

**ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ**

**Laboratoire de Génétique Moléculaire, Génomique et Microbiologie - UMR7156**

**THÈSE** présentée par :

**Cyrielle REISSER**

soutenue le : 31 janvier 2014

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

Discipline/ Spécialité : Sciences du Vivant

**Analyse de la variabilité intraspécifique  
chez les levures : résistance à  
l'ammonium et aux composés azolés**

**THÈSE dirigée par :**

**Mr SCHACHERER Joseph**  
**Mr DE MONTIGNY Jacky**

Maître de conférence, université de Strasbourg  
Professeur, université de Strasbourg

**RAPPORTEURS :**

**Mme NEUVÉGLISE Cécile**  
**Mr DEVAUX Frédéric**

Directeur de recherches, INRA/AgroParisTech  
Professeur, université Pierre et Marie Curie

---

**AUTRES MEMBRES DU JURY :**

**Mme LINSTER Carole**  
**Mr CHARVIN Gilles**

Collaborateur scientifique, Université du Luxembourg  
Chargé de recherches, IGBMC/CNRS



## Remerciements

Le travail présenté dans ce mémoire a été réalisé au sein du laboratoire de Génétique Moléculaire, Génomique et Microbiologie UMR7156 UDS/CNRS. Je tiens tout d'abord à remercier Serge Potier de m'avoir accueilli au sein du laboratoire. Je remercie également Jean-luc Souciet, Serge Potier et Joseph Schacherer de m'avoir respectivement reçu dans l'équipe Évolution des Génomes de Micro-organismes puis dans l'équipe Variation Intraspécifique et Évolution des Génomes. Merci au CNRS et à la région Alsace d'avoir financé ces travaux de thèse.

Je tiens à exprimer toute ma reconnaissance aux membres du jury Cécile Neuveglise, Carole Linster, Frédéric Devaux et Gilles Charvin d'avoir accepté d'évaluer la qualité scientifique de ce travail.

J'adresse un immense merci à mes deux directeurs de thèse ; Joseph Schacherer et Jacky de Montigny. Jacky, merci d'avoir été là pour me soutenir et m'encourager aux moments où j'en avais le plus besoin. Joseph, c'est un immense honneur d'avoir été ta 1<sup>ère</sup> thésarde. Tu auras été LE modèle pour moi durant ces années passées au laboratoire !

Je tenais à adresser un remerciement spécial à André Goffeau qui est à l'origine d'une partie de ma thèse.

Merci à toutes les personnes de l'équipe haplo ! Claudine j'ai été très heureuse d'avoir partagé le même bureau avec toi pendant 4 ans. Grâce à toi, les habits gris à rayures verticales (ou horizontales ?!) auront enfin pris tout leur sens. Anne, merci pour tes conseils, ton soutien bioinformatique à toute épreuve, merci d'avoir réinitialisé 2 ou 3 fois mon mot de passe gem200 (oups !), mais surtout merci pour tous les bons moments qu'on aura passé ensemble que ce soit au labo ou dans les déplacements ! Merci à tous les thésards de l'équipe : Zlatyo, Samuyl, Jackson, Jing à qui je dis « Saint Dié », et Anastasie sur qui l'honneur des microbio repose désormais ! Christian, merci pour ton positivisme à toute épreuve ! Merci aux autres personnes qui font ou ont fait partie de l'équipe : Denise, Sophie, David, Romain, Frédéric, Kelle et Marine. Une pensée pour mon super stagiaire allemand Henri, nous avons beaucoup ri ensemble et nos manip nous auront appris à ne surtout jamais trop espérer ! Je tiens également à remercier Lucie, Aurélie et Fabien qui ont passé leur stage de master au laboratoire en même temps que moi, et avec lesquels j'ai passé de très bons moments ! Enfin merci à toi Pa(ul) de m'avoir tout appris au labo, de m'avoir légué ta paillasse (« avant qu'il ne soit trop tard » !), d'avoir pris tous ces cafés avec moi et également de m'avoir soutenue, consolée et fait beaucoup rire.

## Remerciements

---

Je tiens à remercier toutes les personnes des autres équipes. Sandrine, merci pour nos nombreux bons moments qui me font rire dès que j'y repense, mais merci également d'avoir « râlé » avec moi sur les anomalies régulièrement rencontrées dans les pièces laboratoires ;) Jérémy, Antoine, Pauline, Omniea et Julien, vous avez été de super collègues thésards, courage à vous pour la suite ! Sandrine et Nathalie, je vous remercie pour vos nombreuses aides administratives. Christelle et Thierry, c'est vraiment génial de votre part de toujours passer dire bonjour dans les bureaux du 6<sup>ème</sup> ! Une pensée à Elena qui est arrivée en même temps que moi au laboratoire ! Arigato à Marie Claire de m'avoir donné un aperçu de l'univers japonais. Merci à Benoît, pour son logiciel « courbes de croissances 1.4 », ses conseils, ses critiques, son disque dur externe, son humour et « ses » pâtisseries !

Je tiens également à remercier toutes les personnes du département 2 dont Daphné, Mélanie, Johan, et Ludovic. Un immense merci à Yann et Anne-Marie pour leur nombreux conseils sur la qPCR et autres. Votre présence a illuminé mes expériences de qPCR, c'est dire votre mérite !

Pour finir j'aimerais remercier les personnes qui ne font pas parties du laboratoire mais qui étaient présentes pour moi tout au long de cette thèse. Merci à mes amis de fac avec une mention ++ pour Fanny et Delphine ! Un petit coucou à mes copines de la danse qui me voyaient arriver ces derniers lundis de plus en plus déprimée mais repartir avec la pêche ! Merci à mes amis que je connais depuis plus longtemps encore dont Ludovic, Estelle, Florence, Delphine et ma super Débo (« elle est là ! »). Un grand merci à toute ma famille, en particulier pour les corrections du manuscrit de dernières minutes et pour les plats et desserts déjà cuisinés ! Merci à mon frère Joffrey et à ma soeurette Auryane. Enfin, merci à toi Gaétan, de m'avoir soutenue durant cette thèse, grâce à toi je ne prononcerai plus *Saccharomykes* et *klouyveri* de la même façon ;).

D'une façon générale merci à tous ceux qui m'auront fait avancer durant cette thèse et qui auront cru en moi !

# Table des matières

Introduction.....	1
<b>1. De la génétique mendélienne à la génétique quantitative .....</b>	<b>1</b>
<b>2. Les levures des organismes modèles en génétique, génomique comparative et génomique des populations .....</b>	<b>3</b>
2.1. La levure <i>Saccharomyces cerevisiae</i> .....	3
2.1.1. Une levure modèle .....	3
2.1.2. La diversité écologique et géographique au sein de l'espèce <i>S. cerevisiae</i> .....	5
2.2. La levure <i>Lachancea kluyveri</i> .....	5
<b>3. L'étude de la diversité génétique.....</b>	<b>6</b>
3.1. La diversité interspécifique ou génomique comparative.....	6
3.2. La diversité intraspécifique ou génomique des populations .....	7
3.2.1. Les variations génétiques base de la diversité entre les souches .....	7
3.2.2. La diversité génétique au sein de l'espèce <i>S. cerevisiae</i> .....	9
3.2.3. La diversité génétique au sein de l'espèce <i>L. kluyveri</i> .....	13
3.2.3.1. La diversité du profil des caryotypes des isolats de <i>L. kluyveri</i> .....	14
3.2.3.2. La diversité des génomes mitochondriaux de 18 isolats de <i>L. kluyveri</i> .....	15
3.2.3.3. La diversité du génome nucléaire de 29 isolats de <i>L. kluyveri</i> .....	17
3.3. La diversité interspécifique ou génomique comparative.....	17
3.3.1. La diversité phénotypique au sein des souches de <i>S. cerevisiae</i> .....	18
3.3.1.1. Une pléiade de phénotypes d'intérêt.....	18
3.3.1.2. Les analyses à haut débit .....	20
3.3.2. La diversité phénotypique au sein des souches de <i>L. kluyveri</i> .....	23
<b>4. L'étude de la relation génotype - phénotype .....</b>	<b>25</b>
4.1. Les analyses de liaison.....	25
4.1.1. Le principe de la méthode.....	25
4.1.2. Quelques exemples d'analyse de liaison chez <i>S. cerevisiae</i> .....	27
4.1.3. Les analyses de pool de ségréants .....	27
4.1.3.1. Le principe de la méthode.....	28
4.1.3.2. Quelques exemples d'analyse de pool de ségréants chez <i>S. cerevisiae</i> .....	30
4.2. Les analyses d'association.....	30
4.3. Du pic au gène, à la mutation .....	30
4.3.1. La validation expérimentale des gènes candidats .....	31
4.3.1.1. Le remplacement allélique.....	31

---

4.3.1.2. Le test d'hémizygoté réciproque.....	32
<b>5. Les objectifs de ma thèse.....</b>	<b>32</b>
<b>Matériels et Méthodes.....</b>	<b>35</b>
<b>1. Organismes et méthodes de culture.....</b>	<b>35</b>
1.1. Organismes de l'étude.....	35
1.1.1. <i>Saccharomyces cerevisiae</i> .....	35
1.1.2. <i>Lachancea kluyveri</i> .....	36
1.2. Milieux de culture.....	38
1.3. Conditions de culture.....	39
1.4. Évaluation de la concentration cellulaire.....	39
1.5. Conservation des souches.....	39
<b>2. Techniques génétiques chez la levure.....</b>	<b>39</b>
2.1. Obtention de cellules diploïdes.....	39
2.2. Obtention de cellules haploïdes.....	39
2.2.1. La microdissection.....	39
2.2.2. Le test de spores en vrac.....	40
2.2.2.1 Le test éther pour <i>S. cerevisiae</i> .....	40
2.2.2.2 Le test par billes de verre pour <i>L. kluyveri</i> .....	40
2.3. Détermination du signe sexuel.....	40
2.3.1. Détermination du signe sexuel par croisement.....	40
2.3.2. Détermination du signe sexuel halo mating type.....	40
2.4. Détermination des phénotypes.....	41
2.4.1. Analyse de la croissance par suivi de microcultures.....	41
2.4.1.1. La courbe de croissance et les caractères quantitatifs mesurés.....	41
2.4.1.2. Analyse des courbes de croissance.....	43
<b>3. Préparation des acides nucléiques.....</b>	<b>43</b>
3.1. Préparation de l'ADN.....	43
3.1.1. Extraction d'ADN génomique de levure par la méthode phénol chloroforme.....	43
3.1.2. Extraction d'ADN génomique total par kit Qiagen.....	44
3.1.3. Préparation de chromosomes entiers de levure.....	44
3.2. Extractions des ARN totaux.....	44
<b>4. Purification, quantification et analyse des acides nucléiques.....</b>	<b>45</b>
4.1. Purification de l'ADN.....	45
4.1.1. Précipitation alcoolique de l'ADN.....	45
4.1.2. Purification de l'ADN à l'hexamine cobalt.....	46

---

4.2. Purification des ARN par traitement à la DNase .....	46
4.3. Contrôle et dosage des acides nucléiques .....	46
4.3.1. Analyse des acides nucléiques par électrophorèse sur gel d'agarose .....	46
4.3.2. Quantification des acides nucléiques .....	46
4.3.3. Séparation des chromosomes par électrophorèse en champs alternés .....	47
<b>5. Amplification d'un fragment d'ADN par PCR .....</b>	<b>48</b>
5.1. Principe.....	48
5.2. Milieux de PCR.....	48
<b>6. Dosage d'ARN par RT-PCR quantitative.....</b>	<b>49</b>
6.1. Réverse transcription des ARN totaux par amorçage aléatoire.....	50
6.2. Réaction de qPCR .....	50
6.3. Analyse des résultats de qPCR.....	51
<b>7. Séquençage .....</b>	<b>52</b>
7.1. Séquençage des génomes par la technique Illumina HiSeq 2000 .....	52
7.2. Séquençage des fragments PCR.....	52
<b>8. Techniques de transformation des levures.....</b>	<b>53</b>
8.1. Transformation par électroporation.....	53
8.1.1. Protocole pour <i>S. cerevisiae</i> .....	53
8.1.1.1. Préparation des cellules électro-compétentes .....	53
8.1.1.2. Électroporation.....	53
8.1.2. Protocole pour <i>L. kluyveri</i> .....	53
8.1.1.1. Préparation des cellules électro-compétentes .....	53
8.1.1.2. Électroporation.....	53
8.2. Transformation par choc thermique avec acétate de lithium .....	54
8.2.1. Préparation des cellules compétentes.....	54
8.2.2. Choc thermique.....	54
<b>9. Bioinformatique.....</b>	<b>54</b>
9.1. Les banques de données .....	54
9.2. Analyse des données de séquençage de l'ADN génomique .....	55
9.2.1. Nettoyage des séquences.....	55
9.2.3. Détermination des SNP.....	55
9.2.4. Analyse des couvertures .....	55
9.2.5. Représentations graphiques des données de séquençage grâce au logiciel R.....	56
9.3. Alignement et assemblage des séquences issues du séquençage de fragments PCR.....	56
9.4. Construction d'arbres phylogénétiques.....	56
9.5. La représentation des données de résistance .....	56

Chapitre 1 : Les bases génétiques de la résistance à l'ammonium : un cas mendélien .....	57
<b>1. Introduction</b> .....	<b>57</b>
<b>2. Manuscrit</b> .....	<b>61</b>
<b>3. Discussion et conclusion</b> .....	<b>69</b>
Chapitre 2 : Analyse de la variabilité intraspécifique de la résistance aux composés azolés chez <i>S. cerevisiae</i> et <i>L. kluyveri</i> .....	71
<b>Introduction</b> .....	<b>71</b>
<b>1. <i>S. cerevisiae</i>, un pathogène opportuniste</b> .....	<b>71</b>
<b>2. Les antifongiques</b> .....	<b>72</b>
2.1. Les polyènes .....	72
2.2. Les analogues de la pyrimidine .....	73
2.3. Les composés azolés .....	73
2.4. Les candines .....	74
2.5. Les allylamines, thiocarbamates et morpholines .....	75
2.6. Quels traitements pour <i>S. cerevisiae</i> ? .....	75
<b>3. Les mécanismes de résistance aux composés azolés chez les levures</b> .....	<b>75</b>
3.1. Modification de la cible .....	75
3.2. Surexpression de la cible .....	76
3.3. L'altération de la voie de biosynthèse des stérols .....	76
3.4. L'entrée de l'antifongique dans la cellule .....	76
3.5. Les mécanismes de résistance multiple aux drogues (MDR) .....	77
3.5.1. La formation de biofilms .....	77
3.5.2. L'expulsion de l'antifongique par des transporteurs .....	77
3.5.2.1. Les transporteurs PDR .....	77
3.5.2.2. Les facteurs de transcription de type PDR .....	81
<b>Résultats</b> .....	<b>83</b>
<b>1. Étude de la variabilité des gènes de résistance multiple</b> .....	<b>83</b>
1.1. Analyse des gènes <i>PDR</i> chez <i>S. cerevisiae</i> .....	83
1.1.1. Analyse du gène <i>PDR5</i> .....	84
1.1.1.1. Analyse de la distribution du polymorphisme du gène <i>PDR5</i> des souches cliniques .....	85
1.1.1.2. Analyse des séquences du gène <i>PDR5</i> .....	86

1.1.1.3. Hypothèses sur l'origine de la version allélique de <i>PDR5</i> des 5 souches cliniques.....	87
1.1.1.4. Analyse de la séquence protéique de Pdr5p .....	88
1.1.1.5. Analyse par région de la séquence protéique de Pdr5p .....	89
1.1.1.6. Recherche sur l'origine des régions hautement polymorphiques .....	91
1.2. Analyse des gènes <i>PDR</i> chez <i>L. kluyveri</i> .....	91
<b>2. Étude de la résistance aux antifongiques au sein des deux espèces.....</b>	<b>93</b>
2.1. Stratégie de phénotypage par microcultures.....	93
2.1.1. Choix du milieu de référence .....	94
2.1.2. Estimation de la résistance .....	94
2.2. Variation intraspécifique de la résistance aux composés azolés chez <i>S. cerevisiae</i> .....	95
2.3. Variation intraspécifique de la résistance aux composés azolés chez <i>L.kluyveri</i> .....	96
<b>3. Détermination de l'origine génétique de la résistance par analyse de pool de ségrégants ...</b>	<b>97</b>
3.1. Design expérimental .....	97
3.1.1. Choix des antifongiques.....	97
3.1.2. Choix des parents .....	98
3.1.2.1. <i>S. cerevisiae</i> des souches parentales aux phénotypes extrêmes.....	98
3.1.2.2. <i>L. kluyveri</i> une sélection de parents aux phénotypes intermédiaires .....	98
3.1.3. Génération des ségrégants.....	99
3.1.4. Phénotypage des ségrégants.....	99
3.1.4.1. La résistance des ségrégants au fluconazole.....	100
3.1.4.2. La résistance des ségrégants au kétoconazole .....	101
3.2. Analyse des séquences des pools de ségrégants du couple Sace_clinique.....	102
3.2.1. Stratégie d'analyse du séquençage.....	103
3.2.2. Résultats du pool de résistance au fluconazole.....	105
3.2.2.1. Identification d'un pic au niveau du chromosome XV.....	106
3.2.2.2. Contenu génétique du pic .....	106
3.2.2.3. Analyse de l'expression du gène <i>PDR5</i> .....	108
3.2.2.4. La séquence promotrice du gène <i>PDR5</i> .....	109
3.2.3. Résultats du pool de résistance au kétoconazole.....	112
3.2.3.1 Identification des QTL.....	113
3.2.3.1.1. Détection d'un pic sur le chromosome XI.....	113
3.2.3.1.2. Détection d'un pic sur le chromosome XV .....	113
3.3. Analyse des pools de ségrégants du couple de <i>L. kluyveri</i> Lakl_Japon .....	113
3.3.1. Résultats du pool de résistance au fluconazole et au kétoconazole .....	114
3.3.2. Conclusion de l'analyse pour le couple Sakl_Japon et perspectives .....	115

3.4. Analyse des séquences des pools de ségréants issus du couple de <i>L. kluyveri</i> Sakl_Europe .....	116
3.4.1. Les résultats du séquençage du groupe contrôle .....	116
3.4.2. Comparaisons entre le groupe contrôle et les groupes de ségréants résistants au fluconazole et au kétoconazole .....	117
3.4.3. Conclusion des pools de ségréants issus du couple Lakl_Europe.....	117
<b>Conclusions générales et perspectives .....</b>	<b>119</b>
<b>Chapitre 3 : Expériences d'évolution en présence de fluconazole de souches de <i>S. cerevisiae</i> et <i>L. kluyveri</i>.....</b>	<b>121</b>
<b>1. Introduction .....</b>	<b>121</b>
<b>2. Manuscrit .....</b>	<b>125</b>
<b>3. Discussion et perspectives .....</b>	<b>155</b>
3.1. L'expérience d'évolution.....	155
3.2. Le mécanisme d'adaptation .....	155
3.3. Les origines de la résistance .....	156
<b>Références bibliographiques .....</b>	<b>159</b>
<b>Annexes .....</b>	<b>169</b>
Annexe MM1 : Liste des amorces utilisées dans le chapitre 2 .....	169
Annexe 2.1 : Divergence génétique entre les souches de <i>L. kluyveri</i> .....	170
Annexe 2.2 : Liste des gènes présents dans le QTL du chromosome XV des analyses de pool de ségréants du couple Sace_clinique au fluconazole et au kétoconazole .....	171
Publications, communications et activité d'enseignement .....	173

# Liste des figures

## Introduction

Figure 1 : Arbre phylogénétique des Hémiascomycètes.....	4
Figure 2 : Arbre de type Neighbour-joining basé sur les sites polymorphiques détectés dans la collection de 63 souches de <i>S. cerevisiae</i> .....	11
Figure 3 : Arbre de type Neighbour-joining obtenu à partir des séquences concaténées de 9 gènes et de 4 régions intergéniques .....	13
Figure 4 : Caryotype en champs pulsés des 29 souches de <i>L. kluyveri</i> séquencées.....	14
Figure 5 : Arbre de type Neighbor-joining de 18 souches de <i>L. kluyveri</i> basé sur les séquences codantes des gènes mitochondriaux .....	16
Figure 6 : Arbre de type Neighbor-joining basé sur les 881 427 sites polymorphiques détectés au sein de la collection des <i>L. kluyveri</i> .....	18
Figure 7 : Variations phénotypiques au sein du genre des <i>Saccharomyces sensu stricto</i> .....	21
Figure 8 : Variabilité intraspécifique de trois caractères morphologiques au sein de 37 souches de <i>S. cerevisiae</i> .....	23
Figure 9 : Regroupement hiérarchique des profils phénotypiques de 40 souches de <i>L. kluyveri</i> .....	24
Figure 10 : Schéma d'une analyse de liaison chez <i>S. cerevisiae</i> .....	26
Figure 11 : Corrélation entre les marqueurs génétiques contenus dans les gènes 1 et 2 et le phénotype d'intérêt .....	27
Figure 12 : Schéma d'une analyse de pool de ségréants basée sur un trait simple.....	28
Figure 13 : Principe de l'analyse d'association chez la levure .....	31
Figure 14 : Schéma du remplacement allélique .....	31
Figure 15 : Schéma du test d'hémizygocité réciproque.....	32

## Matériels et Méthodes

Figure MM1 : Cycle cellulaire de <i>Saccharomyces cerevisiae</i> .....	35
Figure MM2 : Stratégie de phénotypage employée permettant de tester la résistance des souches aux composés azolés.....	42
Figure MM3 : Illustration d'une courbe de croissance avec ses 3 premières phases caractéristiques ..	42
Figure MM4 : Schéma récapitulatif des expériences à réaliser, en partant d'une culture cellulaire, pour quantifier des ARN .....	49
Figure MM5 : Profil de fluorescence d'une réaction de qPCR.....	51

