain. In fact, these genomes are mosaics of large regions identical to S288c interspersed with small regions of high sequence divergence.

Nevertheless to have a better insight into natural population variation, two major projects generated genome-wide maps at the nucleotide level of large *S. cerevisiae* collections, sampled from a diverse array of sources (beer, bread, vineyards, immunocompromised individuals, various fermentations and soil) and from different continents (Fig. 1). Two different strategies were used: high-density arrays [24] and low coverage whole genome sequencing [22]. In the first study, a total of 1.89 million single nucleotide polymorphisms (SNPs) grouped into 101,343 distinct segregating sites were identified in a sample of 63 *S. cerevisiae* strains. In the second study, the *Saccharomyces* Genome Resequencing Project (SGRP) described 235,127 SNPs and 14,051 indels from 1–4 fold, or more, coverage whole genome sequences of 36 *S. cerevisiae* and 35 *S. paradoxus* strains. The SGRP created a genome assembly for each strain by developing an iterative process called Parallel ALignment and ASsembly (PALAS). This process implemented an imputation method to computationally infer missing (or poor quality) information (both nucleotide and INDELS) from related strains by means of a method based on ancestral recombination graphs. This sequence survey also discovered 38 new hypothetical open reading frame (mostly subtelomeric) that are absent in the S288c reference genome.

Both analyses provided a deep insight into the population structure as well as the evolutionary history of this yeast species. From the sequence polymorphisms survey [24], the authors observed clear population structure at the level of major ecological subgroups. These data strongly supported the presence of different clusters such as strains from vineyards. These clusters represent separate domestication events, but *S. cerevisiae* as a whole is not domesticated. The SGRP analysis proposed the presence of five “clean” lineages (specific to geographic location or ecological niches), with the majority of segregating sites private to one population and uniformly distributed along the genome, as well as many recombinant (mosaics) strains originated from the various clean lineages. An alternative scenario to the domestication process was also proposed, with humans having utilised natural existing variants for different fermentation processes and offered opportunity to interbreeding (generating the mosaics). Surveying additional strains is needed to fully resolve the roles of ecology versus geography in the genetic differentiation of this species.

The determination of linkage disequilibrium (LD) – non random association of alleles at two or more loci – also provided information about recombination and evolutionary history of the species. Linkage disequilibrium falls to half of its maximum value at about 11 kb [24]. Nevertheless, the architecture of linkage disequilibrium is variable within the different subpopulations: laboratory, wine, clinical and distillery strains. Most of the laboratory strains are derivatives of the sequenced reference strain S288c. Their genomes are mosaics of large regions identical to S288c interspersed with small regions of high sequence divergence. As a result, high linkage disequilibrium can be observed. Linkage disequilibrium falls to half of its maximum value at about 23.8 kb. By contrast, the low level of linkage disequilibrium (about 2.5 kb) in the wine strains probably reflects the extended length of time since the most recent common ancestor of these strains, and perhaps a higher frequency of outcrossing events. Finally, the linkage disequilibrium decay and the fact that segregating sites are located every ~100–200 bp suggest that *S. cerevisiae* could be a good model organism for genome-wide association studies.

In parallel to these large-scale polymorphism surveys, genomes of several *S. cerevisiae* strains were sequenced at high coverage (Fig. 1). These studies focused on strains from different genetic backgrounds: a clinical strain (YJM789), wine strains (RM11-1a and EC1118) a strain involved in biofuel production (JAY291) and laboratory strains (SK1, Y55, W303) [22,23,34,35]. Exploration and extensive analysis of these complete genomes was very fruitful. The genome of the wine yeast EC1118 differs from other *S. cerevisiae* strains by three large regions originated from either a closely related or a non-*Saccharomyces* species [23]. These introgressions encompass 34 genes and potentially play a key function in fermentation, such as metabolism of sugars and nitrogen. Similarly, the genome of the JAY270 strain pointed out specific gene polymorphisms and introgressions important for bioethanol production. These could explain desirable phenotypes such as ethanol and cell mass production as well as high

