



HAL
open science

Origines génétiques de la variation de tolérance au stress au sein de populations naturelles de levures

Anastasie Sigwalt

► **To cite this version:**

Anastasie Sigwalt. Origines génétiques de la variation de tolérance au stress au sein de populations naturelles de levures. Génétique des populations [q-bio.PE]. Université de Strasbourg, 2016. Français. NNT : 2016STRAJ024 . tel-01492857

HAL Id: tel-01492857

<https://theses.hal.science/tel-01492857>

Submitted on 20 Mar 2017

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

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

ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

**Unité de recherche génétique moléculaire, génomique,
microbiologie (GMGM) - UMR7156**

THÈSE présentée par :
Anastasié SIGWALT

soutenue le : **3 juin 2016**

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

Discipline / Spécialité : Sciences du Vivant / Aspects moléculaires et cellulaires de la Biologie

**Origines génétiques de la variation de tolérance au
stress au sein de populations naturelles de levures**

THÈSE dirigée par :

M. SCHACHERER Joseph

Maître de conférences, Université de Strasbourg

M. DE MONTIGNY Jacky

Professeur, Université de Strasbourg

RAPPORTEURS :

M. FISCHER Gilles

Directeur de recherches, Université Pierre et Marie Curie, Paris

M. MARULLO Philippe

Chargé de recherches, Université de Bordeaux

AUTRES MEMBRES DU JURY :

M. HUGUENEY Philippe

Directeur de recherches, Université de Strasbourg

M. LEMAIRE Marc

Professeur, Université Lyon I

Résumé

Une question centrale de la génétique moderne est de mieux comprendre comment la variation génétique présente au sein d'individus d'une même espèce influence la diversité phénotypique et l'évolution. La levure modèle *Saccharomyces cerevisiae* offre une occasion unique d'apporter des éléments de réponse à cette question à travers la dissection de l'architecture génétique de la variation de tolérance à des stress environnementaux à l'échelle d'une population. Mon étude révèle un niveau supplémentaire de complexité de la relation génotype-phénotype où finalement les caractères supposés les plus simples, dits Mendéliens (déterminisme strictement monogénique) peuvent se révéler être complexes (déterminisme multigénique) selon le fonds génétique en raison de l'action de gènes modificateurs, d'interactions épistatiques et/ou de suppresseurs. Toutefois, les processus évolutifs peuvent être bien différents en fonction des espèces. Afin de mieux les décrypter, je me suis également intéressée à *Lachancea kluyveri*, une levure phylogénétiquement distante de *S. cerevisiae*. Cette espèce présente une diversité génétique plus élevée et constitue une ressource encore peu exploitée. L'exploration de la diversité phénotypique et la détermination de ses origines génétiques initiées dans cette étude sont extrêmement prometteuses et apportent de solides fondations pour l'étude à la fois de l'architecture génétique des caractères et de l'évolution de la relation génotype-phénotype au sein de diverses espèces de levures.

Remerciements

Le travail présenté dans ce mémoire a été réalisé au sein du laboratoire de génétique moléculaire, génomique et microbiologie (GMGM), UMR7156/CNRS, Université de Strasbourg, sous la direction du Pr. Jacky de Montigny et du Dr. Joseph Schacherer dans l'équipe variation intra-spécifique et évolution des génomes dite haplo. Je souhaite remercier Pr. Serge Potier, ancien directeur du laboratoire et actuel directeur de l'École doctorale des Sciences de la Vie et de la Santé (ED414), Université de Strasbourg, de m'avoir accueillie dans le laboratoire et de m'avoir autorisée à soutenir ma thèse. Je remercie les Instituts américains de la Santé (NIH), la Région Alsace et l'Université de Strasbourg d'avoir financé mes travaux de thèse.

Je tiens à remercier Dr. Gilles Fischer (Université Pierre et Marie Curie, Paris), Dr. Philippe Marullo (Université de Bordeaux), Pr. Marc Lemaire (Université de Lyon 1) et Dr. Philippe Huguéney (Université de Strasbourg/INRA de Colmar) d'avoir accepté d'examiner mon travail. Je souhaite adresser toute ma gratitude à Philippe Huguéney pour avoir participé à mon comité de mi-thèse.

Je remercie particulièrement mes directeurs de thèse, Joseph Schacherer et Jacky de Montigny. Merci de m'avoir fait confiance depuis le master et de m'avoir donné la chance de poursuivre en thèse. J'ai été très heureuse de vous avoir comme directeurs de thèse. Joseph, je te remercie de m'avoir parfaitement encadrée, guidée et conseillée tout au long de ma thèse et de m'avoir permis de mener mon projet de bout à bout. Jacky, je te remercie de m'avoir permis d'acquérir une première expérience dans l'enseignement supérieur et pour les nombreux conseils que tu m'as apportés concernant mon projet professionnel.

Je remercie toute l'équipe haplo sans qui mon travail n'aurait pu aboutir sans les compétences scientifiques et les qualités humaines des personnes la composant. Merci beaucoup pour toutes les lunettes au flan qui m'ont permis de recharger mes batteries. J'ai beaucoup apprécié les bons moments passés avec vous au laboratoire et en dehors. Je garderai en mémoire les ekidens où j'ai plus brillé en tant que supportrice que coureuse. Je tiens à remercier les personnes qui ont contribué à ce travail, Anne, Christian, Claudia, David, Jackson, Jing et Paul. Jing, un grand merci pour tes nombreux conseils et ton aide précieuse qui m'ont permis d'avancer, pour les MilkyWay qui m'ont permis de me redonner la pêche et pour ta bonne humeur à toute épreuve. Je garde un bon souvenir des petites expéditions à Illkirch quelquefois mouvementées mais au final bien

amusantes. Je te souhaite tout le meilleur pour la suite et une brillante carrière de docteur, tu le mérites. Claudia, merci beaucoup d'avoir contribué à mon travail avec *ENA2*, je t'en suis très reconnaissante. Anne, David et Jackson, je vous remercie pour votre soutien bioinformatique et les nombreuses analyses réalisées. Anne et Claudia, merci beaucoup pour votre écoute bienveillante. J'ai été très contente de partager le bureau avec vous. Christian, merci pour tes blagues étranges mais efficaces pour mettre l'ambiance dans le bureau. Paul, merci de m'avoir initiée aux techniques de génétique de la levure et de m'avoir encadrée pendant mon stage de master. Cyrielle, merci beaucoup pour tes encouragements depuis mes débuts au laboratoire. Claudine et Kelle, merci pour votre joie de vivre. Je souhaite bonne chance et beaucoup de succès aux doctorants de l'équipe Jackson, Jean-Sébastien et Téo pour la suite de leur thèse et une belle réussite à Arnaud pour son master.

Merci également à tous les autres membres du laboratoire. Je remercie Benoît Kammerer qui m'a permis d'acquérir des connaissances solides en microbiologie tout au long de mon cursus universitaire depuis la licence jusqu'au doctorat.

Je voudrais remercier toute l'équipe pédagogique de virologie, Salah Bouzoubaa, Maria Dimitrova, Alyssa Flobinus, David Gilmer et Corinne Keichinger. Merci de m'avoir intégrée dans votre équipe et de m'avoir confié des responsabilités d'enseignement. Cette expérience a été très enrichissante et j'en garderai toujours un très bon souvenir. Alyssa, j'ai été contente de partager un poste d'ATER avec toi. Je te souhaite une très belle réussite pour ta thèse.

Je remercie toute ma famille, mes parents, qui m'ont toujours encouragée et accompagnée dans mes choix ainsi que mes sœurs Catherine, Delphine et Anne-Stéphanie, mes beaux-frères Fernand, Christian et Flavien, ma nièce Camille et mes neveux, Pierre, Nathan et Valentin. Un grand merci à mon petit lutin.

Je remercie de tout mon cœur Frédéric qui a toujours été là pour moi et qui m'a toujours soutenue. Ta présence m'est extrêmement précieuse. Sans toi, je ne serais jamais parvenue à réaliser ce parcours.

Table des matières

ETAT DE L'ART	1
Introduction	3
1. Les levures comme modèles en génétique et en génomique	5
1.1. La levure <i>Saccharomyces cerevisiae</i>	7
1.2. La levure <i>Lachancea kluyveri</i>	8
2. L'étude de la diversité génétique.	9
2.1. La diversité interspécifique ou génomique comparative	9
2.2. La diversité intraspécifique ou génomique des populations	11
2.2.1. Les variations génétiques	11
Le polymorphisme nucléotidique	11
Les insertions et les délétions (indels)	12
La variation de nombre de copies (CNV, copy number variation)	12
Les réarrangements chromosomiques	13
Les aneuploïdies	13
Les mécanismes épigénétiques	14
2.2.2. La diversité génétique au sein de l'espèce <i>Saccharomyces cerevisiae</i>	14
2.2.3. La diversité génétique au sein de l'espèce <i>Lachancea kluyveri</i>	16
La diversité des génomes mitochondriaux	16
La diversité du génome nucléaire	18
3. L'étude de la diversité phénotypique entre souches d'une même espèce	19
3.1. Une multitude de phénotypes d'intérêt	19
3.2. Les analyses à haut débit	20
L'analyse de la croissance	21
L'analyse de la morphologie des cellules	22
4. L'étude de la relation génotype-phénotype	23
4.1. Les analyses de liaison	24
4.1.1. Le principe de la méthode	24
4.1.2. Quelques exemples d'analyse de liaison	26
4.1.3. Les analyses par pool de ségréants	27
4.1.4. Quelques exemples d'analyse par pool de ségréants	29
4.2. L'analyse d'association	30
4.2.1. Le principe de la méthode	30
4.2.2. Les études d'association pangénomiques	31
4.3. L'identification et la validation fonctionnelle des gènes candidats	31
4.3.1. Le remplacement allélique	32
4.3.2. Le test d'hémizygocité réciproque	33
Références	35
VUE D'ENSEMBLE DU PROJET	45

Projet-----	47
Publications relatives à ce travail -----	50
CHAPTER 1-----	51
The hidden complexity of Mendelian traits across yeast natural populations-----	51
Abstract -----	53
Introduction -----	54
Results -----	57
Global landscape of Mendelian traits in <i>S. cerevisiae</i> . -----	57
Molecular characterization of identified Mendelian traits-----	60
Fitness distribution of identified Mendelian traits across large natural populations -----	63
Hidden complexity of a rare Mendelian variant across different genetic backgrounds -----	64
Discussion -----	68
Biased genomic distribution of causal Mendelian variants in yeast-----	68
Stability of Mendelian inheritance and the functional nature of causal variant -----	68
From Mendelian to complex, a continuum -----	69
References -----	71
CHAPTER 2-----	75
Large-scale survey of intraspecific fitness and cell morphology variation in a protoploid yeast species -----	75
Abstract -----	77
Introduction -----	78
Results -----	80
Fitness variation within a natural population of <i>L. kluyveri</i> -----	80
Cell morphology variation within a natural population of <i>L. kluyveri</i> -----	86
Trait variation and population history -----	88
Relationship between genetic and phenotypic diversity-----	91
Comparison of trait variation across different yeast species-----	93
Discussion -----	95
References -----	97
CHAPTER 3-----	99
Dissection of quantitative traits by bulk segregant mapping in a protoploid yeast species -----	99
Abstract -----	101
Introduction -----	102
Results -----	103
Description of the developed strategy -----	103

Fluorescence-based isolation of a large descendance-----	104
Effect of the fluorescent marker and <i>CHS3</i> deletion on growth phenotype -----	106
Fluorescence-based sorting of <i>MATa</i> progeny -----	107
Bulk segregant analysis under selective pressure -----	108
Functional validation of <i>ENA2</i> gene involved in resistance to osmotic stress-----	111
Conclusion -----	112
References -----	113
MATERIEL ET METHODES -----	115
1. Les organismes de l'étude -----	117
1.1. La levure <i>Saccharomyces cerevisiae</i> -----	117
1.1.1. Généralités sur le cycle de vie de <i>Saccharomyces cerevisiae</i> -----	117
1.1.2. Souches de <i>Saccharomyces cerevisiae</i> étudiées -----	117
1.2. La levure <i>Lachancea kluyveri</i> -----	117
1.2.1. Généralités sur le cycle de vie de <i>Lachancea kluyveri</i> -----	117
1.2.2. Souches étudiées de <i>Lachancea kluyveri</i> -----	118
2. Les méthodes de culture-----	119
2.1. Milieux de culture -----	119
2.2. Conditions de culture-----	119
2.3. Evaluation de la concentration cellulaire -----	119
2.4. Conservation des souches -----	119
3. Les techniques de génétique de la levure-----	120
3.1. Obtention de cellules diploïdes -----	120
3.2. Obtention de cellules haploïdes et analyse de la viabilité -----	120
4. La détermination des phénotypes-----	120
4.1. Analyse de la croissance en microcultures-----	120
4.1.1. Stratégie de phénotypage-----	120
4.1.2. Analyse des courbes de croissance -----	121
4.2. Analyse de la croissance en milieu solide-----	122
4.2.1. Stratégie de phénotypage-----	122
4.2.2. Traitement et analyse des images -----	123
4.3. Test en gouttes -----	123
4.4. Analyse de la morphologie des cellules -----	123
4.4.1. Culture et fixation des cellules-----	123
4.4.2. Marquage des cellules -----	123
4.4.3. Observation et acquisition des images de microscopie à fluorescence -----	124
4.4.4. Traitement des images -----	124
5. Préparation de l'ADN -----	125
5.1. Extraction d'ADN génomique par la méthode phénol chloroforme -----	125

5.2. Extraction d'ADN génomique total par kit Qiagen -----	125
5.3. Purification d'ADN plasmidique -----	126
5.4. Extraction des ARN totaux -----	126
5.5. Contrôle et dosage des acides nucléiques -----	126
5.5.1. Analyse des acides nucléiques par électrophorèse sur gel d'agarose -----	126
5.5.2. Quantification des acides nucléiques -----	127
6. Amplification d'un fragment d'ADN par PCR -----	127
6.1. La réaction de PCR -----	127
6.1.1. Principe de la PCR -----	127
6.1.2. Milieux de PCR -----	128
6.1.3. Purification de fragments PCR -----	128
6.2. La réaction de RT-PCR -----	128
7. Système de clonage Gateway® -----	129
7.1. Principe de la technologie Gateway® -----	129
7.2. Réaction de recombinaison BP -----	129
7.3. Réaction de recombinaison LR -----	130
7.4. Cassette Gateway® -----	130
7.4.1. Sélection négative des clones d'entrée et d'expression -----	130
7.4.2. Sélection positive des clones d'entrée et d'expression -----	131
8. Techniques de transformation des levures -----	131
8.1. Transformation par électroporation -----	131
8.1.1. Protocole de transformation de <i>Saccharomyces cerevisiae</i> -----	131
Préparation de cellules électro-compétentes -----	131
Electroporation -----	132
8.1.2. Protocole de transformation de <i>Lachancea kluyveri</i> ¹⁷ -----	132
Préparation de cellules électro-compétentes -----	132
Electroporation -----	132
8.2. Transformation par choc thermique -----	133
9. Observation de cellules marquées par microscopie à fluorescence -----	133
10. Analyse des cellules par cytométrie en flux -----	133
10.1. Principe de la cytométrie en flux -----	133
10.2. Contrôle de la ploïdie des souches -----	134
10.2.1. Préparation des cellules -----	134
10.2.2. Mesure semi-quantitative de la fluorescence -----	134
10.3. Tri cellulaire -----	135
10.3.1. Préparation des cellules (adapté de Keller et Knop, 2009) ¹⁸ -----	135
10.3.2. Tri des cellules de signe <i>MATa</i> -----	135
10.3.3. Enumération des cellules triées -----	136
10.3.4. Enrichissement des cellules triées -----	136

11. Séquençage	136
12. Bioinformatique	137
12.1. Les banques de données biologiques	137
12.2. Analyse des données de séquençage de l'ADN génomique	138
12.2.1. Nettoyage des séquences	138
12.2.2. Alignement des séquences	138
12.2.3. Détermination des SNP	138
12.2.4. Analyse des couvertures	138
12.2.5. Représentations graphiques des données de séquençage	139
12.3. Méthode d'ajustement du modèle pour les caractères quantitatifs à distribution bimodale	139
12.4. Analyses statistiques des données de croissance et de morphologie cellulaire	142
CONCLUSION ET PERSPECTIVES	149
Conclusion et perspectives	151
Origines et complexité génétique de la diversité phénotypique intra-spécifique	151
Evolution de la relation génotype-phénotype chez la levure	154
Références	156
ANNEXES	159
Liste des publications	161
Liste des communications	163
Activité d'enseignement	165

ETAT DE L'ART

Introduction

L'élucidation de la relation entre génotypes et phénotypes est un défi majeur de la génétique moderne. Chez l'Homme, la plupart des maladies génétiques (dégénérescence maculaire liée à l'âge, maladie de Crohn, lupus érythémateux systémique par exemple) sont déterminées par la combinaison de plusieurs gènes et par des facteurs environnementaux¹⁻³. La valeur phénotypique (P) d'un individu dépend donc de sa valeur génétique (G) et de son environnement (E) et par conséquent $P = G + E$. La composante génétique (G) se décompose elle-même en trois types d'effets génétiques : des effets additifs (A), des effets d'interaction entre gènes (interactions épistatiques, I) et entre allèles (dominance, D) et donc $G = A + I + D$. On parle de caractères multigéniques ou complexes.

A l'heure actuelle, les études d'association génotype-phénotype pangénomiques chez l'Homme ne permettent de détecter qu'une minorité des allèles impliqués dans des maladies correspondant à des traits complexes⁴. Les variants génétiques identifiés n'expliquent en effet qu'une infime fraction de l'héritabilité à savoir la part de variance phénotypique expliquée par des facteurs génétiques⁵. Plusieurs facteurs peuvent potentiellement expliquer l'héritabilité manquante dont l'existence d'allèles non identifiés, soit communs avec des effets petits à modestes soit rares à forts effets, la variation du nombre de copies de gènes, des interactions génétiques ou encore des facteurs non héréditaires (environnementaux, épigénétiques)⁶⁻⁸. Par exemple, les loci identifiés dans la dégénérescence maculaire liée à l'âge n'expliquent que 50% de l'héritabilité totale estimée et seulement 20% et 15% dans le cas de la maladie de Crohn et du lupus érythémateux systémique⁸, respectivement. A ces facteurs s'ajoute un manque de puissance statistique pour détecter les allèles rares, en partie dû à des populations de tailles trop réduites. Les études d'association génotype-phénotype restent donc à ce jour insatisfaisantes en raison de l'incapacité à déterminer complètement les bases génétiques de la plupart des maladies génétiques humaines pour des applications cliniques à savoir la détermination des facteurs de risque, l'établissement d'un diagnostic génétique ou encore la mise en place d'un traitement et de potentiels outils thérapeutiques.

De l'autre côté du spectre, la détermination de l'origine génétique de caractères Mendéliens gouvernés par l'expression d'un seul gène (mucoviscidose, myopathie de Duchenne, galactosémie par exemple) est plus facile à déterminer et disséquer. Le séquençage de génomes entiers ou d'exomes au sein de grandes populations d'individus permettent en effet la découverte efficace et précise des gènes pour plusieurs maladies^{3,9}. Par ailleurs, l'étude des formes monogéniques de

maladies génétiques complexes s'avère utile pour révéler des variants génétiques rares et évaluer leur contribution à l'héritabilité des maladies à caractères plus complexes. De récentes études ont en effet permis de révéler des allèles rares, par exemple associés à l'obésité sévère⁹ ou encore des allèles prédisposant à deux des facteurs de risques impliqués dans les maladies cardiovasculaires à savoir l'hypercholestérolémie et l'hypertriglycéridémie^{10,11}. Ces dernières stratégies semblent être par conséquent la meilleure marche à suivre à l'heure actuelle pour augmenter le répertoire des gènes et variants associés aux maladies génétiques à caractères complexes.

Les études des caractères complexes chez l'homme, bien qu'extrêmement importantes, sont entravées non seulement par la complexité génétique, la pléiotropie et les interactions gène-environnement, mais également par la taille conséquente des génomes. Afin d'avoir une meilleure compréhension des caractères complexes, l'utilisation d'organismes modèles tels que les levures est essentielle. Dans l'équipe « variation intraspécifique et évolution des génomes », au laboratoire de Génétique Moléculaire, Génomique et Microbiologie, la relation génotype-phénotype est étudiée au sein de plusieurs espèces de levures dont *Saccharomyces cerevisiae* et *Lachancea kluyveri*. L'étude de cette relation est réalisée à travers plusieurs phénotypes d'intérêt médical, environnemental ou évolutif tels que l'isolement reproductif ou encore la tolérance aux stress environnementaux. Le but de nos recherches est de mieux comprendre les règles qui gouvernent la relation génotype-phénotype. Pour ce faire, nous combinons des méthodes de séquençage et de phénotypage à haut débit. La première étape consiste à obtenir des mesures quantitatives pour l'ensemble des phénotypes d'intérêt afin de décrire les variations phénotypiques des différentes souches. La deuxième étape repose sur la cartographie des gènes responsables de la variation phénotypique observée. Dans le cadre de ma thèse, je me suis focalisée sur l'analyse de la variation de tolérance au stress ainsi que leurs origines génétiques au sein de deux espèces de levure à savoir *S. cerevisiae* et *L. kluyveri*.

1. Les levures comme modèles en génétique et en génomique

Les levures sont des champignons unicellulaires apparus à partir de deux embranchements distincts du règne des Mycètes : les Ascomycètes et les Basidiomycètes. Les levures du sous-phylum des Saccharomycotina anciennement appelées Hémiascomycètes, dont font partie *S. cerevisiae* et *L. kluyveri*, constituent l'une des trois classes des Ascomycètes. A ce jour près d'un millier d'espèces Saccharomycotina au mode de vie très divers ont été décrites. Le sous-phylum comprend entre autres des pathogènes opportunistes de l'Homme dont la plupart font partie du genre *Candida*, des pathogènes de plantes principalement du genre *Eremothecium*, la levure de boulangerie *S. cerevisiae* ou encore des levures ayant des propriétés physiologiques notables, des levures osmotolérantes dont *Pichia sorbitophila*, des levures méthylophiles dont *Pichia pastoris* ou encore des levures assimilant des sources de carbone et d'azote moins communes telles que *Kluyveromyces lactis*, *Yarrowia lipolytica* et *L. kluyveri*.

La structure des génomes des levures Saccharomycotina, de petite taille (< 21 Mb) et compacts, permet un séquençage relativement peu coûteux et une analyse aisée. Ces deux dernières décennies le génome de plus de soixante quinze espèces a ainsi été partiellement ou entièrement séquencé¹². L'analyse comparative de ces génomes a révélé l'importance des réarrangements chromosomiques au cours de l'évolution avec notamment la duplication totale du génome survenue dans l'histoire évolutive des levures Saccharomycotina il y a environ 100 millions d'années^{13,14} (Figure 1). Les espèces apparues après la duplication totale du génome sont dites post-dupliquées par opposition aux levures dites pré-dupliquées dont le génome n'a pas subi cet événement majeur. Les levures post-dupliquées regroupent le genre *Saccharomyces* et possèdent 10 à 16 chromosomes. Le genre des *Saccharomyces sensu stricto* auquel appartient *S. cerevisiae* possède un nombre de chromosomes doublé (16 chromosomes) par rapport aux levures pré-dupliquées. Les autres levures post-dupliquées se distinguent par une réduction du nombre de chromosomes¹⁵. *L. kluyveri* a quant à elle divergé de la lignée des *Saccharomyces* avant la duplication totale du génome, elle est dite pré-dupliquée ou protoploïde.

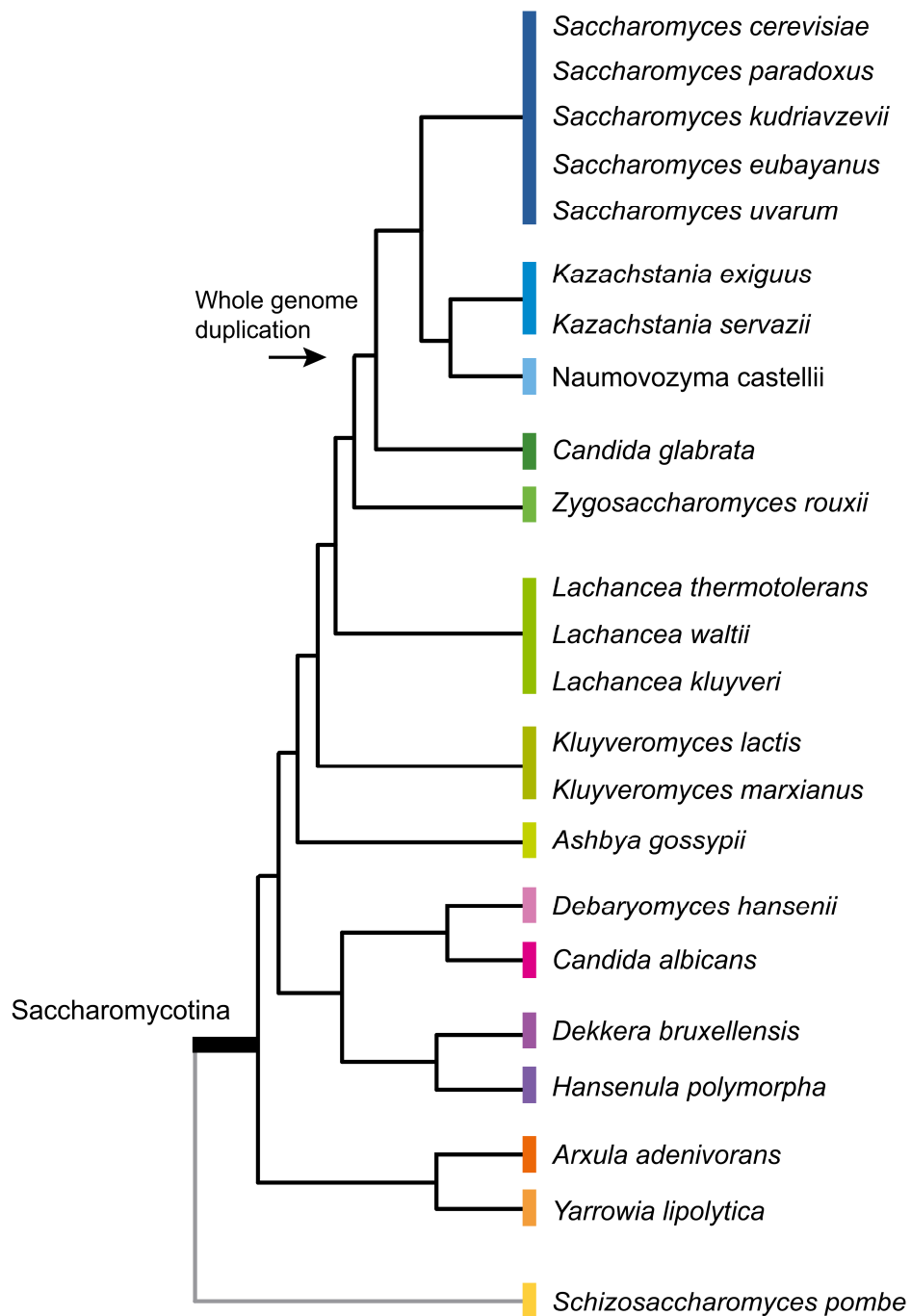


Figure 1 : Arbre phylogénétique des levures du sous-phylum des Saccharomycotina (adaptée de Peter et Schacherer, 2016)¹⁶. Cette figure représente un aperçu des espèces dont le génome a été séquencé. Les différents clades sont représentés par une couleur. L'événement de duplication totale du génome est indiqué par une flèche.

1.1. La levure *Saccharomyces cerevisiae*

Depuis les temps anciens, la levure *S. cerevisiae* est utilisée par l'Homme dans divers processus de fermentation alcoolique comme la panification, la vinification ou le brassage de la bière. La première preuve archéologique de l'utilisation des levures pour la fermentation d'une boisson alcoolisée à base d'un mélange de riz, de miel et de fruits en Chine est datée de 7000 ans av. J.-C.¹⁷. L'analyse d'anciennes jarres datant de l'Égypte Antique prouve leur utilisation dans la fabrication du vin dès 3150 av. J.-C.¹⁸. *S. cerevisiae* n'est cependant pas une espèce domestiquée dans son ensemble, elle occupe également de nombreux habitats naturels comme les fruits, le sol et les exsudats d'arbres. Elle peut par ailleurs se révéler être un pathogène opportuniste de l'Homme en colonisant les muqueuses de patients immunodéprimés¹⁹⁻²².

De par son caractère unicellulaire et sa facilité de manipulation, la levure *S. cerevisiae* s'est imposée dès le milieu du 20^{ème} siècle comme un organisme modèle en génétique puis en biologie cellulaire et moléculaire. Elle représente le modèle eucaryote le plus simple pour étudier les principaux mécanismes cellulaires qui sont conservés à travers les eucaryotes. Elle combine également les avantages des micro-organismes. *S. cerevisiae* est facilement cultivable en laboratoire sous différentes conditions, a un temps de génération court et peut être croisée afin d'étudier la descendance. Elle se prête aisément à de nombreuses techniques de biologie moléculaire et de génétique.

En 1996, l'espèce *S. cerevisiae* et plus précisément la souche haploïde de laboratoire S288c fut le premier eucaryote dont le génome a été complètement séquencé²³. Son génome d'une taille de 12,1 Mb est compact et contient environ 5500 gènes répartis sur 16 chromosomes (nommés I à XVI). Depuis avec l'émergence des techniques de séquençage de nouvelle génération, les séquences complètes du génome de nombreuses souches au fonds génétique très divers sont disponibles dont la souche de la vigne RM11, très étudiée²⁴⁻²⁶, la souche EC1118, utilisée à travers le monde dans la vinification²⁷, la souche YJM789, isolée du poumon d'un patient atteint du syndrome d'immunodéficience acquise²⁸⁻³⁰ et la souche JAY291 impliquée dans la production de bioéthanol³¹. Du fait de ses origines variées et du nombre croissant de génomes séquencés, *S. cerevisiae* est utilisée pour l'étude de la diversité génétique intraspécifique, encore appelée la génomique des populations^{32,33}.

S. cerevisiae est rapidement devenue un organisme modèle en génomique fonctionnelle avec l'achèvement de la première collection de délétion systématique de gènes³⁴. Pour laquelle près de 96% des gènes composant le génome de *S. cerevisiae* ont été remplacés par la cassette de délétion *KanMX* conférant une résistance à la généticine. La délétion des gènes essentiels et non-essentiels

à la survie cellulaire a été réalisée respectivement dans un contexte diploïde hétérozygote et haploïde. Plus de 80% des gènes ont des fonctions caractérisées, soit un pourcentage supérieur à n'importe quel autre eucaryote³⁵. Plus récemment, *S. cerevisiae* s'est révélée être un organisme modèle privilégié pour le développement de méthodes à haut débit dans divers domaines : génomique, phénotomique, métabolomique, protéomique ou encore transcriptomique³⁶.

1.2. La levure *Lachancea kluyveri*

L. kluyveri (anciennement appelée *Saccharomyces kluyveri*) a été décrite pour la première fois en 1956^{37,38}. Il s'agit d'une levure protoploïde c'est-à-dire qu'elle n'a pas d'ancêtre ayant subi une duplication totale du génome contrairement à *S. cerevisiae*^{38,14}. Cette espèce se retrouve dans une grande partie du monde dans des habitats d'origine naturelle dont les exsudats d'arbre, l'intestin de drosophiles et le sol. Elle partage le même cycle de vie que *S. cerevisiae*. Cependant elle a l'avantage d'être une levure hétérothallique ce qui signifie que les haploïdes *MATa* et *MAT α* sont stables et ne peuvent pas changer de signe sexuel^{39,40}. Les cassettes silencieuses du signe sexuel *HML* (Hidden *MAT* Left) et *HMR* (Hidden *MAT* Right) ainsi que le gène *HO* (Homothallic switching endonuclease) codant l'endonucléase responsable du changement de signe sexuel sont en effet absents du génome de *L. kluyveri*.

Cette levure possède un génome d'une taille de 11,3 Mb comprenant 5321 gènes répartis sur 8 chromosomes (nommés A à H)⁴¹. L'analyse de son génome a révélé une région de 1 Mb située sur le bras gauche du chromosome C qui se démarque du reste du génome en termes de composition nucléotidique. Son contenu en guanine (G) et en cytosine (C) atteint dans cette région 52,9% contre 40,4% sur le restant du génome. L'étude des relations phylogénétiques, de la conservation de la synténie et du contenu en gènes entre *L. kluyveri* et des espèces apparentées suggère que cette région pourrait être le résultat d'un événement d'hybridation entre deux espèces du clade des *Lachancea*⁴⁰.

Par ailleurs, *L. kluyveri* présente un intérêt biologique en raison du caractère singulier de son métabolisme. À l'opposé de *S. cerevisiae*, elle ne fermente pas les sucres comme le glucose en présence d'oxygène⁴². Elle se différencie de la plupart des levures du groupe des *Saccharomycetaceae* par sa capacité à utiliser les pyrimidines et ses dérivés comme unique source d'azote⁴³⁻⁴⁵. Récemment une nouvelle voie de dégradation des pyrimidines a été partiellement décrite chez *L. kluyveri*⁴⁶. Son étude est facilitée par l'applicabilité de nombreuses techniques de génétique de la levure comme la transformation par des vecteurs plasmidiques, le remplacement allélique, la séparation des chromosomes par électrophorèse en champs pulsés et l'obtention de descendants haploïdes^{47,44,48}.

2. L'étude de la diversité génétique.

Le développement des nouvelles techniques de séquençage de ces dernières années a permis de donner une nouvelle dimension à l'exploration des génomes. Ces avancées technologiques permettent désormais d'étudier de manière plus exhaustive la diversité génétique interspécifique, c'est-à-dire les variations génétiques entre isolats appartenant à des espèces différentes (interspécifique) et la diversité génétique intraspécifique entre souches d'une même espèce.

2.1. La diversité interspécifique ou génomique comparative

Le séquençage systématique de génomes complets est essentiel pour étudier l'évolution des génomes, un domaine où les levures Saccharomycotina sont des organismes de choix¹². De par les caractéristiques de leur génome, petit et compact, de nombreuses souches appartenant à des espèces différentes de levures Saccharomycotina ont été séquencées et mises à la disposition de la communauté scientifique⁴⁹. Un projet de séquençage d'environ 1000 espèces supplémentaires de levures Saccharomycotina a récemment été initié (Projet Y1000+, <http://www.y1000plus.org/>). A ce jour, le sous-phylum des Saccharomycotina compte le plus grand nombre d'espèces dont le génome a été séquencé. Ces levures ont ainsi permis d'initier des études de la diversité génétique interspécifique encore appelée génomique comparative⁵⁰⁻⁵². Les différences observées entre les espèces au niveau moléculaire ont révélé une distance évolutive comparable à celle de l'embranchement des Chordés regroupant pourtant des espèces très diverses ne serait-ce qu'au niveau de leur morphologie⁵².

L'importante échelle évolutive couverte a permis de mettre en évidence des mécanismes d'évolution majeurs. Le plus remarquable est l'évènement de duplication totale du génome survenu dans l'histoire évolutive des levures Saccharomycotina il y a environ 100 millions d'années vraisemblablement par allopolyploidisation⁵³. Suite au séquençage du génome de *S. cerevisiae* en 1996, la mise en évidence de blocs de duplication a permis d'émettre l'hypothèse que son génome ait évolué à partir d'un ancêtre commun dont le génome a été dupliqué¹³. Par la suite la comparaison avec le génome des levures pré-dupliquées *Asbyya gossypii* et *Lachancea waltii* a permis de valider l'hypothèse d'une duplication totale du génome^{54,14}. Cet évènement majeur s'est produit juste après la séparation des genres *Zygosaccharomyces* et *Saccharomyces* puis s'est suivi d'une perte massive des gènes dupliqués. De ce fait, seules 551 paires de gènes ohnologues issues de cet évènement majeur subsistent encore dans le génome actuel de *S. cerevisiae*⁵⁵. Cette perte massive de gènes a conduit à une taille similaire du génome des levures pré- et post-dupliquées. En effet, la taille du génome de *S. cerevisiae* est seulement 13% supérieure à celle du génome de *L. waltii* et comprend 10% de gènes supplémentaires¹⁴.

L'analyse comparative des génomes a par ailleurs permis de décrire différents groupes parmi les Saccharomycotina dont les *Dipodascaceae*, le groupe des CTG et les *Saccharomycetaceae* (Figure 2).

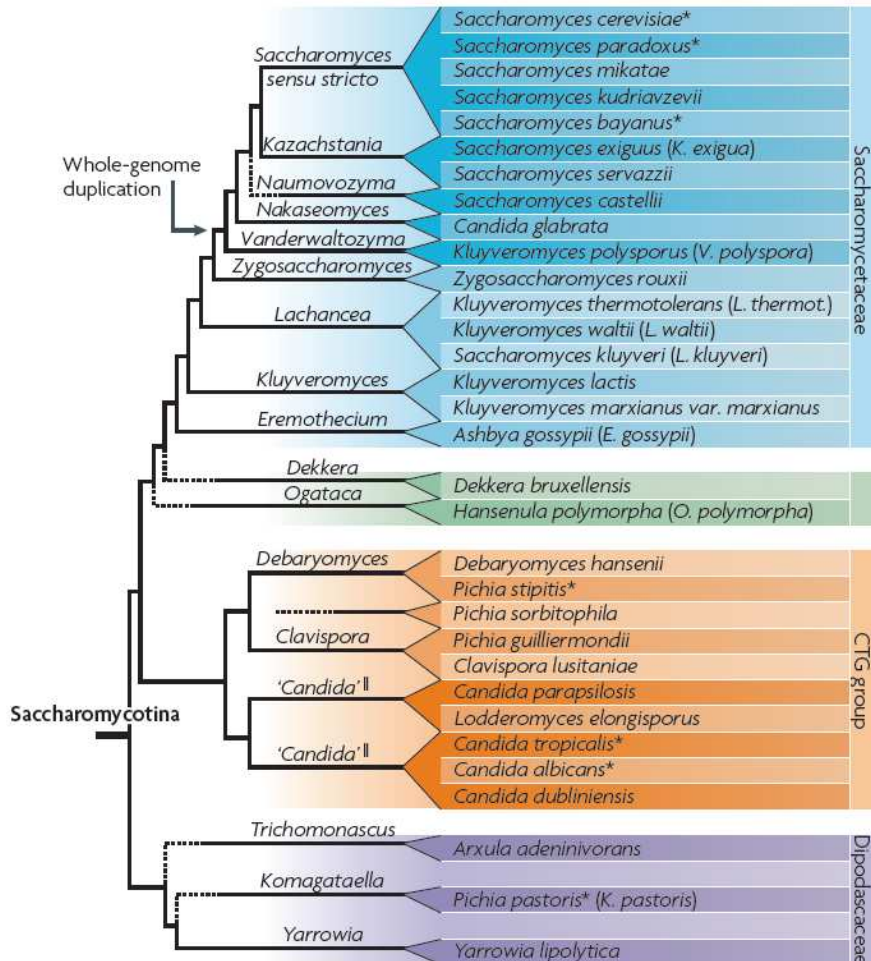


Figure 2 : Arbre phylogénétique des levures Saccharomycotina (adaptée de Dujon *et al.* 2010)¹². Les trois groupes majeurs sont chacun représentés par une couleur : en bleu les *Saccharomycetaceae*, en orange le groupe des CTG et en violet les *Dipodascaceae*. Chaque triangle regroupe les espèces d'un même clade.

Chaque groupe se distingue par des caractéristiques génomiques particulières telles qu'un faible nombre de chromosomes pour le groupe des Dipodascaceae, un changement du code génétique pour le groupe des CTG (le codon CUG code pour la sérine au lieu de la leucine) et des centromères ponctuels pour les *Saccharomycetaceae*. Certaines espèces résultent de différents mécanismes d'évolution comme des insertions en nombre variable dans le génome de divers éléments transposables, des hybridations intra- ou inter-espèce suivi d'une perte de l'hétérozygotie et des introgressions^{12,56}.

Bien que la génomique comparative apporte des enseignements précieux sur la dynamique et l'évolution des génomes, la séquence d'un génome unique n'est cependant pas représentative de la diversité d'une espèce. C'est pourquoi l'étude de la diversité génétique intraspécifique encore appelée génomique des populations connaît un grand essor depuis quelques années¹⁶. L'exploration de la diversité génétique a en effet été menée au sein de plusieurs espèces de levures dont *S. cerevisiae*, *S. paradoxus*, *Saccharomyces kudriavzevii*, *Saccharomyces uvarum*, *L. kluyveri*, *Candida albicans* et *Schizosaccharomyces pombe*^{32,33,57-64}.

2.2. La diversité intraspécifique ou génomique des populations

L'identification des variations génétiques entre isolats d'une même espèce est une étape indispensable pour déterminer les processus évolutifs intraspécifiques impliqués dans la variation de séquence mais également pour comprendre les différences de phénotypes observées au sein d'une espèce. L'analyse comparative des génomes révèle plusieurs types de variations génétiques pouvant varier en taille du nucléotide au chromosome entier (Figure 3).

2.2.1. Les variations génétiques

Le polymorphisme nucléotidique

Le polymorphisme nucléotidique ou polymorphisme d'un seul nucléotide (SNP, *Single Nucleotide Polymorphism*) est une mutation au niveau d'un seul nucléotide. Une telle mutation peut être provoquée par des agents mutagènes, des intermédiaires toxiques du métabolisme ou peut survenir suite à une erreur dans la réplication ou la réparation de l'ADN. La mutation peut avoir un impact sur le phénotype lorsqu'elle se situe dans une région codante ou promotrice. Si le polymorphisme nucléotidique se situe dans une région codante et qu'il entraîne un changement d'acide aminé dans la séquence protéique alors il est dit non-synonyme. Une telle modification peut affecter la conformation et la fonction biologique de la protéine. En revanche, s'il n'y a pas de changement d'acide aminé, le polymorphisme nucléotidique est dit synonyme. D'autre part, si le polymorphisme nucléotidique se situe dans une région promotrice alors il peut provoquer une modification de l'expression du gène. Chez *S. cerevisiae*, la densité du polymorphisme nucléotidique varie de 1 à 6 SNP/kb entre la souche de référence et les autres souches^{29,33} soit en moyenne de 2,8 SNP/kb³³. Chez *L. kluyveri*, le génome des souches est hautement polymorphique avec une densité moyenne de 28 SNP/kb dans les régions intergénomiques et de 17 SNP/kb dans les régions codantes⁵⁹.

Les insertions et les délétions (indels)

Les indels sont des insertions ou des délétions de taille limitée à quelques nucléotides pouvant se produire suite à une erreur dans la réplication de l'ADN, à des insertions d'éléments transposables ou à des répétitions de type microsatellites. Lorsque les insertions ou les délétions se situent dans une région codante et ne sont pas multiples de trois, il se produit un décalage du cadre de lecture. La séquence de la région codant la protéine est alors modifiée à partir du décalage pouvant entraîner la synthèse d'une protéine tronquée prématurément ou induire un changement conformationnel de la protéine. Cette dernière peut alors perdre sa fonction ou avoir une activité modifiée. Un total de 5 438 463 indels distribués sur 125 701 sites polymorphiques ont été détectés dans une collection de plus d'un millier de souches de *S. cerevisiae* (Projet 1002 génomes de levures : <http://1002genomes.u-strasbg.fr/>). Un total de 243 899 indels distribués sur 44 578 sites polymorphiques ont été détectés dans une collection de 28 souches de *L. kluyveri*⁶⁰. La majorité des indels (88%) chez cette espèce sont localisés dans des régions intergéniques contre seulement 12% (28 254 indels) dans des régions codantes.

La variation de nombre de copies (CNV, copy number variation)

Il s'agit d'une région du génome qui est présente à un nombre variable de copies à travers les individus d'une même espèce. L'analyse comparative du génome de 93 souches d'origines variées de *S. cerevisiae* a permis d'estimer précisément le nombre de copies des gènes *ENA* et *CUP1* connus pour être impliqués dans la résistance au chlorure de sodium (NaCl) et au cuivre respectivement⁶⁵. Le nombre de copies des gènes *ENA* basé sur la couverture de séquençage des génomes varie d'une à quatorze copies. Dans les souches multi-copies, la tolérance au NaCl et au lithium serait associée à une introgression, c'est-à-dire à un transfert d'une version allélique des gènes *ENA* de *Saccharomyces paradoxus*^{66,67}. L'amplification du nombre de copies de ces gènes a pour conséquence d'augmenter la quantité de transcrits et par conséquent de protéines conduisant à une résistance plus importante au NaCl. De manière similaire, le nombre de copies des gènes *CUP1* varie d'une à dix-huit copies. Une telle dynamique peut également être observée lors d'expériences d'évolution en laboratoire. Une expérience d'évolution initiée à partir d'une souche prototrophe de *S. cerevisiae* exposée à un milieu limité en sulfate pendant 200 générations a permis de détecter par puce d'hybridation génomique comparative (aCGH, array comparative genomic hybridization) une amplification du gène *SUL1* dans le génome des souches évoluées. Le nombre de copies de ce gène codant un transporteur de sulfate varie de trois à cinq copies. Cette étude a révélé que les copies surnuméraires du gène *SUL1* ont un effet bénéfique sur la croissance des souches évoluées par rapport à la souche ancestrale⁶⁸.