## Chapitre 1 : Les bases génétiques de la résistance à l'ammonium : un cas mendélien

Figure 1.1 : Modèle de la toxicité à l'ammonium et de la détoxification chez la levure ..... 58

## Chapitre 2 : Analyse de la variabilité intraspécifique de la résistance aux composés azolés chez *S. cerevisiae* et *L. kluyveri*

Figure 2.1 : La voie de biosynthèse de l'ergostérol chez les levures et les étapes d'inhibition des composés azolés..... 74

Figure 2.2 : Schéma de la topologie d'un transporteur de type PDR avec ses principales caractéristiques..... 78

Figure 2.3 : Relations entre les gènes *PDR sensu stricto* basées sur l'analyse des gènes adjacents ..... 79

Figure 2.4 : Distribution du polymorphisme entre les souches YJM789 et S288c du gène *PDR5* ..... 80

Figure 2.5 : La voie PDR classique chez la levure ..... 81

Figure 2.6 : Polymorphisme en SNP/kb de 8 gènes *PDR* au sein de la collection de *S. cerevisiae* ..... 84

Figure 2.7 : Représentation du polymorphisme du gène *PDR5* pour les souches cliniques..... 86

Figure 2.8 : Exemple d'alignement de la région hautement polymorphique 1 sur Jalview ..... 87

Figure 2.9 : Cladogramme de type Neighbor-joining basé sur la séquence protéique du gène *PDR5* complet..... 88

Figure 2.10 : Cladogrammes de type Neighbor-joining basés sur la séquence protéique de la région conservée 2 (A) et la région hautement polymorphique 2 (B) de *PDR5*. ..... 90

Figure 2.11 : Polymorphisme en SNP/kb des gènes impliqués dans la résistance multiple chez *L. kluyveri*..... 92

Figure 2.12 : Les antifongiques azolés utilisés dans le criblage des collections ..... 93

Figure 2.13 : Comparaison des vitesses spécifiques de croissance  $\mu$  ( $h^{-1}$ ) de 40 *S. cerevisiae* sur YPG et sur YNB ..... 94

Figure 2.14 : Résistance aux composés azolés de 28 souches de *S. cerevisiae* ..... 95

Figure 2.15 : Résistance aux composés azolés de 40 souches de *L. kluyveri* ..... 97

Figure 2.16 : Vitesse spécifique de croissance de 88 ségréants par couple sur milieu complet supplémenté de  $4.10^{-5}$  M de fluconazole ..... 101

Figure 2.17 : Vitesse spécifique de croissance de 88 ségréants par couple sur YPG supplémenté de  $5.10^{-6}$  M de kétoconazole..... 102

Figure 2.18 : Schéma de la stratégie globale employée, débutant par le séquençage de l'ADN du pool jusqu'à la détection des QTL ..... 104

Figure 2.19 : Analyse de la fréquence allélique du pool de résistance au fluconazole Sace\_clinique 105

Figure 2.20 : Expression des gènes du QTL par rapport à la souche de référence S288c ..... 107

---

Figure 2.21 : Analyse des séquences PDRE dans le promoteur de <i>PDR5</i> .....	110
Figure 2.22 : Cladogramme de type Neighbor-joining basé sur la séquence nucléique de la région intergénique <i>YOR152c</i> et <i>PDR5</i> .....	111
Figure 2.23 : Représentation de la fréquence allélique le long du génome de référence dans le pool de résistance Sace_clinique au kétoconazole .....	112
Figure 2.24 : Représentation de la fréquence allélique le long du génome dans le pool de résistance Lakl_Japon au fluconazole .....	114
Figure 2.25 : Représentation de la fréquence allélique le long du génome dans le pool de résistance Lakl_Japon au kétoconazole .....	115
Figure 2.26 : Représentation de la fréquence allélique du couple contrôle .....	116
Figure 2.27 : Analyse des séquences des pools de ségréants Sakl_Europe. ....	118

### Chapitre 3 : Expériences d'évolution en présence de fluconazole de souches de *S. cerevisiae* et *L. kluyveri*

Figure 3.1 : Expérience d'évolution et stratégie d'analyse de la résistance .....	123
Figure 3.2 : Stratégie des délétions par région des chromosomes surnuméraires.....	158

# Liste des tableaux

## Introduction

Tableau 1 : Exemple de souches de <i>S. cerevisiae</i> complètement séquencées .....	12
--	----

## Matériels et Méthodes

Tableau MM1 : Les de souches de <i>S. cerevisiae</i> utilisées ainsi que leurs origines écologiques et géographiques .....	36
Tableau MM2 : Les de souches de <i>L. kluyveri</i> utilisées ainsi que leurs origines écologiques et géographiques .....	37
Tableau MM3 : Antifongiques additionnés au milieu YPG et concentrations finales .....	38
Tableau MM4 : Amorces utilisées pour la qPCR .....	51

## Chapitre 2 : Analyse de la variabilité intraspécifique de la résistance aux composés azolés chez *S. cerevisiae* et *L. kluyveri*

Tableau 2.1 : Liste des gènes <i>PDR</i> sélectionnés pour l'analyse de la variabilité génétique .....	83
Tableau 2.2 : Souches dont le gène <i>PDR5</i> a été séquencé .....	87
Tableau 2.3 : Mesure de l'expression du gène <i>PDR5</i> chez YJM326 par rapport à la souche YJM280 en YPG et YPG supplémenté de fluconazole .....	108

## Chapitre 3 : Expériences d'évolution en présence de fluconazole de souches de *S. cerevisiae* et *L. kluyveri*

Tableau 3.1 : Liste des gènes candidats pour expliquer la résistance des souches évoluées au fluconazole .....	156
--	-----

## Liste des abréviations

5-FC	5-fluorocytosine
A <sub>600</sub>	Absorbance à 600 nm
aa	Acides aminés
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
AmB	Amphotéricine B
ARN	Acide ribonucléique
BET	Bromure d'éthidium
C	Cytosine
CM	Clotrimazole
Ct	Threshold cycle
DMSO	Diméthylsulfoxyde
dN	Mutations Non synonymes
dNTP	2'-désoxyribonucléoside-5'-triphosphate
dS	Mutations Synonymes
DTT	Dithiothréitol
E	Efficacité / biomasse générée
EDTA	Éthylène diamine tétraacétate
FC	Fluconazole
g	Temps de génération
G	Guanine
h	Heures
IC	Itraconazole
Indel	Insertion/Déletion
K <sup>+</sup>	Ion potassium
kb	Kilobase
KC	Kétoconazole
L	Phase de latence
MC	Miconazole
MDR	Multi Drug Resistance
min	Minutes
MQ	Eau MilliQ
NH <sub>4</sub> <sup>+</sup>	Ion ammonium
ORF	Open Reading Frame
pb	Paire(s) de base
PCR	Polymerase Chain Reaction
PDR	Pleiotropic Drug Resistance
PDRE	Pleiotropic Drug Reponse Elements
qsp	Quantité suffisante pour
QTL	Quantitative Trait Locus
rpm	Rotations par minute
RT	Rétro-transcription

s	Secondes
SNP	Single Nucleotide Polymorphism
T <sub>m</sub>	Temperature Melting
Tris	Tris-hydroxyméthyl-aminométhane
UV	Ultraviolet
WGD	Whole Genome Duplication
Δ	Délétion
μ	Vitesse spécifique de croissance

# Introduction

## 1. De la génétique mendélienne à la génétique quantitative

En 1856, le botaniste Gregor Mendel (1822-1884) débute des expériences de croisements entre des variétés homozygotes de petits pois. Il analyse la descendance hétérozygote en se basant sur sept caractères facilement identifiables (forme et couleur de la graine, couleur de l'enveloppe, forme et couleur de la gousse, position des fleurs et longueur de la tige). Il démontre le rôle équivalent des deux parents dans l'hérédité mais certains caractères transmis sont dominants alors que d'autres sont récessifs. En 1866, il publie les lois de l'hérédité qui proposent notamment que chaque parent transmette des facteurs différents qui gouvernent les caractères. Malgré leur importance, ces travaux n'ont été réellement reconnus par la communauté scientifique que 30 ans plus tard en 1900 grâce aux publications indépendantes de trois botanistes : Hugo De Vries (1848-1935) aux Pays-Bas, Carl Erich Correns (1864-1933) en Allemagne et Erich Tschermak-Seysenegg (1871-1962) en Autriche. Les expériences de Mendel se basent sur des traits discrets qui se distinguent nettement les uns des autres. La variation de ces traits est dite discontinue. C'est en hommage au généticien que ces caractères sont également appelés mendélien. Cependant, la plupart des caractères sont quantitatifs, la variation des traits est alors dite continue car ils s'expriment à travers une palette de phénotypes à l'image de la taille des individus chez l'Homme, c'est pourquoi les lois de Mendel sont dans un premier temps controversées par ses contemporains.

Charles Darwin qui est un contemporain de Gregor Mendel avait quant à lui mis en évidence la variation continue des traits et y voyait un impact sur l'adaptation des espèces à un nouvel environnement. Il écrivit à ce sujet dans son livre « De l'origine des espèces » paru en 1859 que si un individu varie légèrement d'une manière qui lui est profitable, il possèdera un avantage sur les autres et aura de meilleures chances pour survivre. Charles Darwin est le père fondateur de ce qui est appelé aujourd'hui la théorie de l'évolution (1859). Selon cette théorie, tous les êtres vivants ont évolué au fil du temps grâce au processus de « sélection naturelle ». Ce processus repose sur trois principes 1) les individus issus d'une même espèce présentent des variations, 2) les individus les plus adaptés au milieu survivront et se reproduiront davantage que les autres et 3) les caractères conférant un avantage sont héréditaires. La nature de ces variations était à l'époque inconnue et le lien entre les théories de Mendel et de Darwin n'a été fait que quelques décennies plus tard.

Les travaux du statisticien Ronald Aylmer Fisher ont permis d'établir le lien entre les lois de l'hérédité basées sur des traits discontinus et les nombreux caractères continus observés dans la nature en prouvant que plusieurs facteurs mendéliens pouvaient impacter sur l'expression d'un même caractère et expliquer une variation de type continue (Fisher 1918).

Comme expliqué précédemment, la majorité des caractères exhibent une variation continue. Dans la plupart des cas, ces caractères sont dus à plusieurs gènes qui ont un impact partiel sur le phénotype. La présence d'une variation phénotypique continue est donc un bon indicateur d'un trait quantitatif. Pour expliquer la génétique quantitative nous allons nous baser sur un exemple simple. Si un trait quantitatif est contrôlé par 5 gènes dans un génome haploïde, et que pour chaque gène il existe deux allèles, l'allèle – et +. L'allèle + ajoutera une unité au phénotype tandis que l'allèle – ne rajoutera rien. Pour chaque gène il y aura donc deux possibilités ce qui donne  $2^5$  soit 32 génotypes possibles. Le génotype (-----) donnera le phénotype 0 tandis que le génotype (+++++) aura le phénotype 5. Au total 10 génotypes permettent un phénotype 2 qui sera constitué par la combinaison de 2 + et de 3 – tels que dans les génotypes (---++), (-+--+). Chaque classe phénotypique comporte donc ses propres variations qui constituent en elles-mêmes une variation continue.

Dans l'équipe variation intra-spécifique et évolution des génomes, au laboratoire de génétique moléculaire, génomique et microbiologie, les liens entre le phénotype et le génotype sont étudiés en utilisant les levures *Saccharomyces cerevisiae* et *Lachancea kluyveri* (anciennement appelée *Saccharomyces*) comme organismes modèles (Kurtzman 2003). L'étude de ce lien est réalisée à travers plusieurs exemples concrets d'intérêt médical ou évolutif tels que la résistance à l'ammonium, la résistance aux antifongiques, l'isolement reproductif et la diversité du contenu en éléments transposables des génomes. Le but de nos recherches est de comprendre les règles qui gouvernent les relations entre le phénotype et le génotype. Pour cela, nous combinons des méthodes de génotypage et de phénotypage à haut débit. La première étape consiste à obtenir des mesures quantitatives pour l'ensemble des phénotypes. La deuxième étape consiste en la cartographie des gènes impliqués dans la variation phénotypique observée. Le travail de ce manuscrit s'est focalisé sur l'analyse des variations de résistance à l'ammonium et aux composés azolés ainsi que de leurs origines génétiques au sein des deux levures modèles.

## 2. Les levures : des organismes modèles en génétique, génomique comparative et génomique des populations

Les levures sont des organismes formant un groupe polyphylétique car elles sont apparues à partir de plusieurs embranchements du règne des mycètes. Elles sont réparties en deux phylums, les *Basidiomycètes* et les *Ascomycètes*. Les levures Hémiascomycètes nouvellement appelées les *Saccharomycotina* constituent un des trois sous-clades des *Ascomycètes*. Il existe près de 963 espèces déjà identifiées dans le clade des Hémiascomycètes dont 61 espèces ont leur génome séquencé. Elles sont communément appelées le groupe des « levures bourgeonnantes ». Ce nom est donné en référence à leur mode de reproduction asexué dans lequel la cellule fille se crée à partir d'une division asymétrique en forme de bourgeon depuis la cellule mère.

Des levures aux modes de vie très différents font partie de ce clade comme les pathogènes opportunistes *Candida albicans* et *Candida glabrata*, le pathogène de plante *Ashbya gossypii* ou bien la levure de boulangerie *S. cerevisiae*. Dans l'histoire évolutive des Hémiascomycètes, il y a eu une duplication totale du génome (figure 1) (Wolfe and Shields 1997). Les espèces dont les ancêtres ont divergé après la duplication totale du génome sont dites post-dupliquées. Une partie de ces levures (les *Saccharomyces sensu stricto*) possèdent un nombre de chromosomes doublé (16 chromosomes) par rapport aux levures pré-dupliquées. Les autres levures post-dupliquées ont quant à elles subi une réduction de chromosomes (Gordon *et al.* 2011). Dans ce travail, les deux levures modèles *S. cerevisiae* et *L. kluyveri* ont été utilisées : la première est une levure post dupliquée tandis que la seconde est une levure protoploïde (un des phylum des levures Hémiascomycètes pré-dupliquées).

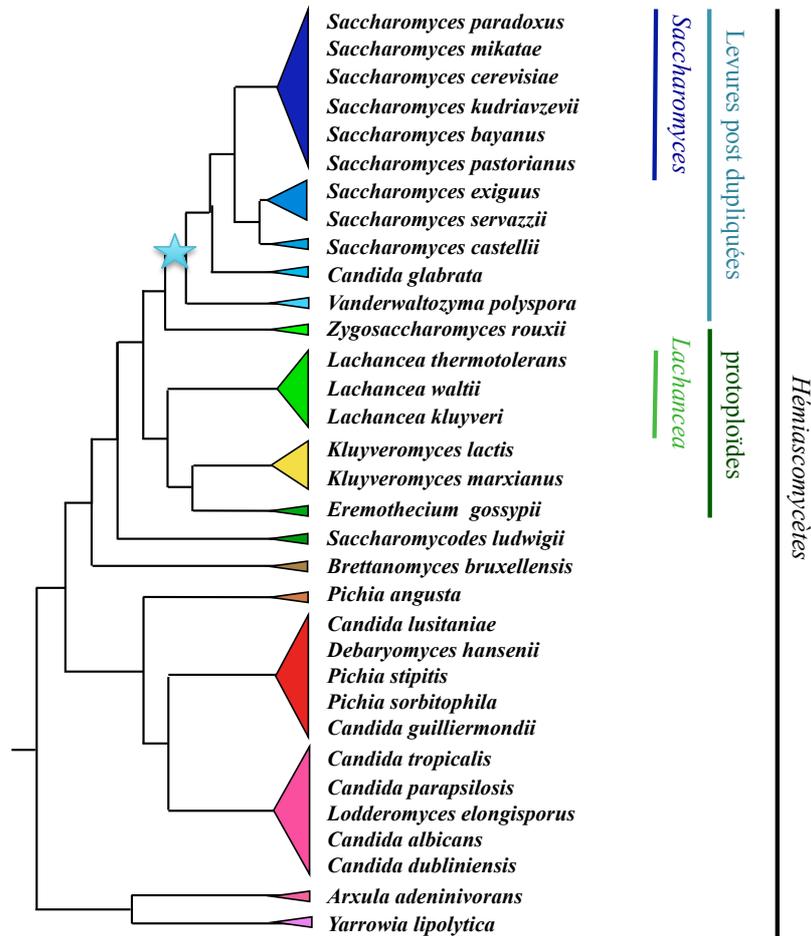
### 2.1. La levure *Saccharomyces cerevisiae*

#### 2.1.1. Une levure modèle

La levure *S. cerevisiae* est utilisée à travers le monde dans des processus de panification et de fermentation alcoolique (vin, bière et autres breuvages alcoolisés). L'analyse d'amphores vieilles de 3150 ans retrouvées en Égypte indique qu'elle était déjà utilisée à cette époque dans la fabrication du vin (Cavalieri *et al.* 2003). Une autre étude prouve même l'utilisation des levures en Chine 9000 ans avant JC pour la fermentation d'autres aliments tels que le riz, le miel et des fruits (McGovern *et al.* 2004). Cette levure est également isolée à partir de milieux naturels comme le sol, les exsudats d'arbres et les intestins de drosophiles. Elle peut également se révéler être un pathogène opportuniste en infectant des patients immunodéprimés (Wei *et al.* 2007).

Dans les années 1950, la levure s'est d'abord imposée comme un organisme modèle en génétique. Plusieurs concepts de la génétique classique tels que l'étude de la recombinaison méiotique et les cartes génétiques ont ainsi pu être étudiés. Dans les années 1970, elle a été utilisée comme un modèle

en biologie cellulaire et moléculaire notamment de part son caractère unicellulaire et de part un grand nombre de processus cellulaires hautement conservés à travers les eucaryotes. Son génome de 12,1 Mb répartis sur 16 chromosomes est compact : il ne possède qu'un faible nombre d'éléments répétés non codants et un faible nombre d'introns. Tous ces atouts lui ont permis d'être le premier eucaryote dont le génome a été entièrement séquencé (Goffeau *et al.* 1996).



**Figure 1 :** Arbre phylogénétique des Hémiascomycètes. Sur cette figure sont représentées les espèces dont les génomes ont été séquencés. Chaque couleur regroupe les levures d'un même clade (adapté de Dujon 2006). L'étoile bleue symbolise la duplication totale du génome.

D'un point de vue expérimental, *S. cerevisiae* est facilement cultivable et de nombreuses techniques de génétique existent pour cet organisme. Il est ainsi possible de construire des souches mutées grâce à des techniques classiques de biologie moléculaire. Une collection de délétions systématiques de 96% des gènes a été construite dans le but de réaliser une analyse fonctionnelle globale de l'ensemble du génome. Pratiquement tous les gènes de la levure ont été délétés et remplacés par le gène *KanMX* entraînant une résistance à la généticine. Les gènes non-essentiels à la survie cellulaire ont été délétés dans un contexte haploïde. Les gènes essentiels l'ont été dans un

contexte diploïde hétérozygote (Giaever *et al.* 2002). Près de 85% des gènes ont des fonctions caractérisées ce qui est un pourcentage bien supérieur à tous les autres eucaryotes.

Pour tous ces avantages, *S. cerevisiae* est l'organisme modèle privilégié pour le développement des nouvelles méthodes à haut débit, que ce soit en génotypage, en phénotypage, en transcriptomique ou en protéomique (Botstein and Fink 2011).

### **2.1.2. La diversité écologique et géographique au sein de l'espèce *S. cerevisiae***

Bien qu'étant une des premières espèces domestiquées par l'homme pour sa capacité à dégrader les sucres en alcool et CO<sub>2</sub>, *S. cerevisiae* est avant tout une levure d'origine naturelle. Les recherches sur cette espèce se sont essentiellement limitées aux souches de laboratoire qui ont été isolées depuis des dizaines d'années à partir de leur environnement naturel et cultivées durant des centaines de générations en conditions de laboratoire. L'utilisation exclusive de ces souches s'explique par leurs phénotypes propices à la culture en condition de laboratoire comme notamment l'absence de floculation et de croissance hyphale. Le séquençage du génome de la souche de laboratoire S288c a également joué un grand rôle dans son utilisation massive (Goffeau *et al.* 1996). Les souches de laboratoire ne reflètent pourtant pas toute la diversité présente au sein de l'espèce et elles ne représentent pas non plus son mode de vie dans la nature. Certaines de ces souches de laboratoire se révèlent même être génétiquement très proches (Schacherer, Ruderfer, *et al.* 2007). Ces dernières années de nombreux isolats de *S. cerevisiae* provenant de différents milieux naturels ont été caractérisés (Fay and Benavides 2005, Wang *et al.* 2012, Aa *et al.* 2006, Sniegowski *et al.* 2002). Ils révèlent que *S. cerevisiae* est avant tout une espèce d'origine naturelle.

## **2.2. La levure *Lachancea kluyveri***

*Lachancea kluyveri* est une levure bourgeonnante protoploïde, c'est à dire qu'elle n'est pas issue d'une espèce dans laquelle le génome a été dupliqué. Elle a donc divergé de la lignée des *Saccharomyces* avant la duplication totale du génome (figure 1). *L. kluyveri* est une espèce retrouvée dans différents écosystèmes d'origine naturelle, des souches ont notamment été isolées à partir d'exsudats d'arbres, d'intestin de drosophile, d'une caverne, de sol, de mousse et de feuilles en décomposition. Contrairement à *S. cerevisiae*, *L. kluyveri* est une levure décrite comme étant hétérothallique ce qui signifie que les cellules haploïdes  $a$  et  $\alpha$  sont stables et ne peuvent pas changer de signe sexuel (Payen *et al.* 2009; McCullough and Herskowitz 1979). Ceci est notamment dû au fait qu'elles ne possèdent pas les cassettes HML et HMR ainsi que le gène codant l'endonucléase HO permettant le switch du signe sexuel (Payen *et al.* 2009).

