Species-wide survey reveals the complex landscape of the genetic origin of reproductive isolation within natural yeast populations

Jing Hou

UNIVERSITÉ DE STRASBOURG

ÉCOLE DOCTORALE VIE ET SANTÉ

Laboratoire de Génétique Moléculaire, Génomique et Microbiologie - UMR7156

THÈSE présentée par :

Jing HOU

soutenue le : **24 mars 2016**

pour obtenir le grade de : **Docteur de l'université de Strasbourg**

Discipline/Spécialité : **Science de la vie et de la santé / Aspect moléculaire et cellulaire de la biologie**

Multiplicité des origines génétiques de l'isolement reproductif au sein de populations naturelles de levures

THÈSE dirigée par :

M. SCHACHERER Joseph	Maître de Conférences HDR, Université de Strasbourg

RAPPORTEURS :

M. WOLFE Kenneth	Professeur, University College Dublin
Mme. DELNERI Daniela	Professeur, University of Manchester

AUTRES MEMBRES DU JURY :

Mme. BLANDIN Stéphanie	Chargé de Recherche, INSERM, Strasbourg
M. DUJON Bernard	Professeur, Université de Pierre et Marie Curie & Institut Pasteur, Membre de l'Académie des sciences

Thèse de Doctorat de l'Université de Strasbourg

Discipline : Science du Vivant

Spécialité : Biologie Moléculaire et Génétique

Multiplicité des origines génétiques de l'isolement reproductif au sein de populations naturelles de levures

Présentée par

Jing Hou

pour l'obtention du grade de Docteur de l'Université de Strasbourg

UNIVERSITÉ DE STRASBOURG

Soutenue le 24 mars 2016, devant la commission d'examen :

Dr. Joseph Schacherer — **Directeur de thèse**
Maître de Conférences HDR, Université de Strasbourg

Pr. Kenneth Wolfe — **Rapporteur externe**
Professeur, University College Dublin, Ireland

Pr. Daniela Delneri — **Rapporteur externe**
Professeur, University of Manchester, UK

Dr. Stéphanie Blandin — **Examinateur interne**
Chargé de Recherche, INSERM, Strasbourg

Pr. Bernard Dujon — **Membre invité**
Professeur, Université de Pierre et Marie Curie & Institut Pasteur,
Membre de l'Académie des sciences

Doctoral dissertation from the University of Strasbourg

Discipline: Life Science

Specialty: Molecular Biology and Genetics

Species-wide survey reveals the complex landscape of the genetic origin of reproductive isolation within natural yeast populations

By

Jing Hou

in partial fulfillment of the requirements of a degree of Doctor of Philosophy

March 24, 2016

Members of the committee:

Dr. Joseph Schacherer **Thesis advisor**
Assistant Professor, University of Strasbourg

Pr. Kenneth Wolfe **External reporter**
Professor, University College Dublin, Ireland

Pr. Daniela Delneri **External reporter**
Professor, University of Manchester, UK

Dr. Stéphanie Blandin **Internal examiner**
Researcher, INSERM, Strasbourg

Pr. Bernard Dujon **Invited member**
Professor, University of Pierre and Marie Curie & Institut Pasteur,
Member of the Institut de France

Abstract

Elucidating the genetic origin of phenotypic diversity among individuals within the same species is essential to understand evolution. By combining classical genetic analyses and high-throughput genomic strategies, we performed species-wide surveys and dissected in depth the molecular basis of the onset of reproductive isolation - a phenotype that constitutes a key step in the formation of new species - across the yeast *Saccharomyces cerevisiae*. We showed that the raw potential of reproductive isolation could readily segregate at the intraspecific level, which is governed by various molecular mechanisms ranging from large-scale chromosomal changes to incompatible epistatic genetic interactions. While phenotypes like reproductive isolation are cryptic and can only be revealed by testing different combinations of parental backgrounds, other phenotypes such as monogenic Mendelian traits are thought to be simple in terms of their phenotypic penetrance and genetic constitution. However, our survey showed that the expressivity of monogenic mutations and hence the inheritance pattern of a Mendelian trait could also depend on parental combinations, transitioning from simple to complex trait due to the presence of modifiers and genetic interactions in specific genetic backgrounds. Overall, using the power of genetics, our studies unveiled the multiplicity and complexity of the genetic origin of phenotypes within a population, from the origin of reproductive isolation to the hidden complexity of Mendelian traits.

Acknowledgments

First, I would like to thank all members of the committee, Pr. Kenneth Wolfe, Pr. Daniela Delneri, Pr. Bernard Dujon and Dr. Stephanie Blandin for accepting to evaluate this work. Special gratitude goes to Dr. Blandin for already being a member of my mid-thesis committee, and to Pr. Dujon, with whom I was fortunate to have had the chance to discuss my project and results on multiple occasions during my Ph.D.

This work was completed at the department of genetics, genomics and microbiology, UMR 7156/CNRS, University of Strasbourg, under the supervision of Dr. Joseph Schacherer in the group intraspecific variation and genome evolution.

I would like to thank Pr. Serge Potier, for welcoming me into the department as the director at the time, and authorizing me to defend my thesis as the current director of the doctoral school of life science ED 414. Serge, thank you for coming with me and Joseph to Paris. We had fancy Chinese food and met the Chinese ambassador. It was a fun day.

I would like to thank my family, especially my mom, for the inconstant but durable support they've given me throughout the nine years I spent in France. To this day, my grandma still wonders how much school I have left to do in order to get a government job back home. My answer to her? I just change the subject!

I would like to thank my friends back in China. Zou Dewen, my best friend, you've always believed in me, helped me and supported my choices, personally or otherwise. With you back home I know I have less things to worry about. Jiang Hao, we came to France on the same plane, and have been good friends ever since. Thanks for taking care of me after I got my head open by a car! Wang Dahai, my buddy in class for two years in Le Mans. Zhang Cong and Zhang Yuanliang, you guys endured the first year of Master with me in Strasbourg. The "Dongbei" gang, Wang Chunliang, Yu Wei, Lin Lei, talking to you was pleasantly nostalgic. Ray Wong, thank you for all the weekends and holidays that you've sneaked me into the lab of Barbara, you were my first introduction to yeast research! Liu Zhenhua, my neighbor on the 5th floor,

thank you for our discussions on the rooftop, best of luck on your research and your little family. And finally, Bai Xueke, getting to know you has been such a pleasant experience. What can I say? Thank you for being in my life, and please come back!

Big thanks, of course, to all members of the group, former or current. Paul, thank you for teaching me so many bad jokes and curses in multiple languages and dialects. But seriously, thank you for being the big brother in the lab, I've learned a lot from you ever since I started as a master student in the group. Kelle, thank you for all the adventures we had together. The highlight must be driving around Sicily without my passport, and on the way back, got stranded in a weird van together with a weird guy in a weird place called "Woippy", at 2 o'clock in the morning. Ok scratch that, the best will always be Legoland! And speaking of Legoland, thanks Christian, *a.k.a.* Dr. Loulou Brion, together we built something cool there, although it might have been more fun without all the little kids around "stealing" our pieces!

The best moments with Loulou, however, will always be the heated discussions and "fights" we had over science. I remember the frightened face of Laure when we argued in front of her, that was priceless. Was it about the definition of QTLs? Mosaic population structure? Mac over PC? I can't even remember. Anyway, thanks Loulou, discussing with you was inspiring and enriching, and those are the memories I shall always cherish. Heck, to show my gratitude, I'll from now on submit my sequencing data to your database - the Extractable Nuclear Antigen (ENA)!

Thanks to all the people who have contributed to this work, Anne, David, Jean-Seb, Jackson, and especially Anastasie, without whom the *PDR1* story wouldn't exist. Anastasie, thank you for all the fun conversations we had during our little trips to Illkirch. I remember the days when we started our trainings together in the lab. Best of luck on finishing your thesis! David, thank you for all your help with R, it has been always a pleasure working with you. And, Anne, where do I start, you were always there for me when I needed help, advice or comfort (and of course data). It's hard to imagine not being around you. Under whose desk am I going to hide now? And just for the record, you will always be my favorite teacher in biochemistry.

To Claudia, the “maman” at the bench. You are such a meticulous person, any lab will be lucky to have you! Thank you for ordering toys for me and “little Loulou”, he’s my favorite little buddy at the age group 3 to 10. To the youngsters who call Claudia “maman” in the lab, Fifou, Elodie and Teo, your dynamisms always amaze me, keep it up! Thank you Teo for helping me with the figures and other stuffs. Do something great, you are the one.

To all the people who have been around that I failed to mention, Jacky, Jean-Luc, Claudine, Cyrielle, and many others. Also to M. Kammerer. The conversations with him, scientific or not, were always interesting.

Writing this thesis has been a personal recollection of the journey I took on. Among all the things I know I could have done better, that I wish I could’ve improved upon, choosing my advisor is definitely not one of them. Boss, I saved you for the last, because I can’t even start to express how much you mean to me, how much you’ve made me grown. To all the days and nights discussing science with you, all the hours we spent staring on the white board and trying to find solutions to an interesting problem. To every “eureka” and every disappointment. The four years I spent as your student have been the best experience I could ever imagine. To me, nothing gives the same level of excitement as the pursuit of science, and I was so lucky to have you on my side, to be guided by you. You’ve shown me what it is to be a good scientist: to be driven, determined, creative and resilient. You are and always will be my mentor. Every little achievement I’ve made, every progress I have now, and every success I may or may not have in the future, all will be impossible without you. You raised me up, to more than I can be. Thank you Joseph. Hope I can make you proud some day.

Table of Contents

STATE OF THE ART

The various flavors of reproductive isolation in yeast

Introduction

Speciation, the evolutionary process by which new species emerge, lies at the heart of the observed biodiversity. Under the biological species concept, the formation of new species requires the establishment of reproductive barriers that limit the gene flow among populations[1]. In other words, new species form when individuals from diverging populations become reproductively isolated and unable to produce viable or fertile offspring, eventually allowing nascent species to be genetically and phenotypically distinct. While this concept is widely applied for sexually reproducing organisms, it is not until the past two decades that precise molecular characterizations of the genetic basis of reproductive isolation have become possible[2].

Reproductive isolation can act prior to mating (pre-zygotic), which prevent the formation of a zygote; or soon after mating (post-zygotic) leading to reduced offspring viability or fertility[1]. While many external factors, such as differences in life history and temporal patterns may cause pre-zygotic isolation[1], genetic analyses mainly focused on intrinsic post-zygotic isolation[2,3]. During the past years, much progress has been made on the subject, leading to the identification of multiple isolating mechanisms between closely related species in various taxa[2-9]. However, contrasting to the perception of its mechanistic multiplicity, only a few examples have been characterized to the molecular level, and the tempo and mode of reproductive isolation were still poorly understood. Are the identified mechanisms the original cause of reproductive isolation, or just a consequence of subsequent divergence within nascent species? Which types of genetic changes are of particular interest in the onset of reproductive isolation? What is the relative role of selection *vs.* drift through initial stage to the completion reproductive isolation?

To address these questions, it is essential to systematically explore the onset and accumulation of reproductive isolation at various evolutionary scales over a broad taxonomic range. Within the past few years, such efforts have started to be undertaken (Table 1). In fact, with the increased availability of large collections of isolates from various species, cases of partial reproductive isolation at the intraspecific scale were recently observed in model systems such as *Drosophila*[10-12], *Arabidopsis*[13-15], *Caenorhabditis*[16-18] and *Saccharomyces*[19-24]. While the number of cases identified is still low, it has been clear that the raw potential for

speciation segregates readily within populations, which seems to be a rule rather than an exception in a broad context.

Here, we briefly recapitulate the genetic origins of intrinsic post-zygotic reproductive isolation in major model organisms including *A. thaliana*[13,25], *C. elegans*[16,26], *D. melanogaster*[10] and their close relatives[11,12,27-29]. We then concentrate on yeasts and conduct a more comprehensive review on the current state of the genetic basis of post-zygotic reproductive isolation in the *Saccharomyces* genus and recent advances at the intraspecific scales within multiple species of this group[21,22,30], especially in *S. cerevisiae*[19,20,23,24]. We focus on the mechanistic diversity as well as their underlying evolutionary origins that act intra-specifically, and try to provide a comparative view on the onset of reproductive isolation along a continuum of genetic differentiation, which encompasses intraspecific populations, recent delineating nascent species as well as closely related sister species of the same subphylum.

Brief overview of reproductive isolation in different model organisms

On the conceptual ground, the most prominent genetic explanation of intrinsic post-zygotic reproductive isolation is the presence of genetic incompatibilities, popularized in the 1940s by Theodosius Dobzhansky and Hermann Müller[31,32]. The hitherto known as the Dobzhansky-Müller model posits that populations could evolve independently and accumulate different mutations that are well adapted in their original genetic backgrounds but do not function properly together in hybrids. The loss of viability or fertility in the offspring may simply be caused by the accumulation of such incompatible mutations, which arose as a by-product of genomic differentiation[33]. Not only that this model offers an elegant solution on how genetic basis for reproductive isolation could originate from an inter-mating population, it also integrates the notion that incompatible alleles may accumulate with increased genomic divergence[33,34]. Examples of such incompatible gene pairs have been identified between closely related species in various taxa[2-9], and more recently among populations of the same species in several major genetic models such as *Drosophila melanogaster*[10,11], *Arabidopsis thaliana*[13,25], *Caenorhabditis elegans*[16,26] (Table 1).

Different evolutionary forces could putatively explain the observed incompatibilities. Adaptive processes such as niche specialization to pathogens were of particular importance in the evolution of plant immune systems, where defense-related genes acquired in different populations could cause hybrid

	Species pair	Evidence	Genes	Chromosome	Phenotype	References
Arabidopsis	*A. thaliana* ×*A. thaliana*	TRD	*HPA1,HPA2*	autosome	viability	25
	A. thaliana × *A. thaliana*	Diallele cross	*DM1-9,SRF3*	autosome	fitness	13,35
	A. lyrata × *A. lyrata*	TRD	-	autosome	viability	15
Caenorhabditis	*C. elegans* × *C. elegans*	TRD	*PEEL1, ZEEL1*	autosome	viability	37
	C. elegans × *C. elegans*	TRD	-	autosome	viability	17
	C. briggsae × *C. briggsae*	Cybrid	-	cyto-nuclear	fitness	16
	C. briggsae × *C. briggsae*	TRD	-	autosome	viability	18
	C. briggsae × *C. nigoni*	Introgression	-	X-linked	viability	28
Drosophila	*D. melanogaster* × *D. melanogaster*	TRD	-	autosome	viability	10
	D. melanogaster × *D. simulans*	Suppressor	*Lhr, Hmr, gzgf*	X-linked	viability	11,12

Table 1. Evidences of reproductive isolation within and between species in model organisms

necrosis through autoimmune responses[13,14,35,36]. Many other cases were related to neutral processes such as genetic drift or the propagation of selfish genetic elements[25,37]. For example, reciprocal inactivation of a duplicated essential gene pair *HPA1/HPA2* in natural accessions of *A. thaliana* could lead to seed abortion in the F2 offspring when none of the functional copies were present[25]. Differences in recombination rates or mutation loads could also put emphasizes on certain types of mechanisms, for example cyto-nuclear incompatibilities involving interacting genes located on mitochondrial and nuclear genomes[7,16,24,38-40]. Such cyto-nuclear incompatibilities were found to cause F2 sterility in interspecific yeast hybrids[38,40], and could be a common cause of hybrid weakness in *Drosophila*[39].

Besides incompatibilities at the gene level, large genomic changes could also lead to post-zygotic reproductive isolation[1]. For example, differences of the ploidy level or chromosome numbers among parental species were common in causing reproductive isolation in plants[41] and animals[1], where unbalanced gene dosages in the offspring could lead to inviability or sterility. Between *A. thaliana* and its sister species *A. lyrata*, differences in chromosome numbers (5 for *A. thaliana* and 8 for *A. lyrata*) were accounted for the observed reproductive isolation, where F1 hybrids were viable but sterile[9]. Other localized chromosomal rearrangements, for example translocations and inversions, have also drawn much interest, as parental species that differ by such structural variation would most likely produce offspring with unequal distribution of essential genes upon meiosis[1]. The role of chromosomal rearrangements is indeed well established leading to post-zygotic reproductive isolation in various taxa, especially in plants[42] and *Drosophila*[43].

However, while extremely insightful, reproductive isolation studies in complex model systems suffered from several drawbacks. Studies using such models often restricted to a low number of parental combinations due to experimental workloads, therefore large-scale analysis has rarely been undertaken[13]. Moreover, due to considerable genome size and complexity, precise identifications of the molecular mechanisms involved are challenging especially for structural variations such as inversions and translocations. As a result, only a low number of cases have been fully characterized to date, and an overview of the relative importance of different mechanisms to the onset and propagation of reproductive isolation across the observed natural diversity in these species is far from reached.

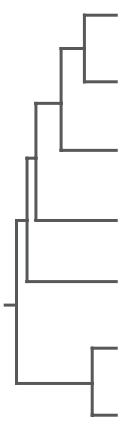
Yeasts, ideal model to explore inter- and intraspecific reproductive isolation

An emerging model system, which allows the integration of the genetic and genomic diversity within and between closely related species, is the budding yeast *S. cerevisiae* and its close relatives. Compare to other complex models, yeasts present numerous advantages due to their short generation time, small and compact genomes and laboratory amenable sexual reproductions. Rather than relying on a low number of crosses and transmission ratio distortion in the offspring as it is the case for complex organisms, yeasts offer the possibility to systematically examine a large number of crosses and use pooled mapping strategies that require much less sequencing efforts. Natural populations of multiple yeasts species can be isolated from various biotopes and geographical locations[44,45]. The vast natural distribution with the ever-growing availability of whole genome sequencing data make yeasts an ideal model system to obtain a comprehensive view on how reproductive isolation emerges at different evolutionary scales by taking into account the roles of ecology, domestication and other selective or neutral processes.

Reproductive isolation in the *Saccharomyces* genus

The *Saccharomyces sensu stricto* complex comprises seven known species (*S. cerevisiae, S. paradoxus, S. arboricolus, S. kudiavzevii, S. mikatae, S. eubayanus and S. uvarum*) to date, all of which could cross to form viable hybrids under laboratory conditions, with no evident mating preferences[6,46,47] (Figure 1). Although outcrossing events are rare[46,48], natural hybrids between closely related *Saccharomyces* species are readily observed[49-51]. Many of the hybrids were found to be involved in industrial related processes, such as beer[50] and wine making[51]. Among *Saccharomyces* species, introgressions are frequently reported, for example between *S. cerevisiae* and *S. paradoxus*[52]. Recent evidence indicates that such events could occur recurrently across multiple populations within a species[22]. In fact, genome-wide screen in a large number of *S. uvarum* isolates identified multiple chromosomal regions ascribed to different *Saccharomyces* species, such as *S. kudriavzevii*, *S. cerevisiae* and *S. eubayanus*[22].

Strong post-zygotic reproductive isolations are observed between species within this complex. Interspecific hybrids typically yield less than 1% of viable meiotic offspring in most parental combinations[6], and ~7% for the least diverged species pair *S. eubayanus* and *S. uvarum*[46]. High levels of DNA sequence divergence are considered as the main cause of loss of hybrid offspring viability,



	Intraspecific divergence %	# of crosses	# of offspring per cross	Offspring viability range	Origin of reproductive isolation
S. cerevisiae	0.1 - 1.5	> 100	80 - 1000	10% - 100%	Chromosomal rearrangement; mutator effect; genetic incompatibility; cyto-nulear interaction; uncharacterized
S. paradoxus	0.1 - 4.9	> 100	20 - 100	5% - 95%	Chromosomal rearrangement; uncharacterized
S. mikatae	-	-	-	-	-
S. kudriavzevii	0.1 - 4.2	15	20	0% - 88%	uncharacterized
S. arboricolus	-	-	-	-	-
S. eubayanus	-	-	-	-	-
S. uvarum	0.3 - 6.3	15	200	7.3% - 95%	uncharacterized

Figure 1. Intraspecific divergence and evidences of post-zygotic reproductive isolation within species of the *Saccharomyces* genus

which impairs proper chromosomal segregation through mechanism of anti-recombination by the mismatch repair system (MMR)[53,54]. In fact, viable hybrid offspring often show high numbers of aneuploidies and reduced recombination rate[55,56], the effect of which could be rescued by deleting components of the MMR[54,55].

It is widely admitted that the degree of post-zygotic isolation is correlated with the level of divergence between the parental pair, as the effect of anti-recombination progresses with increased DNA sequence divergence[6,46,53]. However, this overly simplified generalization might be due to sampling bias of the parental isolate pairs. As a matter of fact, it is increasingly evident that multiple mechanisms leading to reproductive isolation operate concurrently at both intra- and interspecific levels. The correlation between sequence divergence and reproductive isolation could then be plausible only at the maximum level of offspring viability when other mechanisms were absent, and likely to play a relatively minor role to the initial stage of reproductive isolation.

For example, besides sequence divergence, chromosomal rearrangements[56,57] and cyto-nulear incompatibilities[38,40] also contribute to the observed post-zygotic reproductive isolation in the *Saccharomyces* genus. Classic Dobzhansky-Müller incompatibilities could also play a role, although no clear examples have been found so far[58-60]. In fact, autotetraploid hybrids between *S. cerevisiae* and *S. paradoxus* (with two copies of each parental genome) showed high offspring viability contrasting to diploid hybrids, strongly suggesting that dominant

Dobzhansky-Müller incompatibilities were not present between this species pair[58]. Nevertheless, while chromosome replacements in *S. cerevisiae* with its *S. paradoxus* homologs in haploids were mostly viable[60], the existence of complex recessive epistatic interactions impacting offspring fertility and fitness cannot be ruled out[61].

Nevertheless, as interspecific reproductive isolation is nearly complete in *Saccharomyces*, it has been difficult to disentangle the effect of simple divergence from functional genetic differentiation. As a result, the role of mechanisms identified using interspecific approaches remains largely indecisive at the incipient stage of reproductive isolation[56,57], and recent works have turned their focus on intraspecific studies using natural populations within the same species of *Saccharomyces* yeasts.

Evidences of intraspecific reproductive isolation within yeast natural populations

With the increasing availability of whole genome sequencing data, multiple yeast species have become the workhorses for functional and evolutionary genomic studies[22,62-65]. Evidences of intraspecific reproductive isolation leading to offspring loss upon crosses were quite frequently observed within collection of isolates of various yeast species[53,66-68]. However, such cases were often found when generating recombined offspring for linkage mapping and were usually dismissed as "annoying crosses" without further dissection of the underlying causes. Again, these observations indicate that mechanisms leading to reproductive isolation segregate readily at the intraspecific scale. Using various approaches, many recent studies addressed specifically how such mechanisms could emerge and lead to intraspecific reproductive isolation, which will be discussed individually in the following section (Figure 2).

The role of chromosomal rearrangements in intrinsic post-zygotic isolation

Although it is well accepted that large chromosomal rearrangements such as translocations and inversions could contribute at least partly to the observed offspring loss in *Saccharomyces* hybrids[56,57], their role at the incipient stage of speciation has received much debate[56]. For one reason is that the overall distribution of chromosomal rearrangements usually does not correlate with the

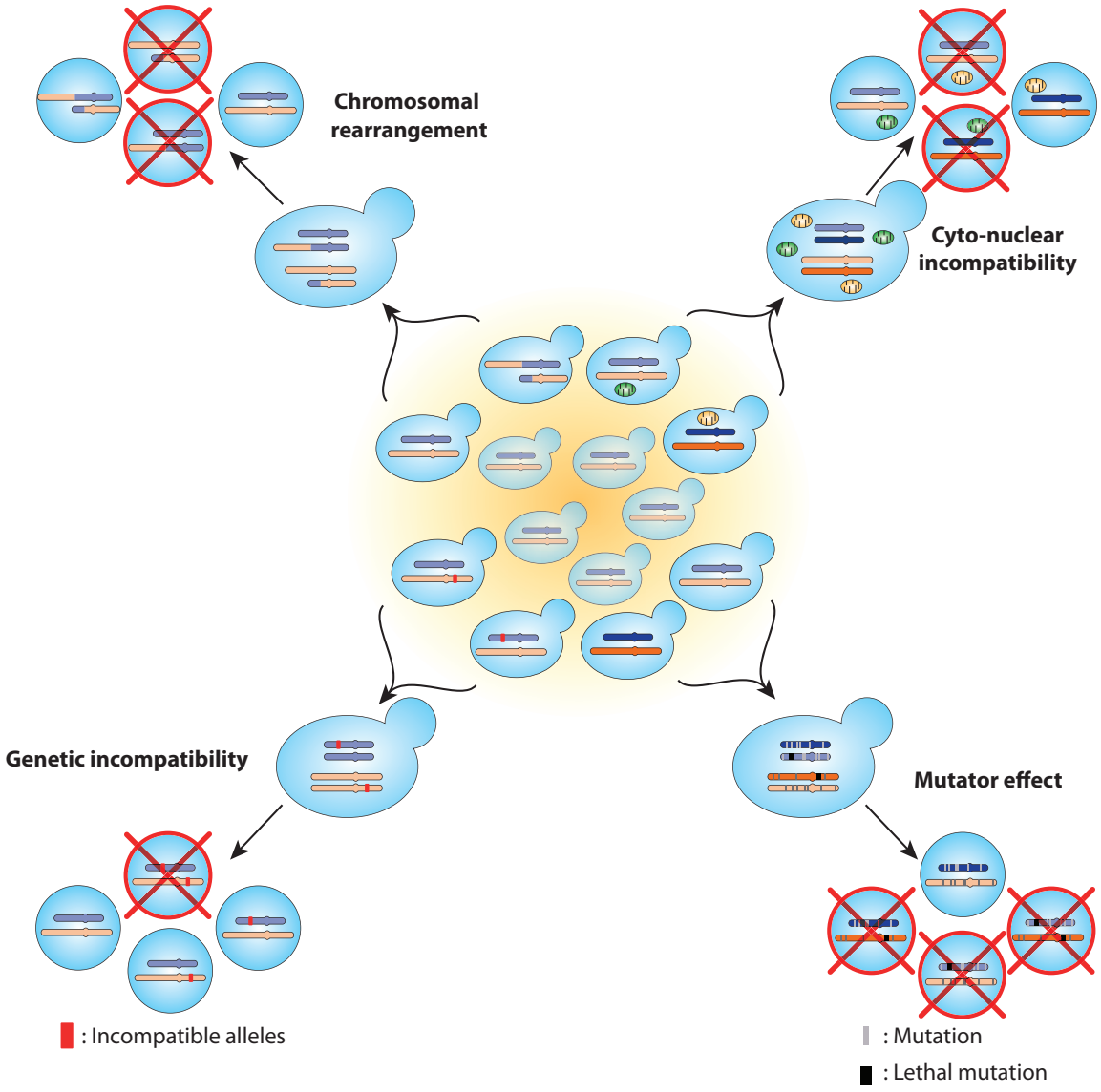


Figure 2. Molecular mechanisms leading to reproductive isolation within natural populations of *S. cerevisiae*. Courtesy of Teo Fournier

level of reproductive isolation or with the scales of genetic divergence observed[56,57]. In fact, many species pairs within the *Saccharomyces sensu stricto* complex harbor individuals with collinear genomes but are completely reproductively isolated[6]. Moreover, artificially generated collinear parental pairs by reverting the observed translocation between the *S. cerevisiae* and *S. mikatae* species clearly showed that translocation events only have a marginal effect on the loss of offspring viability[56].

Paradoxically, it appears that while genomic configurations were sometimes conserved between species, individuals from the same species could be surprisingly diverse in terms of chromosomal profiles[19,21]. When studying a

large collection of natural isolates in *S. cerevisiae*, 3 different types of translocation were identified in 10 out of 60 isolates, which explains the total effect of reproductive isolation observed[19]. Similar observations have been made in *S. paradoxus* populations, where the level of chromosomal rearrangements was partly but significantly correlated with the level of reproductive isolation across 25 isolates[21]. Most identified rearrangements arise through neutral events such as ectopic recombination between repetitive sequences like *Ty* elements, with few exceptions that were adaptive in specific environmental contexts by altering expression patterns of genes present at the junction of the rearranged regions[69].

While chromosomal rearrangements cannot be the only explanation of reproductive isolation observed in many *Saccharomyces* species pairs, there has been a recent example illustrating that such rearrangements could be directly involved during incipient speciation in allopatric populations of North American *S. paradoxus*[30]. In this case, two allopatric populations separated by glaciation differing by a translocation and an inversion in their genomes, gave rise to a hybrid population upon secondary contact. These rearrangements were then fixed in the hybrids and introgressed by repeated backcrossing events with one of the parental population that does not have these rearrangements. Eventually, the hybrid population became reproductively isolated with both parental species through time, illustrating that speciation through chromosomal rearrangement could indeed be possible in yeast[30].

Cyto-nuclear incompatibility and offspring fitness

Compiling evidences suggest that cyto-nuclear incompatibilities could play a disproportionate role during speciation. Incompatible combinations between mitochondrial and nuclear genomes were found to lead to hybrid problems in a wide range of species[16,39], including yeast[38,40]. In fact, as mitochondrial genomes were more prone to mutations, nuclear genomes have to evolve accordingly because proper interactions between the two were often essential for survival or fitness[7]. Because of such constant arms race between mitochondrial and nuclear genomes, cyto-nuclear incompatibilities were more likely to evolve as diverging population could take different trajectories of cyto-nuclear co-evolution. Within the *Saccharomyces* genus, examples of cyto-nuclear incompatibilities were found between *S. cerevisiae* and *S. paradoxus* as well as between *S. cerevisiae* and *S. uvarum* (formerly *S. bayanus*), each with independent origins[40]. For example, the incompatibility between *S. cerevisiae* and *S. paradoxus* was due to inefficient splicing of the *COX1* intron in *S. paradoxus* by the *S. cerevisiae* version of the *MRS1* gene,

which arose with the loss of corresponding intron in *S. cerevisiae*. In all cases, the observed cyto-nuclear incompatibility dampens offspring respiratory capacities, leading to partial sterility[40].

At the intraspecific level, no evident cases of specific cyto-nuclear gene pairs leading to reproductive isolation were found so far. However, global epistasis between nuclear and mitochondrial genomes were commonly observed leading to increased phenotypic variance within *S. cerevisiae*[24]. By testing pairwise combinations of mitochondrial and nuclear genomes in 10 divergent isolates, it was shown that novel combinations often lead to reduced fitness, and the effect of which was not correlated with the level of genetic divergence between the tested pairs. These observations suggest co-evolutions between mitochondrial and nuclear genomes were already significant at the intraspecific scale, however whether such observed fitness variation could eventually lead to reproductive isolation remains unclear.

Antagonistic effects of mutator phenotype related to mismatch repair

In addition to mitochondrial and nuclear genome pairs, genes in the nuclear genome could also co-evolve and lead to hybrid problems. One classic example was illustrated by the mismatch repair system (MMR) in *S. cerevisiae*. Using allelic survey across a number of natural isolates, it has been shown that specific combinations of the *PMS1* and *MLH1* genes, essential players of MMR system, were conserved across the species and possibly maintained by balancing selection[70]. When the original combinations are disrupted, interaction between incompatible allelic pairs could result in a mutator phenotype due to mal-functioning in the MMR. Accumulation of undesired mutations could then lead to sporadic offspring loss, the effect of which could depend on specific genetic backgrounds[71].

Other than offspring loss related to the accumulation of deleterious mutations, the mutator phenotype could also lead to accelerated adaptation in stress conditions due to increased mutation rates[23]. In fact, it was shown that strains with incompatible combination of *PMS1* and *MLH1* thrives more rapidly on high osmotic stress condition by acquiring advantageous mutations in the *PMR1* gene earlier than strains with compatible combinations.

In principle, such mutator phenotype could be considered as a special form of Dobzhansky-Müller incompatibility. However, the effect of incompatible allelic

combination results in increased mutation rates and only act indirectly on offspring viability. The most curious feature of this type of interaction is that the incompatibility could lead to opposite effects on offspring viability depending on different environmental contexts. The interplay between genetic interactions and environmental selections could therefore be important in the onset of reproductive isolation, at least in this specific configuration.

Condition specific genetic incompatibilities and the role of selection

Similarly, different environmental conditions could also have an impact on the effect of classic Dobzhansky-Müller incompatibilities in yeast. Genetic incompatibilities related to negative epistatic interactions were mostly invisible on permissive laboratory conditions that optimize growth in *S. cerevisiae*[19]. However, by taking into account of different environmental factors, such interactions were much more common than previously thought, leading to condition specific reproductive isolations. In fact, systematic survey across 25 crosses on 20 conditions revealed over 24% of the cases showing offspring loss with various severities, the effect of which were specific to independent crosses and conditions[20]. Using segregation analysis followed by pooled sequencing strategy, the first example of two loci Dobzhansky-Müller incompatibility within a yeast species was identified related to respiratory conditions. In this case, the incompatibility was due to a nonsense mutation in the *COX15* gene and a tRNA suppressor *SUP7*, leading to the loss of 25% of the offspring in conditions that require respiration[20]. Both mutations were extremely rare across natural populations in this species, although there were some evidences suggesting that the specific derived combination were maintained by positive selection[20].

Interestingly, most identified cases of negative epistasis appeared not to be related to two loci interactions but instead showing a higher genetic complexity even at the intraspecific scale[20,72]. Moreover, despite the relatively high frequency of occurrence, most incompatible cases were not shared across different isolates, indicating unique genetic origins[72]. These observations highlight again the role of environmental selection to the onset of reproductive isolation in yeast. It is worth noting that natural populations within the same species govern raw speciation potential through condition specific epistatic interactions. Nevertheless, precise molecular dissection of more such incompatibility cases is still required to get a global view of types and distributions of genes involved.

Conclusion and perspectives

Yeasts along with other major genetic model organisms provide unique comparative insights into the genetic basis of post-zygotic reproductive isolation across a broad evolutionary scale. In particular, systematic exploration by looking at large natural yeast populations across multiple environmental conditions was particular useful and fruitful to dissect the mechanistic complexity of reproductive isolation. Nevertheless, despite significant advances, intraspecific reproductive isolation is still underexplored. In particular, most molecular exploration of reproductive isolation cases were restricted to *S. cerevisiae*, and even so many evident cases in *S. cerevisiae* still have to be characterized. Would the patterns be different in other yeast species where the level of intraspecific genetic diversity is usually higher? How relevant are the mechanisms found in yeast to the onset of post-zygotic reproductive isolation in other species in general? What are the roles of the genetic bases of reproductive isolation in shaping other phenotypic traits? Further explorations of the natural population diversity across a broad taxonomic range will be promising to provide some answers to these questions, and may impart deeper understandings regarding the patterns and constraints of genetic differentiation as well as their role in speciation and biodiversity.

References

1. Coyne, J.A. & Orr, H.A. *Speciation*, xiii, 545, Sinauer Associates, Sunderland (2004).
2. Presgraves, D.C. The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175-80 (2010).
3. Maheshwari, S. & Barbash, D.A. The genetics of hybrid incompatibilities. *Annu Rev Genet* **45**, 331-55 (2011).
4. Chen, C. *et al.* A two-locus interaction causes interspecific hybrid weakness in rice. *Nat Commun* **5**, 3357 (2014).
5. Ouyang, Y. & Zhang, Q. Understanding reproductive isolation based on the rice model. *Annu Rev Plant Biol* **64**, 111-35 (2013).
6. Greig, D. Reproductive isolation in *Saccharomyces*. *Heredity (Edinb)* **102**, 39-44 (2009).
7. Chou, J.Y. & Leu, J.Y. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays* **32**, 401-11 (2010).
8. Zanders, S.E. *et al.* Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *Elife* **3**, e02630 (2014).
9. Bomblies, K. & Weigel, D. *Arabidopsis*: a model genus for speciation. *Curr Opin Genet Dev* **17**, 500-4 (2007).
10. Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. & Ayroles, J.F. Genetic incompatibilities are widespread within species. *Nature* **504**, 135-7 (2013).
11. Phadnis, N. *et al.* An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. *Science* **350**, 1552-5 (2015).
12. Brideau, N.J. *et al.* Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. *Science* **314**, 1292-5 (2006).
13. Chae, E. *et al.* Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* **159**, 1341-51 (2014).
14. Alcazar, R. *et al.* Analysis of a plant complex resistance gene locus underlying immune-related hybrid incompatibility and its occurrence in nature. *PLoS Genet* **10**, e1004848 (2014).
15. Leppala, J., Bokma, F. & Savolainen, O. Investigating incipient speciation in *Arabidopsis lyrata* from patterns of transmission ratio distortion. *Genetics* **194**, 697-708 (2013).
16. Chang, C.C., Rodriguez, J. & Ross, J. Mitochondrial-Nuclear Epistasis Impacts Fitness and Mitochondrial Physiology of Interpopulation *Caenorhabditis briggsae* Hybrids. *G3 (Bethesda)* **6**, 209-19 (2015).
17. Snoek, L.B. *et al.* Widespread genomic incompatibilities in *Caenorhabditis elegans*. *G3 (Bethesda)* **4**, 1813-23 (2014).
18. Ross, J.A. *et al. Caenorhabditis briggsae* recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. *PLoS Genet* **7**, e1002174 (2011).
19. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* **24**, 1153-9 (2014).
20. Hou, J., Friedrich, A., Gounot, J.S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat Commun* **6**, 7214 (2015).
21. Charron, G., Leducq, J.B. & Landry, C.R. Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Mol Ecol* **23**, 4362-72 (2014).
22. Almeida, P. *et al.* A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nat Commun* **5**, 4044 (2014).

23. Bui, D.T., Dine, E., Anderson, J.B., Aquadro, C.F. & Alani, E.E. A Genetic incompatibility accelerates adaptation in yeast. *PLoS Genet* **11**, e1005407 (2015).