Les réarrangements chromosomiques

Les réarrangements chromosomiques affectent des segments de chromosomes qui peuvent être amplifiés, délétés, inversés ou transloqués. Ces événements surviennent suite à des mécanismes de réparation des cassures double brins de l'ADN soit par recombinaison homologue soit par recombinaison non homologue (NHEJ, *non homologous end joining*). Les réarrangements chromosomiques peuvent avoir un impact sur l'expression des gènes et leurs fonctions. Un exemple de réarrangement chromosomique présentant un avantage adaptatif chez *S. cerevisiae* est une translocation réciproque entre le chromosome XV et XVI conduisant à la permutation du promoteur du gène *SSU1* intervenant dans la résistance aux sulfites avec le promoteur du gène *ADH1*. Cette translocation permet d'augmenter l'expression du gène *SSU1* et par conséquent la capacité de résistance aux sulfites⁶⁹. Par ailleurs, les réarrangements chromosomiques ont une fréquence d'apparition élevée dans les expériences d'évolution. De tels mécanismes peuvent par conséquent permettre une adaptation rapide d'un isolat à son nouvel environnement. Une expérience d'évolution initiée sur une souche de laboratoire de *S. cerevisiae* exposée à un milieu carencé en glucose pendant 100 à 500 générations a permis d'identifier des réarrangements chromosomiques dans six des huit souches évoluées. Parmi ces mécanismes d'évolution, un même réarrangement chromosomique est apparu indépendamment dans trois souches évoluées soulignant son avantage sélectif⁷⁰. Cependant, les réarrangements chromosomiques ne confèrent pas nécessairement un avantage sélectif. Récemment, il a été mis en évidence l'existence de translocations réciproques à l'origine d'incompatibilités génétiques chez *S. cerevisiae*^{71,72}.

Les aneuploïdies

L'aneuploïdie désigne un nombre anormal de chromosomes dans la cellule, soit inférieur ou supérieur au lot normal de chromosomes. Elle peut survenir suite à une mauvaise ségrégation des chromosomes lors d'une division cellulaire ou à une erreur lors de la réplication de l'ADN. Il s'agit d'un mécanisme adaptatif fréquemment détecté dans les cellules en réponse à un stress environnemental ou à une délétion de gènes⁷³⁻⁷⁷. Un exemple est l'acquisition de la résistance au radical, un inhibiteur de la protéine de choc thermique (Hsp90), chez *S. cerevisiae*. La résistance à ce composé est en effet associée à une disomie du chromosome XV portant les gènes *STI1* et *PDR5* (codant respectivement une protéine chaperon de Hsp90 et un transporteur de drogues multiples) ayant pour conséquence un effet de dosage génique⁷⁶. Une étude récente d'évolution expérimentale suggère que l'aneuploïdie n'est pas une solution durable en raison de son fort coût énergétique pour la cellule. Elle représente un état transitoire d'adaptation rapide à un nouvel environnement avant l'émergence de solutions moins coûteuses pour la cellule⁷⁷.

Les mécanismes épigénétiques

Il s'agit de mécanismes qui régulent le niveau d'expression des gènes sans modification de la séquence de l'ADN comme la variation de la structure de la chromatine. Lorsque la chromatine est condensée (hétérochromatine), les gènes présents dans la région chromosomique concernée ne sont pas accessibles à la machinerie transcriptionnelle et par conséquent ne sont pas exprimés. En revanche, si la chromatine est décondensée (euchromatine), les gènes sont accessibles et sont donc exprimés. L'état de la chromatine dépend notamment de l'état post-traductionnel des protéines d'histones. Chez *S. cerevisiae*, il a été mis en évidence un fort polymorphisme des positions des modifications post-traductionnelles des histones entre la souche BY (isogénique à la souche de référence S288c) et la souche de la vigne RM11. Par exemple, dans le cas de l'acétylation de la lysine H14 de l'histone H3, 40% de la chromatine présente un polymorphisme entre ces deux souches⁷⁸.

2.2.2. La diversité génétique au sein de l'espèce *Saccharomyces cerevisiae*

L'exploration de la variabilité génétique présente au sein de l'espèce *S. cerevisiae* a été réalisée grâce à des analyses comparatives des génomes dans le but de déterminer la structure de la population et d'élucider son histoire évolutive. L'étude du polymorphisme de séquence des souches s'est initialement concentrée sur des parties restreintes du génome en utilisant différents marqueurs moléculaires^{19,79-84} avant de s'étendre au génome entier par des techniques d'hybridation du génome sur puces à ADN et de séquençage^{32,33,85,86}. Plus récemment, les génomes de plus grandes collections de souches ont été séquencés et représentent à ce jour les jeux de données les plus complets pour cette espèce^{65,87-89} (Projet 1002 génomes de levures : <http://1002genomes.u-strasbg.fr/>). L'ensemble de ces données révèle une diversité génétique bien supérieure à celle observée chez l'Homme avec une diversité maximale de 0,9%.

En 2009, l'étude de la diversité génétique intra-spécifique a été entreprise chez *S. cerevisiae*. Deux approches différentes ont été employées pour l'exploration de la diversité de ces génomes. Dans le cadre d'une première étude, les variations de séquences ont été déterminées par hybridation de l'ADN génomique sur des puces à ADN à haute densité (Yeast Tilling Microarray)³³. Une collection de 63 souches de *S. cerevisiae* d'origines écologique et géographique différentes a été analysée (Figure 3). Dans le cadre d'une deuxième étude, une collection de 36 souches de *S. cerevisiae* a été séquencée à faible couverture (la couverture allant de 1 à 3 X) par une stratégie de séquençage classique de type Sanger³².

Dans la 1^{ère} étude³³, un total de 1 299 811 SNP distribués sur 101 343 sites polymorphiques ont été détectés entre les souches de la collection et la souche de référence S288c soit en moyenne 30 097 SNP (2,8 SNP par kb) en excluant les souches de laboratoire qui sont très proches de la souche de référence. L'estimation de la fréquence et de la distribution des variations de séquence d'ADN a permis de mettre en évidence des régions conservées et des régions hautement polymorphiques au sein de l'espèce *S. cerevisiae*. Cependant les cartes de polymorphisme génétique établies n'informent pas sur la nature des changements nucléotidiques mais uniquement sur leur localisation sur le génome. Cette étude a également permis de déterminer la proximité ou non en termes de séquence des différents génomes analysés et de déterminer des sous populations. L'analyse phylogénétique des souches basée sur les 101 343 sites polymorphiques révèle très nettement un regroupement des souches de la vigne, du saké et de laboratoire au sein de clusters (Figure 3).

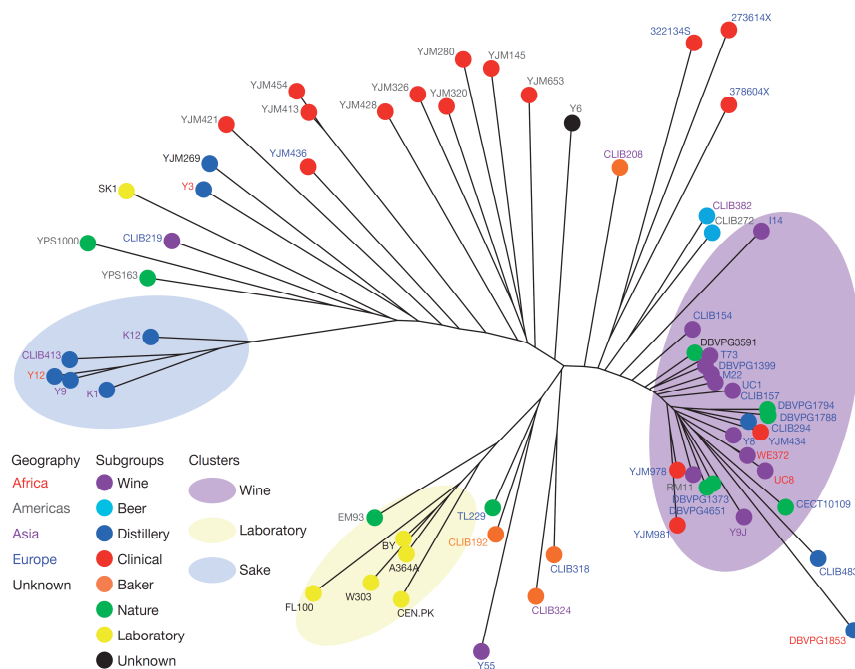


Figure 3 : Arbre de type Neighbour-joining construit sur la base des sites polymorphiques identifiés dans la collection de 63 souches de *S. cerevisiae* (d'après Schacherer *et al.* 2009)³³.

Par conséquent, les souches les plus proches proviennent d'une même origine écologique. La proximité de ces dernières peut s'expliquer simplement par une origine commune : la souche de référence S288c séquencée en 1996^{23,85} et peut également traduire un processus de domestication^{19,89}.

Dans la 2^{ème} étude³², l'analyse comparative des génomes a permis de détecter un total de 235 127 SNP et 14 051 indel chez *S. cerevisiae*. L'analyse phylogénétique des souches basée sur les différences de SNP montre une proximité de séquence des génomes en fonction de l'origine géographique pour les souches d'origine naturelle d'Amérique du Nord, de Malaisie et d'Afrique de l'Ouest mais également en fonction de l'origine écologique pour les souches industrielles isolées à partir du saké et de la vigne. L'analyse de la structure de la population indique que les sous populations de Malaisie, d'Afrique de l'Ouest, du saké, d'Amérique du Nord et de la vigne/d'Europe présentent une homogénéité génétique importante, formant 5 lignées dites pures. Les génomes des autres lignées sont dits mosaïques car ils présentent une importante hétérogénéité génétique. La présence de souches mosaïques au sein de cette espèce suggère un échange important de matériel génétique entre différentes souches issues de diverses niches probablement favorisé par la domestication de la levure par l'Homme.

En 2013, l'exploration de la diversité génétique intraspécifique a été étendue à une plus large collection de souches de *S. cerevisiae*⁸⁷. Au total 262 souches d'origines écologiques et géographiques variées ont été génotypées par une stratégie de séquençage de régions d'ADN associées à des sites de restriction (RAD-seq, Restriction site-Associated sequencing). Par cette technique une même fraction du génome des souches a été séquencée soit environ 1% correspondant à 116 880 pb. L'alignement multiple des séquences générées a permis de détecter un total de 5 868 sites polymorphiques. L'analyse phylogénétique des souches et de la structure de la population montre globalement un regroupement des souches selon l'origine géographique et confirme l'importance des lignées mosaïques et l'expansion de ces dernières à travers l'activité humaine.

2.2.3. La diversité génétique au sein de l'espèce *Lachancea kluyveri*

L'étude de la variabilité génétique intraspécifique présente au sein de l'espèce *L. kluyveri* a été réalisée grâce à des analyses comparatives des génomes mitochondriaux⁹⁰ et nucléaires⁶⁰ suite à une stratégie de séquençage à haut débit (Illumina HiSeq 2000).

La diversité des génomes mitochondriaux

La diversité intraspécifique des génomes mitochondriaux a été décrite chez *L. kluyveri* en 2012 soit pour la première fois au sein d'une espèce appartenant aux levures Saccharomycotina⁹⁰. Le génome mitochondrial de 18 souches de *L. kluyveri* d'origines écologiques et géographiques variées a été analysé suite à un projet de reséquençage. Le génome mitochondrial de référence a une taille de 51 525 pb soit plus de deux fois la taille de celui de l'espèce la plus proche

phylogénétiquement, *Lachancea thermotolerans*⁹¹. Malgré la différence de taille, ce génome comprend le même ensemble de 35 gènes que celui décrit chez *L. thermotolerans* parmi lesquels 8 gènes codent pour des protéines et 27 gènes pour des ARN non codants. Les gènes codants pour des protéines codent pour trois sous-unités de la cytochrome c oxydase (*COX1* à *3*), trois sous-unités de l'ATP synthase (*ATP6*, *ATP8* et *ATP9*), l'apocytochrome b (*COB*) et une protéine ribosomale (*VAR1*). Les autres gènes codent pour les deux sous-unités ribosomiques (*SSU* et *LSU*), la sous-unité de la ribonucléase P (*RPM1*) et des ARN de transfert. L'ordre des 35 gènes sur le génome est globalement bien conservé entre les deux espèces apparentées, à l'exception de deux gènes.

La comparaison du génome mitochondrial de référence avec celui de quatre autres souches montre une conservation de la synténie des 35 gènes et par conséquent une absence de réarrangements au niveau des génomes mitochondriaux chez *L. kluveri*. Le taux de polymorphisme au niveau des régions codantes est variable allant de 0 à 13,96 SNP/kb. Parmi les 1101 SNP identifiés dans les régions codantes, 930 sont synonymes et 171 sont non-synonymes ce qui indique que ces régions sont conservées. Les mutations synonymes sont bien plus répandues que les mutations non synonymes, vraisemblablement en raison de l'élimination de nombreuses substitutions délétères au niveau des gènes par sélection négative. Bien que les séquences des régions codantes soient hautement conservées parmi les souches de *L. kluveri* (en moyenne de 5,2 SNP/kb et de 0,19 indel/kb), les régions intergéniques sont bien plus variables (en moyenne de 37,2 SNP/kb et de 9,8 indels/kb). La construction d'un arbre phylogénétique de type Neighbor-joining basée sur la concaténation des alignements de séquences des gènes *ATP6-ATP8-ATP9-COX1-COX2-COX3-COB* indique une proximité de séquence des génomes mitochondriaux spécifique à la localisation géographique. L'analyse de la structure de la population va également dans ce sens. Les souches les plus proches génétiquement ont par conséquent une même origine géographique. Une certaine variation est également observée dans le contenu en introns et la longueur des régions intergéniques où un fort polymorphisme entre les souches est détecté. La variation de ces deux paramètres permet d'expliquer la différence de taille des génomes mitochondriaux entre les souches et par conséquent révèle une forte plasticité de ces génomes. Par ailleurs, le contenu en introns est corrélé à la phylogénie des souches. Le contenu en introns est donc similaire entre les souches provenant d'une même origine géographique.

Une telle étude a ainsi permis de retracer l'histoire évolutive de l'espèce *L. kluveri* à travers la comparaison des génomes mitochondriaux très conservés au niveau des régions codantes et bien plus variables au niveau du contenu en introns et des séquences intergéniques.

La diversité du génome nucléaire

Le génome nucléaire d'une collection de 28 souches de *L. kluyveri* représentative de la quasi totalité des souches naturelles disponibles à ce jour a été analysé suite à un projet de reséquençage⁶⁰. L'analyse comparative de ces génomes indique un polymorphisme remarquablement élevé avec un total de 6 515 704 SNP distribués sur 881 427 sites polymorphiques soit 5 fois supérieurs à chez *S. cerevisiae*³³. Les régions codantes comprennent en moyenne 17 SNP/kb dont les deux tiers sont des SNP synonymes contre 28 SNP/kb dans les régions intergéniques. L'analyse phylogénétique des souches et de la structure de la population basée sur les 881 427 sites polymorphiques montre une proximité de séquences des génomes en fonction de l'origine géographique ce qui concorde avec les résultats de l'analyse des génomes mitochondriaux (Figure 4).

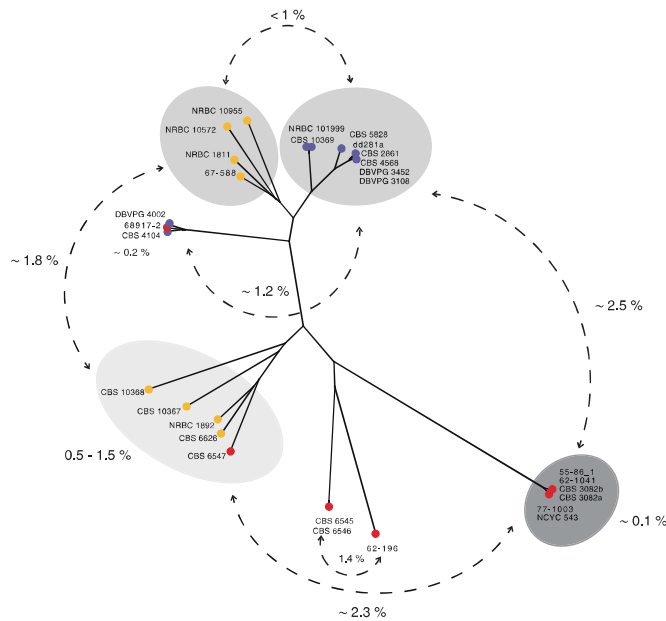


Figure 4: Arbre de type Neighbour-joining construit sur la base des 881 427 sites polymorphiques identifiés au sein de la collection de 28 souches de *L. kluyveri* (d'après Friedrich *et al.* 2015)⁶⁰. La longueur des branches est proportionnelle au nombre de sites discriminants chaque paire de souches. Les flèches en pointillés désignent les polymorphismes inter et intra groupes. L'origine géographique des souches est représentée par un code couleur, en jaune les souches d'Asie, en bleu les souches d'Europe, en rouge les souches d'Amérique.

Par ailleurs, l'analyse du contenu en GC le long des génomes a permis de confirmer la présence d'une région GC riche de 1 Mb située sur le bras gauche du chromosome C dans l'ensemble des génomes⁴⁰. Cette région comprend le locus *MAT* qui contrôle le signe sexuel et qui a conservé une composition nucléotidique normale. L'arbre phylogénétique des souches basé sur les sites polymorphiques localisés dans cette région GC riche présente la même topologie que la phylogénie des souches basée sur l'ensemble du génome. Ceci indique que cette région était bien présente dans le génome du dernier ancêtre commun de l'espèce et que celle-ci a évolué de la même manière que le restant du génome. Cette étude suggère que cette région a été acquise suite à un évènement d'introgession depuis une espèce proche phylogénétiquement de *L. kluyveri* dont

le génome devait avoir une synténie conservée, un fort contenu en GC et être dépourvu des cassettes du signe sexuel. En effet, les cassettes silencieuses du signe sexuel (*HML* et *HMR*) sont absentes du génome de *L. kluyveri* mais bien présentes dans les régions subtélomériques des bras chromosomiques orthologues au bras gauche du chromosome C d'espèces apparentées du genre *Lachancea*⁴⁰. En raison de la perte de ces cassettes silencieuses, *L. kluyveri* est hétérothallique et ne peut de ce fait pas changer de signe sexuel. De manière intéressante, le patron d'évolution de cette région est caractérisé par des taux de recombinaison et de mutations différents du reste du génome.

Ces résultats révèlent l'importance évolutive d'un évènement ancestral d'introgression au sein de l'espèce *L. kluyveri*, un mécanisme évolutif reconnu comme étant répandu chez les levures⁹².

L'étude de la diversité génétique intraspécifique chez *S. cerevisiae* et *L. kluyveri* constitue une première étape indispensable vers la compréhension de la relation génotype-phénotype. La deuxième étape est l'étude de la diversité phénotypique des souches au sein de ces deux espèces.

3. L'étude de la diversité phénotypique entre souches d'une même espèce

La diversité phénotypique peut être étudiée à différentes échelles, au niveau macroscopique (par exemple la croissance invasive, la formation de biofilm et la morphologie des colonies), au niveau cellulaire (par exemple la morphologie et la croissance des cellules) et au niveau moléculaire (par exemple le métabolome, le protéome et le transcriptome).

3.1. Une multitude de phénotypes d'intérêt

Chez les levures Saccharomycotina, la diversité phénotypique est considérable entre souches d'une même espèce et reflète la grande diversité des stress environnementaux rencontrés par les souches.

La levure *S. cerevisiae* peut induire deux programmes de différenciation cellulaire à savoir la croissance invasive sous la forme pseudo-hyphale et la sporulation en réponse à un environnement devenant défavorable à la croissance. Les souches cliniques isolées de patients atteints de pathologies possèdent des phénotypes associés à la pathogénicité et en communs avec la levure pathogène *C. albicans*, dont la capacité à pousser à de hautes températures (38-42°C) et à envahir les tissus⁹³. En effet, les souches cliniques sont capables d'effectuer une transition morphologique entre la forme unicellulaire dite levure et la forme filamenteuse dite pseudo-hyphale. Cette dernière forme leur permet de résister et de proliférer dans le corps humain jusqu'à envahir différents tissus.

La sporulation est un processus induisant la division par méiose d'une cellule diploïde en quatre spores haploïdes en réponse à une carence en azote et en source de carbone. Les spores représentent une forme de résistance aux conditions défavorables du milieu. L'efficacité de sporulation, c'est-à-dire la fraction de cellules qui sporulent, est variable entre les souches de *S. cerevisiae*. Par exemple, les souches isolées à partir d'exsudats de chênes en Amérique du Nord et du sol ont une efficacité de sporulation supérieure à 95% à l'opposé des souches du vin qui sporulent faiblement^{94,95}. L'adaptation des souches à leur environnement permet d'expliquer un tel phénotype. En effet, le sol ou les exsudats d'arbres représentent des milieux hostiles où les nutriments sont rares ou peu accessibles contrairement au vin qui est un milieu riche en nutriments.

La floculation est également un phénotype très étudié en raison de son intérêt en industrie. Il s'agit d'un phénomène dans lequel les cellules s'agrègent pour former des agrégats composés de milliers de cellules^{96,97}. Les agrégats ainsi formés ont la propriété de sédimenter, ce qui facilite grandement la séparation des cellules du milieu dans lequel elles se trouvent. Cette propriété a une application en industrie dans l'élaboration de boissons alcoolisées telles que la bière ou le vin afin de séparer les levures du milieu à la fin de la fermentation. Bien que la floculation soit un phénomène pouvant être estimé visuellement, il existe différents types de mesure pour la quantifier dont le test de sédimentation de Helm qui mesure le nombre de cellules sédimentées⁹⁸. Sur la base de ce test, les souches industrielles utilisées dans le brassage de la bière ont un pourcentage de floculation de l'ordre de 40 à 90% contre 0 à 15% pour les souches de laboratoire pour lesquelles la floculation est un phénotype indésirable⁹⁹. Cependant la plupart de ces études ne se concentrent que sur le phénotype d'une dizaine de souches de *S. cerevisiae* et par conséquent ne représentent pas l'étendue de la diversité phénotypique au sein de l'espèce. A ce jour l'exploration de la variabilité phénotypique au sein d'une population telle qu'une collection de souches a été entreprise pour quelques phénotypes dont la croissance et la morphologie des cellules.

3.2. Les analyses à haut débit

Des progrès considérables ont été réalisés durant ces dernières années pour analyser la croissance et la morphologie des cellules avec le développement de stratégies de phénotypage à haut-débit basées sur des procédures partiellement automatisées et une analyse informatisée des données. L'exploration de tels phénotypes est possible par le suivi de la croissance dans différentes conditions soit en microcultures^{100,101} soit sur milieu solide^{102,103} et par l'acquisition d'images de

microscopie¹⁰⁴. Ces avancées rendent possible la comparaison d'un plus grand nombre de souches de manière reproductible et efficace.

L'analyse de la croissance

La croissance des levures dépend de plusieurs caractères quantitatifs dont la durée de la phase de latence c'est-à-dire le temps d'adaptation d'une souche à un nouvel environnement, de la vitesse spécifique de croissance et de la biomasse totale formée. Deux études se sont intéressées à la capacité de croissance de souches de levures face à des stress environnementaux pouvant influencer sur un ou plusieurs de ces caractères^{105,67}.

La première étude¹⁰⁵ a exploré la diversité phénotypique au sein d'une collection de 52 souches de *S. cerevisiae* isolées de diverses niches écologiques et soumises à des stress environnementaux. La croissance des souches a été suivie sur milieu solide en présence de 14 facteurs de stress et comparée à la souche de référence S288c. Les résultats mettent en évidence une forte variation phénotypique au sein de l'espèce *S. cerevisiae* et donnent des indications sur les relations entre les souches et leur environnement. Les souches isolées d'une même niche écologique présentent des profils phénotypiques similaires probablement dus à des caractéristiques communes et propres à chacune des niches. Par exemple, les souches isolées de la vigne sont résistantes au sulfate de cuivre, un composé chimique couramment utilisé dans le traitement de la vigne comme fongicide. Le sulfate de cuivre présent dans cette niche exercerait une pression de sélection conduisant à l'apparition de la résistance des souches de la vigne à ce composé. De manière globale, les souches de la vigne résistent relativement bien à l'ensemble des conditions de stress, cette caractéristique reflète leur capacité à se développer dans un environnement variable par opposition aux souches utilisées dans la fermentation du saké plus adaptées à des conditions bien définies de fermentation du saké. L'étude a par ailleurs montré que le profil phénotypique de la souche de laboratoire S288c diffère des autres souches et révèle donc l'importance de l'analyse de la variabilité intraspécifique plus représentative de la diversité de l'espèce.

La deuxième étude⁶⁷ s'est quant à elle intéressée à la diversité phénotypique dans une collection de 39 souches de *S. cerevisiae* et de 39 souches de *S. paradoxus*. Afin d'avoir une vue globale de la diversité phénotypique au sein du genre *Saccharomyces*, quelques souches de *S. bayanus*, *S. arboricolus*, *S. kudriavzevii*, *S. castelli* et *S. mikatae* ont également été incluses dans l'étude. La croissance des souches pour plus de 200 conditions de stress a été analysée en microcultures. L'analyse des données a mis en évidence une variabilité phénotypique plus importante chez *S. cerevisiae* par rapport à *S. paradoxus* probablement en raison de son utilisation par l'Homme et d'un nombre plus élevé de CNV chez *S. cerevisiae*⁸⁸. La variation phénotypique semble en effet refléter

l'histoire de la population plutôt que l'adaptation à une niche spécifique. Les souches de *S. cerevisiae* ont globalement une plus grande tolérance à la plupart des conditions testées, celles-ci ont notamment une meilleure capacité d'assimilation du maltose et de résistance à une haute température, à des toxines ou encore à l'éthanol. La souche de laboratoire S288c a un profil phénotypique atypique comme le suggérait l'étude précédente. De manière générale, les souches de laboratoire ne sont donc pas les plus représentatives de la diversité de l'espèce.

L'analyse de la morphologie des cellules

La morphologie des cellules est déterminée par de nombreux processus cellulaires dont la progression du cycle cellulaire, l'établissement de la polarité de la cellule et la régulation du contrôle de sa taille. Elle dépend de centaines de caractères quantitatifs comprenant la taille et la forme de la cellule-mère, la taille et l'orientation du bourgeon, et l'organisation intracellulaire (Figure 5).

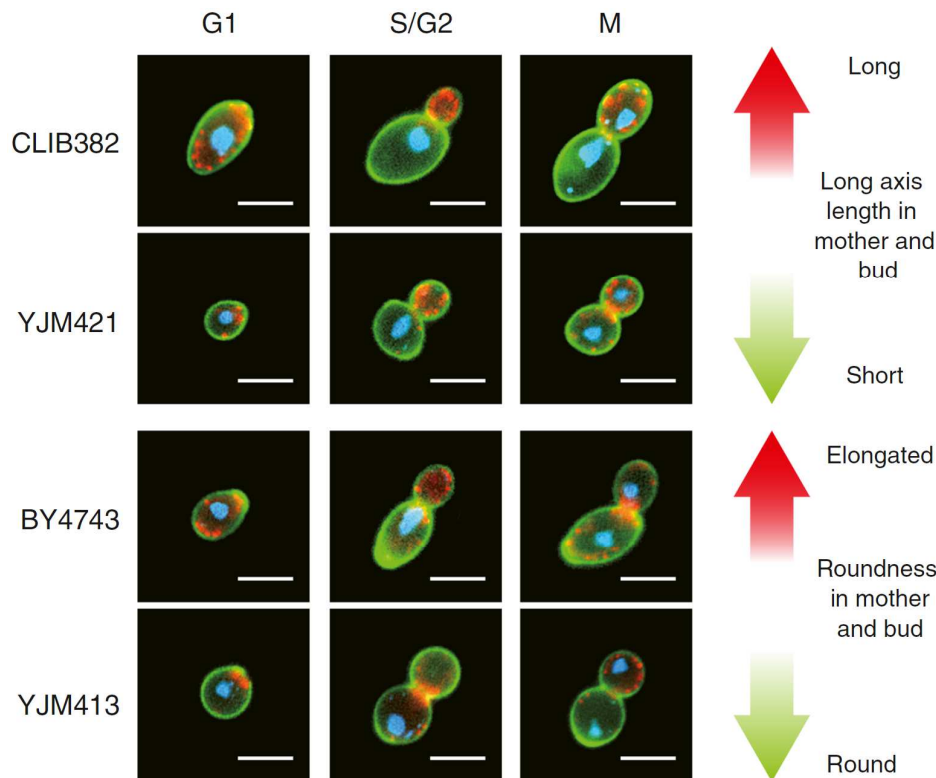


Figure 5 : Variation intraspécifique de la morphologie cellulaire chez *S. cerevisiae* au cours du cycle cellulaire (adaptée de Yvert *et al.* 2013)¹⁰⁷. Observation au microscope à fluorescence des souches CLIB382, YJM421, BY4743 et YJM413 après marquage des mannoprotéines constituant la paroi des cellules, du cytosquelette d'actine et de l'ADN à la concanavaleine A couplée à l'isothiocyanate de fluorescéine, à la phalloïdine couplée à la rhodamine et au 4',6'-diamidino-2-phénylindole respectivement aux phases G1, S/G2 et M du cycle cellulaire. Echelle : 5 μ m.

L'exploration des caractères morphologiques chez la levure est possible grâce à une plateforme d'imagerie à haut débit permettant l'acquisition d'images de microscopie à fluorescence¹⁰⁶. Cette technique consiste à marquer les cellules par des fluorochromes afin de visualiser la paroi cellulaire, l'ADN et le cytosquelette d'actine. Après acquisition et analyse des images, 501 paramètres quantitatifs sont mesurés reflétant divers aspects de la morphologie des cellules dont la forme de la cellule, la morphologie nucléaire et le cytosquelette d'actine. Cette même technique a été utilisée pour explorer la variabilité morphologique à travers une collection de 37 souches d'origines naturelles de *S. cerevisiae*¹⁰⁷. La majorité des caractères morphologiques sont variables entre les souches. Le caractère pour lequel la variabilité est la plus forte est l'allongement de la cellule. Par exemple, la souche de boulangerie CLIB192 a une forme ronde tandis que la souche de distillerie YJM269 est complètement allongée. La position du bourgeon et la taille de la cellule mère varient également grandement entre les souches. Par opposition aux phénotypes de croissance en présence de différents stress, aucun impact de l'environnement écologique ni de l'histoire de la population n'est observé sur la variation de la morphologie cellulaire. De manière intéressante, l'étude de la morphologie des cellules fournit une observation directe du comportement des cellules individuellement contrairement à celle de la croissance qui reflète la contribution d'une population de cellules¹⁰⁴.

Une fois la diversité génétique et phénotypique intraspécifique explorée, la dernière étape est l'étude de l'origine génétique de la variation phénotypique.

4. L'étude de la relation génotype-phénotype

L'identification des variations génétiques entre individus d'une même espèce est essentielle pour avoir une compréhension approfondie des mécanismes à l'origine des différences de phénotypes observées. La puissance de la génétique de la levure fait de *S. cerevisiae* un organisme de choix pour étudier la relation génotype-phénotype. De par son génome petit et compact, il est plus aisé de cartographier les régions impliquées dans un phénotype. D'un point de vue expérimental, il est possible de réaliser facilement des croisements et d'obtenir en peu de temps une descendance importante (plus d'une centaine d'individus). Le taux de recombinaison méiotique au sein de cette levure est élevé (en moyenne 90 crossing-over par méiose) et participe à l'augmentation de la variabilité génétique, et donc phénotypique, au sein des descendants¹⁰⁸. Deux méthodes principales de cartographie des QTL existent à savoir l'analyse de liaison et l'analyse d'association¹⁰⁹. Un QTL (Quantitative Trait Locus) est une région génomique plus ou moins grande portant un ou plusieurs gènes dont le polymorphisme est associé à la variation d'un

caractère quantitatif ou complexe. On distingue les QTL majeurs et les QTL mineurs ayant respectivement un fort et un faible impact sur le phénotype.

4.1. Les analyses de liaison

4.1.1. Le principe de la méthode

L'analyse de liaison repose sur l'étude de la descendance issue d'un croisement entre deux souches parentales (Figure 6). Le choix des souches porte sur plusieurs critères. Les deux parents doivent avoir une divergence génétique suffisante, une diversité phénotypique, pouvoir se croiser et créer une descendance de viabilité élevée possédant des points de recombinaison.

Grâce aux événements de recombinaison au cours de la méiose, le génome des descendants contient différentes combinaisons des génomes parentaux. L'hybridation sur puces à ADN ou le séquençage complet des souches parentales permettent d'identifier des sites polymorphiques (locus où il y a deux versions alléliques) dispersés le long du génome et donc de définir des marqueurs (SNP et indels). Ces derniers permettent de déterminer l'origine parentale des régions génomiques dans la descendance. La précision de détection du QTL dépend par conséquent de la densité des marqueurs et de leur répartition homogène sur l'ensemble du génome mais également du taux de recombinaison dans la région où se trouve le locus causal. Il est de ce fait important que la divergence génétique entre les parents soit suffisante. Le génotype de l'ensemble des descendants est caractérisé par la détermination de la version allélique au niveau des sites polymorphiques différenciant les parents. En parallèle, une mesure quantitative du phénotype de chaque descendant est déterminée.

Si le phénotype est monogénique, la descendance présente un caractère Mendélien où seuls les phénotypes parentaux sont observés et en proportion égale. La distribution de la variation phénotypique est bimodale (Figure 7). Au contraire si le phénotype est complexe, la descendance est composée d'individus présentant des phénotypes divergents, c'est-à-dire d'individus exprimant ou non le phénotype mais également d'individus présentant des phénotypes intermédiaires. La distribution de la variation phénotypique suit une loi normale (Figure 7).

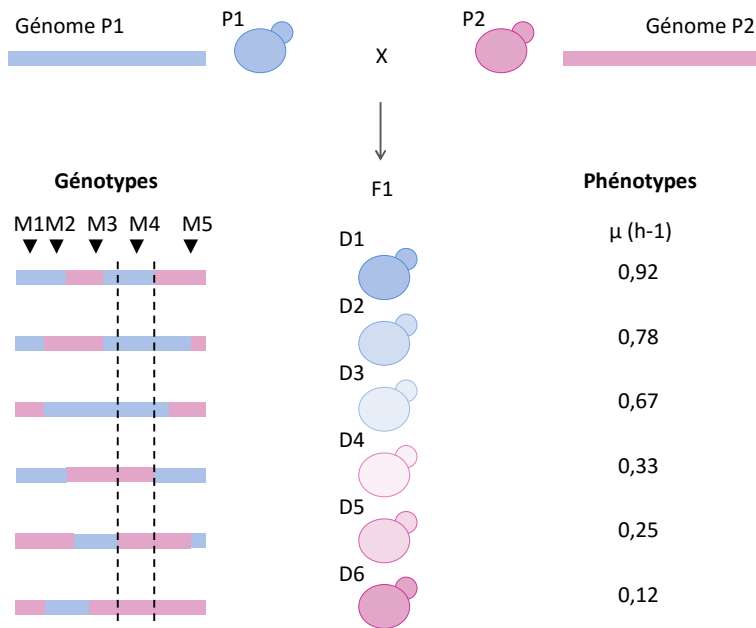


Figure 6 : Schéma d'une analyse de liaison. Le génotype et la mesure quantitative du phénotype (la vitesse spécifique de croissance par exemple) des descendants D1 à D6 issus du croisement entre les parents P1 et P2 sont déterminés. Les marqueurs M1 à M5 sont utilisés pour distinguer les génomes parentaux au sein de la descendance, de manière à savoir de quel parent provient chacune des régions du génome. Les pointillés délimitent la région impliquée dans le phénotype étudié.

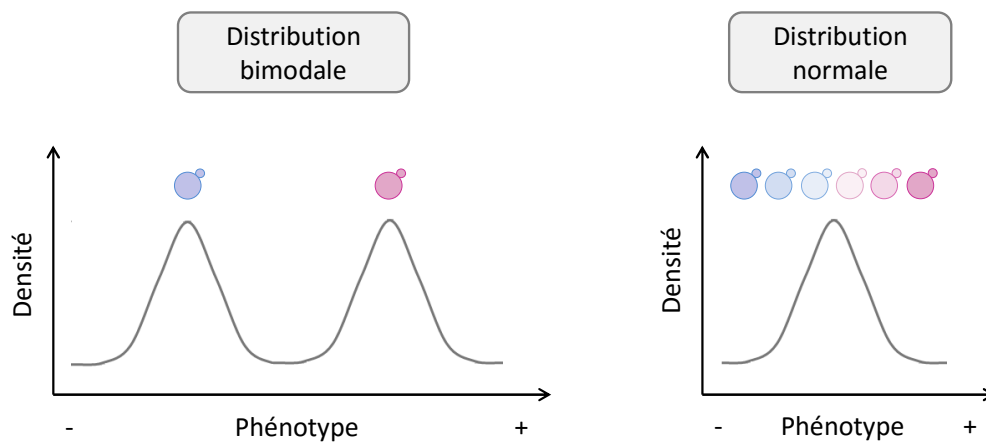


Figure 7 : Schéma des distributions phénotypiques au sein de la descendance.

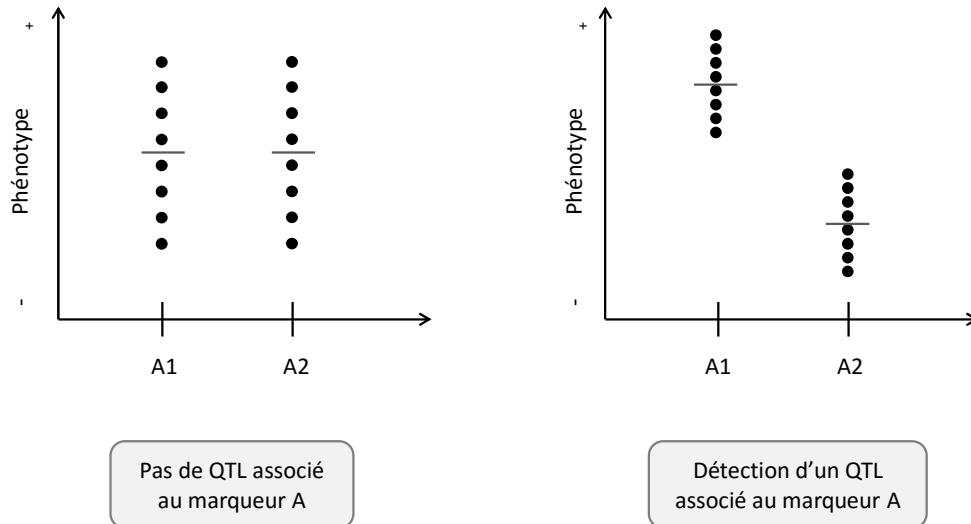


Figure 8 : Corrélation entre le phénotype d'intérêt et le marqueur.

Pour chaque descendant, on dispose alors du génotype aux marqueurs, d'une part, et de la valeur phénotypique, d'autre part. La suite de la méthode consiste à établir une corrélation entre certains marqueurs et la valeur du phénotype. La valeur phénotypique moyenne des descendants possédant à titre d'exemple l'allèle A1 et celle des descendants possédant l'allèle A2 sont calculées. Si la différence entre ces deux moyennes est significative alors un QTL est détecté autour du marqueur A (Figure 8). Dans le cas contraire, l'analyse de liaison ne permet pas de mettre en évidence de QTL associé au marqueur A.

4.1.2. Quelques exemples d'analyse de liaison

L'analyse de liaison a permis de cartographier des QTL impliqués dans des phénotypes complexes d'intérêts industriel, biologique ou encore médical. *S. cerevisiae* est un micro-organisme très utilisé par l'homme dans la fermentation alcoolique pour l'élaboration de boissons alcoolisées (vin, bière et autres breuvages). Afin de mieux contrôler la fermentation, de nombreuses analyses de liaison ont été réalisées sur des phénotypes fermentaires dont la tolérance à l'éthanol¹¹⁰, la production élevée d'éthanol et de composés aromatiques^{111,112}, la production limitée de composés indésirables dont l'acide acétique et le sulfure d'hydrogène, et la durée de la phase de latence^{113,114}.

S. cerevisiae peut également se révéler être un pathogène opportuniste de faible virulence en co-infectant des patients dont les défenses immunitaires sont amoindries. Bien que rares, les nouveaux cas recensés d'infection à *S. cerevisiae*, en association avec d'autres pathogènes, sont en

augmentation. Des analyses de liaison ont de ce fait été réalisées sur des phénotypes d'intérêt médical qui confèrent un avantage sélectif aux souches responsables d'infections opportunistes chez des patients immunodéprimés. La capacité de croissance à haute température et par conséquent à la température corporelle d'un patient fébrile¹¹⁵ ou encore la résistance à des antifongiques et des médicaments psychotropes cytotoxiques chez la levure¹¹⁶ ont fait l'objet d'analyses de liaison.

D'autres analyses de liaison ont porté sur des phénotypes d'intérêt biologique à l'échelle de la cellule tels que les différences de morphologies cellulaires¹¹⁷, l'agrégation des cellules¹¹⁸, l'efficacité de sporulation¹¹⁹ et à l'échelle moléculaire, la variation du niveau d'expression des gènes^{120,121} et du protéome¹²², la longueur des télomères¹²³.

Ces études se sont basées pour la plupart sur l'utilisation d'une centaine de descendants et près de 3000 marqueurs génétiques. Deux études récentes se sont focalisées sur de plus larges échantillons de près de 1000 descendants et des phénotypes variés dont la réponse à divers composés toxiques, des variations de pH et de températures ou encore l'utilisation de différentes sources de carbone, la morphologie des colonies et la floculation^{124,125}. De telles études à grande échelle ont l'avantage de détecter des QTL avec un fort pouvoir statistique et de mesurer leurs effets, d'estimer directement la part d'héritabilité de caractères complexes et d'examiner les sources d'héritabilité manquante.

Cependant l'analyse de liaison, a le désavantage d'être une méthode de cartographie onéreuse et laborieuse car elle nécessite de déterminer le génotype de tous les descendants. Afin de réduire le coût du génotypage et les analyses des données de séquençage, une variante de l'analyse de liaison classique a été développée à savoir l'analyse par pool de ségrégants dite Bulk Segregant Analysis¹²⁶.

4.1.3. Les analyses par pool de ségrégants

L'analyse par pool de ségrégants permet de caractériser en une seule fois le génotype d'un groupe de descendants présentant le phénotype d'intérêt et non plus de chaque descendant un à un (Figure 9). Cette méthode débute de la même manière qu'une analyse de liaison classique, à l'exception près qu'un ou deux groupes (ou pool) de descendants sont formés en fonction de leur valeur phénotypique. Le premier groupe rassemble des descendants exprimant le phénotype d'intérêt tandis que le deuxième groupe réunit des descendants sélectionnés au hasard de manière à constituer un groupe contrôle. L'ADN total des deux groupes est en général complètement séquencé par des stratégies de séquençage à haut-débit. Pour chaque site polymorphique

différenciant les souches parentales, les fréquences alléliques des parents sont estimées dans la descendance. Pour les sites polymorphiques situés dans une région non concernée par le phénotype d'intérêt, 50% des allèles proviennent d'un parent et 50% de l'autre parent. La fréquence allélique est égale à 50%. Par contre, les régions du génome impliquées dans le phénotype présentent un déséquilibre de la fréquence allélique. La fréquence allélique est égale à 100%. En d'autres termes, les régions provenant majoritairement d'un parent sont liées au phénotype. Enfin dans le groupe contrôle la fréquence allélique doit être proche de 50% tout le long du génome.