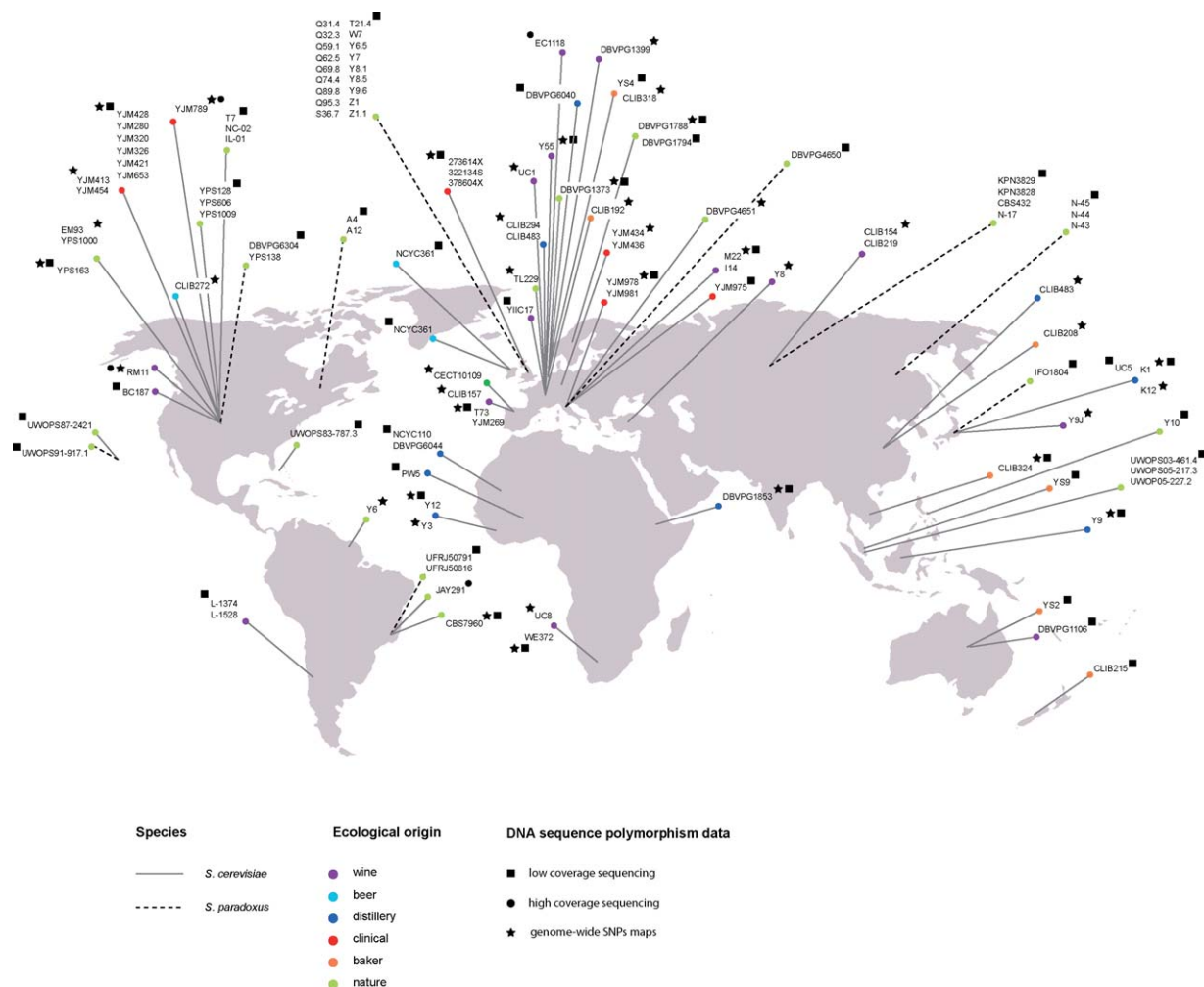


Fig. 1. Exploration of the intraspecific diversity of yeast isolates. Geographical and ecological origins of yeast strains that were used in large-scale polymorphism surveys. The line denotes the species and circle color denotes ecological niche as specified in the key.

temperature and oxidative stress tolerance [34]. In addition to this genotype-phenotype relationship exploration, whole genome sequences allowed a better insight into the yeast life cycle. Based on the distribution of segments of shared genealogy among three strains: YJM789, RM11 and S288 C, it was estimated that only 314 outcrossing events have occurred during approximately 16 million cell divisions [36]. This results show that outcrossing is relatively infrequent in *S. cerevisiae*: roughly once every 50,000 divisions.

4. *Saccharomyces paradoxus*: a rising evolutionary model

In the field of yeast population genomics, *S. paradoxus* is becoming an attractive model organism. *S. paradoxus* is the closest known species to *S. cerevisiae* and is remarkably similar in terms of genome organization and physiology. Most of the *S. paradoxus* strains were isolated from the bark and surrounding soil of oak trees, where it can co-exist with its sibling species *S. cerevisiae* [37,38]. *S. paradoxus* has

been isolated from many locations worldwide and it is considered as a non domesticated wild species.