En 2009, dans le but d'explorer les génomes de levures n'ayant pas subi la duplication de génome, le Consortium Génolevures a analysé les génomes de trois espèces dont *L. kluyveri* (Souciet *et al.* 2009). Cette levure possède un génome de 11,3 Mb répartis sur 8 chromosomes. Le séquençage

systematique du génome a permis de montrer qu'une région de 1 Mb située sur le bras gauche du chromosome C se distingue du reste du génome en terme de composition nucléotidique (Souciet *et al.* 2009). Cette région est dite GC riche car 53% des nucléotides sont des guanine (G) ou cytosine (C) contrairement au reste du génome où ce pourcentage est de 40%. L'origine de cette région reste pour l'instant inconnue.

De nombreux outils et méthodes de génétiques développés chez *S. cerevisiae* sont utilisables chez *L. kluyveri* comme la transformation par des vecteurs de type plasmidique, le remplacement allélique, la séparation des chromosomes en champs pulsés et l'obtention de ségrégants haploïdes (Fujimura 1991; Neuvéglise *et al.* 2000). Grâce à ces nombreuses méthodes et à la connaissance d'un génome de référence déjà séquencé, nous avons commencé à utiliser cette levure comme modèle au sein de notre équipe.

### 3. L'étude de la diversité génétique

#### 3.1. La diversité interspécifique ou génomique comparative

Le séquençage de *S. cerevisiae* en 1996 a permis de mettre en évidence des blocs de duplication dans le génome qui ont conduit à poser l'hypothèse que ce génome est issu d'une duplication totale du génome (Wolfe and Shields 1997). La comparaison avec le génome de la levure protoploïde *Lachancea waltii* a permis de valider cette hypothèse de duplication totale du génome dans l'histoire évolutive des Hémiascomycètes (Kellis *et al.* 2004). Bien que cet événement ait doublé le nombre de chromosomes au sein des levures du genre *Saccharomyces*, il a été suivi d'une perte massive de nombreux gènes dupliqués. Au total chez *S. cerevisiae*, seules 550 paires de gènes homologues issues de cet événement subsistent encore (Byrne and Wolfe 2005). Cette perte explique que le génome de *S. cerevisiae* soit seulement 13% supérieur en taille par rapport à celui de *L. waltii*, et présente seulement 10% de gènes supplémentaires (Kellis *et al.* 2004).

Depuis le séquençage de *S. cerevisiae* en 1996 et grâce à l'évolution des méthodes de séquençage, près de 80 génomes de levures ont été séquencés. Les Hémiascomycètes contiennent le plus grand nombre d'espèces séquencées au sein d'un phylum avec au total 61 espèces. Les analyses comparatives ont permis de mettre en évidence des mécanismes d'évolution majeurs autre que la duplication totale du génome entre les Hémiascomycètes comme des introgressions, de la diversité des éléments transposables et des hybridations intra ou inter-espèces suivies de la perte de l'hétérozygoté (Dujon 2010; Bleykasten-Grosshans and Neuvéglise 2011).

Les mécanismes évolutifs qui ont lieu chez ces levures sont en partie communs à ceux des animaux (Keeling *et al.* 2005) comme par exemple la présence d'événements de duplication totale du génome. Avec leurs génomes compacts et les nombreuses séquences disponibles dans les banques de données, elles constituent un modèle de génomique comparative et évolutive au sein des eucaryotes

(Dujon 2006). De plus, leurs génomes couvrent une grande distance évolutive. Si l'on se base sur la divergence en acides aminés des protéines orthologues par comparaison 2 à 2 entre les espèces, il y a plus de divergence au sein des Hémiascomycètes qu'au sein des Chordés qui sont pourtant très divers morphologiquement (Dujon *et al.* 2004). Cependant il ne faut pas oublier que les levures se multiplient principalement de façon clonale et que ce mode de reproduction doit avoir un impact sur les processus d'évolution.

La séquence d'un unique génome n'est pas représentative de l'espèce, c'est pourquoi durant cette dernière décennie des études de la structure des génomes au sein de populations de *S. cerevisiae* et *Saccharomyces paradoxus* ont été menées (Johnson *et al.* 2004; Aa *et al.* 2006; Liti *et al.* 2009). Grâce aux nouvelles technologies de séquençage à haut débit notamment, l'analyse de la diversité génétique intraspécifique, encore appelée la génomique des populations, a pu émerger et connaît actuellement un grand essor.

### **3.2. La diversité intraspécifique ou génomique des populations**

Diverses variations génétiques existent entre les génomes de souches d'une même espèce. Elles peuvent se situer aussi bien au niveau nucléotidique (mutation, petite délétion/insertion) que chromosomique (aneuploïdie, réarrangement chromosomique). Ces variations génétiques sont à l'origine de variations phénotypiques mais elles constituent également de précieux indices sur la structure des populations au sein de l'espèce et donc sur son histoire évolutive.

#### **3.2.1. Les variations génétiques base de la diversité entre les souches**

##### **Le polymorphisme nucléotidique :**

Le polymorphisme nucléotidique ou SNP (Single Nucleotide Polymorphism) est une mutation observée au niveau d'un nucléotide. La mutation survient suite à une erreur dans la réplication de l'ADN. Toutes les mutations peuvent potentiellement avoir un impact sur le phénotype. Si le SNP se situe dans la phase codante d'un gène et que le changement de base induit un changement d'acide aminé dans la protéine elle est dite non synonyme. Elle peut alors avoir un impact sur la conformation ou fonction de la protéine. Dans le cas contraire, s'il n'y a pas de changement d'acide aminé, elle est dite synonyme. Une mutation dans la région promotrice peut entraîner une modification de l'expression du gène. Le polymorphisme nucléotidique est de loin la source de diversité génétique la plus fréquente entre les individus d'une même espèce. Le polymorphisme chez *S. cerevisiae* entre la souche de référence et les autres souches varie de 1 à 6 SNP/kb (Schacherer *et al.* 2009; Wei *et al.* 2007), la moyenne de polymorphisme étant de 2,8 SNP/kb (Schacherer *et al.* 2009).

**Les indels :**

Les indels pour **insertions** **deletions** sont un ajout ou une perte de quelques nucléotides. Dans une séquence codante, si elles ne sont pas multiples de trois, elles créeront un décalage de la phase de lecture. La séquence protéique sera alors totalement changée à partir du décalage. La protéine risque donc de changer de conformation, d'être tronquée prématurément et peut par ces changements perdre sa fonctionnalité ou bien avoir une activité modifiée. Les indels peuvent aussi correspondre à des insertions d'éléments transposables ou à des répétitions de type microsatellites. La souche YJM789 isolée à partir d'une infection d'un patient atteint du SIDA contient près de 6 000 indels et 60 000 SNP par rapport à S288c, dont certains vont être responsables de gènes tronqués donc non fonctionnels (Wei *et al.* 2007).

**Les CNV (copy number variations) :**

Il s'agit d'une modification de l'ADN qui conduit à un nombre anormal de copies d'une région du génome. La comparaison génomique par hybridation sur puce à ADN (array comparative genomic hybridation, aCGH) permet de mettre en évidence les amplifications ou délétions de régions. Cette technique a été utilisée sur 16 souches de *S. cerevisiae* d'origines écologiques variées et a permis de mettre en évidence de nombreuses CNV. Des gènes codant des alcools déshydrogénases impliquées dans la tolérance à l'éthanol comme *ADH7* et *AAD3* sont présents en plusieurs copies chez des souches isolées de la vigne. Ce plus grand nombre de copies aura pour conséquence un plus grand nombre de transcrits et une quantité de protéines plus importante entraînant une résistance à l'éthanol plus grande (Carreto *et al.* 2008).

**Les réarrangements chromosomiques :**

Les réarrangements chromosomiques affectent des portions de chromosomes. Ces régions peuvent être délétées, amplifiées, inversées ou encore être transloquées. Ils impliquent souvent des séquences répétées comme des éléments transposables qui permettront la recombinaison entre deux régions chromosomiques. Des réarrangements chromosomiques ont pu être fréquemment sélectionnés chez *S. cerevisiae* (Schacherer *et al.* 2007; Tourrette *et al.* 2007). Les réarrangements chromosomiques sont fréquemment détectés dans les expériences d'évolution en laboratoire car ils permettent l'adaptation rapide de la souche à son nouvel environnement. Par exemple, lors d'une expérience d'évolution dans laquelle une souche de laboratoire a été cultivée en chemostat, dans un milieu limité en nutriment, durant 100 à 500 générations, 8 souches évoluées ont été caractérisées (Dunham *et al.* 2002). Parmi les 8 souches évoluées, 6 contenaient des réarrangements chromosomiques. Un même réarrangement chromosomique est apparu indépendamment 3 fois, ce qui souligne l'avantage sélectif de ce mécanisme d'évolution.

**Les aneuploïdies :**

L'aneuploïdie définit un nombre anormal de chromosomes dans la cellule. La présence de chromosomes surnuméraires peut engendrer un plus grand nombre de transcrits des gènes présents sur ce chromosome. C'est également un mécanisme adaptatif très présent dans les expériences d'évolution mais c'est également un moyen d'acquérir la résistance à un traitement antifongique en une génération chez les levures pathogènes (Dunham *et al.* 2002; Selmecki *et al.* 2009). L'aneuploïdie du fait de son fort coût énergétique pour la cellule est un moyen transitoire de s'adapter au nouvel environnement. Dans une expérience d'évolution à long terme les souches évoluées de *S. cerevisiae* ont remplacé les aneuploïdies par d'autres solutions moins coûteuses pour la cellule (Yona *et al.* 2012).

**Les mécanismes épigénétiques :**

Ces caractères influent sur le patron d'expression des gènes sans modification de la séquence nucléotidique. Ils peuvent être causés par la variation de la structure de la chromatine par le biais de modifications des protéines associées à la chromatine (les histones) ou bien par la méthylation des cytosines présentes dans la séquence de l'ADN ce qui va changer l'expression des gènes au sein de la région chromosomique concernée (Margueron *et al.* 2005). Certains de ces événements sont héréditaires. Ils peuvent être causés par un facteur environnemental, par un autre locus ou même être spontanés. Les mécanismes épigénétiques peuvent intervenir dans la différenciation cellulaire chez les organismes multicellulaires comme par exemple dans la maturation du lymphocyte (Margueron *et al.* 2005).

Les épimutations se définissent comme les différences du patron des méthylations de l'ADN entre individus. Par exemple chez la levure *S. cerevisiae* il a été démontré un fort polymorphisme des positions des modifications traductionnelles des histones entre la souche BY (isogénique à la souche de référence S288c) et la souche de la vigne RM11. Plus précisément rien que pour l'acétylation de la lysine H14 de l'histone H3, 40% de la chromatine présente un polymorphisme entre ces 2 souches (Nagarajan *et al.* 2010). Les épimutations jouent donc un rôle dans la variabilité intraspécifique.

**3.2.2. La diversité génétique au sein de l'espèce *S. cerevisiae***

Ces dix dernières années, la variabilité intraspécifique présente au sein des *S. cerevisiae* a été analysée grâce à des analyses comparatives des génomes. Ces études se sont basées sur des techniques d'hybridation du génome sur puces à ADN et de séquençage partiel ou total du génome. Elles ont eu pour but de déterminer la structure de la population de l'espèce et d'explorer son histoire évolutive.

**L'origine des souches de fermentation utilisées par l'homme**

La première étude sur la variabilité génétique intraspécifique s'est basée sur une collection de 81 souches de *S. cerevisiae* dont 60 souches sont associées aux processus de fermentation industrielle et 19 proviennent d'infection ou d'un environnement naturel (Fay and Benavides 2005). Le séquençage

de 5 loci pour toutes les souches a permis d'estimer la variabilité génétique qui règne au sein de la collection. Les souches isolées du saké et du vin ont une faible diversité génétique par rapport aux autres souches non associées aux productions de fermentations industrielles. D'après la diversité génétique mesurée dans les souches utilisées en fermentation alcoolique, l'estimation du nombre de génération que réalise une levure par année et le taux de mutation de l'espèce, les souches du vin et du saké auraient été employées par l'homme pour les processus de fermentation il y a 11 900 ans. Les souches auraient par la suite suivi un événement de domestication : c'est-à-dire qu'elles auraient acquis des spécificités conférant des avantages pour l'homme dans la fermentation du raisin et du riz.

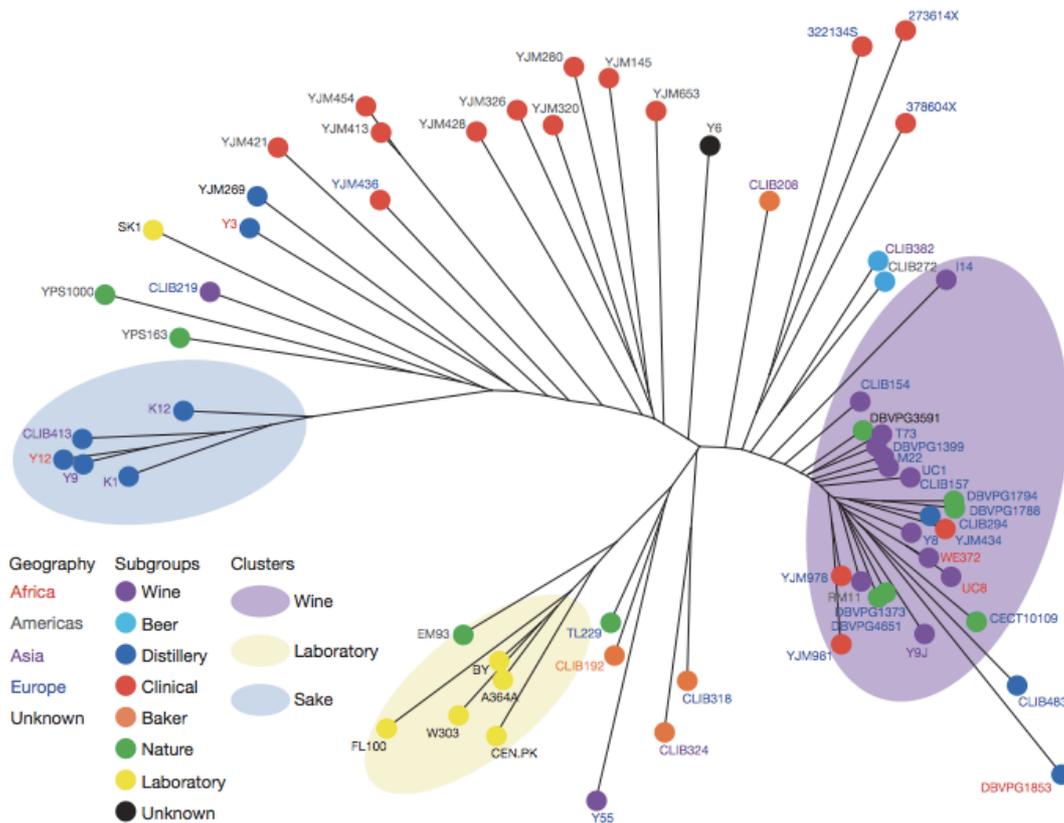
### **La diversité génétique est corrélée à l'origine écologique**

En 2009, deux études ont permis d'établir des cartes de polymorphisme génétique au niveau nucléotidique pour de grandes collections de souches de *S. cerevisiae*. Elles ont ainsi permis d'analyser la structure de la population. Pour ce faire deux méthodes ont été employées : l'utilisation des puces ADN à haute densité et le séquençage du génome à faibles couvertures (Schacherer *et al.* 2009; Liti *et al.* 2009). Des souches d'origines écologiques (environnement naturel, vigne, bière, clinique, boulangerie et laboratoire) et géographiques (Europe, Amérique et Asie) variées ont été utilisées pour les deux études.

Dans la 1<sup>ère</sup> étude, la diversité génétique de 63 souches de *S. cerevisiae* a été analysée par hybridation de l'ADN génomique sur puces à ADN (high-density Affymetrix Yeast Tiling Microarray) (Schacherer *et al.* 2009). Cette technique a permis de détecter 1 299 811 SNP entre les souches de la collection et la souche de référence S288c. Les mutations ont pu être regroupées en 101 343 sites polymorphes. Si les souches de laboratoire, qui sont très proches de la souche de référence, ne sont pas prises en compte, cela fait en moyenne 30 097 SNP soit 2,8 SNP par kb. Cette étude a également permis d'estimer la fréquence et la distribution des variations génétiques mettant en évidence des régions conservées et des régions hautement polymorphes au sein de l'espèce *S. cerevisiae*. Cependant, il est à noter que les cartes génomiques ne permettent pas de connaître la nature des changements nucléotidiques mais seulement de les localiser. Sur la figure 2, la diversité des souches de *S. cerevisiae* est représentée sur un arbre de type Neighbour-Joining construit sur la base des 101 343 sites polymorphes. Les souches provenant de la vigne, du saké et du laboratoire sont regroupées au sein de clusters. Les souches partageant le plus de sites polymorphes sont donc celles provenant d'une même origine écologique. Cette observation est contraire à ce qui est observé chez *S. paradoxus*, l'espèce la plus proche de *S. cerevisiae*, dans laquelle les souches les plus proches proviennent d'une même origine géographique (Liti *et al.* 2009). L'hypothèse avancée est que cette évolution est probablement due à la domestication de *S. cerevisiae*.

La 2<sup>ème</sup> étude a permis le séquençage de 36 souches de *S. cerevisiae* et 35 souches de *S. paradoxus*. Un total de 235 127 SNP et de 14 051 indel ont été détectés à travers les 36 souches. L'analyse phylogénétique basée sur les SNP a révélé un regroupement des souches à la fois selon leur

origines géographiques pour les souches d'Amérique du nord américaines et de Malaisie, mais également écologique avec les souches isolées du saké et les souches issues de la vigne. L'analyse de la structure des populations a permis de mettre en évidence 5 lignées dites pures, les génomes des autres lignées sont dits mosaïques car ils sont constitués de plusieurs de ces lignées pures. La domestication de la levure par l'Homme a pu favoriser les croisements entre différentes souches appartenant à des niches différentes et donc permettre l'apparition de ces génomes mosaïques.



**Figure 2 :** Arbre de type Neighbour-joining basé sur les sites polymorphes détectés dans la collection des 63 souches de *Saccharomyces cerevisiae* (Schacherer *et al.* 2009).

### Projets de séquençage de souches d'origine écologique variée

Parallèlement à ces études globales sur le polymorphisme des souches de *S. cerevisiae* et dans un souci d'une meilleure compréhension de l'espèce, les génomes de plusieurs souches d'origines écologiques variées ont été séquencés à haute couverture (tableau 1). Ces séquençages de génomes ont permis une comparaison plus approfondie des souches entre elles.

Le séquençage de la souche commerciale du vin EC1118 a permis de mettre en évidence l'introgession de 3 régions (Novo *et al.* 2009). Ces régions contiennent 34 gènes impliqués dans les processus de fermentation du vin. Une des introgessions provient d'une espèce proche des *Saccharomyces* tandis que les deux autres proviennent d'espèces plus éloignées. Une de ces régions a été identifiée comme provenant de la levure *Zygosaccharomyces bailii* qui est un contaminant majeur

du vin. Certaines de ces introgressions sont présentes chez d'autres souches de la vigne ce qui indique qu'elles sont apparues récemment dans l'histoire évolutive de l'espèce.

Les souches commerciales utilisées pour la fermentation alcoolique du raisin ou du houblon ont des caractéristiques qui leurs sont propres comme une haute tolérance à l'éthanol ou la capacité à supporter des températures fluctuantes. Dans le but de comprendre ces spécificités phénotypiques, les génomes de 6 souches impliquées dans la fermentation alcoolique ont été séquencés (Borneman *et al.* 2011). En plus d'avoir mis en évidence de nombreux SNP et indel par rapport aux souches de laboratoires, entre 36 et 110 nouvelles ORF (Open Reading Frame) ont été mises en évidence selon les souches. Certaines de ces ORF sont communes et ont un impact sur le phénotype comme par exemple le gène *RTMI*, commun aux souches de brasserie et qui confère une résistance à un inhibiteur chimiquement non caractérisé contenu dans mélasse. Les souches de brasseries possédant *RTMI* ont acquis un avantage sur leur environnement. Certaines des caractéristiques phénotypiques propres aux souches de fermentation ont donc pu être expliquées par ce séquençage.