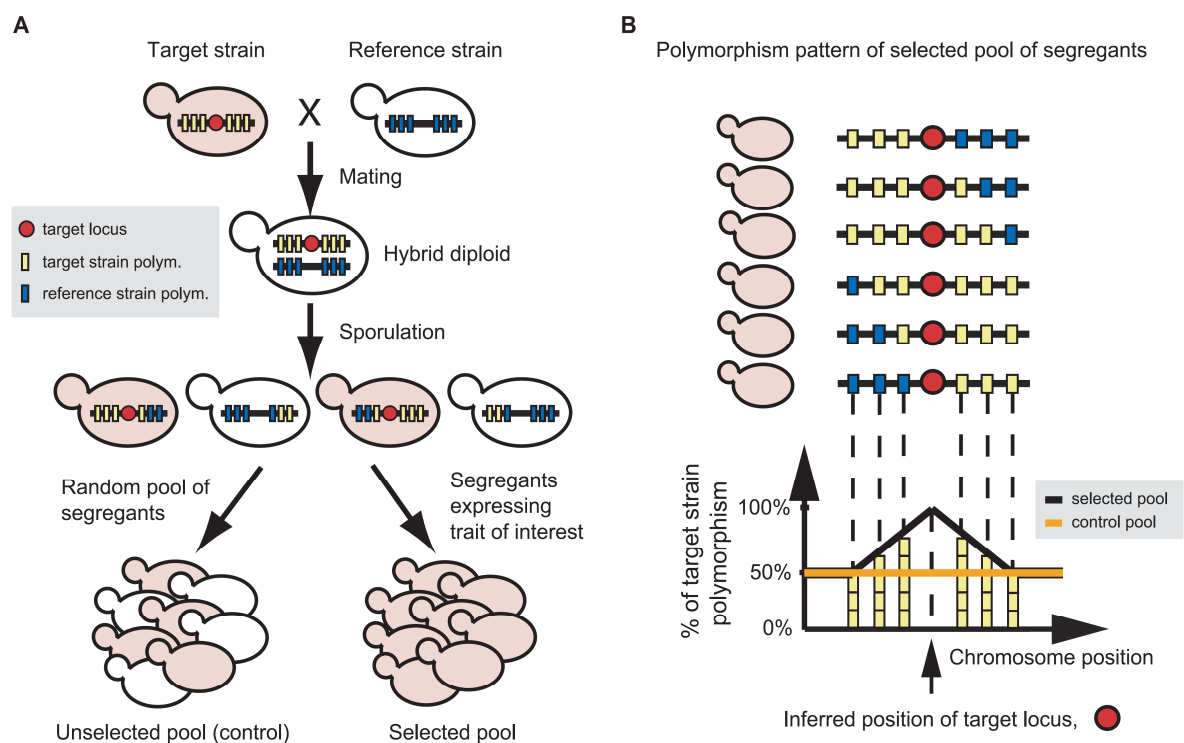


Figure 9 : Schéma de l'analyse de pool de ségrégants (d'après Segre *et al.* 2006)¹²⁶. (A) Une souche exprimant un phénotype d'intérêt (en rouge) est croisée avec une souche ne l'exprimant pas (en blanc). Le polymorphisme entre les deux souches parentales représenté par des rectangles jaunes et bleus définis les marqueurs. L'hybride obtenu est mis sur milieu de sporulation afin de générer une descendance. Un pool contrôle composé de descendants sélectionnés aléatoirement et un pool de descendants exprimant le phénotype d'intérêt sont constitués. (B) La (les) région(s) génomique(s) liée(s) au phénotype d'intérêt est (sont) cartographiée(s) par l'analyse des fréquences alléliques dans la descendance pour chaque site polymorphique différenciant les deux parents. Le groupe contrôle permet de s'assurer que la fréquence allélique de chaque parent est de 50% tout le long du génome. La déviation de la fréquence allélique dans le groupe d'intérêt permet de localiser la (les) région(s) impliquée(s) dans le phénotype d'intérêt.

La résolution de détection de QTL étant fonction du nombre de points de recombinaisons, de nouvelles analyses visent à augmenter le nombre de descendants dans le but de cartographier des QTL avec une plus grande précision. Les analyses de type X-QTL et iQTL dérivant de l'analyse de pool de ségrégants permettent la détection de QTL avec une plus grande sensibilité et une meilleure résolution^{127,128}. La 1^{ère} méthode de type X-QTL permet de générer une descendance F1 de très grande taille grâce à un système de sélection basé sur des auxotrophies¹²⁷. La 2^{nde} méthode de type iQTL réalise quant à elle plusieurs générations de descendants et utilise de ce fait les recombinaisons qui s'accumulent au cours des générations¹²⁸. Chez cette dernière les descendants présentant le phénotype d'intérêt sont sélectionnés en masse sous pression de sélection.

4.1.4. Quelques exemples d'analyse par pool de ségrégants

L'analyse par pool de ségrégants a d'abord été développée chez la levure en utilisant la technologie des puces à ADN^{129,126,127} puis a été adaptée au séquençage à haut débit^{130,131}. Cette méthode de cartographie a permis d'identifier chez *S. cerevisiae* les origines génétiques de nombreux phénotypes contrôlés par un ou plusieurs gènes et ayant des intérêts dans bien des domaines.

S. cerevisiae est un micro-organisme de choix en industrie notamment pour la production de bioéthanol comme source d'énergie à partir de la lignocellulose. Les souches utilisées dans un tel procédé industriel doivent par conséquent être capable de fermenter le principal pentose de l'hémicellulose à savoir le xylose et pouvoir résister à de fortes concentrations en éthanol. Une première analyse par pool de ségrégants a permis d'identifier un gène responsable de l'utilisation du xylose comme source de carbone dans de nombreuses souches isolées de la vigne à savoir le gène *XDH1* codant la xylitol déhydrogénase¹³¹. Une seconde analyse a permis de cartographier trois gènes comme étant impliqués dans la tolérance à de fortes concentrations en éthanol chez une souche utilisée dans la production de bioéthanol soit les gènes *MKT1*, *SWS2* et *APJ1*¹³².

L'explosion oxydative des polynucléaires neutrophiles, libérant des formes réactives de l'oxygène tels que le peroxyde d'hydrogène (H₂O₂), représente un puissant mécanisme de défense du système immunitaire humain contre les agents pathogènes dont les levures. C'est pourquoi, il est primordial de définir les déterminants génétiques pouvant conduire à une résistance au stress oxydatif. Ainsi, une analyse par pool de ségrégants a été réalisée chez une souche isolée d'un patient infecté par *S. cerevisiae* dans le but de déterminer l'origine génétique de la résistance au peroxyde d'hydrogène¹³³. Cette étude a permis de révéler l'importance du facteur de transcription Rds2 dans la réponse au stress oxydatif.

Certaines mutations peuvent avoir des conséquences directes sur la capacité reproductive et par conséquent conduire à l'apparition d'isolements reproductifs voir de nouvelles espèces. Récemment, une étude s'est intéressée aux mécanismes responsables de l'isolement reproductif intraspécifique¹³⁴. L'analyse par pool de ségrégants a permis de mettre en évidence deux allèles incompatibles dans le cas d'une croissance sur une source de carbone non fermentescible. L'incompatibilité génétique détectée provoque une déficience au niveau de la chaîne respiratoire dans une partie de la descendance et se produit entre un gène codant une protéine mitochondriale *COX15* et un ARNt suppresseur *SUP7*¹³⁴.

4.2. L'analyse d'association

4.2.1. Le principe de la méthode

L'analyse d'association est basée sur le même principe que l'analyse de liaison à l'exception près qu'elle est réalisée directement sur une population existante, comme une collection de souches isolées de différents environnements par exemple¹⁰⁹ (Figure 10).

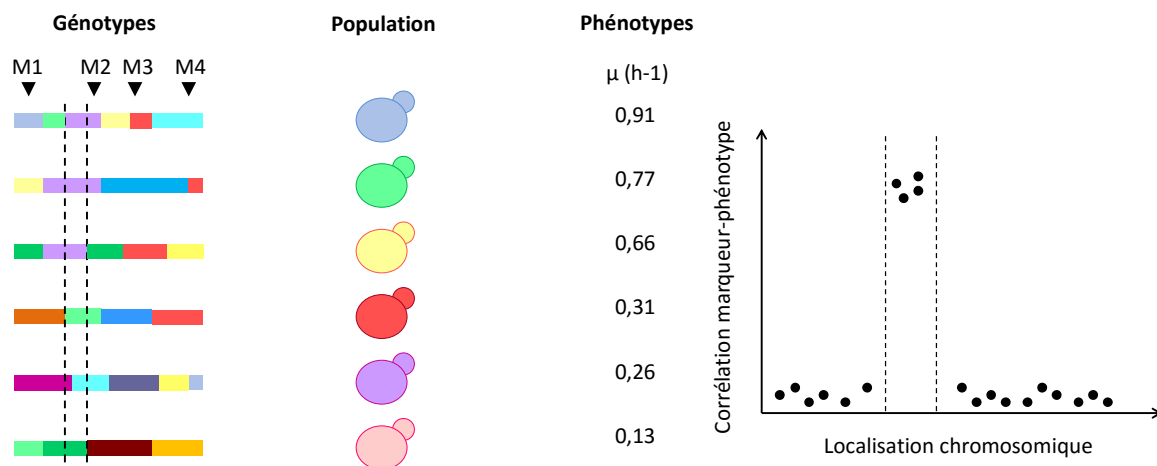


Figure 10 : Schéma d'une analyse d'association. Le génotype et le phénotype d'une population de souches sont déterminés. Le polymorphisme entre les souches de la population définit les marqueurs. La (les) région(s) génomique(s) associée(s) au phénotype d'intérêt est (sont) cartographiée(s) par l'analyse de la corrélation entre chaque marqueur et le phénotype des descendants l'exprimant.

Les génomes sont par conséquent issus de l'accumulation de divers événements de recombinaison ancestrale, c'est-à-dire d'échanges de matériels génétiques entre des génomes ancestraux, et divergent au niveau de leur séquence. Le génotype de chaque isolat est déterminé en parallèle de la mesure quantitative du phénotype étudié. Comme décrit précédemment,

L'association entre chaque marqueur et le phénotype d'intérêt dans la population est testée, permettant d'identifier les régions ayant un impact sur la variation phénotypique. L'analyse d'association a plus particulièrement recours au déséquilibre de liaison génétique au sein du génome, qui est défini par une association non aléatoire des allèles présents sur des loci différents. Cette méthode de cartographie a l'avantage d'explorer la diversité génétique présente au sein d'une population de souches et pas uniquement le patrimoine génétique de deux souches parentales.

4.2.2. Les études d'association pangénomiques

Le développement des techniques de séquençage et de génotypage à haut débit de ces dernières années a permis l'obtention rapide d'un très grand nombre de génotypes et par conséquent d'une haute densité de marqueurs génétiques. Cette opportunité technologique a ouvert la voie à des études d'association à très grande échelle, les études d'association pangénomiques (GWAS, Genome-Wide Association Study). Les études d'association pangénomiques se révèlent être une méthode efficace pour cartographier les gènes à l'origine de caractères complexes.

Cependant, à l'heure actuelle peu d'analyses d'association ont été menées chez la levure en raison d'un manque de puissance statistique, en partie dû à la taille relativement faible des populations de souches étudiées et de la complexité de la structure de la population^{32,33}. Afin de faciliter de telles études chez cet organisme modèle, il est primordial de contrôler et de corriger la structure de la population à l'aide d'outils statistiques afin de limiter les fausses associations entre un marqueur génétique et un phénotype^{135,136}. Une autre approche possible pour contrer l'effet de la structure de la population est d'utiliser un très grand nombre de souches et de sélectionner des sous-populations de souches homogènes au niveau de la structure de la population voir avec un niveau de structuration faible.

L'espèce *S. cerevisiae* fait actuellement l'objet d'une telle étude à grande échelle dans le cadre du projet « les 1002 génomes de levures : fondations pour des analyses pangénomiques d'association génotype-phénotype » (<http://1002genomes.u-strasbg.fr/>) au sein de l'équipe et avec le soutien de collaborations. Pour ce projet, le génome de plus de 1002 souches de *S. cerevisiae* a été séquencé dans son intégralité (Illumina HiSeq 2000) afin d'avoir une vision globale de la diversité génétique au sein de cette espèce et de définir les origines génétiques de la diversité phénotypique.

4.3. L'identification et la validation fonctionnelle des gènes candidats

Les régions génomiques cartographiées par analyse de liaison ou par analyse d'association peuvent contenir une centaine de gènes ayant des fonctions variées. Une étude approfondie de la

fonction et de la séquence des gènes présents dans l'intervalle de localisation du QTL s'avère donc nécessaire pour établir une liste plus restreinte de gènes candidats au sein de chaque région. L'analyse fonctionnelle des gènes consiste à interroger des banques de données biologiques telles que Saccharomyces Genome Database (SGD, <http://www.yeastgenome.org/>) ou encore Genome Resources for Yeast Chromosomes (GRYC, gryc.inra.fr) respectivement consacrées à l'organisme modèle *S. cerevisiae* et à 24 espèces de levures Saccharomycotina dont *L. kluyveri*. Ces banques de données permettent de lister les gènes présents dans la région d'intérêt et de donner des indications quant à leur fonction dans la cellule à travers le phénotype de mutants de délétion, au processus biologique dans lequel ils sont impliqués et à leurs interactions. L'analyse du polymorphisme permet d'acquérir de nouvelles indications sur les potentielles variations nucléotidiques présentes dans les gènes candidats par comparaison des séquences des deux souches parentales. Si tel est le cas, les séquences nucléiques sont traduites en séquences protéiques afin de discriminer les mutations synonymes, des mutations non-synonymes. Ces dernières sont susceptibles d'avoir un impact direct sur la conformation et le fonctionnement de la protéine et donc des conséquences au niveau phénotypique. L'analyse de l'expression des gènes peut également compléter cette étude. Il n'est pas à exclure la présence de mutations dans des régions non codantes qui peuvent notamment impacter le niveau d'expression des gènes.

Une fois le ou les gènes candidats identifiés ainsi que les variants génétiques, il reste à conclure quant à leur rôle ou non dans le phénotype. La validation fonctionnelle des gènes candidats est réalisée expérimentalement par remplacement allélique ou par un test d'hémizygotie réciproque^{137,115}. Ces deux méthodes permettent d'identifier de manière définitive parmi les gènes candidats lesquels ont un rôle de déterminant génétique et de mesurer leur impact sur le phénotype.

4.3.1. Le remplacement allélique

Le remplacement allélique permet de tester la version allélique d'un gène candidat dans un autre fonds génétique (Figure 11A). Plus particulièrement, cette méthode consiste à remplacer le gène candidat dans le parent n'exprimant pas le phénotype par la version allélique du parent exprimant le phénotype. La souche obtenue après remplacement allélique conserve ainsi la totalité de son patrimoine génétique hormis le gène candidat. De ce fait, si le remplacement allélique restaure entièrement ou en partie le phénotype d'intérêt, alors le gène candidat est impliqué dans la variation phénotypique.

4.3.2. Le test d'hémizygocité réciproque

Le test d'hémizygocité réciproque permet de tester la version allélique d'un gène candidat dans un fonds génétique diploïde (Figure 11B). Deux souches parentales sont croisées afin de former une souche diploïde hétérozygote à partir de laquelle sont générées deux souches hémizygotes c'est-à-dire délétées pour l'allèle d'un des parents. Les phénotypes des trois souches diploïdes sont alors comparés. Si les phénotypes des trois souches sont différents alors le gène candidat est impliqué dans la variation du phénotype.

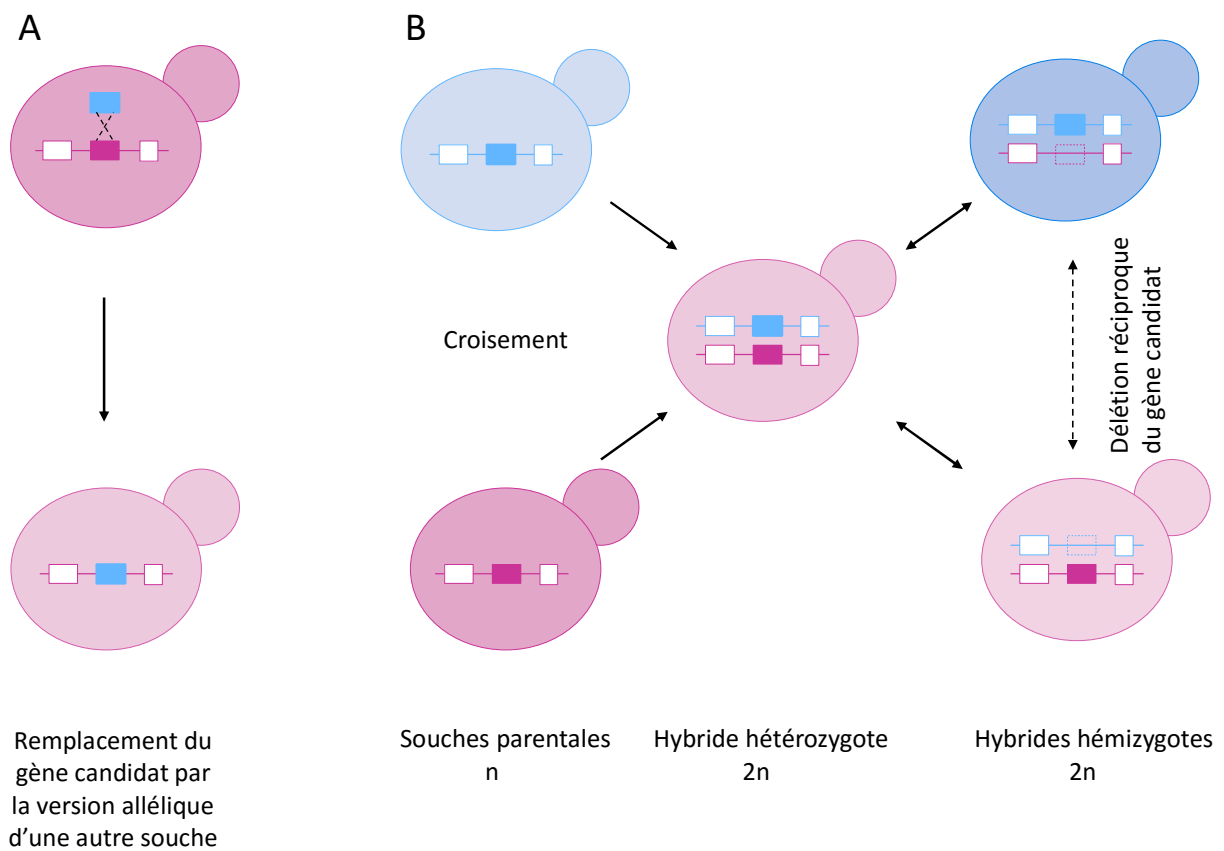


Figure 11 : Validation fonctionnelle d'un gène candidat. A. Schéma d'un remplacement allélique. B. Schéma d'un test d'hémizygocité réciproque.

Références

1. Badano, J. L. & Katsanis, N. Beyond Mendel: an evolving view of human genetic disease transmission. *Nature reviews. Genetics* **3**, 779–89 (2002).
2. Botstein, D. & Risch, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nature genetics* **33 Suppl**, 228–37 (2003).
3. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
4. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 9362–7 (2009).
5. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nature reviews. Genetics* **9**, 255–66 (2008).
6. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *American journal of human genetics* **90**, 7–24 (2012).
7. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1193–8 (2012).
8. Ku, C.-S., Naidoo, N. & Pawitan, Y. Revisiting Mendelian disorders through exome sequencing. *Human genetics* **129**, 351–70 (2011).
9. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
10. Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–6 (2015).
11. Lange, L. A. *et al.* Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *American journal of human genetics* **94**, 233–45 (2014).
12. Dujon, B. Yeast evolutionary genomics. *Nature reviews. Genetics* **11**, 512–24 (2010).
13. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–13 (1997).
14. Kellis, M., Birren, B. W. & Lander, E. S. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**, 617–24 (2004).
15. Gordon, J. L., Byrne, K. P. & Wolfe, K. H. Mechanisms of chromosome number evolution in yeast. *PLoS genetics* **7**, e1002190 (2011).

16. Peter, J. & Schacherer, J. Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast (Chichester, England)* **33**, 73–81 (2016).
17. McGovern, P. E. *et al.* Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17593–8 (2004).
18. Cavalieri, D., McGovern, P. E., Hartl, D. L., Mortimer, R. & Polsinelli, M. Evidence for *S. cerevisiae* fermentation in ancient wine. *Journal of molecular evolution* **57 Suppl 1**, S226–32 (2003).
19. Fay, J. C. & Benavides, J. A. Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS genetics* **1**, 66–71 (2005).
20. Landry, C. R., Townsend, J. P., Hartl, D. L. & Cavalieri, D. Ecological and evolutionary genomics of *Saccharomyces cerevisiae*. *Molecular Ecology* **15**, 575–591 (2006).
21. Aa, E., Townsend, J. P., Adams, R. I., Nielsen, K. M. & Taylor, J. W. Population structure and gene evolution in *Saccharomyces cerevisiae*. *FEMS yeast research* **6**, 702–15 (2006).
22. 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**, 5404–5417 (2012).
23. Goffeau, A. *et al.* Life with 6000 genes. *Science (New York, N.Y.)* **274**, 546, 563–7 (1996).
24. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science (New York, N.Y.)* **296**, 752–5 (2002).
25. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature genetics* **35**, 57–64 (2003).
26. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS biology* **6**, e83 (2008).
27. Novo, M. *et al.* Eukaryote-to-eukaryote gene transfer events revealed by the genome sequence of the wine yeast *Saccharomyces cerevisiae* EC1118. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 16333–8 (2009).
28. Tawfik, O. W., Papasian, C. J., Dixon, A. Y. & Potter, L. M. *Saccharomyces cerevisiae* pneumonia in a patient with acquired immune deficiency syndrome. *Journal of clinical microbiology* **27**, 1689–91 (1989).
29. Wei, W. *et al.* Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 12825–30 (2007).
30. Borneman, A. R., Forgan, A. H., Pretorius, I. S. & Chambers, P. J. Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS yeast research* **8**, 1185–95 (2008).
31. Argueso, J. L. *et al.* Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol

- production. *Genome research* **19**, 2258–70 (2009).
32. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–41 (2009).
 33. Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–5 (2009).
 34. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–91 (2002).
 35. Pena-Castillo, L. & Hughes, T. R. Why are there still over 1000 uncharacterized yeast genes? *Genetics* **176**, 7–14 (2007).
 36. Botstein, D. & Fink, G. R. Yeast: an experimental organism for 21st Century biology. *Genetics* **189**, 695–704 (2011).
 37. Phaff, H. J., Miller, M. W. & Shifrine, M. The taxonomy of yeasts isolated from *Drosophila* in the Yosemite region of California. *Antonie van Leeuwenhoek* **22**, 145–61 (1956).
 38. Kurtzman, C. P. Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygorhynchus*. *FEMS yeast research* **4**, 233–45 (2003).
 39. McCullough, J. & Herskowitz, I. Mating pheromones of *Saccharomyces kluyveri*: pheromone interactions between *Saccharomyces kluyveri* and *Saccharomyces cerevisiae*. *Journal of bacteriology* **138**, 146–54 (1979).
 40. Payen, C. *et al.* Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome research* **19**, 1710–21 (2009).
 41. Souciet, J.-L. *et al.* Comparative genomics of protoploid *Saccharomycetaceae*. *Genome research* **19**, 1696–709 (2009).
 42. Moller, K., Sharif, M. Z. & Olsson, L. Production of fungal alpha-amylase by *Saccharomyces kluyveri* in glucose-limited cultivations. *Journal of biotechnology* **111**, 311–8 (2004).
 43. Gojkovic, Z., Paracchini, S. & Piskur, J. A new model organism for studying the catabolism of pyrimidines and purines. *Advances in experimental medicine and biology* **431**, 475–9 (1998).
 44. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *Journal of molecular biology* **295**, 1073–87 (2000).
 45. Beck, H., Dobritzsch, D. & Piskur, J. *Saccharomyces kluyveri* as a model organism to study pyrimidine degradation. *FEMS yeast research* **8**, 1209–13 (2008).
 46. Andersson Rasmussen, A. *et al.* Global expression analysis of the yeast *Lachancea* (*Saccharomyces*)

- kluuyveri reveals new URC genes involved in pyrimidine catabolism. *Eukaryotic cell* **13**, 31–42 (2014).
47. Fujimura, H. Transformation of the yeast *Saccharomyces kluuyveri* by *Saccharomyces cerevisiae*-based plasmids. *FEMS microbiology letters* **66**, 149–52 (1991).
 48. Neueglise, C. *et al.* Genomic exploration of the hemiascomycetous yeasts: 9. *Saccharomyces kluuyveri*. *FEBS letters* **487**, 56–60 (2000).
 49. Dujon, B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in genetics* : *TIG* **22**, 375–87 (2006).
 50. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science (New York, N.Y.)* **301**, 71–6 (2003).
 51. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–54 (2003).
 52. Dujon, B. *et al.* Genome evolution in yeasts. *Nature* **430**, 35–44 (2004).
 53. Marcet-Houben, M. & Gabaldón, T. Beyond the Whole-Genome Duplication: Phylogenetic Evidence for an Ancient Interspecies Hybridization in the Baker's Yeast Lineage. *PLoS Biol* **13**, e1002220 (2015).
 54. Dietrich, F. S. *et al.* The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science (New York, N.Y.)* **304**, 304–7 (2004).
 55. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome research* **15**, 1456–61 (2005).
 56. Bleykasten-Grosshans, C. & Neueglise, C. Transposable elements in yeasts. *Comptes rendus biologiques* **334**, 679–86 (2011).
 57. Leducq, J.-B. *et al.* Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1**, 15003 (2016).
 58. Hittinger, C. T. *et al.* Remarkably ancient balanced polymorphisms in a multi-locus gene network. *Nature* **464**, 54–8 (2010).
 59. Almeida, P. *et al.* A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature communications* **5**, 4044 (2014).
 60. Friedrich, A., Jung, P., Reisser, C., Fischer, G. & Schacherer, J. Population genomics reveals chromosome-scale heterogeneous evolution in a protoploid yeast. *Molecular biology and evolution* **32**, 184–92 (2015).
 61. Ford, C. B. *et al.* The evolution of drug resistance in clinical isolates of *Candida albicans*. *eLife* **4**,

- e00662 (2015).
62. Hirakawa, M. P. *et al.* Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome research* **25**, 413–25 (2015).
 63. Fawcett, J. A. *et al.* Population genomics of the fission yeast *Schizosaccharomyces pombe*. *PLoS one* **9**, e104241 (2014).
 64. Jeffares, D. C. *et al.* The genomic and phenotypic diversity of *Schizosaccharomyces pombe*. *Nature genetics* **47**, 235–41 (2015).
 65. Strobe, 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 research* **25**, 762–74 (2015).
 66. Doniger, S. W. *et al.* A Catalog of Neutral and Deleterious Polymorphism in Yeast. *PLoS Genet* **4**, e1000183 (2008).
 67. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS genetics* **7**, e1002111 (2011).
 68. Payen, C. *et al.* The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. *G3 (Bethesda, Md.)* **4**, 399–409 (2014).
 69. Zimmer, A. *et al.* QTL dissection of Lag phase in wine fermentation reveals a new translocation responsible for *Saccharomyces cerevisiae* adaptation to sulfite. *PLoS one* **9**, e86298 (2014).
 70. Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 16144–9 (2002).
 71. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal rearrangements as a major mechanism in the onset of reproductive isolation in *Saccharomyces cerevisiae*. *Current biology*: CB **24**, 1153–9 (2014).
 72. Clowers, K. J., Heilberger, J., Piotrowski, J. S., Will, J. L. & Gasch, A. P. Ecological and Genetic Barriers Differentiate Natural Populations of *Saccharomyces cerevisiae*. *Molecular biology and evolution* **32**, 2317–27 (2015).
 73. Hughes, T. R. *et al.* Widespread aneuploidy revealed by DNA microarray expression profiling. *Nature genetics* **25**, 333–7 (2000).
 74. Rancati, G. *et al.* Aneuploidy underlies rapid adaptive evolution of yeast cells deprived of a conserved cytokinesis motor. *Cell* **135**, 879–93 (2008).
 75. Pavelka, N. *et al.* Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321–5 (2010).

76. Chen, G., Bradford, W. D., Seidel, C. W. & Li, R. Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy. *Nature* **482**, 246–50 (2012).
77. Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 21010–5 (2012).
78. Nagarajan, M. *et al.* Natural single-nucleosome epi-polymorphisms in yeast. *PLoS genetics* **6**, e1000913 (2010).
79. Azumi, M. & Goto-Yamamoto, N. AFLP analysis of type strains and laboratory and industrial strains of *Saccharomyces sensu stricto* and its application to phenetic clustering. *Yeast (Chichester, England)* **18**, 1145–54 (2001).
80. Hennequin, C. *et al.* Microsatellite typing as a new tool for identification of *Saccharomyces cerevisiae* strains. *Journal of clinical microbiology* **39**, 551–9 (2001).
81. Winzeler, E. A. *et al.* Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics* **163**, 79–89 (2003).
82. Ayoub, M.-J., Legras, J.-L., Saliba, R. & Gaillardin, C. Application of Multi Locus Sequence Typing to the analysis of the biodiversity of indigenous *Saccharomyces cerevisiae* wine yeasts from Lebanon. *Journal of applied microbiology* **100**, 699–711 (2006).
83. Liti, G., Barton, D. B. H. & Louis, E. J. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics* **174**, 839–50 (2006).
84. Legras, J.-L., Merdinoglu, D., Cornuet, J.-M. & Karst, F. Bread, beer and wine: *Saccharomyces cerevisiae* diversity reflects human history. *Molecular ecology* **16**, 2091–102 (2007).
85. Schacherer, J., Tourrette, Y., Potier, S., Souciet, J.-L. & de Montigny, J. Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. *DNA repair* **6**, 1441–52 (2007).
86. Muller, L. A. H. & McCusker, J. H. Nature and distribution of large sequence polymorphisms in *Saccharomyces cerevisiae*. *FEMS Yeast Res* **11**, 587–594 (2011).
87. Cromie, G. A. *et al.* Genomic sequence diversity and population structure of *Saccharomyces cerevisiae* assessed by RAD-seq. *G3 (Bethesda, Md.)* **3**, 2163–71 (2013).
88. Bergstrom, A. *et al.* A high-definition view of functional genetic variation from natural yeast genomes. *Molecular biology and evolution* **31**, 872–88 (2014).
89. Almeida, P. *et al.* A population genomics insight into the Mediterranean origins of wine yeast domestication. *Molecular ecology* **24**, 5412–27 (2015).
90. Jung, P. P., Friedrich, A., Reisser, C., Hou, J. & Schacherer, J. Mitochondrial genome evolution in a single protoploid yeast species. *G3 (Bethesda, Md.)* **2**, 1103–11 (2012).
91. Talla, E., Anthouard, V., Bouchier, C., Frangeul, L. & Dujon, B. The complete mitochondrial

- genome of the yeast *Kluyveromyces thermotolerans*. *FEBS letters* **579**, 30–40 (2005).
92. Morales, L. & Dujon, B. Evolutionary role of interspecies hybridization and genetic exchanges in yeasts. *Microbiology and molecular biology reviews* : *MMBR* **76**, 721–39 (2012).
 93. de Llanos, R., Fernandez-Espinar, M. T. & Querol, A. A comparison of clinical and food *Saccharomyces cerevisiae* isolates on the basis of potential virulence factors. *Antonie van Leeuwenhoek* **90**, 221–31 (2006).
 94. Gerke, J. P., Chen, C. T. L. & Cohen, B. A. Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* **174**, 985–97 (2006).
 95. Gerke, J., Lorenz, K. & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. *Science (New York, N.Y.)* **323**, 498–501 (2009).
 96. Stratford, M. Evidence for two mechanisms of flocculation in *Saccharomyces cerevisiae*. *Yeast (Chichester, England)* **5 Spec No**, S441–5 (1989).
 97. Bony, M., Thines-Sempoux, D., Barre, P. & Blondin, B. Localization and cell surface anchoring of the *Saccharomyces cerevisiae* flocculation protein Flo1p. *Journal of bacteriology* **179**, 4929–36 (1997).
 98. Helm, E., Nohr, B. & Thorne, R. The measurement of yeast flocculence and its significance in brewing. *Wallerstein Lab Commun* **16**, 315–325 (1953).
 99. Verstrepen, K. J., Derdelinckx, G., Verachtert, H. & Delvaux, F. R. Yeast flocculation: what brewers should know. *Applied microbiology and biotechnology* **61**, 197–205 (2003).
 100. Warringer, J. & Blomberg, A. Automated screening in environmental arrays allows analysis of quantitative phenotypic profiles in *Saccharomyces cerevisiae*. *Yeast (Chichester, England)* **20**, 53–67 (2003).
 101. 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* **10**, e0119807 (2015).
 102. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nature reviews. Genetics* **8**, 437–49 (2007).
 103. Dittmar, J. C., Reid, R. J. & Rothstein, R. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC bioinformatics* **11**, 353 (2010).
 104. Ohya, Y., Kimori, Y., Okada, H. & Ohnuki, S. Single-cell phenomics in budding yeast. *Molecular biology of the cell* **26**, 3920–5 (2015).
 105. Kvitek, D. J., Will, J. L. & Gasch, A. P. Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLoS genetics* **4**, e1000223 (2008).
 106. Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proceedings of the*

- National Academy of Sciences of the United States of America* **102**, 19015–20 (2005).
107. Yvert, G. *et al.* Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC systems biology* **7**, 54 (2013).
 108. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* **454**, 479–85 (2008).
 109. Mackay, T. F. C., Stone, E. A. & Ayroles, J. F. The genetics of quantitative traits: challenges and prospects. *Nature reviews. Genetics* **10**, 565–77 (2009).
 110. Hu, X. H. *et al.* Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **175**, 1479–87 (2007).
 111. Katou, T., Namise, M., Kitagaki, H., Akao, T. & Shimoi, H. QTL mapping of sake brewing characteristics of yeast. *Journal of bioscience and bioengineering* **107**, 383–93 (2009).
 112. Steyer, D. *et al.* QTL mapping of the production of wine aroma compounds by yeast. *BMC genomics* **13**, 573 (2012).
 113. Marullo, P., Yvert, G., Bely, M., Aigle, M. & Dubourdieu, D. Efficient use of DNA molecular markers to construct industrial yeast strains. *FEMS yeast research* **7**, 1295–306 (2007).
 114. Marullo, P. *et al.* Single QTL mapping and nucleotide-level resolution of a physiologic trait in wine *Saccharomyces cerevisiae* strains. *FEMS yeast research* **7**, 941–52 (2007).
 115. Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–30 (2002).
 116. Perlstein, E. O., Ruderfer, D. M., Roberts, D. C., Schreiber, S. L. & Kruglyak, L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nature genetics* **39**, 496–502 (2007).
 117. Nogami, S., Ohya, Y. & Yvert, G. Genetic complexity and quantitative trait loci mapping of yeast morphological traits. *PLoS genetics* **3**, e31 (2007).
 118. Li, J. *et al.* Polygenic molecular architecture underlying non-sexual cell aggregation in budding yeast. *DNA research* : an international journal for rapid publication of reports on genes and genomes **20**, 55–66 (2013).
 119. Deutschbauer, A. M. & Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature genetics* **37**, 1333–40 (2005).
 120. Brem, R. B., Storey, J. D., Whittle, J. & Kruglyak, L. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436**, 701–3 (2005).
 121. Ehrenreich, I. M., Gerke, J. P. & Kruglyak, L. Genetic dissection of complex traits in yeast: insights from studies of gene expression and other phenotypes in the BYxRM cross. *Cold Spring Harbor symposia on quantitative biology* **74**, 145–53 (2009).

122. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature genetics* **39**, 1369–75 (2007).
123. Gatbonton, T. *et al.* Telomere length as a quantitative trait: genome-wide survey and genetic mapping of telomere length-control genes in yeast. *PLoS genetics* **2**, e35 (2006).
124. Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V. & Kruglyak, L. Finding the sources of missing heritability in a yeast cross. *Nature* **494**, 234–7 (2013).
125. Wilkening, S. *et al.* An evaluation of high-throughput approaches to QTL mapping in *Saccharomyces cerevisiae*. *Genetics* **196**, 853–65 (2014).
126. Segre, A. V., Murray, A. W. & Leu, J.-Y. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS biology* **4**, e256 (2006).
127. Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039–42 (2010).
128. Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome research* **21**, 1131–8 (2011).
129. 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).
130. Birkeland, S. R. *et al.* Discovery of mutations in *Saccharomyces cerevisiae* by pooled linkage analysis and whole-genome sequencing. *Genetics* **186**, 1127–37 (2010).
131. Wenger, J. W., Schwartz, K. & Sherlock, G. Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS genetics* **6**, e1000942 (2010).
132. Swinnen, S. *et al.* Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. *Genome research* **22**, 975–84 (2012).
133. Diezmann, S. & Dietrich, F. S. Oxidative stress survival in a clinical *Saccharomyces cerevisiae* isolate is influenced by a major quantitative trait nucleotide. *Genetics* **188**, 709–22 (2011).
134. Hou, J., Friedrich, A., Gounot, J.-S. & Schacherer, J. Comprehensive survey of condition-specific reproductive isolation reveals genetic incompatibility in yeast. *Nature communications* **6**, 7214 (2015).
135. Connelly, C. F. & Akey, J. M. On the prospects of whole-genome association mapping in *Saccharomyces cerevisiae*. *Genetics* **191**, 1345–53 (2012).
136. Diao, L. & Chen, K. C. Local ancestry corrects for population structure in *Saccharomyces cerevisiae* genome-wide association studies. *Genetics* **192**, 1503–11 (2012).
137. Liti, G. & Louis, E. J. Advances in quantitative trait analysis in yeast. *PLoS genetics* **8**, e1002912 (2012).

VUE D'ENSEMBLE DU PROJET

Projet

L'identification des variations génétiques entre individus d'une même espèce est une étape essentielle pour comprendre les différences phénotypiques observées à différentes échelles au sein d'une population. A ce jour, la relation génotype-phénotype reste peu claire et très difficile à explorer chez l'Homme. Il est donc nécessaire de se tourner vers des organismes modèles plus faciles à étudier. Dans ce contexte, ma thèse a eu pour objectif d'explorer et d'approfondir nos connaissances sur les règles qui gouvernent cette relation en utilisant deux espèces de levures à savoir *Saccharomyces cerevisiae* et *Lachancea kluyveri* qui possèdent les avantages pratiques des micro-organismes tout en étant des eucaryotes. Le thème principal de ce projet a été d'élaborer une stratégie permettant de disséquer efficacement et spécifiquement les origines génétiques de la variation de tolérance à des stress environnementaux et d'évaluer la mesure par laquelle ces traits d'intérêt varient à travers des populations naturelles de *S. cerevisiae* et de *L. kluyveri*.

Bien qu'au cours de ces dernières années, la levure *S. cerevisiae* soit devenue un modèle puissant en génétique quantitative, les origines génétiques de la tolérance à des facteurs de stress n'ont été disséquées que pour un nombre limité de souches et ne reflètent par conséquent pas la diversité phénotypique globale au sein d'une population. Dans le **premier chapitre**, la variabilité phénotypique et l'architecture génétique de la tolérance au stress chez *S. cerevisiae* ont été décrites au sein de la descendance issue de croisements entre 41 souches d'origines variées et la souche de laboratoire $\Sigma 1278b$. Afin d'évaluer la diversité phénotypique, la variation de croissance des descendants a été mesurée face à un panel de 30 conditions de stress choisies de manière à induire diverses réponses physiologiques au niveau de la cellule et ayant un impact par exemple sur la réplication de l'ADN, sur la stabilité des protéines ou encore sur la membrane cellulaire. A partir de l'exceptionnel jeu de données généré, la distribution des phénotypes et la ségrégation dans la descendance ont été analysées de manière à déterminer si la différence phénotypique observée est contrôlée par un gène à effet majeur (caractère Mendélien) ou par plusieurs gènes (caractère complexe). Les critères de l'étude ont par conséquent permis d'estimer pour la première fois la complexité génétique de la variation phénotypique observée à l'échelle d'une espèce. Au total, 8,9 % (98/1105) des cas étudiés ont été identifiés comme Mendéliens et ont fait l'objet d'une étude de l'architecture génétique plus approfondie. Des gènes à effets majeurs ont pu être localisés sur le génome de diverses souches et l'effet phénotypique d'une mutation monogénique en particulier a été évalué dans différents fonds génétiques. De manière

intéressante, une déviation du caractère Mendélien vers des caractères de complexités intermédiaires à élevées a été observée illustrant la complexité cachée d'une mutation monogénique à travers une population naturelle de levures. L'effet d'une telle mutation peut en effet être modulé par l'action de gènes modificateurs selon le fonds génétique. Plus de tels gènes modificateurs sont nombreux et exercent une influence prononcée sur le phénotype, plus le caractère Mendélien se complexifie. Cette étude montre qu'il n'existe pas réellement de distinction claire entre les caractères Mendéliens et les caractères complexes mais plutôt un continuum génétique¹.

Afin d'établir des parallèles importants entre différentes espèces de levures et d'observer des comportements communs ou propres à chaque espèce, il est essentiel de caractériser la variabilité phénotypique dans des paysages génétiques variés. Dans le **second chapitre**, la diversité phénotypique a par conséquent été entreprise chez la levure non conventionnelle *L. kluyveri* à la fois pour la capacité de croissance des souches en réponse à des stress environnementaux et pour les changements de morphologie cellulaire afin de la comparer à celle observée chez *S. cerevisiae*. L'exploration de la variation de tolérance au stress a été réalisée par le suivi de la croissance d'une collection de 27 souches isolées à partir d'habitats naturels dans 55 conditions de croissance dont différentes températures, sources de carbone et composés toxiques. L'analyse de la morphologie des cellules a quant à elle été réalisée sur 501 paramètres extraits à partir d'images de microscopie et reflétant divers aspects morphologiques dont la forme de la cellule, la morphologie nucléaire et le cytosquelette d'actine. De manière intéressante, le comportement général de ces deux phénotypes d'intérêt est très différent au sein et entre espèce et n'est pas défini par l'histoire évolutive de la population ni par l'environnement. Les analyses phénotypiques ont permis de caractériser pour la première fois la diversité qui règne au sein d'une espèce lointainement apparentée à la levure modèle *S. cerevisiae*. Elles constituent une première étape importante dans la compréhension de la relation génotype-phénotype non encore explorée chez *L. kluyveri*².

Compte tenu de sa diversité génétique et phénotypique, *L. kluyveri* a fait l'objet dans le **troisième chapitre** d'une étude sans précédent de la relation génotype-phénotype. Dans le but de détecter avec précision et à moindre coût les gènes impliqués dans la diversité phénotypique et plus particulièrement dans la tolérance au stress, une analyse de pool de ségrégants (Bulk Segregant Analysis) a été privilégiée. Une telle méthode de cartographie repose sur la caractérisation en une seule étape du génotype de centaines voire de milliers de descendants haploïdes exprimant le

phénotype d'intérêt. Chez la levure, les descendants peuvent habituellement être isolés par microdissection de tétrades. Cependant l'inconvénient majeur de cette technique est l'efficacité limitée à générer rapidement une large descendance haploïde pour une analyse de pool de ségrégants. Par conséquent, un marqueur fluorescent spécifique des cellules de signe *MATa* a directement été intégré dans le génome de souches parentales haploïdes afin de rapidement isoler par cytométrie en flux une large population de descendants *MATa* à cartographier ($> 10^5$ cellules).

Les descendants avec des phénotypes extrêmes ont ensuite été sélectionnés en présence de fortes concentrations en sels comme modèles expérimentaux. Le séquençage complet de pools de descendants osmorésistants a permis d'identifier des gènes directement associés à l'halotolérance. Ces résultats prometteurs montrent que l'analyse de pool de ségrégants est une méthode puissante pour cartographier les bases génétiques de la variation phénotypique au sein de levures non conventionnelles et plus précisément chez *L. kluyveri*³.

L'intérêt de mon travail de thèse a résidé à la fois dans l'originalité du thème de recherche et des méthodes utilisées. L'acquisition de nouvelles techniques automatisées de caractérisation phénotypique et génétique ont permis de déterminer rapidement l'identité et les propriétés spécifiques de souches de levures.

Publications relatives à ce travail

1. 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* 39693 (2016). doi:10.1101/039693
2. Jung, P. P., Sigwalt, A., Ohnuki, S., de Montigny, J., Ohya, Y. & Schacherer, J. Large-scale survey of intraspecific fitness and cell morphology variation in a protoploid yeast species. *G3 (Bethesda)* (2016). doi:10.1534/g3.115.026682
3. Sigwalt, A., Caradec, C., Brion, C., Hou, J., de Montigny, J., Jung, PP., Fischer, G., Llorente, B., Friedrich, A. & Schacherer, J. Dissection of quantitative traits by bulk segregant mapping in a protoploid yeast species. *FEMS Yeast Research* (Under review).

CHAPTER 1

The hidden complexity of Mendelian traits across yeast natural
populations

Abstract

Mendelian traits are considered as the lower end of the complexity spectrum of heritable phenotypes. However, more than a century after the rediscovery of Mendel's law, the global landscape of monogenic variants as well as their effects and inheritance patterns within natural populations is still not well understood. Using the yeast *Saccharomyces cerevisiae*, we performed a species-wide survey of Mendelian traits across a large population of isolates. We generated offspring from 41 unique parental pairs, and analyzed 1,105 cross/trait combinations. We found that 8.9% of the cases were Mendelian. Most were caused by common variants showing stable inheritances in a natural population. By contrast, a rare monogenic variant related to drug resistance displayed a significant and variable expressivity across different genetic backgrounds, leading to modified inheritances ranging from intermediate to high complexities. Our results illustrate for the first time the continuum of the hidden complexity of a monogenic mutation, where genotype is hardly predictive of phenotype.