Initially, multilocus sequence analysis was used to infer phylogenetic relationships between *S. paradoxus* geographic subpopulations [28,39,40]. These studies indicated that strains within this species are both genetically divergent and partially reproductively isolated. In addition, three highly diverged geographic subpopulations were identified: Europe, Far East Asia and America lineages.

Subsequently, a sequence variation study scaled up on the entire third chromosome (approximately 2.3% of the genome) [41]. The authors analysed the sequence of 20 isolates from two subpopulations: 12 from Europe and eight from Far East Asia. This allowed a precise quantification of the life cycle. Mutational and recombinational diversity along this chromosome clearly shows that *S. paradoxus* is primarily asexual (like *S. cerevisiae*) [41,42]. In both subpopulations, the sexual cycle occurs approximately every 1000 asexual generations. Moreover,

these population genomic data have allowed the identification of recombination hotspots. Interestingly, multiple recombination hotspots are conserved between *S. paradoxus* and *S. cerevisiae* probably as a result of the low frequency of sex in these two yeasts [42].

The SGRP initiative sequenced 35 *S. paradoxus* strains, representative of the major geographical subpopulations, with the aim of drawing comparisons with *S. cerevisiae* (Fig. 1). Half of the strains were isolated from the same geographical location (England) as representatives of a single recombining population. Genome sequences revealed a previously unknown fourth diverged lineage from Hawaii. The sequence divergence between lineages is variable: 1.2% (Far Eastern/European), 2.3% (American/Hawaii) and 3.7% (European-Far Eastern/North America-Hawaii). Genetic diversity in *S. paradoxus* is considerable higher than in *S. cerevisiae*. The large majority of *S. paradoxus* SNPs are private to each subpopulation resulting in a marked population structure that follows geographic boundaries. The South American isolates (previously regarded as *S. cariocanus*) show the highest Ty counts [22], consistent with the rapid accumulation of chromosomal translocation in this lineage [43]. Genome-wide analysis also confirmed the presence of a large region (23 kb) introgressed from *S. cerevisiae* into the European subpopulation [28], with identical break-points in all strains, perhaps originated from a single event. Finally, this study also showed that *S. paradoxus* strains have limited phenotypic variation compared to *S. cerevisiae* [22]. This observation stands in contrast with the genetic diversity. A possible explanation could be that *S. cerevisiae* strains exhibit a higher phenotypic variation because they occupy a larger ecological niche.

5. Dissecting the functional variants

The genetic differences among individuals lead to broad quantitative phenotypic variation [44]. Dissecting the genetic mechanisms underlying this natural phenotypic variation is a major challenge in modern biology. There are two major approaches to mapping the quantitative differences in phenotypic traits: linkage analysis and genome-wide association studies. In the first case, the causative loci are mapped using the segregating progeny of crosses between genetically divergent strains. The aim is to identify segregating genetic variants that contribute to phenotypic variation in progenies. In contrast, genome-wide association studies use a large sample of unrelated individuals from the same population.

In the past decade, *S. cerevisiae* has become a primary model for dissecting the complex architecture of quantitative traits using linkage mapping. The genetic basis of a number of interesting phenotypes have been studied in yeast, including growth at high temperature [45], gene expression [46], and response to drugs [47]. Nevertheless, these studies focused on pairs of specific strains and systematically used the S288c laboratory strain (or its derivative BY). They were instrumental and laid the foundation of yeast forward genomics but the whole genetic diversity of a large population has not yet been systematically explored. The new *Saccharomyces* population genomic

datasets help to characterize the genetic diversity of the species. In fact, these data can facilitate the mapping of functional variants in multiple ways. The characterised genomic relationships among strains can guide an accurate experimental design for both linkage and association mapping.

Recent studies have used strains obtained from multiple lineages and sampled a larger fraction of the variation. Two strains, representative of diverged wine and North America oak populations, were recently used to accurately dissect the variation in the efficiency of sporulation [48]. The authors found that the observed difference is mostly explained by allelic variation of three transcription factors: *IME1*, *RME1*, and *RSF1*. Interestingly, this study nicely illustrates how genetic interactions between transcription factors might have an impact on phenotypic diversity.