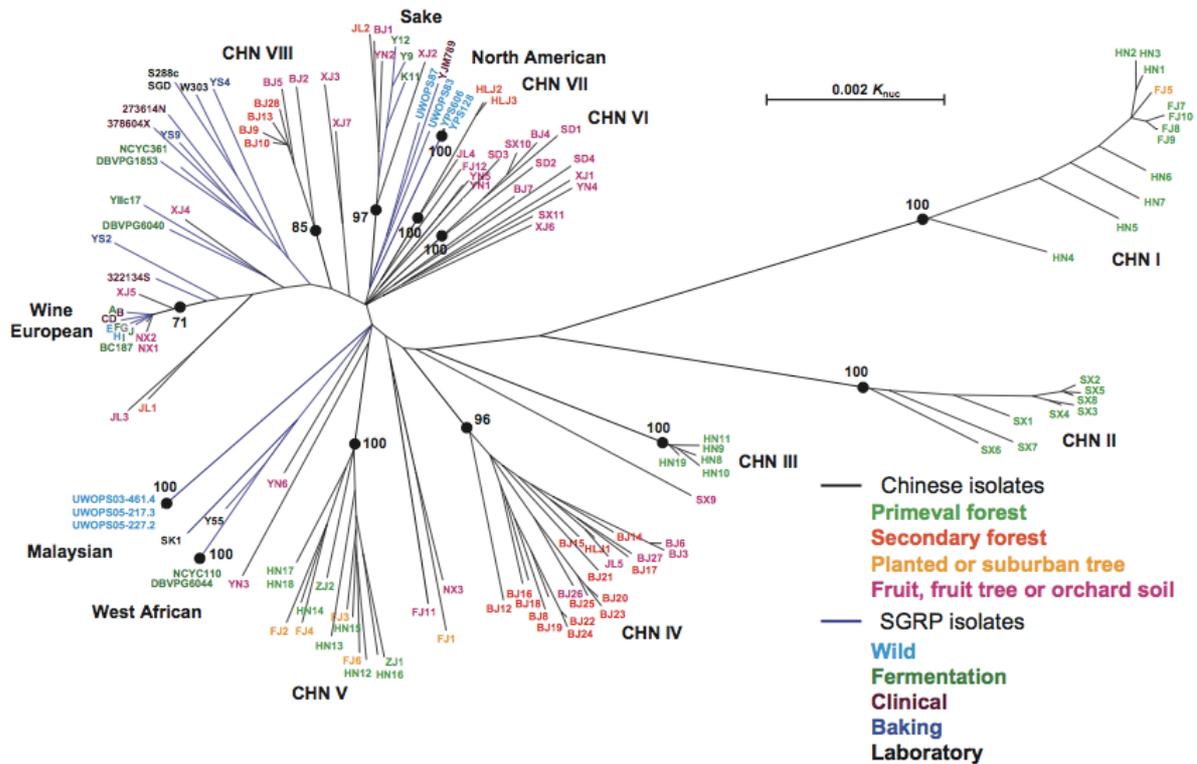
Origine écologique	Souches	Références
Nature	YPS163	(Doniger <i>et al.</i> 2008)
Clinique	YJM789	(Wei <i>et al.</i> 2007)
Vin	M22	(Doniger <i>et al.</i> 2008)
	RM11	( <a href="http://www.broadinstitute.org/">http://www.broadinstitute.org/</a> )
	EC1118	(Novo <i>et al.</i> 2009)
	AWRI1631	(Borneman <i>et al.</i> 2008)
	AWRI1796	(Borneman <i>et al.</i> 2011)
	VL3	(Borneman <i>et al.</i> 2011)
	LanLvin QA23	(Borneman <i>et al.</i> 2011)
Laboratoire	SK1	(Van Overbeek <i>et al.</i> 2012)
	W303	(Ralser <i>et al.</i> 2012)
Bière	FostersO	(Borneman <i>et al.</i> 2011)
	FostersB	(Borneman <i>et al.</i> 2011)
Biotechnologie	JAY291	(Argueso <i>et al.</i> 2009)
	YJS329	(Zheng <i>et al.</i> 2012)

**Tableau 1 :** Exemple de souches de *S. cerevisiae* complètement séquencées.

### La diversité des isolats naturels est supérieure à celle des souches domestiquées

De 2006 à 2009, une campagne de collecte d'isolats naturels provenant en grande partie de forêts primaires (forêts intactes, non exploitées par l'Homme) situées en Chine a eu lieu afin de trouver des souches naturelles n'ayant eu aucun contact avec l'Homme. Suite à cette collecte, 99 souches issues de différentes localisations et substrats ont été sélectionnées et 13 loci ont été séquencés (Wang *et al.* 2012). L'alignement des séquences chez les 99 souches chinoises ainsi que ceux de 38 autres souches déjà caractérisées ont permis d'obtenir un arbre de type Neighbour-joining (figure 3). Les 99 souches chinoises sont réparties en 8 groupes distincts (désignés par CHN I – VIII). La longueur des branches proportionnelle au polymorphisme présent dans les gènes séquencés révèle une très grande diversité

présente dans ces souches naturelles, en particulier pour les groupes CHN I et CHN II. Pour les souches chinoises, leurs regroupements sont corrélés à leurs niches écologiques mais également géographiques desquelles elles ont été isolées. Cette étude nous montre la diversité génétique qui réside au sein des souches *S. cerevisiae* et qui est en particulier très riche pour les souches d'origine naturelle. *S. cerevisiae* a un fort potentiel de modèle écologique non exploité. Cette diversité doit être mise à profit dans les études à venir en utilisant non plus une souche mais une collection de souches.



**Figure 3 :** Arbre de type Neighbour-joining obtenu à partir des séquences concaténées de 9 gènes et de 4 régions intergéniques (longueur totale de 13 500 pb environ) (Wang *et al.* 2012). Les valeurs de bootstrap pour un pourcentage supérieur à 50% sur 1000 répliquats sont indiquées au niveau des nœuds. L'échelle calculée correspond au modèle Kimura à deux paramètres et représente 0,002 substitutions par nucléotide.

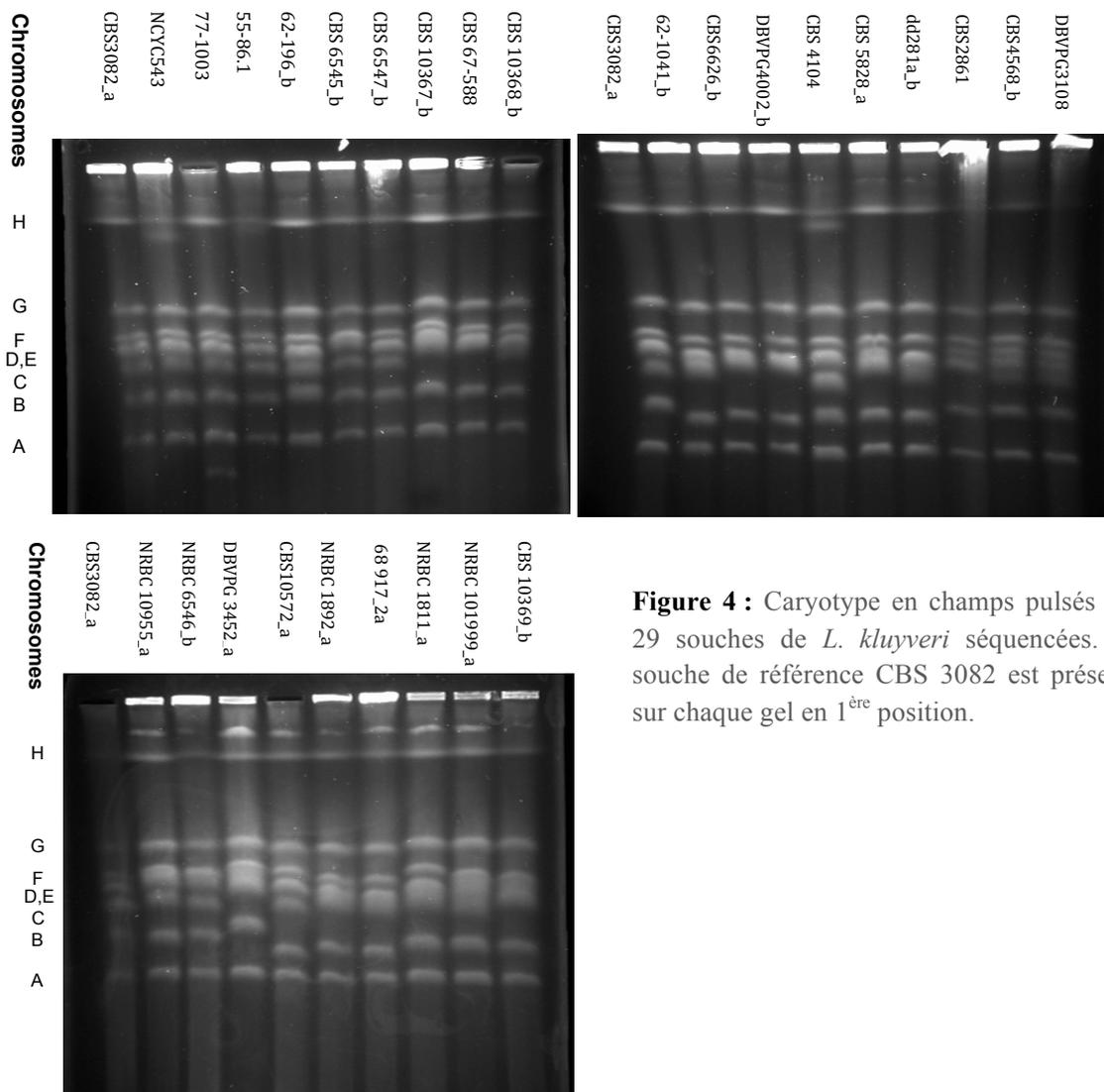
### 3.2.3. La diversité génétique au sein de l'espèce *L. kluyveri*

Afin d'étudier la variabilité intraspécifique de *L. kluyveri*, l'un des premiers objectifs de cette thèse a été de constituer une collection de souches pour cette espèce. Les isolats collectés proviennent de différentes origines géographiques (Amérique, Asie, Europe) mais également de différentes niches écologiques (sol, exsudats d'arbres, feuilles en décomposition, drosophile, ...). Les souches ont été caractérisées comme appartenant à l'espèce grâce au séquençage de la région D1D2 du rDNA. L'un des premiers objectifs de cette analyse était de séquencer plusieurs souches haploïdes, c'est dans ce but que nous avons généré des cellules haploïdes à partir de celles diploïdes qui ont été isolées de la

nature. Parmi les 45 souches qui constituent cette collection, 29 ont été sélectionnées pour un séquençage complet du génome. En plus d'être utilisée dans cette thèse, cette collection sert à d'autres projets tels que l'analyse des cartes de recombinaison, le contenu en éléments transposables, et l'analyse de la variabilité génétique.

### 3.2.3.1. La diversité du profil des caryotypes des isolats de *L. kluyveri*

Parallèlement au séquençage des génomes, j'ai réalisé une analyse des caryotypes par électrophorèse des chromosomes en champ pulsé (PFGE) (figure 4). Certaines souches ont des profils de migration de chromosomes similaires à la souche de référence CBS 3082 mais une majorité des profils sont divergents à celle-ci. L'apparition d'une bande chez la souche 77-1003 ou encore le changement de position de certaines bandes indique que les souches de *L. kluyveri* sont déjà différentes de par leur profil caryotypique. Des remaniements chromosomiques peuvent expliquer la présence de ces différents profils de migration. Cette observation pose la question de la diversité au niveau nucléotidique qui règne au sein de l'espèce. Nous pouvons également nous demander si les souches qui ont des profils caryotypiques proches sont des souches génétiquement proches. Le séquençage du génome de 29 souches de la collection a permis de répondre à ces questions.



**Figure 4 :** Caryotype en champs pulsés des 29 souches de *L. kluyveri* séquencées. La souche de référence CBS 3082 est présente sur chaque gel en 1<sup>ère</sup> position.

Le séquençage des génomes, nucléaires et mitochondriaux, a été effectué par la technique Illumina HiSeq 2000 en multiplexant 10 souches par ligne de séquençage. Les lectures paires obtenues ont une moyenne de 102 pb et ont conduit à une couverture moyenne du génome de 100 X.

### 3.2.3.2. La diversité des génomes mitochondriaux de 18 isolats de *L. kluyveri*

Grâce au projet de réséquençage des souches de *L. kluyveri*, nous avons pu analyser la diversité génétique des génomes mitochondriaux. C'est la première étude de la variation intraspécifique des génomes mitochondriaux menée au sein d'une espèce appartenant aux Hémiascomycètes.

Dans un premier temps, les lectures de la souche NCYC543 (souche très proche génétiquement de la référence CBS3082) ont été assemblées grâce au logiciel *SOAPdenovo* afin de générer la séquence de référence pour l'espèce. Le génome mitochondrial de 51 525 pb fait deux fois la taille de celui de *Lachancea thermotolerans* qui est l'espèce la plus proche mais contient le même groupe de 35 gènes dont 8 gènes codant des protéines et 27 gènes à ARN non codant. Parmi ces 35 gènes, la synténie entre les deux espèces est conservée, mise à part une seule région inversée.

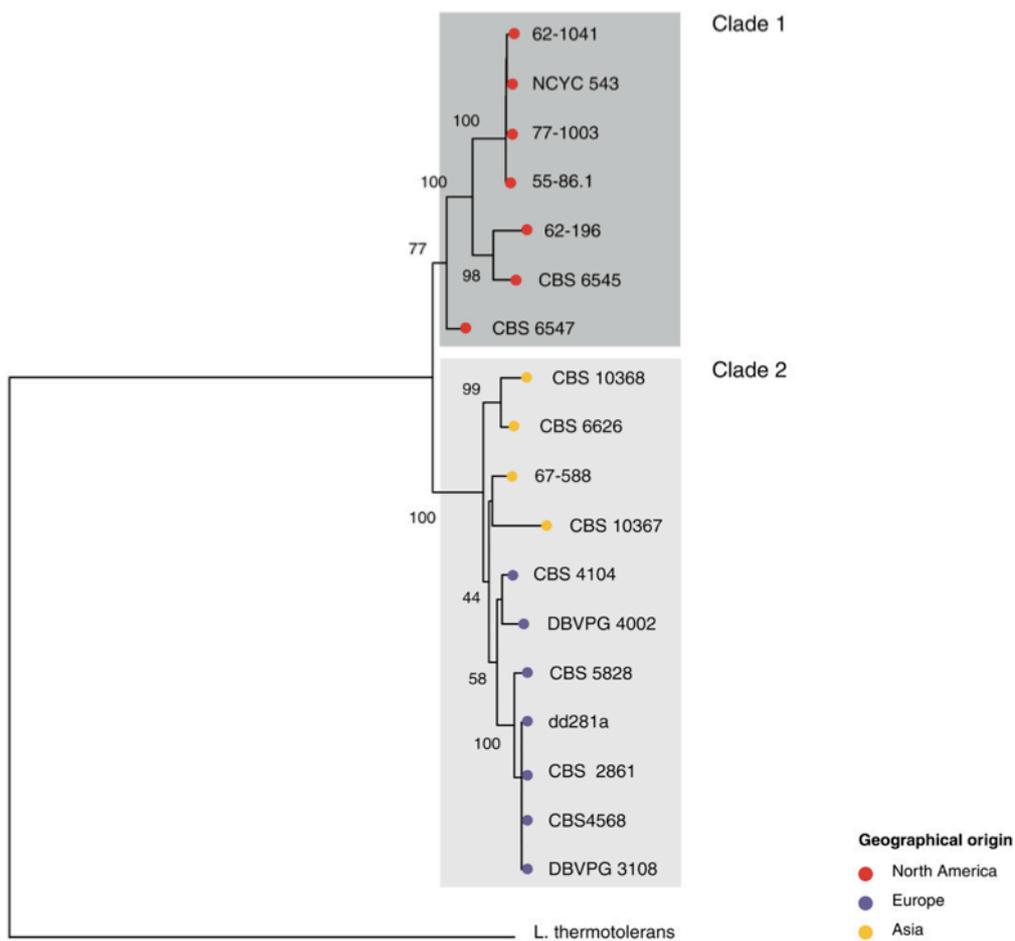
Les génomes mitochondriaux de 18 souches ont ensuite été comparés afin de réaliser une analyse intraspécifique. Dans les régions codantes, le polymorphisme varie de 0 à 13,96 SNP/kb par rapport à la souche de référence. Sur les 1101 SNP détectés au sein des régions codantes, 171 sont non synonymes et 930 sont synonymes ce qui signifie que les régions codantes sont conservées. Pour savoir si le faible nombre de mutations non synonymes par rapport aux mutations synonymes est dû à une pression de sélection négative s'exerçant sur les séquences codantes, le calcul du rapport des mutations non synonymes (dN) sur les mutations synonymes (dS) a été réalisé. La moyenne des rapports par gène a ensuite été calculée. Pour tous les gènes le dN est inférieur au dS ce qui est la marque d'une sélection négative des gènes mitochondriaux. La sélection négative a pour but d'éliminer les substitutions délétères apparaissant au sein des gènes. L'analyse des introns a permis de révéler un fort polymorphisme entre les souches. Le contenu en introns des souches semble être corrélé à l'origine géographique.

Une étude phylogénétique basée sur la concaténation des séquences codantes des gènes *ATP6-ATP8-ATP9-COX1-COX2-COX3-COB* a été réalisée. Les 5475 sites polymorphes détectés lors de l'alignement ont permis de construire un arbre phylogénétique de type Neighbor-joining (figure 5). Deux groupes se distinguent très nettement dans cet arbre, un premier constitué des souches d'Amérique du nord et le second incluant les souches européennes et asiatiques. Les souches européennes sont ensuite regroupées dans un sous-clade. Les souches les plus proches génétiquement seraient contrairement à *S. cerevisiae*, les souches provenant d'une même origine géographique.

Afin d'analyser le patron des sites polymorphes, le génome mitochondrial de référence a été comparé à celui de trois autres souches pour lesquelles les lectures ont été assemblées en une seule séquence par génome. L'ensemble des 35 gènes est bien présent pour les 4 souches et la synténie est conservée. Le polymorphisme est faible dans les régions codantes avec une moyenne de 5,19 SNP/kb

tandis que dans les régions intergéniques la moyenne est beaucoup plus élevée avec 37,2 SNP/kb. Le contenu en indel suit le même profil avec 0,19 indel/kb dans les séquences codantes et 9,8 indel/kb dans les régions intergéniques. Le contenu en introns et la variation des régions intergéniques permet d'expliquer la différence totale de taille des génomes mitochondriaux entre les 4 souches.

En conclusion de cette étude, les séquences des régions codantes des génomes mitochondriaux sont très conservées parmi les souches de *L. kluyveri* alors que les séquences introniques et intergéniques sont bien plus variables (nombreux indel et SNP).



**Figure 5 :** Arbre de type Neighbor-joining de 18 souches de *L. kluyveri* basé sur les séquences codantes des gènes mitochondriaux (Jung *et al.* 2012). Les chiffres au niveau des nœuds sont les valeurs de bootstrap basées sur 1000 répliques.

### 3.2.3.3. La diversité du génome nucléaire de 29 isolats de *L. kluyveri*

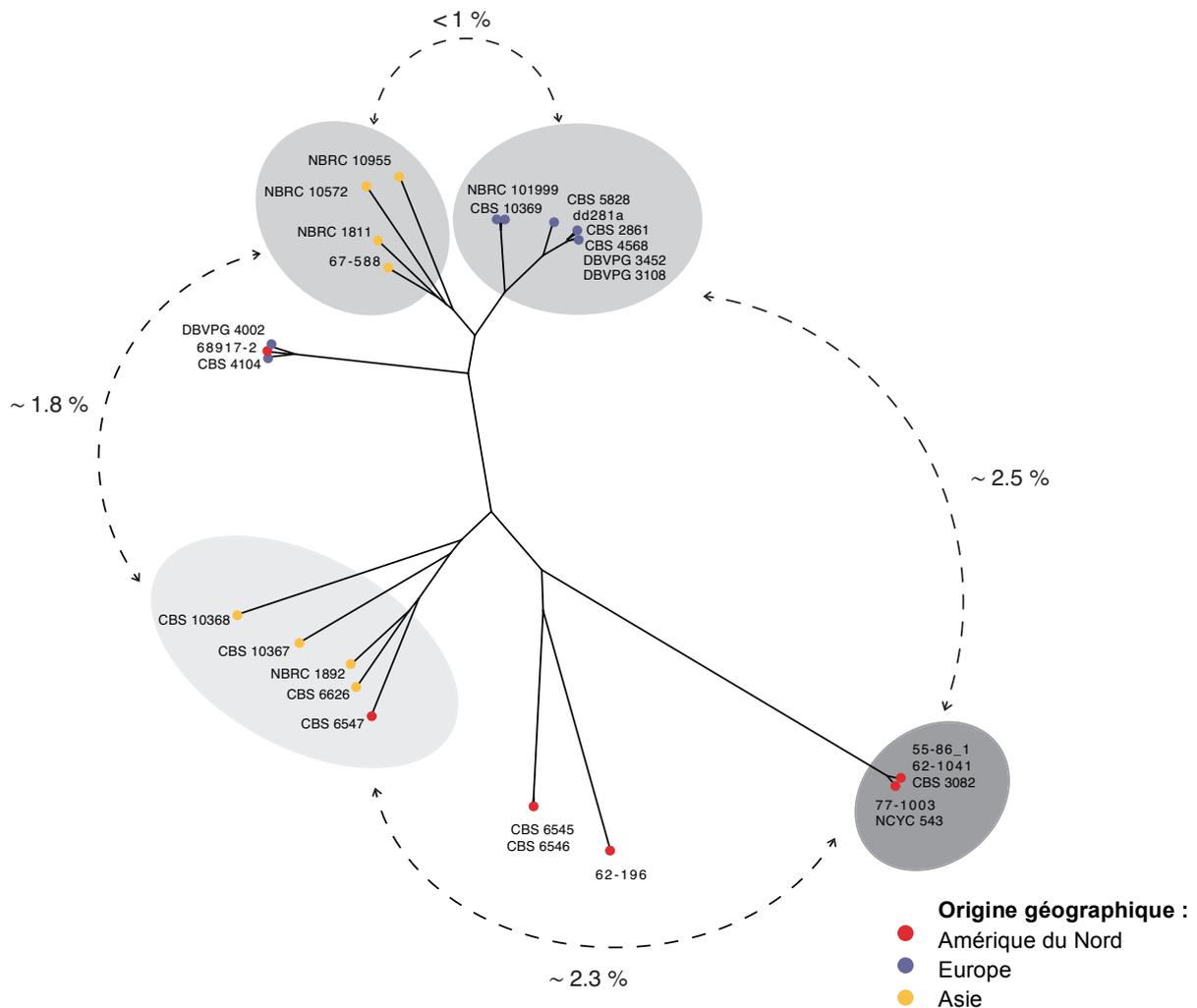
L'analyse des génomes nucléaires a permis la détection d'un polymorphisme élevé avec un total de 6 515 704 SNP répartis sur 881 427 sites polymorphe. Un total de 243 899 indel répartis sur 44 578 sites a été détecté. Le nombre de SNP pour la collection de *L. kluyveri* est supérieur aux 101 343 SNP dénombrés pour les 63 souches de *S. cerevisiae* (Schacherer *et al.* 2009). Cependant le polymorphisme entre souches est variable. Il va de 9822 SNP chez NCYC 543 pour la souche la plus proche de la référence à 309 398 chez la souche NBRC 10955. Un très grand nombre de singletons (344 365), c'est-à-dire de mutations présentes dans le génome d'une seule souche, est détecté, ce qui représente une moyenne de 13 % des sites polymorphes par souche. Ce patron de mutations propre à chaque souche reflète l'histoire évolutive de chacune.