Introduction

To date the genetic causes and complexity underlying the astonishing phenotypic diversity observed in natural populations of *S. cerevisiae* still remain unknown and require further inquiry. In fact, the inheritance patterns of phenotypic traits can be classified as either monogenic or complex. While many traits are complex resulting from variation within multiple genes, their interaction and environmental factors¹, some traits are primarily monogenic and conform to a simple Mendelian inheritance². Nevertheless, while useful, this overly simplistic dichotomic view could potentially mask the continuous level of the underlying genetic complexity³⁻⁵. Indeed, complex traits can be predominantly controlled by variation in a single gene³. Similarly, monogenic traits can be influenced by multiple genes in specific genetic backgrounds such as modifiers⁵⁻⁹. Background effects have been observed in many human disorders^{7,8}, as well as in loss-of-function mutations in various model systems¹⁰⁻¹³ and human cell lines¹⁴⁻¹⁶. However, such specific cases do not reflect the overall genetic diversity and complexity observed in natural populations¹⁷⁻²⁰. More than a century after the rediscovery of Mendel's law, we still lack a global overview of the spectrum of genetic complexity of phenotypic variation within any natural population.

Here, we carried out a first species-wide identification of causal variants of Mendelian traits in the yeast *S. cerevisiae* to characterize in depth their phenotypic effects and transmission patterns across various genetic backgrounds. We generated a large number of crosses using natural isolates, and analyzed the fitness distribution and 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 inheritances across the *S. cerevisiae* species. Interestingly, global phenotypic distribution patterns of multiple Mendelian traits across an extremely large population (~1000 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 <i>Quercus alba</i>	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 <i>Quercus sp.</i>	USA		0.41	35
YPS163	Soil beneath <i>Quercus rubra</i>	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 <i>Quercus sp.</i>	North Carolina, USA		0.43	35
T7	Exudates <i>Quercus sp.</i>	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

Global landscape of Mendelian traits in *S. cerevisiae*.

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 Σ 1278b (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% bactopectone; 1% yeast extract; 2% glucose; 2% agar	Rich medium
YP acetate 2%	2% bactopectone; 1% yeast extract; 2% acetate; 2% agar	Carbon sources
YP EtOH 2%	2% bactopectone; 1% yeast extract; 2% ethanol; 2% agar	Carbon sources
YP glycerol 2%	2% bactopectone; 1% yeast extract; 2% glycerol; 2% agar	Carbon sources
YP sorbitol 2%	2% bactopectone; 1% yeast extract; 2% sorbitol; 2% agar	Carbon sources
YP galactose 2%	2% bactopectone; 1% yeast extract; 2% galactose; 2% agar	Carbon sources
YP ribose 2%	2% bactopectone; 1% yeast extract; 2% ribose; 2% agar	Carbon sources
YP xylose 2%	2% bactopectone; 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 200 μ g/ml	YPD; benomyl 200 μ g/ml	Subcellular organization
YPD benomyl 500 μ g/ml	YPD; benomyl 500 μ g/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 CuSO₄ 10mM	YPD; CuSO ₄ 10mM	Osmotic stress
YPD CuSO₄ 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.

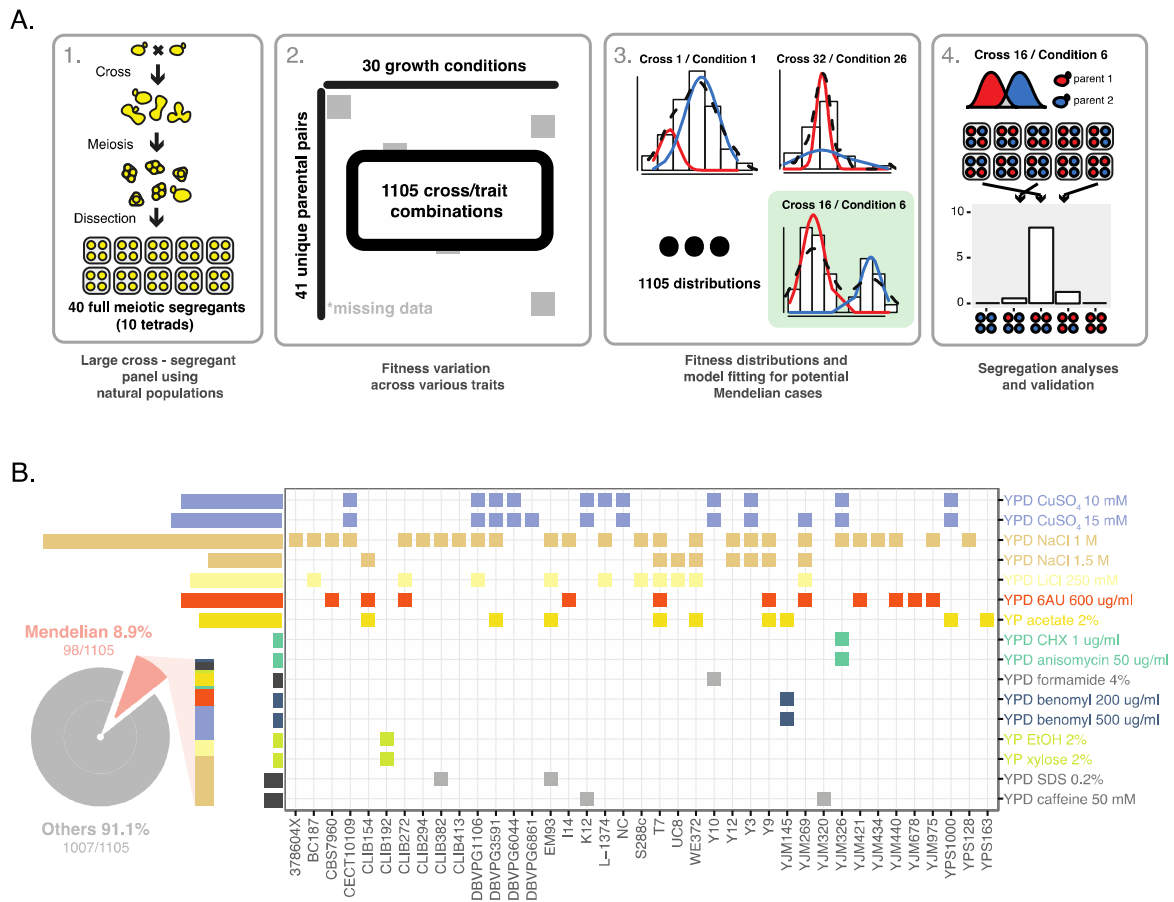


Figure 1. Comprehensive landscape of Mendelian traits in *S. cerevisiae*. (A) Workflow of the detection of Mendelian traits. The workflow was defined as 4 steps, consisting with offspring generation, fitness measurements, model fitting and segregation analysis as indicated. (B) Distribution of all identified 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). 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), CuSO₄ (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 CuSO₄ 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, 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 tolerance²¹, 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 $\Sigma 1278b$ was resistant to both concentrations of CuSO_4 tested and the allelic version of *CUP1* in $\Sigma 1278b$ led to stable Mendelian inheritance across multiple genetic backgrounds (Figure 1B).

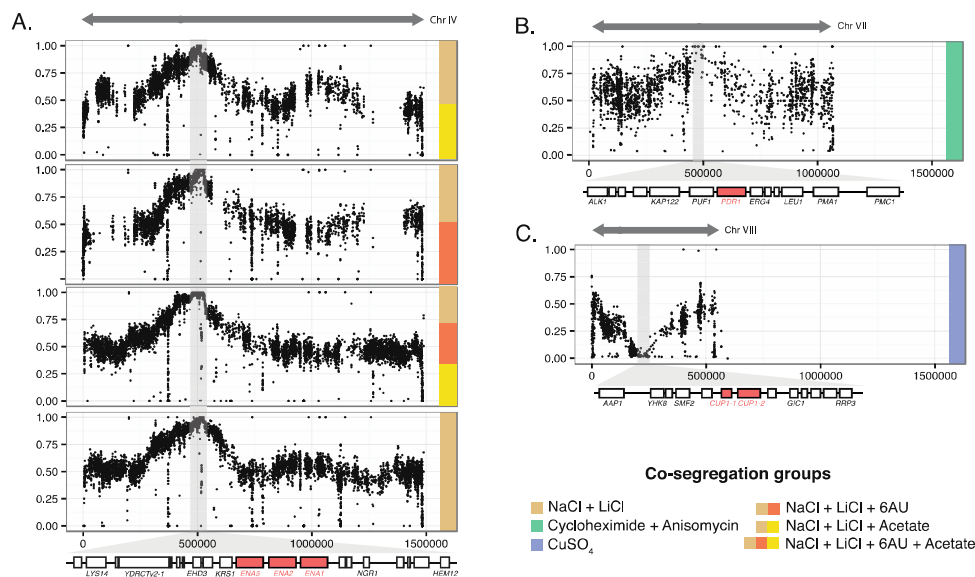


Figure 2. Identification of the genomic regions involved in identified Mendelian traits. Bulk segregant analysis identified causal genomic regions in traits related to co-segregation with NaCl (A), cycloheximide and anisomycin (B) and copper sulfate (C). One chromosomal region with significantly skewed 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 $\Sigma 1278b$. 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 $\Sigma 1278b$ (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 hemizyosity 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.

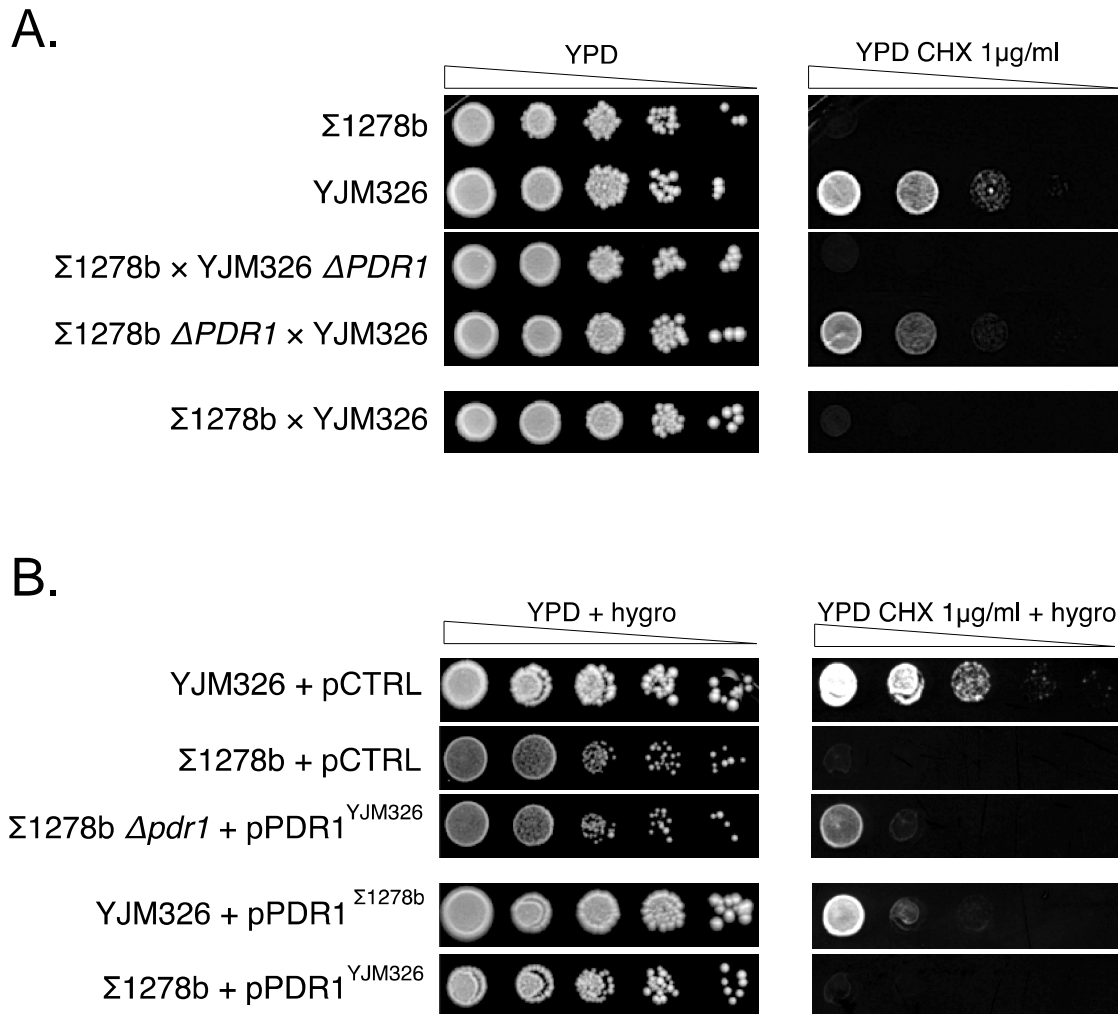


Figure 3. Functional validation of the gene involved in drug resistance. (A) Reciprocal hemizyosity test for the candidate gene *PDR1*. Sensitive (Σ1278b) and resistant (YJM326) parental isolates as well as hybrids that are wild type or hemizygous for the *PDR1* gene are spotted in 5 dilutions onto YPD (left panel) and YPD CHX 1µg/ml (right panel). Cells were grown for 48 hours at 30°C. (B) Ectopic expression of the resistance allele *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 (p*PDR1*^{YJM326}) or sensitive (p*PDR1*^{Σ1278b}) allele was tested in the absence (left panel) or presence (right panel) of cycloheximide. All media were supplemented with 200µg/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 ~1000 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, CuSO₄ and cycloheximide (Figure 4).

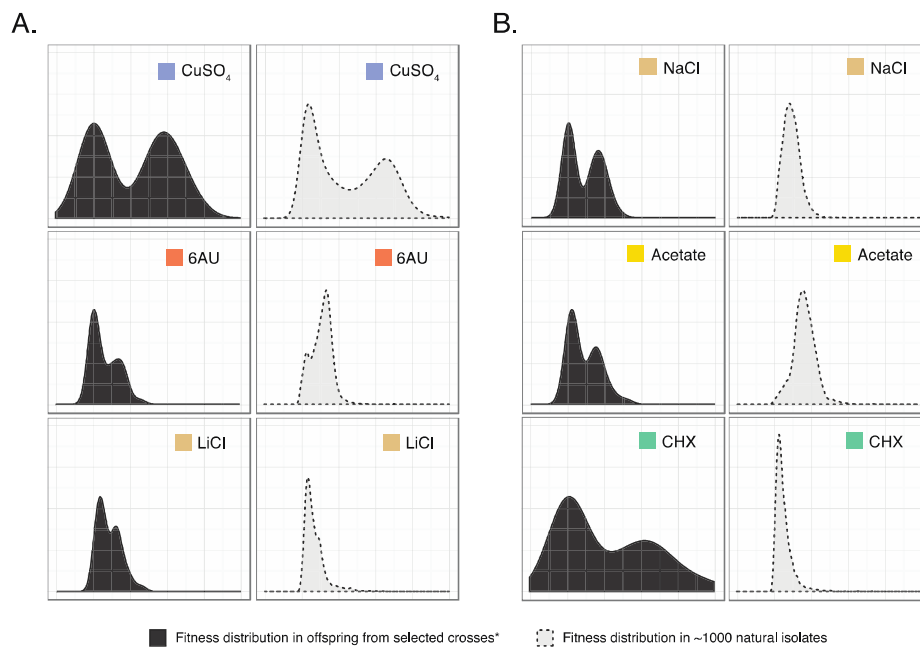


Figure 4. Fitness distribution patterns of identified Mendelian traits within large natural population. Comparisons of the fitness distribution on 6 selected conditions in individual crosses (left panel, N=40) and across ~1000 natural isolates of *S. cerevisiae* (right panel, N=960) are shown. Conditions tested are color-coded. (A) Bimodal distribution patterns observed both in crosses and at the population level. (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 CuSO₄ 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 (Fig. 4B).

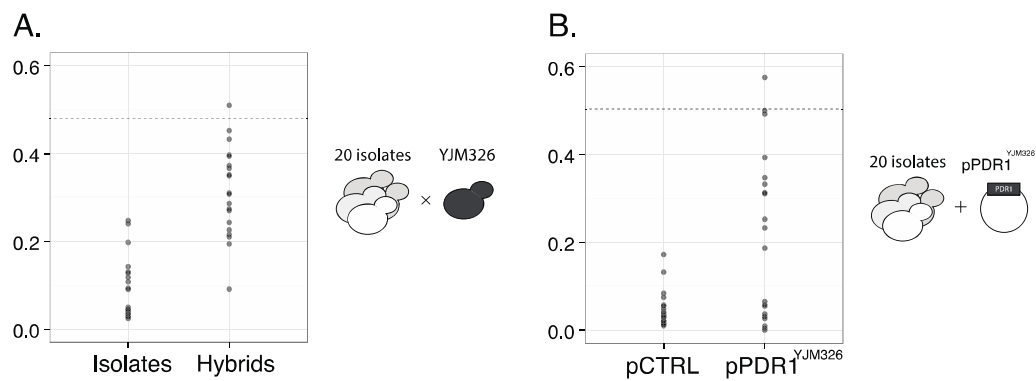


Figure 5. Effects of the *PDR1*^{YJM326} allele in different genetic backgrounds. (A) Fitness variation of 20 isolates (left panel) in comparison with the same set of strains hybridized with YJM326 in the presence of drug. Fitness values (y-axis) correspond to the ratio between the growth in the presence of cycloheximide (YPD CHX 1 μ g/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 1 μ g/ml) with hygromycin to maintain plasmid stability. Dashed line indicates the fitness value of YJM326 carrying the plasmid pPDR1^{YJM326}. See Figure S6 for detailed comparison for effect of hybrid and plasmid in individual genetic backgrounds.

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). Counter intuitively, 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 $\rho = 0.434$), indicating that allelic interactions at the *PDR1* locus only partly accounted for the observed variation (Figure 6).

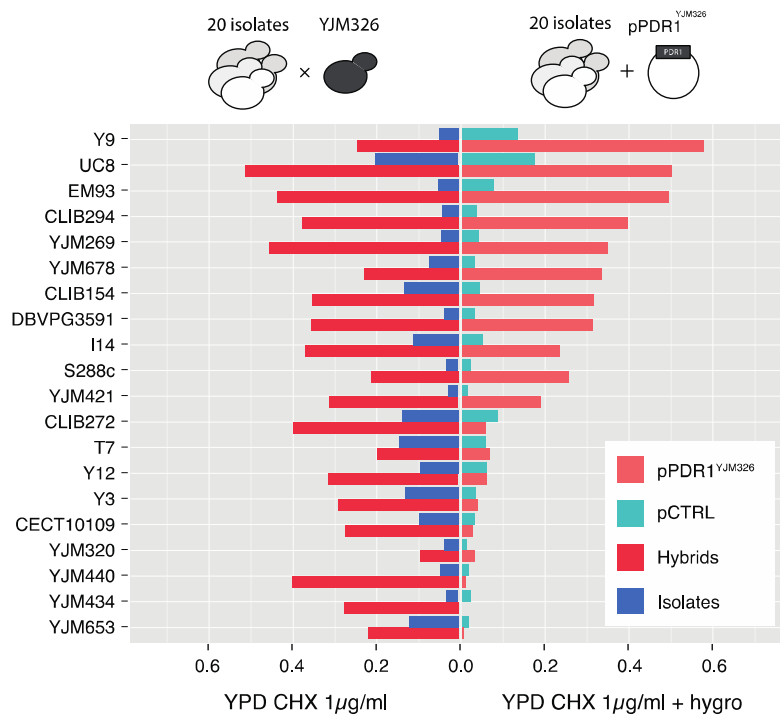


Figure 6. Fitness comparison in strains with ectopic expression of *PDR1*^{YJM326} and in hybrid contexts. The fitness values for 20 isolates in the presence of cycloheximide were compared after hybridizing with the resistant isolate YJM326 (left panel) or after transformed with plasmid carrying the resistant allele pPDR1^{YJM326} (right panel). Strains are indicated on y-axis, with color codes correspond to different configurations (hybrid or plasmid).

The lack of correlation between hybrids and isolates carrying the plasmid with the *PDR1*^{YJM326} allele led us to hypothesize that 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 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.

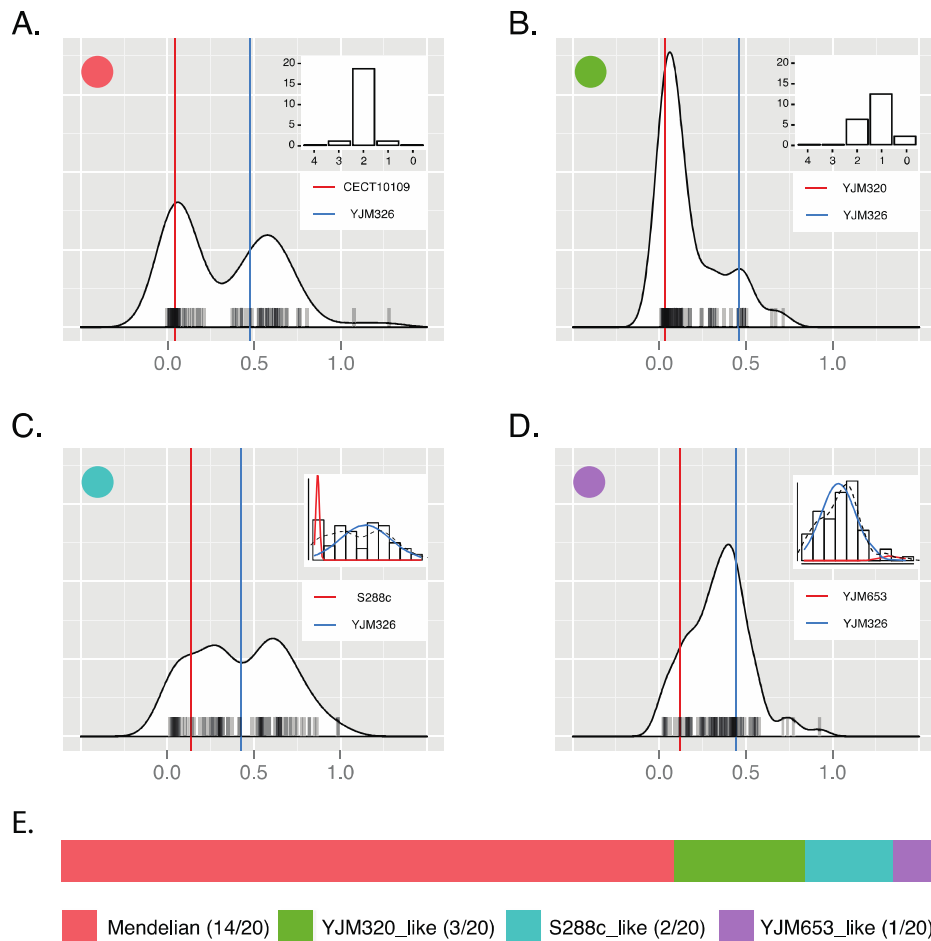


Figure 7. Post-Mendelian inheritance patterns of drug resistance in different hybrid contexts.

(A-D) Offspring fitness distribution patterns observed in hybrids originated from 20 sensitive isolates and YJM326 in the presence of cycloheximide (YPD CHX 1 μg/ml). 80 offspring were tested for each case, and examples of Mendelian (A) and non-Mendelian (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. See Figure S7 for offspring fitness distributions for each of the 20 crosses.

Interestingly, while most of the tested crosses (14/20) displayed Mendelian segregation as was observed in the cross between YJM326 and Σ 1278b, several crosses showed clear deviation of the expected phenotypic distribution (Figure 7). In addition to Mendelian cases, 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 $PDR1^{YJM326}$ 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, which were of medical, industrial and evolutionary interests²²⁻²⁶. 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 identified²⁷. 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 involved²⁸. The stable inheritance patterns of traits caused by alleles with direct phenotypic effect could potentially be 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 salt²⁹, copper^{30,31}, sulfate³² and glucose limitations³³.

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 evidenced 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 genes³⁴, 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* 39693 (2016). doi:10.1101/039693

* These authors contributed equally to this work.

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 conditionspecific 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).

CHAPTER 2

Large-scale survey of intraspecific fitness and cell morphology variation
in a protoploid yeast species

Abstract

It is now clear that the exploration of the genetic and phenotypic diversity of non-model species greatly improves our knowledge in biology. In this context, we recently launched a population genomic analysis of the protoploid yeast *Lachancea kluyveri* (formerly *Saccharomyces kluyveri*), highlighting a broad genetic diversity ($\pi = 17 \times 10^{-3}$) compared to the yeast model organism, *Saccharomyces cerevisiae* ($\pi = 4 \times 10^{-3}$). Here, we sought to generate a comprehensive view of the phenotypic diversity in this species. In total, 27 natural *L. kluyveri* isolates were subjected to trait profiling using the following independent approaches: (i) analyzing growth in 55 growth conditions and (ii) investigating 501 morphological changes at the cellular level. Despite higher genetic diversity, the fitness variance observed in *L. kluyveri* is lower than that in *S. cerevisiae*. However, morphological features show an opposite trend. In addition, there is no correlation between the origins (ecological or geographical) of the isolate and the phenotypic patterns, demonstrating that trait variation follows neither population history nor source environment in *L. kluyveri*. Finally, pairwise comparisons between growth rate correlation and genetic diversity show a clear decrease in phenotypic variability linked to genome variation increase, whereas no such a trend was identified for morphological changes. Overall, this study reveals for the first time the phenotypic diversity of a distantly related species to *S. cerevisiae*. Given its genetic properties, *L. kluyveri* might be useful in further linkage mapping analyses of complex traits, and could ultimately provide a better insight into the evolution of the genotype-phenotype relationship across yeast species.

Introduction

Using a species-wide genetic survey, we previously provided a first general description and characterization of genetic variants involved in Mendelian traits and their phenotypic effects across various genetic backgrounds of the most commonly used yeast model *S. cerevisiae*. However, exploring the genetic basis of trait variation within non-conventional yeast species would probably be valuable and bring in the long term a better insight into the genetic architecture of traits. To date, the degree to which phenotypes vary across non-model Saccharomycotina yeasts has been poorly investigated. In this context, we decided to first characterize the phenotypic diversity within the unexplored yeast species *L. kluyveri* as a first step toward understanding the genotype-phenotype relationship within non-model yeasts species. In this latter, little is known about the phenotypic variation spectrum, despite some known differences comparing to *S. cerevisiae* such as the inability to ferment sugars in the presence of oxygen¹ or the use of pyrimidines and its derivatives as unique nitrogen source^{2,3}.

To gain a better overview of the phenotypic diversity within *L. kluyveri*, we used high-throughput analyses based on growth fitness and cellular morphology. Fitness investigations relied on a microcultivation approach, where the growth of each strain was recorded under a large panel of 55 various growth conditions. In addition, intraspecific cellular morphology characterization was carried out using a single-cell high-dimensional phenotyping based on microscopic images, where 501 morphological parameters were determined⁴. This large-scale analysis provided a first estimation of the phenotypic variation within a non-*Saccharomyces* species. Against all expectations, the growth variation was lower in *L. kluyveri* than in *S. cerevisiae* despite a higher genetic diversity, whereas an opposite trend was found for cellular morphology. Comparison of phenotypic patterns between isolates showed that trait variations (for fitness or morphological traits) follow neither the population evolutionary history nor the source environment. However, fitness profiles are more similar between closely related strains than distant isolates. Altogether, our study provides multiple insights into the phenotypic diversity of *L. kluyveri*, which will be useful to future studies of natural genetic variation in this organism.

Strain	Ecological niches	Geographical origins
62-1041	Willow exudate	USA, California, Davis
CBS 3082	<i>Drosophila pinicola</i>	USA, California
77-1003	Unknown	USA, California
NCYC 543	<i>Drosophila pinicola</i>	USA, California
62-196	<i>Taraxacum officinale</i>	Canada, Saskatoon
CBS 6545	Salix exudate	USA, California
CBS 6546	Exudate of Populus sp.	USA, California
CBS 6547	<i>Drosophila pseudobscura</i>	USA, California, Gualala Creek
CBS 6626	Exudate of tree	Japan, Mt. Takamatsu
NRBC 1892	Tree exudate	Japan
CBS 10367	<i>Quercus Mongolica</i> exudate	Russia, Kedrova pas natural reserve
CBS 10368	<i>Quercus Mongolica</i> exudate	Russia, Sikhote-Alinsky nature reserve
CBS 4104	Soil	Netherlands, Wageningen
68.917-2	Exudate of Populus trichocarpa	Canada
DBVPG 4002	Cavern	Italy
67-588	<i>Ulmus japonica</i> exudate	Japan, Yamabe
NRBC 1811	Partially decaying leaf	Japan
NRBC 10572	Moss	Thailand
NRBC 10955	Decaying leaf of Rhizophora mucronata	Japan
CBS 10369	Unknown	Spain
NRBC 101999	Decaying leaf of Rhizophora mucronata	Japan
CBS 5828	Soil	Denmark
dd281a	Forest	Germany, Heidelberg
CBS 2861	Soil	Sweden
CBS 4568	Soil	Sweden
DBVPG 3452	Soil	Sweden, Ultana
DBVPG 3108	Soil	Netherlands

Table 1. Origin of strains used in this study.

Results

Fitness variation within a natural population of *L. kluyveri*

The ability to grow and the estimation of doubling time in specific environmental conditions (*e.g.* different carbon sources or temperatures) are common parameters allowing for the characterization of the variation between and within species, particularly in yeasts. To obtain an insight concerning the intraspecific variation within the *L. kluyveri* species, we determined the specific growth rate, the lag time, and the yield of biomass produced during the cultivation process of 27 *L. kluyveri* strains (listed in Table 1) grown in 55 environmental conditions, which were classified into 3 main groups: carbon utilization, toxins and environment and metabolites (Table 2). With the exception of the strain 55-86.1, which was a strain prone to the flocculation phenotype, the same set of strains as those used for unraveling the genetic architecture within *L. kluyveri* were used⁵. Overall the reproducibility between biological duplicate is high (Pearson correlation $R=0.95$) suggesting low experimental variation. In YPD condition, the specific growth rates ranged from 0.18 h^{-1} for the strain 68917.2 to 0.72 h^{-1} for the strain NRBC1811. For the reference strain CBS3082a, the doubling time is $\sim 1.15\text{ h}$, which is similar to *S. cerevisiae*.

Condition	Composition	Class
YPD	2% bactopectone; 1% yeast extract; 2% glucose; 2% agar	Rich medium
YP Acetate 2%	2% bactopectone; 1% yeast extract; 2% acetate; 2% agar	Carbon source
YP Ethanol 2%	2% bactopectone; 1% yeast extract; 2% ethanol; 2% agar	Carbon source
YP Ethanol 5%	2% bactopectone; 1% yeast extract; 5% ethanol; 2% agar	Carbon source
YP Galactose 2%	2% bactopectone; 1% yeast extract; 2% galactose; 2% agar	Carbon source
YP Glycerol 2%	2% bactopectone; 1% yeast extract; 2% glycerol; 2% agar	Carbon source
YP Mannitol 2%	2% bactopectone; 1% yeast extract; 2% mannitol; 2% agar	Carbon source
YP Raffinose 2%	2% bactopectone; 1% yeast extract; 2% raffinose; 2% agar	Carbon source
YP Sorbitol 2%	2% bactopectone; 1% yeast extract; 2% sorbitol; 2% agar	Carbon source
YP Succinate 2%	2% bactopectone; 1% yeast extract; 2% succinate; 2% agar	Carbon source
YP Sucrose 2%	2% bactopectone; 1% yeast extract; 2% sucrose; 2% agar	Carbon source
YP Xylose 2%	2% bactopectone; 1% yeast extract; 2% xylose; 2% agar	Carbon source
YPD CaCl ₂ 600 mM	YPD; CaCl ₂ 600 mM	Environment
YPD CaCl ₂ 100 mM	YPD; CaCl ₂ 100 mM	Environment

Cold 23°C	YPD; 23 °C	Environment
YPD CoSO₄ 0.1 mM	YPD; CoSO ₄ 0.1 mM	Environment
YPD CoSO₄ 0.05 mM	YPD; CoSO ₄ 0.05 mM	Environment
YPD CuSO₄ 7.5 mM	YPD; CuSO ₄ 7.5 mM	Environment
YPD CuSO₄ 5 mM	YPD; CuSO ₄ 5 mM	Environment
YPD Ethanol 10%	YPD; ethanol 10 %	Environment
YPD Ethanol 5%	YPD; ethanol 5 %	Environment
Heat 40°C	YPD; 40 °C	Environment
Heat 37°C	YPD; 37 °C	Environment
Heat 38°C	YPD; 38 °C	Environment
YPD LiCl 5 mM	YPD; LiCl 5 mM	Environment
YPD LiCl 1 mM	YPD; LiCl 1 mM	Environment
YPD Methanol 8%	YPD; methanol 8 %	Environment
YPD Methanol 5%	YPD; methanol 5 %	Environment
YPD NaCl 1 M	YPD; NaCl 1 M	Environment
YPD NaCl 0.5 M	YPD; NaCl 0.5 M	Environment
YPD NaCl 0.75 M	YPD; NaCl 0.75 M	Environment
YPD NiSO₄ 10mM	YPD; NiSO ₄ 10 mM	Environment
YNB	6.7% yeast nitrogen base; 2% glucose	Environment
YPD 5-FU 10⁻⁴ M	YPD; 5-FU 10 ⁻⁴ M	Toxins
YPD 5-FU 10⁻⁶ M	YPD; 5-FU 10 ⁻⁶ M	Toxins
YPD 6-azauracil 1250 mg/ml	YPD; 6-azauracil 1250 mg/ml	Toxins
YPD 6-azauracil 500 mg/ml	YPD; 6-azauracil 500 mg/ml	Toxins
YPD Arsenic 0.5 mM	YPD; arsenic 0.5 mM	Toxins
YPD Arsenic 0.05 mM	YPD; arsenic 0.05 mM	Toxins
YPD B-mercaptoethanol 10 mM	YPD; B-mercaptoethanol 10 mM	Toxins
YPD Caffein 2.25 mg/ml	YPD; caffein 2.25 mg/ml	Toxins
YPD Caffein 1.5 mg/ml	YPD; caffein 1.5 mg/ml	Toxins
YPD Cycloheximide 10⁻⁵ M	YPD; cycloheximide 10 ⁻⁵ M	Toxins
YPD Cycloheximide 10⁻⁶ M	YPD; cycloheximide 10 ⁻⁶ M	Toxins
YPD DMSO 8 %	YPD; DMSO 8 %	Toxins

YPD DMSO 6 %	YPD; DMSO 6 %	Toxins
YPD DMSO 4 %	YPD; DMSO 4 %	Toxins
YPD DTT 1.4 mM	YPD; DTT 1.4 mM	Toxins
YPD Nystatin 10 ⁻⁶ M	YPD; nystatin 10 ⁻⁶ M	Toxins
YPD Nystatin 5.10 ⁻⁷ M	YPD; nystatin 5.10 ⁻⁷ M	Toxins
YPD Rapamycin 10 ⁻⁷ M	YPD; rapamycin 10 ⁻⁷ M	Toxins
YPD Rapamycin 10 ⁻⁸ M	YPD; rapamycin 10 ⁻⁸ M	Toxins
YPD SDS 0.03 %	YPD; SDS 0.03 %	Toxins
YPD SDS 0.01 %	YPD; SDS 0.01 %	Toxins
YPD SDS 0.003 %	YPD; SDS 0.003 %	Toxins

Table 2. Detailed media composition used in this study.

As described in both *S. cerevisiae* and *S. paradoxus*, a strong correlation was found when comparing the growth rate and the biomass yield (Pearson correlation $R=0.78$) in most of the experimental conditions tested (with the exception of the substitution of glucose by mannitol or ethanol showing a null correlation). By contrast, the lag phase was independent from these fitness features ($R=-0.11$ and -0.08 for growth rate and yield of biomass, respectively) (Figure 1)⁶. Together, these data suggest that a common genetic architecture governs both growth rate and biomass formation. To gain a better overview of the impact of the different environmental conditions on the fitness, the growth of every strain in each condition was characterized and normalized to the standard YPD medium.

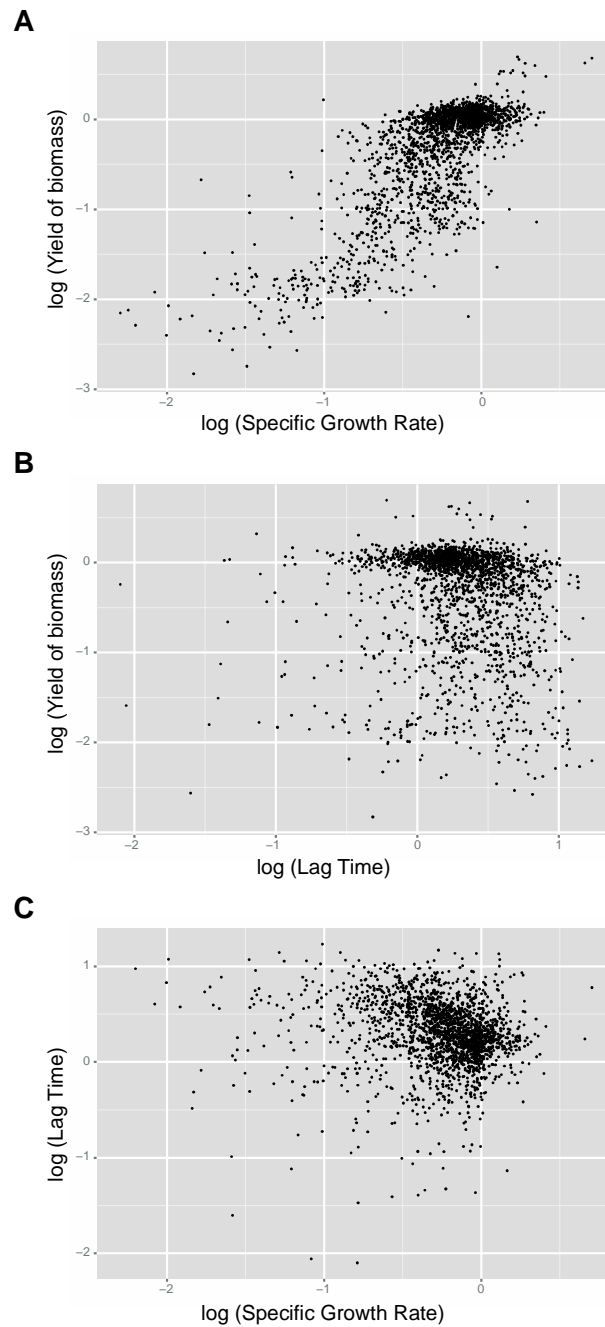


Figure 1. Relationships between specific growth rate, lag phase and yield of biomass. The fitness parameters were extracted from high-density growth curves. A high correlation is found when comparing growth rates and yield of biomass (A) whereas the lag phase feature do not correlate with neither yield of biomass (B) nor growth rates (C).

Barring xylose, growth on fermentable sugars generated little difference in growth compared to glucose. In contrast, non-fermentable sources led to a decrease in the specific growth rates, perhaps due to different uptake mechanisms and metabolization of these sugars (Figure 2). Similarly, 3 subgroups of isolates were found for the environments and metabolites as well as the toxins classes: one corresponded to very low growth reflecting high sensitivity, as it is the case for conditions which alter osmolarity, one corresponded to weak growth such as in the case where temperature was altered, and the last subgroup corresponded to growth similar to that observed in YPD (Figure 2). Strikingly, the 5-FU metabolite generated an enhanced growth rate at concentrations that *S. cerevisiae* strains are sensitive⁷, probably due to the ability of *L. kluyveri* to use pyrimidine nucleotides and derivatives as a nitrogen source⁸.

The estimated variance for both of the 5-FU concentrations was at least twice as high as than in the other tested conditions, suggesting that these isolates are characterized by a very broad resistance to this metabolite. Generally, the variance within the different conditions was low but some growth environments led to higher variation such as succinate or the presence of metal ions (e.g. NiSO₄ or CuSO₄) (Figure 2). This suggests that mechanisms of resistance to specific substrate are variable between different strains. The low variance estimated in the standard YPD condition (0.013) may reflect a very similar growth profile among all isolates in this condition. However, the strain 68.917.2 is a clear outlier as it explains almost 50% of the total variance. Indeed, by removing the growth rate data of 68.917.2, the variance decreased from 0.013 to 0.008. Over all, these data suggest that intraspecies evolution allows for differential resistance to various stresses among some strains.

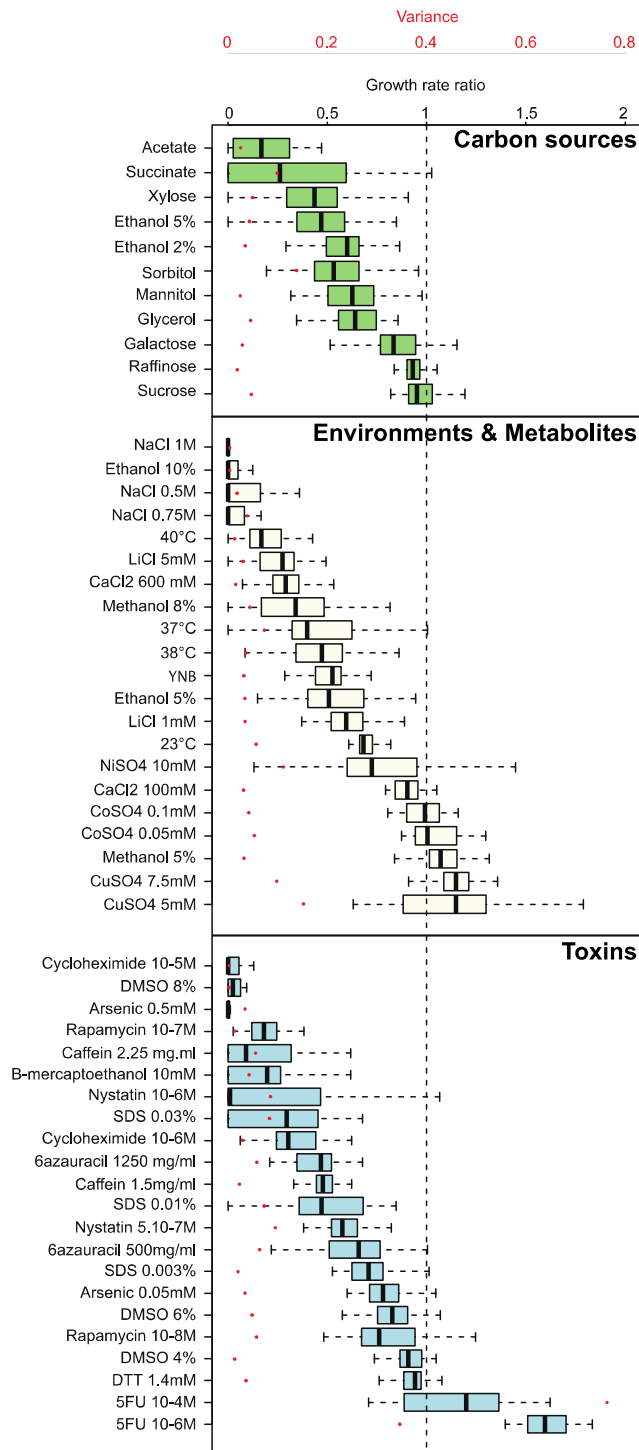


Figure 2. Intraspecific variance of fitness. Growth rates were determined in each condition and normalized to YPD for the 27 strains. A ratio of 1 (dotted line) represents no variation in comparison to YPD. Growth conditions were classified as carbon sources (green), environment and metabolites (ivory) or toxins (blue). The variance among all strains in each condition is depicted by red dots corresponding to the secondary axis. Conditions were ranked according to their median calculated for the 27 strains. Three groups of strains were determined depending on the resistance: low (median < 0.3, dark grey), mid (median between 0.3 and 0.8, intermediate grey) and high resistance (median > 0.8, light grey).

Cell morphology variation within a natural population of *L. kluyveri*

Analysis of morphology includes a set of phenotypic traits among which variation was previously characterized in the budding *S. cerevisiae* yeast⁹. We performed high-dimensional morphological phenotyping of the same 27 *L. kluyveri* isolates after cultivation in complete medium at 30°C. These strains predominantly contained unattached individual cells rather than aggregates of unseparated cells, making it possible to perform semi-automated image analysis with CalMorph⁴. Briefly, each strain was cultivated in five biological replicates. Cells were then fixed with formaldehyde and their cell wall, nuclear DNA, and actin were stained with various specific fluorescent dyes. Images of at least 200 cells per culture were acquired with fluorescent microscopy and analyzed with the CalMorph software to quantify 501 traits reflecting the size, shape, orientation, and intracellular organization of the cells. Despite a low variance of the growth rate determined in complete medium, 384 morphological traits revealed significant intraspecies variability out of the total 501 traits tested ($p < 0.05$ after Bonferroni correction by Kruskal-Wallis test, Figure 3A). Detecting a great number of differences (77%) across the 27 strains suggested that the majority of the morphological organization in *L. kluyveri* is subjected to intraspecies quantitative variation.