24. Paliwal, S., Fiumera, A.C. & Fiumera, H.L. Mitochondrial-nuclear epistasis contributes to phenotypic variation and coadaptation in natural isolates of *Saccharomyces cerevisiae*. *Genetics* **198**, 1251-65 (2014).

25. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**, 623-6 (2009).

26. Seidel, H.S. *et al.* A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol* **9**, e1001115 (2011).

27. Matute, D.R., Gavin-Smyth, J. & Liu, G. Variable post-zygotic isolation in *Drosophila melanogaster/D. simulans* hybrids. *J Evol Biol* **27**, 1691-705 (2014).

28. Bi, Y. *et al.* A Genome-wide hybrid incompatibility landscape between *Caenorhabditis briggsae* and *C. nigoni*. *PLoS Genet* **11**, e1004993 (2015).

29. Kozlowska, J.L., Ahmad, A.R., Jahesh, E. & Cutter, A.D. Genetic variation for postzygotic reproductive isolation between *Caenorhabditis briggsae* and *Caenorhabditis sp. 9*. *Evolution* **66**, 1180-95 (2012).

30. Leducq, J.-B. *et al.* Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1**, 15003 (2016).

31. Muller, H. Isolating mechanisms, evolution and temperature. in *Biol. Symp* Vol. 6 71-125 (1942).

32. Dobzhansky, T. *Genetics and the origin of species*, xvi, 364 p. Columbia Univ. Press (1937).

33. Seehausen, O. *et al.* Genomics and the origin of species. *Nat Rev Genet* **15**, 176-92 (2014).

34. Nosil, P., Harmon, L.J. & Seehausen, O. Ecological explanations for (incomplete) speciation. *Trends Ecol Evol* **24**, 145-56 (2009).

35. Alcazar, R., Garcia, A.V., Parker, J.E. & Reymond, M. Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc Natl Acad Sci U S A* **106**, 334-9 (2009).

36. Bomblies, K. & Weigel, D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet* **8**, 382-93 (2007).

37. Seidel, H.S., Rockman, M.V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589-94 (2008).

38. Lee, H.Y. *et al.* Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065-73 (2008).

39. Rand, D.M., Clark, A.G. & Kann, L.M. Sexually antagonistic cytonuclear fitness interactions in *Drosophila melanogaster*. *Genetics* **159**, 173-87 (2001).

40. Chou, J.Y., Hung, Y.S., Lin, K.H., Lee, H.Y. & Leu, J.Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol* **8**, e1000432 (2010).

41. Bomblies, K. & Madlung, A. Polyploidy in the *Arabidopsis* genus. *Chromosome Res* **22**, 117-34 (2014).

42. Hoffmann, A.A. & Rieseberg, L.H. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Evol Syst* **39**, 21-42 (2008).

43. Kulathinal, R.J., Stevison, L.S. & Noor, M.A. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet* **5**, e1000550 (2009).

44. Schacherer, J., Shapiro, J.A., Ruderfer, D.M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342-5 (2009).

45. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-41 (2009).

46. Hittinger, C.T. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet* **29**, 309-17 (2013).

47. Murphy, H.A., Kuehne, H.A., Francis, C.A. & Sniegowski, P.D. Mate choice assays and mating propensity differences in natural yeast populations. *Biol Lett* **2**, 553-6 (2006).

48. Knop, M. Evolution of the hemiascomycete yeasts: on life styles and the importance of inbreeding. *Bioessays* **28**, 696-708 (2006).

49. Belloch, C. *et al.* Chimeric genomes of natural hybrids of *Saccharomyces cerevisiae* and *Saccharomyces kudriavzevii*. *Appl Environ Microbiol* **75**, 2534-44 (2009).

50. Baker, E. *et al.* The Genome Sequence of *Saccharomyces eubayanus* and the Domestication of Lager-Brewing Yeasts. *Mol Biol Evol* **32**, 2818-31 (2015).

51. Gonzalez, S.S., Barrio, E., Gafner, J. & Querol, A. Natural hybrids from *Saccharomyces cerevisiae*, *Saccharomyces bayanus* and *Saccharomyces kudriavzevii* in wine fermentations. *FEMS Yeast Res* **6**, 1221-34 (2006).

52. Zhang, H., Skelton, A., Gardner, R.C. & Goddard, M.R. *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* reside on oak trees in New Zealand: evidence for migration from Europe and interspecies hybrids. *FEMS Yeast Res* **10**, 941-7 (2010).

53. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**, 839-50 (2006).

54. Greig, D., Travisano, M., Louis, E.J. & Borts, R.H. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J Evol Biol* **16**, 429-37 (2003).

55. Hunter, N., Chambers, S.R., Louis, E.J. & Borts, R.H. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J* **15**, 1726-33 (1996).

56. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68-72 (2003).

57. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. & Louis, E.J. Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451-4 (2000).

58. Greig, D., Borts, R.H., Louis, E.J. & Travisano, M. Epistasis and hybrid sterility in *Saccharomyces*. *Proc Biol Sci* **269**, 1167-71 (2002).

59. Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet* **3**, e21 (2007).

60. Kao, K.C., Schwartz, K. & Sherlock, G. A genome-wide analysis reveals no nuclear dobzhansky-muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genet* **6**, e1001038 (2010).

61. Li, C., Wang, Z. & Zhang, J. Toward genome-wide identification of bateson-dobzhansky-muller incompatibilities in yeast: a simulation study. *Genome Biol Evol* **5**, 1261-72 (2013).

62. Strope, P.K. *et al.* The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* **25**, 762-74 (2015).

63. Friedrich, A., Jung, P., Reisser, C., Fischer, G. & Schacherer, J. Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Mol Biol Evol* **32**, 184-92 (2015).

64. Brion, C., Pflieger, D., Friedrich, A. & Schacherer, J. Evolution of intraspecific transcriptomic landscapes in yeasts. *Nucleic Acids Res* **43**, 4558-68 (2015).

65. Jeffares, D.C. *et al.* The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nat Genet* **47**, 235-41 (2015).

66. Wang, Q.M., Liu, W.Q., Liti, G., Wang, S.A. & Bai, F.Y. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol Ecol* **21**, 5404-17 (2012).

67. Cubillos, F.A. *et al.* Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol Ecol* **20**, 1401-13 (2011).

68. Hittinger, C.T., Rokas, A. & Carroll, S.B. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. *Proc Natl Acad Sci U S A* **101**, 14144-9 (2004).

69. Perez-Ortin, J.E., Querol, A., Puig, S. & Barrio, E. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res* **12**, 1533-9 (2002).

70. Heck, J.A. *et al.* Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3256-3261 (2006).

71. Demogines, A., Wong, A., Aquadro, C. & Alani, E. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet* **4**, e1000103 (2008).

72. Hou, J. & Schacherer, J. Negative epistasis: a route to intraspecific reproductive isolation in yeast? *Curr Genet* (2015).

Project Summary

Species-wide genetic survey to elucidate the genotype-phenotype relationship in yeast

Overview of the project

Natural populations of yeast display considerable levels of genetic and phenotypic diversity, in part allowing them to adapt to changing environments and to new ecological niches. Using a large number of natural isolates originated from various biotopes across the globe, we sought to have a deeper insight into the genetic origin of phenotypic variation within the yeast *Saccharomyces cerevisiae*. More specifically, we dissected the molecular basis of the onset of reproductive isolation, which leads to partial offspring lethality upon crosses (Figure 1A). In addition to discrete characters like offspring lethality, we also focused on quantitative traits such as the global pattern of offspring fitness variation across multiple parental crosses and culture conditions (Figure 1B).

An overarching theme of this project has been the integration of classical yeast genetics with high-throughput genotyping and phenotyping strategies in order to survey the genetic basis of species-wide phenotypic variation. In the **first chapter**, we discuss the theoretical principles of quantitative traits, and how to apply classical yeast segregation analyses to these principles in order to infer the underlying genetic complexity of phenotypes. We particularly focused on the impact of epistasis, *i.e.* the non-additive interactions among genes, to the patterns of phenotypic distribution and segregation in the offspring. Based on the phenotypic segregation patterns, we proposed different mapping strategies to precisely identify the genomic loci involved in two loci epistatic interactions. The methods proposed in this chapter served as a general framework for systematic identification and variant mapping of epistatic interactions within yeast populations[1].

In the **second chapter**, we performed the first species-wide survey of post-zygotic reproductive isolation using natural isolates of *S. cerevisiae* by applying the strategies described in the first chapter (Figure 1C). We selected 60 isolates originated from soil, tree barks, immuno-compromised patients and various fermentations across different continents, and performed systematic crosses with the laboratory reference strain S288c. By analyzing the offspring viability of each cross, 16 reproductive isolation cases were identified, with reduced offspring viabilities ranging from 44% to 86%. Further analyses led to the identifications of large-scale reciprocal translocations in 10 out of 16 cases[2]. In this scenario, loss of offspring viability was due to unbalanced distribution of essential genes upon meiosis - a mechanism which is widely distributed both

within and between closely related yeast species and contribute to reproductive isolation at both evolutionary scales.

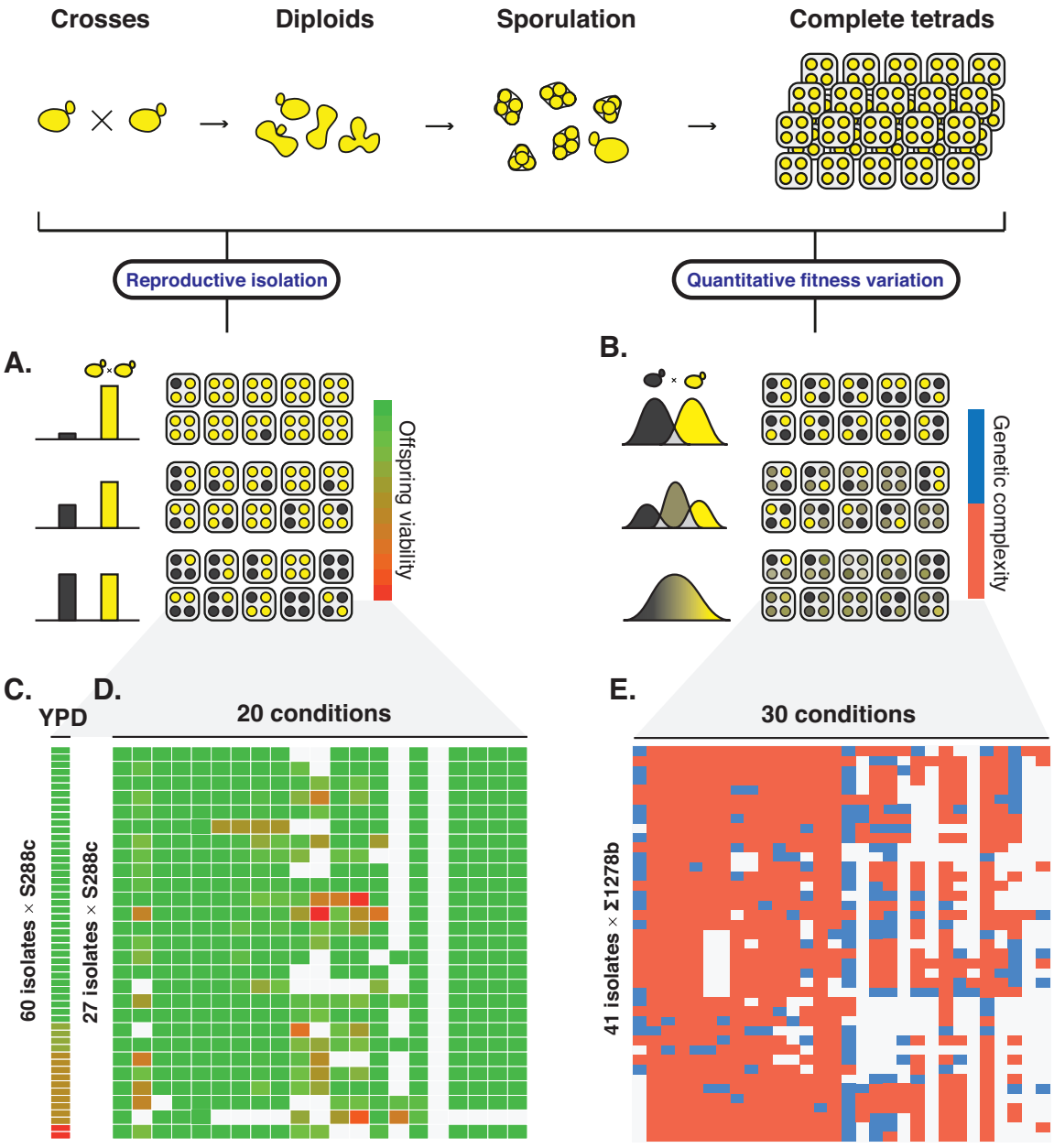


Figure 1. Schematic description of the project

Nevertheless, our first survey was, in some way, biased because of the use of standard rich media that optimize yeast growth. In the **third chapter**, we investigated the impact of environmental factors on the onset of reproductive isolation by selecting 27 crosses that were previously shown to yield high offspring viability on rich media, and tested these crosses on different culture conditions (*e.g.* carbon sources, chemicals that impact various cellular processes

and temperatures) (Figure 1D). In total, 481 cases spanning 27 crosses on 20 conditions were assessed, and 24.3% of all cases (117/481) showed different degrees of condition specific loss of offspring viability ranging from 1% to 62%, indicating the presence of potential negative epistatic interactions. Using the mapping strategy previously described in Chapter 1, we further identified and characterized the first example of two loci Dobzhansky-Müller incompatibility in yeast related to respiratory conditions. These results revealed that negative epistasis are segregating within yeast natural populations and could contribute to reproductive isolation as well[3,4].

Altogether, these first large-scale surveys demonstrated that different genetic variants, either large chromosomal changes or point mutations, could lead to the onset of reproductive isolation within a single species. However, such discrete and severe phenotype is specific and does not reflect the overall phenotypic diversity within a population. In the **fourth chapter**, we conducted a new quantitative phenotypic survey by crossing 41 natural isolates with the strain Σ1278b to assess the global patterns of phenotypic diversity and inheritance (Figure 1E). We measured the offspring fitness variations across a panel of 30 stress conditions (including osmotic stress and various drugs that impact transcription and translation integrity, signal transduction, and cell wall stability), and analyzed the fitness distribution as well as the segregation. By applying the principles described in chapter 1, we identified 8.9% (98/1,105) of the surveyed cases as Mendelian traits. We further traced the effect of a causal Mendelian variant across different genetic backgrounds. Interestingly, significant deviations from the Mendelian expectation were observed ranging from intermediate to high complexities illustrating the hidden complexity of a monogenic mutation across a natural population[5].

Publications related to this work

1. Hou, J. & Schacherer, J. On the mapping of epistatic genetic interactions in natural isolates: Combining classical genetics and genomics. *Methods Mol Biol* **1361**, 345-60 (2016).
2. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* **24**, 1153-9 (2014).
3. Hou, J., Friedrich, A., Gounot, J.S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat Commun* **6**, 7214 (2015).
4. Hou, J. & Schacherer, J. Negative epistasis: a route to intraspecific reproductive isolation in yeast? *Current Genetics*, 1-5 (2015).
5. Hou, J., Sigwalt, A., Pflieger, D., Peter, J., de Montigny, J., Dunham, M. & Schacherer, J. The hidden complexity of Mendelian traits across yeast natural populations. *BioRxiv* (2016).

Chapter 1

Combining classical genetics and genomics to map epistatic genetic interactions in yeast

Introduction

Elucidating the phenotypic outcomes of natural genetic variation within a species remains a major challenge in modern biology. Phenotypes, ranging from growth ability in microorganisms to disease susceptibility in human, could vary quantitatively and continuously within a population. As such, complex traits are not only determined by the additive effect of individual genes, but also are influenced by gene-gene interactions, environment and other non-genetic factors[1]. The nonlinear interaction between genes, or epistasis, has long been recognized as an important biological feature in complex genetic systems, from understanding regulatory networks to identifying mechanisms that are involved in generating and maintaining diversity[2].

The term "epistasis" was first deployed by William Bateson[3] in 1909, describing the deviation of phenotypic segregation pattern from expected allelic effects of individual genes in a hybrid cross. More than 100 years after, our understanding of epistasis still largely benefits from classical genetic approaches. Evidences of epistatic interactions between natural genetic variants have been found through transmission ratio distortion in crosses from multiple major model organisms, including *Arabidopsis thaliana*, *Drosophila melanogaster*, *Mus musculus* and *Caenorhabditis elegans*[4-7]. For example, by examining a population of ~300 recombinant inbred lines (RILs) in *C. elegans*, originated from a cross between Hawaii and Bristol isolates, biased transmission ratio was found for a region on chromosome 1[7]. Further analysis revealed a lethal interaction between alleles of a paternal-delivered toxin *peel-1* and a zygotic-acting antidote *zeel-1*, resulting in transmission ratio distortion and F2 offspring inviability[8]. Nevertheless, while fruitful, studies using such complex models are often restricted to a low number of crosses and only few examples were dissected up to a molecular level[6,8].

In the past decade, the yeast *Saccharomyces cerevisiae* has contributed to a better understanding of the basis of epistasis at a larger scale. A remarkable development has been the synthetic genetic array (SGA), which was based on automated crosses and selections of pairs of deletion mutants at the genome-scale. As genes involved in the same biological process would more likely to interact and display aggravating phenotype when deleted together, this method allowed for genome-wide profiling of the functional connections of genes at an organismal level[9]. In experimental evolution studies, adaptive mutations that

occurred in different lineages have been shown to lead to epistatic interaction as well[10,11]. For example, when populations of the same progenitor strain were evolved in divergent environments (low glucose *vs.* high salinity), adaptive mutations fixed independently in different populations could interact epistatically, causing the hybrids to be unfit in either ancestral environments[11]. In addition to laboratory-induced mutations, epistatic interactions from genetic variants within natural populations are also prevalent, leading to various phenotypes such as mutator effect[12,13] and complex colony morphologies[14].

S. cerevisiae strains are universally isolated from various environments, including soil, tree barks, different insects, immunodepressed patients and fermentation processes related to wine, bread, beer and bioethanol (The 1002 yeast genomes project, http://1002genomes.u-strasbg.fr/). Genetic variation acquired in their natural context including sequence differences, regulatory changes and structural variations, could potentially lead to epistatic interactions when tested in another genetic background. However, the impact of the overall genetic diversity of yeast to the onset of epistasis remains under explored.

Here, we discuss the theoretical basis to identify epistatic interactions existing between loci in different yeast genetic backgrounds. We first introduce the basic principles of quantitative genetics, and discuss how to supplement these principles with classical segregation analysis in the context of yeast crosses. We focus on the contribution of epistatic interactions on the overall phenotypic variance, and illustrate the impact of such interactions on the offspring phenotypic distribution pattern using simulations. We then concentrate on the distribution and segregation patterns in the case of two loci interactions, and propose different mapping strategies based on bulk segregant analysis or consecutive backcrosses followed by high-throughput genome sequencing in order to identify loci involved. Our method allows high resolution mapping of interacting loci that govern various DNA polymorphisms from single nucleotide mutations to large-scale structural variations.

Variance components contributing to trait heritability

In quantitative genetics, the heritability of a trait describes how genetic components contribute to the observed phenotypic variation. In which, broad-sense heritability (H^2) corresponds to the total genetic variance of a trait, and narrow-sense heritability (h^2) is the proportion explained by additive genetic components[15]. Understanding trait heritability constitute an essential step toward deciphering how genotype determines phenotype, and requires precise estimation of the contribution of both additive and non-additive genetic components. However, most recent works, ranging from linkage mapping in model organisms to genome-wide association studies (GWAS) in humans, are largely biased toward the identification of additive variants. While these strategies allowed for direct inferences of causal variants from genomic and population data, the identified variants only explain a small fraction of the total heritable variation[16]. One possible explanation of the unexplained fraction, called the "missing heritability", could be due to the lack of power in identifying non-additive genetic components, for which the detection using these designs are statistically and computationally challenging[2,16].

$$P = \underbrace{A + D + I}_{G} + E \qquad \text{eq. 1}$$

$$P = A + I + E_{error} \qquad \text{eq. 2}$$

Figure 1. Variance components in quantitative traits

On the theoretical ground, the phenotypic variance for a given trait is determined by the combined effect of genetic (G) and environmental (E) components[15]. The genetic component can be further partitioned into the effects of additivity (A), dominance (D) and epistatic interactions (I) (Figure 1, equation 1). Causal variants act additively when the effect of multiple alleles equals to the cumulative effects of individuals. By contrast, non-additive genetic components act non-linearly, including intralocus interactions or dominance (D), and interlocus interactions or epistasis (I). In a practical sense, different variance components could be estimated using populations of progeny originated from experimental crosses in model organisms. In particular, yeast represents a very powerful model. In a yeast cross, the effect of environment (E) is highly controlled due to standardized conditions and can be entirely

attributed to experimental errors (E_{error})[17]. Moreover, recombinant offspring (segregants) as well as the parents are haploids, as a result the dominance effect (D) is removed. Therefore, the phenotypic variance for a trait is simplified as additivity (A), epistatic interactions (I) and measurement errors (E_{error}) (Figure 1, equation 2).

Using an extremely large segregant panel of more than 4,000 individuals from a cross originated between the laboratory strain S288c and a wine isolate RM11, a recent study precisely estimated broad- and narrow-sense heritability of 46 traits using a linkage design, and showed that epistatic interactions explaining 9% of total phenotypic variance observed[18]. Then again, such strategies require large sequencing and computational efforts, which is impossible to apply on a species-wide scale at the present time. In fact, little is known about the relative importance of epistatic interactions contributing to trait variation across the *S. cerevisiae* species, where genetic and phenotypic diversity is decidedly broad[19-21].

Classical genetic analysis and the complexity of traits

To identify epistatic interactions in yeast, the key might be to supplement previous study designs with classical genetic approaches. A unique feature of yeast genetics is the possibility of tetrad analysis. Haploid yeast cells with opposite mating types can cross and undergo meiosis, forming four haploid progeny grouped in a tetrad where the complete information of any meiotic event is preserved. Detecting epistasis in a yeast cross can then be achieved by looking at phenotypic transmission distortion in individual tetrads without sequencing large recombinant populations as it is the case for other model organisms[4-7]. In fact, the pattern of phenotypic distribution and segregation in a tetrad might not only reveal the possible presence of epistasis, but also could be an indicator of the underlying genetic architecture and complexity of a trait (Figure 2).

For example, consider the simplest case where a trait is controlled by a single locus, with contrasting allelic effects from either parental isolate (Figure 2A). As during each meiosis half of the offspring within the same tetrad would inherit randomly one allele from each parent, the overall phenotypic distribution would follow a bimodal pattern. In this scenario, the total heritable variance could be explained by additive effect, therefore the distribution of the mean phenotypic value within each tetrad would be centered around the parental mean (Figure 2A).

However, a trait that is controlled by multiple genes, each with a smaller allelic effect, would display a continuous variation within the offspring (Figure 2B). Providing that all genes act additively, different combinations of alleles are reshuffled in the offspring, resulting in a normal distribution pattern. However, as each tetrad contains the total genetic information from their parents, the distribution of the mean phenotypic value in the tetrads equals to the parental mean (Figure 2B).

In the presence of epistasis, however, the distribution and the segregation pattern could display unique signatures (Figure 2C). Consider a trait determined by two interacting loci. Parental genetic combinations yield similar phenotypic effects, whereas novel allelic combinations interact non-additively and cause lower phenotypic values in the offspring (negative epistasis). In this scenario, a fraction of the offspring present phenotypic values that are strongly deviated from the mean, resulting in a bimodal distribution of the trait with both parental strains in the same phenotypic cluster. A shift of the mean tetrad phenotypic value from the parental mean could also be observed, corresponding to tetrads with epistatic allelic combinations (Figure 2C).

While the offspring phenotypic distribution could reflect the underlying genetic complexity to a certain degree, it cannot precisely infer the number and effect size of variants involved except for monogenic Mendelian cases. On the other hand, this method is powerful to detect the presence of epistatic interaction by the means of comparing the average phenotypic value in the tetrads with the parental mean, even in cases with high genetic complexities resulting from both additive and epistatic components. Systematic detection of epistatic interactions contributing to quantitative traits at the species level can then be achieved by sampling a large number of crosses with different parental combinations. However, efficient methods to identify loci involved in epistatic interactions with high genetic complexities require further investigations.

Characterizing low complexity interactions using segregation analyses

For complex interactions, the difficulty for efficient mapping of the loci stems from the lack of recognized phenotypic segregation pattern in tetrads. Nevertheless, low complexity interactions, such as the case involving two loci with large effect, could have distinct segregation, which are detectable using tetrad analyses. For example, consider a basic interaction model involving two unlinked loci A and B. Locus A has two alleles **A** and **a**; and locus B has two

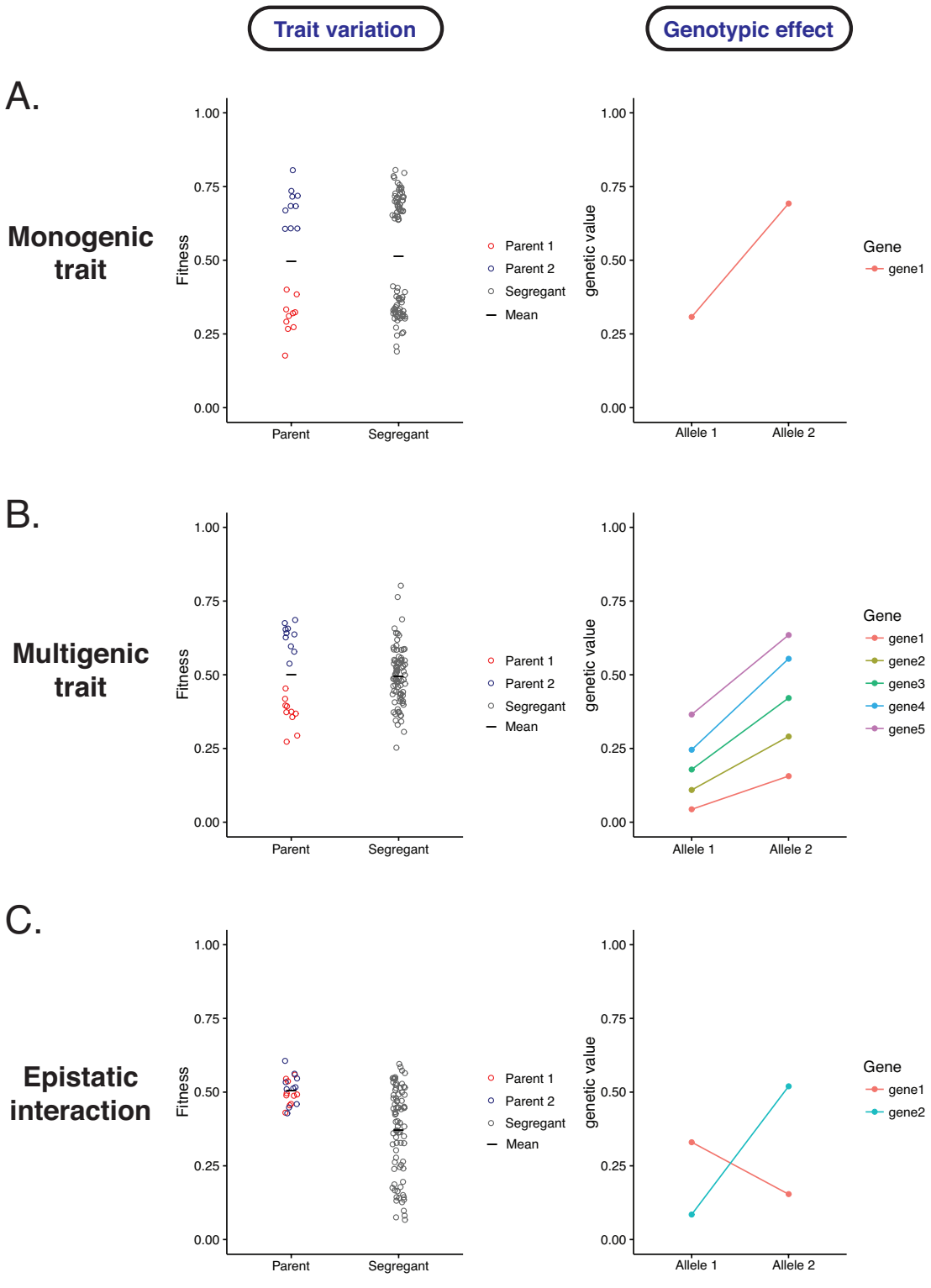
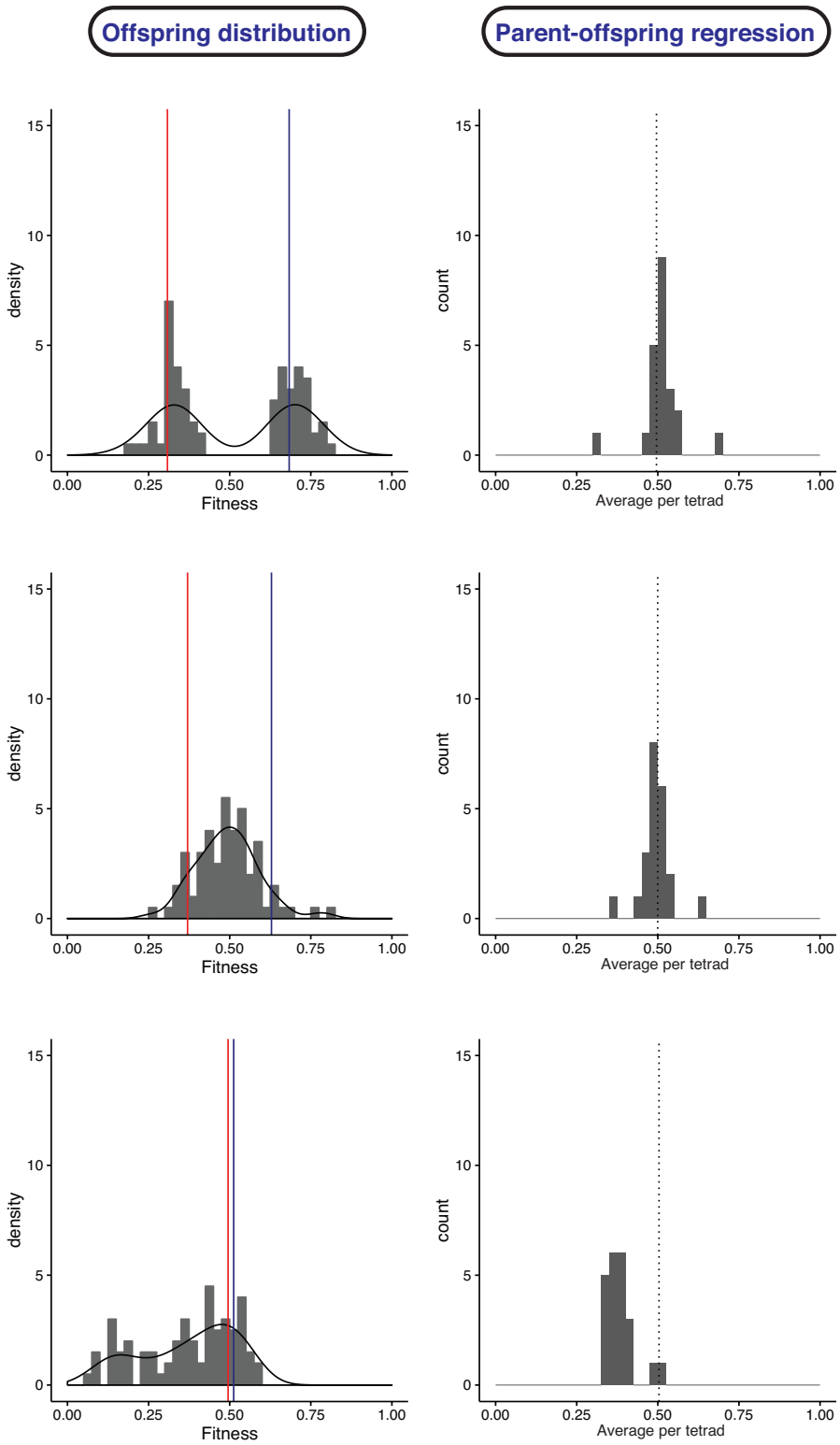


Figure 2. Offspring distribution patterns in monogenic, multigenic and traits with epistatic control. Plots were generated by simulating allelic effects of different complexity categories of traits in 20 complete tetrads (80 offspring). Experimental noise (E_{error}) was incorporated as a standard variation of 0.05 for all individuals. From left to right: overall phenotypic variation in the parents and offspring; genotypic effects; offspring fitness distribution; distribution of the average fitness value in the tetrads relative to parental mean.


Offspring distribution
Parent-offspring regression
density
count
Fitness
Average per tetrad
0
5
10
15
0.00
0.25
0.50
0.75
1.00

alleles **B** and **b**. Suppose we have two parental strains P1 and P2 with genotypes **Ab** and **aB**, respectively. Given that allele **a** and allele **b** interact recessively, the genotypic effects of all allelic combinations from loci A and B can be summarized as Figure 3. In this case, both parental combinations **Ab** and **aB** have the same phenotypic value as well as the recombinant genotype **AB**, whereas **ab** shows an epistatic effect resulting in a lower phenotypic value. If we admit that the allele frequencies of **A**, **B**, **a** and **b** are equal in the offspring, which is expected from the cross between P1 and P2, the overall frequency of the parental phenotype will be 75% (Figure 3).

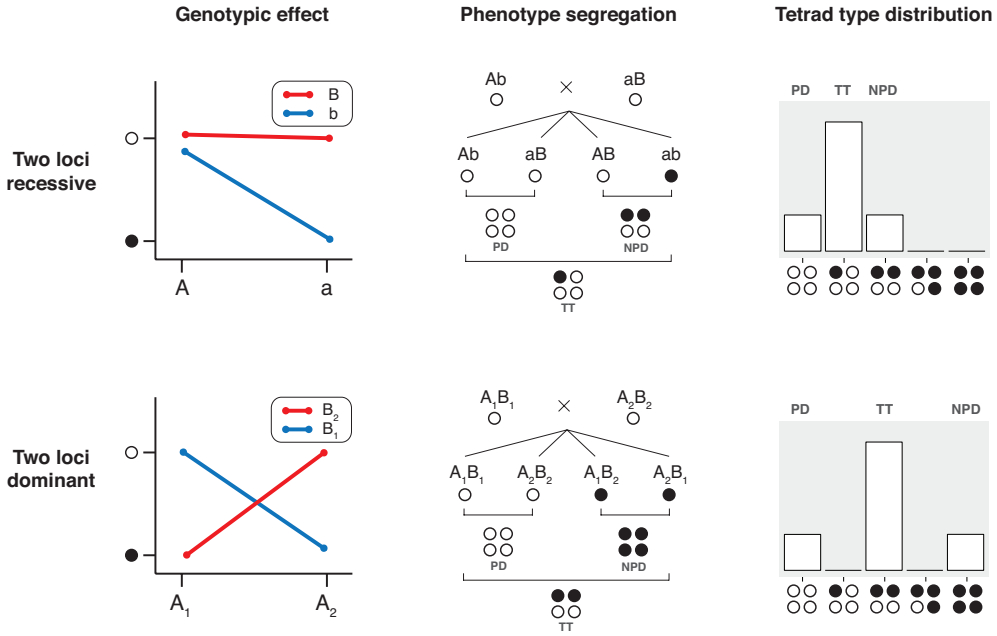


Figure 3. Phenotypic segregation pattern and tetrad type distribution in the offspring. Segregation in the F1 offspring for two loci recessive (upper) and dominant (lower) interactions are presented. Genotypic effects of allelic combinations are shown in left panel. The phenotypic segregation in each type of tetrad PD, TT and NPD are shown on the middle panel, and the distribution of the tetrad types are shown on the right panel supposing that the two loci involved are independent.

Another model is possible when the involved loci interact dominantly. Consider two independent loci A and B each with two alleles denoted $\mathbf{A_1}/\mathbf{A_2}$ and $\mathbf{B_1}/\mathbf{B_2}$. Suppose we have two parental strains P1 and P2 with the genotypes $\mathbf{A_1B_1}$ and $\mathbf{A_2B_2}$. If A and B interact dominantly, any non-parental genotype combination would have an epistatic effect. The genotypic effects of loci A and B could then be summarized in Figure 3, where the parental combinations $\mathbf{A_1B_1}$ and $\mathbf{A_2B_2}$ have the same phenotypic value, and the recombinant genotypes $\mathbf{A_1B_2}$ and $\mathbf{A_2B_1}$ have a lower phenotype value. Again, as equal allelic frequencies of $\mathbf{A_1}$,

A_2, **B_1** and **B_2** are expected, the overall frequency of the parental phenotype in the offspring will be 50% (Figure 3).

For both models considering the segregation of two loci for a given cross, the resulting tetrads could be assigned to different types according to the allelic combination. There are 3 possible types of tetrads: parental ditype (PD) contains only parental alleles, non-parental ditype (NPD) contains only recombined alleles, and tetratype (TT) contains all four possible allelic combinations. As all spores in a tetrad are haploids from a single meiosis event, the phenotype distribution in the tetrad could thus reflect directly the type of interaction of the loci in question (Figure 3).

Mapping strategies for two interacting loci

In the case of two loci interaction, different mapping strategies could be used to identified the loci involved. Since its first implementation in yeast[22], bulk segregant analysis strategy has become increasingly popular among yeast geneticists. The principle of the strategy is to group segregants from a mapping cross according to their phenotypes, and then genotype this pool of segregants all together[23]. Genomic regions containing the causative loci will have a skewed allele frequency whereas the rest of the genome will have an equal proportion of alleles from each parent.