L'analyse basée sur les 881 427 sites polymorphes montre la présence de 4 groupes de souches distincts (figure 6). Le groupe 1 et 4 est formé des souches asiatiques, le groupe 2 contient les souches européennes et enfin le groupe 3 est formé des souches d'Amérique du nord parmi lesquelles on retrouve la souche de référence CBS 3082. Les souches sont donc regroupées par origine géographique ce qui coïncide avec les résultats de l'analyse des génomes mitochondriaux. Au sein de ces sous-populations, la diversité génétique a été mesurée. La divergence maximale au sein d'un groupe est de 1,5% alors qu'entre les groupes cette diversité peut atteindre les 2,5% du génome (figure 6). Le polymorphisme présent au sein des souches de *L. kluyveri* est extrêmement élevé par rapport à ce qui est connu chez les autres espèces d'Hémiascomycètes. En effet, certaines sous-populations de souches de *L. kluyveri* possèdent une divergence nucléotidique égale voir supérieure à celles présentes au sein d'autres espèces (0,5% au sein des souches de *S. cerevisiae* et 1,5% entre celles de *S. paradoxus*).

L'ensemble de ces résultats sera publié dans l'article en préparation (Friedrich *et al.* 2014). L'analyse des génomes des souches de *S. cerevisiae* et *L. kluyveri* est la première étape vers la compréhension du lien phénotype – génotype. La deuxième étape consiste en l'analyse phénotypique des souches.

### 3.3. Étude de la diversité phénotypique au sein d'une espèce

Les phénotypes entre les souches d'une même espèce peuvent varier à bien des niveaux (morphologie de la cellule morphologie des colonies, croissance, ...). Ces variations doivent être mesurées grâce à des méthodes quantitatives et reproductibles. La diversité phénotypique mesurée sera ensuite utilisée afin de cartographier les gènes à l'origine des variations de phénotype observées.



**Figure 6 :** Arbre de type Neighbour-joining basé sur les 881 427 sites polymorphes détectés au sein de la collection des *L. kluyveri*. Les lignes en pointillés désignent le polymorphisme inter et intra groupe (Friedrich et al. 2014).

### 3.3.1. La diversité phénotypique au sein des souches de *S. cerevisiae*

#### 3.3.1.1. Une pléiade de phénotypes d'intérêt

Les souches de l'espèce *S. cerevisiae* sont très diverses phénotypiquement. La diversité phénotypique retrouvée dans une espèce est le reflet de la diversité des stress et environnements rencontrés par les souches. De précédentes études se sont focalisées sur certaines de ces variations phénotypiques comme par exemple l'efficacité de sporulation, la floculation, la morphologie des colonies, et la tolérance à la chaleur.

La sporulation chez les levures est un processus qui génère 4 spores issues de la méiose. Les spores permettent une meilleure résistance à la dessiccation et à un milieu pauvre en nutriment. Elle est une réponse à un environnement changeant devenant défavorable à la croissance de levure. De nombreuses cascades de signalisations sont impliquées dans l'entrée en sporulation et du fait de ces

multiples facteurs, l'efficacité de sporulation est variable au sein des souches de *S. cerevisiae*. Les souches isolées d'exsudats de chênes en Amérique du nord ont une efficacité de sporulation proche de 100% tandis que les souches du vin ne sporulent que faiblement. Ces phénotypes peuvent être causés par l'adaptation des souches à leur environnement. En effet, les souches de vigne vivent en milieu riche en nutriments par rapport à celles se développant dans les exsudats d'arbre qui est un milieu plus pauvre en nutriment pouvant dans certaines conditions être soumis à la sécheresse (Gerke *et al.* 2009). Les souches isolées des exsudats de chênes en Amérique du nord ont également une tolérance plus forte au gel que celles isolées depuis d'autres environnements (Kvitek *et al.* 2008).

Un autre phénotype également très étudié du fait de son intérêt en industrie est la floculation. La majorité des souches utilisées dans les processus de fermentation alcoolique possèdent la capacité de floculer. Ce phénomène permet aux cellules de s'agréger entre elles, ce qui va avoir pour conséquence leur sédimentation. Lors de la fabrication de la bière par exemple, les levures pourront ainsi facilement être séparées du milieu à la fin de la fermentation. Le processus de floculation est complexe : il est assuré par de nombreux gènes *FLO* (pour floculation) mais dépend également du milieu extérieur (nutriments, stress). Lors des fermentations industrielles, le contrôle du milieu permet de maîtriser la floculation. Ce phénomène est cependant souche spécifique car toutes les souches n'auront pas le même phénotype dans des conditions similaires. L'étude de la floculation présente un intérêt industriel évident. Pour quantifier ce phénomène complexe, le test de sédimentation de Helm est le plus couramment utilisé (Helm *et al.* 1953). Le pourcentage de floculation se calcule par la relation :  $1 - (\text{cellules libres} / \text{total des cellules}) \times 100 \%$ . Basé sur ce test, les souches industrielles utilisées pour la fabrication de la bière ont un pourcentage de floculation compris entre 40 et 90 % (Verstrepen *et al.* 2003).

Les souches utilisées en laboratoires ont également des caractéristiques phénotypiques propres. Pour ces souches, la floculation est généralement un phénotype non désiré. Elles ont donc un faible pourcentage de sédimentation compris entre 0 et 15% (Verstrepen *et al.* 2003). Les souches de laboratoires sont également sélectionnées pour leur propriété à pousser rapidement sur milieux complets tandis que les souches utilisées dans la fermentation alcoolique supportent un environnement de pH bas, une faible teneur en nutriments, de fortes concentrations en éthanol et des températures fluctuantes (Mortimer and Johnston 1986).

Les souches isolées d'infections chez l'Homme possèdent quant à elles des phénotypes communs avec d'autres levures pathogènes telles que *Candida albicans* et *Cryptococcus neoformans*. Elles possèdent notamment la capacité à pousser à haute température (38-42°C) ce qui correspond à la température corporelle d'une personne infectée et soumise à une fièvre. Leur capacité à croître sous la forme pseudohyphale leur permet d'être plus invasives que les cellules sous la forme dite levure (de Llanos *et al.* 2006).

Ces études ne sont que quelques exemples parmi d'autres menées sur la diversité phénotypique de *S. cerevisiae*. Ce sont autant de phénotypes variés qui caractérisent les souches dont certains sont

partagés par des souches provenant d'un même environnement écologique. Ces études se basent pour la plupart sur l'utilisation d'une dizaine de souches. Ces dernières années, des stratégies de phénotypage à haut débit ont été mises au point afin de pouvoir comparer un plus grand nombre d'isolats entre eux et ce de manière hautement reproductible en utilisant des procédures partiellement automatisées suivies d'une analyse informatisée des résultats.

### 3.3.1.2. Les analyses à haut débit

Des analyses à haut débit ont été utilisées afin de mesurer la variabilité phénotypique de plus larges collections de souches. De nombreux caractères ont été étudiés, mais dans le cadre de ce travail, nous allons nous intéresser à deux phénotypes en particulier : la croissance et la morphologie des cellules.

#### Analyses de la croissance

La croissance de la levure *S. cerevisiae* dépend de plusieurs caractères quantitatifs à savoir : la phase de latence (temps d'adaptation de la souche à un nouveau milieu), la biomasse totale formée et la vitesse spécifique de croissance. Certains stress peuvent agir sur l'un ou plusieurs de ces caractères. La croissance peut être suivie en milieu solide et en milieu liquide.

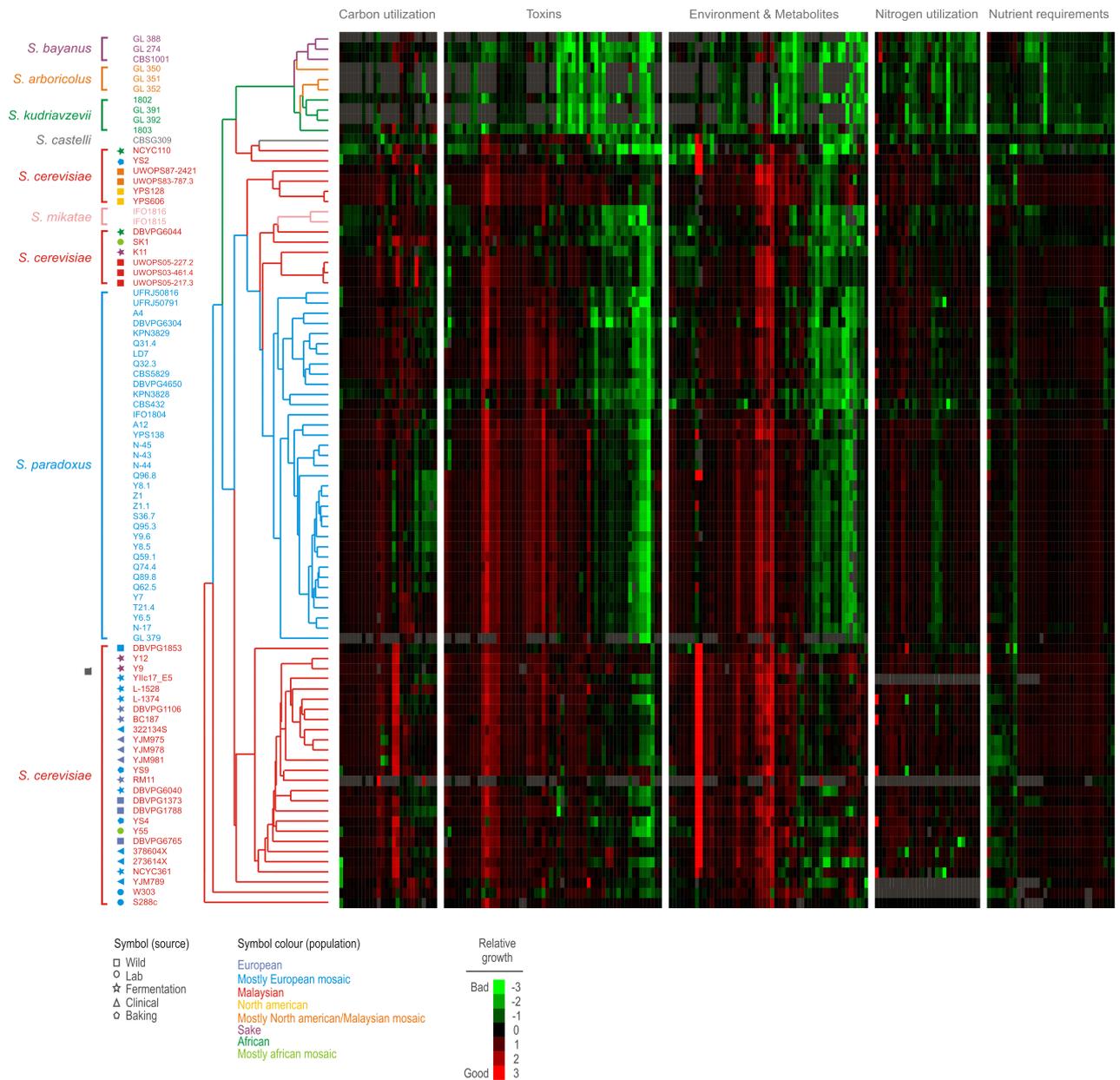
En 2008, une étude a eu pour but d'explorer la diversité phénotypique de 52 isolats de *S. cerevisiae* soumis à des stress environnementaux (Kvitek *et al.* 2008). Au total, 14 conditions (variation de pH, de températures, ...) ont été testées en suivant la croissance des souches en milieu solide. Brièvement, pour chaque souche trois dilutions ont été déposées sur le milieu classique YPD et sur chacune des 14 conditions environnementales. Après incubation, la croissance des souches a été comparée à la souche de référence S288c ce qui a permis d'attribuer une valeur de résistance à chaque condition.

L'analyse des données révèle une forte variation phénotypique entre les isolats. Les souches d'une même niche écologique ont des profils de résistance non identiques mais similaires ce qui suggère des caractéristiques communes et propres à chaque niche. Par exemple, les souches de la vigne sont résistantes au sulfate de cuivre. Ce composé est utilisé comme traitement antifongique des vignes. Sa présence au sein de cette niche écologique va générer une pression de sélection qui peut expliquer la résistance accrue des souches de la vigne à ce composé.

L'étude a également montré que la souche de laboratoire S288c a un profil phénotypique clairement distinct des autres souches. D'ailleurs les études actuelles utilisent également des souches autres que S288c car elle n'est pas toujours représentative de l'espèce.

Dans une étude réalisée en 2011, la diversité phénotypique a été étudiée dans une collection de 39 souches de *S. cerevisiae*, de 39 souches de *S. paradoxus*, et de souches appartenant aux espèces du

genre *Saccharomyces* telles que *S. kudriavzevii*, *S. mikatae*, *S. bayanus* et *S. arboricolus* ainsi que *S. castellii* (Warringer *et al.* 2011). Leur résistance à 200 conditions environnementales différentes a été testée par suivie de microcultures (figure 7).



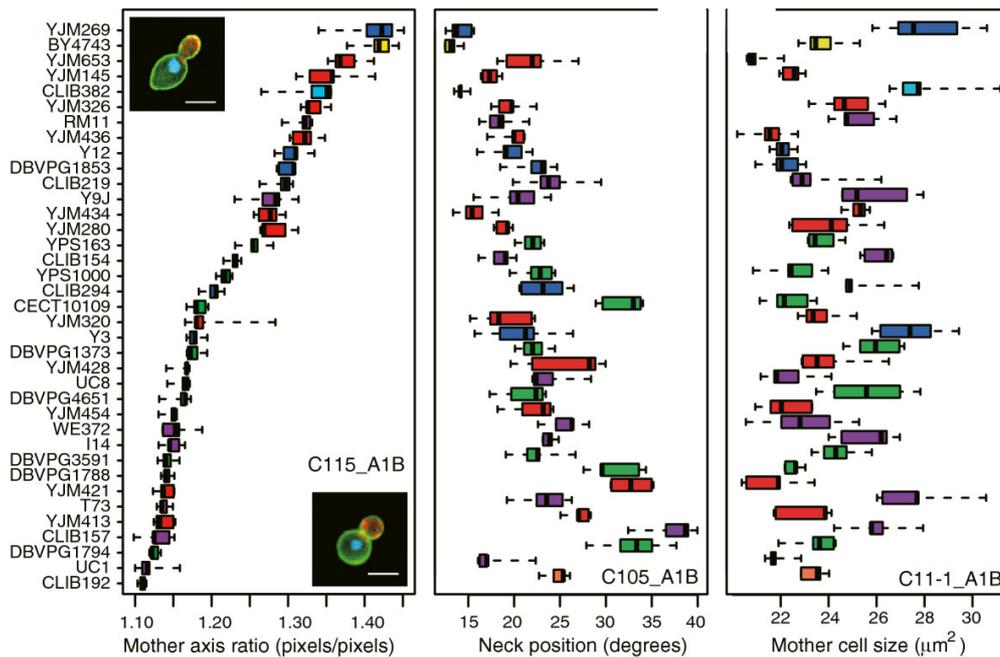
**Figure 7 :** Variations phénotypiques au sein du genre des *Saccharomyces sensu stricto*. Un regroupement hiérarchique des isolats basé sur 200 conditions a été réalisé en utilisant la corrélation de Pearson. Chaque espèce est représentée par une couleur des branches. Chez *S. cerevisiae*, l'origine écologique des souches est indiquée par la présence d'un symbole et l'origine géographique par la couleur du symbole. La carte d'intensité représente le taux de prolifération relatif ( $\text{Log}_2$  [S288c/souche]). La couleur verte indique une croissance inférieure à S288c, en rouge une croissance supérieure, en noir une croissance égale et le gris symbolise des valeurs manquantes.

Les souches de *S. cerevisiae* se distinguent des autres par une meilleure croissance dans de nombreuses conditions parmi lesquelles l'utilisation du maltose, la résistance à une haute température, la capacité à résister à des toxines et la tolérance à l'éthanol. Les souches aux profils phénotypiques les plus proches ont été regroupées grâce au calcul de la corrélation de Pearson (figure 7). Il est intéressant de noter que les souches de *S. cerevisiae* ne sont pas regroupées au sein d'un seul cluster contrairement à celles des autres espèces ce qui reflète une variabilité phénotypique plus importante au sein de cette espèce. Cette diversité de résistance pourrait être due à la pression de sélection qu'exercent les diverses niches écologiques sur les souches de *S. cerevisiae*. Les souches de laboratoires ont un profil phénotypique très différent comparé aux autres souches de l'espèce ce qui suggère que son utilisation par l'Homme a eu un impact sur son phénotype. Comme le suggérait l'étude précédente avec la souche S288c, les souches de laboratoire en général ne sont donc pas les plus représentatives de l'espèce.

### **Mesure des caractères morphologiques**

La morphologie des levures reflète de nombreux processus biologiques comme par exemple la progression du cycle cellulaire, la polarité de la cellule et le contrôle de sa taille. Elle dépend de centaines de caractères quantitatifs dont notamment la longueur de la cellule mère, la longueur du bourgeon, l'angle que forme l'axe du bourgeon par rapport à celui de la cellule mère. Une plateforme d'imagerie à haut débit créée en 2005 permet l'acquisition systématique et quantitative de 501 caractères morphologiques chez la levure à partir d'images de microscopie (Ohya *et al.* 2005). Grâce à des marquages fluorescents de l'ADN, de la paroi cellulaire et du cytosquelette d'actine, les 501 paramètres mesurés englobent la forme de la cellule, le cytosquelette d'actine et la morphologie nucléaire.

Cette technique à haut débit a été utilisée afin d'analyser la variabilité morphologique d'une collection de 37 souches de *S. cerevisiae* d'origines écologiques et géographiques variées (Yvert *et al.* 2013). La plupart de ces caractères sont variables entre souches. Dans la figure 8, trois exemples de variabilité intraspécifique sont illustrés. Le caractère possédant la plus forte variabilité est l'allongement de la cellule. La souche de distillerie YJM269 a une forme totalement allongée tandis que la souche CLIB192 est ronde. L'angle d'apparition du bourgeon par rapport à l'axe de la cellule mère ainsi que la longueur de la cellule mère montrent également une forte variabilité. Si les souches sont regroupées par profils morphologiques similaires, les souches les plus proches d'un point de vue phénotypique ne sont pas celles provenant d'un même environnement écologique ou géographique contrairement à ce qui avait été observé pour la résistance à différents stress (Kvitek *et al.* 2008).

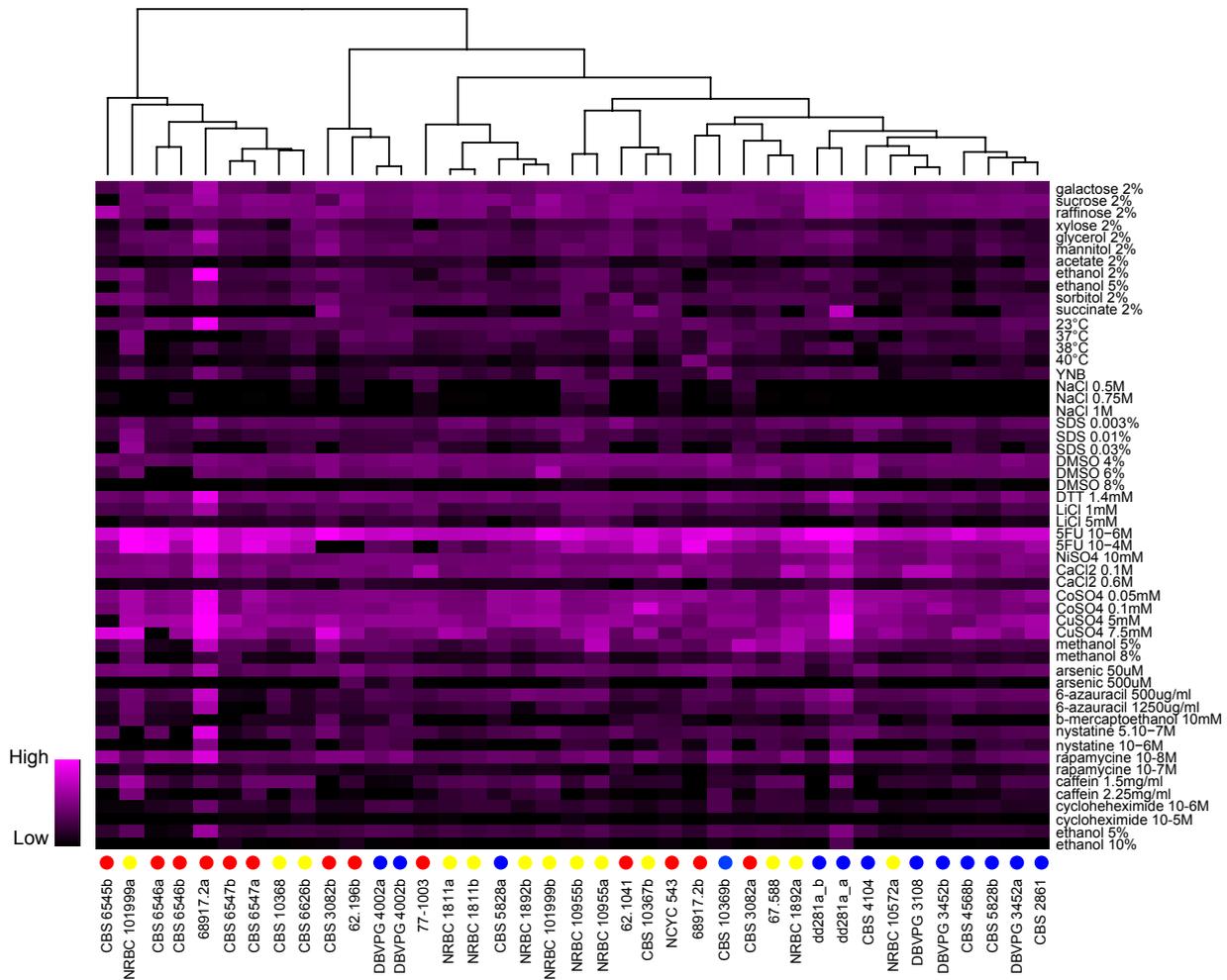


**Figure 8 :** Variabilité intraspécifique de trois caractères morphologiques au sein de 37 souches de *S. cerevisiae* (Yvert *et al.* 2013).