One of the most striking variation in morphology was among the whole cell size (Figure 3A). In Figure 3B, the *L. kluyveri* strains were aligned in descending order of cell size with CBS10368 characterized by the largest cell size, approximately 1.7 times larger than that of the smallest strain, CBS3082a. Two additional traits that dramatically varied across strains were related to the nuclear shape (nuclear fitness for ellipse) and the elongated cell shape (mother axis ratio). The values of these three traits ranked strains in three different orders, indicating that there was no correlation among cell size, nuclear shape, and elongated cell shape. Thus, the variation of *L. kluyveri* cellular morphology represents a set of multiple independent traits with different sources of variability.

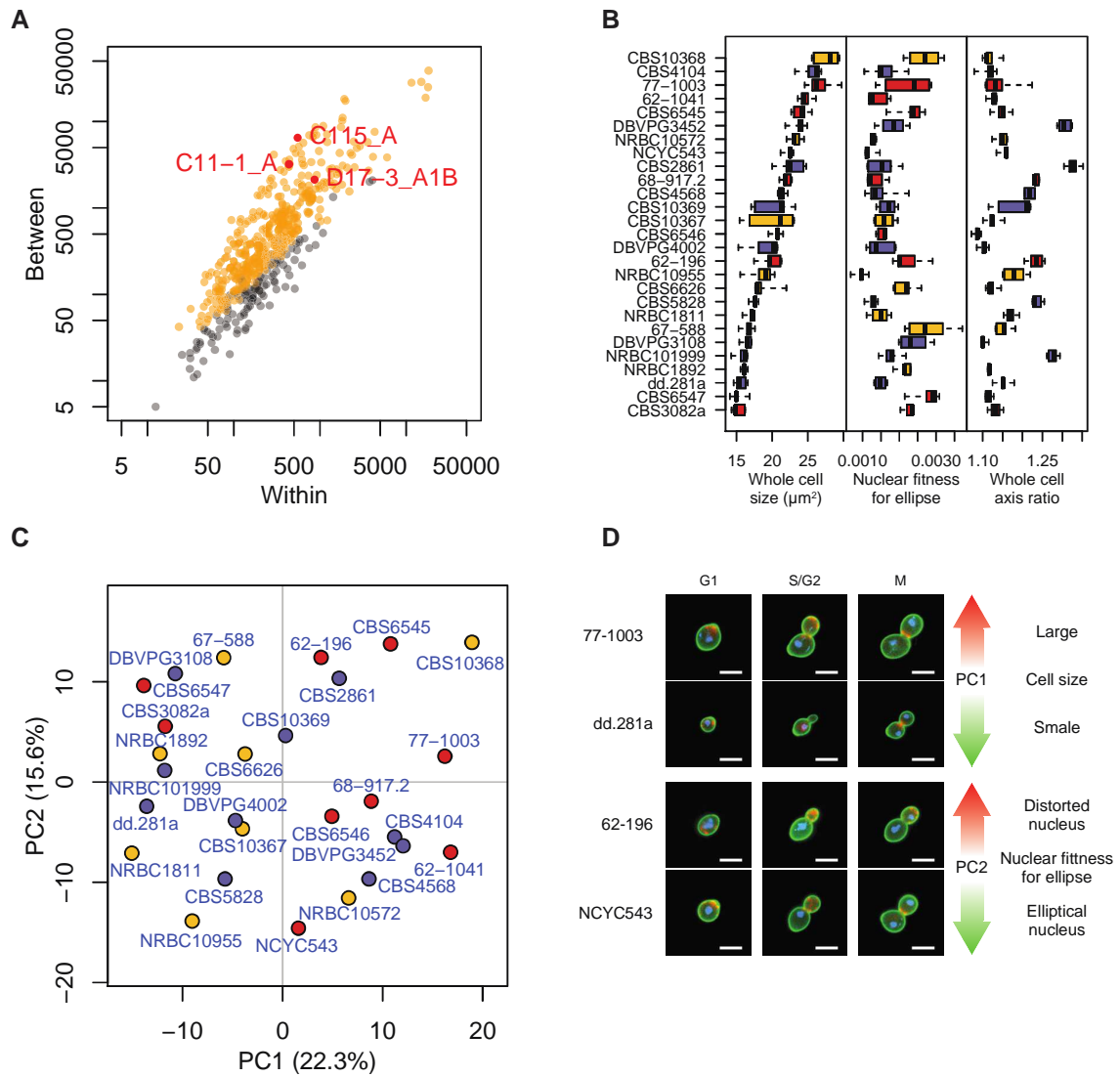


Figure 3. Morphological parameter variation within the *L. kluyveri* species. (A) Intra- versus inter-strain variability of morphological traits. Each dot represents one of 501 measured parameters. Orange and grey distinguish the traits that were called significant and non-significant by the Kruskal-Wallis test at $P < 0.05$ after Bonferroni correction, respectively. For the purpose of visual clarity, each parameter was transformed by $f(x) = (x - \mu_{\text{CBS10367}}) / \sigma_{\text{CBS10367}}$, where μ_{CBS10367} and σ_{CBS10367} are the mean and standard deviation of the parameter across 5 replicates of the CBS10367 strain. Note that significance inference was determined from ranks of raw values and was therefore not affected by this transformation. The sum of squares across replicates (x-axis) and across strains (y-axis) was then computed. The three parameters highlighted in red reflect distinct cellular properties: whole cell size of the G1 cell, nuclear fitness for ellipse of the S/G2 cell, and long over short axis ratio of the ellipse fitted to the G1 cell. (B) Boxplot representation of these morphological parameters calculated for the 27 *L. kluyveri* strains. Box colors represent geographical origins with red for America, blue for Europe, and yellow for Asia. (C) Principal component analysis of *L. kluyveri* morphological variation. Strains are represented by their coordinates along the first two principal components, using the same colors as in Figure 2B. (D) Representative cells illustrating the traits contributing to the first two principal components. Bar: 5 μm .

Trait variation and population history

The genetic divergence within the *L. kluyveri* species ranges up to 2.5%⁵, whereas trait profiling and morphological feature characterization show little variation. Hierarchical clustering of fitness data obtained for the 27 natural isolates of *L. kluyveri* reveal 2 main clusters that are not correlated with either their ecological or geographical origins (Figure 4A).

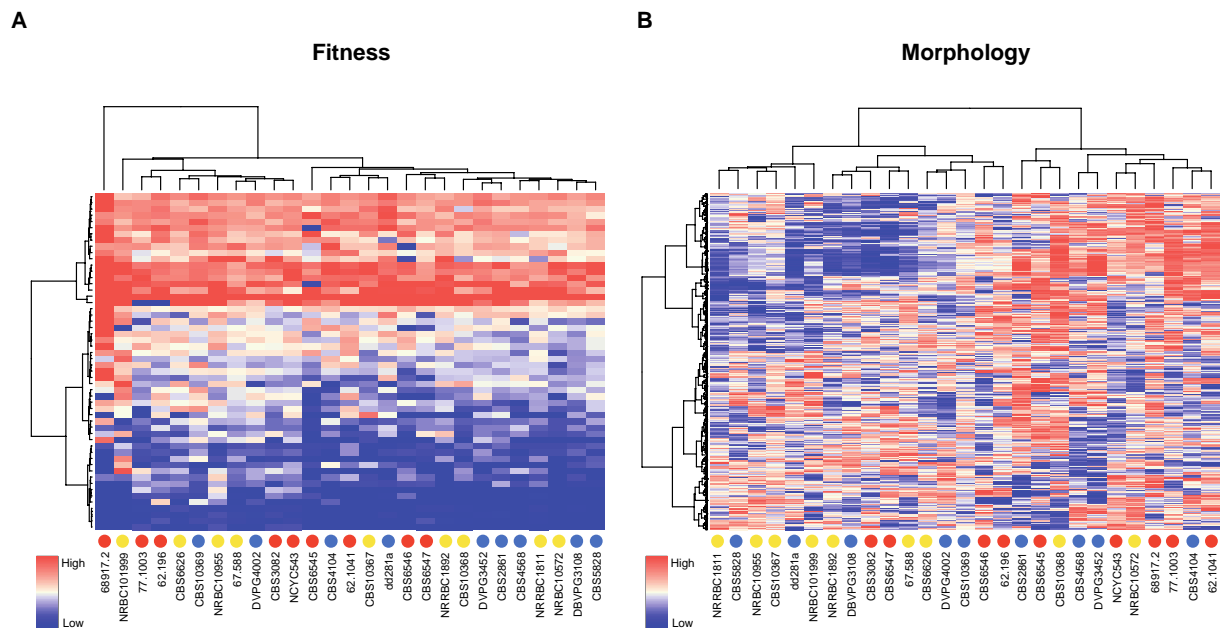


Figure 4. Trait variation within *L. kluyveri* is not related to geographical or ecological niches. Hierarchical clustering of trait profiles was performed using a centered Pearson correlation metric and average linkage mapping for growth (A) as well as morphological parameters (B) analyses. In both heatmaps, rows and columns correspond to conditions (or morphological parameters) and strains, respectively. Growth rates were normalized for each condition to YPD; the high growth rates are depicted in red and slow growth rates in blue in the color scale. Similarly, in the heatmap corresponding to morphological data, low to high morphological score values are depicted by a gradient from blue to red. Colored circles represent geographical origins of each strain: Europe (blue), America (red) and Asia (yellow). Similarly, stars depict ecological origins of each strain: tree exudate and leaf (green), soil (brown), insect gut (purple) and unknown (grey).

Except for the strains isolated from Europe that are mainly found in the group 2 (7 out of 9 strains), the strains are evenly distributed between clusters. It is notable that strain 68.917.2 does not belong to either of the two groups, and appears to be an outlier within the *L. kluyveri* species, probably due to its slow growth on YPD. Hierarchical clustering carried out with the different classes of environmental conditions analyzed here, identified similar groups clearly suggesting that trait evolution is independent of strain niche according to the conditions tested (Figure 5).

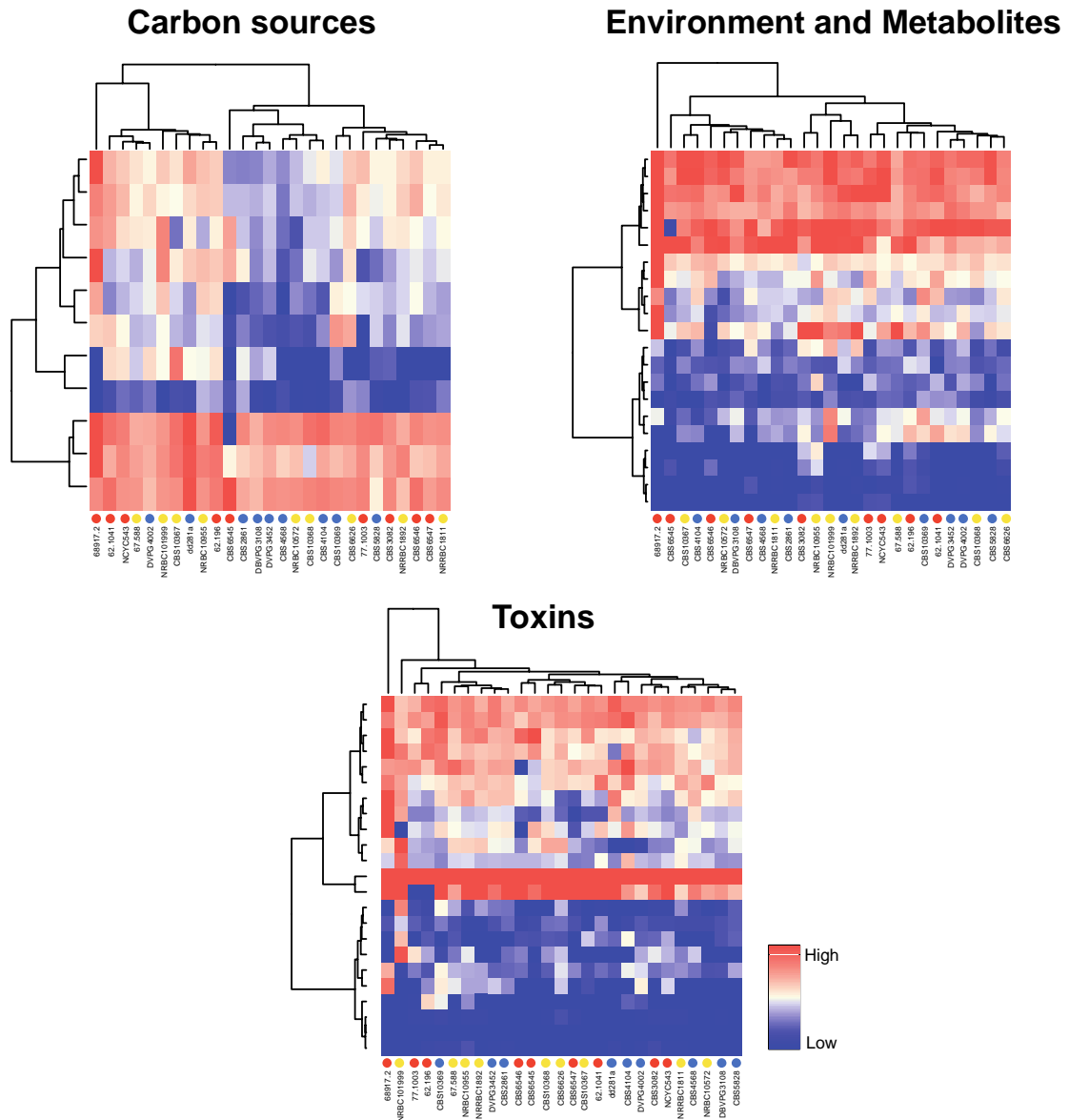


Figure 5. Trait variation with *L. kluveri* according to the environmental condition classes. Hierarchical clustering of trait profiles was performed using a centered Pearson correlation metric and average linkage mapping for carbon sources, toxins and environment and metabolites classes. The color scale represents results where high growth rates are depicted in red and low growth rates in blue.

Similarly, hierarchical clustering of the 501 morphological parameters revealed 2 clusters, each divided into 2 subgroups where no correlation with strain origin was distinguishable (Figure 4B). We then investigated the properties of the variation in morphological features variation using principal component analysis (PCA). The first two components were correlated with cell size and nuclear fitness for ellipse, respectively (Figure 3C and Table 3). We found that the strains were nearly evenly spaced with no particular subgroups apparent. Analysis of all spaces represented by the first five components (63% of the cumulative contribution ratio, Figure 6) led to similar

results (Figure 7), suggesting that *L. kluveri* is characterized by a continuum of morphological features rather than discrete classes of distinct morphologies. In accordance with hierarchical clustering, strains from common ecological origins did not group together, indicating that differences in the morphological traits do not simply reflect adaptation to the simulated environments.

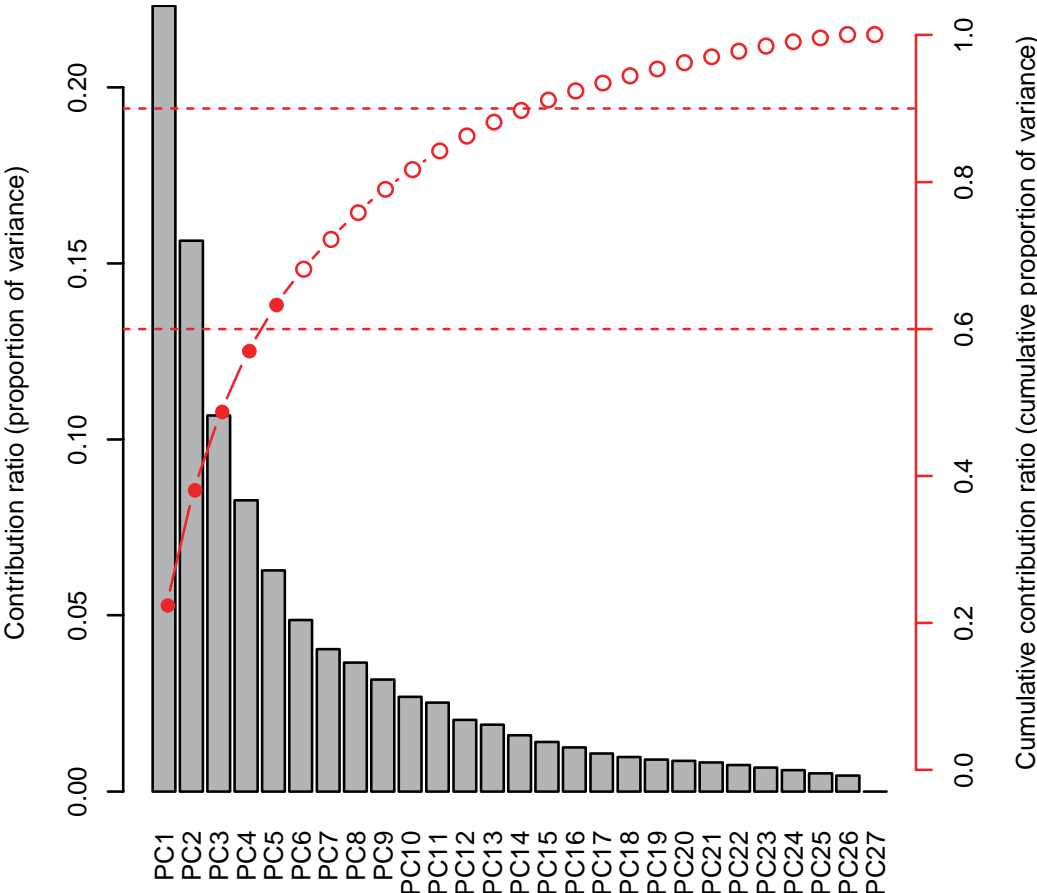


Figure 6. Cumulative proportion of variance of the principal component analysis for the phenotypes. Grey bars indicate the proportion of variance (left axis) explained by the PCs. Red circles indicate the cumulative proportion of variance (right axis) explained by the PCs. The horizontal dashed red line indicates 60% and 90% of the cumulative proportion of variance.

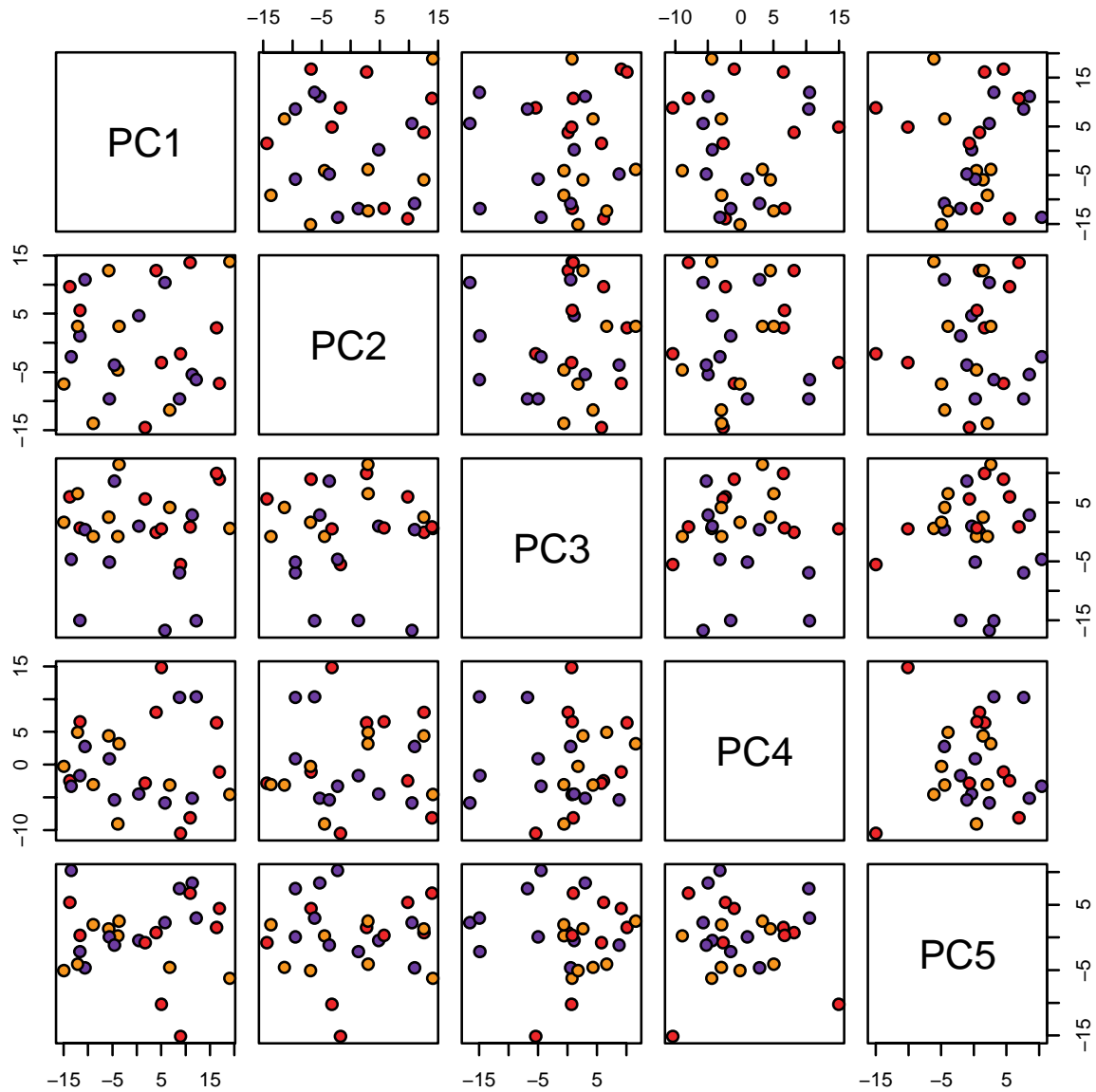


Figure 7. Pair plots of PC scores for the 27 *L. kluveri* species. Principal component analysis of *L. kluveri* morphological variation. Strains are represented by their coordinates along the first five principal components. Box colors represent geographical origins as red with America, purple with Europe, and orange with Asia, using the same colors as in Figure 3B.

Relationship between genetic and phenotypic diversity

According to 881,427 polymorphic sites determined previously within the *L. kluveri* species (Friedrich et al. 2015), genetic architecture led to the definition of 6 subgroups, containing the 27 strains used in this study (Figure 8A). Among these subgroups, there were 2 clearly defined populations: one corresponded to the strains, which had originated from America (group 1) and the second to those isolated from Europe (group 6) (Figure 8A)⁵. The average pairwise comparison of trait profiles within the *L. kluveri* species gave a correlation of 0.8. This parameter

appears to be variable within and between the different groups. Indeed, the estimated correlation in the group 6 is approximately 0.85 whereas it decreases to 0.7 in the group 2 (Figure 8B). In contrast, comparison of morphological parameters revealed an average correlation of nearly 0 (-0.034), due to the significant differences among the 384 cellular features. In spite of this almost null correlation, we were also able to determine variation between and within the different groups, for example between the groups 3 and 4 a weak anti-correlation was estimated ($R = -0.2$) (Figure 8C).

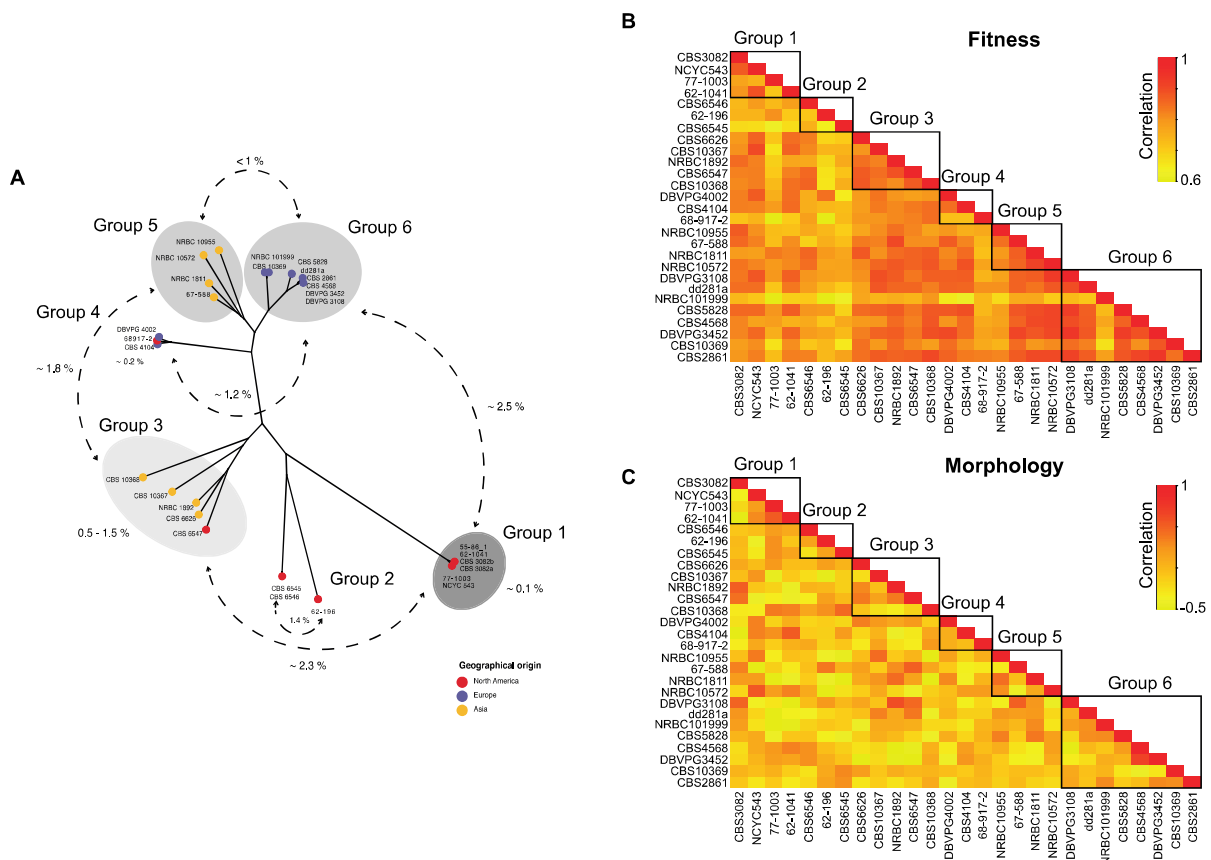


Figure 8. Correlation between and within specific groups of *L. kluyveri*. (A) Genetic diversity within the *L. kluyveri* based on more than 881,427 segregating sites leads to the distinction of 6 different subgroups. Pairwise comparison for growth rate (B) and morphological parameters (C) revealed high correlation to fitness independent of the subgroups, whereas no clear correlation was found for morphological data. The color scale represents levels of correlation where high correlation is depicted in red and low correlation in yellow.

To better understand how genetic variation drives phenotypic variability, we determined the relationship between genetic diversity and trait profiles. Strains belonging to the group 1 were

characterized by less similar trait profiles (average correlation of 0.8) despite their high genetic conservation. However, a significant anti-correlation (Kendall's test, p -value = 3.94×10^{-9}) was found when comparing fitness data among all the 27 strains, indicating that trait variation increases with the enhancement of genetic diversity (Figure 9A). In contrast, no such correlation was determined between morphological features and genetic variation (Kendall's test, p -value = 0.296) (Figure 9B). These results suggest that genetic diversity within *L. kluyveri* influences trait profiles but cannot explain all of the phenotypic variability given the lower correlation among North American strains.

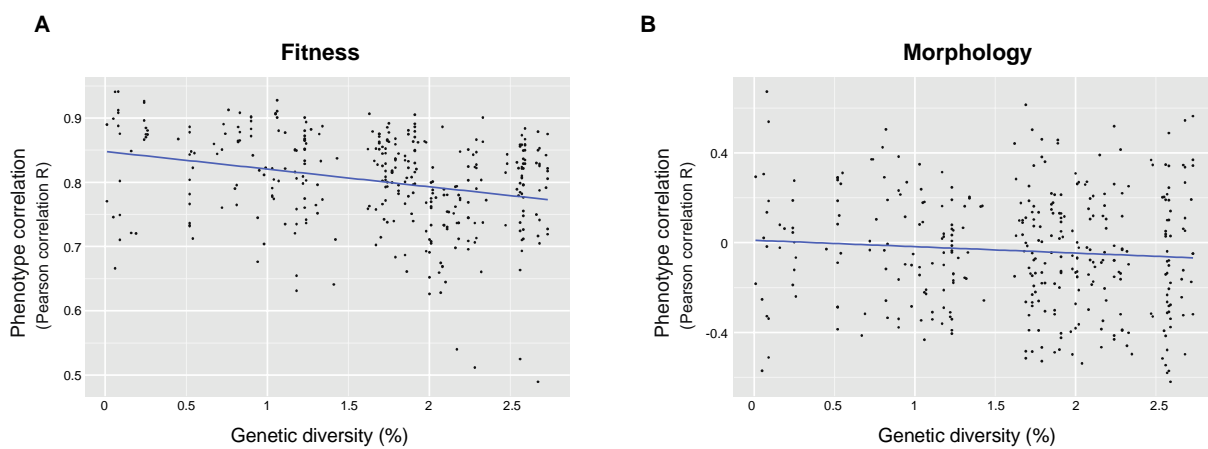


Figure 9. Correlation between genetic diversity and trait profiles variation. Pairwise genetic diversity is plotted against pairwise Pearson correlation calculated for both fitness (A) and morphological parameters (B). The blue line corresponds to linear regressions.

Comparison of trait variation across different yeast species

Trait variation analysis within *S. cerevisiae* revealed a correlation with their genomic structure rather than to adaptation associated with their ecological niche. For example, *S. cerevisiae* strains isolated from Europe are more prone to be resistant to Na^+ and Li^+ ions given the pleiotropic roles of causal loci. In addition, clean lineages within the *S. paradoxus* species show similar trait profiles⁶. As mentioned above, we previously identified 2 clean lineages within the *L. kluyveri* species: one corresponding to strains originated from North America (group 1), the second comprised of strains isolated from Europe (group 6)⁵. In contrast to *S. cerevisiae*, trait profile and morphological feature analyses revealed variation independently from the genetic structure of the *L. kluyveri* species. Moreover, the evolution of resistance to various stresses including heavy

metals or high osmolarity is not related to strain niche. The lack of correlation between growth phenotype or cellular parameters on one hand, and the population structure or the origins of the strains on the other could be related to the small number of isolates available from *L. kluyveri*. Despite this low number of strains, the overall genetic diversity is greater than 2.5% whereas it is approximately 0.5% for *S. cerevisiae*. In sharp contrast, the phenotypic variance is lower in *L. kluyveri* than in *S. cerevisiae* (Figure 10), which has also been described for both *S. paradoxus* and *S. pombe* species^{6,10}. However, morphological variance determined in this study for *L. kluyveri* shows a clear trend where it is much higher than in *S. cerevisiae* (Figure 10)⁹. Altogether, these data suggest that cellular morphology and growth are clearly two independent features given the impact of genetic diversity on their variation. Nevertheless, the small number of strains analyzed could lead to biased results and a larger collection would be ideal to definitely characterize morphological variation.

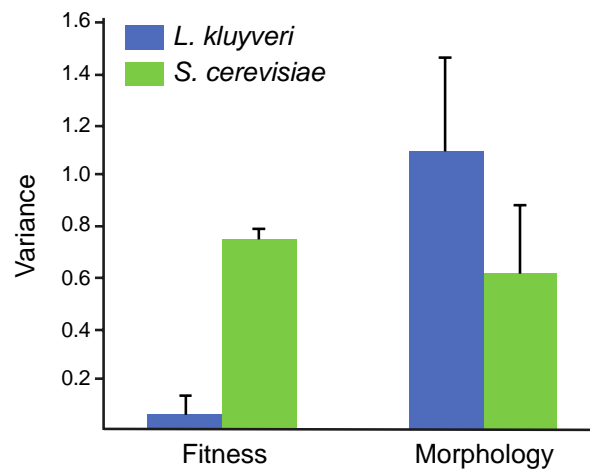


Figure 10. Fitness and morphological parameters variance. Comparison of phenotypic variance within *L. kluyveri* (blue) and *S. cerevisiae* (green) for growth data and morphological parameters.

Discussion

Assessing the diversity among individuals within a species is now feasible for non-model organisms and essential for understanding population history as well as obtaining a better insight into the genotype-phenotype relationship¹¹. The Saccharomycotina subphylum (budding yeasts), which includes the baker's yeast *S. cerevisiae*, includes ideal non-model organisms to explore diversity in terms of genetic and phenotypic variation¹²⁻¹⁴. In this study, we investigated phenotypic variation within the protoploid *L. kluyveri* species by analyzing variation in both growth and morphological parameters, allowing for a comparison with previous studies performed in *S. cerevisiae*^{9,6}. Interestingly, the general behavior of these two categories of phenotype is very different within as well as between species.

Among the 501 morphological features investigated in logarithmically growing cells, 384 showed significant variation among the strains studied suggesting that these traits are subjected to large intraspecific variation within *L. kluyveri*. Similarly, 440 morphological parameters were previously described as significantly variable among 37 strains of *S. cerevisiae*⁹. Comparison of cell morphology between *L. kluyveri* and *S. cerevisiae* allowed us to clearly distinguish these two species according to the size of the nuclei or the buds, for instance. *L. kluyveri* has an extremely high variance in cell morphology of 1.1 compared to 0.62 determined with previously published data generated in *S. cerevisiae*⁹. By contrast, growth analysis of the same set of *L. kluyveri* isolates in 55 different environmental conditions showed a very low variance (0.06) demonstrating that growth profiles are more similar than that for *S. cerevisiae* species where the variance was estimated to be approximately 0.75⁶.

In contrast to fitness patterns in *S. cerevisiae*, trait variation (for fitness and cell morphology) is not defined by population history in *L. kluyveri*. Using PCA analysis and hierarchical clustering, we did not observe any correlation between the origin (ecological or geographical) and the phenotypic patterns, showing that trait variations are not linked to either the population structure or the source environment. However, the number of *L. kluyveri* strains used in this study is relatively low and their ecological origins are somehow similar. Increasing the number of strains would probably reveal more structured trait profiles.

Finally, we determined the correlation between pairwise trait variation and genetic diversity in order to have a better view of how genomes and phenotypes are related. Interestingly, there is strong evidence that genetic diversity is anti-correlated with fitness variation. However, this is not the case for differences among morphological parameters, showing a different global behavior of

the two types of traits. We are here considering genetic diversity as SNPs but other genome variation can lead to trait profile differences. Genomes of the 27 strains of *L. kluyveri* were obtained through mapping against a reference genome⁵. Although this method gives deep information on genome variation, important features are still missing such as transposon content, structural variants or the presence of specific genes for a distinct group of strains (pan genome). Moreover, prion proteins can also serve as a motor for phenotypic inheritance in *S. cerevisiae* but are still not well known in *L. kluyveri*¹⁵. A better knowledge of these features would help to gain a better understanding of phenotypic diversity.

Given the genetic characteristics of *L. kluyveri* and its trait variability, quantitative genetics is definitively applicable to this species and would lead to a better understanding of the genotype-phenotype relationship across diverse yeasts. Moreover, laboratory evolution experiments carried out for both *S. cerevisiae* and the pathogenic *Candida albicans* species have shown that the presence of additional chromosome is linked to stress resistance acquisition^{16,17}. Similar experiments in *L. kluyveri* would reveal if the acquisition of aneuploidy is a common mechanism across yeast species to improve fitness in stress condition. Finally, further analysis will without a doubt provide deeper insight into the functional diversity in this non-model organism.

Publication related to this chapter

Jung, P.P.*, Sigwalt, A.*, Ohnuki, S.*, de Montigny, J., Ohya, Y. & Schacherer, J. Large-Scale Survey of Intraspecific Fitness and Cell Morphology Variation in a Protoploid Yeast Species. *G3 (Bethesda)* (2016). doi:10.1534/g3.115.026682

* These authors contributed equally to this work.

References

1. Møller, K., Langkjaer, R. B., Nielsen, J., Piskur, J. & Olsson, L. Pyruvate decarboxylases from the petite-negative yeast *Saccharomyces kluyveri*. *Mol. Genet. Genomics* **270**, 558–568 (2004).
2. Gojkovic, Z., Paracchini, S. & Piskur, J. A new model organism for studying the catabolism of pyrimidines and purines. *Adv. Exp. Med. Biol.* **431**, 475–479 (1998).
3. Beck, H., Dobritzsch, D. & Piškur, J. *Saccharomyces kluyveri* as a model organism to study pyrimidine degradation. *FEMS Yeast Research* **8**, 1209–1213 (2008).
4. Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 19015–19020 (2005).
5. 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**, 184–192 (2015).
6. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS Genet.* **7**, e1002111 (2011).
7. Jaquet, L., Lollier, M., Souciet, J. L. & Potier, S. Genetic analysis of yeast strains lacking negative feedback control: a one-step method for positive selection and cloning of carbamoylphosphate synthetase-aspartate transcarbamoylase mutants unable to respond to UTP. *Mol. Gen. Genet.* **241**, 81–88 (1993).
8. Andersen, G. *et al.* Catabolism of pyrimidines in yeast: a tool to understand degradation of anticancer drugs. *Nucleosides Nucleotides Nucleic Acids* **25**, 991–996 (2006).
9. Yvert, G. *et al.* Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst Biol* **7**, 54 (2013).
10. Brown, W. R. A. *et al.* A Geographically Diverse Collection of *Schizosaccharomyces pombe* Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3 (Bethesda)* **1**, 615–626 (2011).
11. Peter, J. & Schacherer, J. Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* **33**, 73–81 (2016).
12. Freel, K. C., Friedrich, A., Hou, J. & Schacherer, J. Population genomic analysis reveals highly conserved mitochondrial genomes in the yeast species *Lachancea thermotolerans*. *Genome Biol Evol* **6**, 2586–2594 (2014).

13. Freel, K. C., Friedrich, A. & Schacherer, J. Mitochondrial genome evolution in yeasts: an all-encompassing view. *FEMS Yeast Res.* **15**, fov023 (2015).
14. Freel, K. C. *et al.* Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea quebecensis*. *Genome Biol Evol* (2016). doi:10.1093/gbe/evv262
15. Halfmann, R. *et al.* Prions are a common mechanism for phenotypic inheritance in wild yeasts. *Nature* **482**, 363–368 (2012).
16. Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B. & Berman, J. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet.* **5**, e1000705 (2009).
17. Yona, A. H. *et al.* Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 21010–21015 (2012).

CHAPTER 3

Dissection of quantitative traits by bulk segregant mapping in a
protoploid yeast species

Abstract

Since more than a decade ago, *Saccharomyces cerevisiae* has been used as a model to dissect complex traits, revealing the genetic basis of a large number of traits in fine detail. However, to have a more global view of the genetic architecture of traits across species, the examination of the molecular basis of phenotypes within non-conventional species would undoubtedly be valuable. In this respect, the Saccharomycotina yeasts represent ideal and potential non-model organisms. Here we sought to assess the feasibility of genetic mapping by bulk segregant analysis in the protoploid *Lachancea kluyveri* (formerly *Saccharomyces kluyveri*) yeast species, a distantly related species to *S. cerevisiae*. For this purpose, we designed a fluorescent mating-type marker, compatible with any mating-competent strains representative of this species, to rapidly create a large population of haploid segregants ($> 10^5$ cells). Quantitative trait loci can be mapped by selecting and sequencing an enriched pool of progeny with extreme phenotypic values. As a test bed, we applied this strategy and mapped the causal loci underlying halotolerance phenotypes in *L. kluyveri*. Overall this study demonstrates that bulk segregant mapping is a powerful way for investigating the genetic basis of natural variations in non-model yeast organisms and more precisely in *L. kluyveri*.

Introduction

In the previous chapter, the large-scale phenotypic survey on growth ability in response to different stresses and cellular morphology changes revealed a broad phenotypic diversity within the *L. kluyveri* yeast species. Nevertheless, at this stage of the study, the genetic origins of the phenotypic variation in *L. kluyveri* remain unknown and deserve to be explored. In this context, we sought to lay the foundation for the exploration of the genotype-phenotype relationship within this non-model yeast species. Mapping and identification of Quantitative Trait Loci (QTLs) has been enhanced by bulk segregant analysis coupled with whole genome sequencing (BSA-seq), well suited to *S. cerevisiae*¹⁻⁶. One such method relies on the simultaneous analysis of several hundreds, even thousands, of recombinant descendants (segregants) with extreme phenotypes⁷. In yeast, segregants can be classically isolated by manual tetrad dissection. But the time consuming nature of the traditional manual method makes difficult its utilization in large scale. Different systems were thereby designed based on the introduction of fluorescent reporter genes to permit rapid isolation of labeled tetrads⁸⁻¹⁰ or individual spores^{11,12} by fluorescence-activated cell sorting (FACS).

Here, we performed a first QTL mapping using a bulk segregant strategy in the non-explored yeast species *L. kluyveri*. The approach is based on the stable integration of a fluorescent mating-type marker into the genome of haploid strains. We demonstrated that the fluorescent marker is well suited for isolation of a large mapping population of haploid cells by FACS and can be used as a mating-type indicator. Using BSA coupled with whole-genome sequencing, we identified QTLs underlying halotolerance phenotypes as test cases. Altogether, the genetic approach developed in this framework could be useful for future studies of the genotype-phenotype relationship within *L. kluyveri* as well as in other yeast species.

Results

Description of the developed strategy

Detection accuracy of quantitative trait locus is related to single nucleotide polymorphism (SNP) marker density and homogeneous repartition of such markers throughout the genome, it was therefore essential to select parental strains with sufficient genetic divergence. Population genomic analysis has enabled to estimate the genetic diversity among 28 haploid strains of *L. kluyveri* ranging from ~ 0.1 % to ~ 2.5 %¹². Among this sample, the *MAT α* strain NBRC10955 and the *MAT α* strain 67-588 were retained to illustrate the principle of the method (Table 1).

Strain	Ecological niches	Geographical origins	Ploidy	Mating-type
62-1041	Willow exudate	USA, California, Davis	n	<i>MATα</i>
CBS 3082	<i>Drosophila pinicola</i>	USA, California	2n	
NCYC 543	<i>Drosophila pinicola</i>	USA, California	n	<i>MATα</i>
67-588	<i>Ulmus japonica</i> exudate	Japan, Yamabe	n	<i>MATα</i>
NRBC 10955	Decaying leaf of <i>Rhizophora mucronata</i>	Japan	n	<i>MATα</i>

Table 1. Origin, ploidy and mating-type of strains used in this study.

The genetic divergence between the two strains reaches 0.7 %, corresponding to a total of 85,705 polymorphic sites, with an average density of one SNP every 130 pb. In addition, the cross led to an acceptable offspring viability of 73 % on standard rich medium YPD that avoid eventual biases, which could be due to chromosomal rearrangements or genetic incompatibilities^{5,6}, during QTL analysis. The rate-limiting step in BSA being the generation of a large descendance, an approach was designed to rapidly and efficiently isolate a mapping population (Figure 1).

First, a fluorescent mating-type marker was stably integrated into the genome of selected parental strains NBRC10955 and 67-588. Second, FACS was used to select thousands of fluorescently labeled haploid progeny resulting from the cross between the two parental strains. Third, the isolated pool of segregants was placed under osmotic pressure to enrich the more resistant segregants with beneficial alleles. Finally, the resulting subset of osmoresistant segregants was sequenced to determine the genetic origins of halotolerance in *L. kluyveri* as proof of principle.

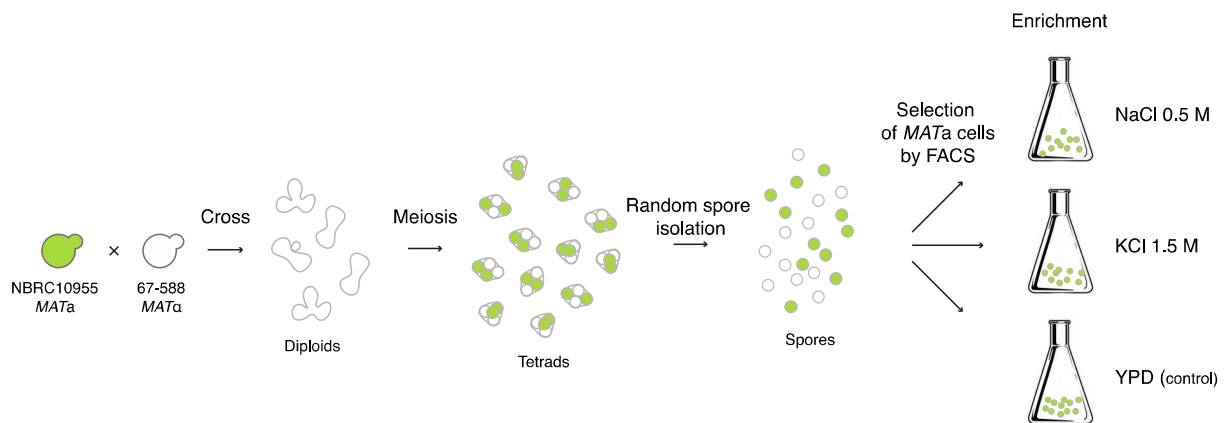


Figure 1. Workflow of the mapping strategy. The a-type-specific promoter *STE2p* was fused to the fluorescent marker *sfGFP* (superfolder GFP) and introduced in both strains NBRC10955 and 67-588. This construct results in green fluorescence in *MATa* cells. After random spore isolation the marker is expressed and enables the selection of *MATa* populations of segregants.