A new collection of genetically tractable parentals [49] and F1 segregants from four highly diverged *S. cerevisiae* are also available [50]. Using this set of F1 segregants, linkage analysis clearly showed that subtelomeric regions play a major role in quantitative natural variation, supporting their importance in adaptive variation.

Furthermore, 16 strains were selected and crossed in all pairwise combination to create a large library of F1 hybrids for genetic analysis such as heterosis [51]. Finally, SNPs dataset aid the prediction of functional and deleterious variants using a conservation based approach like the one implemented by SIFT [52].

6. Conclusions and perspectives

The advent of next generation sequencing technologies has changed the face of genomics and given access to population genomic data. Although yeast has proven important in developing essential tools for analysing multiple individuals, this field is only just beginning. A new set of genome sequenced 27 strains is now available (<http://www.genetics.wustl.edu/jflab/data4.html>) and deep coverage (40–50X) of 42 SGRP *S. cerevisiae* and *S. paradoxus* strains will soon be released (Fig. 1). Expanding the genome-wide sequence datasets will increase the number of variants and add power for genome-wide association approach. In addition, the availability of multiple *S. paradoxus* wild isolates characterised at the genome level offer an attractive model to study ecology and evolution. Furthermore, most of the genetic and molecular techniques as well as the QTL mapping are readily available in *S. paradoxus* [53] and a deletion collection is underway.

The characterization of SNPs is only the first step toward characterising the genetic make-up of a natural population. The high coverage sequencing data will allow characterization of heterozygous diploid genomes without the need of creating homozygous derivatives. This is a major step forward given the majority of *Saccharomyces* isolates are diploids or have a more complex ploidy configuration (e.g. aneuploidy, polyploids). However, the ability to sporulate and generate viable gametes remains an essential feature for subsequent forward and reverse genetics analysis and should be considered during selection of strains. The high

genome coverage will enable detection of copy number differences, perhaps relevant for genome-wide association studies.

A major challenge that also remains is to produce a complete *de novo* end-to-end assembly. The subtelomeric regions are particularly difficult due to their repetitive nature [54], intrinsic high genomic instability [55] and functional divergence [56]. Resolving the structure of chromosome ends is crucial as they contain many genes involved in secondary metabolisms and, therefore, play key roles in individual variation. The generation of full genome assemblies will also yield a complete picture for other polymorphisms such as structural variants and their impact in evolution and fitness can be determined.

Finally, parallel explorations of new species as well as individuals from the same species will offer a precise view of the evolution of the genotype-phenotype relationship in yeast. Today, yeast population genomic studies are underway in many other species including *Schizosaccharomyces pombe*, *Lachancea kluyveri* and *Candida albicans*. These new datasets will allow specific analysis within species as well as drawing of important parallels between species. Yeasts with their compact and well-characterised genome are likely to play a key role in the rising population genomic field.

Disclosure of interest

The author declare that he has no conflicts of interest concerning this article.

Acknowledgements

The authors are grateful to Dr C. Nieduszynski and M. Hawkins for critical reading of the manuscript. GL is funded by The Biotechnology and Biological Sciences Research Council (grant number BB/F015216/1). JS was supported by a CNRS PEPS grant.