While traditional design in BSA-seq (**B**ulk **S**egregant **A**nalysis combined with new generation **seq**uencing) for mapping quantitative trait loci (QTL) usually focuses on pools of segregants with upper and lower extreme phenotypes, the same design is less applicable when mapping epistatic interactions. Take for example the case of a recessive two loci interaction (Figure 3). The lower phenotypic group $P_{episitasis}$ contains segregants with only one possible genotype **ab**. Sequencing of this pool will efficiently localize the causative loci, as the allele frequency of loci A will be biased toward the allelic version of **a** (**A/a** = 0/1) and the allele frequency of loci B will be biased toward the allelic version of **b** (**B/b** = 0/1) (Figure 4B). However, when the epistatic effect is strong enough to affect the viability of the segregants, mapping using the lower phenotypic group will simply be impossible. In this case, only the upper phenotypic group P_{parent} could be used, which contains equal proportion of segregants with the genotype **Ab**, **aB** and **AB**. As a result, only a small variation of allele frequency at both loci could be observed (**A/a** = 0.67/0.33, **B/b** = 0.67/0.33) (Figure 4A), and the power of detecting these loci will be extremely limited due to the

presence of experimental noise (that is, random allele frequency variation at unassociated loci).

For efficient mapping of the aforementioned scenario, the segregation pattern of the phenotype has to be taken account of. Suppose that the combination **ab** cause a lethal phenotype, then the distribution of viable segregant in the tetrads will be PD:TT:NPD = 4:3:2. Knowing that the lethal combination **ab** is absent in the NPD tetrads, the mapping could be achieved by pooling segregants from independent NPD tetrads. In this case, as all segregants in this pool will only have the genotype **AB**, the allele frequency at A locus will bias toward **A** (**A/a** =1/0) and the allele frequency at B locus will bias toward **B** (**B/b** =1/0) (Figure 4C).

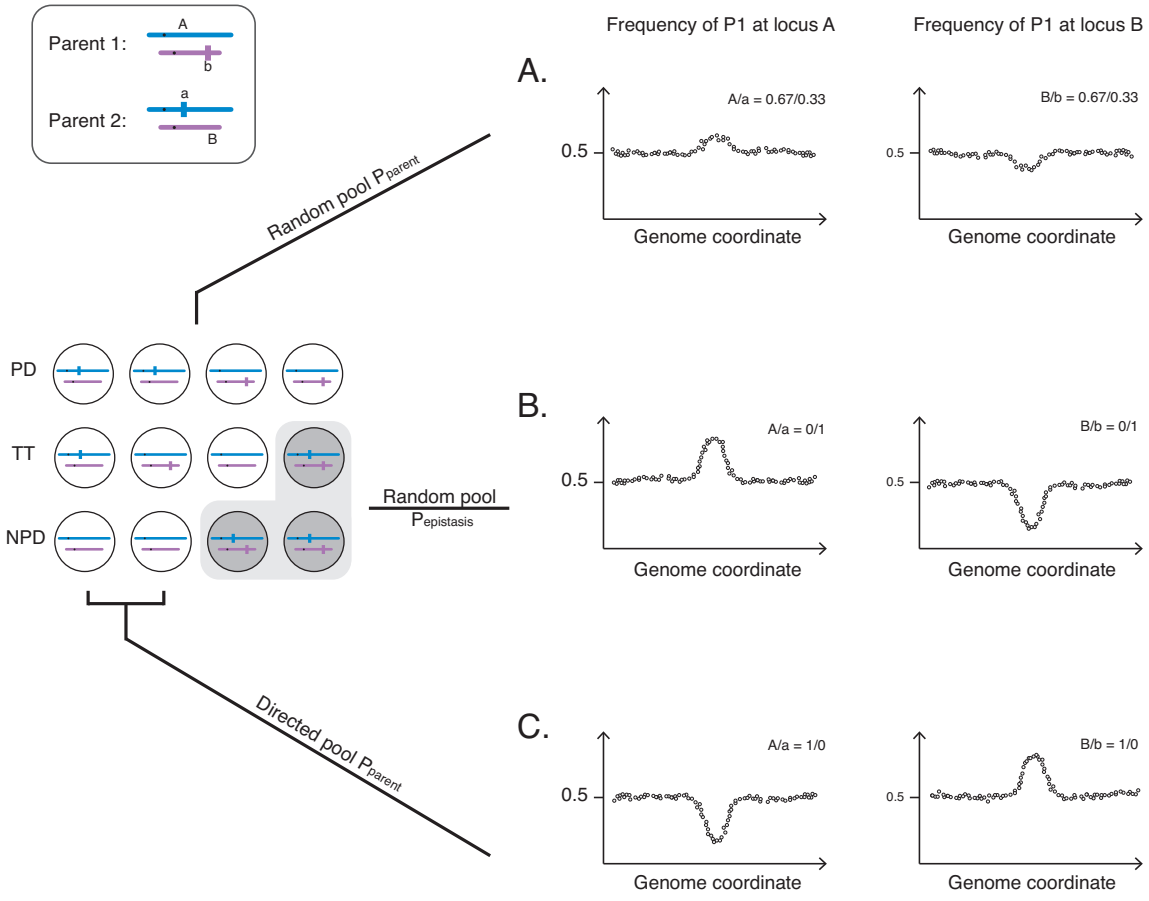


Figure 4. Comparison of random and segregation-directed bulk segregant analysis strategy in mapping a two loci recessive interaction. Allele frequencies of (A) random F1 pool of P_{parent} segregants; (B) random F1 pool of $P_{epistasis}$ segregants and (C) segregation directed pool of P_{parent} segregants in NPD tetrads are represented. The allele frequency equal to 1 means only the allele from parent P1 is present.

Nevertheless, a major limit of this method is that it relies on the ability of selecting a pool of segregant with biased genotype. For example, in the case of a dominant interaction between two loci, upper and lower phenotypic groups as well as different types of tetrads will always have the same frequencies of each allele. The application of BSA-seq is simply powerless in this scenario and another mapping strategy is required.

Introgression of alleles with major phenotypic effects by consecutively backcrossing one strain to another is not new, especially in organisms such as yeast where backcrossing is timely effective. However, the use of introgression in mapping epistatic interaction is not yet common. The concept here is to treat the segregation pattern as a phenotype itself, and simultaneously introduce all interacting loci into a single genetic background. The identification of the causative loci is then possible by sequencing only one backcrossed segregant and looking for introgressed regions. Even though this strategy is somewhat more labor intensive, it allows for efficient mapping of dominant interactions, which compensate the major short coming of BSA-seq.

Take for example a dominant interaction between two loci (Figure 3). Parental strains P1 and P2 have the genotype $\mathbf{A_1B_1}$ and $\mathbf{A_2B_2}$, which result in the same phenotypic value. Therefore, the distribution of the parental phenotype in different tetrad types would be PD:TT:NPD = 4:2:0 (Figure 3). To map these loci, the idea is to introduce both alleles $\mathbf{A_1}$ and $\mathbf{B_1}$ into the genetic background of the parental strain P2 (Figure 5). To do so, one PD tetrad in the generation F1 is selected, and all four spores from this tetrad are backcrossed with P2. For all four crosses, the segregation pattern of the phenotype is scored again. Since PD tetrads contain only segregants with the parental genotype $\mathbf{A_1B_1}$ and $\mathbf{A_2B_2}$, half of these segregants ($\mathbf{A_1B_1}$) will retain the epistatic segregation, whereas the other half of the segregants ($\mathbf{A_2B_2}$) will show no phenotypic effect when backcrossed to P2. Then, one segregant that retained the 4:2:0 segregation is selected, and again one PD tetrad is taken to perform another round of backcross. By repeating this procedure for several generations, the genome of the backcrossed segregant will be highly enriched by the allele of P2, except for the regions containing the causative loci (Figure 5).

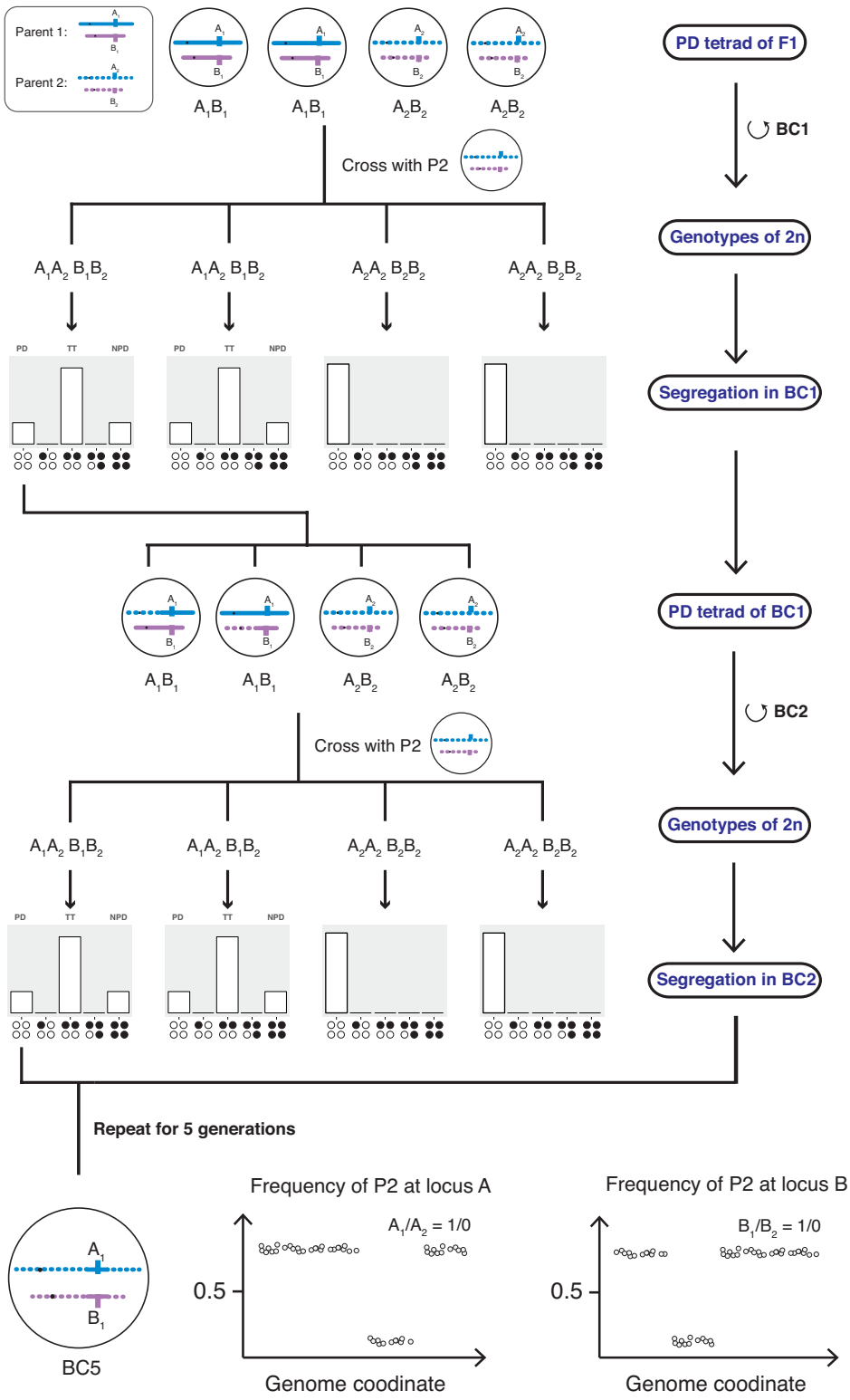


Figure 5. Overview of double introgression of interacting loci using successive backcrossing. For each generation the tetrad type segregation is treated as a phenotype. One full PD tetrad from F1 is selected and backcrossed to the parental strain P2 and the segregation pattern is scored. One segregant that retained the segregation pattern is then selected and 5 subsequent backcrosses are performed. The resulting BC5 segregant will be enriched for the P2 genome except for regions involved in the interaction.

Conclusion

In this first chapter, we discussed the theoretical framework to identify epistatic interactions. We first considered the signature of epistasis to the overall offspring phenotypic distribution pattern for any given cross. We further focused on the segregation pattern in the tetrads for two loci interactions and discussed methods, which can be used to map the genomic loci involved in such interactions. The strategies discussed here were applied to our species-wide genetic surveys for various phenotypes, from the onset of reproductive isolation (Chapter 2 and 3) to the dissection of the genetic complexity of traits in general using *S. cerevisiae* (Chapter 4).

Publication related to this chapter:

Hou, J. & Schacherer, J. On the mapping of epistatic genetic interactions in natural isolates: combining classical genetics and genomics. *Methods Mol Biol* **1361**, 345-60 (2016).

References

1. Phillips, P.C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855-67 (2008).
2. Mackay, T.F. Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nat Rev Genet* **15**, 22-33 (2014).
3. Bateson, W. *Mendel's principles of heredity*, Cambridge Univ. Press (1909).
4. Chae, E. *et al.* Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* **159**, 1341-51 (2014).
5. Corbett-Detig, R.B., Zhou, J., Clark, A.G., Hartl, D.L. & Ayroles, J.F. Genetic incompatibilities are widespread within species. *Nature* **504**, 135-7 (2013).
6. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**, 623-6 (2009).
7. Seidel, H.S., Rockman, M.V. & Kruglyak, L. Widespread genetic incompatibility in C. elegans maintained by balancing selection. *Science* **319**, 589-94 (2008).
8. Seidel, H.S. *et al.* A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol* **9**, e1001115 (2011).
9. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425-31 (2010).
10. Kvitek, D.J. & Sherlock, G. Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet* **7**, e1002056 (2011).
11. Dettman, J.R., Sirjusingh, C., Kohn, L.M. & Anderson, J.B. Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature* **447**, 585-8 (2007).
12. Demogines, A., Wong, A., Aquadro, C. & Alani, E. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet* **4**, e1000103 (2008).
13. Bui, D.T., Dine, E., Anderson, J.B., Aquadro, C.F. & Alani, E.E. A genetic incompatibility accelerates adaptation in yeast. *PLoS Genet* **11**, e1005407 (2015).
14. Taylor, M.B. & Ehrenreich, I.M. Genetic interactions involving five or more genes contribute to a complex trait in yeast. *PLoS Genet* **10**, e1004324 (2014).
15. Falconer, D.S. & Mackay, T.F.C. *Introduction to quantitative genetics*, Longman, Harlow (1996).
16. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
17. Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-7 (2013).
18. Bloom, J.S. *et al.* The role of genetic interactions in yeast quantitative traits. *Nat Commun.* **6**, 8712 (2015).
19. Skelly, D.A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-504 (2013).
20. Schacherer, J., Shapiro, J.A., Ruderfer, D.M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342-5 (2009).
21. Bergstrom, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* **31**, 872-88 (2014).
22. Brauer, M.J., Christianson, C.M., Pai, D.A. & Dunham, M.J. Mapping novel traits by array-assisted bulk segregant analysis in Saccharomyces cerevisiae. *Genetics* **173**, 1813-6 (2006).
23. Dunham, M.J. Two flavors of bulk segregant analysis in yeast. *Methods Mol Biol* **871**, 41-54 (2012).

Chapter 2

Species-wide survey of reproductive isolation reveals the role of chromosomal rearrangements on rich media

Introduction

Understanding the molecular basis of how reproductive isolation evolves between individuals from the same species offers valuable insight into patterns of genetic differentiation as well as the onset of speciation[1,2]. The yeast *Saccharomyces cerevisiae* constitutes an ideal model partly due to its vast ecological range, high level of genetic diversity[3-6] and laboratory amendable sexual reproduction. Between *S. cerevisiae* and its sibling species in the *Saccharomyces sensu stricto* complex, reproductive isolation acts post-zygotically and could be attributed to chromosomal rearrangements[7], cyto-nuclear incompatibility[8,9] and anti-recombination[10,11]; although the implication of these mechanisms at the incipient stage of speciation remains unclear due to further divergence in the nascent species. Recently, several studies assessed the onset of intraspecific reproductive isolation in *S. cerevisiae* by evaluating the effect of the mismatch repair system[12-14] or by fostering incipient speciation using the same initial genetic background[15-18]. Nevertheless, the overall genetic diversity within this species was largely overlooked and no systematic evaluation has been performed.

To this end, we carried out the first species-wide survey of post-zygotic reproductive isolation within a natural population of *S. cerevisiae*. We selected 60 isolates originated from soil, tree barks, immuno-compromised patients and various fermentations across different continents, and performed systematic crosses with the laboratory reference strain S288c. By measuring the offspring viability for each cross, we identified 16 cases of reproductive isolation with reduced offspring viabilities ranging from 44% to 86%. Depending on the segregation pattern in each case, we applied different mapping strategies based on bulk segregant analysis and successive backcrossing along with next-generation sequencing, which allowed full characterizations of the observed cases to a molecular resolution. We identified reciprocal translocations in a large fraction of all isolates surveyed, which lead to the observed loss of offspring viability due to unbalanced distribution of essential genes upon meiosis. These data suggest that large-scale chromosomal rearrangements might play a major role to the onset of reproductive isolation in this species, at least on permissive laboratory conditions.

Strains	Source	Location	Offspring viability	Divergence to S288c (%)	Reference
273614X	Clinical	UK	100%	0.34	3
CLIB192	Baker	France	100%	0.11	3
CLIB413	Fermented Rice		100%	0.36	3
T7	Oak tree	US	100%	0.49	Justin Fay*
YJM421	Clinical	US	100%	0.35	3
EM93_3	Rotting fig	US	98%	0.14	3
WE372	Wine	South Africa	98%	0.26	3
YJM320b	Clinical	US	96%	0.32	3
CLIB272	Beer	US	95%	0.23	3
DBVPG6861	Polluted water		95%		
YJM434	Clinical	Europe	95%	0.28	3
YJM678	Clinical		95%		
CLIB483	Fermentation	France	94%	0.29	3
RM11	Vineyard	US	93%	0.36	Justin Fay*
CLIB154	Wine	Russia	93%	0.21	3
CLIB382	Beer	Japan	93%	0.25	3
DBVPG3591	Cocoa beans		93%	0.23	3
I14	Vineyard	Italy	93%	0.25	3
UC8	Wine	South Africa	93%	0.28	3
DBVPG1106	Grapes	Australia	92%	0.35	50
YJM975	Clinical	Italy	92%	0.36	50
CLIB219	Wine	Russia	92%	0.44	3
YJM280	Clinical	US	92%	0.35	3
Y9	Ragi	Indonesia	92%	0.34	3
Y12	Palm wine	Africa	92%	0.35	3
YJM269	Apple juice		92%	0.38	3
K12	Sake	Japan	91%	0.25	3
YJM326	Clinical	US	91%	0.32	3
BC187	Barrel fermentation	US	91%	0.37	50
L-1374	Wine	Chile	91%	0.36	50
PW5	Palm wine	Nigeria	91%	0.59	Justin Fay*
Y3	Palm wine	Africa	91%	0.38	3
CBS7960	Sugar cane	South Africa	90%	0.39	Justin Fay*
CECT10109	Prickly pear	Spain	90%	0.31	3
CLIB294	Fermentation	France	90%	0.25	3
DBVPG1794	Soil	Finland	90%	0.29	3
NC_02	Forest	US	90%	0.43	Justin Fay*
UC1	Wine	France	90%	0.24	3
YJM145	Clinical	US	90%	0.37	3
YJM413	Clinical		90%	0.33	3
YJM440	Clinical	US	90%		
YPS1000	Oak exudate	US	90%	0.41	3
YPS128	Oak	US	90%	0.53	50
YPS163	Oak exudate	US	90%	0.36	3
DBVPG1788	Soil	Finland	86%	0.36	Justin Fay*
Y55	Wine	France	86%	0.54	50
DBVPG6044	Bili wine	West Africa	86%	0.60	Justin Fay*
378604X	Clinical	UK	84%	0.41	50
YPS1009	Oak tree	US	84%	0.50	Justin Fay*
Y10	Coconut	Philippines	83%	0.49	Justin Fay*
YJM981	Clinical	Italy	73%	0.29	3
T73	Wine	Spain	73%	0.23	3
Y9J	Wine	Japan	73%	0.28	3
L-1528	Wine	Chile	72%	0.35	Justin Fay*
M22	Wine	Italy	71%	0.24	3
YJM978	Clinical	Italy	71%	0.26	3
DBVPG4651	Tuber Magnatum	Italy	71%	0.28	3
DBVPG1339	Grape must	Netherland	70%	0.24	3
CECT10266	Tanning Liquor	Spain	48%	0.44	This study
YJM454	Clinical	US	44%	0.48	This study
FY5	Lab	US	98%	0	This study

Table 1. Origin, divergence and offspring viability of strains used in this study. The offspring viability for crosses between listed strains and FY4 (isogenic to S288c) was estimated by dissecting 20 tetrads. Cross between FY4 and FY5 (isogenic to S288c) was performed as control. *Publically available sequences from the Fay lab: http://www.genetics.wustl.edu/jflab/

Results

To obtain a global view of the landscape of intraspecific reproductive isolation in *S. cerevisiae*, we selected 60 natural isolates from diverse ecological and geographical niches (Table 1). Estimated genetic divergence within these strains ranges from 0.11% to 0.60%, which is a relatively comprehensive representation of the genetic diversity observed in this species (Figure 1), with the exception of the highly divergent Chinese strains (~1%) recently isolated from Southern island of China[5]. We crossed all isolates with the reference strain S288c and estimated the offspring viability for each cross. A relatively large fraction of crosses (16 out of 60) qualified as cases of reproductive isolation, with reduced offspring viabilities ranging from 44% to 86%. No apparent correlation was observed between the estimated genetic divergence of the parental pairs and the resulting offspring viability (Figure 2), indicating that general DNA sequence differences were not sufficient to explain the observed reproductive isolation.

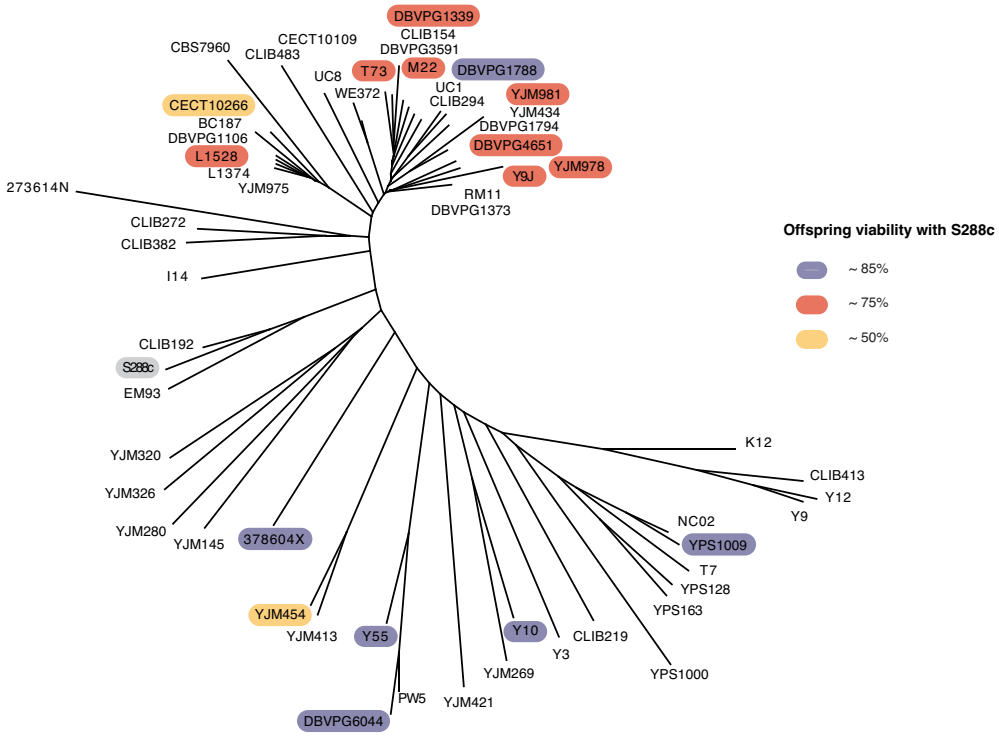


Figure 1. Neighbor-joining tree of 60 studied *S. cerevisiae* isolates. A majority-rule consensus tree of the surveyed strains was built based on the 101,343 segregating sites identified in[3]. Branch lengths are proportional to the number of segregating sites that differentiate each pair of strains. Isolates that are incompatible were color-coded according to the offspring viability resulting from the cross with the reference S288c.

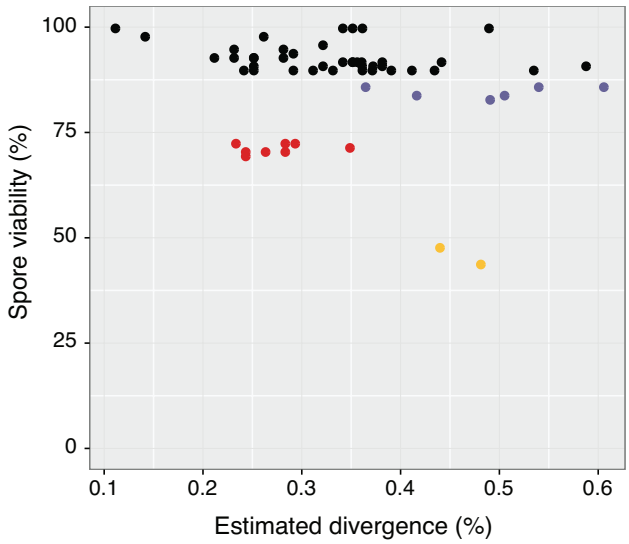


Figure 2. Sequence divergence does not correlate with offspring viability. Sequence divergence between each pair of parental strains was plotted against the observed offspring viability. Incompatible cases were color coded. Blue: crosses with offspring viability of ~85%. Red: crosses with offspring viability of ~75%. Yellow: crosses with offspring viability of ~50%.

To understand the molecular basis and complexity underlying the identified cases, additional tetrads were dissected for all 16 incompatible crosses and the segregation of the lethal phenotype was analyzed (Figure 3). In total, 6 cases showed mild reduction of offspring viability (78% to 87%, mean=82%; 65 tetrads analyzed on average) (Figure 2), which resulted in a Poisson distribution with decreasing number of full tetrads (4 viable spores, Figure 3). This segregation pattern suggests that these cases were probably caused by a mutator[13,14] or anti-recombination effect of the mismatch repair system[12], as previously observed. The remaining 10 cases with a higher degree of progeny loss (44% to 74%) were further analyzed.

Bulk segregant analysis revealed a unique reciprocal translocation responsible for cases of reduced offspring viability of ~75%

According to the segregation, 8 crosses (between S288c and DBVPG1339, DBVPG4651, M22, T73, Y9J, L-1528, YJM978 and YJM981) showed predominantly 3 types of tetrads with 4, 3 or 2 viable spores (Figure 1, Figure 3). The ratio between these tetrad types was roughly 1:4:1, resulting in reduced spore viability of ~75% (66% to 74%, mean=71%; 228 tetrads analyzed on average). In addition, pairwise crosses among all 8 strains showed offspring viabilities higher than 90%, indicating that these cases represented a unique genetic origin.

The particular segregation pattern observed could be explained by two possible mechanisms. First, lethal epistatic interaction between two independently

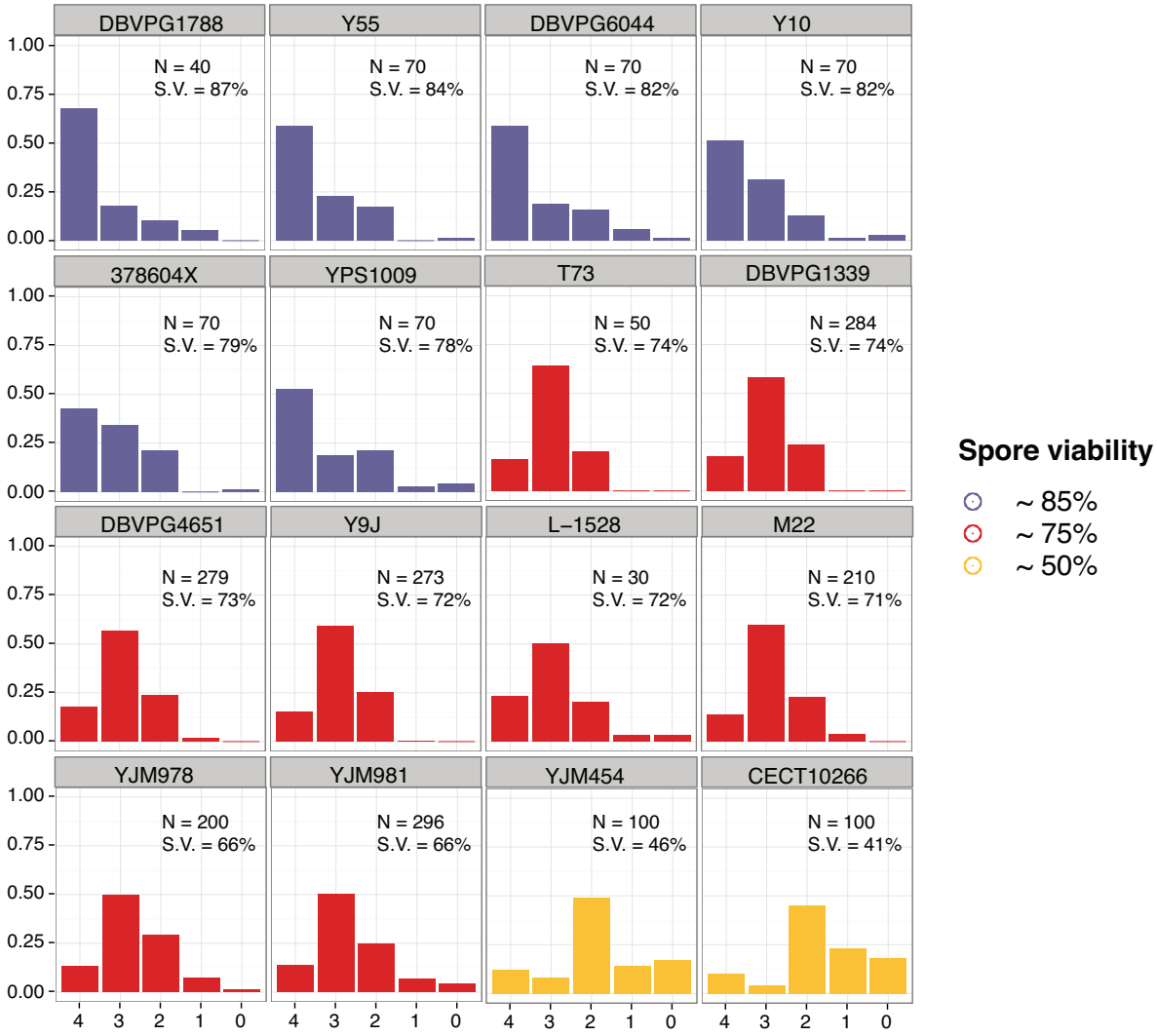


Figure 3. Segregation patterns of the lethal phenotype. Each plot represents the distribution of tetrad types from crosses between S288c and the strain indicated. Crosses are color coded according to their segregation pattern and the observed offspring viability. The number of tetrad dissected (N) and estimated spore viability (S.V.) are indicated. The horizontal axis corresponds to the number of viable spores per tetrad, and the vertical axis the fraction of each type of tetrad observed.

evolved parental alleles, known as the Dobzhansky-Müller incompatibility, could explain this segregation if alleles involved were recessive and unlinked[19,20]. Alternatively, large-scale chromosomal rearrangements such as reciprocal translocations could also lead to this type of segregation, under the condition that only one of the exchanged chromosome arm contains essential genes. In both genic and chromosomal scenarios, lethal genotype combination will follow Mendelian segregation and be united in 1/4 of the offspring, in light of the fact that tetrads with 4 viable spores being parental ditypes (PD), 3 viable spores being tetratypes (TT) and 2 viable spores being non-parental ditypes (NPD).

Since all crosses observed in this category represented essentially a single genetic origin for the observed incompatibility, we selected one strain, DBVPG1339, to map the genomic regions involved. Briefly, ~300 tetrads were dissected, then 50 segregants from independent NPD tetrads were cultured and pooled by equal O.D. readings at 600 nm. Genomic DNA isolated from the pool was sequenced using Illumina Hiseq 2000 technology, with an average coverage of 100X. Sequence reads were aligned against the genome of S288c, and the allele frequency of S288c was scored at each polymorphic (SNP) position. As the lethal genotype combination is absent in viable spores from NPD tetrads, genomic regions involved would have allele frequencies that are skewed from 0.5, whereas the rest of the genome would have equal proportion of alleles from each parent.

Using this strategy, we were able to map two genomic regions with significantly skewed allele frequencies (Figure 4). The first one was located at the left-arm region of chromosome VIII (position 15000 to 71000; allele frequency of S288c near 0); and the second one near the centromeric region of chromosome XVI (position 374000 to 453000; allele frequency of S288c near 1) (Figure 4A). In addition, the coverage for these two chromosomes revealed significant abnormalities: the end of chromosome VIII (~15 kb) showed a very low coverage (~30X) whereas the left-arm of chromosome XVI (~370 kb) showed a coverage that was nearly 200X, indicating that two copies of the left-arm of chromosome XVI might be present (Figure 4B). This unbalanced inheritance of the aforementioned regions strongly suggests the presence of a putative reciprocal translocation between chromosome VIII and XVI in the genome of incompatible strains. In fact, when crossing strains bearing the putative translocation with the reference strain S288c, offspring would have inherited either balanced set of chromosomes (spores in PD tetrads for example), or unbalanced set of chromosomes (spores in NPD tetrads for example). As the region involved on chromosome VIII was near the telomere and does not contain any essential genes, only unbalanced spores with two copies of the left-arm of chromosome XVI were viable, as was evident by the abnormal coverage observed. The existence of this putative translocation was confirmed by PCR in DBVPG1339 as well as in all other 7 strains with reduced offspring viability of ~75% (Figure 5). This translocation occurred between the promoter region of *ECM34* (YHL043W) on chromosome VIII and the promoter region of *SSU1* (YPL092C) on chromosome XVI (Figure 5).

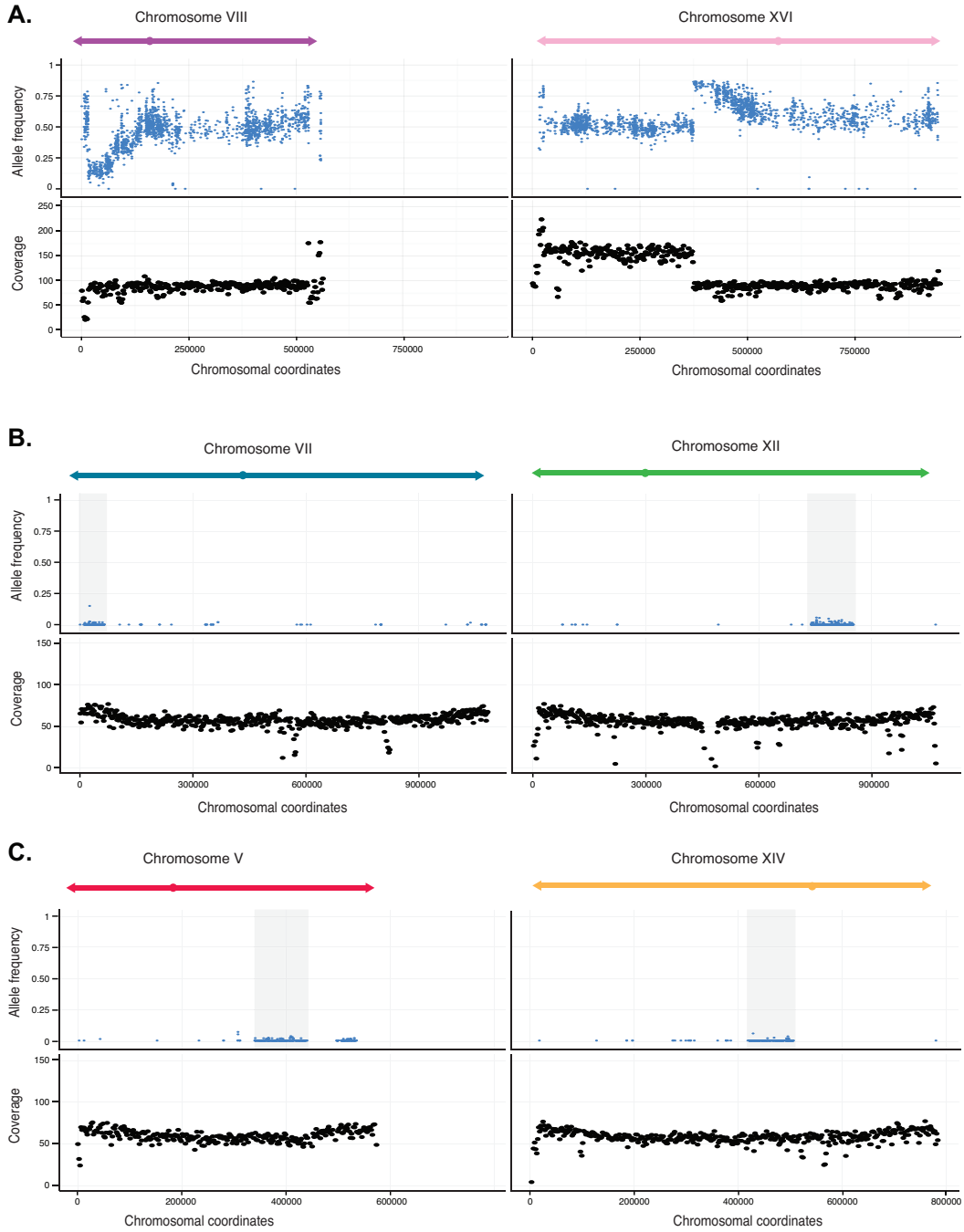


Figure 4. Mapping of the genomic regions involved. A. Regions involved in 75% spore viability. The horizontal axis represents the coordinates of chromosome VIII and XVI. The upper vertical axis represents the allele frequencies of S288c: values close to 1 imply that only alleles of S288c are present and vice versa. The lower vertical axis represents the sequencing coverage in a 2 kb window. The coverage was expected to be 100X. **B. Regions involved in CS-B5.** The first region was ~60 kb, located on the left-arm of chromosome VII (85000 - 145000). The second was a 120 kb region on the right-arm of chromosome XII (725000 - 845000). Coverage ~50X indicates the presence of a single genome copy. **C. Regions involved in YS-B5.** In total, two regions were mapped, the first one was a 105 kb region on the right-arm of chromosome V (385000 - 490000) and the second was a 85 kb region on the left-arm of chromosome XIV (415000 - 500000). Coverage ~50X indicates the presence of a single genome copy.