### 3.3.2. La diversité phénotypique au sein des souches de *L. kluyveri*

Une étude menée au sein de notre laboratoire a eu pour but d'analyser la diversité phénotypique au sein d'une collection de 40 souches de *L. kluyveri* cultivées dans différentes conditions. Pour ce faire, la croissance des souches a été suivie en microcultures. La vitesse spécifique de croissance a été déterminée pour chaque culture. Le rapport des vitesses spécifiques de croissance pour chaque souche entre le milieu complet et les différentes conditions a permis de comparer les souches entre elles. Ainsi un rapport de 1 signifie que la souche pousse aussi bien en milieu complet que dans la condition testée. À l'inverse un rapport proche de 0 signifie que la souche pousse très lentement, elle est donc sensible à la condition testée.

Au total, 52 conditions de culture réparties en trois catégories ont été testées : les sources de carbone (telles que le galactose et le sorbitol), les conditions physico-chimiques environnementales (températures et milieu minimum) et les composés toxiques (métaux, antifongiques, détergents, ...) (figure 9). Les résultats mettent en évidence une diversité phénotypique importante au sein de l'espèce *L. kluyveri*. Les profils phénotypiques des souches dans leur globalité sont différents et propres à chaque souche testée.



**Figure 9 :** Regroupement hiérarchique des profils phénotypiques de 40 souches de *L. kluyveri*. Le code couleur en dessous des noms de souches indique l'origine géographique des souches : le rouge symbolise les souches d'Amérique du nord, le bleu l'Europe et le jaune l'Asie. Les vitesses spécifiques de croissance dans chaque condition ont été normalisées avec le milieu standard YPG. Le rapport des deux valeurs est symbolisé par un dégradé violet – noir ; le violet symbolisant une forte croissance et le noir une croissance fortement ralentie.

La hiérarchisation des souches par la corrélation de Pearson en fonction de leur profil phénotypique a mis en évidence le fait que les souches provenant d'un même environnement écologique ou géographique ne sont pas plus proches sur le plan phénotypique qu'avec les autres souches. L'ensemble de ces travaux est en cours de publication (Jung *et al.* 2014).

Les analyses phénotypiques à haut débit ont permis de caractériser la diversité qui règne au sein des espèces *S. cerevisiae* et *L. kluyveri*. Elles constituent une première étape importante dans la compréhension de la relation entre le génotype et le phénotype. À ce stade de l'étude, les origines génétiques de la variation des phénotypes observée ne sont pas encore connues et méritent d'être explorées.

## 4. L'étude de la relation génotype – phénotype

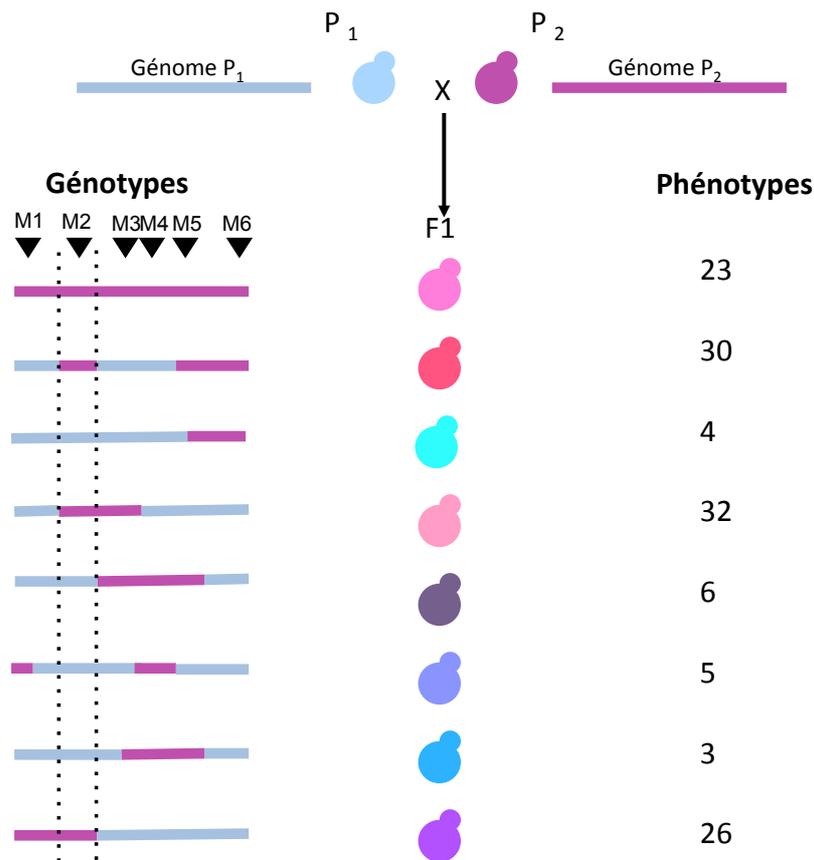
Les variations phénotypiques présentes au sein d'une espèce sont en partie causées par les variations génétiques entre les génomes. Déterminer les origines génétiques causant ces variations est primordial afin de trouver les mécanismes généraux sous-jacents à cette variation naturelle. Grâce à la capacité de réaliser facilement des croisements, à obtenir une descendance importante (des centaines de ségrégants) en peu de temps et à son fort taux de recombinaison durant la méiose (90 crossing-over par méiose en moyenne), *S. cerevisiae* est un organisme de choix pour réaliser une analyse de cartographie des QTL (Mancera *et al.* 2008). Un QTL (Quantitative Trait Locus) est une région génomique plus ou moins grande qui est associée à un caractère quantitatif. La région peut contenir un ou plusieurs gènes impliqué(s) dans le phénotype.

Identifier les gènes responsables d'un phénotype peut se révéler être une tâche difficile du fait des différents QTL qui agissent sur un même phénotype, des phénomènes d'épistasie et de l'impact de l'environnement sur l'expression des gènes. Les QTL ayant un effet majeur sur le phénotype sont appelés QTL majeurs tandis que ceux avec un faible impact sur le phénotype sont appelés QTL mineurs. Deux principales méthodes de cartographie des QTL existent : l'analyse de liaison et l'analyse d'association (Mackay *et al.* 2009).

### 4.1. Les analyses de liaison

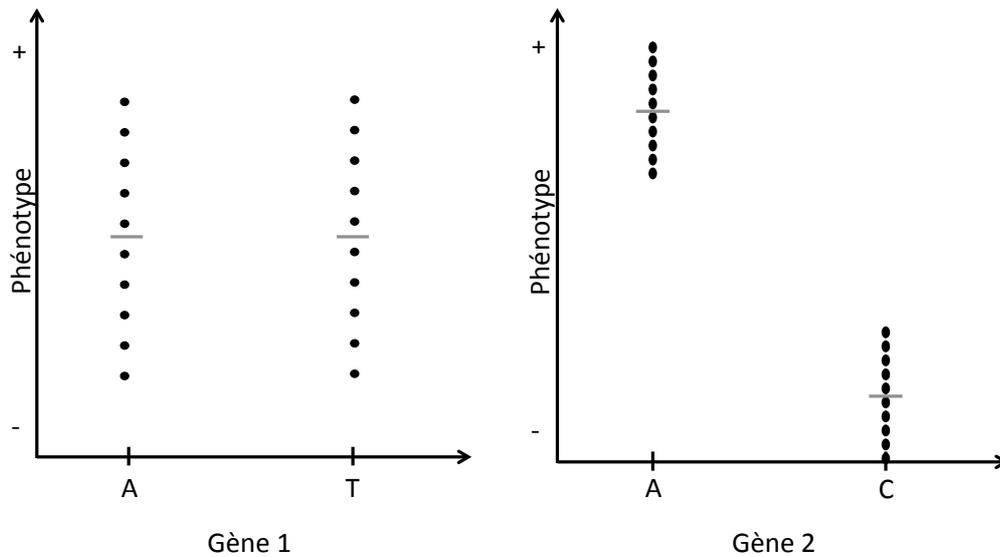
#### 4.1.1. Le principe de la méthode

L'analyse de liaison est une méthode qui repose sur l'étude de la descendance issue d'un croisement entre deux souches dites parentales de signe opposé (*MATa* et *MATα*) (figure 10). À partir de ce croisement, une population de descendants est générée. Grâce aux événements de recombinaison qui ont lieu durant la méiose, les génomes des descendants contiennent différentes combinaisons des génomes parentaux. Le génotype des souches parentales doit être déterminé afin de définir des marqueurs (SNP et indels) qui permettront de différencier les deux génomes parentaux dans la descendance. Le génotype peut être analysé par hybridation sur puces à ADN ou par séquençage. Si le phénotype est complexe, la descendance issue du croisement présentera un trait continu de variations phénotypiques. Le génotype de chaque descendant est alors déterminé, c'est-à-dire que la version allélique au niveau des sites polymorphes dispersés sur l'ensemble du génome est caractérisée. En parallèle, la valeur quantitative du phénotype d'intérêt de chaque descendant est également mesurée. La comparaison des génotypes des descendants au phénotype d'intérêt va permettre de mettre en évidence des régions du génome ne suivant pas une ségrégation aléatoire (c'est à dire que 50 % des allèles proviennent du parent 1 et 50% du parent 2). Les régions provenant majoritairement d'un parent sont impliquées dans le phénotype d'intérêt.



**Figure 10 :** Schéma d'une analyse de liaison chez *S. cerevisiae*. Deux parents P<sub>1</sub> et P<sub>2</sub> sont croisés afin de générer une descendance. Les génotypes des parents et des descendants sont établis. Les marqueurs M1 à M6 vont permettre de distinguer les deux génomes parentaux au sein de la descendance de manière à savoir quelle région du génome provient de quel parent. En parallèle, une mesure quantitative pour le phénotype des descendants est déterminée. Les lignes en pointillées délimitent la région impliquée dans le phénotype d'intérêt.

Pour savoir si une région est impliquée dans le phénotype d'intérêt, la corrélation entre chaque marqueur et le phénotype d'intérêt est analysée par un test statistique. Pour illustrer cela, prenons l'exemple de l'analyse des marqueurs des gènes 1 et 2 (figure 11). Le gène 1 contient le marqueur nucléotidique A chez le parent P<sub>1</sub> et T chez le parent P<sub>2</sub>. Aucune différence phénotypique entre les descendants possédant le marqueur A ou T n'est visible dans la descendance. Le gène 2 contient le marqueur nucléotidique A chez le parent P<sub>1</sub> et C chez le parent P<sub>2</sub>. La moyenne des valeurs des phénotypes des descendants possédant le marqueur A est significativement plus élevée que celle possédant le marqueur C. Le gène 2 est donc lié au phénotype d'intérêt.



**Figure 11 :** Corrélation entre les marqueurs génétiques contenus dans les gènes 1 et 2 et le phénotype d'intérêt. Le gène 1 n'est pas lié au phénotype d'intérêt car le phénotype des descendants n'est pas corrélé à la version allélique des parents. Le gène 2 est lié au phénotype d'intérêt car il y a une différence significative de la moyenne des valeurs du phénotype entre descendants possédant le marqueur A et C.

#### 4.1.2. Quelques exemples d'analyse de liaison chez *S. cerevisiae*

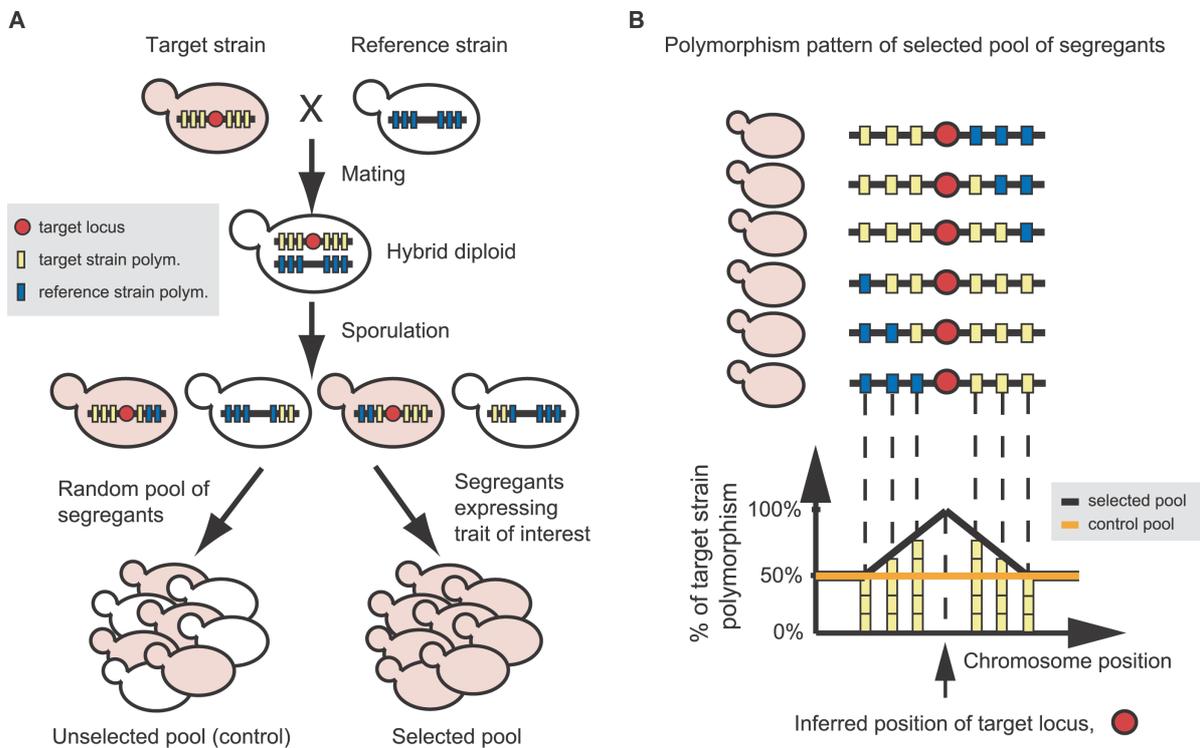
Cette méthode a permis de déterminer les origines génétiques de plusieurs phénotypes complexes comme la sporulation entre les souches de laboratoire S288c et SK1 (Deutschbauer and Davis 2005) et les variations du protéome entre les souches S228c et RM11 (vigne) (Foss *et al.* 2007). Des analyses de liaison ont été réalisées pour des phénotypes d'intérêt industriel tels que la tolérance à l'éthanol (Hu *et al.* 2007), la production élevée d'éthanol et d'arômes dans la bière (Katou *et al.* 2009) ou encore la production d'arômes de vin (Steyer *et al.* 2012). Des phénotypes d'intérêt médical tels la croissance à haute température (41°C) (Steinmetz *et al.* 2002), et la résistance à des médicaments dont des antifongiques et antidépresseurs ont également fait l'objet d'analyses de liaison (Perlstein *et al.* 2007). Cette dernière étude basée sur l'analyse de 104 ségrégants a permis d'identifier un total de 124 QTL en réponse à 83 composés testés.

L'analyse de liaison classique nécessite de connaître le génotype de tous les descendants. Elle est donc coûteuse et génère de nombreuses données de séquençage à analyser. Des variantes de la méthode ont donc été mises au point dans le but de l'améliorer.

#### 4.1.3. Les analyses de pool de ségrégants

L'analyse de pool de ségrégants est une variante de l'analyse de liaison (Segrè *et al.* 2006). Plutôt que de caractériser le génotype de chaque descendant un à un, elle repose sur la caractérisation en une seule fois du génotype de tous les ségrégants ayant le phénotype d'intérêt. En plus de baisser le

coût de l'analyse, cette technique permet également d'utiliser un plus grand nombre de ségrégants ce qui améliorera la précision de détection des QTL.



**Figure 12 :** Schéma de l'analyse de pool de ségrégants basée sur un trait simple (Segrè *et al.* 2006). (A) Une souche parentale exprimant le phénotype d'intérêt est croisée avec une souche ne le possédant pas. Le polymorphisme définissant les marqueurs entre les souches est représenté par les traits jaunes et bleus. L'hybride obtenu est mis à sporuler afin de générer une descendance. Un pool contrôle et un pool de souches exprimant le phénotype d'intérêt sont constitués. (B) L'analyse des fréquences alléliques provenant du parent 1 et 2 pour chaque site polymorphe va permettre de cartographier les régions génomiques liées au phénotype. Sur le graphique le groupe contrôle permet de vérifier que la fréquence allélique est de 50% pour chaque parent tout le long du génome. Dans le groupe d'intérêt, la déviation de la fréquence de ces 50% permet de localiser la région liée au phénotype.

#### 4.1.3.1. Le principe de la méthode

La technique d'analyse de pool de ségrégants permet la cartographie des gènes impliqués dans un phénotype à partir d'un groupe de descendants.

La méthode débute de la même manière que l'analyse de liaison classique, deux souches parentales haploïdes de signe opposé sont croisées ensemble afin de former un diploïde hybride qui sera placé sur milieu carencé pour induire la méiose et obtenir des descendants haploïdes (figure 12). La mesure quantitative du phénotype d'intérêt est déterminée pour chacun des descendants. Un ou deux groupes (ou pool) de ségrégants peuvent être formés grâce aux données phénotypiques. Le premier groupe réunit les descendants possédant le phénotype d'intérêt tandis que le deuxième réunit des souches sélectionnées au hasard dans le but de constituer le groupe contrôle. Nous pouvons nous

attendre à ce qu'au sein des génomes descendants 50% des allèles proviennent du parent *MATa* et les 50% restant du parent *MATα*.

L'ADN total des deux groupes est génotypé. Grâce aux marqueurs (SNP) dispersés dans tout le génome, les fréquences alléliques correspondant au parent *MATa* et au parent *MATα* au sein de l'ADN des groupes peuvent être estimées. Dans le groupe de souches au phénotype d'intérêt, les régions du génome impliquées dans le phénotype montreront un déséquilibre de la fréquence allélique. En d'autres termes, les régions pour lesquelles le génome d'un parent est sur-représenté contiennent un ou des QTL. Par contre, pour un marqueur non concerné par le phénotype d'intérêt, la fréquence de chaque allèle parental dans la descendance sera de 50%. Dans le groupe contrôle la fréquence allélique doit donc être proche de ces 50% tout le long du génome.

#### **4.1.3.2. Quelques exemples d'analyse de pool de ségréants chez *S. cerevisiae***

La première analyse de pool de ségréants chez *S. cerevisiae* a été réalisée en 2006 (Brauer *et al.* 2006). Afin de tester la méthode, trois études basées sur des phénotypes de complexité différente ont été menées dans ce travail. Le 1<sup>er</sup> phénotype dont l'origine génétique a dû être cartographiée est l'auxotrophie à l'arginine, le 2<sup>ème</sup> est un défaut de la croissance sur acétate, et le 3<sup>ème</sup> est la floculation. La technique d'analyse de pool de ségréants a permis de déterminer les bases génétiques de ces trois phénotypes, et a ainsi permis l'amélioration de la technique d'analyse de liaison chez la levure.

Une étude de 2010 a pu identifier l'origine génétique d'un phénotype d'intérêt industriel : la dégradation de la lignocellulose. Le xylose est un sucre constituant du bois. Très peu de souches d'origine naturelle de *S. cerevisiae* peuvent le fermenter, mais certaines ont été modifiées génétiquement afin d'utiliser la biomasse lignocellulosique pour produire de l'éthanol. Une analyse de pool de ségréant a mis en évidence le gène *XDHI* permettant d'utiliser le xylose comme seule source de carbone (Wenger *et al.* 2010). Ce gène n'est pas présent chez les souches de laboratoire mais il est retrouvé dans de nombreuses souches isolées de la vigne.

Une étude de 2011 a eu pour but de trouver les origines génétiques de la résistance à un stress de type oxydatif chez une souche isolée d'un patient infecté à *S. cerevisiae* (Diezmann and Dietrich 2011). Le stress oxydatif est un mécanisme de défense employé par le système immunitaire humain c'est pourquoi il est important de comprendre quels sont les déterminants génétiques entraînant une résistance chez des pathogènes potentiels. Grâce à cette étude il a été prouvé que l'allèle codant le facteur de transcription Rds2 contribuait à la réponse au stress oxydatif. Il confère donc un avantage sélectif pour les souches infectant un patient.

L'analyse de liaison a d'ores et déjà permis d'identifier les origines génétiques de phénotypes d'intérêt chez *S. cerevisiae*. Toutefois, elle possède le désavantage de ne se baser que sur le fond

génétiq ue de deux souches parentales. Elle n'exploite donc pas toute la diversité génétique présente au sein d'une collection de souches. L'analyse d'association est une autre méthode de cartographie des QTL qui se base sur l'analyse d'une population de souches.