References

- [1] M. Nordborg, T.T. Hu, Y. Ishino, J. Jhaveri, C. Toomajian, H. Zheng, E. Bakker, P. Calabrese, J. Gladstone, R. Goyal, M. Jakobsson, S. Kim, Y. Morozov, B. Padhukasahasram, V. Plagnol, N.A. Rosenberg, C. Shah, J.D. Wall, J. Wang, K. Zhao, T. Kalbfleisch, V. Schulz, M. Kreitman, J. Bergelson, The pattern of polymorphism in *Arabidopsis thaliana*, *PLoS Biol* 3 (2005) e196.
- [2] R.M. Clark, G. Schweikert, C. Toomajian, S. Ossowski, G. Zeller, P. Shinn, N. Warthmann, T.T. Hu, G. Fu, D.A. Hinds, H. Chen, K.A. Frazer, D.H. Huson, B. Scholkopf, M. Nordborg, G. Ratsch, J.R. Ecker, D. Weigel, Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*, *Science* 317 (2007) 338–342.
- [3] S. Kim, V. Plagnol, T.T. Hu, C. Toomajian, R.M. Clark, S. Ossowski, J.R. Ecker, D. Weigel, M. Nordborg, Recombination and linkage disequilibrium in *Arabidopsis thaliana*, *Nat Genet* 39 (2007) 1151–1155.
- [4] D. Weigel, R. Mott, The 1001 genomes project for *Arabidopsis thaliana*, *Genome Biol* 10 (2009) 107.
- [5] S. Ossowski, K. Schneeberger, R.M. Clark, C. Lanz, N. Warthmann, D. Weigel, Sequencing of natural strains of *Arabidopsis thaliana* with short reads, *Genome Res* 18 (2008) 2024–2033.
- [6] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Showkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendt, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordtsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jiang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kasprzyk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M. Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrino, M.J. Morgan, P. de Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y.J. Chen, Initial sequencing and analysis of the human genome, *Nature* 409 (2001) 860–921.
- [7] K.A. Frazer, D.G. Ballinger, D.R. Cox, D.A. Hinds, L.L. Stuve, R.A. Gibbs, J.W. Belmont, A. Boudreau, P. Hardenbol, S.M. Leal, S. Pasternak, D.A. Wheeler, T.D. Willis, F. Yu, H. Yang, C. Zeng, Y. Gao, H. Hu, W. Hu, C. Li, W. Lin, S. Liu, H. Pan, X. Tang, J. Wang, W. Wang, J. Yu, B. Zhang, Q. Zhang, H. Zhao, J. Zhou, S.B. Gabriel, R. Barry, B. Blumenstiel, A. Camargo, M. Defelice, M. Faggart, M. Goyette, S. Gupta, J. Moore, H. Nguyen, R.C. Onofrio, M. Parkin, J. Roy, E. Stahl, E. Winchester, L. Ziaugra, D. Altshuler, Y. Shen, Z. Yao, W. Huang, X. Chu, Y. He, L. Jin, Y. Liu, W. Sun, H. Wang, Y. Wang, X. Xiong, L. Xu, M.M. Waye, S.K. Tsui, H. Xue, J.T. Wong, L.M. Galver, J.B. Fan, K. Gunderson, S.S. Murray, A.R. Oliphant, M.S. Chee, A. Montpetit, F. Chagnon, V. Ferretti, M. Leboeuf, J.F. Olivier, M.S. Phillips, S. Romy, C. Sallee, A. Verner, T.J. Hudson, P.Y. Kwok, D. Cai, D.C. Koboldt, R.D. Miller, L. Pawlikowska, P. Taillon-Miller, M. Xiao, L.C. Tsui, W. Mak, Y.Q. Song, P.K. Tam, Y. Nakamura, T. Kawaguchi, T. Kitamoto, T. Morizono, A. Nagashima, Y. Ohnishi, A. Sekine, T. Tanaka, T. Tsunoda, P. Deloukas, C.P. Bird, M. Delgado, E.T. Dermitzakis, R. Gwilliam, S. Hunt, J. Morrison, D. Powell, B.E. Stranger, P. Whittaker, D.R. Bentley, M.J. Daly, P.I. de Bakker, J. Barrett, Y.R. Chretien, J. Maller, S. McCarroll, N. Patterson, I. Pe'er, A. Price, S. Purcell, D.J. Richter, P. Sabeti, R. Saxena, S.F. Schaffner, P.C. Sham, P. Varilly, L.D. Stein, L. Krishnan, A.V. Smith, M.K. Tello-Ruiz, G.A. Thorisson, A. Chakravarti, P.E. Chen, D.J. Cutler, C.S. Kashuk, S. Lin, G.R. Abecasis, W. Guan, Y. Li, H.M. Munro, Z.S. Qin, D.J. Thomas, G. McVean, A. Auton, L. Bottolo, N. Cardin, S. Eyheramendy, C. Freeman, J. Marchini, S. Myers, C. Spencer, M. Stephens, P. Donnelly, L.R. Cardon, G. Clarke, D.M. Evans, A.P. Morris, B.S. Weir, J.C. Mullikin, S.T. Sherry, M. Feolo, A. Skol, H. Zhang, I. Matsuda, Y. Fukushima, D.R. Macer, E. Suda, C.N. Rotimi, C.A. Adebamowo, I. Ajayi, T. Aniagwu, P.A. Marshall, C. Nkwodimma, C.D. Royal, M.F. Leppert, M. Dixon, A. Peiffer, R. Qiu, A. Kent, K. Kato, N. Niikawa, I.F. Adewole, B.M. Knoppers, M.W. Foster, E.W. Clayton, J. Watkin, D. Muzny, L. Nazareth, E. Sodergren, G.M. Weinstock, I. Yakub, B.W. Birren, R.K. Wilson, L.L. Fulton, J. Rogers, J. Burton, N.P. Carter, C.M. Clee, M. Griffiths, M.C. Jones, K. McLay, R.W. Plumb, M.T. Ross, S.K. Sims, D.L. Willenby, Z. Chen, H. Han, L. Kang, M. Godbout, J.C. Wallenburg, P. L'Archeveque, G. Bellemare, K. Saeki, D. An, H. Fu, Q. Li, Z. Wang, R. Wang, A.L. Holden, L.D. Brooks, J.E. McEwen, M.S. Guyer, V.O. Wang, J.L. Peterson, M. Shi, J. Spiegel, L.M. Sung, L.F. Zacharia, F.S. Collins, K. Kennedy, R. Jamieson, J. Stewart, A second generation human haplotype map of over 3.1 million SNPs, *Nature* 449 (2007) 851–861.
- [8] R.M. Durbin, G.R. Abecasis, D.L. Altshuler, A. Auton, L.D. Brooks, R.A. Gibbs, M.E. Hurles, G.A. McVean, A map of human genome variation from population-scale sequencing, *Nature* 467 (2010) 1061–1073.