Successive backcross strategy identified multiple reciprocal translocations responsible for the reduced offspring viability of ~50%

The remaining 2 crosses (CECT10266 and YJM454 with S288c) showed a reduced spore viability of 50% (44% to 48%, mean=46%; 100 tetrads analyzed on average), where 3 major types of tetrad were observed, each contained 4, 2 or 0 viable spores with a ratio of 1:4:1 (Figure 3). Based on the segregation pattern, we reasoned that the most plausible explanation was the presence of a reciprocal translocation involving two large chromosomal regions, each of which contains at least one essential gene[21]. In this context, any meiotic recombination will lead to mis-segregation of essential genes and consequently only the progeny that inherited a balanced set of chromosomes would be viable. Moreover, the cross between CECT10266 and YJM454 demonstrated a further reduction of offspring viability (~25%), indicating that these two strains probably underwent different events leading to the observed reproductive isolation.

Since in these cases, all viable F1 segregants would have equal probability of inheriting either balanced parental genome, no allele frequency variation would be observed by simply pooling the F1 segregants, as opposed to previously seen in cases with ~75% offspring viability. To efficiently map the translocation junctions, we used a strategy based on successive backcrossing and next-generation sequencing as described previously in Chapter 1. Basically, for both crosses, F1 segregants that have maintained the phenotype of 50% offspring viability were successively backcrossed to S288c for 5 generations, in order to obtain a single segregant enriched for the S288c genome but still retained the original translocation. Each 5th generation backcross segregant, namely CS-B5 (segregant originated from the cross between CECT10266 and S288c) and YS-B5 (segregant originated from the cross between YJM454 and S288c), was completely sequenced using Illumina Hiseq 2000 technology, with an average coverage of 50X. Reads alignment and SNP callings were performed as previously. Due to limited recombination around the translocation junctions, the genome of these backcrossed segregants would be otherwise allelic to S288c except for regions involved in the translocation.

Identification of a reciprocal translocation between chromosome VII and XII in CECT10266

Genome sequencing of the segregant CS-B5 (derived from the cross between CECT10266 and S288c) revealed two regions that are polymorphic to S288c. The first region was approximately 60 kb in length, located on the left-arm of chromosome VII (85000 - 145000) and the second was a 120 kb region on the right-arm of chromosome XII (725000 - 845000) (Figure 4B). Using genomic DNA from the parental strain CECT10266 as template, several PCRs were performed to identify the breakpoints of the putative translocation. The first breakpoint was located between *MCM6* (YGL200C) and *EMP24* (YGL201C) on chromosome VII and the second breakpoint was located between YLR326W and *NMA1* (YLR328W) on chromosome XII. Considering the relative position of the centromeres on those two chromosomes, the translocation would likely have occurred between the left-arm of chromosome VII and the right-arm of chromosome XII (Figure 4B), leading to two new chimeric chromosomes with functional centromeres. The junctions of this putative translocation were confirmed using PCR amplification. Sequencing of the amplified fragments revealed a full-length Ty2 transposon at both junctions (Figure 5), suggesting that the translocation was likely originated by homologous recombination (HR) between *Ty* elements.

Identification of a reciprocal translocation between chromosome V and XIV in YJM454

Similarly, we also mapped two regions that were polymorphic to S288c in the genome of YS-B5 (derived from the cross between YJM454 and S288c). The first one was a 105 kb region on the right-arm of chromosome V (385000 - 490000) and the second was a 85 kb region on the left-arm of chromosome XIV (415000 - 500000) (Figure 4C). By the same principle, we checked potential breakpoints within these regions by PCR using genomic DNA of YJM454 as template. The first breakpoint was located between *PMD1* (YER132C) and *GLC7* (YER133W) on chromosome V; and the second one was between *PHO23* (YNL097C) and *RPS7B* (YNL096C) on chromosome XIV. In this case, the right-arm of chromosome V was likely exchanged with the left-arm of chromosome XIV to ensure centromeric functions of the chimeric chromosomes (Figure 5). Indeed, PCR amplifications confirmed the presence of both junctions for this putative translocation (Figure 5). Sequence analysis of the

junctions revealed a full-length Ty2 transposon at both junctions, and an additional 3 kb fragment containing a partial length *Ty4* element at the junction uniting the right-arms of chromosome V and XIV (Figure 5). The presence of multiple *Ty* elements suggests that the breakpoints might overlap with potential *Ty* insertion hotspots. This translocation probably was also mediated by homologous recombination (HR) through *Ty* elements.

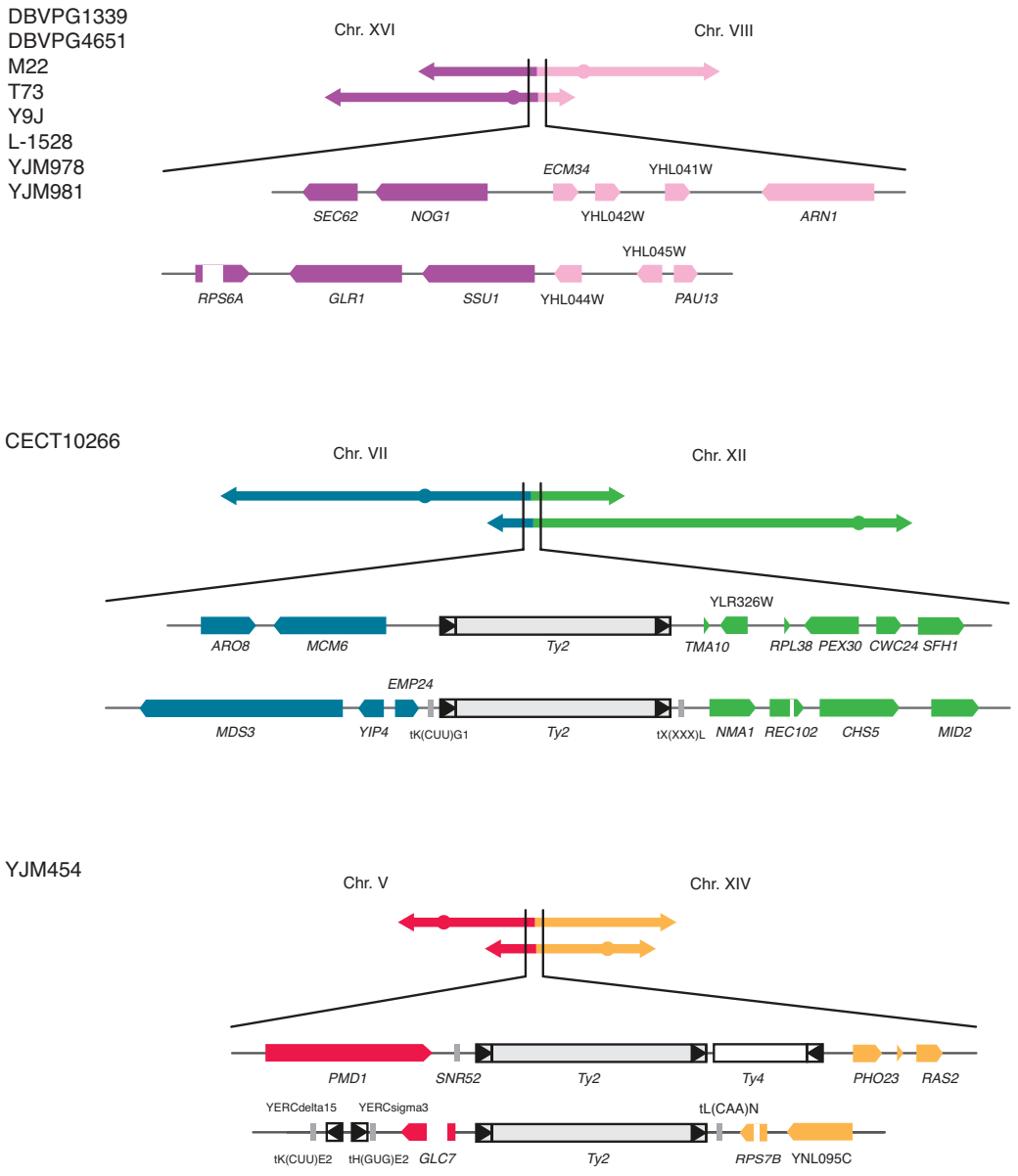


Figure 5. Identified translocations responsible for the observed reproductive isolation. Schematics of translocations identified in this study. Chromosome pairs involved are color-coded. Chromosome and gene sizes are scaled according to SGD annotations.

Discussions

The process of speciation is often quantitative, as the strength of reproductive isolation varies continuously at different levels of divergence[2]. The yeast *Saccharomyces cerevisiae* and its close relatives in the *Saccharomyces sensu strico* complex offer a unique opportunity to explore the possible mechanisms leading to the onset of intrinsic reproductive isolation at both "short" (within species) and "long" (between species) evolutionary scales. Including *S. cerevisiae*, six species are currently circumscribed in this group[22], all of which readily cross with each other to form viable hybrids[23]. Yet, interspecific hybrids showed strong post-zygotic reproductive isolation, producing only ~1% of viable offspring[23,24]. Many species in this group differ by chromosomal rearrangements[7,24,25], however, as this only partially explains the substantial loss of hybrid progeny due to the extant high interspecific divergence, the relative role of translocations in the onset of reproductive isolation and speciation in these species was largely debated[7,26,27].

By performing a systematic survey across a large collection of natural isolates, we found that chromosomal rearrangements, especially reciprocal translocations, play a substantial role in the onset of reproductive isolation in *S. cerevisiae*. The fact that this type of mechanism exists at different temporal levels of genetic divergence, both within and between species, suggests that reciprocal translocations might have a larger impact to the onset of speciation in yeast than previously thought.

Adaptation through chromosomal rearrangements is common in *S. cerevisiae*

Chromosomal rearrangements including polyploidies, aneuploidies, segmental duplications and translocations, are frequently observed in wild and domesticated strains of *S. cerevisiae*[28-31]. Such rearrangements could readily be associated with adaptation to environmental stress. One textbook example was a reciprocal translocation between chromosome VIII and XVI, which has been observed in several wine strains[32]. This translocation was also identified in different strains in this study, explaining the reduced offspring viability of ~75% when crossed with S288c (Figure 5). In fact, this translocation has led to reorganization of the promoter region for gene adjacent to the junction on chromosome XVI, which permits the consecutive overexpression of the gene *SSU1*[32]. This overexpression results in a sulfite resistant phenotype, which conferred to an adaptive advantage for wine strains, as sulfite was a commonly

used compound in wine making[32,33]. Another example has been recently found in strains isolated from a copper contaminated site (Evolution Canyon) in Israel [34]. In those strains, translocation and segmental duplication involving chromosome VII and VIII were repeatedly observed, leading to increasing copy number of copper resistance related genes *CUP1*, *CUP2* and *COX23*, which were essential for strains to survive in high copper concentration environment.

Not only do these genomic changes occur frequently in nature, adaptive chromosomal rearrangements are also commonly observed in short-term evolution experiments in laboratory settings[35-37]. For example, in a chemostat experiment, a beneficial translocation between chromosome VII and XV was appeared in a glucose restricted environment through only ~150 generations, leading to tricarboxylic pathway repression by re-modulating a key regulator, *CIT1*, at the junction of the translocation [35]. These observations, in agreement with our data, suggest that chromosomal rearrangements might offer a mechanism of rapid response to environmental stress in *S. cerevisiae*. When beneficial, such rearrangements could overcome the strong selective disadvantage regarding the potential loss of progeny and become fixed in the population.

Do Dobzhansky-Müller incompatibilities exist in yeast?

In theory, the Dobzhansky-Müller model of genetic incompatibility offers the inherent link between divergent adaptation and reproductive isolation. If two populations are evolved to adapt different environments, mutations accumulated independently in each specialization may cause negative interactions which reduce hybrid fitness or viability[19]. To date, a few pairs of "Dobzhansky-Müller genes" have been identified in plants, insects and animals, both among and within species[1,38-42]. Curiously, between different yeast species, no nuclear genetic incompatibility has ever been found despite repeated effort[43-45].

As natural populations of *S. cerevisiae* are highly structured according to their ecological niches[3], it is tempting to speculate that mutations accumulated during their adaptation to different environment can lead to genetic incompatibility. However, by screening a large collection of ecologically diverse strains of *S. cerevisiae*, we found no Dobzhansky-Müller genetic incompatibilities, suggesting that genetic incompatibility may play a minor role to the onset of reproductive isolation in yeast. Nevertheless, the lack of awareness concerning

such incompatibilities in yeast might be due to the incomplete penetrance of antagonic genetic interactions on permissive rich media. Indeed, more than a few studies have shown that sometimes incompatibilities can only be observed in specific environmental conditions, such as media supplemented with different carbon sources[8,9,46] or defined temperatures[47-49]. In a well-designed experimental evolution in *S. cerevisiae*, authors independently evolved two populations from the same ancestor in high-salt or low-glucose media to foster allopatric speciation[15,16] and demonstrated that divergent evolution of these populations has led to genetic incompatibilities, which can be observed in either high-salt or low-glucose conditions, but not in rich media. Considering the complexity of environmental fluctuations in their natural habitat, condition related genetic incompatibilities in yeast might be common, which in turn could contribute, at least partly, to barriers to gene flow in nature. Future research should explore more environmental factors such as temperature, media composition or exposure to various chemical compounds in order to get a more complete picture of the molecular mechanisms involved in the onset of intraspecific reproductive isolations in *S. cerevisiae*.

Publication related to this chapter:

Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* **24**, 1153-9 (2014).

References

1. Maheshwari, S. & Barbash, D.A. The genetics of hybrid incompatibilities. *Annu Rev Genet* **45**, 331-55 (2011).
2. Nosil, P. *Ecological speciation*, xvii, 280 p. Oxford University Press (2012).
3. Schacherer, J., Shapiro, J.A., Ruderfer, D.M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342-5 (2009).
4. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-41 (2009).
5. Wang, Q.M., Liu, W.Q., Liti, G., Wang, S.A. & Bai, F.Y. Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Mol Ecol* **21**, 5404-17 (2012).
6. Cromie, G.A. *et al.* Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda)* **3**, 2163-71 (2013).
7. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. & Louis, E.J. Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451-4 (2000).
8. Chou, J.Y., Hung, Y.S., Lin, K.H., Lee, H.Y. & Leu, J.Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol* **8**, e1000432 (2010).
9. Lee, H.Y. *et al.* Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065-73 (2008).
10. Chambers, S.R., Hunter, N., Louis, E.J. & Borts, R.H. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol Cell Biol* **16**, 6110-20 (1996).
11. Hunter, N., Chambers, S.R., Louis, E.J. & Borts, R.H. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J* **15**, 1726-33 (1996).
12. Greig, D., Travisano, M., Louis, E.J. & Borts, R.H. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. *J Evol Biol* **16**, 429-37 (2003).
13. Heck, J.A. *et al.* Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3256-3261 (2006).
14. Demogines, A., Wong, A., Aquadro, C. & Alani, E. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet* **4**, e1000103 (2008).
15. Dettman, J.R., Sirjusingh, C., Kohn, L.M. & Anderson, J.B. Incipient speciation by divergent adaptation and antagonistic epistasis in yeast. *Nature* **447**, 585-8 (2007).
16. Anderson, J.B. *et al.* Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. *Curr Biol* **20**, 1383-8 (2010).
17. Maclean, C.J. & Greig, D. Reciprocal gene loss following experimental whole-genome duplication causes reproductive isolation in yeast. *Evolution* **65**, 932-45 (2011).
18. Kroll, E. *et al.* Starvation-associated genome restructuring can lead to reproductive isolation in yeast. *PLoS One* **8**, e66414 (2013).
19. Dobzhansky, T. *Genetics and the origin of species*, xvi, 364 p. Columbia Univ. Press (1937).
20. Muller, H. Isolating mechanisms, evolution and temperature. *Biol. Symp* Vol. 6 71-125 (1942).
21. Loidl, J., Jin, Q.W. & Jantsch, M. Meiotic pairing and segregation of translocation quadrivalents in yeast. *Chromosoma* **107**, 247-54 (1998).
22. Scannell, D.R. *et al.* The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the *Saccharomyces sensu stricto* genus. *G3 (Bethesda)* **1**, 11-25 (2011).

23. Naumov, G.I. *et al.* Natural polyploidization of some cultured yeast *Saccharomyces sensu stricto*: auto- and allotetraploidy. *Syst Appl Microbiol* **23**, 442-9 (2000).

24. Liti, G., Barton, D.B. & Louis, E.J. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**, 839-50 (2006).

25. Ryu, S.L., Murooka, Y. & Kaneko, Y. Genomic reorganization between two sibling yeast species, Saccharomyces bayanus and *Saccharomyces cerevisiae*. *Yeast* **12**, 757-64 (1996).

26. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68-72 (2003).

27. Greig, D. Reproductive isolation in *Saccharomyces*. *Heredity (Edinb)* **102**, 39-44 (2009).

28. Albertin, W. *et al.* Evidence for autotetraploidy associated with reproductive isolation in *Saccharomyces cerevisiae*: towards a new domesticated species. *J Evol Biol* **22**, 2157-70 (2009).

29. Rachidi, N., Barre, P. & Blondin, B. Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. *Mol Gen Genet* **261**, 841-50 (1999).

30. Bidenne, C., Blondin, B., Dequin, S. & Vezinhet, F. Analysis of the chromosomal DNA polymorphism of wine strains of *Saccharomyces cerevisiae*. *Curr Genet* **22**, 1-7 (1992).

31. Casaregola, S. *et al.* A family of laboratory strains of *Saccharomyces cerevisiae* carry rearrangements involving chromosomes I and III. *Yeast* **14**, 551-64 (1998).

32. Perez-Ortin, J.E., Querol, A., Puig, S. & Barrio, E. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res* **12**, 1533-9 (2002).

33. Goto-Yamamoto, N. *et al. SSU1-R*, a sulfite resistance gene of wine yeast, is an allele of SSU1 with a different upstream sequence. *Journal of Fermentation and Bioengineering* **86**, 427-433 (1998).

34. Chang, S.L., Lai, H.Y., Tung, S.Y. & Leu, J.Y. Dynamic large-scale chromosomal rearrangements fuel rapid adaptation in yeast populations. *PLoS Genet* **9**, e1003232 (2013).

35. Dunham, M.J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci U S A* **99**, 16144-9 (2002).

36. Rancati, G. *et al.* Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135**, 879-93 (2008).

37. Yona, A.H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress. *Proc Natl Acad Sci U S A* **109**, 21010-5 (2012).

38. Coyne, J.A. & Orr, H.A. *Speciation*, xiii, 545, Sinauer Associates, Sunderland (2004).

39. Seidel, H.S. *et al.* A novel sperm-delivered toxin causes late-stage embryo lethality and transmission ratio distortion in *C. elegans*. *PLoS Biol* **9**, e1001115 (2011).

40. Seidel, H.S., Rockman, M.V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589-94 (2008).

41. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**, 623-6 (2009).

42. Presgraves, D.C. The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175-80 (2010).

43. Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet* **3**, e21 (2007).

44. Kao, K.C., Schwartz, K. & Sherlock, G. A genome-wide analysis reveals no nuclear dobzhansky-muller pairs of determinants of speciation between *S. cerevisiae* and *S. paradoxus*, but suggests more complex incompatibilities. *PLoS Genet* **6**, e1001038 (2010).

45. Li, C., Wang, Z. & Zhang, J. Toward genome-wide identification of bateson-dobzhansky-muller incompatibilities in yeast: a simulation study. *Genome Biol Evol* **5**, 1261-72 (2013).

46. Chou, J.Y. & Leu, J.Y. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays* **32**, 401-11 (2010).

47. Bomblies, K. & Weigel, D. Hybrid necrosis: autoimmunity as a potential gene-flow barrier in plant species. *Nat Rev Genet* **8**, 382-93 (2007).

48. Alcazar, R., Garcia, A.V., Parker, J.E. & Reymond, M. Incremental steps toward incompatibility revealed by *Arabidopsis* epistatic interactions modulating salicylic acid pathway activation. *Proc Natl Acad Sci U S A* **106**, 334-9 (2009).

49. Alcazar, R. & Parker, J.E. The impact of temperature on balancing immune responsiveness and growth in *Arabidopsis*. *Trends Plant Sci* **16**, 666-75 (2011).

50. Skelly, D.A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-504 (2013).

Chapter 3

Comprehensive survey of condition specific reproductive isolation reveals genetic incompatibility in yeast

Introduction

Using a large number of natural isolates of *S. cerevisiae,* we carried out a first comprehensive effort characterizing the onset of reproductive isolation within a species[1]. We identified chromosomal rearrangements segregating in diverse populations and acting as the major mechanism leading to reduced offspring viabilities observed in 16% of the crosses. In parallel, a study within *S. paradoxus* populations reached similar conclusion[2]. While chromosomal rearrangements seemed to be widely distributed both within and between closely related yeast species[3,4] and could contribute to reproductive isolation at both evolutionary scales, no evident case of classic Dobzhansky-Müller incompatibility has been found so far in natural populations of yeast, contrasting to other model organisms[5,6]. Nevertheless, all studies up until now were performed under laboratory conditions, which consist of estimating the offspring viability on a rich permissive media that optimize yeast growth. Considering the vast ecological range that natural populations of yeasts encounter in nature[7,8], our view of reproductive isolation cases restricted to laboratory conditions might be overly simplified.

To assess the impact of environmental factors to the onset of reproductive isolation within *S. cerevisiae*, we selected 27 crosses previously shown to yield high offspring viability on rich media, and tested their offspring viability on a large number of conditions (different carbon sources, chemicals and temperatures). Interestingly, these highly compatible crosses could be sometimes incompatible on other conditions. In fact, over 24% (117/481) of the cases tested showed potential negative epistasis, among which 6.7% (32/481) were severe, with at least 20% of progeny loss on tested conditions. We analyzed the segregation patterns of identified cases and focus on one case that demonstrated a potential recessive two loci Dobzhansky-Müller incompatibility related to the loss of offspring viability on media containing non-fermentable carbon sources (e.g. glycerol and ethanol). By further analyzing the genes and mutations involved, we showed that the incompatibility is due to the presence of a nonsense mutation in a nuclear-encoding mitochondrial gene and a tRNA suppressor. We provided evidence that this precise configuration could be adaptive in fluctuating environments, highlighting the potential role of ecological selection in the onset of genetic incompatibility and reproductive isolation in yeast.

Results

As described previously in Chapter 2, a total of 27 natural isolates compatible with the reference strain S288c (offspring viability > 90% on YPD) were selected (Table 1). All isolates were crossed with S288c and offspring viability was scored and confirmed on YPD (Table 1). For each cross, 20 full tetrads (containing only viable spores) were chosen to be tested on 20 conditions, including different temperatures, carbon sources, and various chemical compounds (Figure 1, Table 2). This summed up to a total of 540 instances spanning 27 crosses on 20 conditions (Figure 1). Among all 540 instances assessed, 59 involved at least one parental strain being non-viable on the condition tested and were excluded for further analysis (Figure 1). Overall, 24.3% of all instances (117/481) showed signs of negative epistasis with different degrees of loss of offspring viability ranging from 1% to 62% (Figure 1). Among these cases, 6.7% (32/481) showed moderate to severe incompatibility, with at least 20% of the segregants being non-viable on the condition tested (Figure 1).

Strains	Source	Location	Offspring viability	Divergence to S288c(%)	Experiment	Reference
CECT10109	Prickly pear	Spain	97%	0.31	Screen & Stress tolerance	1
CLIB192	Baker	France	92%	0.11	Screen	1
CLIB219	Wine	Russia	92%	0.44	Screen & Stress tolerance	1
CLIB272	Beer	US	95%	0.23	Screen & Stress tolerance	1
CLIB294	Fermentation	France	94%	0.25	Screen & Stress tolerance	1
CLIB382	Beer	Japan	92%	0.25	Screen & Stress tolerance	1
DBVPG3591	Cocoa beans		93%	0.23	Screen & Stress tolerance	1
DBVPG6861	Polluted water		96%		Screen	1
EM93	Rotting fig	US	91%	0.14	Screen	1
I14	Vineyard	Italy	94%	0.25	Screen & Stress tolerance	1
K12	Sake	Japan	91%	0.25	Screen	1
PW5	Palm wine	Nigeria	91%	0.59	Screen	1
T7	Oak tree	US	100%	0.49	Screen & Stress tolerance	1
UC8	Wine	South Africa	99%	0.28	Screen & Stress tolerance	1
Y12	Palm wine	Africa	92%	0.35	Screen & Stress tolerance	1
Y3	Palm wine	Africa	91%	0.38	Screen & Stress tolerance	1
Y9	Ragi	Indonesia	92%	0.34	Screen & Stress tolerance	1
YJM269	Apple juice		92%	0.38	Screen & Stress tolerance	1
YJM280	Clinical	US	92%	0.35	Screen	1
YJM320	Clinical	US	96%	0.32	Screen & Stress tolerance	1
YJM326	Clinical	US	91%	0.32	Screen & Stress tolerance	1
YJM421	Clinical	US	95%	0.35	Screen & Stress tolerance	1
YJM440	Clinical	US	90%		Screen & Stress tolerance	1
YJM653	Clinical	US	94%	0.32	Screen & Stress tolerance	1
YJM678	Clinical		95%		Screen & Stress tolerance	1
YPS1000	Oak exudate	US	90%	0.41	Screen & Stress tolerance	1

Table 1. Origin, divergence and offspring viability estimates of strains used in this study

Potential case of two loci Dobzhansky-Müller incompatibility

To assess the genetic complexity of the observed cases, we analyzed the segregation patterns of the lethal phenotype. Most of the cases were consistent with complex epistasis and were not characterized in this study (~103/117). The remaining 14 cases were found in six parental combinations, related to various conditions including the presence of NaCl (1M), high temperature (42°C), SDS (0.03% and 0.06%), cycloheximide (0.1 µg/ml) and non-fermentable carbon sources. We focused on one cross between a clinical isolate YJM421[9] and S288c, which showed a clear pattern of recessive two loci Dobzhansky-Müller incompatibility in several conditions related to respiration efficiency (YP sorbitol 2%, YP glycerol 2%, YP ethanol 2% and YP galactose

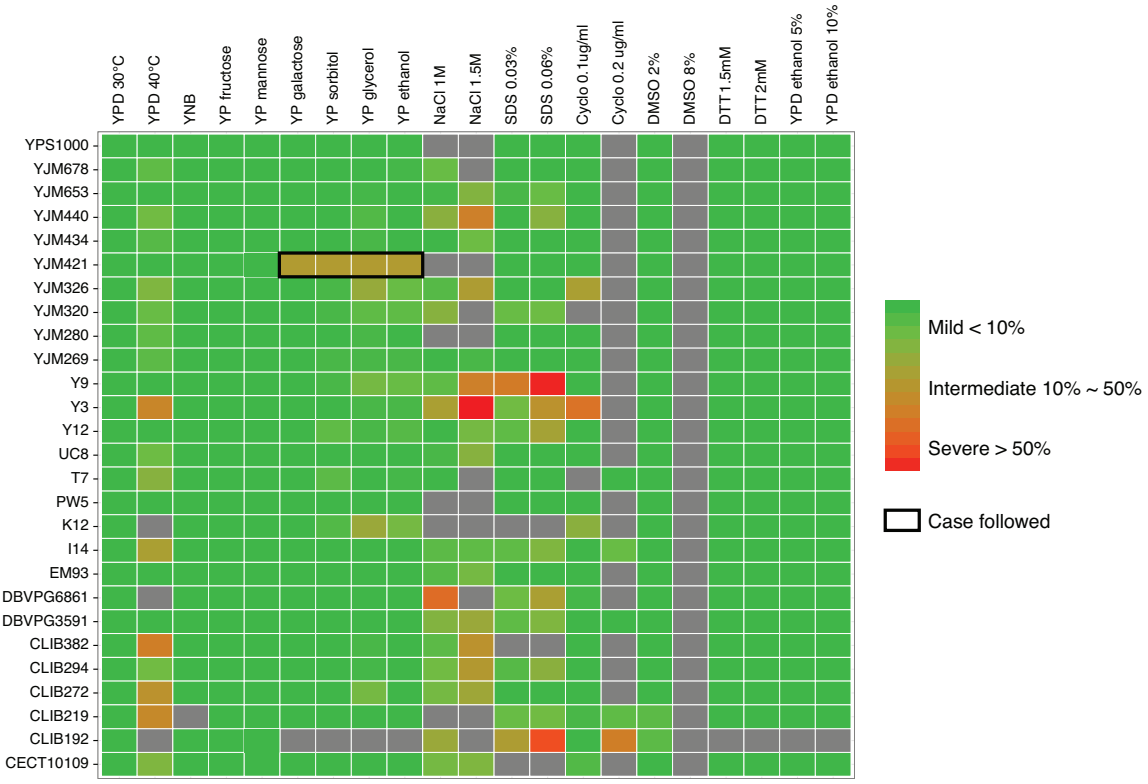


Figure 1. Offspring viability of 27 natural isolates crossed with S288c on 20 conditions. Offspring viabilities estimated based on 20 full tetrads are color-coded with the vertical axis representing isolates crossed and horizontal axis representing the 20 conditions tested. All isolates were previously shown to produce high offspring viability (> 90%) on YPD when crossed with S288c[1]. Conditions where either one or both parental strains were non viable are colored in grey. The case followed is circled in black.

2%). In this scenario, the lethal allelic combination should follow Mendelian segregation which leads to 1/4 in the loss of viability in the offspring, resulting

in a ratio of 1:4:1 for tetrads containing 4, 3 or 2 viable spores, respectively, assuming the interacting loci are independent (Figure 2). For this cross, 100 additional tetrads were tested on YP glycerol 2% and the segregation pattern was confirmed (Figure 2). Approximately 25% of the offspring were respiration deficient and were unable to grow on YP glycerol 2%.

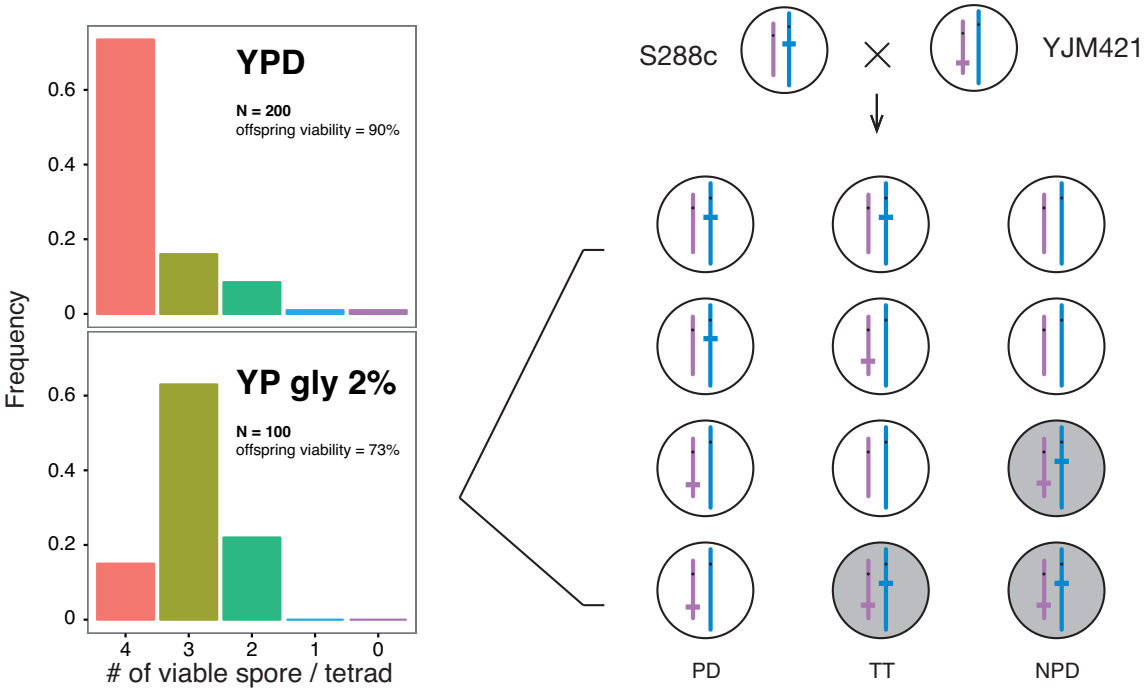


Figure 2. Phenotypic segregation pattern of the incompatibility between YJM421 and S288c. The frequency of tetrads containing 4, 3, 2, 1 or 0 viable segregants was presented for YPD (upper plot) and YP glycerol 2% (lower plot). The number of tetrads tested are as indicated.

Mapping of the loci involved using bulk segregant analysis

To map the loci involved, we used a bulk segregant analysis strategy followed by whole genome sequencing (BSA-seq). Briefly, 80 segregants that were non-viable on YP glycerol 2% from independent tetrads were pooled and sequenced as described previously in Chapter 2. The sequences obtained were aligned to the genome of S288c and the allele frequency of S288c was scored at each polymorphic position. For most genomic regions, the expected allele frequency for both parental strains was ~0.5, whereas the loci involved in the incompatibility would have deviated allele frequencies. Using this strategy, we mapped two regions with significant allele frequency deviation, one located on the right arm of chromosome V (position 413107 to 458959) and the second on

the left arm of chromosome X (position 331633 to 364022), spanning approximately 46 kb and 33 kb, respectively (Figure 3).

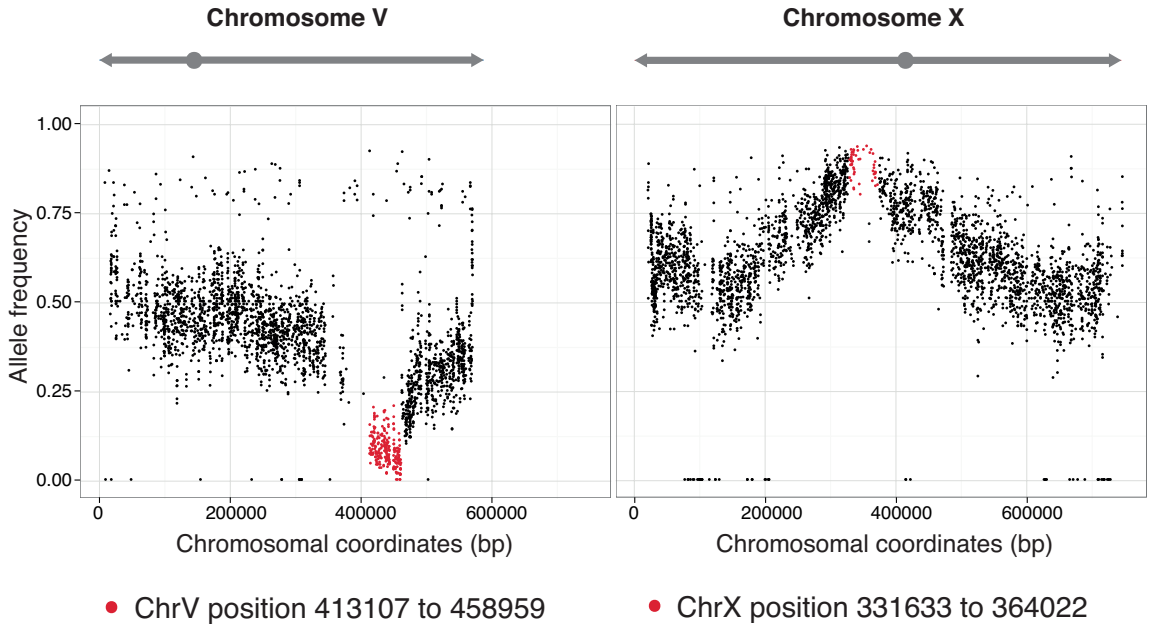


Figure 3. Genomic regions with skewed allele frequencies identified using bulk segregant analysis. A total of 80 non-viable segregants on YP glycerol 2% were pooled and sequenced and two candidate regions were identified. The horizontal axis represents the coordinates of chromosome V and X. The vertical axis corresponds to the allele frequencies of S288c. Identified regions are colored in red.

Identification and functional validation of candidate genes

To identify the causative genes for the observed respiratory deficiency, we closely examined the mapped regions for potential candidates. In total, five genes in these regions were potentially involved in respiration according to the SGD annotations (http://www.yeastgenome.org/), among which three were found in the region on chromosome V (*EMP65*, *COX15* and *FTR1*) and two in the region on chromosome X (*TIM54* and *AIM22*). We examined the DNA sequences of these genes in the YJM421 background and found a nonsense mutation at the position +115 in the open reading frame of *COX15* (CAA to TAA; position 453574 on chromosome V). *COX15* encodes an inner membrane cargo protein in the mitochondria, the function of which is essential for respiration[10]. The observation of a nonsense mutation in this gene was surprising as the presence of such mutation would likely abolish the function of *COX15* and lead to respiratory deficiency, whereas the strain YJM421 carrying the mutation was respiratory competent.