#### **4.2. Les analyses d'association**

L'analyse d'association se base sur une population d'individus (de souches) contrairement à l'analyse de liaison qui se base sur une population de descendants (figure 13) (Mackay *et al.* 2009). Le phénotype de chaque souche est quantifié. Parallèlement à cela, le génotype pour chacun des isolats est déterminé. Les génomes des souches sont issus de différents événements de recombinaison entre les génomes ancestraux. Comme précédemment expliqué (figure 11), la corrélation entre chaque marqueur et la moyenne des phénotypes des souches le possédant est analysée. L'ensemble de ces données de corrélation va permettre de délimiter les régions du génome impliquées dans le phénotype d'intérêt.

L'avantage de l'analyse d'association réside dans le fait qu'elle ne se base pas uniquement sur le fond génétique de deux souches parentales. Cependant, jusqu'à présent très peu d'études de ce type ont été menées chez la levure notamment du fait du manque d'une représentativité statistique due à une population de souches trop réduite (Diao and Chen 2012).

#### **4.3. Du pic au gène, à la mutation**

Les régions génomiques mises en évidence par les méthodes de cartographie des QTL (analyse de liaison et analyse d'association) peuvent contenir une centaine de gènes. Plus on se situe au centre du pic, plus la probabilité que le gène à l'origine de la variation phénotypique s'y trouve est élevée. Une analyse approfondie est donc nécessaire afin d'identifier le ou les gènes candidats au sein de chaque région.

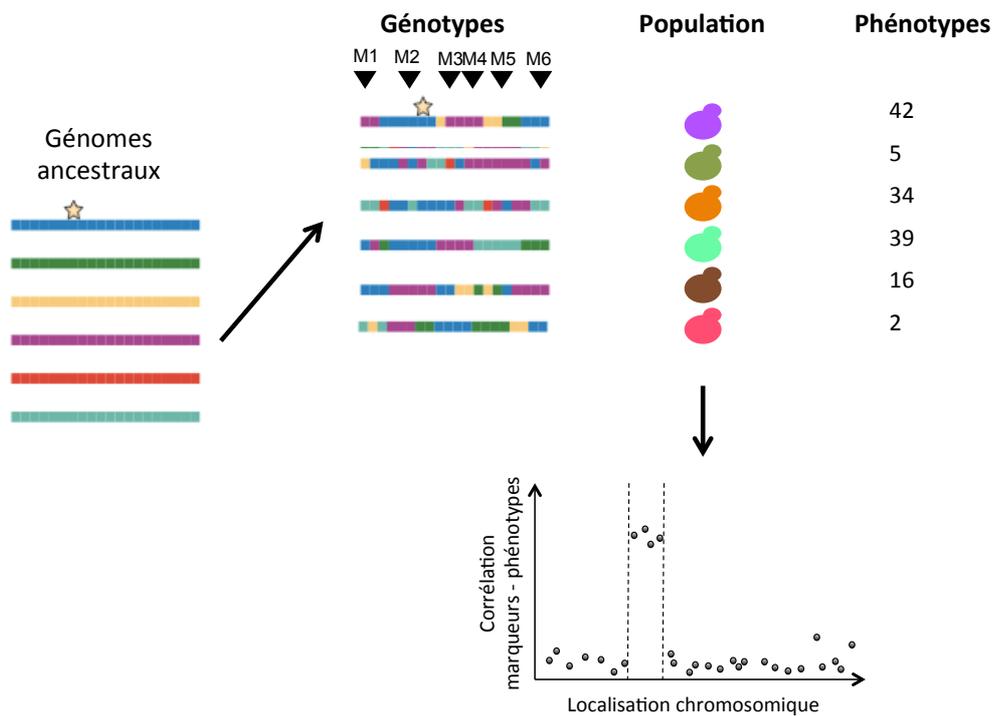
La fonction et les séquences des gènes présents au sein des QTL sont analysées. L'analyse fonctionnelle des gènes contenus dans le QTL, consiste à considérer la fonction des gènes, les facteurs de transcription les régulant, le phénotype des mutants de délétion de ces gènes et les processus dans lesquels les gènes sont impliqués. À partir de ces informations, une liste de gènes candidats pour le QTL est établie. Les séquences de ces gènes sont ensuite comparées chez les deux parents afin de savoir s'il y a des variations nucléotidiques. Les mutations non synonymes peuvent avoir un impact direct sur le fonctionnement et la structure de la protéine ce qui peut avoir des conséquences au niveau phénotypique.

Pour déterminer si le gène candidat est impliqué dans le phénotype d'intérêt, la validation expérimentale sera réalisée par un remplacement allélique ou par un test d'hémizygoté réciproque.

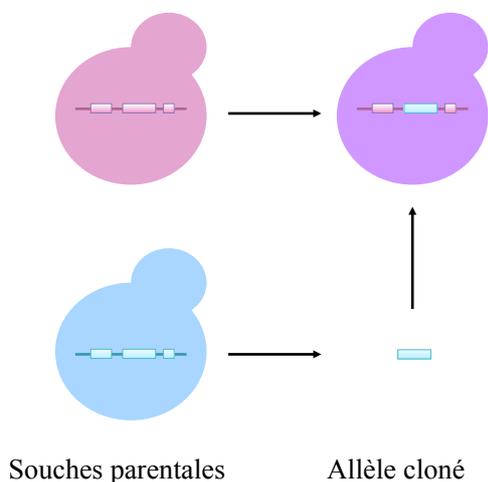
### 4.3.1. La validation expérimentale des gènes candidats

#### 4.3.1.1. Le remplacement allélique

Si deux versions alléliques du gène candidat existent, l'implication du gène candidat peut être contrôlée par remplacement allélique. Cette méthode consiste à remplacer l'allèle du parent qui ne possède pas le phénotype, par celui du parent qui le possède (figure 14). Ainsi, la souche obtenue après remplacement conservera son patrimoine génétique excepté pour le gène candidat. Si le remplacement est suffisant pour restaurer le phénotype d'intérêt dans sa totalité ou au moins en partie, son rôle en tant que déterminant génétique sera confirmé.



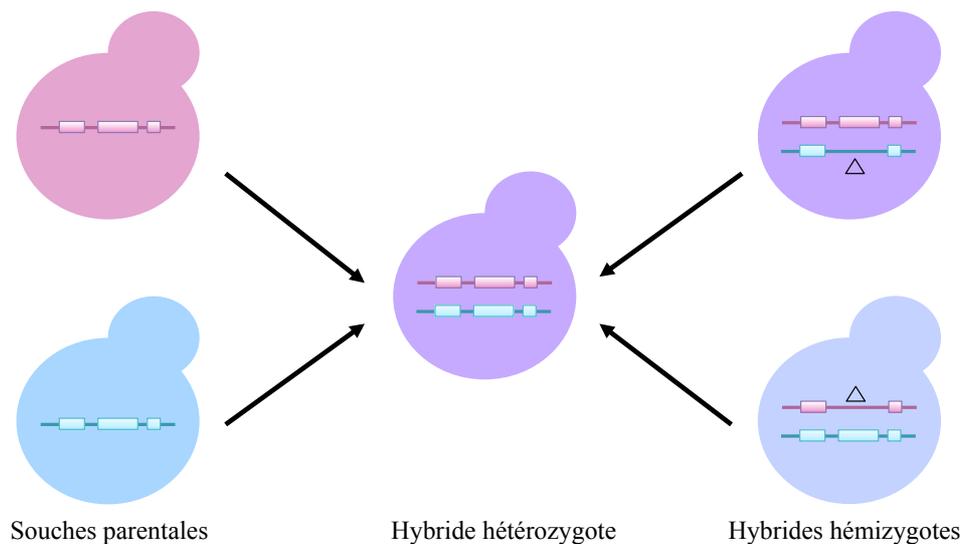
**Figure 13 :** Principe de l'analyse d'association chez la levure (schéma modifié de Mackay *et al.* 2009). Le phénotype et le génotype de chaque souche de la population sont déterminés. Les marqueurs détectés vont permettre d'identifier les différents génotypes. La corrélation entre chaque marqueur et le phénotype des descendants le possédant sera analysée. Cette approche va permettre d'identifier les marqueurs liés au caractère analysé.



**Figure 14 :** Schéma du remplacement allélique. Le remplacement allélique permet de tester la version allélique d'un gène candidat dans une souche possédant un autre fond génétique.

#### 4.3.1.2. Le test d'hémizygoté réciproque

Une souche diploïde est formée à partir du croisement entre les deux souches parentales (figure 15). Deux souches hémizygotés seront construites à partir de ce diploïde. La première sera délétée pour l'allèle potentiellement impliqué dans le phénotype du parent *MATa* et la deuxième sera délétée pour l'allèle du parent *MATα*. Les phénotypes de la souche diploïde hétérozygote et des souches hémizygotés seront comparés. Si les phénotypes sont différents entre les trois souches, alors le gène est impliqué dans la variation phénotypique des souches parentales.



**Figure 15 :** Schéma du test d'hémizygoté réciproque.

## 5. Les objectifs de ma thèse

L'étude de la diversité des génomes et des phénotypes au sein d'une espèce permet de comprendre son histoire évolutive. Les levures constituent un excellent organisme modèle pour cette thématique car elles réunissent les avantages des microorganismes (faibles besoins nutritifs, temps de génération court, génome contenant peu d'introns, mutants, possibilité de stockage des souches à long terme) et ceux des eucaryotes (compartimentation de la cellule, présence du cytosquelette, présence de la mitochondrie, compaction de l'ADN). De plus, de nombreux isolats sont disponibles chez plusieurs espèces comme *S. cerevisiae* ou *L. kluyveri*. Avec l'intérêt croissant qui leur est apporté, de plus en plus de souches seront isolées à travers le monde depuis des niches écologiques diverses. La variabilité intraspécifique génétique et phénotypique pourra être analysée grâce aux techniques à haut débit décrites précédemment. Il est possible de déterminer les origines génétiques de variations

phénotypiques grâce aux méthodes de cartographie des QTL que nous avons décrites ci-dessus. Ces stratégies ont déjà été utilisées chez *S. cerevisiae*. Elles seront mises au point et développées chez *L. kluyveri* dans le cadre de ce travail de thèse. En parallèle, nous nous sommes également focalisés sur l'étude de la variabilité de la résistance à des composés toxiques chez *S. cerevisiae* et *L. kluyveri*.

### **La variabilité intraspécifique de résistance à un composé toxique d'origine naturelle**

Dans un premier temps, nous nous sommes intéressés à l'analyse de la résistance à l'ammonium d'une collection de 63 souches de *S. cerevisiae*. L'ammonium est un composé qui peut être toxique lorsqu'il est présent en fortes concentrations dans le milieu. Ces concentrations élevées sont retrouvées dans des moûts de raisin à partir desquelles *S. cerevisiae* a pu être isolée montrant ainsi que des souches peuvent présenter un phénotype de résistance. L'étude de la variabilité de résistance à ce composé a été entreprise en utilisant une collection d'isolats provenant de différentes niches écologiques et géographiques. La souche K12 isolée à partir du saké a été caractérisée comme étant résistante à de fortes concentrations d'ammonium. Elle nous a permis, grâce à une analyse de pool de ségrégants, de trouver l'origine génétique du phénotype de la résistance à l'ammonium. Ce premier travail est présenté dans le chapitre 1 de ce manuscrit.

### **La variabilité intraspécifique de résistance à des composés toxiques d'intérêt médical**

*S. cerevisiae* a été isolé à de nombreuses reprises lors d'infections de patients immunodéprimés. En effet, *S. cerevisiae* est un pathogène opportuniste qui peut infecter des personnes dont les défenses immunitaires sont amoindries. C'est pourquoi la résistance à des antifongiques azolés, qui sont largement prescrits en cas d'infection fongique chez *S. cerevisiae*, a été analysée dans ce manuscrit. La variabilité de résistance au sein de la nouvelle collection de souches de *L. kluyveri* a été testée afin de voir si une levure d'origine naturelle qui n'a probablement aucune pression de type antifongique dans son environnement présente ou non des variabilités de résistance. Si oui quels mécanismes sont impliqués ? C'est dans cette optique que j'ai réalisé une analyse de pool de ségrégants chez *S. cerevisiae* et *L. kluyveri*. Les résultats sont présentés dans le chapitre 2 de ma thèse.

### **Des expériences d'évolution sur fluconazole ont généré des mécanismes communs d'adaptation entre *S. cerevisiae* et *L. kluyveri***

L'étude de la résistance à l'ammonium et aux antifongiques azolés a permis d'analyser la résistance naturellement présente au sein des collections de souches de *S. cerevisiae* et *L. kluyveri*. Cette diversité est le fruit de milliers d'années d'évolution et reflète l'histoire évolutive des deux levures. Les mécanismes moléculaires à l'origine de la résistance sont en partie connus mais peu d'études ont jusqu'à présent été réalisées sur les processus d'adaptation conduisant à la résistance.

C'est donc dans le but d'explorer les mécanismes évolutifs conduisant à la résistance chez *S. cerevisiae* et *L. kluyveri* que nous avons réalisé une expérience d'évolution adaptative sur fluconazole.

Une souche sensible de chaque espèce a été repiquée toutes les 48 heures durant 200 générations (environ 30 jours) en présence de fluconazole. Le séquençage des génomes de 6 souches évoluées a permis de mettre en évidence un mécanisme d'adaptation commun entre les deux espèces : la formation de chromosomes supplémentaires.

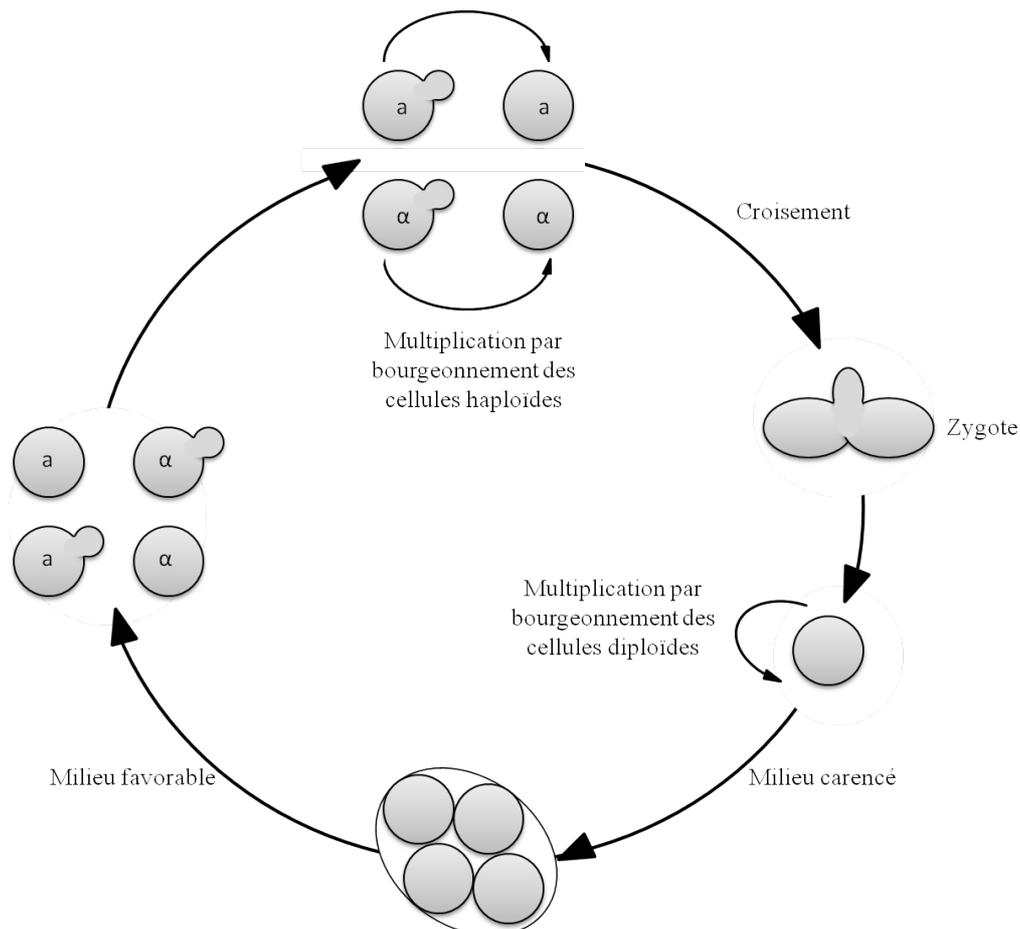
## Matériels et Méthodes

### 1. Organismes et méthodes de culture

#### 1.1. Organismes de l'étude

##### 1.1.1. *Saccharomyces cerevisiae*

*Saccharomyces cerevisiae* est un champignon unicellulaire Hémiascomycète homothallique. Son cycle de vie est haplodiplobiontique c'est-à-dire qu'il peut se maintenir sous forme haploïde et diploïde par multiplication végétative (bourgeoisement) (figure MM1). Lorsque les conditions de vie deviennent défavorables (carence en source de carbone ou d'azote) les levures diploïdes vont sporuler et donner naissance à un asque contenant quatre spores non ordonnées (Freese *et al.* 1982). Deux de ces spores sont de signe a, deux de signe alpha. Cependant les levures haploïdes homothalliques peuvent changer de signe sexuel. Ce phénomène est contrôlé par le gène codant l'endonucléase Ho (Herskowitz 1988). Quand les conditions redeviennent favorables les spores peuvent germer et ainsi former des cellules haploïdes. Deux cellules de signe opposé peuvent alors fusionner de manière à reformer un diploïde.



**Figure MM1 :** Cycle de vie de *Saccharomyces cerevisiae*.

Les caractéristiques des différentes souches de *S. cerevisiae* utilisées lors de ce travail sont répertoriées dans le tableau MM1. Les souches de *S. cerevisiae* utilisées sont délétées pour le gène codant l'endonucléase Ho par remplacement allélique avec le gène KanMx conférant la résistance à la généticine G418. Cette délétion empêche la souche de changer de signe sexuel, les souches deviennent alors hétérothalliques (Herskowitz 1988).

Souches	Origine écologique	Origine géographique	Ploidie
CLIB272	Bière	États-Unis	alpha
YJM145	Clinique. Patient SIDA	États-Unis	alpha
YJM280	Clinique. Fluide péritonéale	États-Unis	alpha
YJM320	Clinique. Sang	États-Unis	alpha
YJM326	Clinique. Inconnue	États-Unis	alpha
YJM421	Clinique. Fluide ascites	États-Unis	alpha
YJM678	Clinique. Inconnue		alpha
YJM428	Clinique. Fluide de paracentèse	États-Unis	alpha
YJM653	Clinique. Lavage broncho alvéolaire	États-Unis	alpha
YJM981	Clinique. Vagin	Italie	alpha
CECT10109	Figue	Espagne	alpha
DBVPG1788	Sol	Finlande	alpha
DBVPG1794	Sol	Finlande	alpha
DBVPG3591	Fèves de cacao	Inconnue	alpha
DBVPG4651	Tuber Magnatum	Italie	alpha
EM93	Figue pourrie	Californie	alpha
TL229	Fromage	France	alpha
YPS1009	Exudats de chêne	Inconnue	alpha
YPS163	Exudats de chêne	États-Unis	alpha
K12	Saké	Japon	alpha
Y10	Noix de coco	Phillippines	alpha
YJM269	Jus de pomme	Inconnue	alpha
FY4	Laboratoire	Inconnue	a
FY5	Laboratoire	Inconnue	alpha
CLIB154	Vin	Russie	alpha
I14	Sol du vignoble	Italie	alpha
UC8	Vin	Afrique du Sud	alpha
Y55	Vin	France	alpha
Y9J	Vin	Japon	alpha

**Tableau MM1 :** Les souches de *S. cerevisiae* utilisées ainsi que leurs origines écologiques et géographiques.

### 1.1.2. *Lachancea kluyveri*

*Lachancea kluyveri* est une levure Hémiascomycète hétérothallique. Les deux signes sexuels a et alpha sont présents et contrairement à *S. cerevisiae*, les haploïdes sont stables (hétérothallisme) du fait de la perte des cassettes silencieuses HML et HMR, et de l'absence du gène codant l'endonucléase

Ho (Payen *et al.* 2009). Son cycle de vie est, tout comme *S. cerevisiae*, haplodiplobiontique (figure MM1). Les souches de *L. kluyveri* utilisées sont listées dans le tableau MM2.