Fluorescence-based isolation of a large descendance

To design the fluorescent mating-type marker, it was primordial to place the reporter gene under the control of a mating-type specific promoter. In *S. cerevisiae*, *STE2* gene is well known to be specifically expressed in *MATa* haploid cells¹⁴. This gene encodes a receptor of α -factor pheromone located on the surface of *MATa* cells that is essential for the mating process and similar to that found in *L. kluyveri*¹⁵. In order to confirm the specificity of *STE2* gene for *MATa* cells in *L. kluyveri*, a reverse transcription polymerase chain reaction (RT-PCR) was performed. The presence or the absence of a *STE2* transcript in *MATa/MATα* (CBS3082 2n), *MATα* (62-1041 n) and *MATa* (NCYC543 n) strains was ascertained (Table 1, Figure 2).

The *STE2* gene was effectively expressed as a full transcript only in the *MATa* strain and repressed in *MATα* and *MATa/MATα* strains. As a result, the a-type-specific promoter *STE2p* of *L. kluyveri* was fused to the fluorescent marker *sfGFP* (superfolder GFP) coupled to a selection marker in order to fluorescently label *MATa* cells. The *KanMX* resistant cassette was preferred to an auxotrophic marker to avoid undesirable large phenotypic effects. Manual tetrad dissection being clearly laborious in *L. kluyveri* due to a high compaction of tetrads and to spores tightly linked by interspore bridges, the choice of the target gene for the integration of the fluorescent mating-type marker into the genome of parental strains NBRC10955 and 67-588 fell on the non-essential gene *CHS3* (Chitin synthase III)¹⁶. Indeed, the deletion of *CHS3* gene involved in chitin

formation of outer layers of spore wall promotes a better separation of sister spores in *L. kluyveri*. Consequently, the *CHS3* gene was deleted and replaced by the fluorescent mating-type marker in the two parental strains NBRC10955 and 67-588 with a view to label *MATa* cells and to facilitate the isolation of spores.

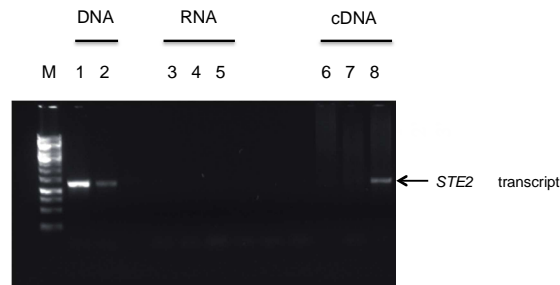


Figure 2. RT-PCR analysis of *STE2* transcript. Agarose gel electrophoresis showing *STE2* product (indicated by an arrow) amplified by PCR from CBS3082 *a/α* DNA as control (lines 1 and 2), from CBS3082 *a/α*, 62-1041 *α* and NCYC543 *a* RNA, as control to ensure the absence of DNA (lines 3, 4 and 5 respectively) and *STE2* transcript amplified by RT-PCR from CBS3082 *a/α*, 62-1041 *α* and NCYC543 *a* cDNA (lines 6, 7 and 8 respectively) with *STE2*-specific primers. M : 250 bp DNA Ladder.

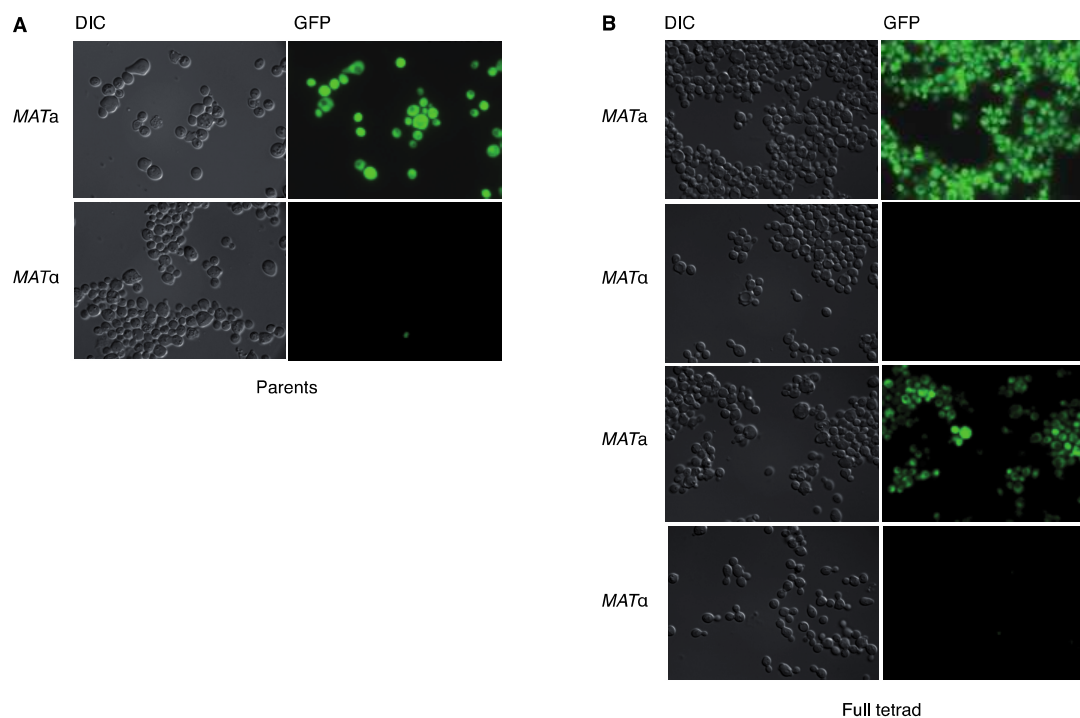


Figure 3. Validation of the fluorescent mating-type marker. Cells were visualized by fluorescence microscopy to monitor the expression of the fluorescent mating-type marker *sfGFP*. A. Parental strains NBRC10955 *MATa chs3Δ::STE2p-sfGFP-KanMX* and 67-588 *MATα chs3Δ::STE2p-sfGFP-KanMX*. B. Four germinating spores after tetrad dissection of *MATa chs3Δ::STE2p-sfGFP-KanMX / MATα chs3Δ::STE2p-sfGFP-KanMX* diploid.

To validate the fluorescent mating-type marker, mutant parental strains and the offspring resulting from the cross were observed by fluorescence microscopy. As expected, the presence of the fluorescent marker into parental genomes resulted in *MATa* specific expression of *sfGFP* and no expression in the *MAT α* parent (Figure 3A). After sporulation of diploid between mutants NBRC10955 and 67-588, traditional tetrads dissection and germination, *MATa* haploid progeny expressed the fluorescent marker (Figure 3B). We observed a Mendelian segregation of the marker in each tetrad (N = 11), *i.e.* 2 *sfGFP*⁺ and 2 *sfGFP*⁻ spores corresponding to *MATa* and *MAT α* cells, respectively. These results suggested that *MATa* haploid segregants could be isolated by FACS using *sfGFP* fluorescent marker expression as an indicator of *MATa* mating-type.

Effect of the fluorescent marker and *CHS3* deletion on growth phenotype

To control if the fluorescent mating-type marker as well as the deletion of the *CHS3* gene had no impact on the phenotype of interest, growth of wild-type and mutant parental strains were recorded in YPD medium alone and with a concentration range of NaCl (0.25 to 0.75 M) and KCl (1 to 2 M) using a microcultivation approach (Table 2).

Condition	Composition
YPD	2% bactopectone; 1% yeast extract; 2% glucose; 2% agar
YPD NaCl 0.25 M	YPD; NaCl 0.25 M
YPD NaCl 0.5 M	YPD; NaCl 0.5 M
YPD NaCl 0.75 M	YPD; NaCl 0.75 M
YPD KCl 1 M	YPD; KCl 1 M
YPD KCl 1.5 M	YPD; KCl 1.5 M
YPD KCl 2 M	YPD; KCl 2 M

Table 2. Detailed media composition used in this study

Wild-type and mutant parental strains had similar growth rates in YPD and in response to osmotic stress (Figure 4). The deletion of the *CHS3* gene as well as the integration of the fluorescent marker into parental genomes did not influence growth rate and consequently had no large phenotypic effect. The two conditions YPD NaCl 0.5 M and YPD KCl 1.5 M revealed

contrasted resistance phenotypes between parental strains (Figure 4). These stress conditions were consequently retained to determine the genetic origins of tolerance to high salt concentrations.

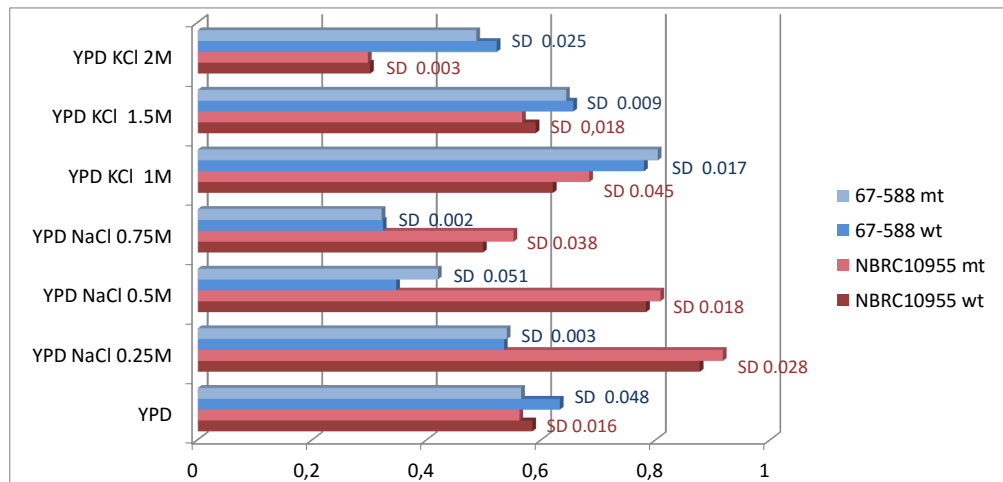


Figure 4. Measurement of growth phenotype. Specific growth rates (μ) were determined in each condition (YPD, YPD NaCl 0.25 to 0.75 M and YPD KCl 1 to 2 M) for wild-type (wt) and mutant (mt) strains NBRC10955 and 67-588. Specific growth rates in osmotic stress conditions were normalized to YPD. Mean values between biological replicates (N=2) were calculated and used to determine standard deviation between wild-type and mutant strains.

Fluorescence-based sorting of *MATa* progeny

To demonstrate the feasibility of separating *MATa* haploid segregants, random spore preparation from the sporulated diploid between mutants NBRC10955 and 67-588 was subjected to FACS (Figure 5).

A restrictive FACS gating was performed to select a population of single cells small in size (FSC-A), with low complexity (SSC-A) and expressing *yGFP*. After sorting, an aliquot of the selected population was tested for cell contamination and enumeration. Microscopic examination of a total of 80 isolated cells revealed exclusively green fluorescence, showing that all the sorted cells were *MATa* haploids. As a result, the approach enabled the isolation of a pure and large population of *MATa* haploids ($> 10^5$ cells) suitable for a subsequent BSA-seq mapping strategy.

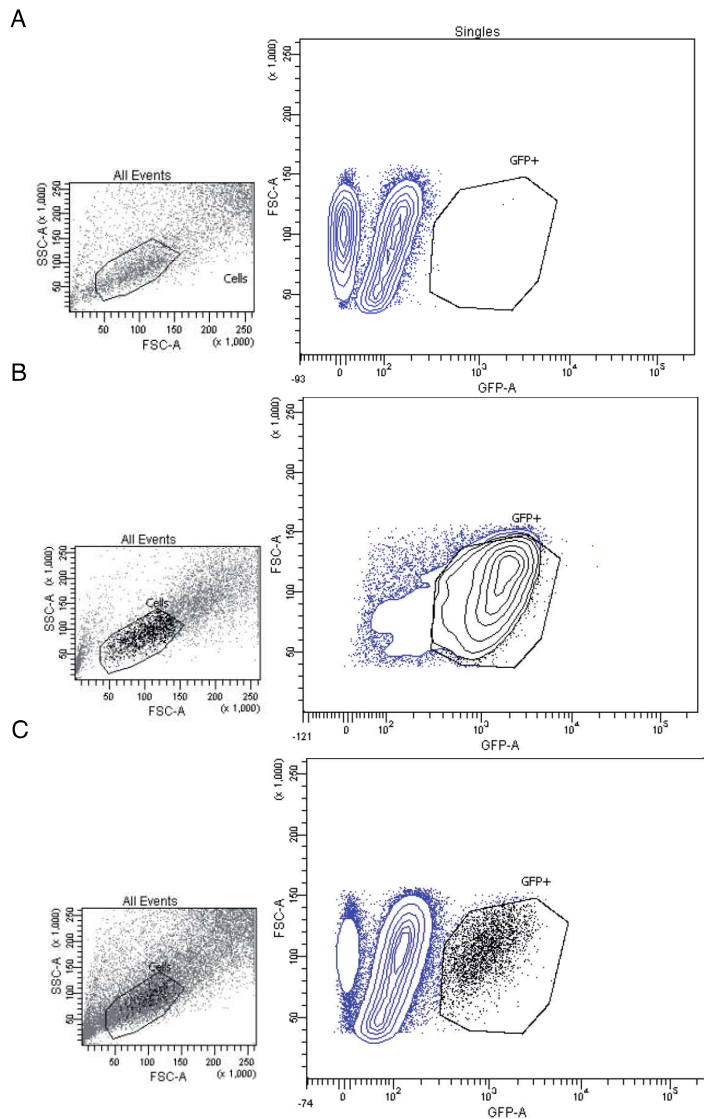


Figure 5. Fluorescence-based sorting of *MATa* progeny. A. Negative control: random spores preparation from the wild-type diploid NBRC10955 *MATa*/67-588 *MATα*. B. Positive control: NBRC10955 *chs3Δ::STE2p-sfGFP-KanMX*. C. Target cell population: random spores preparation from the mutant diploid *MATa chs3Δ::STE2p-sfGFP-KanMX* / *MATα chs3Δ::STE2p-sfGFP-KanMX*. Cells were first gated based on side (SSC-A) and forward scatter area (FSC-A) parameters to exclude both cellular debris and aggregates and then on GFP positive signal to sort single GFP⁺ cells. The target cell subpopulation was afterwards sorted.

Bulk segregant analysis under selective pressure

For the purpose of dissecting the genetic origin of halotolerance phenotypes, the resulting population of *MATa* haploid progeny was grown in YPD supplemented with NaCl 0.5 M or KCl 1.5 M to enrich for individuals with extreme phenotypes, and without selection as control, during

two days. Genomic DNA from the three pools was extracted and sequenced using Illumina HiSeq 2000 technology. The sequences of the pooled recombinant haploids were aligned to the genome of the NBRC10955 parental strain. To detect the QTLs, NBRC10955 allele frequency was scored across the 85,705 polymorphic positions. Following the selection, genomic regions involved in salt-tolerance were expected to have allele frequencies skewed from 0.5, whereas the rest of the genome should have similar proportions of alleles from each parent. In addition, because only *MATa* segregants have been selected for phenotypic analysis, the *MAT* locus region located on left arm of the chromosome C should come from the *MATa* strain NBRC10955 in the three pools generated. This is clearly illustrated by an allele frequency equal to 1 on this part of the genome (Figure 6), which also confirms that the fluorescent marker can be used to isolate a pure population of *MATa* haploids. Moreover, a very large part of the chromosome C corresponding to the entire left arm of the NBRC10955 strain is inherited by all the segregants, suggesting a very low recombination rate in this region.

Two genomic regions with skewed allele frequency were accurately identified for NaCl selective pool of segregants (Figure 6). The first putative causative region was mapped into a ~ 250 kb interval (1 000 000 – 1 250 000) on chromosome H spanning *ENA2* candidate gene encoding for a sodium efflux pump (Figure 6). In *S. cerevisiae*, copy number variation in *ENA* gene locus is well known to be associated to NaCl tolerance¹⁷. In *S. cerevisiae* strains with multicopies of *ENA* genes, NaCl tolerance would be linked to an introgression event, namely a transfer of an allele of *Saccharomyces paradoxus* *ENA* genes by hybridization^{18,19}. In general, the increase in copy number of *ENA* genes enhances the corresponding protein abundance and limits intracellular toxicity in high-salt concentrations. Based on average coverage analysis, no variation in terms of *ENA2* copy number was determined between *L. kluyveri* parental strains. Consequently, the resistance-sensitivity to high concentration of NaCl is allele specific in these genetic backgrounds. A total of 49 SNPs were detected in the *ENA2* coding region between NBRC10955 and 67-588 strains including 19 missense SNPs. The upstream region contains 12 SNPs differentiating the two strains. In addition, expression data in standard YPD medium revealed that *ENA2* was overexpressed constitutively (~ 1.6 x) in NBRC10955 compared to 67-588²⁰.

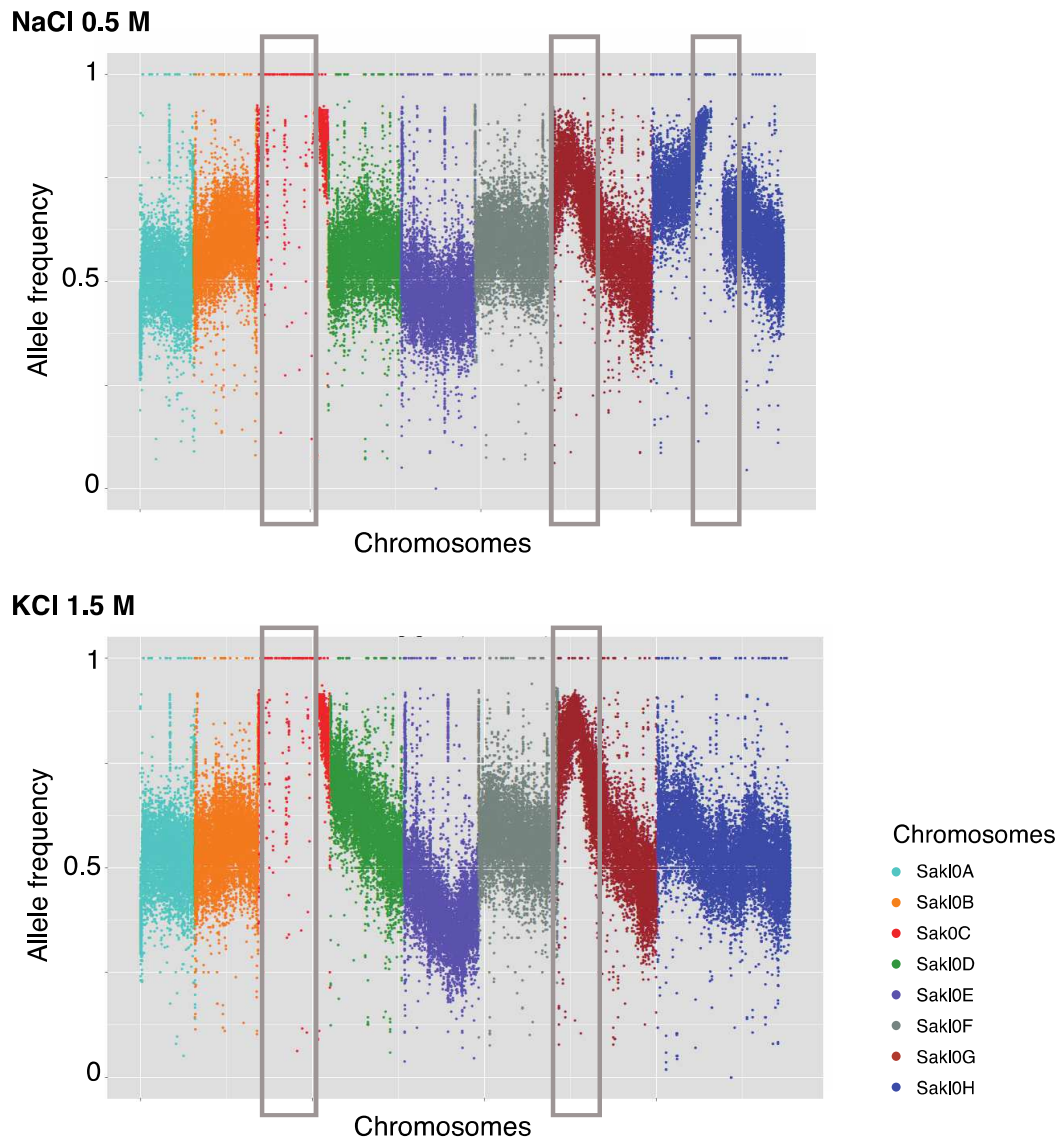


Figure 6. Identification of causal genomic regions determining halotolerance. Overview of the allele frequency of the parental strain NBRC10955 *MATa* (y-axis) on the eight chromosomes depicted by a color code (x-axis) for NaCl 0.5 M and KCl 1.5 M osmotic pools. Shaded boxes correspond to chromosomal regions with high skewed allele frequency.

A second putative causative region was mapped into a ~ 120 kb interval (250 000 – 370 000) on chromosome G (Figure 6). This latter region, shared with KCl selective pool of segregants, includes the *RVS167* candidate gene potentially related to NaCl and KCl tolerance. The *RVS167* gene encodes for an actin-associated protein that regulates notably viability following an osmotic stress. This gene is known to play a role in the hyperosmotic stress resistance including high concentrations of KCl and NaCl salts within *S. cerevisiae*²¹.

Functional validation of *ENA2* gene involved in resistance to osmotic stress

To definitively demonstrate that QTL mapping is applicable to this non-model organism, the involvement of *ENA2* in the tolerance to high concentration of NaCl was confirmed by a reciprocal hemizyosity analysis. The resistance-sensitivity to NaCl of hemizygote hybrids for *ENA2*^{NBRC10955} or *ENA2*⁶⁷⁻⁵⁸⁸ was tested. Hemizygous hybrids (NBRC10955 Δ *ena2*/67-588 and NBRC10955/67-588 Δ *ena2*), wild-type haploid strains (NBRC10955 and 67-588) as well as wild-type hybrid (NBRC10955/67-588) were tested on YPD NaCl 0.5 M (Figure 7).

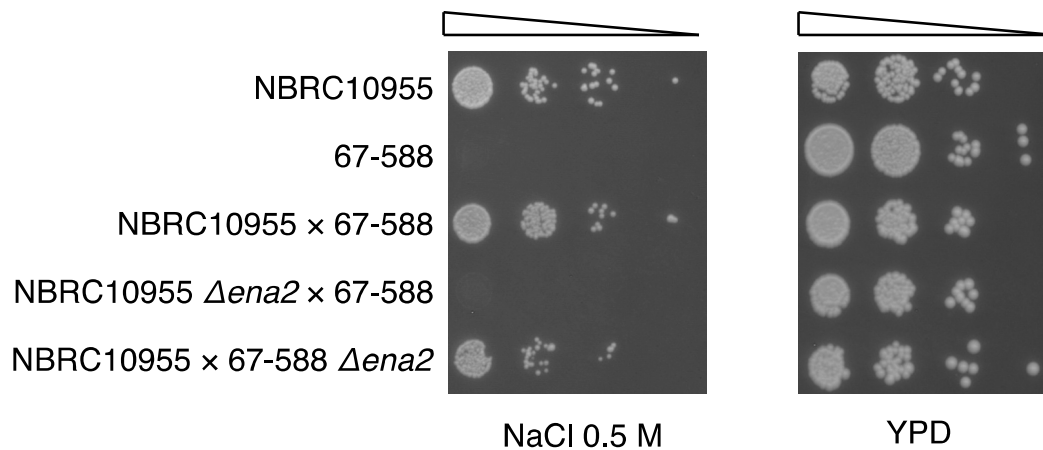


Figure 7. Reciprocal hemizyosity test on halotolerance. Sensitive (67-588) and resistant (NBRC10955) strains as well as wild-type and hemizygous hybrids for the *ENA2* gene were spotted in 5 dilutions onto YPD (right) and YPD NaCl 0.5 M (left) and observed after 48 hours at 30°C.

The results showed that the hemizygous diploid NBRC10955 Δ *ena2*/67-588 was sensitive in the presence of 0.5 M NaCl but not the hemizygous diploid NBRC10955/67-588 Δ *ena2*, indicating that the presence of *ENA2*^{NBRC10955} was sufficient for the resistance.

Conclusion

In conclusion, the method described in this study made possible the rapid isolation of a large mapping population and selection of segregants with extreme phenotypes in the non-model yeast species *L. kluyveri*. The use of the fluorescent marker expression as an indicator of a mating-type enabled facile creation of *MATa* segregant pools and consequently avoided eventual problem of cell mating during cultivation. The resulting mapping population under osmotic pressure allowed to dissect genetic origins of salt resistance and to identify genes directly associated with halotolerance. As a result, quantitative genetics is definitively affordable in *L. kluyveri* and shows promising results for further genetic dissection of phenotypes in more strain backgrounds. Larger-scale study will thereby provide a global view of the mechanisms leading to exceptional phenotypes within this non-model organism and lead to a better understanding of the evolution of the genotype-phenotype relationship across diverse yeast species.

Publication related to this chapter

Sigwalt, A., Caradec, C., Brion, C., Hou, J., de Montigny, J., Jung, PP., Fischer, G., Llorente, B., Friedrich, A. & Schacherer, J. Dissection of quantitative traits by bulk segregant mapping in a protoploid yeast species. *FEMS Yeast Research* (Under Review).

References

1. Birkeland, S. R. *et al.* Discovery of Mutations in *Saccharomyces cerevisiae* by Pooled Linkage Analysis and Whole-Genome Sequencing. *Genetics* **186**, 1127–1137 (2010).
2. Wenger, J. W., Schwartz, K. & Sherlock, G. Bulk Segregant Analysis by High-Throughput Sequencing Reveals a Novel Xylose Utilization Gene from *Saccharomyces cerevisiae*. *PLoS Genet* **6**, e1000942 (2010).
3. Parts, L. *et al.* Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res.* **21**, 1131–1138 (2011).
4. Magwene, P. M., Willis, J. H. & Kelly, J. K. The Statistics of Bulk Segregant Analysis Using Next Generation Sequencing. *PLoS Comput Biol* **7**, e1002255 (2011).
5. Hou, J., Friedrich, A., de Montigny, J. & Schacherer, J. Chromosomal Rearrangements as a Major Mechanism in the Onset of Reproductive Isolation in *Saccharomyces cerevisiae*. *Current Biology* **24**, 1153–1159 (2014).
6. 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).
7. Ehrenreich, I. M. *et al.* Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**, 1039–1042 (2010).
8. Ludlow, C. L. *et al.* High-throughput tetrad analysis. *Nat Meth* **10**, 671–675 (2013).
9. Sirr, A. *et al.* Allelic Variation, Aneuploidy, and Nongenetic Mechanisms Suppress a Monogenic Trait in Yeast. *Genetics* **199**, 247–262 (2015).
10. Scott, A. C., Ludlow, C. L., Cromie, G. A. & Dudley, A. M. BEST: Barcode Enabled Sequencing of Tetrads. *JoVE (Journal of Visualized Experiments)* e51401–e51401 (2014). doi:10.3791/51401
11. Chin, B. L., Frizzell, M. A., Timberlake, W. E. & Fink, G. R. FASTER MT: Isolation of Pure Populations of a and α Ascospores from *Saccharomyces cerevisiae*. *G3* **2**, 449–452 (2012).
12. 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).
13. 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**, 184–192 (2015).
14. Hartig, A., Holly, J., Saari, G. & MacKay, V. L. Multiple regulation of STE2, a mating-type-specific gene of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **6**, 2106–2114 (1986).

15. Marsh, L. & Herskowitz, I. STE2 protein of *Saccharomyces kluyveri* is a member of the rhodopsin/beta-adrenergic receptor family and is responsible for recognition of the peptide ligand alpha factor. *PNAS* **85**, 3855–3859 (1988).
16. Coluccio, A. & Neiman, A. M. Interspore bridges: a new feature of the *Saccharomyces cerevisiae* spore wall. *Microbiology (Reading, Engl.)* **150**, 3189–3196 (2004).
17. Ruiz, A. & Ariño, J. Function and Regulation of the *Saccharomyces cerevisiae* ENA Sodium ATPase System. *Eukaryotic Cell* **6**, 2175–2183 (2007).
18. Doniger, S. W. *et al.* A Catalog of Neutral and Deleterious Polymorphism in Yeast. *PLOS Genet* **4**, e1000183 (2008).
19. Warringer, J. *et al.* Trait Variation in Yeast Is Defined by Population History. *PLOS Genet* **7**, e1002111 (2011).
20. Brion, C., Pflieger, D., Friedrich, A. & Schacherer, J. Evolution of intraspecific transcriptomic landscapes in yeasts. *Nucl. Acids Res.* gkv363 (2015). doi:10.1093/nar/gkv363
21. Ren, G., Vajjhala, P., Lee, J. S., Winsor, B. & Munn, A. L. The BAR Domain Proteins: Molding Membranes in Fission, Fusion, and Phagy. *Microbiol. Mol. Biol. Rev.* **70**, 37–120 (2006).

MATERIEL ET METHODES

1. Les organismes de l'étude

1.1. La levure *Saccharomyces cerevisiae*

1.1.1. Généralités sur le cycle de vie de *Saccharomyces cerevisiae*

Saccharomyces cerevisiae suit un cycle de vie haplodiplobiontique. Selon les conditions de croissance, les cellules se développent sous forme diploïde ($2n$) ou haploïde (n) par multiplication végétative (bourgeonnement). Lorsque les conditions sont défavorables à la croissance, comme une carence en sources d'azote et de carbone, les cellules diploïdes se divisent par méiose en quatre spores haploïdes non ordonnées, dont deux de signe sexuel *MATa* et deux de signe sexuel *MAT α* , regroupées sous forme de tétrade dans un asque¹. Il est à noter que cette espèce est une levure homothallic, certaines cellules haploïdes sont de ce fait capables de changer de signe sexuel sous le contrôle du gène *HO* codant l'endonucléase (Homothallic switching endonuclease)². Lorsque les conditions redeviennent favorables, les spores germent et forment des cellules haploïdes capables de se croiser quand ces dernières sont de signe sexuel opposé et de former de nouvelles cellules diploïdes.

1.1.2. Souches de *Saccharomyces cerevisiae* étudiées

Une collection de 41 souches haploïdes stables (hétérothallics) de *S. cerevisiae* d'origines écologiques et géographiques variées a été utilisée. Le gène *HO* a été délété par insertion de la cassette de délétion *KanMX* conférant la résistance à la généticine G418^{3,4}. Les souches de laboratoire FY4 et FY5 isogéniques à S288c et Σ 1278b ont également été utilisées. La collection de mutants de délétion dans le fonds génétique Σ 1278b provient du laboratoire du Dr. Charles Boone⁵. La souche YJM326 Δ *pr1* a été générée par insertion de la cassette de résistance à l'hygromycine *HygMX* par recombinaison homologue.

1.2. La levure *Lachancea kluyveri*

1.2.1. Généralités sur le cycle de vie de *Lachancea kluyveri*

Lachancea kluyveri suit, tout comme *S. cerevisiae*, un cycle de vie haplodiplobiontique. Cependant elle a l'avantage d'être hétérothallic. Les cellules sont donc stables à la fois à l'état haploïde et diploïde en raison de la perte des cassettes silencieuses *HMR* et *HML* et de l'absence du gène *HO* dans le génome⁶. Toutefois certaines souches haploïdes de *L. kluyveri* possède en plus du cycle classique un pseudo-cycle atypique dans lequel elles sont capables de dupliquer leur génome afin de devenir des diploïdes, un phénomène appelé endoreduplication (Figure 1). Celles-ci sont par

conséquent homozygotes, en particulier pour le locus *MAT* c'est-à-dire *MAT_a/MAT_a* ou *MAT_α/MAT_α* et peuvent dans certain cas sporuler conduisant à la formation de 4 spores du même signe ou se croiser avec des cellules haploïdes de signe opposé, formant de ce fait des cellules triploïdes. Ces dernières peuvent quelquefois sporuler et former des tétrades composées de deux diploïdes et deux haploïdes. Les différentes étapes du pseudo-cycle apparaissent à des fréquences spécifiques à chaque souche.

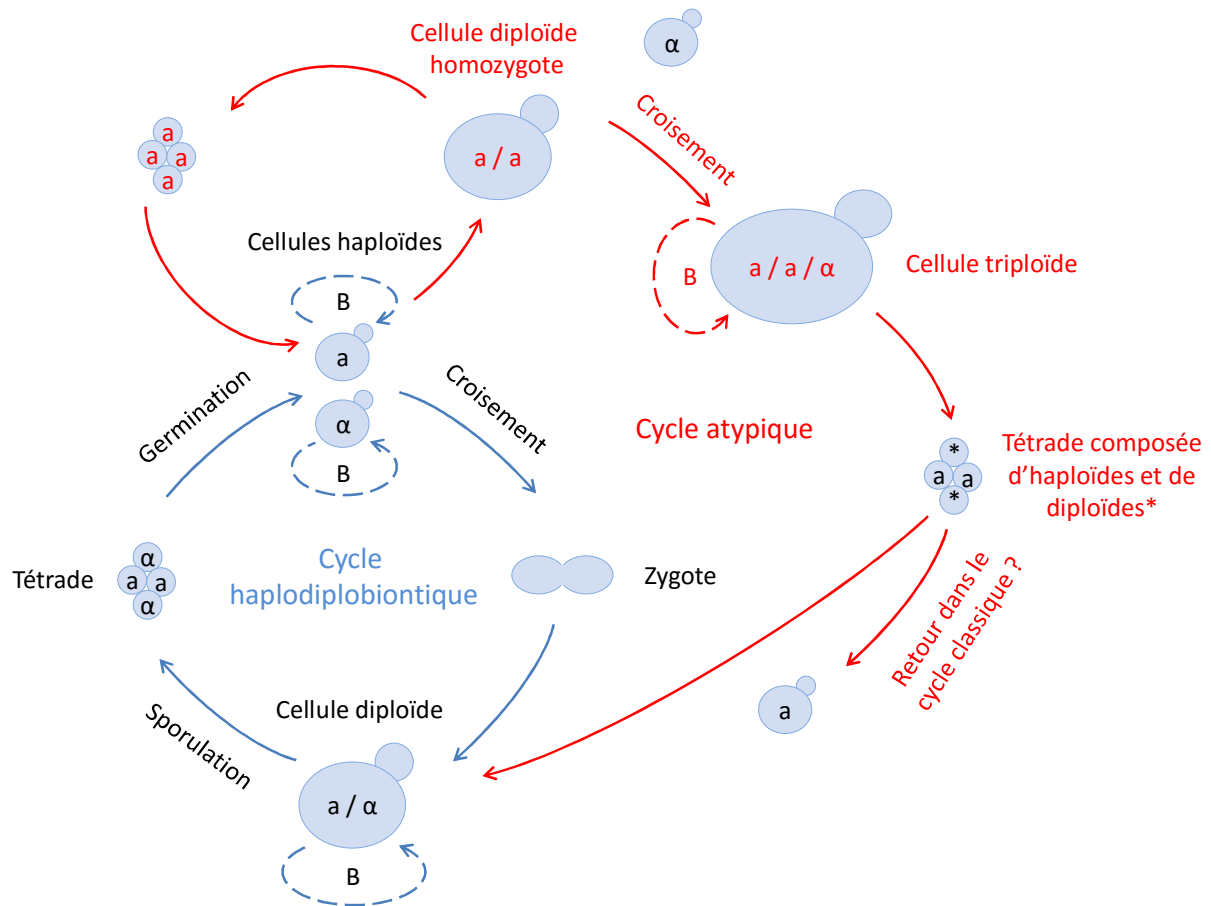


Figure 1. Cycle de vie de *L. kluyveri*. Comme la levure *S. cerevisiae*, *L. kluyveri* possède un cycle haplodiplontique (flèches bleues) où les cellules haploïdes et diploïdes sont capables de se multiplier par bourgeonnement (B). En plus de ce cycle classique, les cellules haploïdes de *L. kluyveri* possèdent un cycle atypique (flèches rouges).

1.2.2. Souches étudiées de *Lachancea kluyveri*

Une collection de 27 souches de *L. kluyveri* d'origines écologiques et géographiques diverses a été utilisée⁷. Les souches NBRC10955 $\Delta chs3$ et 67-588 $\Delta chs3$ ont été générées par insertion du marqueur fluorescent *MAT_a* spécifique fusionné à la cassette de résistance à la généticine *KanMX*

par recombinaison homologue. Les souches NBRC10955 Δ ena2 et 67-588 Δ ena2 ont été générées par insertion de la cassette de résistance à l'hygromycine *HphMX*.

2. Les méthodes de culture

2.1. Milieux de culture

Les milieux de culture sont stérilisés par autoclavage à 110°C pendant 20 minutes. La composition des milieux est la suivante :

- Milieu complet YPG : bactopectone 20 g/l, extrait de levure 10 g/l, glucose 20 g/l
- Milieu de sporulation AcK : acétate de potassium 10 g/l
- Milieu minimum YNB : yeast nitrogen base 6,7 g/l, glucose 20 g/l
- Milieu synthétique SD : yeast nitrogen base without amino acid 6,7 g/l, glucose 20 g/l, adénine 20 mg/l, histidine 20 mg/l, leucine 30 mg/l, méthionine 20 mg/l, uracile 20 mg/l

Ces milieux peuvent être solidifiés par addition de 2 % d'agar. Le milieu complet YPG peut être supplémentés par du chloramphénicol, de la généticine ou de l'hygromycine (200 µg/ml) afin d'obtenir des milieux sélectifs. Le milieu YPG peut être supplémenté par différents composés (antifongiques, métaux, sels, etc) afin d'induire différentes réponses physiologiques au niveau des levures.

2.2. Conditions de culture

Les levures sont cultivées en milieu liquide sous agitation ou en milieu solide à une température optimale de 30°C sauf indication contraire.

2.3. Evaluation de la concentration cellulaire

La concentration en cellules des cultures de *S. cerevisiae* et *L. kluyveri* en milieu liquide a été estimée par la mesure de la turbidité. L'absorbance à 600 nm a été mesurée par un spectrophotomètre Spectronic® 20 Genesys™ où une unité d'absorbance équivaut à $4,5 \cdot 10^7$ cellules haploïdes/ml et à $1 \cdot 10^7$ cellules diploïdes/ml.

2.4. Conservation des souches

Les souches de levure sont conservées à -80°C dans du milieu YPG supplémenté de glycérol à une concentration finale de 20%.

3. Les techniques de génétique de la levure

3.1. Obtention de cellules diploïdes

Deux souches de signes sexuels opposés (*MATa* et *MATα*), en phase exponentielle de croissance, sont croisées sur milieu YPG solide. Les zygotes de forme caractéristique sont observés au microscope 3 à 4 heures après le croisement et sont isolés à l'aide d'un micromanipulateur Singer de type MSM sur milieu YPG solide. Les cellules diploïdes sont incubées à 30°C pendant 2 à 3 jours et contrôlées par repiquage sur milieu de sporulation AcK. S'il y a présence de diploïdes, des tétrades composées de 4 spores sont observées au microscope après 2 à 3 jours d'incubation à 30°C.

3.2. Obtention de cellules haploïdes et analyse de la viabilité

Des cellules diploïdes en phase exponentielle de croissance sont repiquées sur milieu de sporulation AcK et incubées à 30°C pendant 2 à 3 jours. Les cellules diploïdes sporulées sont mises en suspension dans 200 µl d'eau stérile auxquels sont ajoutés 50 µl de zymolyase à 0,5 mg/ml (MP Biomedicals MT ImmunO™ 20 T) et incubées 15 min à 30°C. Après digestion de la paroi des asques sous l'action de la zymolyase, les quatre spores sont séparées à l'aide d'un micromanipulateur Singer MSM-400 et alignées sur milieu YPG solide. Les spores viables capables de germer formeront des colonies au bout de 48 heures à 30°C. La viabilité de la descendance correspond au rapport entre le nombre de spores viables et le nombre total de spores disséquées.

4. La détermination des phénotypes

4.1. Analyse de la croissance en microcultures

4.1.1. Stratégie de phénotypage

Des pré-cultures des souches de *L. kluyveri* étudiées sont réalisées en microplaque de 96 puits contenant 150 µl de milieu YPG liquide par puits. Les souches sont ensemencées en duplicat et certains puits ne sont pas inoculés constituant des témoins. Les pré-cultures sont incubées pendant une nuit à 30°C de manière à disposer de levures en phase exponentielle de croissance. Une nouvelle microplaque contenant 150 µl de milieu liquide d'une condition de croissance donnée est ensuite ensemencée à partir des pré-cultures. La croissance des souches est suivie par un lecteur de microplaques de type TECAN Infinite® F200. L'absorbance à 600 nm est mesurée

toutes les 10 minutes pendant 48 heures à une température de 30°C sauf indication contraire et sous agitation.

4.1.2. Analyse des courbes de croissance

La cinétique de croissance des levures en milieu liquide se compose de 4 phases caractéristiques : la phase de latence, la phase exponentielle, la phase stationnaire et la phase de décroissance (Figure 2).

La phase de latence est d'une durée variable en fonction de l'organisme et des conditions de culture. Elle correspond à la période pendant laquelle la population s'adapte à son environnement et ne se reproduit que faiblement. La phase exponentielle correspond à la période pendant laquelle la population se développe à une vitesse spécifique maximale de croissance selon un modèle exponentiel. La phase stationnaire correspond au maximum du rendement de croissance et se caractérise par une vitesse spécifique de croissance ralentie en raison de l'appauvrissement du milieu. Les nutriments ne sont plus en quantité suffisante dans le milieu pour couvrir les besoins de toute la population. Une partie de la population continue à se multiplier et compense celle qui meurt. La vitesse spécifique de croissance devient nulle. La phase de mortalité correspond à la mort des cellules. Elle se traduit par le déclin de la population.

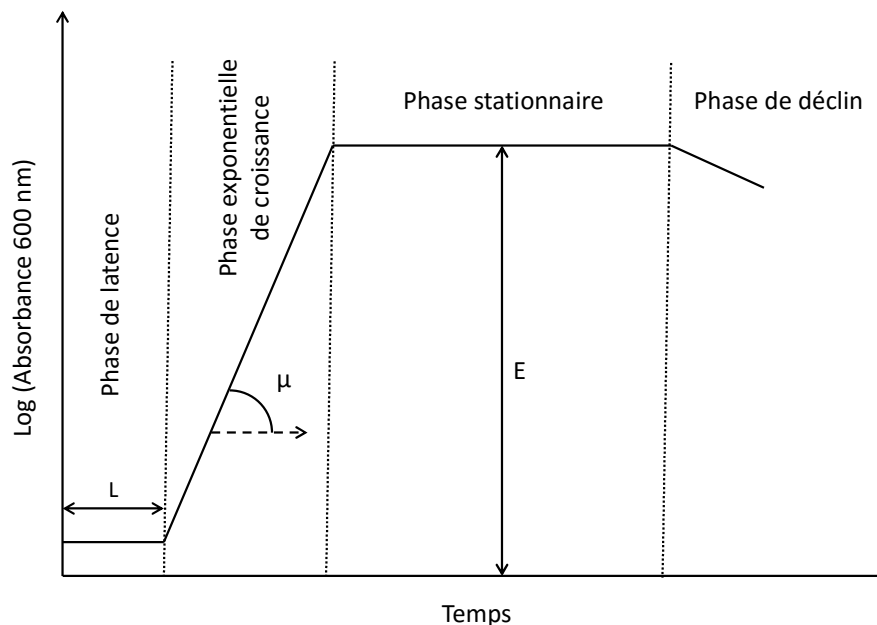


Figure 2 : Illustration d'une courbe de croissance. L : durée de la phase de latence ; μ : vitesse spécifique de croissance ; E : efficacité.

Des courbes de croissance sont établies à partir des mesures d'absorbance avec le logiciel *Courbes de croissance 1.4*, mis au point par Benoît Kammerer afin de déterminer des caractères quantitatifs pour chaque souche dans différentes conditions. La moyenne des absorbances des puits non inoculés est soustraite aux absorbances déterminées pour chacun des puits inoculés. Le logiciel détermine les variables caractéristiques propres à chaque souche dans une condition donnée. Il s'agit du temps de génération (g), de la vitesse spécifique de croissance (μ) et de la durée de la phase de latence (L) (Figure 2). La biomasse générée (E pour efficacité) peut être mesurée par la lecture de l'absorbance à 600 nm. La variable caractéristique choisie dans ce travail comme phénotype ou caractère quantitatif est la vitesse spécifique de croissance qui est directement liée au temps de génération ($\mu = \ln 2 / g$).