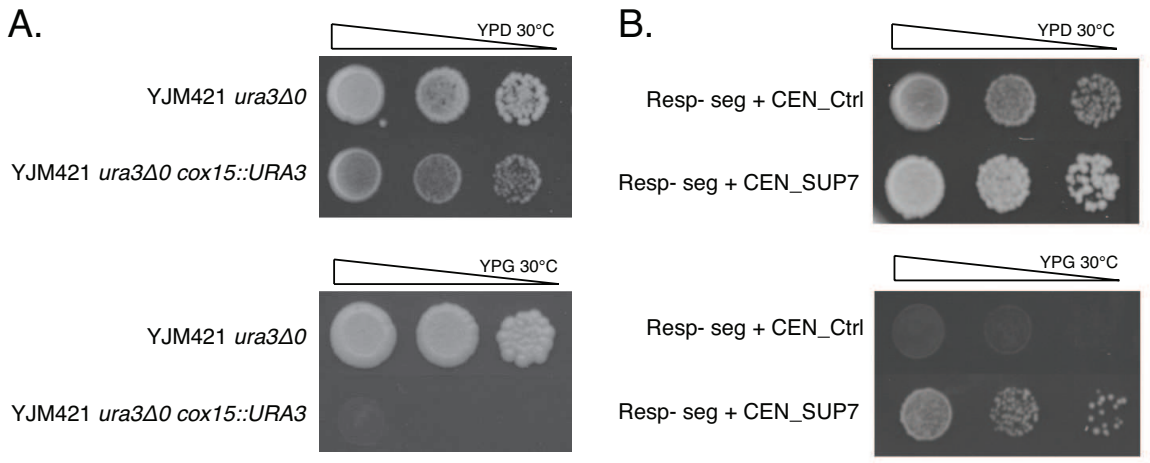


Figure 4. Functional validation of the incompatible gene pair. A. Deletion of *cox15*^{stop} in the YJM421 background results in respiratory deficiency. Strains with or without deletion of *cox15*^{stop} (YJM421 *ura3Δ0 cox15*^{stop}*::URA3* and YJM421 *ura3Δ0*, respectively) in exponential growth phase were spotted in 3 dilutions onto YPD (upper plot) and YP glycerol 2% (lower plot). Cells were grown for 48 hours at 30°C. **B. Rescue of respiration capacity in non-viable segregants on YP glycerol 2% with *SUP7*.** Respiration deficient F1 segregant was transformed with plasmids containing *SUP7* or empty control and grown for 48 hours at 30°C on YPD (upper plot) and YP glycerol 2% (lower plot).

To verify if the YJM421 allele of *COX15* (*cox15*^{stop}) was functional, we deleted the *cox15*^{stop} in YJM421, and the resulting mutant was unable to grow on media containing non-fermentable carbon sources (Figure 4A). Moreover, allele replacement of *cox15*^{stop} with the wild type *COX15* from S288c in the YJM421 background resulted in total rescue of the genetic incompatibility observed: cross between YJM421 *cox15*^{stop}*::COX15* and S288c led to 98.6% offspring viability on YP glycerol 2% (400 segregants tested; Figure 5). These results confirmed that *cox15*^{stop} was functional in the YJM421 background and was involved in the incompatibility between YJM421 and S288c. The fact that *cox15*^{stop} was functionally active strongly suggests the presence of a genetic element at the interacting loci on chromosome X that compensates the effect of the nonsense mutation in YJM421.

Indeed, when examining the DNA sequence of YJM421 in the mapped region on chromosome X, we found a mutation at the anticodon position of a tyrosine tRNA tY(GUA)J1 (GTA to TTA, position 354280 on chromosome X), which in turn transformed this tRNA into a TAA nonsense suppressor (*SUP7*). The presence of this suppressor would effectively read-through the premature stop codon in *cox15*^{stop}, which leads to a functional protein product in YJM421.

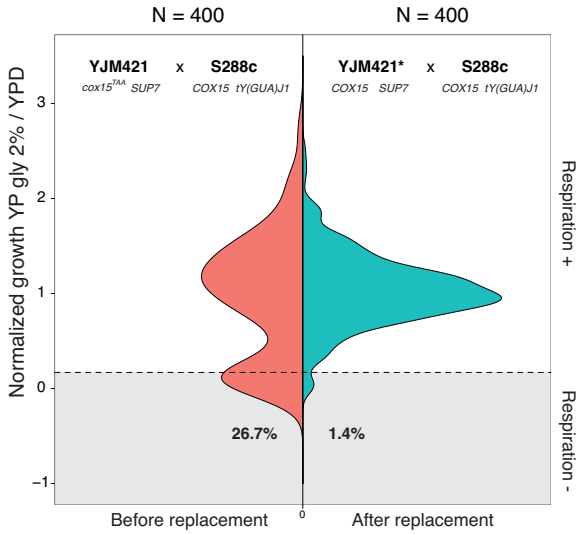


Figure 5. Fitness distribution of segregants obtained before and after allele replacement of *COX15*. Normalized growth ratio of 400 segregants from the cross between YJM421 (*cox15^{stop}/SUP7*; left panel) or YJM421* (*COX15/ SUP7*; right panel) and S288c are presented as color coded frequency distributions. Shaded areas indicate the fractions of segregants that are respiratory deficient.

However, this configuration of *cox15^{stop}/SUP7* in YJM421 renders the strain incompatible when crossed with S288c, as 1/4 of the segregants would inherent only the non-functional *cox15^{stop}* allele but not the suppressor, leading to respiratory deficiency (Figure 6). To confirm this hypothesis, we transformed segregants that are non-viable on YP glycerol 2% with a yeast centromeric plasmid containing the suppressor *SUP7* (CEN_SUP7), and confirmed that the presence of this suppressor restored their respiration capacity (Figure 4B). These results demonstrated the first identified pair of Dobzhansky-Müller incompatibility genes within a yeast species. Nevertheless, the evolutionary and physiological implications of this specific combination of *cox15^{stop}/SUP7* are still unclear.

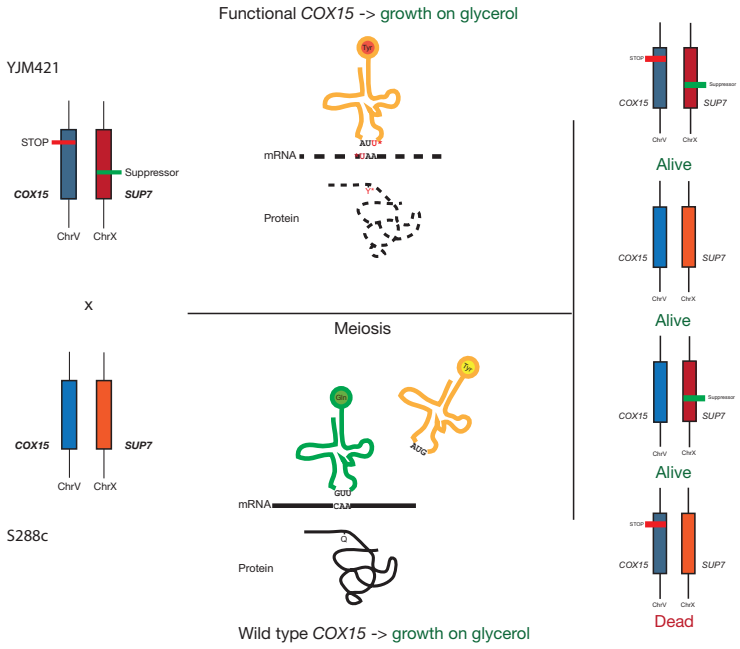


Figure 6. Allelic combination of the incompatible gene pair. Parental combinations were respiratory competent in S288c and YJM421. Segregants that inherited the non-functional *COX15* from YJM421 but not the suppressor cannot restore the function of *COX15* and therefore are non-viable in the presence glycerol.

Differential fitness effect of *SUP7* in diverse isolates

In fact, tRNA suppressors are well known to effectively suppress nonsense mutations by stop codon read-through, although the presence of such suppressors is likely detrimental due to the perturbation of cellular translational fidelity. As the incompatible strain YJM421 did not show any apparent growth defect, we sought to evaluate the effect of the suppressor *SUP7* on growth in different genetic backgrounds. We transformed 23 diverse natural isolates (Table 1) with a plasmid containing *SUP7* (CEN_SUP7) or an empty control plasmid (CEN_Ctrl) and measured their growth rate in a non-stressful condition (YPD at 30°C) using microcultures. A mean reduction of the growth rate across all strains tested was observed (mean reduction 22.8%, N = 138, two-sided t-test p-value << 0.005), with the most severe case of 2.53 folds lower growth in the presence of the suppressor compared to the strain carrying the control plasmid (Figure 7A; strain Y9 with two-sided t-test p-value < 0.05, N = 6). Interestingly, despite an overall deleterious effect of *SUP7*, several isolates, including YJM421, YJM320 and T7, showed similar or higher growth rates in the presence of the suppressor (Figure 7A). These results suggest that the effect of *SUP7* on growth is background dependent and different levels of genetic assimilation could be observed, such as the case for YJM421, thus allowing for the persistence of *SUP7* in this strain.

Condition	Composition	Stress type	Experiment
YPD 30°C	1% yeast extract; 2% peptone; 2% glucose	Rich media	Screen & Stress tolerance
YPD 37°C	1% yeast extract; 2% peptone; 2% glucose	Rich media	Stress tolerance
YPD 40°C	1% yeast extract; 2% peptone; 2% glucose	Rich media	Screen & Stress tolerance
YNB	0.67% yeast nitrogen base w/ ammonium sulfate	Minimum media	Screen
YP gal 2%	1% yeast extract; 2% peptone; 2% galactose	Carbon utilization	Screen
YP fru 2%	1% yeast extract; 2% peptone; 2% fructose	Carbon utilization	Screen
YP man 2%	1% yeast extract; 2% peptone; 2% mannose	Carbon utilization	Screen
YP sorb 2%	1% yeast extract; 2% peptone; 2% sorbitol	Carbon utilization	Screen
YP gly 2%	1% yeast extract; 2% peptone; 2% glycerol	Carbon utilization	Screen & Stress tolerance
YP eth 2%	1% yeast extract; 2% peptone; 2% ethanol	Carbon utilization	Screen & Stress tolerance
YP glu 8%	1% yeast extract; 2% peptone; 8% glucose	Carbon utilization	Stress tolerance
YP glu 0.01%	1% yeast extract; 2% peptone; 0.01% glucose	Carbon utilization	Stress tolerance
NaCl 1 M	YPD; NaCl 1 M	Signal transduction	Screen & Stress tolerance
NaCl 1.5 M	YPD; NaCl 1.5 M	Signal transduction	Screen
KCl 1M	YPD; KCl 1 M	Osmotic stress	Stress tolerance
SDS 0.03%	YPD; SDS 0.03%	Membrane stability	Screen
SDS 0.06%	YPD; SDS 0.06%	Membrane stability	Screen & Stress tolerance
CHX 0.1 µg/ml	YPD; Cycloheximide 0.1 µg/ml	Translational stress	Screen & Stress tolerance
CHX 0.2 µg/ml	YPD; Cycloheximide 0.2 µg/ml	Translational stress	Screen
YPD eth 5%	YPD; ethanol 5%	Proteome stability	Screen
YPD eth 10%	YPD; ethanol 10%	Proteome stability	Screen
YPD eth 15%	YPD; ethanol 15%	Proteome stability	Stress tolerance
DMSO 2%	YPD; DMSO 2%	Oxidative stress	Screen
DMSO 8%	YPD; DMSO 8%	Oxidative stress	Screen
DTT 1.5 mM	YPD; DTT 1.5 mM	Reductive stress	Screen
DTT 2 mM	YPD; DTT 2 mM	Reductive stress	Screen

Table 2. Media compositions used for screening and stress tolerance test

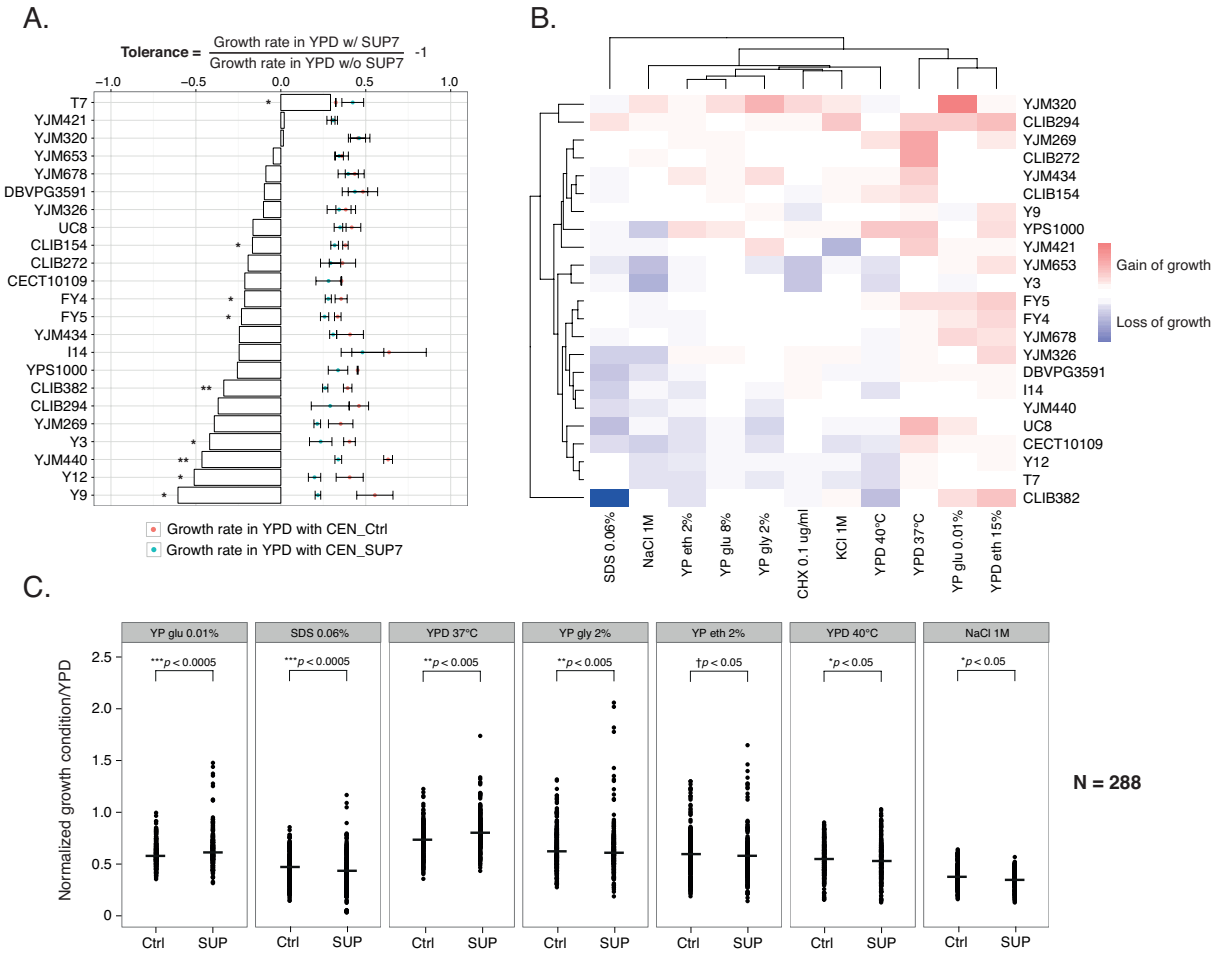


Figure 7. Phenotypic consequences of *SUP7* in various isolates. A. Growth variation in the presence *SUP7* in non-stressful conditions (YPD 30°C). Growth rates of 23 strains measured in liquid YPD at 30°C in the presence of a plasmid containing *SUP7* (blue dots) or empty control vector (red dots) are presented with error bars (mean ± SD; N = 6). Percentage of growth variation is calculated and presented as bars (N = 6; two-sided t-test **p*-value < 0.05; ***p*-value < 0.005). **B. Suppressor induced phenotypic variation in stress conditions.** Significant variation of the normalized growth ratio (> 10%) due to the presence of *SUP7* of 23 strains is presented for 11 stress conditions using a heatmap with blue indicating loss of growth and red for gain of growth compared to strains carrying the empty control vector. **C. Significant increase of phenotypic variance in the presence of *SUP7*.** Distribution of the normalized growth ratio in stress conditions was compared for strains carrying the suppressor *SUP7* (SUP) or control (Ctrl) in seven conditions. Median values for each condition are indicated with a bar. Statistical significance is as shown on the plot (N = 288, two-sided F-test **p*-value < 0.05, ***p*-value < 0.005, ****p*-value < 0.0005; Levene test †*p*-value < 0.05).

Impacts of *SUP7* in stress conditions across natural isolates

To further investigate the phenotypic consequences of *SUP7*, we evaluated the fitness of the same set of 23 isolates (Table 1) carrying the plasmid with *SUP7* (CEN_SUP7) on solid media for various stress conditions (membrane stability, proteome perturbation, osmotic stress, different carbon sources and high temperatures; Figure 7B; Table 2). The normalized growth ratio was calculated by comparing the colony size on tested conditions *vs.* YPD to eliminate the effects of growth differences on YPD and pinning density on solid plates. We then calculated the percentage of fitness variation for each isolate in the presence of *SUP7* compared to the same isolate carrying the control plasmid on each condition. Significant variation due to the presence of *SUP7* was observed in most of the conditions tested, with nearly half of the cases showing a gain of fitness higher than 10% (Figure 7B). These variations appeared to be strain and condition specific, with exceptions for some conditions (YPD 37°C and YPD ethanol 15%) where all strains grew better in the presence of *SUP7*, and some strains (CLIB294, YJM269 and CLIB272) with an overall gain of fitness across all conditions. In addition, for most of the conditions tested (7/11), significantly increased phenotypic variance was observed in the presence of *SUP7* compared to the controls across all strains (Figure 7C). These results suggest that the suppressor *SUP7* contributes to marked phenotypic variation across different genetic backgrounds in stress conditions, and carrying the suppressor might, in turn, offer some selective advantages in the presence of environmental challenges.

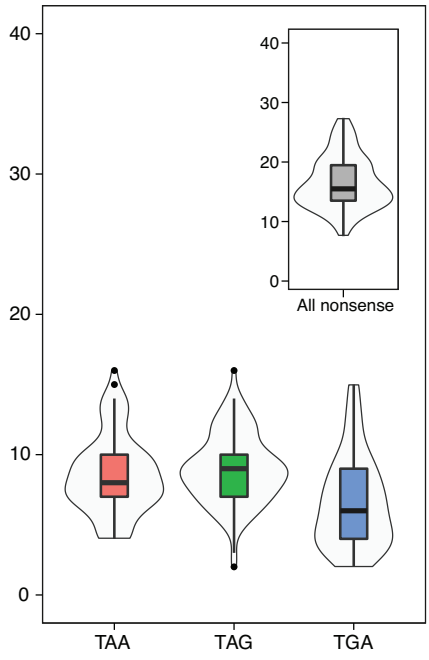


Figure 8. Nonsense mutations present in the natural populations. Distribution and number of nonsense mutations in verified ORFs across 100 sequenced natural isolates[11,12]. Mutations in different stop codon classes are color-coded.

GO Biological Process (2062 categories)

Category	*p*-value	In Category from Cluster	k	f
transmembrane transport [GO:0055085]	2.76E-09	SEO1 VBA2 SUL1 PCA1 PHO89 GEX1 ERS1 GIT1 YDL199C VBA4 CCC2 NHX1 CAN1 HXT13 HVG1 FCY21 ALR2 AGP3 MAL11 DUR3 VMR1 YHK8 QDR1 DAL4 PAM16 TRK1 HXT8 HXT9 MCH2 GAP1 NFT1 GEX2 MMP1 NHA1 YLR152C AQR1 TIM23 BIO5 TAT2 HXT11 YOL162W YOL163W NRT1 THI72 COT1 SSU1 PXA1 YPR011C OPT2	49	303
cellular monovalent inorganic cation homeostasis [GO:0030004]	0.00029	NHX1 NHA1 VHS3	3	3
amino acid transport [GO:0006865]	0.0003189	VBA2 VBA4 CAN1 AUA1 AGP3 AVT3 GAP1 MMP1 BIO5 TAT2	10	42
maltose metabolic process [GO:0000023]	0.000415	MAL33 MAL32 MAL13 MAL11 MAL12	5	11
cellular cell wall organization [GO:0007047]	0.001996	ECM8 ECM11 HLR1 ECM34 ECM12 ECM14 PGU1 ECM4 CTS1 ECM19 ECM30 WSC2 TIR4 SPR1	14	89
flocculation [GO:0000128]	0.002621	FLO1 MUC1 FLO10	3	5
nucleobase transport [GO:0015851]	0.002908	FCY21 DAL4 NRT1 THI72	4	10
amino acid transmembrane transport [GO:0003333]	0.003968	CAN1 AGP3 GAP1 MMP1 BIO5 TAT2	6	24
nucleobase, nucleoside, nucleotide and nucleic acid transport [GO:0015931]	0.004332	FCY21 DAL4 NRT1 THI72	4	11
maltose catabolic process [GO:0000025]	0.004391	MAL32 MAL12	2	2
glutathione transmembrane transport [GO:0034775]	0.004391	GEX1 GEX2	2	2
copper ion export [GO:0060003]	0.004391	PCA1 CCC2	2	2
cinnamic acid catabolic process [GO:0046281]	0.004391	PAD1 FDC1	2	2
triglyceride mobilization [GO:0006642]	0.004391	TGL4 TGL5	2	2
pseudohyphal growth [GO:0007124]	0.00887	CDC39 TMN2 STE12 FKH1 MUC1 PGU1 PHD1 SPH1 DFG16 HMS1	10	64

MIPS Functional Classification (459 categories)

Category	p-value	In Category from Cluster	k	f
amino acid/amino acid derivatives transport [20.01.07]	0.0001246	VBA2 ERS1 CAN1 AUA1 AGP3 AVT3 GAP1 MMP1 BIO5 TAT2 NRT1	11	45
transcription repression [11.02.03.04.03]	0.0003351	CRF1 MIG3 RME1 HYM1 MOT3 WHI5 SFL1 ROX1	8	28
detoxification [32.07]	0.002122	GEX1 GRX3 VBA4 PAD1 ALR2 ROG3 SLH1 YHK8 YJR015W NFT1 GEX2 AQR1 SSU1	13	80
transport facilities [20.03]	0.004513	SEO1 GEX1 MCH2 PTR2 GEX2 TIM23 BIO5 YOL162W YOL163W NRT1 THI72 YPR011C OPT2	13	87
cellular import [20.09.18]	0.006049	APL3 SUL1 CAN1 HXT13 ALR2 DAL4 TRK1 HXT8 HXT9 FRE2 GAP1 TAT2 HXT11	13	90
secondary metabolism [01.20]	0.006162	COQ4 GRE3 AYT1 GCY1	4	12

Table 3. GO term enrichment results of genes carrying TAA nonsense mutation in 100 diverse natural isolates

Frequency of nonsense mutation and tRNA suppressor in yeast

To explore the prevalence of nonsense mutations and tRNA suppressors in natural populations of yeast, we surveyed 100 genomes of *S. cerevisiae* that are publically available[11,12]. Compared to common lab strains, nonsense mutations in natural isolates were quite frequent, with an average of ~10 nonsense mutations in each stop codon class per strain (Figure 8). The frequency of nonsense mutations globally followed a normal distribution, with a maximum frequency of nonsense mutations per strain of ~16 mutations for each anticodon class (Figure 8). Genes bearing any class of nonsense mutations were functionally enriched for stress related activities, such as transmembrane transporter activity, detoxification, and transcription regulation (Table 3). More than 40% (215/500) of the detected genes with nonsense mutations were shared by at least two isolates.

In addition to nonsense mutations, we also looked for the presence of potential tRNA suppressors in these genomes. No tRNA suppressor of any anticodon class was found in this rather large data set. In contrast to the prevalence of nonsense mutations in natural populations, the frequency of tRNA suppressors is extremely rare, possibly suggesting a transient role of the suppressors in adaptation in *S. cerevisiae*.

Discussion

We performed a species-wide survey of environment dependent reproductive isolation and identified the first Dobzhansky-Müller incompatibility gene pair related to offspring respiratory deficiency within the yeast species *Saccharomyces cerevisiae*. We showed the incompatibility was due to a combination of a nonsense mutation in *COX15* and a tRNA suppressor *SUP7* in a single isolate, which leads to 1/4 of the offspring having only a non-functional copy of *COX15* upon crossing with the reference strain S288c. We also provided evidence that the persistence of this particular allelic combination might potentially be related to increased evolutionary potential when facing fluctuating environmental conditions in nature. Our study highlights the importance of understanding the ecological context to hybrid fitness and extends the overview of possible mechanisms involved in the onset of intraspecific post-zygotic reproductive isolation in yeast.

The current landscape of intraspecific reproductive isolation in yeasts

In yeasts, multiple mechanisms such as chromosomal rearrangements, anti-recombination, cyto-nuclear incompatibility and meiotic drive elements, have been identified to explain the observed loss of hybrid fertility between different species[3,4,13-17]. However, the relative role of Dobzhansky-Müller genetic incompatibility to the onset of reproductive isolation in yeasts has long been a subject of debate, primarily due to lack of empirical support[18-21]. At the intraspecific level, large-scale chromosomal rearrangements such as reciprocal translocations were considered to be the major mechanism leading to reduced offspring viability when crossing natural populations[1,2], whereas cases of deleterious genic interactions were found to be rare, with the only example demonstrated in *S. cerevisiae* related to interactions between genes in the mismatch repair system (MMR) leading to sporadic progeny loss[22,23]. In addition to these mechanisms, here we showed that classic two loci Dobzhansky-Müller incompatibilities do exist in natural isolates of yeast and could readily lead to reproductive isolation in different environmental conditions. While the conditions investigated in the current study do not represent the true ecological contexts encountered, it is evident that the overall picture of molecular mechanisms affecting reproductive traits in nature is far more complex than previously envisioned within a yeast species.

Evolutionary origin and maintenance of the identified incompatibility

Nevertheless, the evolutionary forces driving the onset and maintenance of isolating mechanisms are still in question. In the identified case of respiration related genetic incompatibility, both mutations were found in the genome of the incompatible strain YJM421, resulting in a "derived-ancestral" type of interaction. This observation confirmed that the onset of genetic incompatibility does not necessarily require independent fixation of causative mutations in allopatry, as was initially proposed by the Dobzhansky-Müller model[24]. However, while the origin of the onset and maintenance of this particular configuration is unclear, several possibilities could be envisioned. On one hand, as the YJM421 strain showed phenotypic tolerance of the suppressor *SUP7* (Figure 4A), it is possible that this suppressor was acquired in conditions where possessing a suppressor was beneficial. In this scenario, the loss-of-function allele of $cox15^{stop}$ might arise due to random genetic drift, the effect of which was buffered by the preexistence of *SUP7*. Alternatively, the fixation of the $cox15^{stop}$ could arise prior to the apparition of *SUP7*. In fact, the incompatible isolate YJM421 was of clinical origin and it has been shown that deletion of *COX15* could confer higher levels of resistance to antifungal drugs and biofilm formation, two traits that are particularly advantageous for clinical propagation[25]. In this scenario, it is possible that the loss-of-function allele $cox15^{stop}$ originally arose due to selection pressure in clinical conditions. When the strain was replaced in favorable conditions and the original selective pressure was removed, the suppressor could arise to rescue the loss of respiratory capacity due to the adaptation to a new environment. This strain could then become integrated in subsequent genetic and phenotypic assimilation, allowing the particular allelic combination to persist. Even though these scenarios remain only conjectures, it is likely that environmental fluctuation and selection might at least partly contribute to the onset and maintenance of this particular case.

tRNA suppressors in adaptation and the onset of epistasis in yeast

tRNA suppressors perturb the translational fidelity by stop codon read-through, the effect of which resembles the yeast prion [PSI+][26,27]. In *S. cerevisiae*, tRNA suppressors have been frequently selected in numerous genetic screens[28-30]. However, how frequently such suppressors occur in natural isolates was unknown. We surveyed over 100 publically available genomes of *S. cerevisiae* natural isolates and found no tRNA suppressor of any known family in these

genomes except for that of YJM421. The scarcity of tRNA suppressors in natural isolates in contrast with their relatively high frequency of occurrence in the presence of strong selection suggests that tRNA suppressors might represent a transient survival mechanism which could subsequently be lost in the absence of selection[31]. Nevertheless, the evolutionary fate of such suppressors probably depends on the specific genomic and environmental context of the strain in question. In fact, we demonstrated that the presence of the suppressor *SUP7* conferred an increased phenotypic capacity across multiple stress conditions, possibly fueled by the relatively high number of naturally occurring nonsense mutations in natural isolates. Therefore, much like the prion [PSI+][32], tRNA suppressors could offer context dependent selective advantages in yeast. However, as opposed to prions, tRNA suppressors are stably transmitted in a Mendelian manner, which in turn could drive the fixation of allelic combinations leading to the onset of negative epistasis, as is evident in the identified case of genetic incompatibility.

By taking into account environmental factors in the onset of reproductive isolation across a large number of crosses in *S. cerevisiae*, we revealed that context dependent negative epistasis readily segregates in this species and the frequency of which might be more common than previously thought. Nevertheless, cases of the classic two loci Dobzhansky-Müller incompatibility appeared to be rare, with most cases of identified potential negative epistasis apparently reaching a higher complexity, even at an intraspecific scale. The origin of such epistasis was potentially due to the combinatory effect of selection and drift. Further understanding of the onset of intraspecific genetic incompatibilities will extend our perspectives regarding the ongoing phenotypic consequences of genetic diversity within a species, as well as the underlying evolutionary forces that shape the patterns of such variation.

Publications related to this chapter:

Hou, J., Friedrich, A., Gounot, J.S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat Commun* **6**, 7214 (2015).

Hou, J. & Schacherer, J. Negative epistasis: a route to intraspecific reproductive isolation in yeast? *Current Genetics*, 1-5 (2015).

References

1. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* **24**, 1153-9 (2014).
2. Charron, G., Leducq, J.B. & Landry, C.R. Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Mol Ecol* **23**, 4362-72 (2014).
3. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68-72 (2003).
4. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. & Louis, E.J. Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451-4 (2000).
5. Bikard, D. *et al.* Divergent evolution of duplicate genes leads to genetic incompatibilities within *A. thaliana*. *Science* **323**, 623-6 (2009).
6. Seidel, H.S., Rockman, M.V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589-94 (2008).
7. Liti, G. & Schacherer, J. The rise of yeast population genomics. *C R Biol* **334**, 612-9 (2011).
8. Hittinger, C.T. *Saccharomyces* diversity and evolution: a budding model genus. *Trends Genet* **29**, 309-17 (2013).
9. McCusker, J.H., Clemons, K.V., Stevens, D.A. & Davis, R.W. Genetic characterization of pathogenic *Saccharomyces cerevisiae* isolates. *Genetics* **136**, 1261-9 (1994).
10. Merz, S. & Westermann, B. Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in *Saccharomyces cerevisiae*. *Genome Biol* **10**, R95 (2009).
11. Skelly, D.A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-504 (2013).
12. Bergstrom, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* **31**, 872-88 (2014).
13. Hunter, N., Chambers, S.R., Louis, E.J. & Borts, R.H. The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J* **15**, 1726-33 (1996).
14. Zanders, S.E. *et al.* Genome rearrangements and pervasive meiotic drive cause hybrid infertility in fission yeast. *Elife* **3**, e02630 (2014).
15. Chou, J.Y., Hung, Y.S., Lin, K.H., Lee, H.Y. & Leu, J.Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol* **8**, e1000432 (2010).
16. Chou, J.Y. & Leu, J.Y. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays* **32**, 401-11 (2010).
17. Lee, H.Y. *et al.* Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell* **135**, 1065-73 (2008).
18. Li, C., Wang, Z. & Zhang, J. Toward genome-wide identification of Bateson-Dobzhansky-Muller incompatibilities in yeast: a simulation study. *Genome Biol Evol* **5**, 1261-72 (2013).
19. Greig, D. Reproductive isolation in *Saccharomyces*. *Heredity (Edinb)* **102**, 39-44 (2009).
20. Greig, D. A screen for recessive speciation genes expressed in the gametes of F1 hybrid yeast. *PLoS Genet* **3**, e21 (2007).
21. Greig, D., Borts, R.H., Louis, E.J. & Travisano, M. Epistasis and hybrid sterility in *Saccharomyces*. *Proc Biol Sci* **269**, 1167-71 (2002).
22. Demogines, A., Wong, A., Aquadro, C. & Alani, E. Incompatibilities involving yeast mismatch repair genes: a role for genetic modifiers and implications for disease penetrance and variation in genomic mutation rates. *PLoS Genet* **4**, e1000103 (2008).

23. Heck, J.A. *et al.* Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3256-3261 (2006).

24. Coyne, J.A. & Orr, H.A. *Speciation*, xiii, 545, Sinauer Associates, Sunderland (2004).

25. Vandenbosch, D. *et al.* Genomewide screening for genes involved in biofilm formation and miconazole susceptibility in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **13**, 720-30 (2013).

26. Shorter, J. & Lindquist, S. Prions as adaptive conduits of memory and inheritance. *Nat Rev Genet* **6**, 435-50 (2005).

27. Torabi, N. & Kruglyak, L. Genetic basis of hidden phenotypic variation revealed by increased translational readthrough in yeast. *PLoS Genet* **8**, e1002546 (2012).

28. Adams, A.E. & Botstein, D. Dominant suppressors of yeast actin mutations that are reciprocally suppressed. *Genetics* **121**, 675-83 (1989).

29. Novick, P., Osmond, B.C. & Botstein, D. Suppressors of yeast actin mutations. *Genetics* **121**, 659-74 (1989).

30. Ono, B.I., Tanaka, M., Kominami, M., Ishino, Y. & Shinoda, S. Recessive UAA suppressors of the yeast *Saccharomyces cerevisiae*. *Genetics* **102**, 653-64 (1982).

31. Yu, W. & Spreitzer, R.J. Chloroplast heteroplasmicity is stabilized by an amber-suppressor tryptophan tRNA(CUA). *Proc Natl Acad Sci U S A* **89**, 3904-7 (1992).

32. Halfmann, R. *et al.* Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature* **482**, 363-8 (2012).

Chapter 4

The hidden complexity of Mendelian traits across yeast natural populations

Introduction

Using species-wide genetic surveys, we previously showed that independently segregating genetic variants, either large chromosomal changes or point mutations, could have marked phenotypic consequences on offspring viability upon hybridization and contribute to the onset of reproductive isolation within yeast natural populations. However, in addition to such specific phenotype, the genetic origin and complexity underlying the overall phenotypic diversity still require further inquiry. In fact, for any trait, the underlying genetic complexity can be classified as either monogenic or complex. While complex traits are resulted from variation within multiple genes, their interaction and environmental factors[1], some traits are primarily monogenic and conform to a simple Mendelian inheritance[2]. Nevertheless, while useful, this overly simplistic dichotomic view could potentially mask the continuous level of the underlying genetic complexity[3-5]. For instance, traits that appear to have a Mendelian inheritance can be influenced by variation in multiple genes such as modifiers[5-9]. Background effects have been observed in many human disorders[7,8], as well as in loss-of-function mutations in various model systems[10-13] and human[14-16]. However, such specific cases do not reflect the overall genetic diversity and complexity in natural populations[17-20]. More than a century after the rediscovery of Mendel's law, we still lack a global view of the spectrum of genetic complexity of phenotypic variation within any natural population.

Here, we focused on the first species-wide identification of causal variants of Mendelian traits, and characterized in depth their phenotypic effects and transmission patterns across various genetic backgrounds. To do so, we carried out a new survey using large number of crosses, and analyzed the quantitative fitness distribution and phenotype segregation patterns in the offspring for more than 1,100 cross/trait combinations. We found that 8.9% of the cases were Mendelian, among which most were caused by common variants and showed stable inheritance across the *S. cerevisiae* species. Interestingly, global phenotypic distribution patterns of multiple Mendelian traits across an extremely large population (~1,000 isolates) were not necessarily correlated with patterns observed in the offspring from individual crosses. We further characterized a causal variant related to drug resistance and traced its effects across multiple genetic backgrounds. Significant deviations from the Mendelian expectation were observed with variable genetic complexities, illustrating the hidden complexity of a monogenic mutation across a yeast natural population.

Strain	Source	Location	Crossed with YJM326	Divergence to S288c (%)	Reference
BC187	Barrel fermentation	USA		0.37	36
YPS128	Soil beneath *Quercus alba*	Pennsylvania, USA		0.53	36
DBVPG1106	Grapes	Australia		0.35	36
L-1374	Wine	Chile		0.36	36
378604X	Clinical sputum	Newcastle, UK		0.41	36
YJM975	Clinical isolate (Vaginal)	Bergamo, Italy		0.36	36
DBVPG6044	Bili wine	West Africa, Africa		0.60	36
Y55	Wine	France		0.54	36
CLIB192	Bakery	France		0.11	35
CLIB272	Beer	USA	*	0.23	35
CLIB382	Beer	Ireland		0.25	35
YJM145	AIDS patients	USA		0.37	35
YJM280	Peritoneal fluid	USA		0.35	35
YJM320	Blood	California, USA	*	0.32	35
YJM326	Human, clinical	California, USA		0.32	35
YJM421	Ascites fluid	USA	*	0.35	35
YJM434	Human, clinical	Europe	*	0.38	35
YJM440	Human, clinical	NA	*		35
YJM653	Human, clinical	NA	*		35
YJM678	Human, clinical	NA	*		35
CBS7960	Ethanol factory (sugar cane syrup)	São Paulo, Brazil		0.39	35
CECT10109	Prickly pear	Spain	*	0.31	35
DBVPG3591	Cocoa beans	NA	*	0.23	35
DBVPG6861	Poluted stream water	Tijuca forest, Rio de Janeiro, Brazil			35
EM93	Rotting fig	California, USA	*	0.14	35
YPS1000	Exudates *Quercus sp.*	USA		0.41	35
YPS163	Soil beneath *Quercus rubra*	USA		0.36	35
CLIB294	Distillery	France	*	0.25	35
CLIB413	Fermenting rice	China		0.33	35
K12	Sake	Japan		0.25	35
Y10	Coconut	Philippines		0.49	35
Y12	Palm wine	Ivory Coast	*	0.35	35
Y3	Palm wine	Africa	*	0.38	35
Y9	Ragi fermentation	Indonesia	*	0.34	35
YJM269	Red Blauer Portugieser grapes	Austria	*	0.38	35
CLIB154	Wine	Russia	*	0.21	35
I14	Vineyard soil	Italy	*	0.25	35
UC8	Wine	South Africa, Africa	*	0.28	35
WE372	Wine	South Africa, Africa		0.26	35
NC02	Exudates *Quercus sp.*	North Carolina, USA		0.43	35
T7	Exudates *Quercus sp.*	Babler State Park, MO, USA	*	0.49	35

Table 1. Origin and sequence divergence compared to the reference S288c for strains used in this study.