- [9] A. Goffeau, B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston, E.J. Louis, H.W. Mewes, Y. Murakami, P. Philippsen, H. Tettelin, S.G. Oliver, Life with 6000 genes, *Science* 274 (1996) 547–563.
- [10] H.W. Mewes, K. Albermann, M. Bahr, D. Frishman, A. Gleissner, J. Hani, K. Heumann, K. Kleine, A. Maierl, S.G. Oliver, F. Pfeiffer, A. Zollner, Overview of the yeast genome, *Nature* 387 (1997) 7–65.
- [11] B. Dujon, Yeast evolutionary genomics, *Nat Rev Genet* 11 (2010) 512–524.
- [12] J. Souciet, M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara, E. Bon, P. Brottier, S. Casaregola, J. de Montigny, B. Dujon, P. Durrens, C. Gaillardin, A. Lepingle, B. Llorente, A. Malpertuy, C. Neuveglise, O. Ozier-Kalogeropoulos, S. Potier, W. Saurin, F. Tekai, C. Toffano-Nioche, M. Wesolowski-Louvel, P. Wincker, J. Weissenbach, Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies, *FEBS Lett* 487 (2000) 3–12.
- [13] P. Cliften, P. Sudarsanam, A. Desikan, L. Fulton, B. Fulton, J. Majors, R. Waterston, B.A. Cohen, M. Johnston, Finding functional features in Saccharomyces genomes by phylogenetic footprinting, *Science* 301 (2003) 71–76.
- [14] M. Kellis, N. Patterson, M. Endrizzi, B. Birren, E.S. Lander, Sequencing and comparison of yeast species to identify genes and regulatory elements, *Nature* 423 (2003) 241–254.
- [15] C.A. Nieduszynski, Y. Knox, A.D. Donaldson, Genome-wide identification of replication origins in yeast by comparative genomics, *Genes Dev* 20 (2006) 1874–1879.
- [16] M. Kellis, B.W. Birren, E.S. Lander, Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*, *Nature* 428 (2004) 617–624.
- [17] F.S. Dietrich, S. Voegeli, S. Brachat, A. Lerch, K. Gates, S. Steiner, C. Mohr, R. Pohlmann, P. Luedi, S. Choi, R.A. Wing, A. Flavier, T.D. Gaffney, P. Philippsen, The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome, *Science* 304 (2004) 304–307.
- [18] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. De Montigny, C. Marck, C. Neuveglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthonard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J.M. Beckerich, E. Beyne, C. Bleykasten, A. Boisrame, J. Boyer, L. Cattolico, F. Confaniolieri, A. De Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J.M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G.F. Richard, M.L. Straub, A. Suleau, D. Swennen, F. Tekai, M. Wesolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, J.L. Souciet, Genome evolution in yeasts, *Nature* 430 (2004) 35–44.
- [19] G. Fischer, E.P. Rocha, F. Brunet, M. Vergassola, B. Dujon, Highly variable rates of genome rearrangements between hemiascomycetous yeast lineages, *PLoS Genet* 2 (2006) e32.
- [20] A. Gabriel, J. Dapprich, M. Kunkel, D. Gresham, S.C. Pratt, M.J. Dunham, Global mapping of transposon location, *PLoS Genet* 2 (2006) e212.
- [21] G. Liti, A. Peruffo, S.A. James, I.N. Roberts, E.J. Louis, Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu stricto* complex, *Yeast* 22 (2005) 177–192.
- [22] G. Liti, D.M. Carter, A.M. Moses, J. Warringer, L. Parts, S.A. James, R.P. Davey, I.N. Roberts, A. Burt, V. Koufopanou, I.J. Tsai, C.M. Bergman, D. Bensasson, M.J. O'Kelly, A. van Oudenaarden, D.B. Barton, E. Bailes, A.N. Nguyen, M. Jones, M.A. Quail, I. Goodhead, S. Sims, F. Smith, A. Blomberg, R. Durbin, E.J. Louis, Population genomics of domestic and wild yeasts, *Nature* 458 (2009) 337–341.
- [23] M. Novo, F. Bigey, E. Beyne, V. Galeote, F. Gavory, S. Mallet, B. Cambon, J.L. Legras, P. Wincker, S. Casaregola, S. Dequin, Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118, *Proc Natl Acad Sci U S A* 106 (2009) 16333–16338.
- [24] J. Schacherer, J.A. Shapiro, D.M. Ruderfer, L. Kruglyak, Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*, *Nature* 458 (2009) 342–345.
- [25] M. Azumi, N. Goto-Yamamoto, AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces sensu stricto* and its application to phenetic clustering, *Yeast* 18 (2001) 1145–1154.
- [26] C. Hennequin, A. Thierry, G.F. Richard, G. Lecointre, H.V. Nguyen, C. Gaillardin, B. Dujon, Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains, *J Clin Microbiol* 39 (2001) 551–559.
- [27] E.A. Winzeler, C.I. Castillo-Davis, G. Oshiro, D. Liang, D.R. Richards, Y. Zhou, D.L. Hartl, Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays, *Genetics* 163 (2003) 79–89.