La moyenne des vitesses spécifiques de croissance des duplicats de chaque souche pour chaque condition testée est déterminée ainsi que l'écart-type. Ce dernier permet de représenter la variabilité autour de la moyenne entre réplicats. Le rapport de la vitesse spécifique de croissance pour chaque souche et chaque condition testées sur la vitesse spécifique de croissance en milieu complet YPG liquide est calculé. L'objectif est de s'affranchir des variations de croissance des différentes souches sur milieu complet et de comparer les souches entre elles. La variance des vitesses spécifiques de croissance est également déterminée pour chaque souche et chaque condition de croissance.

4.2. Analyse de la croissance en milieu solide

4.2.1. Stratégie de phénotypage

Une stratégie de phénotypage à haut-débit a été développée afin de mesurer la variation de croissance des souches de *S. cerevisiae* sur milieu solide face à différents facteurs de stress. Des pré-cultures des souches étudiées sont réalisées en microplaque de 96 puits contenant 150 μ l de milieu YPG liquide par puits. Les pré-cultures sont incubées pendant une nuit à 30°C de manière à disposer de levures en phase exponentielle de croissance. Les souches sont repiquées sur milieu YPG solide au format 384 à l'aide d'une plateforme robotisée de repiquage RoTor HDA (Singer Instruments). Les souches sont au moins présentes en duplicat à différentes positions sur les matrices 384 obtenues. Ces dernières sont incubées pendant 24 heures à 30°C afin d'avoir une croissance suffisante et repiquées sur différents milieux dont le milieu de référence YPG comme contrôle de repiquage. Les matrices sont incubées pendant 48 heures à 30°C sauf indication contraire et numérisées après 24, 40 et 48 heures d'incubation avec un scanner Perfection V700

Photo (Epson) à une résolution de 600 dpi et 16 bits en niveau de gris. Les images sont enregistrées au format TIFF.

4.2.2. Traitement et analyse des images

La taille des colonies est mesurée avec le plugin ColonyArea du logiciel ImageJ et le progiciel ScreenMill⁸. Ces logiciels de traitement et d'analyse d'images détectent l'intensité des pixels pour identifier et quantifier la taille des colonies. La croissance normalisée des souches est déterminée en calculant le rapport de la taille des colonies entre le milieu testé et le milieu de référence YPG pour chaque souche.

4.3. Test en gouttes

Des pré-cultures des souches étudiées sont réalisées en microplaque 96 puits contenant 150 µl de milieu YPG liquide par puits. Celles-ci sont incubées pendant une nuit à 30°C de manière à disposer de levures en phase exponentielle de croissance. Des dilutions décimales (10^0 à 10^5) sont réalisées à partir des précultures. Les dilutions 10^0 à 10^5 sont repiquées à l'aide d'un répliqueur sur milieu sélectif et sur milieu YPG solide comme contrôle. Les boîtes sont numérisées après 48 heures d'incubation à 30°C avec un scanner Perfection V700 Photo (Epson). Les différences de croissance des souches pour les conditions testées sont observées.

4.4. Analyse de la morphologie des cellules

4.4.1. Culture et fixation des cellules

Les souches de *L. kluyveri* sont cultivées à 30°C sous agitation dans 20 ml de milieu SD jusqu'en phase exponentielle de croissance puis fixées par ajout de 2,5 ml de formaldéhyde 37% et 2,5 ml de tampon phosphate de potassium 1 M. Les cellules sont incubées à 30°C pendant 30 minutes puis centrifugées à 3 000 rpm pendant 5 minutes. Les cellules sont resuspendues dans 1 ml de formaldéhyde 37%, 1 ml de tampon phosphate de potassium 1 M et 8 ml d'eau distillée puis incubées à 30°C pendant 45 minutes. Les cellules sont à nouveau centrifugées à 3000 rpm pendant 5 minutes et reprises dans 1 ml de tampon phosphate salin (PBS) 1x pH 7,4 (NaCl 137 mM, KCl 2,7 mM, Na₂HPO₄ 10 mM, KH₂PO₄ 1,76 mM).

4.4.2. Marquage des cellules

Afin d'acquérir des images de microscopie à fluorescence, les mannoprotéines constituant la paroi des cellules, le cytosquelette d'actine et l'ADN sont respectivement marqués à la

concanaline A couplée à l'isothiocyanate de fluorescéine (FITC-ConA), à la phalloïdine couplée à la rhodamine (rhodamine-phalloïdine) et au 4',6'-diamidino-2-phénylindole.

Les cellules sont centrifugées à 3000 rpm pendant 30 secondes puis lavées dans 600 µl de tampon PBS 1x. Après centrifugation à 3000 rpm pendant 30 secondes, 90 µl de tampon PBS 1x, 10 µl de rhodamine-phalloïdine (200 unités/ml) et 1 µl de Triton X-100 (10%) sont ajoutés aux cellules. Celles-ci sont incubées à l'obscurité à 4°C pendant 1 nuit. Les cellules sont lavées avec 600 µl de tampon PBS 1x, puis avec 600 µl de tampon P (phosphate de sodium 10 mM, NaCl 150 mM, pH 7,2). Après centrifugation à 3000 rpm pendant 30 secondes, les cellules sont resuspendues dans 488 µl de tampon P et 12 µl de FITC-ConA (1 mg/ml) puis incubées à température ambiante pendant 5 minutes. Les cellules sont lavées dans 600 µl de tampon P puis à nouveau resuspendues dans 600 µl de tampon P. Les cellules sont soniquées pendant 5 secondes au niveau 3 (TAITEC BR-5S) puis centrifugées à 3 000 rpm pendant 30 secondes. Les cellules sont resuspendues dans 7 µl de DAPI (1 µg/ml) et 100 µl de tampon de montage conservé à l'obscurité (p-phenylenediamine 20mg, PBS 1x 1 995 µl, NaOH 0,1 M 5 µl, glycérol 18 ml).

4.4.3. Observation et acquisition des images de microscopie à fluorescence

Les cellules marquées sont observées entre lame et lamelle sous un microscope à fluorescence Axioplan2 imaging (Zeiss) avec un objectif à immersion x100 équipé des filtres FITC/DAPI/Cyc3/Cyc5, du contraste interférentiel différentiel (DIC Nomarski) et d'une caméra CoolSnapHQ (Photometrics, Roper Scientific). Au moins 200 cellules sont capturées dans un ensemble d'images acquises à partir d'une culture cellulaire indépendante grâce au logiciel MetaMorph (version 5.0). Un total de 135 ensembles d'images est acquis à partir de 5 réplicats biologiques.

4.4.4. Traitement des images

Les images de microscopie à fluorescence sont traitées avec le logiciel CalMorph (version 1.3) qui mesure 501 paramètres quantitatifs sur des cellules individuelles reflétant divers aspects de la morphologie des cellules dont la forme de la cellule, la morphologie nucléaire et le cytosquelette d'actine⁹.

5. Préparation de l'ADN

5.1. Extraction d'ADN génomique par la méthode phénol chloroforme

Les ADN génomiques destinés à une réaction de PCR ont été extraits par la méthode phénol/chloroforme classique¹⁰. Les souches sont cultivées à 30°C sous agitation dans 15 ml de milieu YPG liquide jusqu'en phase stationnaire de croissance. Les cellules sont centrifugées pendant 5 minutes à 3000 rpm. Le culot est repris dans 1 ml d'eau et centrifugé 2 minutes à 12400 rpm. Les cellules sont remises en suspension dans 250 µl de tampon de broyage (Tris-HCl 10 mM pH 8 ; EDTA 1 mM ; NaCl 100 mM ; Triton X-100 2% ; SDS 1 %). Environ 100 µl de billes de verre d'un diamètre compris entre 0,2 et 0,5 mm et 250 µl de phénol-chloroforme sont ajoutés. Le mélange est vortexé vigoureusement pendant 3 minutes permettant la lyse des cellules puis centrifugé 5 minutes à 11000 rpm. Après centrifugation, un précipité est observé à l'interface qui correspond à des débris cellulaires et des protéines dénaturées. La phase aqueuse, dans laquelle restent en solution les acides nucléiques (ADN et ARN), est prélevée délicatement. Après ajout d'un volume de phénol chloroforme, le mélange est vortexé vigoureusement pendant 45 secondes puis centrifugé pendant 2 minutes à 11000 rpm afin d'éliminer le maximum de débris cellulaires. La phase aqueuse est à nouveau prélevée et 600 µl d'éthanol absolu ainsi que 20 µl d'acétate de sodium 4 mM sont ajoutés. Une précipitation d'une heure à -20°C suivie d'une centrifugation de 5 min à 12400 rpm sont réalisées. Le culot d'acides nucléiques (ADN et ARN) est séché puis repris dans 400 µl de tampon TE (Tris-HCl 10 mM pH 8 ; EDTA 1 mM). Les ARN sont ensuite dégradés par l'ajout de 5 µl de RNase A (10 mg/ml) pendant 10 minutes à 37°C. Les ADN sont à nouveau précipités par l'ajout de 10 µl d'acétate de sodium 4 M et 1 ml d'éthanol absolu pendant au minimum 1 heure à -20°C. Le mélange est centrifugé à 4°C pendant 15 minutes à 12400 rpm. Le culot est ensuite lavé avec 1 ml d'éthanol 70% et centrifugé 5 minutes à 12400 rpm. Le culot contenant l'ADN est séché puis repris dans 50 µl de tampon TE.

5.2. Extraction d'ADN génomique total par kit Qiagen

Les ADN génomiques des pools de descendants destinés au séquençage à haut débit de type Illumina Hi-seq 2000 et MiSeq sont extraits à l'aide du kit Qiagen Genomic-tip 100/G. Chez *S. cerevisiae*, 50 spores indépendantes sont séparément mises en culture dans 5 ml de milieu YPG liquide et incubées à 30°C sous agitation pendant une nuit. Les 50 cultures sont mélangées à des mesures d'absorbance à 600 nm égales. Les cellules sont centrifugées 10 minutes à 6000 rpm et reprises selon les instructions du fabricant données dans la section « Yeast » du protocole

Qiagen. Chez *L. kluyveri*, les cellules sont directement centrifugées à partir des enrichissements puis reprises selon les instructions du fabricant.

5.3. Purification d'ADN plasmidique

Les vecteurs de type plasmidiques ont été purifiés avec le kit Qiagen QIAprep Spin Miniprep en suivant les indications du fabricant.

5.4. Extraction des ARN totaux

Les ARN destinés à une réaction de RT-PCR ont été extraits dans un contexte sans nucléases. Les souches sont cultivées à 30°C sous agitation dans 7 ml de milieu YPG liquide jusqu'en milieu de phase exponentielle de croissance, à une absorbance comprise entre 0,35 et 0,45. Les cellules sont récupérées sur un filtre d'une porosité de 0,45 µm et instantanément congelées dans de l'azote liquide afin de conserver les ARN intacts et de ne pas induire de stress pouvant changer le contenu en ARN de la cellule. Les cellules sont conservées à -80°C jusqu'à l'extraction des ARN.

Les cellules sont traitées avec 750 µl de tampon de lyse (EDTA 10 mM, SDS 0,5 %, TrisHCl 10 mM pH 7,5) et vortexées pendant 3 minutes. Un volume de phénol acide (MP AQUAPHEO01) est ajouté. Le mélange est vortexé à nouveau pendant 3 minutes avant d'être incubé 1h à 65°C. Après un choc thermique de 10 minutes dans la glace, les acides nucléiques contenus dans la phase aqueuse sont transférés après centrifugation à 12400 rpm pendant 5 minutes dans un tube PLG™ (Phase Lock Gel Heavy 2ml, 5 Prime). Les traces de protéines et de débris cellulaires sont éliminées par une nouvelle centrifugation à 12400 rpm de 5 minutes après ajout d'un volume de chloroforme. Les acides nucléiques contenus dans le surnageant sont précipités par ajout de 1,5 ml d'éthanol 70% et 75 µl d'acétate de sodium 3 M puis incubés pendant une nuit à -20°C. Les ARN sont purifiés à l'aide du kit Qiagen RNeasy puis traités à la DNase afin d'éliminer l'ADN résiduel à l'aide du kit Invitrogen DNase I, Amplification Grade en suivant les instructions des fabricants.

5.5. Contrôle et dosage des acides nucléiques

5.5.1. Analyse des acides nucléiques par électrophorèse sur gel d'agarose

La présence d'acides nucléiques et leur non dégradation sont vérifiées par migration électrophorétique sur gel d'agarose 1 % dans du tampon TAE x 1 (acétate de sodium 8,3 mM ; acide acétique 0,135 % ; EDTA 1,1 mM ; Tris-base 40 mM) sous une tension de 90 V. L'ajout de

bromure d'éthidium (BET) dans le gel ou dans un bain après migration permet d'observer l'ADN sous rayonnement UV (UltraViolet).

5.5.2. Quantification des acides nucléiques

La concentration d'ADN, d'ADNc ou d'ARN est estimée par la mesure de l'absorbance à 260 nm avec un spectrophotomètre NanoDrop ND-1000 (Thermo Scientific). A cette longueur d'onde, une unité de densité optique d'ADN bicaténaire, d'ADN monocaténaire ou d'ARN correspond respectivement à une concentration de 50 µg/ml, 33 µg/ml et 40 µg/ml. Le rapport d'absorbance 260/280 nm permet d'évaluer la pureté de l'acide nucléique extrait. Un rapport 260/280 nm d'environ 1,8 indique que l'ADN est pur, et un rapport d'environ 2 indique que l'ARN est pur. Un rapport 260/280 nm nettement inférieur à ces valeurs indique la présence de protéines, de phénol ou d'autres contaminants qui absorbent à 280 nm. Le rapport d'absorbance 260/230 nm est une seconde mesure permettant d'évaluer la pureté de l'ADN et de l'ARN. Un rapport 260/230 nm compris entre 2,0 et 2,2 indique que l'ADN ou l'ARN est pur. Un rapport 260/230 nm nettement inférieur à 2,0-2,2 indique la présence de contaminants qui absorbent à 230 nm.

6. Amplification d'un fragment d'ADN par PCR

6.1. La réaction de PCR

6.1.1. Principe de la PCR

La PCR (Polymerase Chain Reaction) est une technique de biologie moléculaire permettant l'amplification exponentielle et spécifique d'un fragment d'ADN. La réaction de PCR nécessite une ADN polymérase thermostable (la Taq DNA polymérase ou la polymérase de haute fidélité iProof), une matrice d'ADN, des amorces oligonucléotidiques, un mélange des quatre désoxyribonucléotides (dNTP) et un thermocycleur de type Eppendorf Mastercycler EP. Elle comprend trois étapes qui sont répétées sur plusieurs cycles selon la quantité d'ADN matrice initiale : la dénaturation, l'hybridation et l'élongation. La dénaturation de l'ADN bicaténaire en ADN monocaténaire se fait à 94°C pour la Taq DNA polymérase ou à 98°C pour la polymérase de haute fidélité iProof pendant 30 secondes. La température d'hybridation (T_m) des amorces sur les sites spécifiques de l'ADN varie en fonction de la séquence et de la taille des amorces. Celle-ci est généralement comprise entre 50°C et 60°C. La durée de l'étape d'hybridation est de 30 secondes. L'élongation permettant la synthèse du brin complémentaire spécifique au brin matrice est réalisée à une température optimale de 72°C pour les deux ADN polymérases. La durée

d'élongation varie selon la taille du fragment à amplifier, généralement 1 min/kb pour la Taq DNA polymérase et 30 s/kb pour la polymérase de haute fidélité iProof. La réaction de PCR est généralement réalisée sur 30 à 35 cycles. Avant le début des cycles, une étape de dénaturation est réalisée pendant 2 minutes à 94°C pour la Taq DNA polymérase et à 98°C pour la polymérase de haute fidélité iProof. Après la fin des cycles, une étape supplémentaire d'élongation est réalisée pendant 10 minutes à 72°C afin de terminer la synthèse de potentiels brins inachevés.

6.1.2. Milieux de PCR

Deux ADN polymérases ont été utilisées lors de ces travaux : la Taq polymérase et la polymérase de haute fidélité iProof. Cette dernière a été employée dans le cadre de remplacement allélique pour l'amplification de fragments de plus ou moins grande taille nécessitant un faible taux d'erreurs.

Le milieu réactionnel de la PCR fourni avec la Taq polymérase se compose d'environ 100 ng d'ADN à amplifier, 0,5 µl d'une solution de dNTP à 10 mM (MP-QBIOgene), 1 µl de chaque oligonucléotide à 20 µM, 0,2 µl d'enzyme Taq polymérase (5 U/µl, MP-QBIOgene) et 5 µl de tampon concentré 10x (Tris-HCl pH 8,3 100 mM ; KCl 500 mM ; MgCl₂ 15 mM ; gélatine 0,01%). Le mélange réactionnel est complété avec de l'eau à un volume final de 50 µl.

Le milieu réactionnel de la PCR fourni avec la polymérase de haute fidélité iProof se compose d'environ 100 ng d'ADN à amplifier, 0,5 µl d'une solution de dNTP à 10 mM (MP-QBIOgene), 1 µl de chaque oligonucléotide à 20 µM, 0,5 µl d'enzyme polymérase de haute fidélité iProof (2 U/µl, Bio-rad) et 10 µl de tampon HF ou GC concentrés 5x contenant 7,5 mM de MgCl₂. Le tampon GC a été utilisé pour l'amplification de séquences riches en GC. Le milieu réactionnel est complété à un volume final de 50 µl.

6.1.3. Purification de fragments PCR

Les fragments de PCR destinés à du clonage sont purifiés avec le kit Qiagen QIAquick Gel Extraction en suivant les instructions du fabricant.

6.2. La réaction de RT-PCR

La RT-PCR est une technique qui fait intervenir une rétro-transcription (RT) de l'ARN en ADN complémentaire (ADNc) suivie d'une PCR. Ces deux étapes sont réalisées en une seule fois dans le cadre de l'analyse de la transcription du gène *STE2* décrite dans le chapitre 3 à l'aide du kit Invitrogen SuperScript® III One-Step RT-PCR System with Platinum® Taq DNA Polymerase et des amorces spécifiques au gène *STE2* en suivant les instructions du fabricant.

7. Système de clonage Gateway®

7.1. Principe de la technologie Gateway®

La technologie de clonage Gateway® développée par la société Invitrogen® est une méthode de clonage permettant de transférer rapidement et efficacement un à quatre fragments d'ADN (produits de PCR, ADNc, fragments de restriction) dans de multiples vecteurs. La méthode se base sur les propriétés d'intégration et d'excision du génome du bactériophage lambda dans le génome d'*Escherichia coli* (*E. coli*) par un système de recombinaison site-spécifique entre des sites d'attachement (attB, attL, attP et attR)¹¹. L'intégration est catalysée par l'intégrase (Int) produite par le phage et nécessite un facteur d'intégration de l'hôte (IHF). L'excision nécessite quant à elle, une excisionase (Xis) en plus de l'intégrase et du facteur d'intégration.

Dans la technologie Gateway®, les sites naturels d'attachement sont modifiés afin d'augmenter l'efficacité et la spécificité de recombinaison et de permettre un clonage directionnel¹². La technologie met en jeu deux réactions de recombinaison sites spécifiques successives, la réaction BP permettant la construction d'un vecteur d'entrée et la réaction LR permettant la construction d'un vecteur d'expression¹³.

La technologie de clonage Gateway® est utilisée pour le test de complémentation décrit dans le chapitre 1.

7.2. Réaction de recombinaison BP

La réaction de recombinaison BP permet la recombinaison entre une séquence d'intérêt à cloner portant à chacune de ses extrémités un site de recombinaison attB et un vecteur donneur portant une cassette Gateway® entourée de deux sites de recombinaison attP. Au cours de la recombinaison, catalysée par l'intégrase et le facteur d'intégration (BP Clonase™ enzyme mix), il y a échange des fragments d'ADN situés entre les sites de recombinaison. Le produit ainsi obtenu, appelé le vecteur d'entrée, porte la séquence d'intérêt entourée de sites de recombinaison attL. Le sous-produit porte quant à lui la cassette Gateway® entourée des sites attR. Les sites attL et attR sont des sites hybrides des sites de recombinaison attB et attP. Le vecteur d'entrée constitue le point de départ de la seconde réaction de recombinaison.

La technologie de clonage Gateway® permet le clonage d'un à 4 fragments d'ADN dans un même vecteur, il existe par conséquent de nombreuses combinaisons de sites de recombinaison attB selon le nombre de fragments à intégrer et de leur orientation.

La réaction de recombinaison BP a été réalisée à l'aide du kit Invitrogen™ Gateway® BP Clonase™ II Enzyme Mix avec le vecteur donneur pDONR™221 portant les sites de recombinaison attL1 et attL2 en suivant les instructions du fabricant. Les sites de recombinaison attB1 et attB2 ont été ajoutés aux extrémités des fragments d'intérêt par PCR. La correcte intégration des fragments d'intérêt dans les vecteurs d'entrée a été confirmée par analyse des profils de restriction enzymatique.

7.3. Réaction de recombinaison LR

La réaction de recombinaison LR permet la recombinaison entre la (les) séquence(s) d'intérêt à transférer portée(s) par le(s) vecteur(s) d'entrée et entourée(s) des sites attL et un vecteur de destination portant une cassette Gateway® entourée des sites de recombinaison attR. Au cours de la recombinaison, catalysée par les enzymes intégrase et excisionase et le facteur d'intégration (LR Clonase™ enzyme mix), il y a échange des fragments d'ADN situés entre les sites de recombinaison. Le produit ainsi obtenu, appelé le vecteur d'expression, porte la (les) séquence(s) d'intérêt entourée(s) de sites de recombinaison attB. Le sous-produit porte quant à lui la cassette Gateway® entourée des sites de recombinaison attP. Les sites attB et attP sont des sites hybrides des sites attL et attR.

La réaction de recombinaison LR a été réalisée à l'aide du kit Invitrogen™ Gateway® LR Clonase™ II Enzyme Mix avec le vecteur de destination p41 Hyg 1-F GW de type centromérique fourni par le laboratoire de L. Kruglyak¹⁴ en suivant les instructions du fabricant. Le vecteur de destination p41 Hyg 1-F GW porte le gène de résistance à l'hygromycine ainsi qu'une cassette Gateway® à un fragment. La correcte intégration des fragments d'intérêt dans les vecteurs d'expression a été confirmée par analyse des profils de restriction enzymatique.

7.4. Cassette Gateway®

La cassette Gateway® est une séquence d'ADN entourée des sites de recombinaison attP (sur les vecteurs d'entrée) ou attR (sur les vecteurs de destination). Celle-ci porte deux gènes de sélection des clones d'entrée et d'expression à savoir les gènes ccdB et CmR. Lors de la réaction de recombinaison BP et LR, la cassette Gateway® est remplacée par la (les) séquence(s) d'intérêt pour générer des vecteurs d'entrée et d'expression.

7.4.1. Sélection négative des clones d'entrée et d'expression

La présence du gène de résistance au chloramphénicol (CmR) permet une sélection négative des vecteurs d'entrée et d'expression chez la bactérie *E. coli* suite aux réactions de recombinaison et

de transformation. Les bactéries transformées avec un vecteur n'ayant pas réagi (vecteur donneur ou de destination) ou avec un sous-produit sont chloramphénicol résistantes en raison de la présence de la cassette Gateway® et par conséquent du gène cmR. Les bactéries transformées avec un vecteur d'entrée ou d'expression sont quant à elles chloramphénicol sensibles du fait du remplacement de la cassette Gateway® par la (les) séquence(s) d'intérêt.

7.4.2. Sélection positive des clones d'entrée et d'expression

La présence du gène ccdB permet une sélection positive des vecteurs d'entrée et d'expression chez la bactérie *E. coli* suite aux réactions de recombinaison et de transformation. Le gène ccdB code une protéine CcdB qui bloque l'action de l'ADN gyrase (protéine indispensable à la lecture et à la réplication de l'ADN)^{15,16}. En présence de la protéine CcdB, une souche bactérienne sauvage ne peut par conséquent pas répliquer son ADN, sa croissance est de ce fait inhibée. Les bactéries transformées avec un vecteur n'ayant pas réagi (vecteur donneur ou de destination) ou avec un sous-produit ne parviennent pas à croître en raison de la présence de la cassette Gateway® et donc du gène ccdB. Les bactéries transformées avec un vecteur d'entrée ou d'expression peuvent quant à elles répliquer leur ADN et sont donc capables de croître du fait du remplacement de la cassette Gateway® par la(les) séquence(s) d'intérêt. Une telle sélection permet de récupérer uniquement les clones d'entrée et d'expression avec une haute efficacité.

8. Techniques de transformation des levures

8.1. Transformation par électroporation

8.1.1. Protocole de transformation de *Saccharomyces cerevisiae*

Préparation de cellules électro-compétentes

Les cellules sont cultivées à 30°C sous agitation dans 50 ml de milieu YPG liquide jusqu'en fin de phase exponentielle de croissance à une absorbance à 600 nm de 0,6 à 0,8. Après une incubation de 10 minutes dans la glace, les cellules sont centrifugées pendant 5 minutes à 5 000 rpm à 4°C. Les cellules sont lavées deux fois avec 20 ml de sorbitol 1 M (froid) puis une fois avec 5 ml de sorbitol 1 M et centrifugées pendant 5 minutes à 5 000 rpm à 4°C. Les cellules sont reprises dans 200 µl de sorbitol 1 M (froid) puis conservées dans la glace.

Electroporation

A un volume de 50 µl de cellules compétentes en suspension est ajouté 1 µl d'ADN à 1 µg/µl ou de vecteur plasmidique à 10-100 ng/µl. Les cellules sont ensuite déposées dans une cuve d'électroporation de 2 mm et soumises à un choc électrique (1 500 V, 10 µF, 600 W) à l'aide d'un électroporateur Eppendorf 2510. Des pores transitoires sont créés dans la membrane des cellules permettant ainsi l'entrée de l'ADN. Immédiatement après l'électroporation, 1 ml de milieu YPG sorbitol 1 M liquide est ajouté à la cuve. Les cellules sont transférées dans un nouveau tube et régénérées environ 1 heure à 30°C. Les cellules sont brièvement centrifugées puis reprises dans 200 µl de milieu YPG sorbitol 1 M liquide. Un volume de 100 µl de cellules est étalé sur milieu sélectif. Les cellules sont incubées à 30°C et les colonies apparaissent après 2 à 3 jours d'incubation.

8.1.2. Protocole de transformation de *Lachancea kluyveri*⁴⁷

Préparation de cellules électro-compétentes

Les cellules sont cultivées à 30°C sous agitation dans 50 ml de milieu YPG liquide jusqu'en fin de phase exponentielle de croissance à une absorbance à 600 nm de 0,8 à 1,5 et centrifugées pendant 10 minutes à 5000 rpm. Les cellules sont resuspendues dans 25 ml de tampon phosphate de potassium 50 mM pH 7,5 contenant 25 ml de DTT (dithiothreitol) et incubées à 37°C pendant 30 minutes puis lavées deux fois avec 50 ml de tampon STM (froid) (sucrose 270 mM, Tris-HCl 10 mM pH 7,5, MgCl₂ 1 mM). Les cellules sont centrifugées à 5000 rpm pendant 10 minutes à 4°C. Les cellules sont reprises dans 100 µl de tampon STM (froid) puis conservées dans la glace.

Electroporation

A un volume de 50 µl de cellules compétentes en suspension est ajouté 1 µl d'ADN à 1 µg/µl ou de vecteur plasmidique à 10-100 ng/µl. Les cellules sont ensuite déposées dans une cuve d'électroporation de 2 mm et soumises à un choc électrique (1 500 V, 10 µF, 600 W) à l'aide d'un électroporateur Eppendorf 2510. Des pores transitoires sont créés dans la membrane des cellules permettant ainsi l'entrée de l'ADN. Immédiatement après l'électroporation, 1 ml de milieu YPG liquide est ajouté à la cuve. Les cellules sont transférées dans un nouveau tube et régénérées environ 1 heure à 30°C. Les cellules sont brièvement centrifugées puis reprises dans 500 µl de milieu YPG liquide. Un volume de 100 µl de cellules est étalé sur milieu sélectif. Les cellules sont incubées à 30°C et les colonies apparaissent après 2 à 3 jours d'incubation.

8.2. Transformation par choc thermique

La transformation des cellules par choc thermique est réalisée à l'aide du kit EZ-Yeast Transformation commercialisé par MP Biomedicals sur les souches ayant une faible efficacité de transformation. Les cellules sont cultivées à 30°C sous agitation dans 15 ml de milieu YPG liquide jusqu'en fin de phase exponentielle de croissance jusqu'à une absorbance de 0,6 à 0,8. Un volume de 1,5 ml de culture est centrifugé pendant 1 minute à 3 000 rpm. Les cellules sont lavées deux fois avec 1 ml d'eau stérile et centrifugées pendant 1 minute à 3 000 rpm. Les cellules sont reprises dans 125 µl de solution EZ-Transformation, 1 µl d'ADN à 1-3 µg/µl et 5 µl d'ADN porteur (DNA carrier) afin d'augmenter l'efficacité de transformation puis incubées à 42°C pendant 30 minutes. Les cellules sont centrifugées pendant 1 minute à 3 000 rpm puis reprises dans 1 ml de milieu YPG liquide. Les cellules sont régénérées environ 1 heure à 30°C puis brièvement centrifugées et reprises dans 200 µl de milieu YPG liquide. La totalité des cellules sont étalées sur milieu sélectif. Les cellules sont incubées à 30°C et les colonies apparaissent après 2 à 3 jours d'incubation.

9. Observation de cellules marquées par microscopie à fluorescence

Les cellules de *L. khyveri* marquées par une protéine fluorescente verte dérivée de la GFP (superfolder GFP, sfGFP) décrites dans le chapitre 3 sont observées entre lame et lamelle sous un microscope à fluorescence Axiovert200 (Zeiss) avec un objectif à immersion x100 équipé d'un filtre GFP, du contraste interférentiel différentiel (DIC Nomarski) et d'une caméra CoolSnapHQ2 (Photometrics, Roper Scientific).

10. Analyse des cellules par cytométrie en flux

10.1. Principe de la cytométrie en flux

La cytométrie en flux est une technique qui permet de mesurer les caractéristiques individuelles de cellules en suspension telles que la taille, la forme, la complexité et l'intensité de fluorescence. Les cellules entraînées par un flux liquide passent individuellement devant une source d'excitation lumineuse. Des détecteurs captent des signaux optiques émis par chaque cellule tels que la lumière diffusée dans l'axe du rayon incident (Forward Scatter, FSC) qui renseigne sur la taille des cellules, la lumière diffusée sous un angle de 90 degrés (Side Scatter, SSC) qui renseigne sur la forme, la structure interne et la granularité des cellules ou encore les signaux de fluorescence émis par la cellule elle-même (autofluorescence) ou par un marqueur fluorescent. La cytométrie en flux permet l'acquisition de données pour un très grand nombre de cellules sous forme de graphiques

ou d'histogrammes et de statistiques pour les populations cellulaires et les paramètres étudiés. Elle permet d'analyser des sous-populations cellulaires complexes et/ou rares à partir d'une population hétérogène voire de les trier à haute vitesse c'est-à-dire de séparer physiquement les cellules. La cytométrie en flux a l'avantage de pouvoir isoler des cellules avec un taux de pureté élevé et de préserver la viabilité des cellules. Ces dernières parfaitement viables peuvent par conséquent être remises en culture pour des analyses ultérieures. La cytométrie en flux a été appliquée dans ce travail pour l'analyse de la ploïdie de souches de levures et pour le tri d'une sous-population de cellules.

10.2. Contrôle de la ploïdie des souches

La détermination de la ploïdie est réalisée par une mesure cellule par cellule de l'intensité de fluorescence émise suite à un marquage de l'ADN à l'iodure de propidium, un agent intercalant de l'ADN fluorescent. L'intensité de fluorescence qui est équivalente à la quantité d'ADN présente dans chaque cellule est comparée à celle de souches de référence ayant une ploïdie connue (haploïde et diploïde). Le contrôle de la ploïdie est réalisé sur les souches haploïdes de *L. kluyveri* avant/après transformation et croisement afin de s'assurer de l'absence de phénomène d'endoréduplication.

10.2.1. Préparation des cellules

Les cellules sont cultivées à 30°C sous agitation dans 5 ml de milieu YPG liquide jusqu'en phase stationnaire de croissance. La préculture est diluée au dixième et incubée à 30°C pendant 2 heures sous agitation. Les cellules en phase exponentielle de croissance sont centrifugées pendant 5 minutes à 3 000 rpm puis lavées dans 1 ml d'eau et centrifugées pendant 5 minutes à 3 000 rpm. Les cellules sont fixées dans 1 ml d'éthanol à 70 % et centrifugées pendant 5 minutes à 3 000 rpm. Les cellules sont reprises dans 1 ml de tampon citrate de sodium 50 mM pH 7,5 contenant 10 µl de RNase A (100 mg/ml) et incubées pendant 2 heures à 37°C. Après incubation, les cellules sont soniquées pendant 20 secondes à une amplitude de 20 % pour éliminer les agrégats cellulaires. L'ADN des cellules est marqué à l'iodure de propidium (16 µg/ml) contenu dans 1 ml de tampon phosphate de sodium 50 mM pH 7,5. Les cellules sont incubées à 4°C à l'abri de la lumière pendant 12 à 48 heures afin de favoriser le marquage de l'ADN.

10.2.2. Mesure semi-quantitative de la fluorescence

Les cellules marquées à l'iodure de propidium sont passées dans un cytomètre en flux Cyflow® Space de Partec et excitées par un laser émettant à une longueur d'onde de 580 nm. La

fluorescence émise sur le canal FL2 est mesurée. En moyenne, 100 000 cellules sont analysées par échantillon. Les données sont stockées par le logiciel FloMax® de l'appareil et converties sous forme d'histogrammes de fluorescence dont les pics correspondent aux cellules se trouvant dans les différentes phases du cycle cellulaire. La sensibilité de l'appareil est configurée pour permettre une bonne différenciation des différents niveaux de ploïdie avec des gains de 200, 300 et 778 respectivement pour les paramètres FSC, SSC et FL2. Le profil de fluorescence obtenu et plus particulièrement l'intensité de fluorescence du premier pic (phase G₀/G₁) est comparée avec celui de deux souches contrôle l'une haploïde et l'autre diploïde afin de déterminer la ploïdie des souches testées.

10.3. Tri cellulaire

Le tri cellulaire est réalisé à partir d'une population hétérogène de cellules décrites dans le chapitre 3 exprimant et n'exprimant pas le marqueur fluorescent vert *sfGFP* couplé au promoteur *STE2* spécifique des cellules de signe *MATa* et directement intégré dans le génome de souches de *L. kluyveri*. Les cellules fluorescentes vertes de signe *MATa* sont isolées afin de générer une population à cartographier.

10.3.1. Préparation des cellules (adapté de Keller et Knop, 2009)¹⁸

Un hybride en phase exponentielle de croissance issu du croisement entre deux souches marquées est repiqué sur milieu de sporulation AcK et incubé à 30°C pendant 2 à 3 jours. L'hybride est mis en suspension dans 880 µl d'eau stérile additionnées de 20 µl de zymolyase à 10 mg/ml (MP Biomedicals MT ImmunO™ 100 T) et de 100 µl de sulfatase d'*Helix pomatia* (Sigma-Aldrich®) afin d'éliminer les cellules n'ayant pas sporulé puis incubé pendant une heure à 37°C. Environ 200 µl de billes de verre stériles d'un diamètre compris entre 0,2 et 0,5 mm sont ajoutés aux cellules. Le mélange est vortexé pendant 5 minutes afin de libérer les spores de l'asque. Les cellules sont brièvement centrifugées, lavées dans 1 ml de tampon phosphate salin (PBS) 1x pH 7,4 (NaCl 137 mM, KCl 2,7 mM, Na₂HPO₄ 10 mM, KH₂PO₄ 1,76 mM) stérile puis reprises dans 1 ml de tampon PBS 1x. Les cellules sont observées au microscope optique (objectif x40) entre lame et lamelle afin de vérifier la bonne libération des spores de l'asque.

10.3.2. Tri des cellules de signe *MATa*

Les cellules sont passées dans un cytomètre en flux BD FACSAria™ II de BD Biosciences et excitées par un laser émettant à une longueur d'onde de 488 nm. La fluorescence émise est détectée à l'aide d'un filtre passe-bande B525/50. Une sous-population de cellules de petite taille

(FSC), de faible complexité (SSC) et exprimant le gène *yGFP* est sélectionnée afin de s'affranchir de potentiels asques non cassés, de spores encore liées sous forme de triades ou de diades et de n'isoler que des cellules individuelles de signe *MATa*. Les cellules de l'échantillon sont injectées individuellement par l'intermédiaire d'une buse dans un courant continu de tampon PBS 1x stérile. Les cellules entraînées par le flux liquide passent séparément devant le faisceau laser. Le programme de tri traite le signal généré par la lumière déviée et la fluorescence émise afin de décider si la cellule doit être isolée ou non selon les trois critères imposés. Une fois le nombre requis de cellules obtenues ($> 10^5$) le tri est arrêté. Une analyse post-tri est réalisée afin de vérifier la pureté de la sous-population de cellules de signe *MATa* triée.

10.3.3. Enumération des cellules triées

L'énumération des cellules est réalisée par des dilutions au $10^{\text{ème}}$ en cascade et étalement sur milieu YPG solide. Les cellules sont incubées à 30°C pendant 2 jours jusqu'à apparition des colonies à dénombrer. Du chloramphénicol ($200 \mu\text{g/ml}$) est additionné au milieu afin de prévenir tout risque de contamination bactérienne suite au passage des cellules dans le cytomètre en flux.

10.3.4. Enrichissement des cellules triées

L'enrichissement des cellules de signe *MATa* triées est réalisé en milieu complet avec et sans pression de sélection. Un volume de $500 \mu\text{l}$ de cellules en suspension dans du tampon PBS 1x est mis en culture dans 50 ml de milieu YPG liquide supplémenté ou non de NaCl (0,5 M) ou de KCl (1,5 M). L'enrichissement est réalisé à 30°C sous agitation pendant 48 heures. Du chloramphénicol ($200 \mu\text{g/ml}$) est additionné aux milieux afin de prévenir tout risque de contamination bactérienne suite au passage des cellules dans le cytomètre en flux. L'ADN génomique est extrait à partir des cultures cellulaires et séquençé à haut-débit avec le système de séquençage Illumina HiSeq 2000.

11. Séquençage

La stratégie utilisée pour le séquençage des génomes décrit dans les chapitres 1 et 3 est la technique Illumina HiSeq 2000 et la technique Illumina MiSeq. L'avantage de ce type de stratégie est le taux de couverture élevé permettant de détecter la variation de la fréquence allélique à partir des données de pools de ségréants. De manière générale, la technologie Illumina repose sur le séquençage par synthèse¹⁹. La construction d'une librairie pour le séquençage Illumina consiste à fragmenter l'ADN génomique mécaniquement par sonication. Des adaptateurs sont ensuite ajoutés aux deux extrémités des fragments d'ADN. Les fragments de la taille désirée sont

sélectionnés afin de préparer une librairie avec une taille de fragment précise et amplifiés par PCR. Les fragments d'ADN double brin sont dénaturés et fixés sur un support avec des amorces complémentaires aux adaptateurs des fragments d'ADN à séquencer. La cellule est ensuite placée dans la cassette fluïdique du séquenceur. Les fragments sont amplifiés en phase solide par une étape de PCR par pontage. Le brin reverse est clivé et enlevé. Les quatre dNTP sont ajoutés au milieu réactionnel portant chacun un terminateur réversible et un fluorochrome différent (ddNTP). Le ddNTP complémentaire du brin matrice (forward) est incorporé afin de permettre l'élongation du brin complémentaire. Les ddNTP non incorporés sont éliminés du milieu réactionnel. La lecture de la fluorescence émise par le ddNTP incorporé par la polymérase est réalisée par une caméra aux longueurs d'ondes correspondant aux quatre fluorochromes des ddNTP. Le signal lumineux spécifique du ddNTP incorporé est détecté. Le fluorochrome et le terminateur réversible sont éliminés du milieu réactionnel avant de recommencer un nouveau cycle de séquençage. L'étape d'imagerie est répétée à chaque cycle ce qui permet de détecter dans l'ordre les nucléotides et de reconstituer la séquence des fragments. A la fin de l'étape de synthèse du brin reverse, celui-ci est utilisé comme matrice. Le séquençage réalisé aux deux extrémités des fragments permet de générer des lectures pairées ou associées (paired-end). Le multiplexage permet de séquencer plusieurs échantillons simultanément à moindre coût en utilisant des séquences synthétiques de quelques nucléotides appelées index pour identifier chaque échantillon. Le séquençage des échantillons a été réalisé avec la technique Illumina HiSeq 2000 au BGI (Beijing Genomics Institute) excepté pour l'ADN du pool de ségréants résistants au cycloheximide de *S. cerevisiae* décrit dans le chapitre 1 pour lequel la technique Illumina MiSeq a été utilisée.

12. Bioinformatique

12.1. Les banques de données biologiques

Les banques de données biologiques mises à la disposition de la collectivité scientifique permettent d'accéder aux séquences nucléiques et protéiques de gènes d'intérêt et de leurs produits et de réaliser des alignements de séquences (BLAST : Basic Local Alignment Search Tool) afin de mesurer le degré de similarité et de conservation de séquences d'intérêt. Les banques de données permettent également de collecter des informations sur la localisation et la fonction biologique des gènes, leurs interactions ou encore d'interroger une région chromosomique et d'avoir une représentation de la région considérée.

Les banques de données utilisées lors de ce travail sont les suivantes :

- Génolevures : <http://www.genolevures.org/>
- National Center for Biotechnology Information (NCBI) : <http://www.ncbi.nlm.nih.gov/>
- Saccharomyces Genome Database (SGD) : <http://www.yeastgenome.org/>

12.2. Analyse des données de séquençage de l'ADN génomique

12.2.1. Nettoyage des séquences

Les lectures brutes issues du séquençage des ADN génomiques sont nettoyées avec l'outil FASTQ Quality Trimmer inclus dans la suite logicielle FASTX-Toolkit qui permet d'éliminer les nucléotides de mauvaise qualité à la fin des séquences et les lectures trop courtes d'une longueur inférieure à 65 pb.

12.2.2. Alignement des séquences

Les lectures nettoyées sont alignées contre un génome de référence avec le logiciel BWA (Burrows-Wheeler Aligner, version 0.7.4) conçu pour traiter massivement les lectures générées par séquençage à haut-débit²⁰. Le logiciel BWA permet de gérer efficacement la variabilité nucléotidique entre les génomes comparés ainsi que la présence d'insertions et de délétions en autorisant un maximum de 5 mésappariements et 1 gap par lecture chez *S. cerevisiae* et un maximum de 8 mésappariements et 2 gaps chez *L. kluyveri*. L'indicateur '-I' est ajouté dans le cas de l'ADN du pool de ségréants résistants au cycloheximide de *S. cerevisiae* séquencé avec la technique Illumina MiSeq pour lequel les lectures générées sont encodées au format Illumina 1.9.

12.2.3. Détermination des SNP

Le fichier de sortie du logiciel d'alignement BWA constitue le fichier d'entrée du logiciel GATK (Genome Analysis Toolkit, version 3.3-0) utilisé pour identifier les positions polymorphiques par rapport au génome de référence avec les paramètres par défaut²¹. La fréquence allélique des parents est calculée pour chaque position polymorphique en ajoutant le ratio de la balance allélique avec la commande 'VariantAnnotator' de GATK.

12.2.4. Analyse des couvertures

La profondeur de séquençage, ou taux de couverture, est la longueur de l'ensemble des séquences lues mises bout à bout rapportée à la taille du génome séquencé. Plus la profondeur est élevée,

plus le nombre de lectures chevauchantes à assembler (contigs) est important et plus la fraction du génome couverte est grande. Par conséquent, une profondeur de séquençage importante permet d'obtenir une séquence consensus la plus fiable et complète possible, excepté dans le cas particulier d'erreurs d'alignement causées par des séquences répétées ou de faible complexité. La couverture moyenne est calculée sur des fenêtres de 2 kb ou simplement à partir du nombre de lectures alignées pour chaque position polymorphique afin de représenter la couverture le long des génomes séquencés, celle-ci varie de 50X à 100X.

12.2.5. Représentations graphiques des données de séquençage

Les représentations graphiques de l'analyse des couvertures et des fréquences alléliques sont réalisées à partir de l'extension « ggplot2 » du langage de programmation R (<http://www.r-project.org/>).