Results

To obtain a comprehensive view of natural genetic variants leading to Mendelian traits in the *S. cerevisiae* species, we selected 41 diverse natural isolates spanning a wide range of ecological (tree exudates, drosophila, fruits, various fermentation and clinical isolates) and geographical sources (Europe, America, Africa and Asia) and performed systematic crosses with one strain Σ1278b (Table 1). For each cross, we generated 40 offspring representing 10 individual meiosis (full tetrads), summing up to a panel of 1,640 full meiotic segregants from diverse parental origins (Figure 1A, panel 1). All segregants as well as the respective parental isolates were tested for 30 stress responsive traits related to various physiological and cellular processes, including different carbon sources, membrane and protein stability, signal transduction, sterol biosynthesis, transcription, translation, as well as osmotic and oxidative stress (Table 2). In total, we tested 1,105 cross/trait combinations and analyzed the offspring fitness distribution patterns for each combination (Figure 1A, panel 2).

Condition	Composition	Stress type
YPD	2% bactopeptone; 1% yeast extract; 2% glucose; 2% agar	Rich medium
YP acetate 2%	2% bactopeptone; 1% yeast extract; 2% acetate; 2% agar	Carbon sources
YP EtOH 2%	2% bactopeptone; 1% yeast extract; 2% ethanol; 2% agar	Carbon sources
YP glycerol 2%	2% bactopeptone; 1% yeast extract; 2% glycerol; 2% agar	Carbon sources
YP sorbitol 2%	2% bactopeptone; 1% yeast extract; 2% sorbitol; 2% agar	Carbon sources
YP galactose 2%	2% bactopeptone; 1% yeast extract; 2% galactose; 2% agar	Carbon sources
YP ribose 2%	2% bactopeptone; 1% yeast extract; 2% ribose; 2% agar	Carbon sources
YP xylose 2%	2% bactopeptone; 1% yeast extract; 2% xylose; 2% agar	Carbon sources
YPD formamide 4%	YPD; formamide 4%	Protein stability
YPD formamide 5%	YPD; formamide 5%	Protein stability
YPD EtOH 15%	YPD; ethanol 15%	Protein stability
YPD benomyl 200µg/ml	YPD; benomyl 200µg/ml	Subcellular organization
YPD benomyl 500µg/ml	YPD; benomyl 500µg/ml	Subcellular organization
YPD SDS 0.2%	YPD; SDS 0.2%	Membrane stability
YPD DMSO 6%	YPD; DMSO 6%	Membrane stability
YPD KCl 2M	YPD; KCl 2M	Osmotic stress
YPD NaCl 1M	YPD; NaCl 1M	Osmotic stress
YPD NaCl 1.5M	YPD; NaCl 1.5M	Osmotic stress
YPD $CuSO_4$ 10mM	YPD; $CuSO_4$ 10mM	Osmotic stress
YPD $CuSO_4$ 15mM	YPD; $CuSO_4$ 15mM	Osmotic stress
YPD LiCl 250mM	YPD; LiCl 250mM	Osmotic stress
YPD CHX 0.5µg/ml	YPD; cycloheximide 0.5µg/ml	Translation
YPD CHX 1µg/ml	YPD; cycloheximide 1µg/ml	Translation
YPD anisomycin 10µg/ml	YPD; anisomycin 10µg/ml	Translation
YPD anisomycin 20µg/ml	YPD; anisomycin 20µg/ml	Translation
YPD anisomycin 50µg/ml	YPD; anisomycin 50µg/ml	Translation
YPD caffeine 40mM	YPD; caffeine 40mM	Signal transduction
YPD caffeine 50mM	YPD; caffeine 50mM	Signal transduction
YPD 6AU 600µg/ml	YPD; 6-azauracile 600µg/ml	Transcription
YPD nystatin 10µg/ml	YPD; nystatin 10µg/ml	Sterol biosynthesis
YPD Mv 20mM	YPD; methylviologen 20mM	Oxydative stress

Table 2. Detailed media composition used in this study

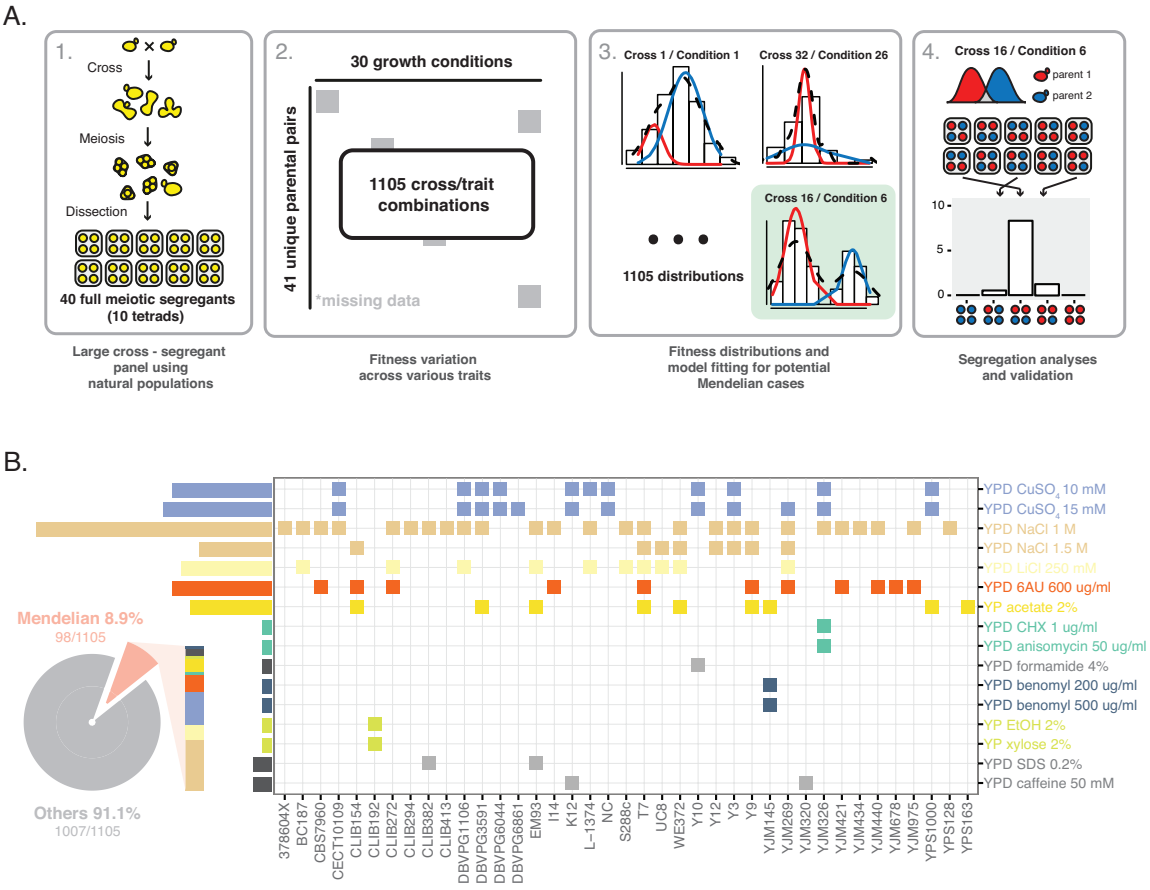


Figure 1. Comprehensive landscape of Mendelian traits in *S. cerevisiae*. A. Workflow of the detection of Mendelian traits. The workflow was defined as 4 steps, consisting with offspring generation, fitness measurements, model fitting and segregation analysis as indicated. **B. Distribution of all identified Mendelian traits spanning different crosses (x-axis) on conditions tested (y-axis).** Each square represents any single Mendelian case and colors indicate different conditions. Pie chart represent the fraction of Mendelian cases relative to the entire dataset.

For a Mendelian trait, contrasting phenotype between the parental isolates was controlled by a single locus, therefore half of the offspring would inherit the causal allele and display a 2:2 segregation in any given tetrads. Consequently, the global offspring fitness distribution would follow a bimodal pattern with equal partitioning of segregants in either parental phenotype cluster. To detect such cases, we first applied a bimodal distribution model with random latent variables for the observed fitness distributions for each cross/trait combination using an Expectation Maximization (EM) algorithm (Figure 1A, panel 3). A case is considered to fit a bimodal distribution when the observed fitness values could be assigned to two non-overlapping clusters (Figure 1A, panel 3). For each fitness distribution observed in a given cross/trait combination, the posterior probability that an individual belongs to either fitness cluster was computed,

and the general features of the fitted bimodal model such as the means and standard deviations for both clusters as well as their relative ratios were extracted. To determine the cutoff values that allow for high confidence calling of bimodal cases and subsequent cluster assignments, we generated a simulated dataset of 1,000 fitness distributions with the same general features compared to the real data, and reapplied the model fitting procedure. Using the simulated data as a training set, we determined that a cutoff of posterior probability > 0.8 for cluster assignment while allowing less than 10% of overlapping between the clusters were the best parameters to maintain a high detection performance (area under the ROC = 0.824) while minimizing case loss. Detailed illustration of data simulation and analyses can be found in the Material & Methods.

By applying these parameters, 318 cross/trait combinations were detected as bimodal, with the parental isolates belonging to distinct clusters. We then analyzed the phenotypic segregation patterns for all bimodal cases (Figure 1A, panel 4). Most bimodal cases showed different patterns of segregation that consists with low genetic complexities. In total, 98 cases were identified as Mendelian, displaying the characteristic 2:2 segregation in the tetrads (Figure 1B). Identified Mendelian cases represented 8.9% (98/1105) across our sample, and were interspersed among various conditions including large number of instances related to NaCl (28 crosses), $CuSO_4$ (13 crosses), 6-azauracil (11 crosses) and acetate (9 crosses) (Figure 1B). Other low frequency cases were found on conditions related to signal transduction (caffeine), carbon sources (ethanol and xylose) various other conditions (formamide, benomyl and SDS) and the antifungal drugs cycloheximide and anisomycin (Figure 1B). In addition, we observed co-segregation of unrelated traits (NaCl, acetate and 6-azauracil; Figure 1B), where the fitness variation patterns in the segregants were highly correlated (Pearson's correlation $\rho > 0.9$). We further characterized cases with co-segregations, high frequency cases related to $CuSO_4$ and the low frequency case related to resistance to the drugs cycloheximide and anisomycin in detail. For the selected cases, 80 additional full tetrads were tested and the 2:2 phenotypic segregation patterns were confirmed.

Molecular characterization of identified Mendelian traits

Using bulk segregant analysis followed by whole genome sequencing as described before, we identified one locus for each case as expected. For all crosses displaying co-segregation with NaCl, the same ~60 kb region (480,000 - 540,000) on chromosome IV was mapped, spanning the *ENA* genes encoding

for sodium and/or lithium efflux pumps (Figure 2A). While variations of the *ENA* genes were known to lead to osmotic stress tolerance[21], the phenotypic associations with other co-segregating traits (acetate and 6-azauracil) were previously unknown. Causal genes related to acetate and 6-azauracil were suspected to be in close genetic proximity with the *ENA* locus, however the precise identities of these genes remained unclear. For cases related to $CuSO_4$, we mapped a 40 kb region on chromosome VIII (190,000 - 230,000; Figure 2C). We identified the *CUP1* gene in this region, which encodes for a copper binding metallothionein (Figure 2C). In this case, the common parental strain Σ1278b was resistant to both concentrations of $CuSO_4$ tested and the allelic version of *CUP1* in Σ1278b led to stable Mendelian inheritance across multiple genetic backgrounds (Figure 1B).

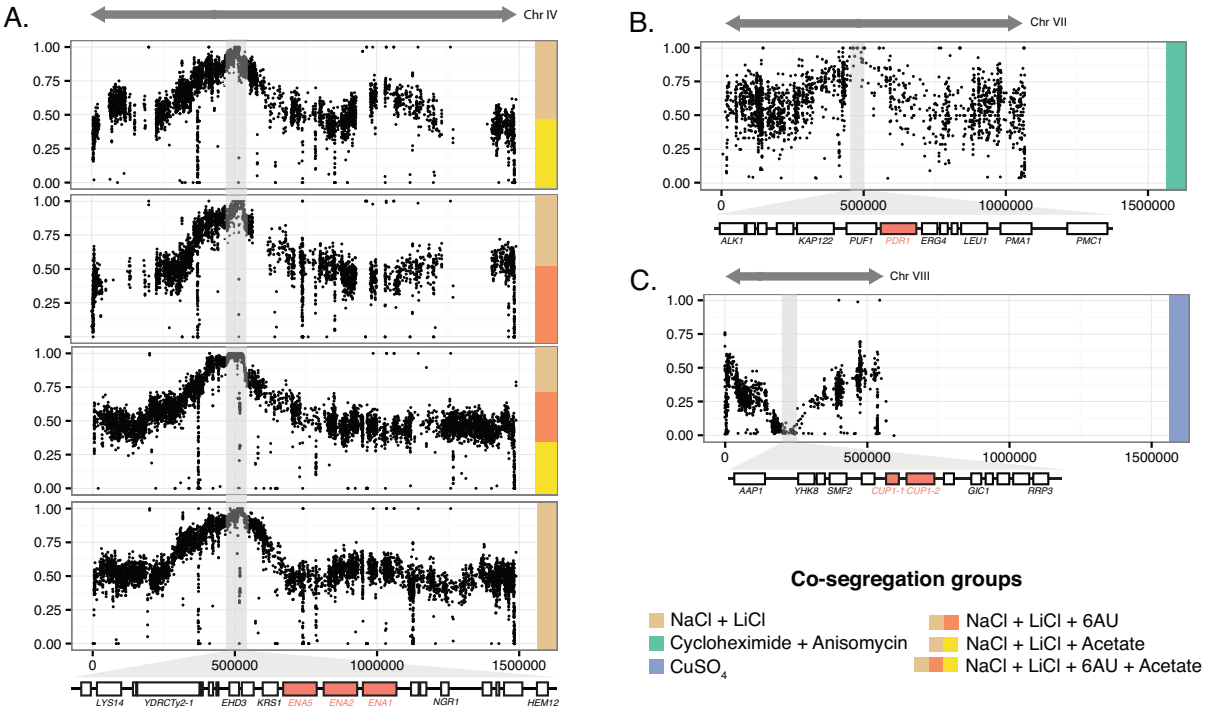


Figure 2. Identification of the genomic regions involved in identified Mendelian traits. A-C. Bulk segregant analysis identified causal genomic regions in traits related to co-segregation with NaCl, cycloheximide and anisomycin and copper sulfate. One chromosomal region with significantly skewed allele frequency was found in each cross, which is presented with color codes. Schematic representations of the chromosome involved are shown, with x-axis corresponding to chromosomal coordinates and y-axis to the allele frequency of the isolates crossed with Σ1278b. Shaded areas correspond to regions with most skewed allele frequencies and genes with these regions are presented to scale.

Finally, the last characterized case involved two anti-fungal drugs cycloheximide and anisomycin, which was found in the cross between a clinical isolate YJM326 and Σ1278b (Figure 1B). Pooled segregants belonging to the higher fitness cluster showed allele frequency enrichment for the YJM326 parent across a ~100 kb region on chromosome VII (420,000 - 520,000; Figure 2B). Further analyses yielded *PDR1* as the potential candidate, which encodes for a transcription factor involved in multidrug resistance. Using reciprocal hemizygosity analysis (Figure 3A) as well as plasmid-based complementation test (Figure 3B), we showed that the *PDR1*^{YJM326} allele was necessary and sufficient for the observed resistance.

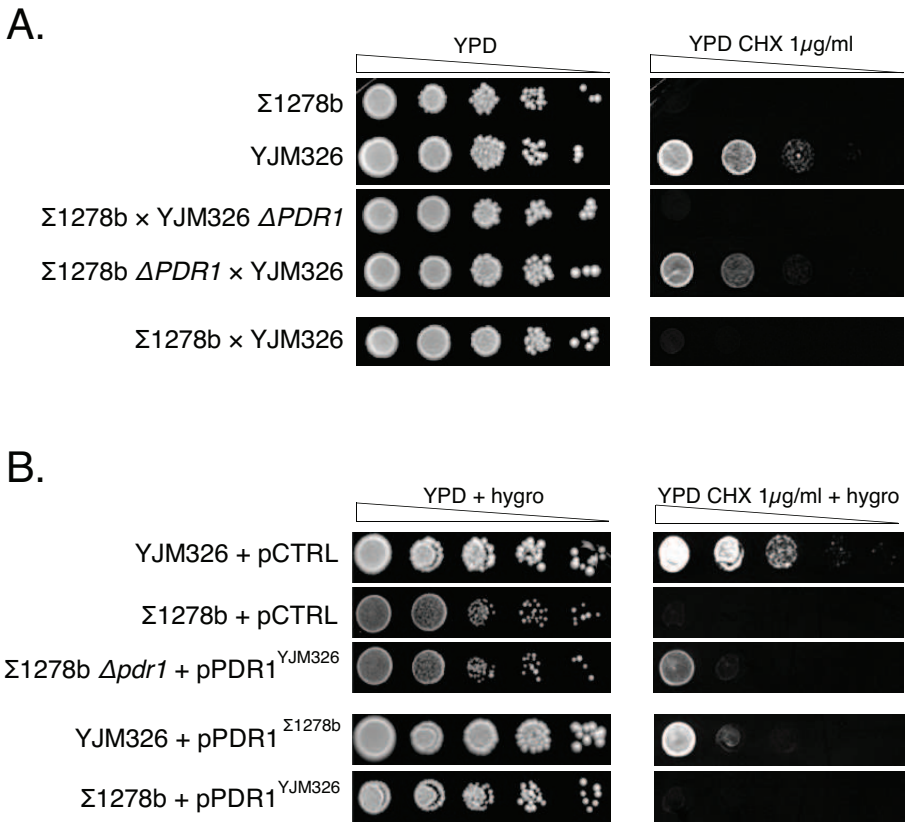


Figure 3. Functional validation of the gene involved in drug resistance. A. Reciprocal hemizygosity test for the candidate gene *PDR1*. Sensitive (Σ1278b) and resistant (YJM326) parental isolates as well as hybrids that are wild type or hemizygous for the *PDR1* gene are spotted in 5 dilutions onto YPD (left panel) and YPD CHX 1 μg/ml (right panel). Cells were grown for 48 hours at 30°C. **B. Ectopic expression of *PDR1*^{YJM326} confers drug resistance in the sensitive strain Σ1278b only with deletion of *PDR1*.** Growth of strains carrying empty control plasmid (pCTRL) or plasmids with the resistant (pPDR1^{YJM326}) or sensitive (pPDR1^{Σ1278b}) allele was tested in the absence (left panel) or presence (right panel) of cycloheximide. All media were supplemented with 200 μg/ml of hygromycin to maintain plasmid stability.

Fitness distribution of identified Mendelian traits across large natural populations

Although Mendelian traits could exhibit distinctive offspring distribution and segregation patterns in individual crosses, the general phenotypic distribution of such traits within a population was unclear. We measured the fitness distribution of an extremely large collection of ~1,000 natural isolates of *S. cerevisiae* (the 1002 yeast genomes project, http://1002genomes.u-strasbg.fr/) on selected conditions related to identify Mendelian traits, including resistance to NaCl, LiCl, acetate, 6-azauracil, $CuSO_4$ and cycloheximide (Figure 4).

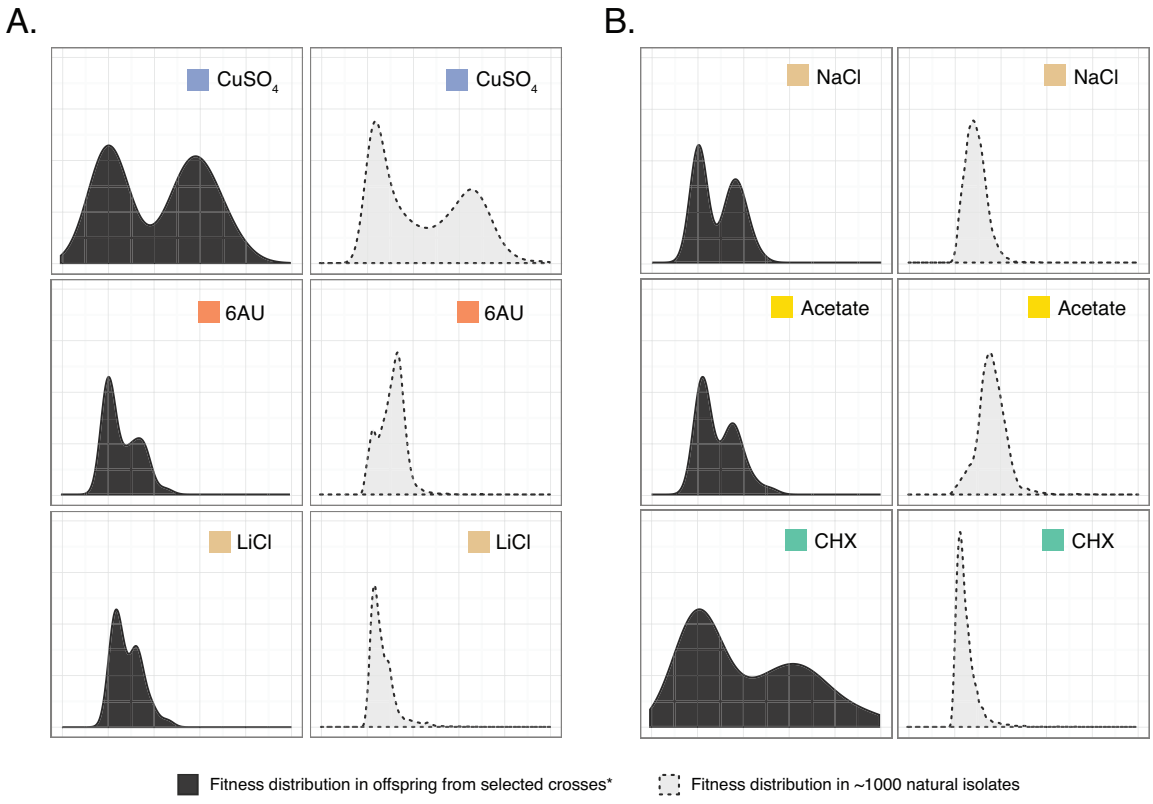


Figure 4. Fitness distribution patterns of identified Mendelian traits within large natural population. A. Bimodal distribution patterns both in crosses and at the population level. Comparisons of the fitness distribution on 6 selected conditions in individual crosses (left panel, N=40) and across ~1000 natural isolates of *S. cerevisiae* (right panel, N=960) are shown. Conditions tested are color-coded. **B. Bimodal distributions observed only in crosses but not within a population.**

Interestingly, while some traits followed the same bimodal distribution model across the population as was observed in offspring from single crosses (Figure 4A), other traits with clear Mendelian inheritance pattern in crosses appeared to vary continuously at the population level (Figure 4B). This observation

suggested that the phenotypic distribution within the population might not necessarily reflect the underlying genetic complexity of traits. Instead, the inheritance pattern for any given trait might largely be determined by specific combinations of parental genetic backgrounds.

Hidden complexity of a rare Mendelian variant across different genetic backgrounds

While focusing on highly frequent cases such as $CuSO_4$ and NaCl provided indications about the transmission stability of common Mendelian variants and revealed previously unknown co-segregations, we were particularly interested in rare cases where the phenotypic effects and the general inheritance patterns across different genetic backgrounds were unknown. The identified Mendelian case related to the anti-fungal drugs cycloheximide and anisomycin could be considered as such. Across our panel, the parent YJM326 was the only highly fit isolate, and few isolates showed similar resistance level within the whole species (Figure 4B).

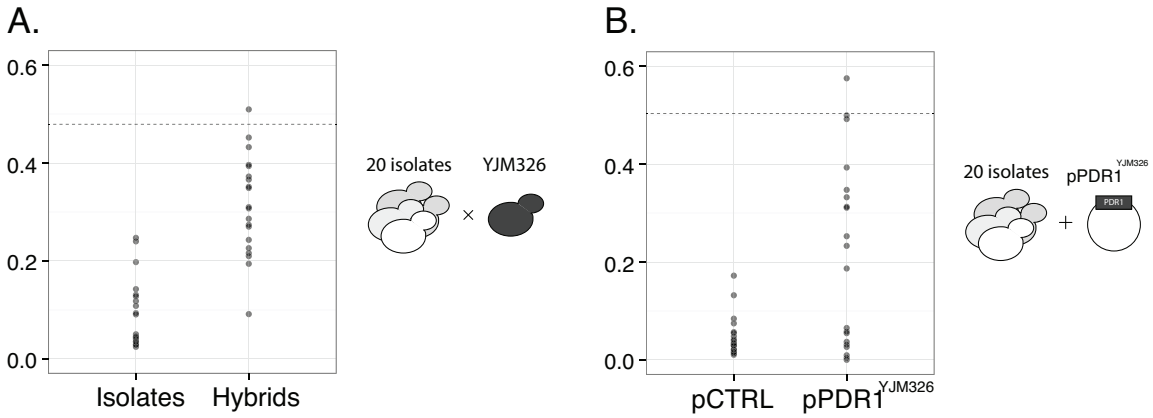


Figure 5. Effects of the *PDR1*^{YJM326} allele in different genetic backgrounds. A. Fitness variation of 20 isolates (left panel) in comparison with the same set of strains hybridized with YJM326 in the presence of drug. Fitness values (y-axis) correspond to the ratio between the growth in the presence of cycloheximide (YPD CHX 1µg/ml) and control media YPD. Dashed line indicates the fitness of the resistant strain YJM326. **B. Fitness variation of 20 isolates carrying empty control plasmid (pCTRL, left panel) or plasmid containing the *PDR1*^{YJM326} allele under its native promoter (pPDR1^{YJM326}, right panel).** Fitness values were measured in the presence of cycloheximide (YPD CHX 1µg/ml) with hygromycin to maintain plasmid stability. Dashed line indicates the fitness value of YJM326 carrying the plasmid pPDR1^{YJM326}.

To test the effect of the *PDR1*^{YJM326} allele in different backgrounds, we crossed the resistant isolate YJM326 with 20 diverse sensitive isolates (Table 1).

Counterintuitively, the resulting hybrids displayed continuous variation of the resistance in the presence of cycloheximide (Figure 5A). To test whether the resistance variation in the hybrids were due to allelic interactions at the *PDR1* locus in different backgrounds, we introduced a plasmid carrying the *PDR1*^{YJM326} allele (pPDR1^{YJM326}) into the same set of isolates, and quantified their fitness in the presence of cycloheximide (Figure 5B). Across all isolates tested, about half (11/20) expressed the resistant phenotype to various degrees (Figure 6). However, fitness between haploid isolates carrying pPDR1^{YJM326} and the corresponding hybrids were only weakly correlated (Pearson's correlation ρ = 0.434), indicating that allelic interactions at the *PDR1* locus only partly accounted for the observed variation (Figure 6).

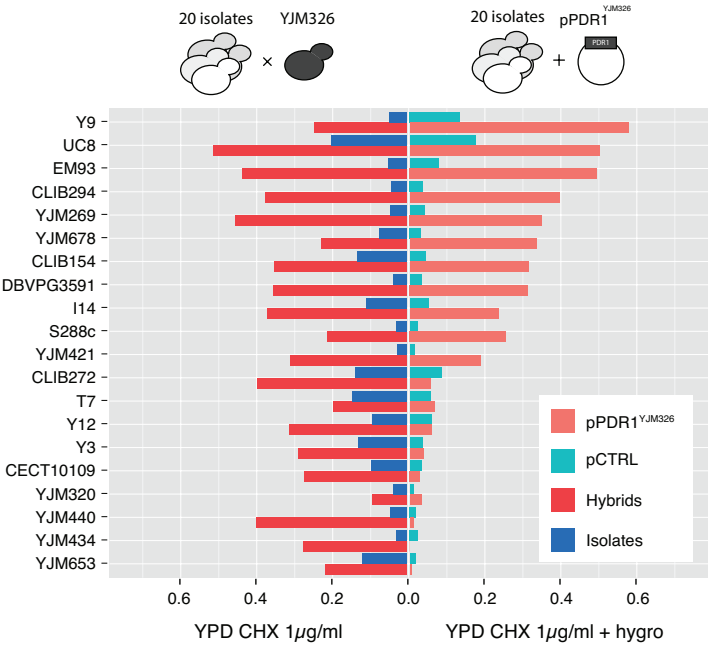


Figure 6. Fitness in strains with ectopic expression of *PDR1*^{YJM326} and in hybrid contexts. The fitness values for 20 isolates in the presence of cycloheximide were compared after crossed with the resistant isolate YJM326 (left panel) or after transformed with plasmid carrying the resistant allele (right panel). Strains are indicated on y-axis, with color codes of different configurations (hybrid or plasmid).

The lack of correlation between hybrids and isolates carrying the plasmid with the *PDR1*^{YJM326} allele led us hypothesize the presence of potential modifiers in various hybrid backgrounds. To test this hypothesis, we evaluated the fitness distributions of the drug resistance in the offspring across the 20 hybrids generated previously. For each hybrid, 20 complete tetrads were tested in the presence of cycloheximide and the fitness distributions as well as the segregation patterns were assessed in the offspring. In the absence of modifiers, haploid segregants are expected to have complete phenotypic penetrance, as the effects

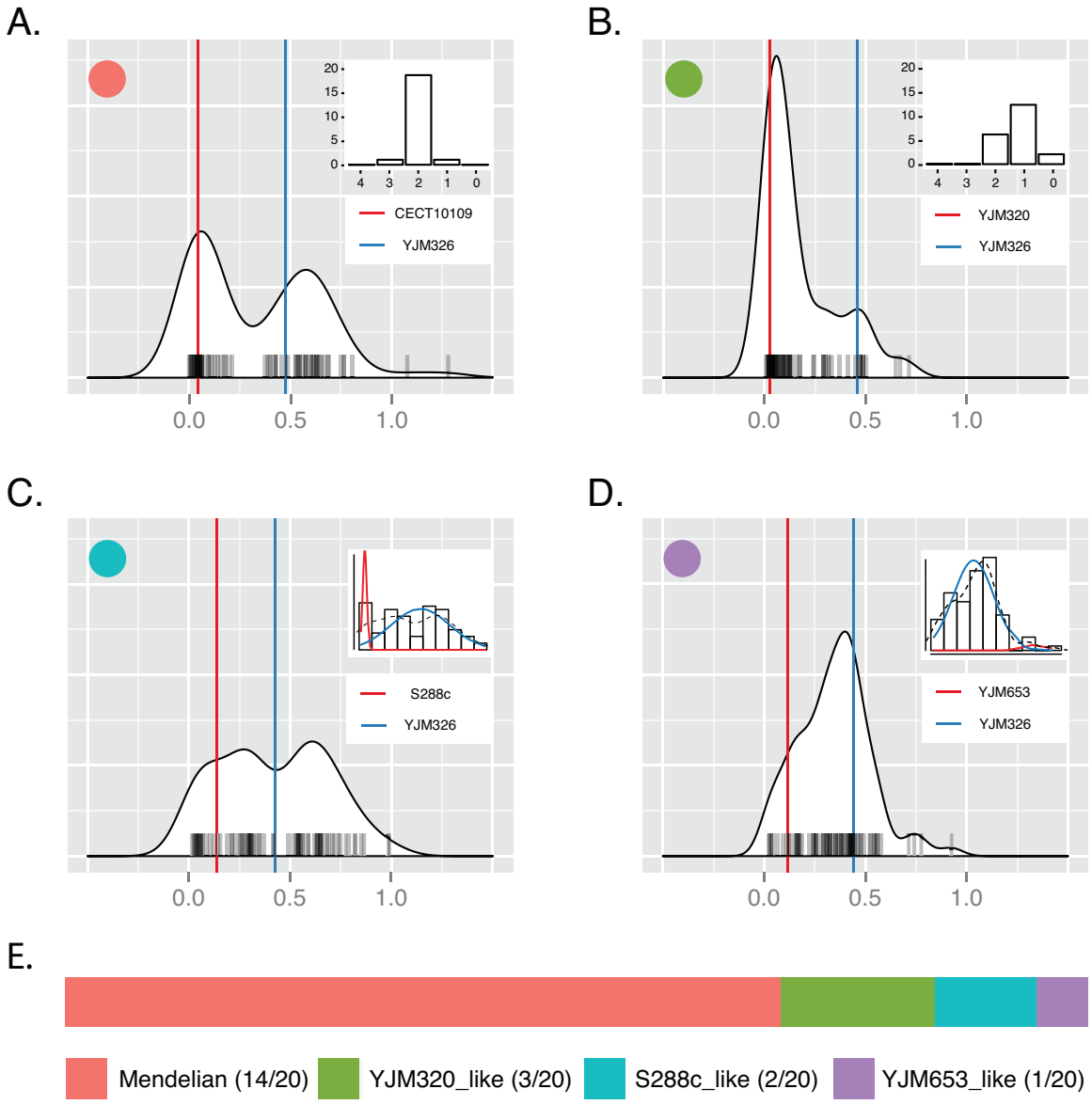


Figure 7. Post-Mendelian inheritance patterns of drug resistance in different hybrid contexts. A-D. Offspring fitness distribution patterns observed in hybrids originated from 20 sensitive isolates and YJM326 in the presence of cycloheximide (YPD CHX 1µg/ml). 80 offspring were tested for each case, and examples of Mendelian (A) and non-Mendedian (C-D) inheritance patterns are shown. Phenotypic segregation is indicated at the upper right side. For non-bimodal cases the model fitting results were shown instead. Parental origins for each cross are shown, and the fitness values of the sensitive (red) or resistant (blue) parental strains are presented as vertical bars. **E. Distribution of different types of inheritance patterns observed.**

of intralocus interaction were eliminated. In this scenario, all crosses between any sensitive parental isolate and YJM326 should display a bimodal distribution in the offspring, with a 2:2 segregation of the phenotype.

Interestingly, while most of the tested crosses (14/20) displayed Mendelian segregation as was observed in the cross between YJM326 and Σ1278b, several crosses showed clear deviation of the expected phenotypic distribution (Figure 7). In addition to Mendelian cases (Figure 7A), 3 other types of distribution were observed (Figure 7B-D). In total, such cases represent ~30% of all crosses tested (Figure 7E). Of these crosses, 15% (3/20, between YJM320, Y3, Y9 and YJM326) showed incomplete penetrance, indicating possible suppressors of the $PDR1^{YJM326}$ allele (Figure 7B). We observed a 1:4:1 ratio between tetrads containing 2, 1 and 0 resistant segregants, possibly indicating that two independent loci, including *PDR1*, were involved (Figure 7B). 10% of the crosses (2/20, between S288c, YJM440 and YJM326) showed enriched high fitness offspring, with an intermediate peak between the sensitive and resistant clusters. This observation suggests the presence of epistatic interactions from these specific genetic backgrounds, resulting as a transitional resistant phenotype cluster with higher genetic complexity (Figure 7C). The levels of genetic complexity in these cases are suspected to be low, but the precise number of the genes involved remained unclear.

In addition to cases with low level of deviations from Mendelian expectations, we also found one cross (between YJM653 and YJM326) with a clear normal fitness distribution in the offspring. In this case, the resistant phenotype was no longer caused by a single Mendelian factor, and the underlying genetic determinants were undoubtedly complex (Figure 7D). Contrasting to other identified Mendelian traits with a stable inheritance patterns across the population, the *PDR1* case represented a perfect example illustrating the hidden complexity of a simple Mendelian trait within natural population of the yeast *S. cerevisiae*.

Discussion

By performing a species-wide survey of monogenic variants in *S. cerevisiae*, we obtained a first estimation of the proportion of Mendelian traits within a natural population. We showed that genes and alleles underlying the onset of Mendelian traits are variable in terms of their type, frequency and genomic distribution at the population level. Remarkably, by tracing the effect of one causal Mendelian variant *PDR1*^{17M326} across the population, we demonstrated that the genetic complexity of traits could be dynamic, transitioning from clear Mendelian to diverse complex inheritance patterns depending on various genetic backgrounds.

Biased genomic distribution of causal Mendelian variants in yeast

Yeasts and more particularly *S. cerevisiae* have been extensively used as a model for dissecting many complex traits that were of medical, industrial and evolutionary interests [22-26]. A trend emerging from studying complex traits in this species was that causal variants do not distribute randomly across the genome, and several hotspots have been identified [27]. As a result, a low number of loci were found to be involved in high numbers of unrelated phenotypes, despite the fact that underlying causal genes could be different. Interestingly, causal variants in Mendelian traits seemed to follow the same trend as supported by our data. In fact, we observed phenotypic co-segregation of unrelated conditions such as resistance to acetate, 6-azauracil and osmotic stress, and showed that only a single region on chromosome IV was involved (Figure 2A). In addition, the observed co-segregations showed relatively high population frequencies, with more than 15% of the crosses co-segregating on at least two different conditions (Figure 1B). This effect of linkage could possibly lead to biased phenotype assortments across the population, although the underlying evolutionary origin is unknown.