- [28] G. Liti, D.B. Barton, E.J. Louis, Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*, *Genetics* 174 (2006) 839–850.
- [29] E. Aa, J.P. Townsend, R.I. Adams, K.M. Nielsen, J.W. Taylor, Population structure and gene evolution in *Saccharomyces cerevisiae*, *FEMS Yeast Res* 6 (2006) 702–715.
- [30] J.C. Fay, J.A. Benavides, Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*, *PLoS Genet* 1 (2005) 66–71.
- [31] M.J. Ayoub, J.L. Legras, R. Saliba, C. Gaillardin, Application of multi locus sequence typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon, *J Appl Microbiol* 100 (2006) 699–711.
- [32] G. Ben-Ari, D. Zenvirith, A. Sherman, G. Simchen, U. Lavi, J. Hillel, Application of SNPs for assessing biodiversity and phylogeny among yeast strains, *Heredity* 95 (2005) 493–501.
- [33] J. Schacherer, D.M. Ruderfer, D. Gresham, K. Dolinski, D. Botstein, L. Kruglyak, Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains, *PLoS ONE* 2 (2007) e322.
- [34] J.L. Argueso, M.F. Carazzolle, P.A. Mieczkowski, F.M. Duarte, O.V. Netto, S.K. Missawa, F. Galzerani, G.G. Costa, R.O. Vidal, M.F. Noronha, M. Dominska, M.G. Andrietta, S.R. Andrietta, A.F. Cunha, L.H. Gomes, F.C. Tavares, A.R. Alcarde, F.S. Dietrich, J.H. McCusker, T.D. Petes, G.A. Pereira, Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production, *Genome Res* 19 (2009) 2258–2270.
- [35] W. Wei, J.H. McCusker, R.W. Hyman, T. Jones, Y. Ning, Z. Cao, Z. Gu, D. Bruno, M. Miranda, M. Nguyen, J. Wilhelmy, C. Komp, R. Tamse, X. Wang, P. Jia, P. Luedi, P.J. Oefner, L. David, F.S. Dietrich, Y. Li, R.W. Davis, L.M. Steinmetz, Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789, *Proc Natl Acad Sci U S A* 104 (2007) 12825–12830.
- [36] D.M. Ruderfer, S.C. Pratt, H.S. Seidel, L. Kruglyak, Population genomic analysis of outcrossing and recombination in yeast, *Nat Genet* 38 (2006) 1077–1081.
- [37] J.P. Sampaio, P. Goncalves, Natural populations of *Saccharomyces kudriavzevii* in Portugal are associated with oak bark and are sympatric with *S. cerevisiae* and *S. paradoxus*, *Appl Environ Microbiol* 74 (2008) 2144–2152.
- [38] P.D. Sniegowski, P.G. Dombrowski, E. Fingerman, *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* 1 (2002) 299–306.
- [39] V. Koufopanou, J. Hughes, G. Bell, A. Burt, The spatial scale of genetic differentiation in a model organism: the wild yeast *Saccharomyces paradoxus*, *Philos Trans R Soc Lond B Biol Sci* 361 (2006) 1941–1946.
- [40] H.A. Kuehne, H.A. Murphy, C.A. Francis, P.D. Sniegowski, Allopatric divergence, secondary contact, and genetic isolation in wild yeast populations, *Curr Biol* 17 (2007) 407–411.
- [41] I.J. Tsai, D. Bensasson, A. Burt, V. Koufopanou, Population genomics of the wild yeast *Saccharomyces paradoxus*: quantifying the life cycle, *Proc Natl Acad Sci U S A* 105 (2008) 4957–4962.
- [42] I.J. Tsai, A. Burt, V. Koufopanou, Conservation of recombination hotspots in yeast, *Proc Natl Acad Sci U S A* 107 (2010) 7847–7852.
- [43] G. Fischer, S.A. James, I.N. Roberts, S.G. Oliver, E.J. Louis, Chromosomal evolution in *Saccharomyces*, *Nature* 405 (2000) 451–454.
- [44] T.F. Mackay, E.A. Stone, J.F. Ayroles, The genetics of quantitative traits: challenges and prospects, *Nat Rev Genet* 10 (2009) 565–577.
- [45] L.M. Steinmetz, H. Sinha, D.R. Richards, J.I. Spiegelman, P.J. Oefner, J.H. McCusker, R.W. Davis, Dissecting the architecture of a quantitative trait locus in yeast, *Nature* 416 (2002) 326–330.
- [46] J. Ansel, H. Bottin, C. Rodriguez-Beltran, C. Damon, M. Nagarajan, S. Fehrmann, J. Francois, G. Yvert, Cell-to-cell stochastic variation in gene expression is a complex genetic trait, *PLoS Genet* 4 (2008) e1000049.
- [47] E.O. Perlsstein, D.M. Ruderfer, D.C. Roberts, S.L. Schreiber, L. Kruglyak, Genetic basis of individual differences in the response to small-molecule drugs in yeast, *Nat Genet* 39 (2007) 496–502.
- [48] J. Gerke, K. Lorenz, B. Cohen, Genetic interactions between transcription factors cause natural variation in yeast, *Science* 323 (2009) 498–501.
- [49] F.A. Cubillos, E.J. Louis, G. Liti, Generation of a large set of genetically tractable haploid and diploid *Saccharomyces* strains, *FEMS Yeast Res* 9 (2009) 1217–1225.
- [50] F.A. Cubillos, E. Billi, E. Zörgö, L. Parts, F.P., S. Omholt, A. Blomberg, J. Warringer, E.J. Louis, G. Liti, Assessing the complex architecture of polygenic traits in diverged yeast populations, *Mol Ecol* 20 (2010) 1401–1413.