12.3. Méthode d'ajustement du modèle pour les caractères quantitatifs à distribution bimodale

Un modèle de mélange gaussien est appliqué aux 1 105 distributions de croissance décrites dans le chapitre 1. Les données de croissance sont séparées en deux composantes normales en utilisant la librairie R "mixtools" avec k (nombre de composantes) = 2 et $maxit$ (maximum nombre d'itération) = 500. Afin de déterminer à quelle composante appartient chaque individu, l'algorithme espérance-maximisation est utilisé (Figure 3).

Il s'agit d'une méthode itérative d'estimation du maximum de vraisemblance à partir d'un modèle statistique dépendant de variables latentes non observables. La moyenne (μ) et l'écart-type (σ) pour chaque composante normale, le rapport entre chaque composante (λ) et la probabilité a posteriori pour chaque individu d'appartenir à l'une des composantes sont extraits à partir du fichier de sortie. Pour déterminer les valeurs de "cut-off" de la probabilité a posteriori, un jeu de données simulé est généré en simulant deux distributions normales avec $n*\lambda$ and $n*(1-\lambda)$ individus pour chaque composante, respectivement, avec une moyenne et un écart-type sélectionnés aléatoirement à partir d'observations de données réelles. Pour chaque jeu de données simulé, les deux distributions normales générées sont combinées, et la procédure est répétée 1 000 fois pour générer un jeu de données dit d'apprentissage avec 1 000 distributions (Figure 3). Le jeu de données simulé est soumis au modèle d'ajustement avec les mêmes paramètres (Figure 3). La moyenne (μ), l'écart-type (σ), le rapport entre chaque composante (λ) et la probabilité a posteriori sont à nouveau extraits, et le jeu de données simulé est évalué contre les données réelles (Figure 4).

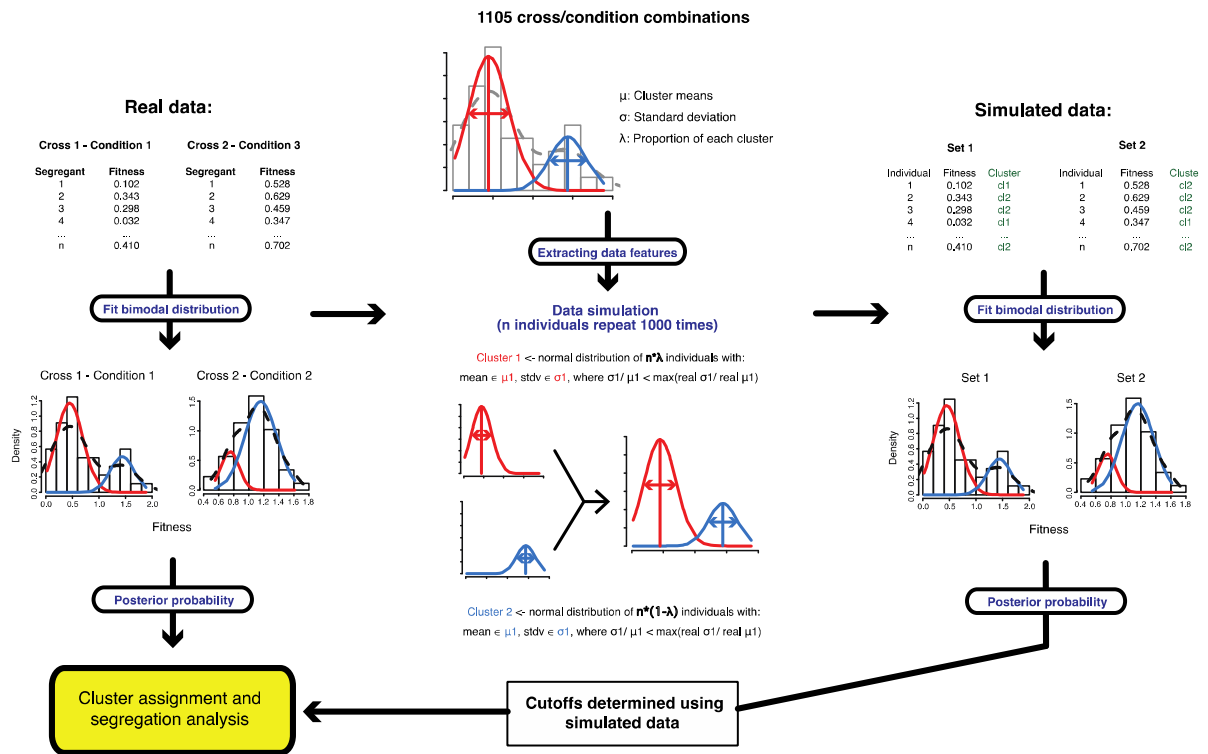


Figure 3: Algorithme espérance maximisation pour la détection de distribution bimodale en utilisant la probabilité a posteriori.

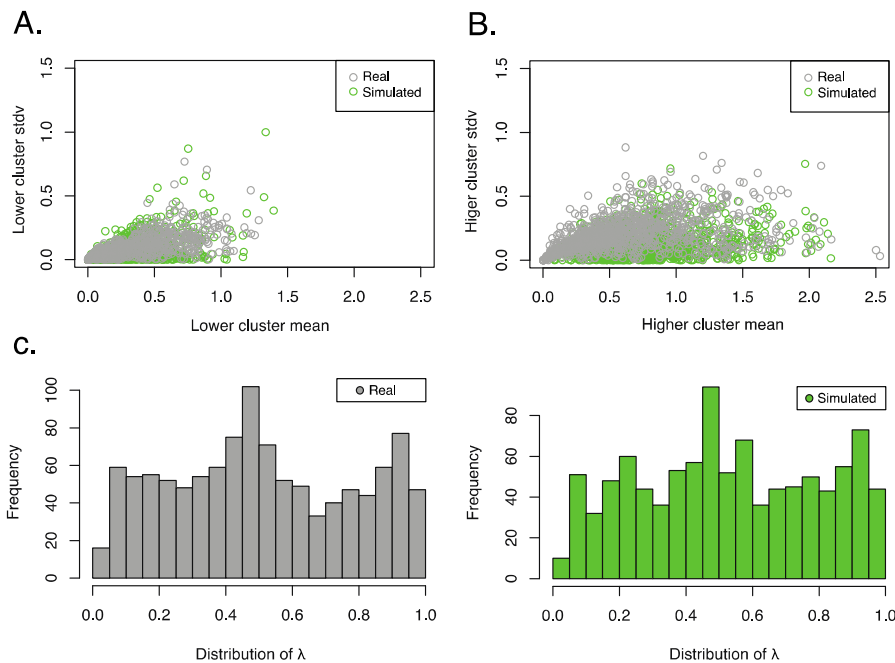


Figure 4. Comparaison des données réelles et des données d'apprentissage

La probabilité a priori étant connue dans ce cas pour chaque individu simulé, il est possible de tester la sensibilité et la spécificité de détection en fonction de la courbe d'efficacité du récepteur (ou courbe ROC, Receiver Operating Characteristic) en utilisant des paramètres de "cut-off" variés. Une séquence allant de 0,5 à 0,95 (incrément de 0,05) pour la probabilité a posteriori et de 0 à 0,9 (incrément de 0,1) pour le pourcentage d'individus ne se chevauchant pas sont testées. Les courbes ROC et l'aire sous la courbe sont calculées pour chaque combinaison de paramètres de "cut-off" en utilisant la librairie R "ROCR"²². Les "cut-off" de 0,8 pour la probabilité a posteriori et de 0,9 pour le pourcentage de non chevauchement sont retenus pour assurer des détections fiables (Figure 5).

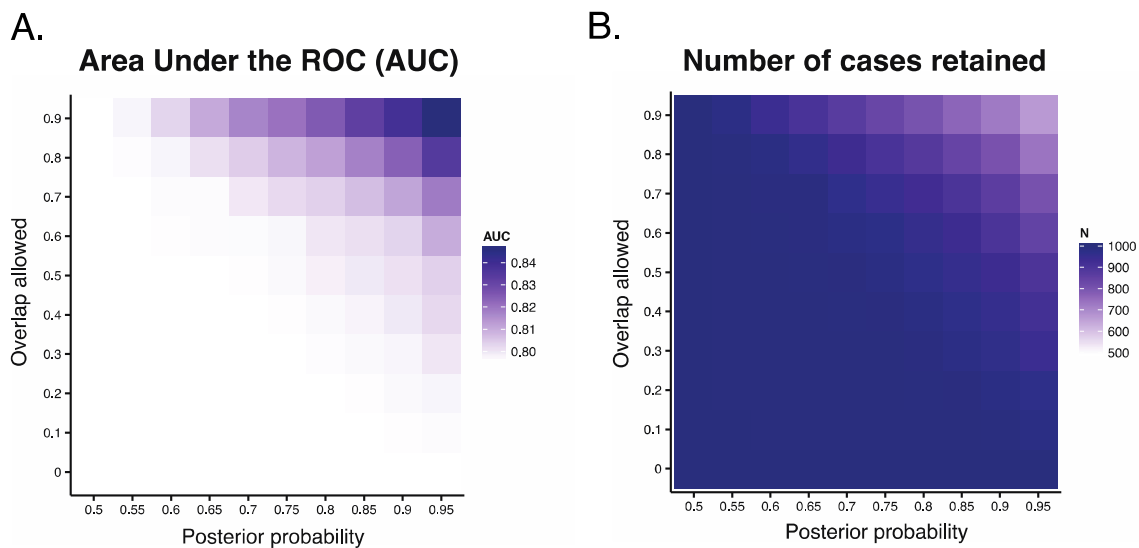


Figure 5. Evaluation de la puissance de détection par rapport aux cas de perte. (A) Evolution de l'aire sous la courbe (AUC) avec différentes combinaisons de paramètres. Axe des abscisses : "cut-offs" de la probabilité a posteriori. Axe des ordonnées : fraction de non chevauchement entre les composantes. (B) Nombre de cas passant la valeur seuil avec différentes combinaisons de paramètres. Axe des abscisses : "cut-offs" de la probabilité a posteriori. Axe des ordonnées : fraction de non chevauchement entre les composantes.

Les paramètres définis sont appliqués sur les données réelles (Figure 3). La ségrégation est déterminée pour les cas de bimodalité. Un caractère est considéré comme Mendélien quand plus de 70% des tétrades présentent une ségrégation 2:2 pour le phénotype léthal. Toutes les analyses sont réalisées dans R (<http://www.r-project.org/>).

12.4. Analyses statistiques des données de croissance et de morphologie cellulaire

Toutes les analyses statistiques décrites dans le chapitre 2 sont réalisées avec le langage de programmation R (<http://www.r-project.org/>). Le test de Kruskal-Wallis, l'analyse en composante principale (principal component analysis, PCA) et l'analyse de classification hiérarchique (hierarchical cluster analysis, HCA) sont réalisées selon la méthode décrite précédemment²³. Pour l'analyse de classification cumulative, 15 composantes principales comptant pour 90% du rapport cumulatif de contribution sont utilisées (Figure 6).

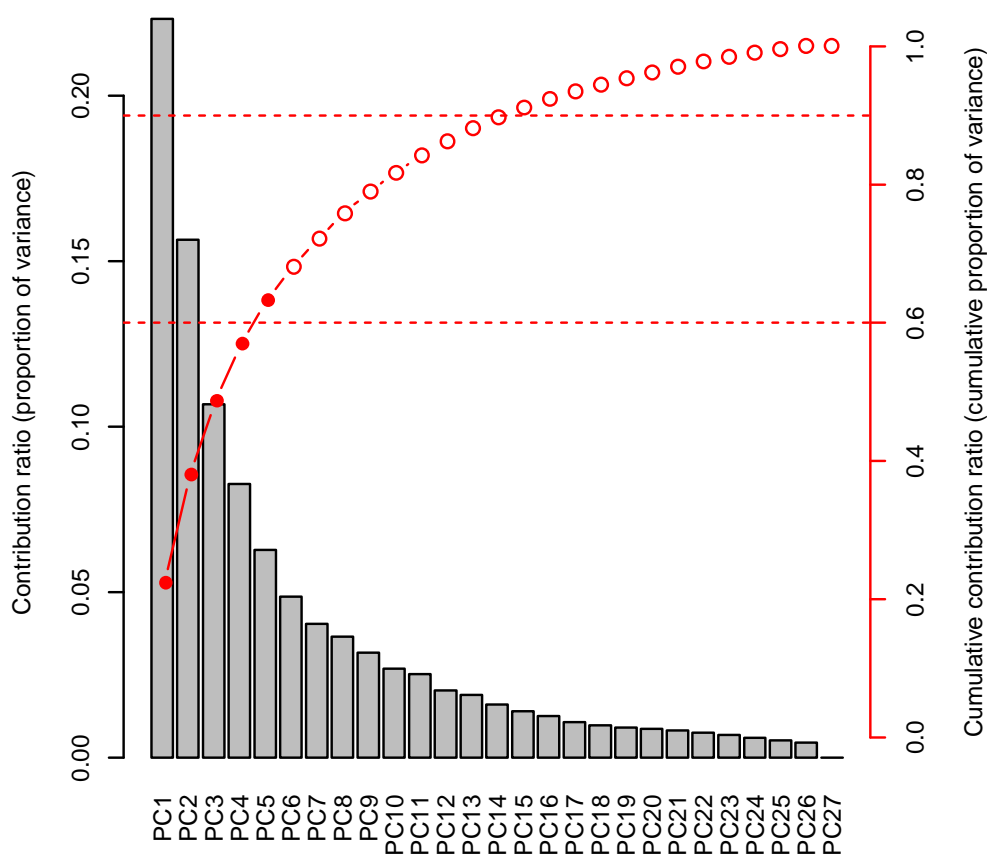


Figure 6. Proportion cumulative de la variance de l'analyse en composante principale pour les phénotypes. Les barres grises indiquent la proportion de la variance (axe de gauche) expliquée par les composantes principales. Les cercles rouges indiquent la proportion cumulative de la variance (axe de droite) expliquée par les composantes principales. La ligne rouge en pointillés indique 60% et 90% de la proportion cumulative de la variance.

Chaque composante principale est décrite dans le Tableau 1.

PC ^a	ID ^b	Loading ^c	P value ^d	Description ^e
PC1a	D127_A	0,958	4,42E-15	Distance between nuclear brightest point and cell tip
PC1a	C12-1 A	0,954	1,34E-14	Whole cell outline length
PC1a	C11-1 A	0,953	1,77E-14	Whole cell size
PC1a	C101 A1B	0,945	1,14E-13	Whole cell size
PC1a	C103 A	0,937	6,77E-13	Long axis length in whole cell
PC1a	C101 C	0,933	1,23E-12	Whole cell size
PC1a	C102 A1B	0,933	1,45E-12	Whole cell outline length
PC1a	C104 A1B	0,932	1,67E-12	Short axis length in mother
PC1a	C11-1 C	0,928	3,10E-12	Mother cell size
PC1a	C104 C	0,928	3,25E-12	Short axis length in mother
PC1a	D102 A	0,922	8,11E-12	Distance between nuclear gravity center and mother tip
PC1a	D135 A	0,922	8,24E-12	Distance between nuclear brightest point and cell center
PC1a	C11-1 A1B	0,918	1,51E-11	Mother cell size
PC1a	C102 C	0,912	3,60E-11	Whole cell outline length
PC1a	C12-1 C	0,908	6,43E-11	Mother cell outline length
PC1a	C104 A	0,905	9,07E-11	Short axis length in whole cell
PC1a	C103 C	0,899	1,84E-10	Long axis length in mother
PC1a	C128 C	0,896	2,83E-10	Distance between middle point of neck and mother hip
PC1a	C12-1 A1B	0,895	2,96E-10	Mother cell outline length
PC1a	C108 A1B	0,894	3,19E-10	Short axis length in bud
PC1a	C112 C	0,894	3,44E-10	Distance between middle point of neck and mother center
PC1a	D185 C	0,894	3,51E-10	Total length of two straight segments D11-1C4-1 and D11-2C4-1
PC1a	D129 A1B	0,893	3,89E-10	Distance between nuclear brightest point and mother tip
PC1a	D142 A1B	0,889	5,93E-10	Distance between nuclear brightest point and mother hip
PC1a	C103 A1B	0,888	6,18E-10	Long axis length in mother
PC1a	C112 A1B	0,888	6,72E-10	Distance between middle point of neck and mother center
PC1a	C11-2 C	0,887	7,32E-10	Bud cell size
PC1a	D104 A1B	0,885	8,90E-10	Distance between nuclear gravity center and mother tip
PC1a	C128 A1B	0,883	1,07E-09	Distance between middle point of neck and mother hip
PC1a	D126 A1B	0,878	1,86E-09	Distance between nuclear gravity center and mother hip

PC1a	D135 C	0,871	3,37E-09	Distance between nuclear brightest point in mother and mother center
PC1a	D186 C	0,859	9,60E-09	Total length of two straight segments D12-1C4-1 and D12-2C4-1
PC1a	C12-2 C	0,857	1,11E-08	Bud cell outline length
PC1a	C108 C	0,857	1,16E-08	Short axis length in bud
PC1a	C107 C	0,848	2,42E-08	Long axis length in bud
PC1a	D136 A1B	0,841	3,88E-08	Distance between nuclear brightest point and mother center
PC1a	D145 A1B	0,813	2,61E-07	Distance between nuclear outline point D7 and mother hip
PC1a	C11-2 A1B	0,811	2,84E-07	Bud cell size
PC1a	C109 C	0,776	2,00E-06	Neck width
PC1a	D118 A1B	0,773	2,29E-06	Distance between nuclear gravity center and mother center
PC1a	D117 A	0,771	2,58E-06	Distance between nuclear gravity center and cell center
PC1a	C109 A1B	0,741	9,72E-06	Neck width
PC1b	D103 C	0,862	7,92E-09	Distance between nuclear gravity center in mother and mother tip
PC1b	D125 C	0,843	3,32E-08	Distance between nuclear gravity center in mother and mother hip
PC1b	D128 C	0,843	3,54E-08	Distance between nuclear brightest point in mother and mother tip
PC1b	D141 C	0,828	9,52E-08	Distance between nuclear brightest point in mother and mother hip
PC1b	D145 C	0,810	3,08E-07	Distance between nuclear outline point D7 in mother and mother hip
PC1c	D121 C	0,955	1,12E-14	Distance between nuclear gravity center in bud and bud tip
PC1c	D139 C	0,922	8,65E-12	Distance between nuclear brightest point in bud and bud tip
PC1c	D146 C	0,920	1,17E-11	Distance between nuclear outline point D8 in bud and bud tip
PC1c	DCV109 C	0,758	4,60E-06	Coefficient of variation of distance between nuclear gravity center in bud and middle point of neck
PC2a	D17-3 A1B	0,864	6,60E-09	Nuclear fitness for ellipse
PC2a	DCV147 C	0,852	1,68E-08	Coefficient of variation of relative distance of nuclear gravity center in mother to mother center
PC2a	DCV181 A1B	0,850	2,03E-08	Coefficient of variation of nuclear minimum radius
PC2a	D123 C	0,813	2,56E-07	Ratio of D121 to C107
PC2a	DCV117 C	0,806	3,83E-07	Coefficient of variation of distance between nuclear gravity center in mother and mother center
PC2a	D175 A1B	0,782	1,45E-06	Maximal distance between nuclear gravity center and nuclear outline
PC2a	D178 A1B	0,747	7,70E-06	Nuclear long axis length
PC2a	D147 C	-0,754	5,66E-06	Relative distance of nuclear gravity center in mother to mother center

PC2a	D113 C	-0,859	1,00E-08	Ratio of D109 to C107
PC2a	D153 C	-0,886	8,13E-10	Mobility of nucleus in bud
PC2b	DCV165 A1B	0,758	4,56E-06	Coefficient of variation of angle between D3-1D4-1 and C1C4-1 or between D3-3D4-3 and C1C4-1
PC2b	D161 A1B	-0,729	1,64E-05	Angle between D3-1D4-1 and C1-1C1-2 or between D3-3D4-3 and C1-1C1-2
PC2b	D165 A1B	-0,736	1,22E-05	Angle between D3-1D4-1 and C1C4-1 or between D3-3D4-3 and C1C4-1
PC2c	DCV147 A	0,817	2,00E-07	Coefficient of variation of relative distance of nuclear gravity center to cell center
PC2c	DCV117 A	0,789	1,03E-06	Coefficient of variation of distance between nuclear gravity center and cell center
PC2d	DCV144 C	0,740	1,04E-05	Coefficient of variation of distance between nuclear outline point D6-2 in bud and middle point of neck
PC3a	CCV12-1 A	-0,793	7,99E-07	Coefficient of variation of whole cell outline length
PC3a	CCV11-1 A	-0,811	2,93E-07	Coefficient of variation of whole cell size
PC3a	CCV104 A	-0,816	2,09E-07	Coefficient of variation of short axis length in whole cell
PC3b	C115 A1B	-0,726	1,81E-05	Mother axis ratio
PC3b	C115 A	-0,737	1,18E-05	Whole cell axis ratio
PC3b	C115 C	-0,749	6,95E-06	Mother axis ratio

^aPC groups for morphological traits

^bIdentifier of morphological traits

^cGreen and red colors indicate negative and positive PC loading values.

^dSignificance of PC loadings

^eFull descriptions and definitions of the morphological traits are provided previously (Ohya et al., 2005)⁹.

Tableau 1. Valeurs de saturation des composantes principales (PC) utilisées pour la caractérisation de chaque composante principale.

La variance phénotypique de la morphologie des cellules est calculée à partir des valeurs de Z des 501 caractères qui sont estimées par le modèle linéaire généralisé (generalized linear model, GLM) comme décrit précédemment²⁴. La variance phénotypique pour les rapports de croissance est estimée comme la moyenne de la variance estimée dans chaque condition environnementale.

Références

1. Freese, E. B., Chu, M. I. & Freese, E. Initiation of yeast sporulation of partial carbon, nitrogen, or phosphate deprivation. *J. Bacteriol.* **149**, 840–851 (1982).
2. Herskowitz, I. Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiol. Rev.* **52**, 536–553 (1988).
3. Liti, G. *et al.* Population genomics of domestic and wild yeasts. *Nature* **458**, 337–341 (2009).
4. Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–345 (2009).
5. 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**, 184–192 (2015).
6. Dowell, R. D. *et al.* Genotype to phenotype: a complex problem. *Science* **328**, 469 (2010).
7. Payen, C. *et al.* Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Res.* **19**, 1710–1721 (2009).
8. Dittmar, J. C., Reid, R. J. & Rothstein, R. ScreenMill: a freely available software suite for growth measurement, analysis and visualization of high-throughput screen data. *BMC Bioinformatics* **11**, 353 (2010).
9. Ohya, Y. *et al.* High-dimensional and large-scale phenotyping of yeast mutants. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 19015–19020 (2005).
10. Hoffman, C. S. & Winston, F. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**, 267–272 (1987).
11. Landy, A. Dynamic, structural, and regulatory aspects of lambda site-specific recombination. *Annu. Rev. Biochem.* **58**, 913–949 (1989).
12. Bushman, W., Thompson, J. F., Vargas, L. & Landy, A. Control of directionality in lambda site specific recombination. *Science* **230**, 906–911 (1985).
13. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Res.* **10**, 1788–1795 (2000).
14. 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).
15. Bernard, P. & Couturier, M. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J. Mol. Biol.* **226**, 735–745 (1992).

16. Miki, T., Park, J. A., Nagao, K., Murayama, N. & Horiuchi, T. Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. Mutants of DNA gyrase subunit A suppress letD (ccdB) product growth inhibition. *J. Mol. Biol.* **225**, 39–52 (1992).
17. Gojkovic, Z., Jahnke, K., Schnackerz, K. D. & Piskur, J. PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *J. Mol. Biol.* **295**, 1073–1087 (2000).
18. Keller, P. J. & Knop, M. Evolution of Mutational Robustness in the Yeast Genome: A Link to Essential Genes and Meiotic Recombination Hotspots. *PLoS Genet* **5**, e1000533 (2009).
19. Mardis, E. R. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* **9**, 387–402 (2008).
20. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
21. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
22. Sing, T., Sander, O., Beerenwinkel, N. & Lengauer, T. ROCR: visualizing classifier performance in R. *Bioinformatics* **21**, 3940–3941 (2005).
23. Yvert, G. *et al.* Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst Biol* **7**, 54 (2013).
24. Yang, M., Ohnuki, S. & Ohya, Y. Unveiling nonessential gene deletions that confer significant morphological phenotypes beyond natural yeast strains. *BMC Genomics* **15**, (2014).

CONCLUSION ET PERSPECTIVES

Conclusion et perspectives

En génétique humaine, l'espoir de trouver l'ensemble des gènes et mutations responsables des maladies communes constituant des traits complexes reste hanté par le mystère de l'héritabilité manquante¹⁻³. Les variants génétiques fréquents identifiés depuis le projet génome humain⁴ à nos jours par des études d'association pangénomiques n'expliquent en effet qu'une faible partie de la variabilité observée chez les patients atteints de maladies complexes¹⁻³. Plusieurs hypothèses sont actuellement à l'étude pour tenter de résoudre l'énigme de l'héritabilité manquante dont l'importance des variants génétiques rares à forts effets sur la prédisposition à de telles maladies⁵⁻⁸ ou encore des interactions gène-gène (épistatiques) et gène-environnement³. De manière générale, la clé de l'énigme passe par une meilleure compréhension de l'architecture génétique des caractères, en déterminant le nombre de variants impliqués, leurs effets à travers diverses combinaisons de variants et leur fréquence au sein d'une population. Les levures comme modèles semblent être les plus à même pour atteindre cet objectif. En comparaison à d'autres organismes eucaryotes modèles, les espèces de levures, notamment *S. cerevisiae*, présentent une vaste diversité génétique, un génome petit et compact, un temps de génération court, des produits de la méiose directement analysables et une facilité de manipulation pratique en laboratoire.

Origines et complexité génétique de la diversité phénotypique intra-spécifique

Dans ce contexte, la première étude s'est attachée à estimer pour la première fois la complexité génétique de la variation phénotypique observée à travers une large population d'isolats de *S. cerevisiae* et a consisté à tout ramener vers le niveau d'organisation le plus simple afin de tracer l'effet d'un variant génétique rare dans différents fonds génétiques. L'expression de l'allèle rare *PDR1*^{YJM326} pourtant à fort effet dans la souche résistante au cycloheximide s'est avéré être variable d'une souche à une autre en raison de la variation de la constitution allélique sur le reste des génomes. Les différents degrés d'expression du phénotype, appelé expressivité, sembleraient être dus à la présence de gènes modificateurs, d'interactions épistatiques et/ou de suppresseurs selon les trois différents degrés de complexité décrits (30% des cas). L'identification et la caractérisation des gènes modificateurs modulant le phénotype sont actuellement en cours suite au séquençage de pools de ségréants. Ceci devrait permettre de mieux comprendre pour ce trait la relation génotype-phénotype et de visualiser l'héritabilité manquante. De manière globale, cette étude a permis de révéler la complexité cachée d'un caractère Mendélien à travers une population où finalement la division établie entre les caractères monogéniques et les caractères complexes n'est pas si absolue.

Cependant, bien que des souches aux fonds génétiques très divers ont été étudiées (autres que S288c et BY), l'estimation de la complexité génétique a été quelque peu biaisée dans cette première étude pour plusieurs raisons. Premièrement, la souche de laboratoire Σ 1278b utilisée comme référence a été impliquée dans tous les croisements étudiés. Par conséquent, la diversité génétique présente au sein de l'espèce n'a pas été pleinement explorée. De plus, les allèles rares à forts effets spécifiques au génome de la souche de référence ont eu un impact sur de nombreux traits (la tolérance au sulfate de cuivre et la sensibilité au chlorure de sodium par exemple) et se sont répercutés sur plusieurs croisements. Deuxièmement, seulement 31 conditions de stress sélectionnées de manière à perturber différentes fonctions cellulaires et induire diverses réponses physiologiques ont été testées. Ainsi augmenter le nombre de conditions de stress apporterait une vision plus large de la diversité phénotypique et de la complexité génétique. D'une part, de nouvelles fonctions cellulaires et réponses physiologiques pourraient être explorées telles que la stabilité de l'ADN en réponse à des agents de dommages à l'ADN ou à des inhibiteurs de sa synthèse (l'actinomycine D et l'hydroxyurée par exemple), la progression du cycle cellulaire en présence d'inhibiteurs de différentes phase du cycle cellulaire (la bléomycine et la rapamycine par exemple) ou encore la mort cellulaire en réponse à des composés toxiques ayant des cibles multiples sur la cellule (le métaarsénite de sodium par exemple). D'autre part, les cibles moléculaires pourraient être diversifiées en utilisant d'autres composés toxiques dans les différentes catégories de stress. A titre d'exemple, dans la catégorie stabilité de la membrane, seuls le dodécylsulfate de sodium (SDS) dénaturant les protéines de membrane et le diméthylsulfoxyde (DMSO) perméabilisant la membrane ont été testés. Le blanc de calcofluor et le cercosporamide, respectivement des inhibiteurs de la synthèse de la chitine et du glucane, pourraient compléter la liste de conditions de stress.

Afin de s'affranchir des différents biais énoncés précédemment et d'obtenir une vue plus complète de l'architecture génétique de la variation phénotypique chez *S. cerevisiae*, un nouveau projet est par conséquent en cours de préparation. Ce dernier est basé sur un système de croisements par paire (diallèles) à partir d'une collection de souches naturelles représentatives de la diversité de l'espèce et sur un plus grand nombre de conditions de stress de manière à avoir une gamme phénotypique plus large (Figure 1).

Avec l'achèvement récent du projet de reséquençage des génomes complets de plus de 1002 souches initié dans le laboratoire avec le soutien de collaborations (Projet 1002 génomes de levures : <http://1002genomes.u-strasbg.fr/>), nous disposons à ce jour de la meilleure compréhension de la diversité génétique et phénotypique naturelle de tous les systèmes modèles eucaryotes. A partir de cette riche ressource, l'objectif est de sélectionner 20 souches diploïdes,

homozygotes, génétiquement diverses et ne présentant pas une structure de la population biaisée. A partir de chaque souche diploïde, des souches haploïdes seront générées et ensuite croisées les unes avec les autres afin de constituer un large panel de 210 hybrides diallèles possédant des combinaisons génomiques uniques. Pour chaque croisement, les ségrégants diallèles issus de 20 tétrades complètes sur milieu complet, seront soumis à du phénotypage à haut-débit sur une 50^{aine} de conditions de stress en même temps que les souches parentales et les hybrides.

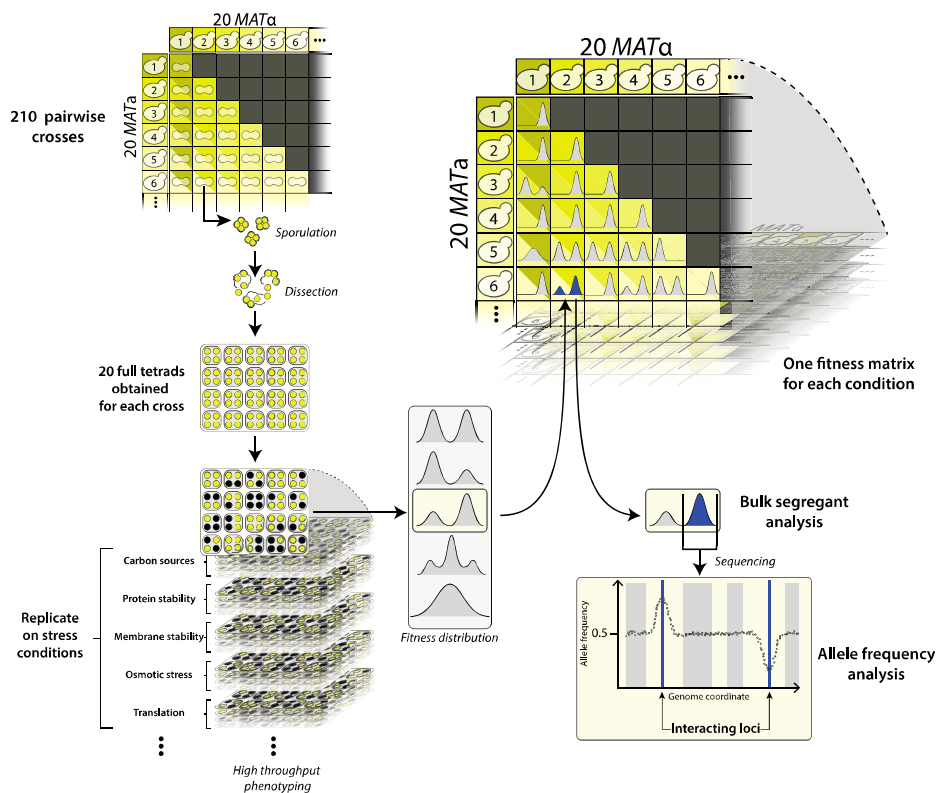


Figure 1. Vue d'ensemble du nouveau projet (réalisée par Téo Fournier)

En comparaison à l'étude précédente, ce système de croisements diallèles apportera non seulement une vue non biaisée de la complexité génétique de la variation phénotypique et informera également quant à la fréquence des variants génétiques causaux dans la population, la dynamique des modèles d'hérédité à travers différents fonds génétiques et la relation entre les combinaisons parentales et les variants génétiques. En plus des analyses de la ségrégation dans la descendance, le panel d'hybrides diallèles offre l'opportunité de tester une analyse d'association pangénomique (GWAS) au sein d'une population possédant des combinaisons alléliques par paire. Il sera également envisageable d'explorer la possibilité de prendre les modèles d'hérédité

comme un caractère complexe et essayer de directement associer la complexité génétique à des loci modificateurs.

De manière globale, le développement fructueux de ce projet apportera sans aucun doute une dimension supplémentaire à la compréhension des effets combinés de variants génétiques sur la variabilité phénotype au sein de populations naturelles de levures et permettra d'élucider l'architecture génétique de caractères complexes.

Evolution de la relation génotype-phénotype chez la levure

Par ailleurs, l'exploration des origines génétiques de la variation phénotypique au sein de levures non conventionnelles serait probablement utile et apporterait à long terme un meilleur aperçu de l'architecture génétique des caractères. En effet, le sous-phylum des levures Saccharomycotina comprend de bons potentiels organismes pour explorer la variation des origines génétiques de la diversité phénotypique pour différentes raisons⁹. Brièvement les espèces de levures Saccharomycotina couvrent une large échelle évolutive au moins aussi vaste que celle de l'embranchement des Chordés¹⁰. De récentes études de génomique des populations ont montré la grande diversité à la fois génétique et phénotypique au sein de multiples espèces de levures, étroitement et lointainement apparentées à *S. cerevisiae*¹¹⁻¹⁹.

Dans ce cadre, nous avons initié l'étude de la relation génotype-phénotype chez *L. kluyveri*, une levure pré-dupliquée éloignée de *S. cerevisiae* dans le but à plus long terme de mettre en évidence des mécanismes moléculaires adaptatifs communs, spécifiques aux levures Saccharomycotina, et des mécanismes propres à l'espèce. Dans un premier temps, l'analyse des variations phénotypiques intra-spécifiques au sein de *L. kluyveri* a représenté une occasion sans précédent d'évaluer la diversité phénotypique au sein d'une espèce protoploïde. Dans le but d'avoir une idée précise de l'étendue de la diversité phénotypique, la variance phénotypique a été comparée à celle de *S. cerevisiae*^{20,21} pour deux types de phénotypes d'intérêt : la croissance et la morphologie des cellules. Il serait à présent pertinent de comparer la variance phénotypique à d'autres espèces de levures pré- et post-dupliquées dont *Saccharomyces paradoxus*, *Saccharomyces uvarum*, *Lachancea thermotolerans*, *Lachancea waltii* et *Kluyveromyces lactis* qui font actuellement l'objet d'une analyse de génomique des populations au sein du laboratoire et pour lesquelles l'analyse de la diversité phénotypique est disponible ou très prochainement. Une étude comparative d'aussi grande envergure permettrait d'avoir une vision plus complète de la diversité phénotypique entre isolats d'une même espèce et entre espèces ainsi que d'identifier des propriétés physiologiques spécifiques à une espèce voir à des isolats en vue de déterminer leurs origines génétiques.

Au vu de ses caractéristiques génétiques¹⁴ et phénotypiques^{15,22}, *L. kluyveri* a dans un second temps fait l'objet d'une première analyse de cartographie de QTL permettant de poser les fondations pour l'étude de la relation génotype-phénotype au sein de cette espèce. L'utilisation du marqueur fluorescent spécifique aux cellules de signe *MATa* s'est avéré utile pour l'isolement rapide d'une population de descendants dans le but de procéder à une cartographie via une analyse de pools de ségrégants. Par cette stratégie des gènes directement associés à la tolérance à de très fortes concentrations en sels ont pu être identifiés dans le fonds génétique de deux souches parentales. Afin d'exploiter pleinement la diversité génétique et phénotypique présente au sein de la population, une perspective serait d'étendre la stratégie à d'autres fonds génétiques et à d'autres conditions de stress. De plus, afin de compléter l'analyse de la relation génotype-phénotype au sein d'espèces différentes de levures, il serait envisageable à plus long terme d'appliquer une telle stratégie à plusieurs espèces pré- et post-dupliquées. Les espèces sélectionnées devraient posséder des cellules haploïdes stables de signes opposés capables de se croiser et de sporuler. Les espèces pré-dupliquées *K. lactis* et *L. waltii* ainsi que les espèces post-dupliquées *S. uvarum* et *S. paradoxus* rassemblent de tels critères.

A plus long terme, l'extension du projet à des paysages génétiques variés permettrait de découvrir plus en détails les règles qui gouvernent la relation génotype-phénotype et d'étudier son évolution à travers le phylum des Saccharomycotina. Ces efforts ne pourront qu'enrichir les connaissances actuelles dans ce domaine et apporter des enseignements précieux pour comprendre les fondements des caractères complexes dans des systèmes eucaryotes supérieurs.

Références

1. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–53 (2009).
2. Visscher, P. M., Hill, W. G. & Wray, N. R. Heritability in the genomics era--concepts and misconceptions. *Nature reviews. Genetics* **9**, 255–66 (2008).
3. Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 1193–8 (2012).
4. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
5. Cirulli, E. T. & Goldstein, D. B. Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* **11**, 415–425 (2010).
6. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
7. Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337**, 100–104 (2012).
8. The UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
9. Peter, J. & Schacherer, J. Population genomics of yeasts: towards a comprehensive view across a broad evolutionary scale. *Yeast* n/a–n/a (2016). doi:10.1002/yea.3142
10. Dujon, B. Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in genetics*: *TIG* **22**, 375–87 (2006).
11. Leducq, J.-B. *et al.* Speciation driven by hybridization and chromosomal plasticity in a wild yeast. *Nature Microbiology* **1**, 15003 (2016).
12. Almeida, P. *et al.* A Gondwanan imprint on global diversity and domestication of wine and cider yeast *Saccharomyces uvarum*. *Nature communications* **5**, 4044 (2014).
13. Peris, D. *et al.* Population structure and reticulate evolution of *Saccharomyces eubayanus* and its lager-brewing hybrids. *Mol. Ecol.* **23**, 2031–2045 (2014).
14. 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**, 184–192 (2015).
15. Brion, C., Pflieger, D., Friedrich, A. & Schacherer, J. Evolution of intraspecific transcriptomic landscapes in yeasts. *Nucl. Acids Res.* gkv363 (2015). doi:10.1093/nar/gkv363
16. Freel, K. C. *et al.* Whole-genome sequencing and intraspecific analysis of the yeast species *Lachancea*

- quebecensis. *Genome Biol Evol* evv262 (2016). doi:10.1093/gbe/evv262
17. Curtin, C. D. & Pretorius, I. S. Genomic insights into the evolution of industrial yeast species *Brettanomyces bruxellensis*. *FEMS Yeast Res.* **14**, 997–1005 (2014).
 18. Hirakawa, M. P. *et al.* Genetic and phenotypic intra-species variation in *Candida albicans*. *Genome Res.* **25**, 413–425 (2015).
 19. Fawcett, J. A. *et al.* Population genomics of the fission yeast *Schizosaccharomyces pombe*. *PloS one* **9**, e104241 (2014).
 20. Warringer, J. *et al.* Trait variation in yeast is defined by population history. *PLoS Genet.***7**, e1002111 (2011).
 21. Yvert, G. *et al.* Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC Syst Biol* **7**, 54 (2013).
 22. Jung, P. P. *et al.* Large-Scale Survey of Intraspecific Fitness and Cell Morphology Variation in a Protoploid Yeast Species. *G3 (Bethesda)* (2016). doi:10.1534/g3.115.026682

ANNEXES

Liste des publications

Sigwalt A, Caradec C, Brion C, Hou J, de Montigny J, Jung PP, Fischer G, Llorente B, Friedrich A & Schacherer J. (In review). Dissection of quantitative traits by bulk segregant mapping in a protoploid yeast species. *FEMS Yeast Research*.

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

Sigwalt A, Jung PP, Ohnuki S, de Montigny J, Ohya Y & Schacherer J. (2016). Large-Scale Survey of Intraspecific Fitness and Cell Morphology Variation in a Protoploid Yeast Species. *G3 (Bethesda)*.

Liste des communications

11^{ème} rencontre internationale des levuristes francophones - Levures, Modèles et Outils

Bordeaux (2014)

Isolation of populations of haploid progeny using fluorescent markers in the *Lachancea kluyveri* yeast species. (Oral)

Séminaire de Microbiologie de Strasbourg

Strasbourg (2013-2015)

A new genetic approach for rapid and efficient selection of haploid cells in the yeast *Saccharomyces kluyveri*. (Poster)

Isolation of populations of haploid progeny using fluorescent markers in *Lachancea kluyveri*. (Poster)

Genetic basis of stress tolerance phenotypes in the yeast *Saccharomyces cerevisiae*. (Oral)

Activité d'enseignement

Poste d'ATER (attaché temporaire d'enseignement et de recherche) en virologie

Faculté des Sciences de la Vie, Université de Strasbourg (2015-2016)

UE Virologie fondamentale (Travaux Dirigés, Licence 3)

UE Techniques pratiquées en virologie (Travaux Pratiques, Licence 3)

UE Formation pratique en virologie moléculaire et cellulaire (Travaux Pratiques, Master 1)

UE Initiation à la démarche scientifique (Exposés, Master 1)

UE Stage en laboratoire en biologie des micro-organismes (Exposés, Master 2)

**Origines génétiques de la variation de tolérance au stress
au sein de populations naturelles de levures**

Résumé

Une question centrale de la génétique moderne est de mieux comprendre comment la variation génétique présente au sein d'individus d'une même espèce influence la diversité phénotypique et l'évolution. La levure modèle *Saccharomyces cerevisiae* offre une occasion unique d'apporter des éléments de réponse à cette question à travers la dissection de l'architecture génétique de la variation de tolérance à des stress environnementaux à l'échelle d'une population. Mon étude révèle un niveau supplémentaire de complexité de la relation génotype-phénotype où finalement les caractères supposés les plus simples, dits Mendéliens (déterminisme strictement monogénique) peuvent se révéler être complexes (déterminisme multigénique) selon le fond génétique en raison de l'action de gènes modificateurs, d'interactions épistatiques et/ou de suppresseurs. Toutefois, les processus évolutifs peuvent être bien différents en fonction des espèces. Afin de mieux les décrypter, je me suis également intéressée à *Lachancea kluyveri*, une levure phylogénétiquement distante de *S. cerevisiae*. Cette espèce présente une diversité génétique plus élevée et constitue une ressource encore peu exploitée. L'exploration de la diversité phénotypique et la détermination de leurs origines génétiques initiées dans cette étude sont extrêmement prometteuses et apportent de solides fondations pour l'étude à la fois de l'architecture génétique des caractères et de l'évolution de la relation génotype-phénotype au sein de diverses espèces de levures.

Mots clés: Levures, tolérance au stress, complexité génétique, phénotypes, croissance, morphologie cellulaire, relation génotype-phénotype, analyse de pool de ségrégants

Summary

Genetic basis of stress tolerance in natural populations of yeast

A central issue of modern genetics is to better understand how genetic variations between individuals within a species influence the phenotypic diversity and the evolution. The budding yeast *Saccharomyces cerevisiae* as a model organism offers a unique opportunity to address this issue through the dissection of the genetic architecture of stress tolerance across a population. My study reveals an additional level of complexity of the genotype-phenotype relationship. Indeed, simple Mendelian traits (monogenic determinism) may become more complex (multigenic determinism) depending on genetic background due to the action of modifier genes, epistatic interactions and / or suppressors. However, evolutionary processes can be very different depending on the species. That is why a non-conventional yeast species namely *Lachancea kluyveri* (formerly *S. kluyveri*) was also studied. This species distantly related to *S. cerevisiae* has a higher genetic diversity and remains a relatively unexplored resource. The exploration of the phenotypic diversity and the determination of the genetic origins initiated in this study lay foundations for the analysis of the genetic architecture of traits and the evolution of the genotype-phenotype relationship within diverse yeast species.

Keywords: Yeast, stress tolerance, genetic complexity, phenotypes, fitness, cell morphology, genotype-phenotype relationship, bulk segregant analysis