Stability of Mendelian inheritance and the functional nature of causal variant

In general, Mendelian traits were considered as rare especially in human disorders, however, no directly estimation of the proportion of Mendelian relative to complex traits was available at the population level, and what types of genes were more susceptible to cause Mendelian inheritance were unknown. Our data showed that across a yeast natural population, causal alleles involved in direct response to stress, such as transporters (*ENA*) or metal-binding genes

(*CUP1*) were more likely to follow Mendelian inheritance. In fact, a large number of Mendelian traits identified in our sample were related to these two loci, and the inheritance patterns were extremely stable, displaying 2:2 segregations with little influence of the genetic backgrounds. Similar pattern was found in a Mendelian trait related ammonium resistance in natural isolates of *S. cerevisiae*, where a transporter gene *TRK1* was involved [28]. The stable inheritance patterns of traits caused by alleles with direct phenotypic effect could potentially due to the lack of regulatory complexity. As was supported by laboratory evolution experiments, amplifications of this type of genes were frequent, conferring to rapid acquisition of resistances in stress conditions such as salt[29], copper[30,31], sulfate[32] and glucose limitations[33].

From Mendelian to complex: a continuum

By contrast, depending on the gene involved, a given Mendelian trait could lead to complex inheritance patterns across different genetic backgrounds, as evidence by the causal allele *PDR1* related to resistance to cycloheximide and anisomycin. By crossing the strain YJM326 carrying the resistant allele $PDR1^{YJM326}$ with diverse natural isolates, we showed that although most crosses retained stable 2:2 segregations, the inheritance pattern of the resistance phenotype in some cases displayed various deviations from Mendelian expectation, including reduced penetrance (3/20), increased genetic complexity (2/20) and in one extreme case, transition from monogenic to complex trait. We propose that the observed post-Mendelian inheritance patterns are due to the functional nature of the *PDR1* gene. In fact, as *PDR1* encodes for a transcriptional factor with complex regulatory networks and impact multiple downstream effector genes[34], the resulting phenotypic expression would possibly be influenced by variations of a large number of genes that are involved in the same network in different genetic backgrounds.

Overall, our data provided a first comprehensive view of natural genetic variants that lead to the onset of Mendelian traits in a yeast population. We showed that monogenic mutations could exhibit post-Mendelian modifications such as pleiotropy, incomplete dominance as well as variations in expressivity and penetrance due to differences in specific genetic backgrounds. Depending on the parental combination, the inheritance might display a Mendelian, intermediate or complex pattern, showing the continuum of the complexity spectrum related to a monogenic mutation, as illustrated by the example of the drug resistance involving $PDR1^{YJM}$. However, while Mendelian traits could be

related to common or rare variants, we found that the overall fitness distribution patterns of such traits at the population level, for some instances if not all, were not informative regarding their genetic complexity. Collectively, phenotypic prediction even for simple Mendelian variants may not be an easy task, in part due to the lack of prediction power using population data and the scarcity of large-scale family transmission information, such as the case for diseases in human. Future studies using pairwise crosses covering a larger panel of conditions in yeasts, or in other model organisms, may provide general trends and a more complete picture regarding the phenotypic predictability of monogenic traits.

Publication related to this chapter:

Hou, J., Sigwalt, A., Pflieger, D., Peter, J., de Montigny, J., Dunham, M. & Schacherer, J. The hidden complexity of Mendelian traits across yeast natural populations. *BioRxiv* (2016)

References

1. Mackay, T.F., Stone, E.A. & Ayroles, J.F. The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* **10**, 565-77 (2009).
2. Antonarakis, S.E. & Beckmann, J.S. Mendelian disorders deserve more attention. *Nat Rev Genet* **7**, 277-82 (2006).
3. Dipple, K.M. & McCabe, E.R. Modifier genes convert "simple" Mendelian disorders to complex traits. *Mol Genet Metab* **71**, 43-50 (2000).
4. Antonarakis, S.E., Chakravarti, A., Cohen, J.C. & Hardy, J. Mendelian disorders and multifactorial traits: the big divide or one for all? *Nat Rev Genet* **11**, 380-4 (2010).
5. Badano, J.L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* **3**, 779-89 (2002).
6. Nadeau, J.H. Modifier genes in mice and humans. *Nat Rev Genet* **2**, 165-74 (2001).
7. Cooper, D.N., Krawczak, M., Polychronakos, C., Tyler-Smith, C. & Kehrer-Sawatzki, H. Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum Genet* **132**, 1077-130 (2013).
8. Dorfman, R. Modifier gene studies to identify new therapeutic targets in cystic fibrosis. *Curr Pharm Des* **18**, 674-82 (2012).
9. Thein, S.L. Genetic modifiers of sickle cell disease. *Hemoglobin* **35**, 589-606 (2011).
10. Dowell, R.D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
11. Hamilton, B.A. & Yu, B.D. Modifier genes and the plasticity of genetic networks in mice. *PLoS Genet* **8**, e1002644 (2012).
12. Paaby, A.B. *et al.* Wild worm embryogenesis harbors ubiquitous polygenic modifier variation. *Elife* **4**(2015).
13. Vu, V. *et al.* Natural variation in gene expression modulates the severity of mutant phenotypes. *Cell* **162**, 391-402 (2015).
14. Blomen, V.A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092-6 (2015).
15. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096-101 (2015).
16. Hart, T. *et al.* High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515-26 (2015).
17. Cao, J. *et al.* Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat Genet* **43**, 956-63 (2011).
18. Strope, P.K. *et al.* The 100-genomes strains, an *S. cerevisiae* resource that illuminates its natural phenotypic and genotypic variation and emergence as an opportunistic pathogen. *Genome Res* **25**, 762-74 (2015).
19. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74 (2015).
20. Hou, J., Friedrich, A., Gounot, J.S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat Commun* **6**, 7214 (2015).
21. Ruiz, A. & Arino, J. Function and regulation of the *Saccharomyces cerevisiae* ENA sodium ATPase system. *Eukaryot Cell* **6**, 2175-83 (2007).
22. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326-30 (2002).
23. Treusch, S., Albert, F.W., Bloom, J.S., Kotenko, I.E. & Kruglyak, L. Genetic mapping of MAPK-mediated complex traits across *S. cerevisiae*. *PLoS Genet* **11**, e1004913 (2015).

24. Bloom, J.S., Ehrenreich, I.M., Loo, W.T., Lite, T.L. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234-7 (2013).

25. Ehrenreich, I.M. *et al.* Genetic architecture of highly complex chemical resistance traits across four yeast strains. *PLoS Genet* **8**, e1002570 (2012).

26. Mukherjee, V. *et al.* Phenotypic evaluation of natural and industrial *Saccharomyces* yeasts for different traits desirable in industrial bioethanol production. *Appl Microbiol Biotechnol* **98**, 9483-98 (2014).

27. Fay, J.C. The molecular basis of phenotypic variation in yeast. *Curr Opin Genet Dev* **23**, 672-7 (2013).

28. Reisser, C. *et al.* Genetic basis of ammonium toxicity resistance in a sake strain of yeast: A Mendelian Case. *G3 (Bethesda)* (2013).

29. Anderson, J.B. *et al.* Determinants of divergent adaptation and Dobzhansky-Muller interaction in experimental yeast populations. *Curr Biol* **20**, 1383-8 (2010).

30. Fogel, S. & Welch, J.W. Tandem gene amplification mediates copper resistance in yeast. *Proc Natl Acad Sci U S A* **79**, 5342-6 (1982).

31. Gerstein, A.C. *et al.* Too much of a good thing: the unique and repeated paths toward copper adaptation. *Genetics* **199**, 555-71 (2015).

32. Gresham, D. *et al.* The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* **4**, e1000303 (2008).

33. Dunham, M.J. *et al.* Characteristic genome rearrangements in experimental evolution of Saccharomyces cerevisiae. *Proc Natl Acad Sci U S A* **99**, 16144-9 (2002).

34. Moye-Rowley, W.S. Transcriptional control of multidrug resistance in the yeast *Saccharomyces*. *Prog Nucleic Acid Res Mol Biol* **73**, 251-79 (2003).

35. Schacherer, J., Shapiro, J.A., Ruderfer, D.M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342-5 (2009).

36. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-41 (2009).

Material & Methods

The lab is where all the fun begins

Wet lab procedures and notes

Strains

A collection of 68 strains isolated from diverse ecological (tree exudate, wine, different fermentations and clinical) and geographical (Europe, Asia, Africa and America) origins were used in this study, including 60 isolates from the collection of Schacherer et al. 2009 and 8 from Liti et al. 2009. All strains are stable haploids with deletion of the *HO* gene[1,2]. Laboratory strains FY4, FY5 (isogenic to S288c) and Σ1278b were used. Deletion mutants in the Σ1278b background were obtained from the gene deletion collection kindly provided by Dr. Charles Boone[3]. A uracil auxotrophic mutant of YJM421 (YJM421 *ura3Δ0*) was generated by deleting the *URA3* gene using 5-FOA selection. YJM326 *Δpdr1* strain was generated by insertion of hygromycin resistance cassette *HygMX* using homologous recombination.

Media and culture conditions

Detailed media compositions for phenotyping of the segregant panel are listed within each chapter. Growth and maintenance of the strains are carried on standard rich media YPD (1% yeast extract, 2% peptone and 2% glucose). A final concentration of 200 µg/ml hygromycin (Euromedex) is supplemented to maintain the plasmids carrying a resistance marker gene *HygMX*. The 5-FOA selection plate is made by supplementing 25 µg/ml uracil and 1 mg/ml 5-FOA in a synthetic complete uracil dropout media (SC-URA). Sporulation is induced on potassium acetate plates (1% potassium acetate, 2% agar). All procedures are performed at 30°C unless otherwise indicated.

Crosses and generation of meiotic offspring

Crosses are carried out on YPD plates by mixing freshly grown cells with opposite mating types. Resulting diploids are put on sporulation medium for 2-3 days (1% potassium acetate, 2% agar). Tetrad dissections are performed using the MSM 400 dissection microscope (Singer instrument) on YPD agar after digestion of the tetrad asci with zymolyase (MP Biomedicals MT ImmunO 20T). For most *S. cerevisiae* crosses, a final concentration of 0.1 mg/ml of zymolyase with 15 min incubation is efficient, however adjustments of the concentration and incubation time can be made in case of difficult digestions.

The stock solution of zymolyase should be kept at -20°C. Fresh stock should be made when digestions become significantly inefficient. After digestion, tetrads are individually picked, broken to spores, aligned on YPD cultured for 48 hours. Viable spores will form colonies and the spore viability corresponds to the ratio between the number of viable spores and the total number of spores dissected.

Gene deletion

Gene deletions are performed by insertion of selective markers, such as drug markers *KanMX, HphMX* and *ClonNAT,* or auxotrophic markers such as *URA3*. Long PCR primers flanked by sequences homologous to the target gene are used to amplify the desired marker. In general, 50 bp length of homology for each flanked end suffices (Figure 1A). Alternatively, regions of homology can be increased using fusion PCR, where individually amplified fragments could be assembled with 18 to 20 bp overlaps (Figure 1B). In the case of inefficient fusion reaction, an additional step can be carried out to increase the size of the overlapping regions (Figure 1C).

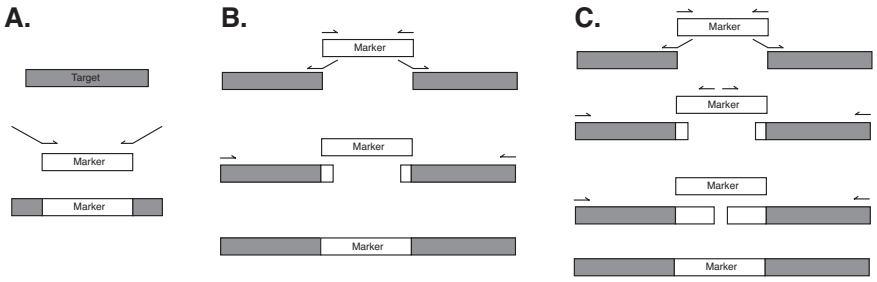


Figure 1. Generation of PCR fragments for gene deletion and allele replacement

Allele replacement

For allele replacements, the generation and assembly of desired fragments follow the same principles. However, the most "clean" way for an allele replacement is the use of the *URA3* marker, which allows for counter-selection using 5-FOA (1g/L) and do not introduce additional marker gene at the locus where the allele replacement is targeted. Several considerations should be taken into account. First, the concentration of uracil in 5-FOA media should be lower than a standard SC media (50 µg/ml), usually around 20 to 25 µg/ml. In fact, 5-FOA selection is extremely sensitive with lots of false positives due to spontaneous mutations, and higher concentration of uracil further decrease the

efficiency. Secondly, to avoid preexisting mutations that confer to 5-FOA resistance prior to the replacement, all precultures should be done using SC-Uracil media instead of YPD. Lastly, instead of directly plating the cells on 5-FOA media after the transformation, cells should be regenerated on solid YPD media O/N. To select transformants, the YPD plate with cells that were grown overnight should be replicated on velvet, and the quantity of cells should be reduced by repeated replication on empty agar plates (7~8 times) before replicate on 5-FOA.

Plasmids constructions

For plasmid constructions in this thesis, we used Gateway cloning technology (Invitrogen). Fragments of interests were amplified from corresponding genomic DNA with attB1/attB2 recombination sites flanked at extremities, and cloned into an empty centromeric plasmid with HphMX resistance marker pCTRL[4]. The resulting plasmids were verified using restriction enzymes and PCR amplification with internal primers.

Transformations

Transformations were performed using EZ transformation kit (MP biomedicals). In general, 2 to 3 µg of fragment DNA are used for gene deletion and allele replacements, and 100 ng are used for plasmid transformation.

Growth measurements using microculture

To efficiently measure the growth rate of strains in parallel, we used microcultures. Strains are pregrown in YPD overnight, and then transferred into 150 µl of the desired media in flat bottom 96 well plates (Nuclon, ThermoFischer) using a long-pin replicator. Each isolate should have at least 2 replicates in the same plate and 2 replicates in another plate to ensure reproducibility. The corner of the 96-well plates should be avoided for insemination of samples due to excessive evaporation at these positions. In general, cultures are followed for 48 hours and the absorbance of each well is read at 595 nm in 10 min intervals (4 spatial positions and 3 flashes) using a TECAN plate reader (Infinite series) with horizontal and orbital shaking. Growth curves can be retrieved using the program GATHODE[5] and the growth rate are calculated using an exponential curve fit. Different media, with drugs to maintain plasmid selections for example, can be used accordingly.

Manual phenotyping on solid media for discrete characters

Crosses with high offspring viability (>90%) on YPD can be tested on different culture conditions to score for discrete phenotypes such as lethality. This procedure was applied for offspring viability estimates in Chapter 3, where 27 crosses were tested on 20 conditions. For each cross, full tetrads (containing 4 viable spores) were suspended in liquid YPD and pinned onto corresponding condition plates as well as onto an YPD control plate using a frogger. Growth was scored by eye after 48 hours where viable segregants formed a patch. The offspring viability for each cross and condition corresponds to the ratio between the number of viable segregants and the total number of segregants viable on YPD.

Quantitative phenotyping using solid media

A high throughput phenotyping procedure was developed to measure fitness variation on solid media. Strains are pregrown in liquid YPD medium and pinned onto a solid YPD matrix plate to a 384 density format using a replicating robot RoTor (Singer instruments). The matrix plates are incubated overnight to allow sufficient growth, which are then replicated on different media conditions including YPD as a pinning control. For each sample, replicates should be present at different matrix positions. After pinning, the plates are incubated for 48 hours at 30°C and scanned at the 24, 40, 48 hour time points with a resolution of 600 dpi at 16-bit grayscale, then analyzed using the R package Gitter[6]. Colony sizes of each strain is measured as the number of pixels present at the corresponding pinning position, and the normalized growth ratios are calculated by normalizing the colony size by the one at the same position on YPD.

Bulk segregant analysis strategy

The principle of bulk segregant analysis is to select a population of segregants with biased genotypes based on their phenotypic segregation pattern, then sequence the DNA of the bulk all together to look for skewed allele frequencies at the causal loci. Different selection criteria can be applied for different mapping scenarios. For reproductive isolation cases with 75% offspring viability on rich media YPD, a subset of viable spores were selected based on lethal phenotype segregation patterns. In this scenario, the segregation of the lethal phenotype resulted in predominantly 3 types of tetrads: tetrads with 4 viable

spores or parental ditypes (PD), 3 viable spores or tetratypes (TT) and 2 viable spores or non-parental ditypes (NPD). As lethal combinations were not present in NPD tetrads, 50 independent spores from this type of tetrad were separately cultured then pooled by equal O.D. readings at 600 nm. For condition specific incompatibility and Mendelian cases, causal genes or genic combinations are presumably united in non-viable segregants, which were then selected and pooled for mapping. Genomic regions involved were subsequently mapped by analyzing the allele frequency variation along the genome for each case.

Successive backcrossing strategy

For reproductive isolation cases with 50% spore viability on YPD, only segregants which inherited either parental genotypes were viable, resulting in a segregation of predominantly 3 types of tetrads: parental ditypes (PD) with four viable spores, tetratypes (TT) with 2 viable spores, and non-parental ditypes (NPD) with 0 viable spores. To map the genomic regions involved, we used a successive backcrossing strategy. For each cross, one F1 parental ditype tetrad (PD, 4 viable spores) was selected, and all four spores were backcrossed to one of the parental strain. Spore viabilities were analyzed, and a segregant, which has retained the 50% spore viability segregation was selected for a subsequent backcross. Five generations of backcrosses were performed and one 5^{th} generation-backcrossed segregant (B5) was obtained and sequenced. In this scenario, causal regions can be identified as the majority of the genome was enriched for the parental alleles except for regions involved in low spore viability.

Sequencing, data processing and other computational analyses

DNA extraction, sequencing, and SNP calling

Genomic DNA can be extracted using the Qiagen Genomic-tip kit, which yield high quality samples for whole genome sequencing. We use a short reads strategy Illumina Hiseq for sequencing. The advantage of this kind of strategy is the high coverage, which allows for detection of allele frequency variation in bulk segregant data. We use paired-end libraries with average size of 280 to 500 bp, 101 bp/read, with 50X to 200X coverage depending on the cases. Quality controlled reads are aligned to a reference genome using BWA[7] with "-n 5 -o 2" options, which allow maximum of 5 mismatches and 2 gaps per read. SNP calling was done using SAMtools[7] or GATK v3.3-0[8], with default parameters. The allele frequency of the reference strain is scored at each polymorphic position. Coverage along the genome can be calculated by averaging the number of reads aligned at each genomic position within a 2 kb window, or simply be calculated at the number of reads aligned at each polymorphic position.

Neighbor joining tree

In chapter 2, a majority-rule consensus tree of the surveyed strains was built based on the 101,343 segregating sites identified by Schacherer et al. 2009. For strains that were not represented in the original tree[1], the publicly available sequences[9] were recovered and aligned against the S288c reference sequence with BWA (-bwasw option), except for the CECT10266 strain, for which we computed our own reads mapping (see *DNA extraction, sequencing, and SNP calling* section). Polymorphic positions were called with SAMtools and used to complete the segregating sites matrix. We constructed a neighbour-joining tree of the strains studied from these SNP data using the software package Splitstree[10], with branch lengths proportional to the number of segregating sites that differentiate each node.

Annotation of nonsense mutations and functional enrichment

In chapter 3, a total of 100 recently sequenced isolates of *S. cerevisiae* (NCBI BioProject PRJNA189847 to PRJNA189936, PRJNA189300, PRJNA188959, PRJEB2299 and[11,12]) were used for nonsense mutation detection. The reads were retrieved and cleaned using cutAdapt, and aligned to the reference

genome S288c. Read alignments and SNP calling were performed as before and SNPs were annotated using the EMBL annotation using a customized Python script. GO term enrichment was performed using FunSpec[13].

Genome assemblies and detection of tRNA suppressor

For the same set of 100 strains, cleaned reads were assembled using SOAPdenovo2[14], version 2.04, with a k-mer size of 75. The assemblies were then surveyed for potential tRNA suppressors using tRNAscan-SE (version 1.3.1) and no suppressor of any codon families were found in this dataset.

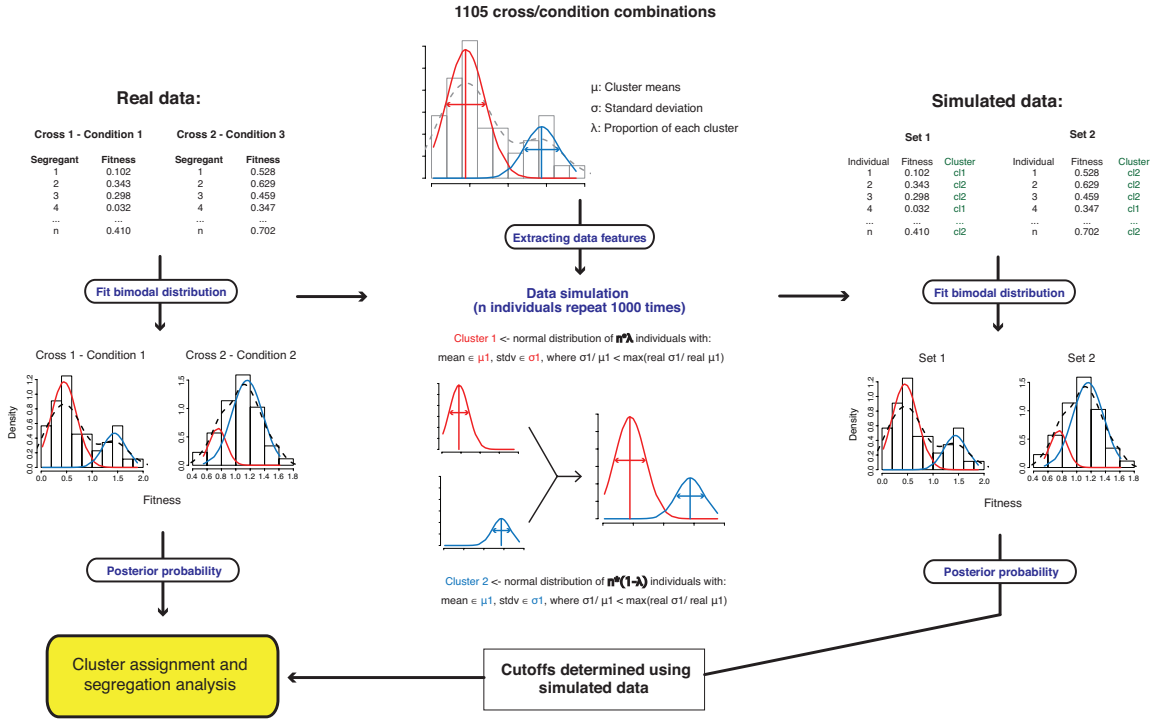


Figure 2. EM-algorithm to detect bimodal distribution using posterior propability

Model fitting procedure for quantitative traits with bimodal distribution

For a trait such as fitness, detection of a bimodal distribution and confident partitioning of the tested individuals could be tricky, especially in large datasets involving multiple cross/trait combinations as described in chapter 4. To this end, we adapted an automated scheme based on Expectation Maximization (EM) algorithm, which is an iterative method for maximum likelihood estimates of predefined statistical model using unobserved latent variables (Figure 2). For the fitness variation data in chapter 4, each of the 1,105 cross/trait combination was fitted to a bimodal distribution using the R package "mixtools"

(https://cran.r-project.org/web/packages/mixtools/index.html) with k = 2 and maxit = 500. Mean (μ), standard deviation (σ), ratio between each cluster (λ) and posterior probability of each cluster for each individual were extracted from the output file. To determine cutoff values of posterior probability for cluster assignment, a simulated dataset was generated, by simulating two normal distributions with n*λ and n*(1-λ) individuals for each cluster, respectively, with mean and standard deviation randomly sampled from observations in real data. For each simulated set, the two normal distributions generated were combined, and the procedure was repeated for 1000 times to generate a training set with 1000 distributions (Figure 2). The training set was then subjected to model fitting with the same parameters (Figure 2). The mean (μ), standard deviation (σ), ratio between each cluster (λ) and posterior probability of each cluster for each individual were extracted again, and the training dataset was evaluated against the real data (Figure 3).

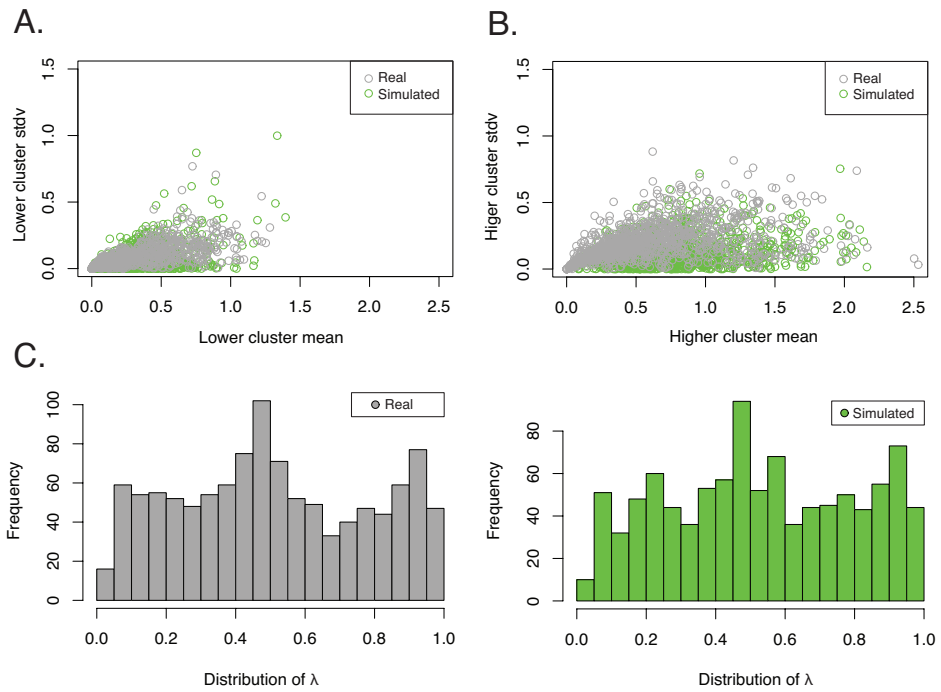


Figure 3. Feature comparison between real and training data

In this case, as the prior probability of cluster assignment was known for each simulated individual, it is possible to test for the detection sensitivity and specificity (Receiver Operating Characteristic or ROC) using varied cutoff parameters. A sequence starting from 0.5 to 0.95 (increment 0.05) for posterior probability and a sequence from 0 to 0.9 (increment 0.1) for percentage of non-overlapping individuals were tested. The ROC curves and area under the curve

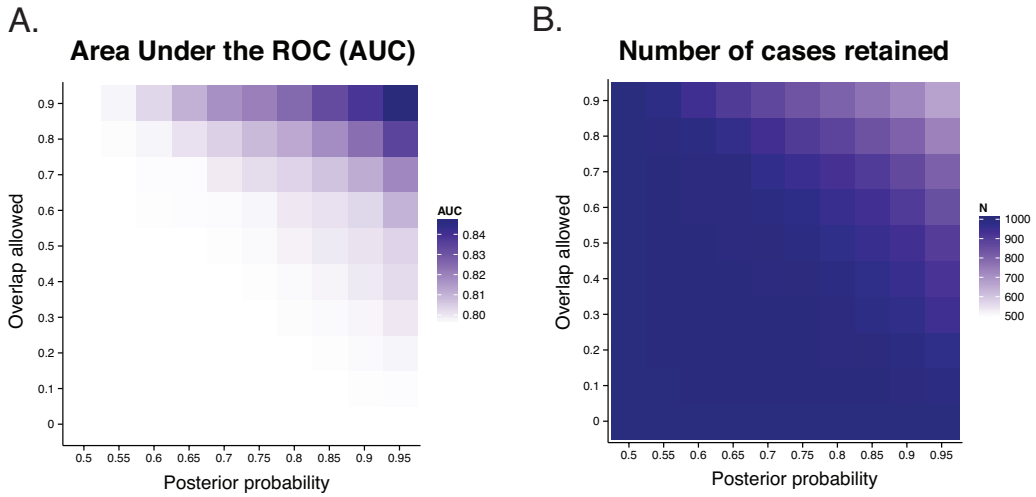


Figure 4. Evaluation of detection power relative to case loss. A. Evolution of area under the curve (**AUC**) with different combinations of parameters. X-axis: posterior probability cutoffs. Y-axis: Fraction of non-overlapping between clusters. **B. Number of cases that passed the threshold** with different combinations of parameters. X-axis: posterior probability cutoffs. Y-axis: Fraction of non-overlapping between clusters.

(AUC) were calculated for each combination of cutoff parameters using R package "ROCR"[15]. Cutoffs of 0.8 for posterior probability and 0.9 for percentage of non-overlapping were retained to ensure confident detections (Figure 4). The defined parameters were applied on real data and cases passed the filter were preceded to cluster assignment (Figure 2). For bimodal cases, the segregation patterns were determined. All analyses were performed in R.

References

1. Schacherer, J., Shapiro, J.A., Ruderfer, D.M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342-5 (2009).
2. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337-41 (2009).
3. Dowell, R.D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
4. Treusch, S., Albert, F.W., Bloom, J.S., Kotenko, I.E. & Kruglyak, L. Genetic mapping of MAPK-mediated complex traits across *S. cerevisiae*. *PLoS Genet* **11**, e1004913 (2015).
5. Jung, P.P., Christian, N., Kay, D.P., Skupin, A. & Linster, C.L. Protocols and programs for high-throughput growth and aging phenotyping in yeast. *PLOS ONE* (in press).
6. Wagih, O. & Parts, L. gitter: a robust and accurate method for quantification of colony sizes from plate images. *G3 (Bethesda)* **4**, 547-52 (2014).
7. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
8. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-303 (2010).
9. Skelly, D.A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* (2013).
10. Huson, D.H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254-67 (2006).
11. Skelly, D.A. *et al.* Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23**, 1496-504 (2013).
12. Bergstrom, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Mol Biol Evol* **31**, 872-88 (2014).
13. Robinson, M.D., Grigull, J., Mohammad, N. & Hughes, T.R. FunSpec: a web-based cluster interpreter for yeast. *BMC Bioinformatics* **3**, 35 (2002).
14. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18 (2012).
15. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940-1 (2005).

Conclusion & Perspectives

On the awesome power of genetics

From genotype to phenotype: insights from species-wide surveys and beyond

In all living organisms, genetic differences constantly emerge and accumulate, providing the raw material for phenotypic variation upon which natural selection operates. Using species-wide surveys, we focused on the ongoing phenotypic consequences of genetic diversity within the *Saccharomyces cerevisiae* yeast species. In particular, our studies unveiled the multiplicity of reproductive isolating mechanisms at the intraspecific scale, which ranges from large-scale chromosomal rearrangements to genetic incompatibilities.

In the aim of obtaining an overview of the genetic origin of intraspecific reproductive isolation in *S. cerevisiae*, our first survey begins with selecting 60 isolates originated from soil, tree barks, immuno-compromised patients and various fermentations across different continents, and systematically crossing these isolates with the laboratory reference strain S288c. By analyzing the offspring viability of each cross, 16 reproductive isolation cases were identified, with reduced offspring viabilities ranging from 44% to 86%. Further analyses led to the identifications of large-scale reciprocal translocations in most cases[1]. It was surprising how widely distributed such rearrangements are both within[2] and among closely related yeast species[3,4]. One extensively studied translocation between chromosome VIII and XVI was found in many *S. cerevisiae* wine isolates, which conferred an advantageous trait of sulfite resistance and could be tightly linked to adaptation of wine making[1,5,6]. While most other documented translocations were mediated by ectopic recombination between *Ty* elements with no apparent selective advantage, the pervasiveness of such rearrangements is still conceivable especially in yeasts. In fact, these events cannot be opposed by natural selection within a clonally expanding population, which represents a large part of the life cycle of *S. cerevisiae*.

Contrary to the prevalence of chromosomal rearrangements, incompatibilities at the gene level seemed to be rare and no evident cases have been found in our first survey. The concept of genetic incompatibility has been formulated almost eight decades ago by Theodosius Dobzhansky[7] and Hermann Müller[8], whereby diverging populations could accumulate independent mutations with no effect in their own genetic backgrounds, yet become incompatible when brought together upon hybridization. The very existence of such genetic incompatibilities among yeasts species has long been a subject of debate, mostly due to the lack of empirical support[9,10]. However, attempts to find such incompatibilities

were mostly performed under laboratory conditions, which involved crossing isolates and then estimate the offspring viability on a rich permissive media that optimize yeast growth. Considering the vast ecological range that natural isolates of yeast encounter in nature, our view of reproductive isolation cases restricted to laboratory conditions might be over simplified.

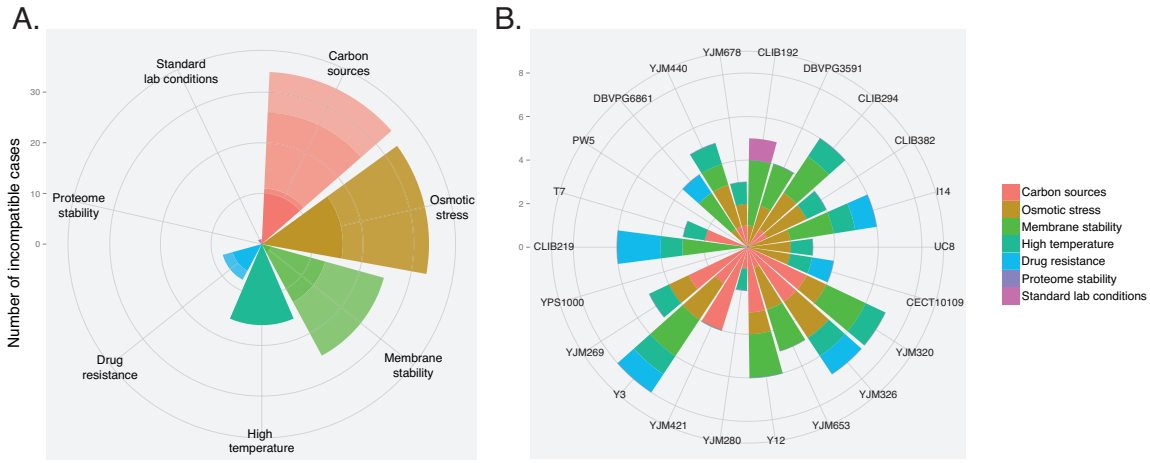


Figure 1 - Negative epistasis in *S. cerevisiae* related to environmental conditions. A. Distribution of epistasis cases according to various stress types. Shades of colors represent different conditions tested that belong to the same category. A total of 117 cases are categorized. **B. Distribution of epistasis cases according to isolates crossed with S288c.** Types of stress are color-coded. Isolates are organized clockwise according to the level of sequence divergence compared to S288c, with CLIB192 being most closely related with a divergence of 0.11%.

With this in mind, we performed subsequently two new rounds of survey with an emphasis on the impact of different culture conditions to the onset of genetic incompatibilities as well as the global landscape of genetic complexity of traits across the *S. cerevisiae* species. Indeed, when taking into account of different environmental factors, negative epistasis involving incompatible genic interactions were much more common than previously considered in *S. cerevisiae*[11]. As a direct extension of our first survey, we selected a subset of 27 crosses that were previously shown to yield high offspring viability on rich media, and tested these crosses on different culture conditions (*e.g.* carbon sources, chemicals that impact various cellular processes and temperatures). In total, 481 cases spanning 27 crosses on 20 conditions were assessed, and 24.3% of all cases (117/481) showed different degrees of condition specific loss of offspring viability ranging from 1% to 62%, indicating the presence of possible negative epistatic interactions. Most cases showed complex segregation patterns of the lethal phenotype, suggesting a higher genetic complexity even at the

intraspecific scale. Within all identified cases, negative epistasis were found in conditions related to various stress types (Figure 1A), and were randomly distributed among different isolates, regardless of their origin or the level of sequence divergence between the parental pair (Figure 1B). In parallel, a new round of crosses between 41 isolates with another strain Σ1278b was performed to avoid potential bias related to the lab strain S288c. In this case, fitness in the offspring were quantitatively measured in 30 stress conditions, resulting in 1,105 cross/trait combinations. A comparable fraction (~20%) of cases were found to be under epistatic control, and similar distribution pattern of instances regarding stress types and parental combinations were observed.

Using a combination of classical genetic analysis and high-throughput genomic mapping strategy, we further identified and characterized the first example of two loci Dobzhansky-Müller incompatibility in yeast related to respiratory conditions. In this case, a clinical isolate YJM421 was incompatible when crossed with the reference S288c due to an interaction between a nonsense mutation in the *COX15* gene and a tRNA suppressor *SUP7*. Curiously, alleles causing this incompatibility may sometimes offer fitness advantages under stress conditions. In fact, when the suppressor mutation *SUP7* in YJM421 was ectopically expressed in other isolates, it conferred to diverse fitness effects. Some isolates displayed significant gain of fitness in some conditions in the presence of the suppressor, and others showed the opposite. This observation suggests that carrying the suppressor might be advantageous in certain environmental conditions, thus balancing the effect of potential offspring loss upon hybridization.

It is now clear that in addition to chromosomal rearrangements, negative epistasis could also lead to the onset of reproductive isolation within yeast natural populations in a condition-specific manner. Nonetheless, it is interesting to note that even though the frequency of potential incompatibilities is relatively high (117/483), most of them were not shared among different isolates, suggesting a rather unique genetic origin for different cases (Figure 1). Taking for example the identified Dobzhansky-Müller case, the allelic combination of $cox15^{stop}$ and *SUP7* were only found in the clinical isolate YJM421, making this isolate universally incompatible with more than 1,000 natural isolates of *S. cerevisiae* that do not possess this combination. Furthermore, same trend was observed when looking at a more substantial dataset of 1,105 cross/trait combinations in the survey with Σ1278b. What do these observations imply in an evolutionary sense?