- [51] W.E. Timberlake, M.A. Frizzell, K.D. Richards, R.C. Gardner, A new yeast genetic resource for analysis and breeding, *Yeast* 28 (2011) 63–80.
- [52] S.W. Doniger, H.S. Kim, D. Swain, D. Corcuera, M. Williams, S.P. Yang, J.C. Fay, A catalog of neutral and deleterious polymorphism in yeast, *PLoS Genet* 4 (2008) e1000183.
- [53] G. Liti, S. Haricharan, F.A. Cubillos, A.L. Tierney, S. Sharp, A.A. Bertuch, L. Parts, E. Bailes, E.J. Louis, Segregating YKU80 and TLC1 alleles underlying natural variation in telomere properties in wild yeast, *PLoS Genet* 5 (2009) e1000659.
- [54] G. Liti, E.J. Louis, Yeast evolution and comparative genomics, *Annu Rev Microbiol* 59 (2005) 135–153.
- [55] K.T. Nishant, W. Wei, E. Mancera, J.L. Argueso, A. Schlattl, N. Delhomme, X. Ma, C.D. Bustamante, J.O. Korbel, Z. Gu, L.M. Steinmetz, E. Alani, The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis, *PLoS Genetics* 6 (2010) e1001109.
- [56] C.A. Brown, A.W. Murray, K.J. Verstrepen, Rapid expansion and functional divergence of subtelomeric gene families in yeasts, *Curr Biol* 20 (2010) 895–903.