While the number of cases that were thoroughly characterized here is far from comprehensive, some general trends can be inferred from our surveys. Firstly, it appeared that non-linear genetic interactions or epistasis could contribute to a substantial part of phenotypic variation observed between individuals from the same species, and most variants leading to such interactions are rare. Secondly, genetic incompatibility that follows strictly the original Dobzhansky-Müller model *i.e.* incompatible alleles reciprocally found in diverging parental populations or "derived-derived" type of interaction seem to be less common, and most incompatibilities may likely to be in a "derived-ancestral" configuration (Figure 2). In fact, most documented cases of genetic incompatibilities were "derived-ancestral" across different model systems[9,10,12-14]. One hypothesis, which could be advanced to explain these observations, is that rare genetic variants could potentially have larger phenotypic effects. When a biological system is perturbed by such effects, compensatory mutations have to subsequently arise to escape from the resulting fitness valley[15]. Consequently, such rare combinations could eventually persist in unique genetic backgrounds and in turn lead to epistatic interactions and incompatibilities when crossed with an ancestral background.

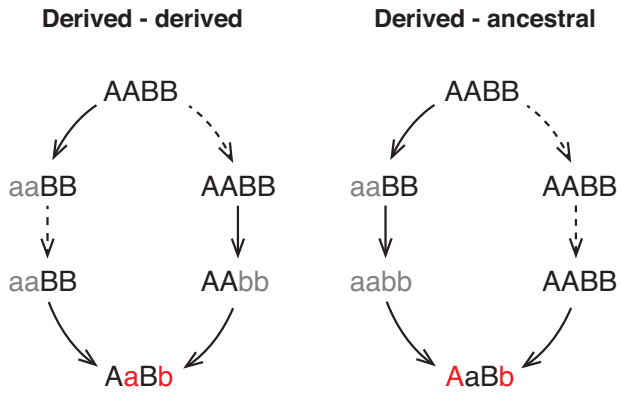


Figure 2. Genetic incompatibility models. Ancestral genotype AABB following different trajectories leading to nascent populations with "derived-derived" or "derived-ancestral" types of incompatibility. Mutations are indicated in grey and incompatible genotypes in red.

Retrospectively, it will be then bold to hypothesize that rare functional variants with large phenotypic effect would be more likely involved in epistasis due to the presence of potential compensatory mutations or modifiers. By analyzing over 1,100 cross/trait combinations in our last survey, we systematically identified and characterized genetic variants leading to monogenic Mendelian inheritance across the species. These variants are solely responsible for the trait observed and therefore have large phenotypic effects. Interestingly, it seems that

causal Mendelian variants that were common in the population tend to have stable inheritance patterns across different parental combinations, whereas rare variants could have biased inheritances, indicating the presence of modifiers and complex interactions. But then again, this hypothesis at present remains purely a conjecture.

Yeasts as a model: the promise of classical genetics in the genomic era

In January 1920, half issue of the journal Genetics was dedicated to the publication of "The genetic basis of truncate wing - an inconstant and modifiable character in *Drosophila*" by Edgar Altenburg and Hermann J. Müller. Merely ten years after the rediscovery of Mendel's law of inheritance, the two young brilliant minds dissected the first identified complex trait in *Drosophila* - the deformation of wing shape known as "truncate" - in amazing genetic detail. The work was completed through sets of carefully designed crossing experiments that were able to trace the effect of individual causal variants by linkage with visible genetic markers, which ultimately led to the resolution of the "truncate" case into 3 loci: a "chief" gene that contribute deterministically to the trait and two "intensifier" genes located on separate chromosomes. Concluding the experiments, the authors stated that:

> *"It is believed by the authors that the general method of attack developed in the truncate case, whereby, by the use of "identifying" genes, a refractory character may be taken apart, put together, or held in a desired combination, will become of more widespread applicability as the linkage groups of the organisms commonly used for genetic study become better known."*
>
> *---- Altenburg E. & Müller H.J., 1920*

Indeed, the beauty of classical genetics has been and still is the ability to "take apart, put together or hold in a desired combination" of causal genetic variants of complex characters. Today, a century after its birth, classical genetics may provide new promises to understand the genotype-phenotype relationship in the genomic era.

With the advent of sequencing technologies, ever growing whole genome data are routinely generated across the tree of life. These advances prompt the rapid development of genome-wide association and linkage mapping strategies, allowing direct inference of causal genomic loci from phenotypic data in a wide range of model and non-model organisms. However, charting the genotype-

phenotype map using these strategies is far from complete. For example, in the hope of predicting phenotypes such as diseases from genomic data, human genome sequencing has been at the forefront of population genomics since its inception[16]. However, most association studies showed that common genetic variants explain relatively little of the phenotypic variance observed[17]. After all, it has been shown that human populations are riddled with rare genetic variants[18], which possibly have larger phenotypic effects[19-21], and perplexed by gene-gene and gene-environment interactions[22] - issues that stems the problem of "missing heritability" in most complex traits[17,22].

Ultimately, the key to the "missing heritability" problem requires a better understanding of the genetic architecture of traits: to take apart, put together or hold in a desired combination of variants involved and elucidate their number, type, effect size and frequency within a population. Yeast models may be in the best position to achieve this goal. Compare to other model organisms, yeast species, especially *S. cerevisiae*, presents vast genetic diversity, small and compact genomes, well-mastered sexual reproduction and efficient laboratory manipulation. Especially, the particularity of the yeast cell cycle allows for tetrad analysis, a unique feature of yeast genetics that offers an unparalleled opportunity to examine the complete product of any single meiosis event. Species-wide genetic surveys, initiated in this work, have been fruitful especially in tackling epistatic genetic interactions, tracing the effects of common or rare variants in different genetic backgrounds and revealing the gradual progression of the hidden complexity of Mendelian traits across a species. For the next step, we will continue to characterize in detail the identified cases of epistasis, including those with more than two loci involved. In parallel, in the aim to get a comprehensive picture of the genetic architecture of phenotypic variation across *S. cerevisiae*, a new project is currently in preparation based on a large-scale unbiased diallel cross design.

The Matrix Reloaded project

In fact, our first view of the genetic complexity of traits is biased in many ways. First, a reference strain (either S288c or Σ1278b) was involved in each of the performed and studied crosses. As a consequence, we have not taken full advantage of the genetic diversity present in the whole *S. cerevisiae* species. In addition, strong effect alleles specific to the reference genome may have an impact on multiple traits and multiple crosses, leading to a biased view. Second, to complete the first phenotypic screen we only tested 31 common and well

studied conditions. Testing more conditions would bring a broader view of the phenotypic and genetic complexity. In this context, the obvious next step is to scale up in every regard. To truly understand the genetic architecture of traits, we need "many by many" crosses (as opposed to the "one by many" crosses we have already performed) and additional traits spanning a broader phenotypic range.

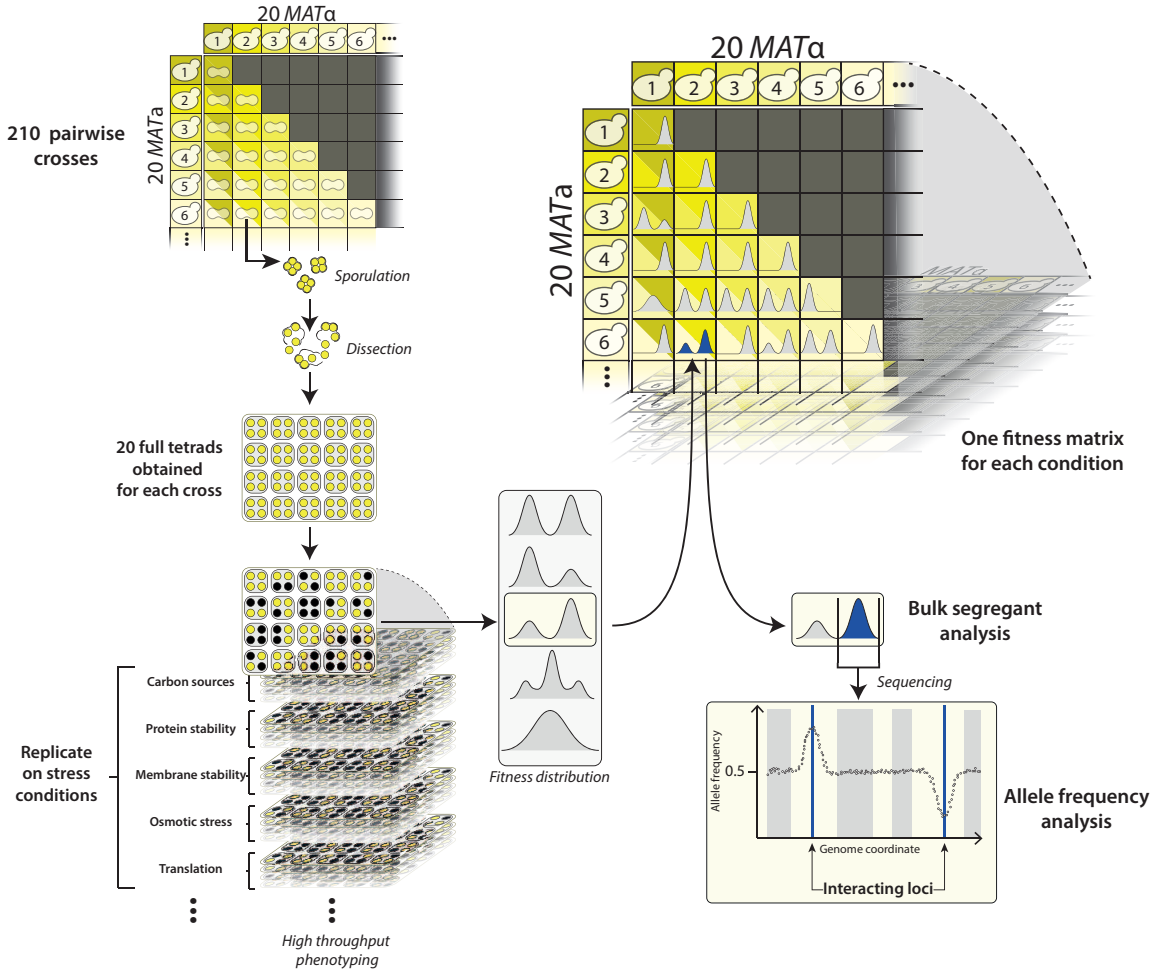


Figure 3. Overview of the "Matrix Reloaded Project". Courtesy of Teo Fournier

Accurately dubbed as the "Matrix Reloaded Project", the proposed study is based on pairwise crosses of a set of natural isolates representative of the whole species diversity. With the recent completion of the whole genome resequencing of over 1,000 natural isolates, initiated by our lab (The 1002 yeast genomes project, http://1002genomes.u-strasbg.fr/), we have currently the best understanding of the natural genetic and phenotypic diversity of any eukaryote model system to date. Taking advantage of this rich resource, we aim to select ~20 isolates that are diploid, homozygous, genetically diverse and present unbiased

population structure. For all selected isolates, stable haploid founder strains will be generated and then crossed to each other, creating a diallel hybrid panel with 210 individuals with unique genomic combinations including homozygote diploids (Figure 3).

For all crosses, 100 haploid recombinant offspring will be generated in the form of complete tetrads for each, and the entire diallel offspring panel, consisting with more than 21,000 individuals, will be subjected to high-throughput phenotyping on ~50 non-correlating conditions along with the parental isolates and hybrids. Compare to our previous surveys, the proposed diallel design will not only provide an unbiased view of the genetic complexity of trait variation, but also informs the population frequency of causal variants, the dynamics of inheritance patterns across different genetic backgrounds and a closed-ring like network relationship between parental combinations and genetic variants.

In addition to segregation analyses in the offspring, the diallel hybrid panel itself offers an exciting opportunity to test genome-wide association in a population with pairwise allelic combinations. It will even be envisionable to explore the possibility of taking patterns of offspring inheritance as a complex trait, and try to directly associate genetic complexity with modifier loci.

Overall, data to be obtained within the frame of this project will undoubtedly offer an additional dimension to the understanding of the phenotypic outcomes of genetic variants within yeast natural populations, and holds great potential to elucidate the genetic architecture of complex traits in broader contexts.

"Neo is the one."

--- The Matrix, 1999

References

1. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Curr Biol* **24**, 1153-9 (2014).
2. Charron, G., Leducq, J.B. & Landry, C.R. Chromosomal variation segregates within incipient species and correlates with reproductive isolation. *Mol Ecol* **23**, 4362-72 (2014).
3. Delneri, D. *et al.* Engineering evolution to study speciation in yeasts. *Nature* **422**, 68-72 (2003).
4. Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G. & Louis, E.J. Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451-4 (2000).
5. Perez-Ortin, J.E., Querol, A., Puig, S. & Barrio, E. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res* **12**, 1533-9 (2002).
6. Clowers, K.J., Heilberger, J., Piotrowski, J.S., Will, J.L. & Gasch, A.P. Ecological and Genetic Barriers Differentiate Natural Populations of *Saccharomyces cerevisiae*. *Mol Biol Evol* (2015).
7. Dobzhansky, T. *Genetics and the origin of species*, xvi, 364 p. Columbia Univ. Press (1937).
8. Muller, H. Isolating mechanisms, evolution and temperature. in *Biol. Symp* (1942).
9. Chou, J.Y., Hung, Y.S., Lin, K.H., Lee, H.Y. & Leu, J.Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biol* **8**, e1000432 (2010).
10. Heck, J.A. *et al.* Negative epistasis between natural variants of the *Saccharomyces cerevisiae MLH1* and *PMS1* genes results in a defect in mismatch repair. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 3256-3261 (2006).
11. Hou, J., Friedrich, A., Gounot, J.S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nat Commun* **6**, 7214 (2015).
12. Seidel, H.S., Rockman, M.V. & Kruglyak, L. Widespread genetic incompatibility in *C. elegans* maintained by balancing selection. *Science* **319**, 589-94 (2008).
13. Phadnis, N. *et al.* An essential cell cycle regulation gene causes hybrid inviability in *Drosophila*. *Science* **350**, 1552-5 (2015).
14. Presgraves, D.C. The molecular evolutionary basis of species formation. *Nat Rev Genet* **11**, 175-80 (2010).
15. Phillips, P.C. Epistasis--the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* **9**, 855-67 (2008).
16. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
17. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-53 (2009).
18. Cirulli, E.T. & Goldstein, D.B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* **11**, 415-25 (2010).
19. Tennessen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64-9 (2012).
20. Nelson, M.R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100-4 (2012).
21. Consortium, U.K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82-90 (2015).
22. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A* **109**, 1193-8 (2012)

Appendices

Résumé de thèse

Dans toute espèce, les individus possèdent une diversité génétique considérable, conduisant à la variation des caractères phénotypiques permettant l'adaptation des organismes aux différents environnements. Un objectif central en biologie est de comprendre la relation entre le génotype et le phénotype, c'est-à-dire comment les variations au niveau des génomes, tels que les mutations ponctuelles et réarrangements chromosomiques, peuvent être à l'origine de l'émergence et le maintien de la biodiversité observée. Mes travaux de thèse s'inscrivent dans cette grande thématique et focalisent sur les conséquences phénotypiques de la variation génétique au sein des populations naturelles d'une même espèce, en utilisant la levure *Saccharomyces cerevisiae* comme système d'étude. Dans un premier temps, je me suis focalisée sur l'effet des mutations présentes dans différents isolats naturels sur l'apparition de l'isolement reproductif, un processus engendrant la perte de la viabilité de la descendance lors du croisement. Dans un second temps, en plus des phénotypes sévères tels que la létalité de la descendance, je me suis aussi intéressée à la caractérisation de la complexité génétique d'un ensemble de traits quantitatifs tels que la croissance sur différents conditions de culture, afin d'avoir une vision globale du nombre de gènes impliqués et de leur mode d'interaction qui sous-tend la variation phénotypique entre individus au sein d'une même espèce.

L'isolement reproductif limite les échanges génétiques entre les populations et est considéré comme une étape clé dans la formation de nouvelles espèces. L'origine de l'isolement reproductif pourrait être génétique, en ce que les populations,

séparés par des barrières géographiques ou écologiques, accumulent indépendamment des mutations qui ne sont pas compatibles dans les hybrides, conduisant à des descendants stériles ou non viables.

Il existe deux types d'isolement reproductif : pré-zygotique et post-zygotique. L'isolement reproductif pré-zygotique se traduit par l'incapacité de former un zygote entre individus appartenant à des espèces différentes. Par exemple, certaines espèces de drosophiles (*Drosophila mauritiana* et *Drosophila sechellia*) ne peuvent pas former d'hybrides de par l'incompatibilité de leur organes sexuels. Dans le cas de l'isolement reproductif post-zygotique, les individus d'espèces différentes peuvent se croiser et former un zygote. Cependant cet hybride est souvent stérile ou présente une faible viabilité de la descendance. C'est le cas des levures appartenant au genre *Saccharomyces*. Les individus appartenant à ces espèces se croisent mais seul 1 % de la descendance est viable.

Les origines de l'isolement reproductif post-zygotique peuvent être multiples. Une de ces origines possibles est la présence d'incompatibilités génétiques définies par le modèle proposé par Theodosius Dobzhansky et Hermann Müller qualifié par conséquent de modèle de Dobzhansky-Müller. Ce modèle met en avant l'hypothèse selon laquelle des interactions entre allèles provenant de différentes espèces peuvent avoir un impact sur la viabilité de l'hybride ou de la descendance. Plus concrètement, une espèce ancêtre ayant un génotype « AABB » peut, au cours de son évolution, coloniser des niches écologiques ou géographiques différentes. Au sein de ces écosystèmes, des mutations « a » et « b » incompatibles peuvent apparaître et se fixer de manière indépendante. Ces mutations n'ont pas de conséquence sur la viabilité de la descendance au sein de leur environnement génétique respectif. Le point critique apparaît uniquement lorsque les deux mutations incompatibles sont réunies. En effet, la réunion de ces

deux mutations incompatibles « a » et « b » conduit à la létalité de l'hybride ou à la réduction de la viabilité de la descendance.

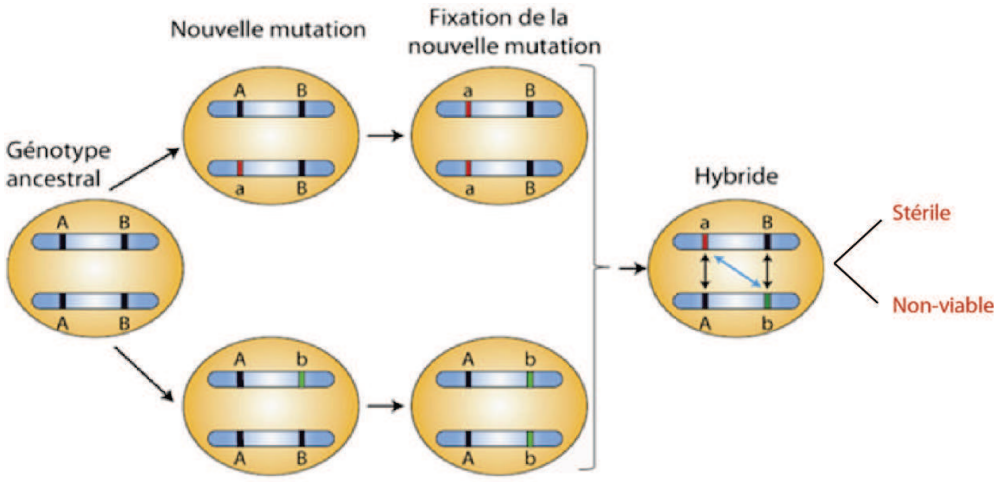


Figure 1. Modèle Dobzhansky-Müller

Bien que ce modèle fut établi dans les années 1940, ce n'est qu'au cours des 10 dernières années que des cas concrets d'incompatibilités génétiques ont été mis au jour entre isolats appartenant à différentes espèces comme par exemple entre *Drosophila melanogaster* et *Drosophila simulans*. Cependant, ces exemples d'incompatibilité inter-spécifique ne permettent pas de conclure quant à l'origine des événements de spéciation. En effet, il est difficile de savoir si ces incompatibilités sont une cause ou une conséquence de la spéciation sachant que les espèces continuent à diverger et évoluer après cet événement.

Au cours des dernières décennies, de nombreuses études ont été focalisées sur les bases génétiques de l'isolement reproductif entre espèces proches de différents taxa, tels que les drosophiles, les souris et les plantes (Presgraves, 2010).

Cependant, il est difficile de savoir si les mécanismes observés chez les espèces différentes sont la cause ou la conséquence. En effet, des études récentes utilisant différents organismes modèles, tels que la plante *Arabidopsis thaliana* (Bikard et al., 2009) et le vers *Caenhorhabditis elegans* (Seidel et al., 2008), ont montré que multiple mécanismes génétiques conduisant à l'isolement reproductif peuvent exister entre isolats au sein d'une même espèce. Néanmoins, bien que des exemples isolés ont été identifiés, étude systématique de l'apparition de l'isolement reproductif au sein de grande population à travers de l'ensemble de la diversité d'une même espèce n'a pas encore réalisée.

Dans cette optique, la levure *S. cerevisiae* semble être un modèle de choix (Figure 2). De nombreux isolats de cette espèce ont été isolés à partir des environnements écologiques divers (vigne, exsudat d'arbre, sol, saké, bière) sur différents continents, et présentent une diversité phénotypique considérable. De par son génome petit (~12 Mb) et compact (~5500 gènes), plus de ~1000 isolats naturels ont été entièrement séquencé à ce jour (1002 yeast genomes project, http://1002genomes.u-strasbg.fr/). Grâce à son cycle sexuel bien caractérisé et un temps de génération courte, il est possible de réaliser facilement un grand nombre de croisements entre isolats d'origines différentes et déterminer de manière précise la viabilité de la descendance (Figure 3A).

Afin d'avoir une vision globale de l'apparition de l'isolement reproductif au sein de cette espèce, nous avons sélectionné une collection de 60 isolats d'origine diverse représentant la diversité globale de cette espèce. L'ensemble de cette collection a été croisé avec une souche de référence au laboratoire S288c, et un grand nombre de descendants ont été analysés (Figure 3C). Au total, 16 cas d'isolement reproductif ont été identifiés, avec une viabilité de descendance

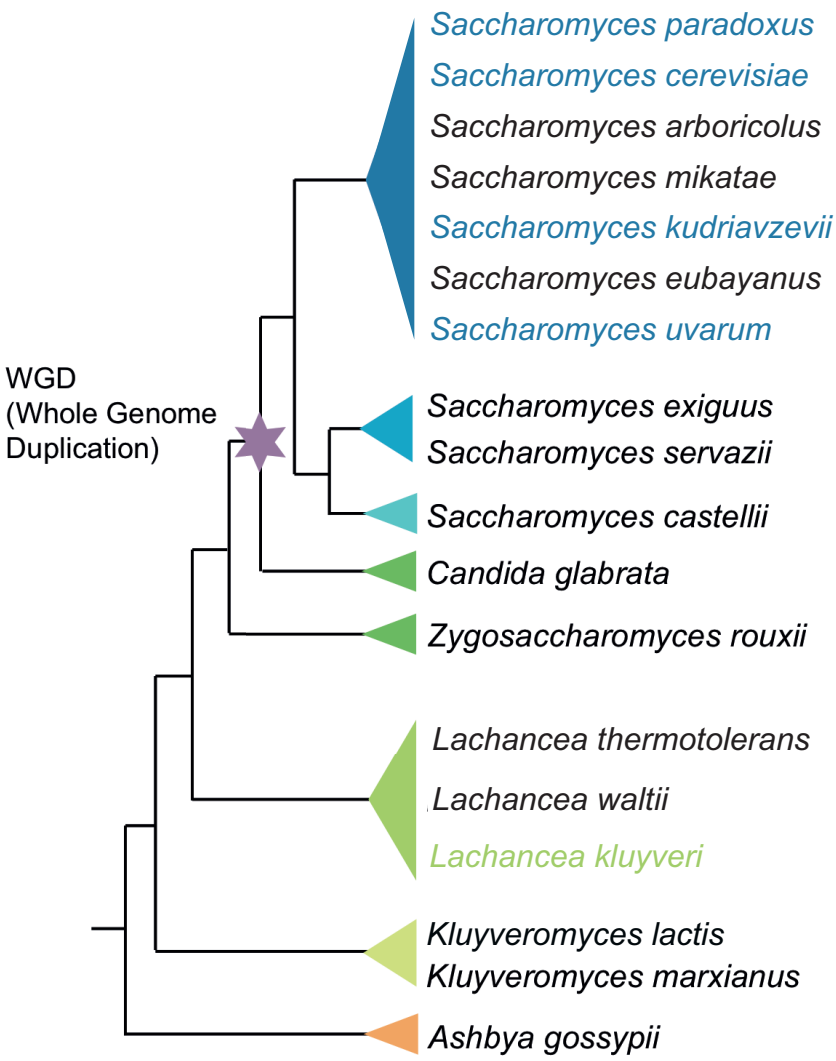


Figure 2. Arbre phylogénétique des levures d'hémiascomycète

réduite allant de 44 % à 86 %. Par la suite, les régions génomiques ont été cartographiées avec une stratégie de séquençage à haut-débit. Les analyses dans ces régions ont permit d'identifier des grands réarrangements chromosomiques de type translocation réciproque dans la plupart des cas (Hou et al., 2014). En effet, ce genre de réarrangement peut conduire à une perte de la viabilité due à la répartition inégale de gènes essentiels dans la descendance lors du croisement. Il est à noter que ce genre de réarrangement a été aussi observé entre isolats

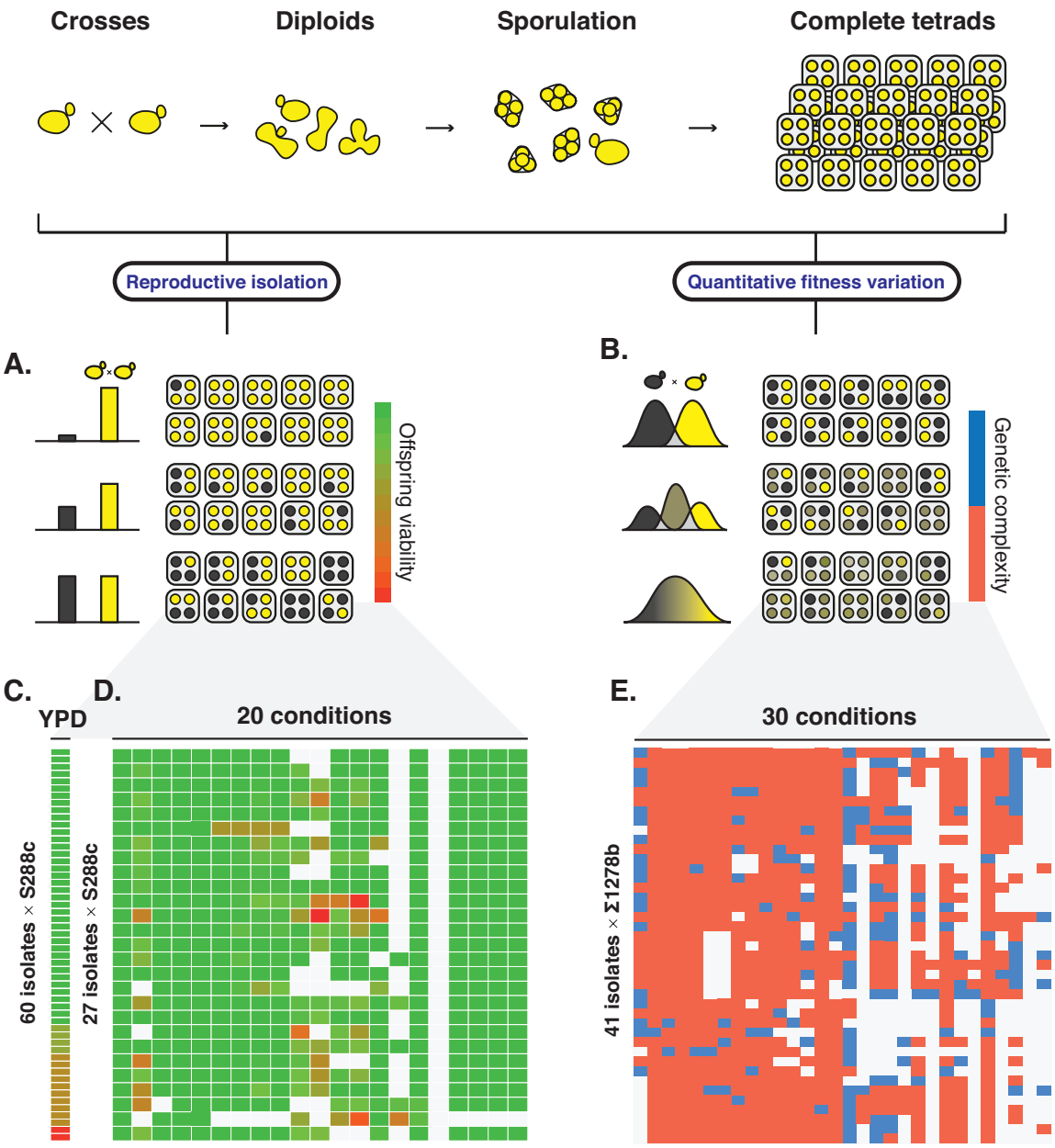


Figure 3. Représentation schématique du projet

d'espèces proches des levures du genre *Saccharomyces* (Fischer et al., 2000), révélant ainsi l'importance de ce mécanisme sur l'apparition d'isolement reproductif et potentiellement la spéciation (Figure 2).

Cette première étude a permis d'avoir une vue globale quand aux mécanismes impliqués dans l'isolement reproductif au sein d'une même espèce, cependant, l'ensemble de ces résultats ont été obtenu seulement sur un milieu de culture riche couramment utilisée au laboratoire et pourrait donc être biaisé. Afin de mettre en lumière l'impact de différents environnements sur l'apparition d'isolement reproductif au sein de cette espèce, nous avons sélectionné 27 croisements présentant une forte viabilité de la descendance (>90%) sur milieu riche, et les avons testé sur 20 conditions de culture différentes (Figure 1D). Les conditions choisies ont des effets sur divers processus cellulaires, tels que la traduction, la transcription, la transduction des signaux *etc*. Au total, 481 cas répartis sur 27 croisements en présence de 20 conditions ont été évalués. Dans 24,3% des cas (117/481), une perte de la viabilité de la descendance spécifique aux conditions de culture a été détectée, allant de 1% à 62%. Cette observation suggère fortement la présence des interactions épistatiques provenant des mutations incompatibles au sein des isolats parentaux. En effet, lors que l'on s'intéresse à un exemple précis lié aux conditions de source de carbone non-fermentescible (glycérol ou éthanol), on a pu mettre en évidence un cas d'incompatibilité génétique classique de type Dobzhansky-Müller qui est due à une interaction épistatique entre deux loci. En utilisant une stratégie d'analyse génétique classique et le séquençage à haut-débit, nous avons identifié les gènes et les mutations qui sont impliqués. Dans ce cas-là, la perte de descendance observée est due à l'interaction entre une mutation nonsense dans le gène *COX15* et un suppresseur de type ARNt (Hou et al., 2015). En combinant avec les résultats obtenus précédemment, notre étude à travers les populations naturelles au sein de *S. cerevisiae* ont mise en évidence la multiplicité des mécanismes moléculaires pouvant être impliqué dans l'isolement reproductif.

Dan l'ensemble, ces premiers efforts à l'échelle de l'espèce ont mis en lumière l'importance de la variation génétique présente dans les populations naturelles, que ce soit des grands réarrangements chromosomiques ou des mutations ponctuelles, sur l'apparition des phénotypes extrêmes telles que la perte de la viabilité de la descendance. Cependant, comment ces variations génétiques peuvent contribuer et influencer les phénotypes quantitatifs reste peu connu. Afin d'avoir un aperçu global de la complexité génétique des traits au sein de *S. cerevisiae*, nous avons mené une étude systématique en analysant la distribution et la ségregation de 30 phénotypes quantitatifs liés à la croissance sur différentes condition de culture chez les descendants d'un grand nombre de croisements (Figure 1E). Concrètement, 41 croisements ont été générés à partir des isolats naturels et 40 descendants issus de chaque croisement ont été testés sur 30 conditions. La croissance de chaque individu a été mesurée de manière quantitative, et la distribution phénotypique de chaque trait a été évaluée. Au total, nous avons pu déduire la complexité génétique sous-jacente de 880 cas. Les résultats ont montré que la plupart des traits sont complexes (82,6%, 727/880), avec 23,6% présentant des interactions épistatiques. En revanche, les traits monogéniques sont rares, représentant 9,8% (86/880). Ces analyses ont permit d'avoir une première estimation de la proportion relative des traits monogéniques *vs.* complexes au sein d'une population naturelle. Les gènes et les mutations impliqués dans ces différents traits sont en cours d'identification, ce qui va sans doute approfondir notre connaissance sur l'impact phénotypique des variations génétiques au sein d'une grande population naturelle.

Références

Bikard, D., Patel, D., Le Mette, C., Giorgi, V., Camilleri, C., Bennett, M.J., and Loudet, O. (2009). Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. Science *323*, 623-626.

Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., and Louis, E.J. (2000). Chromosomal evolution in Saccharomyces. Nature *405*, 451-454.

Hou, J., Friedrich, A., de Montigny, J., and Schacherer, J. (2014). Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in Saccharomyces cerevisiae. Current biology : CB *24*, 1153-1159.

Hou, J., Friedrich, A., Gounot, J.S., and Schacherer, J. (2015). Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. Nature communications *6*, 7214.

Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nature reviews Genetics *11*, 175-180.

Seidel, H.S., Rockman, M.V., and Kruglyak, L. (2008). Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science *319*, 589-594.

List of publications

Hou J, Sigwalt A, Pflieger D, Peter J, de Montigny J, Dunham M & Schacherer J. (2016). The hidden complexity of Mendelian traits across yeast natural populations. ***BioRxiv***.

Hou J, Schacherer J. (2016). On the mapping of epistatic interactions in natural isolates of yeast: combining classical genetics and genomics. ***Methods in Molecular Biology***. 1361:345-60.

Hou J, & Schacherer J. (2015). Negative epistasis: a route to intraspecific reproductive isolation in yeast? ***Current Genetics***. 1-5.

Hou J, Friedrich A, Gounot J-S, Schacherer J. (2015). Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. ***Nature Communications***. 6, 7214.

Freel KC, Friedrich A, **Hou J**, Schacherer J. (2014). Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. ***Genome biology and evolution***. 6(10):2586-94.

Hou J, Friedrich A, de Montigny J, Schacherer J. (2014). Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. ***Current biology***. 24(10):1153-9.

Friedrich A, Jung PP, **Hou J**, Neuveglise C, Schacherer J. (2012). Comparative mitochondrial genomics within and among yeast species of the *Lachancea* genus. ***PloS one***. 7(10):e47834.

Jung PP, Friedrich A, Reisser C, **Hou J**, Schacherer J. (2012). Mitochondrial genome evolution in a single protoploid yeast species. ***G3***. 2(9):1103-11.

List of communications

27th International Conference on Yeast Genetics and Molecular Biology
Levico Terme, Trentino, Italy (2015)

Systematic analysis of the genetic complexity of traits across yeast natural populations. (Oral)

Species-wide survey reveals the complex landscape of the genetic origin of reproductive isolation within natural yeast populations. (Poster)

FEBS-EMBO International Conference
Paris, France (2014)

A two loci genetic incompatibility leading to offspring respiratory deficiency in yeast. (Poster)

GSA Yeast Genetics Meeting
Seattle WA, USA (2014)

A two loci genetic incompatibility leading to offspring respiratory deficiency in yeast. (Poster)

The 11th French national meeting - Levure, Modèles et Outils
Bordeaux, France (2014)

Condition specific reproductive isolation within a yeast species. (Oral)

Séminaire de Microbiologie de Strasbourg
Strasbourg, France (2013-2014)

A two loci genetic incompatibility leading to offspring respiratory deficiency in yeast. (Oral)

Genetic basis of intraspecific reproductive isolation in yeast. (Oral)

Experimental Approaches to Evolution and Ecology using Yeast
Heidelberg, Germany (2012)

Genetic basis of intraspecific reproductive isolation in yeast. (Poster)

Prénom NOM

TITRE de la thèse

Résumé

Un objectif central en biologie est de comprendre la relation entre le génotype et le phénotype. Mes travaux de thèse s'inscrivent dans cette thématique et focalisent sur les populations naturelles d'une même espèce, en utilisant la levure *Saccharomyces cerevisiae* comme système d'étude. Dans un premier temps, je me suis focalisée sur l'effet des mutations présentes dans différents isolats naturels sur l'apparition de l'isolement reproductif, un processus engendrant la perte de la viabilité de la descendance lors du croisement. Dans un second temps, en plus des phénotypes sévères tels que la létalité de la descendance, je me suis aussi intéressée à la caractérisation de la complexité génétique d'un ensemble de traits quantitatifs tels que la croissance sur différents conditions de culture, afin d'avoir une vision globale du nombre de gènes impliqués et de leur mode d'interaction qui sous-tend la variation phénotypique au sein d'une même espèce.

Mots clés : isolement reproductif, complexité génétique, génomique, levure

Résumé en anglais

Elucidating the genetic origin of phenotypic diversity among individuals within the same species is essential to understand evolution. Using the yeast *Saccharomyces cerevisiae*, we showed that reproductive isolation could readily segregate at the intraspecific level, which is governed by various molecular mechanisms ranging from large-scale chromosomal changes to incompatible epistatic genetic interactions. Compared to reproductive isolation, other phenotypes such as monogenic Mendelian traits are thought to be simple in terms of their phenotypic penetrance and genetic constitution. However, our survey showed that the expressivity of monogenic mutations and hence the inheritance pattern of a Mendelian trait could also depend on parental combinations, transitioning from simple to complex. Our studies unveiled the multiplicity and complexity of the genetic origin of phenotypes within a population, from the origin of reproductive isolation to the hidden complexity of Mendelian traits.

Key words: reproductive isolation, genetic complexity, genomics, yeast