Souches	Ploïdie	Origine écologique	Origine géographique	Souche séquencée
dd.281a_b	alpha	Forêt	Allemagne, Heidelberg	Oui
CBS_5828_a	a	Sol	Danemark	Oui
CBS_5828_b	alpha	Sol	Danemark	Non
CBS_10369_b	alpha	Inconnue	Espagne, Cordou, Cruz Conde Park	Oui
NBRC_101999_a	a	Exsudat d'Ulmus	Espagne, Cordou, Cruz Conde Park	Oui
NBRC_101999_b	alpha	Exsudat d'Ulmus	Espagne, Cordou, Cruz Conde Park	Non
DBVPG_4002_a	a	Caverne	Italie	Non
DBVPG_4002_b	alpha	Caverne	Italie	Oui
DBVPG_3108	alpha	Sol	Pays-Bas	Oui
CBS_4104	a	Sol	Pays-Bas, Wageningen	Oui
CBS_2861	a	Sol	Suède	Oui
CBS_4568_b	alpha	Sol	Suède	Oui
DBVPG_3452_a	a	Sol	Suède, Ultuna	Oui
DBVPG_3452_b	alpha	Sol	Suède, Ultuna	Non
CBS_10367_b	alpha	Exsudat de Quercus mongoli	Russie, réserve naturelle de Kedrova pad	Oui
CBS_10368	alpha	Exsudat de Quercus mongoli	Russie, réserve naturelle de Sikhote-Alinsky	Oui
NBRC_105172_a	a	Mousse	Thaïlande	Oui
NBRC_10955_a	a	Feuille de <i>Rhizophora mucronata</i> en décomposition	Japon	Oui
NBRC_10955_b	alpha	Feuille de <i>Rhizophora mucronata</i> en décomposition	Japon	Non
NBRC_1811_a	a	Feuille en décomposition	Japon	Oui
NBRC_1811_b	alpha	Feuille en décomposition	Japon	Non
NBRC_1892_a	a	Exsudat d'arbre	Japon	Oui
NBRC_1892_b	alpha	Exsudat d'arbre	Japon	Non
CBS_6626_b	alpha	Exsudat d'arbre	Japon, Mt. Takam	Oui
67-544_b	alpha	Exsudat de <i>Pterocarya rhoifolia</i>	Japon, Mt. Takamatsu	Non
67-588	alpha	Exsudat d'Ulmus japonica	Japon, Yamabe	Oui
68-917.2_a	a	Exsudat d'érable	Canada	Oui
68-917.2_b	alpha	Exsudat d'érable	Canada	Non
62-196_b	alpha	<i>Taraxacum officinale</i>	Canada, Saskatoone	Oui
NBRC_1685_a	a	<i>Drosophila pinicola</i>	État-Unis	Non
CBS_6547_a	a	<i>Drosophila pseudobscura</i>	État-Unis, California, Gualala Creek	Non

CBS_6547_b	alpha	<i>Drosophila pseudobscura</i>	État-Unis, California, Gualala Creek	Oui
77-1003	alpha	Inconnue	État-Unis, Californie	Oui
CBS_6545_b	alpha	Exsudat de saule	État-Unis, Californie	Oui
CBS_6546_a	a	Peuplier	État-Unis, Californie	Non
CBS_6546_b	alpha	Peuplier	État-Unis, Californie	Oui
NCYC543	a	<i>Drosophila pinicola</i>	État-Unis, Californie	Oui
CBS_3082_a	a	<i>Drosophila pinicola</i>	État-Unis, Californie	Oui
CBS_3082_b	alpha	<i>Drosophila pinicola</i>	État-Unis, Californie	Oui
62-1041_b	alpha	Exsudat de saule	État-Unis, Californie, Davis	Oui

**Tableau MM2 :** Les souches de *L. kluyveri* utilisées ainsi que leurs origines écologiques et géographiques.

## 1.2. Milieux de culture

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

- milieu complet YPG : bactopeptone 20 g/l; yeast extract 10 g/l ; glucose 20 g/l
- milieu minimum YNB : yeast nitrogen base 6,7 g/l; glucose 20 g/l
- milieu de sporulation AcK : acétate de potassium 10 g/l.

Tous ces milieux peuvent être solidifiés par ajout de 2% d'agar. Le milieu YNB peut être supplémenté par ajout de bases ou d'acides aminés (50 µg/ml) en fonction des auxotrophies des souches utilisées. Différents composés antifongiques ont été ajoutés au milieu YPG. La liste de ces composés, ainsi que leurs concentrations finales utilisées dans les milieux de cultures liquides sont indiquées dans le tableau MM3. Les antifongiques sont achetés en poudre et sont ensuite dissouts dans du DMSO (diméthylsulfoxyde). Cette solution mère est ensuite utilisée pour supplémenter le milieu YPG.

Composé	Concentration finale
fluconazole	10 <sup>-5</sup> M et 4.10 <sup>-5</sup> M
kétoconazole	5.10 <sup>-7</sup> M et 5.10 <sup>-6</sup> M
itraconazole	10 <sup>-7</sup> M et 10 <sup>-5</sup> M
miconazole	10 <sup>-7</sup> M et 5.10 <sup>-7</sup> M
clotrimazole	10 <sup>-7</sup> et 5.10 <sup>-6</sup> M
nourcéothricine	100 µg/ml
généticine G418	200 µg/ml
hygromycine B	200 µg/ml

**Tableau MM3 :** Antifongiques ajoutés au milieu YPG et concentrations finales.

### 1.3. Conditions de culture

Les cellules sont cultivées soit en milieu liquide sous agitation soit en milieu solide. La température optimale de croissance est de 30°C pour *S. cerevisiae*, et de 25 à 30°C pour *L. kluyveri*. Afin d'homogénéiser les procédures nous avons décidé de réaliser toutes les cultures à 30°C.

### 1.4. Evaluation de la concentration cellulaire

Afin d'estimer la concentration en cellules des cultures de *S. cerevisiae* et *L. kluyveri* en milieu liquide nous avons choisi la technique de la mesure de la turbidité. Pour ce faire, un spectrophotomètre Spectronic® 20 Genesys™ a été utilisé afin de mesurer l'absorbance à 600 nm. Une unité d'absorbance équivaut à  $4,5 \cdot 10^7$  cellules haploïdes/ml et  $1 \cdot 10^7$  cellules diploïdes/ml.

### 1.5. Conservation des souches

Les souches de levure sont conservées à -80°C dans 1,5 ml de milieu complet contenant 20% de glycérol.

## 2. Techniques génétiques de la levure

### 2.1. Obtention de cellules diploïdes

Des cellules diploïdes peuvent être obtenues par le croisement d'une souche haploïde de signe  $a$  et d'une souche de signe opposé  $\alpha$  en phase exponentielle de croissance. Les zygotes obtenus ont une forme caractéristique facilement distinguable au microscope 3 à 4 heures après croisement. Ils peuvent ensuite être isolés à l'aide d'un micromanipulateur Singer de type MSM puis cultivés sur milieu riche.

### 2.2. Obtention de cellules haploïdes

Des cellules diploïdes en phase exponentielle de croissance sont transférées sur milieu AcK et incubées à 30°C pendant 2 à 4 jours. Sur ce milieu carencé, les diploïdes vont sporuler spontanément. Deux techniques permettent de libérer les spores contenues dans les tétrades, soit de manière ordonnée par la microdissection, soit de manière non-ordonnée par la technique des spores en vrac.

#### 2.2.1. La microdissection

Elle permet de séparer et d'ordonner les spores issues de la méiose. Une anse de milieu de sporulation est mise en suspension dans 200  $\mu$ l d'eau stérile à laquelle sont ajoutés 50  $\mu$ l de zymolyase 20 T à 0,5 mg/ml. Le mélange est incubé 15 min à 30°C afin que la zymolyase digère la paroi des asques. La zymolyase hydrolyse la paroi des asques sans que les tétrades ne se séparent. Les spores d'une tétrade

sont séparées et isolées sur une boîte de milieu complet à l'aide d'un micromanipulateur Singer de type MSM. Les clones des spores apparaissent au bout de 48 h à 96 h et peuvent ensuite être analysés séparément. Ce protocole tel quel, ne fonctionne pas sur *L. kluyveri*.

### **2.2.2. Le test de spores en vrac**

Ces tests permettent de libérer de manière non-ordonnée les spores contenues dans les asques.

#### **2.2.2.1. Le test éther pour *S. cerevisiae***

L'éther tue les diploïdes, détruit la paroi des asques et libère les spores (Dawes and Hardie 1974). Il est ajouté 1 ml d'éther à 0,5 ml d'une suspension de milieu de sporulation. Ce mélange est vortexé pendant 30 s, incubé 20 min à température ambiante puis dilué en cascade au 1/10 et au 1/100. Ces dilutions sont étalées sur milieu complet puis analysées.

#### **2.2.2.2. Le test par billes de verre pour *L. kluyveri***

Les spores sont mises en suspension dans 400 µl de solution de DTT à 5 mM. Le mélange est incubé pendant 30 min à 25°C. Des billes de verre d'environ 200 µm de diamètre sont ajoutées. Le tube est ensuite mis à agiter sur un vortex à moyenne intensité durant 30 min. Puis 150 µl de solution de triton X-100 à 0,1% sont ajoutés. Le mélange est remis en suspension par agitation sur vortex à vitesse maximum durant 2 min. Enfin des dilutions en cascade de la solution sont effectuées, de 10<sup>-1</sup> à 10<sup>-4</sup>, et 150 µl de chaque dilution sont étalés sur milieu YPD.

## **2.3. Détermination du signe sexuel**

### **2.3.1. Détermination du signe sexuel par croisement**

Le signe sexuel d'une souche peut être déterminé en croisant cette souche sur YPG avec des souches contrôles pour lesquelles le signe *MATa* et *MATα* est connu. Les diploïdes qui auront poussé sont mis à sporuler sur le milieu de sporulation AcK. Une observation au microscope permet de déterminer quels sont les croisements ayant abouti à la formation d'un zygote capable de sporuler. Ainsi la présence des spores indique que la souche inconnue est de signe opposé à la souche de référence avec laquelle elle a été croisée.

### **2.3.2. Détermination du signe sexuel par le test de halo mating type**

Ce test permet de déterminer le signe sexuel des souches de *S. cerevisiae*. En plus des cellules dont on veut déterminer le signe sexuel, deux souches sont utilisées :

- DBY7730 (*MATa*) dont le génotype est *ade2 his6 met1 ura1 can1 cyh1 rme sst1-3*.
- DBY7442 (*MATα*) dont le génotype est *sst2* et le phénotype est Leu- Ura- Ade- (le génotype exact n'est pas décrit pour cette souche).

Elles ont la particularité d'être sensibles respectivement au facteur  $\alpha$  et au facteur  $a$ . Ces deux souches sont mises en culture pendant 12 h dans 10 ml de milieu YPD. Elles sont ensuite diluées au 1/10 avec de l'eau stérile et 200  $\mu$ l de ces dilutions sont étalés sur deux boîtes différentes de milieu complet (il faut une boîte par souche). Ces dernières sont alors incubées à 30°C pendant 30 min.

Les souches dont on veut déterminer le signe sexuel sont alors déposées sur ces boîtes, soit par repiquage au cure-dent soit par test en gouttes, puis incubées pendant deux jours à 30°C. Les colonies présentant un halo d'absence de croissance, sur la boîte où la souche DBY7730 est étalée, sont de signe  $\alpha$  alors que celles ayant un halo sur la seconde boîte sont de signe  $a$ .

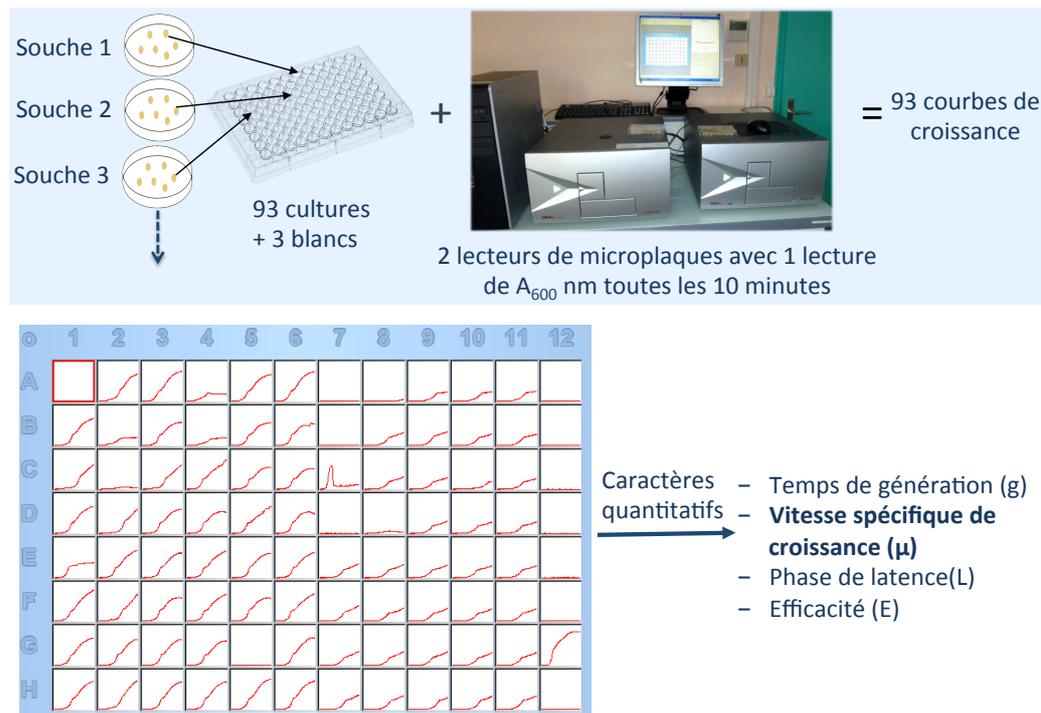
## 2.4. Détermination des phénotypes

### 2.4.1. Analyse de la croissance par suivi de microcultures

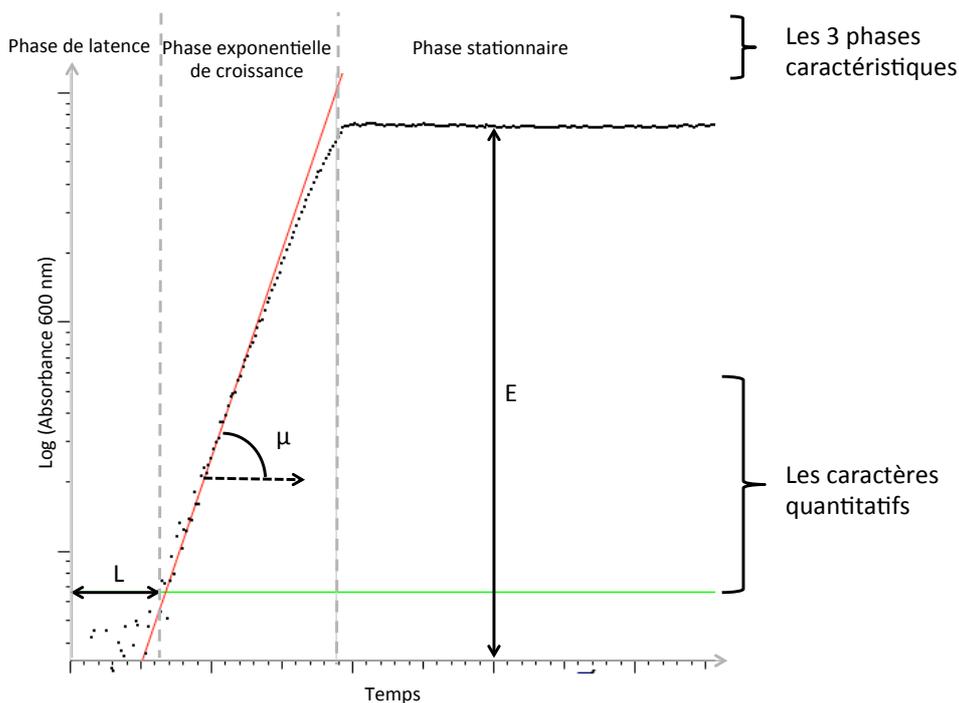
#### 2.4.1.1. La courbe de croissance et les caractères quantitatifs mesurés

Les suivis de croissance sont réalisés par microcultures en microplaques de 96 puits dans un volume final de milieu de 150  $\mu$ l (figure MM2). Le suivi d'une microplaque permet donc de phénotyper jusqu'à 93 souches par plaque, les 3 puits résiduels servant de contrôle. Les microplaques sont incubées à 30°C dans un appareil Tecan Infinite 200 Pro qui permet à la fois d'incuber, d'agiter et de lire les absorbances des cultures. Une mesure de l'absorbance à 600 nm à intervalle de temps régulier durant 48h permet de déterminer des courbes de croissance.

La croissance des micro-organismes se caractérise en quatre phases : la phase de latence, la phase exponentielle, la phase stationnaire et la phase de déclin (figure MM3). Seules les trois premières phases sont étudiées lors de ce travail. La phase de latence est la période pendant laquelle la population s'adapte à son milieu et ne se reproduit que faiblement. La période d'adaptation dépend de l'organisme et du nouveau milieu dans lequel la souche doit se développer. Pendant la phase exponentielle, la population se développe à sa vitesse maximale selon le modèle exponentiel. La phase stationnaire concorde avec une vitesse spécifique de croissance ralentie du fait de l'appauvrissement du milieu. Il n'y a plus assez de nutriments pour couvrir les besoins de toute la population. Une partie de la population ne bourgeonne plus, une partie continue à se multiplier et une autre meurt. Le niveau total de la population apparaît statique tout le long de cette phase. A partir des courbes de croissances, les caractères quantitatifs phase de latence, vitesse spécifique de croissance et efficacité ont pu être extraits. Pour l'analyse de la résistance nous nous sommes uniquement basés sur la vitesse spécifique de croissance  $\mu$ .



**Figure MM2 :** Stratégie de phénotypage employée permettant de tester la résistance des souches aux composés azolés.



**Figure MM3 :** Illustration d'une courbe de croissance avec ses 3 premières phases caractéristiques. La ligne verte représente le niveau minimal d'absorbance que le spectrophotomètre peut détecter. La phase de latence est la durée entre le temps 0 et le moment où l'absorbance a dépassé le niveau basal. La ligne rouge représente la pente de la phase exponentielle de croissance, c'est le moment où la croissance atteint la vitesse la plus élevée. Cette pente correspond à la vitesse spécifique maximale de croissance. Enfin la hauteur de la courbe en phase stationnaire détermine l'efficacité.

### 2.4.1.2. Analyse des courbes de croissance

Les mesures d'absorbance sont extraites et analysées par le logiciel *Courbes de croissance 1.4* mis au point par Benoît Kammerer. La moyenne des absorbances des puits non inoculés est soustraite aux absorbances déterminées pour chacun des puits inoculés. Le logiciel permet le calcul du temps de génération ( $g$ ), de la vitesse spécifique de croissance ( $\mu$ ), et de la durée phase de latence ( $L$ ) (figure MM3). La biomasse générée ( $E$  pour efficacité) peut être mesurée par la lecture de l'absorbance finale à 600 nm.

Chez *S. cerevisiae*, un premier criblage des souches de la collection, nous a permis de sélectionner 28 souches haploïdes sur des critères de reproductibilité et d'absence d'agrégats (tableau MM1). Chez *L. kluyveri*, la résistance de toutes les souches a pu être testée par microcultures car les courbes de croissance sont précises et reproductibles.

## 3. Préparation des acides nucléiques

### 3.1. Préparation de l'ADN

#### 3.1.1. Extraction d'ADN génomique de levure par la méthode phénol chloroforme (Hoffman and Winston, 1987)

Une culture de 15 ml en phase stationnaire dans du milieu YPG est centrifugée pendant 5 minutes à 3000 rpm. Les cellules sont reprises dans 1 ml d'eau distillée. Une centrifugation pendant 2 minutes à 12400 rpm permet de récolter à nouveau les cellules et de les reprendre dans 250  $\mu$ l de tampon de broyage (Tris HCl pH 8 10 mM ; EDTA 1 mM ; NaCl 100 mM ; Triton 2% ; SDS 1%). Environ 100  $\mu$ l de billes de verre et 250  $\mu$ l de phénol chloroforme sont ensuite ajoutés. Le mélange est vortexé pendant 3 minutes puis centrifugé 5 minutes à 11000 rpm. La phase aqueuse dans laquelle se situent les acides nucléiques est prélevée délicatement. Après addition d'un volume de phénol chloroforme, le mélange est vortexé pendant 45 secondes. Il est ensuite centrifugé pendant 2 minutes à 11000 rpm. La phase aqueuse est à nouveau prélevée et 20  $\mu$ l d'acétate de sodium à 4 M et 600  $\mu$ l d'éthanol absolu sont ajoutés. Une précipitation de 1 heure à - 20°C suivit d'une centrifugation 15 minutes à 12400 rpm sont réalisées. Le surnageant est éliminé et le culot séché est repris dans 400  $\mu$ l de tampon TE (Tris HCl pH 8 10 mM et EDTA 1 mM). Les ARN sont ensuite dégradés par l'ajout de 5  $\mu$ l de RNase A (10 mg/ml) par une incubation de 5 minutes à 37°C. Une seconde précipitation est alors effectuée en additionnant 10  $\mu$ l d'acétate de sodium 4 M et de 1 ml d'éthanol absolu et en laissant précipiter l'ADN pendant au minimum 1 heure à - 20°C. Une centrifugation à 4°C pendant 15 minutes à 12400 rpm permet d'obtenir un culot qui est lavé avec 1 ml d'éthanol à 80%. Suite à une centrifugation de 5 minutes à 12400 rpm, le culot contenant l'ADN est séché puis repris dans 50  $\mu$ l de tampon TE.

### 3.1.2. Extraction d'ADN génomique total par kit Qiagen

Les ADN génomiques destinés au séquençage ont été extraits sur colonnes grâce au kit Qiagen genomic DNA en suivant les instructions du fabricant données dans le protocole à la section « Yeast ».

### 3.1.3. Préparation de chromosomes entiers de levure

L'analyse des caryotypes de levures par électrophorèse en champs alternés nécessite d'obtenir des chromosomes intacts (Carle and Olson 1985). Les procédures habituelles de préparation d'ADN ne permettent pas d'obtenir des molécules de haut poids moléculaire. Pour éviter les coupures dues aux forces mécaniques lors des différents mélanges et agitations, les cellules entières incorporées dans des blocs d'agarose sont lysées et déprotéinisées *in situ*. La matrice d'agarose protège l'ADN des forces de scission et fournit des échantillons faciles à manipuler pour l'analyse par électrophorèse en champs alternés.

Une culture contenant  $1-2 \cdot 10^8$  cellules est centrifugée 10 min à 2500 rpm. Les cellules sont ensuite lavées dans 20 ml d'EDTA 50 mM froid puis dans 10 ml de CPES froid formé d'un mélange de CPE×1 (acide citrique 40 mM,  $\text{Na}_2\text{HPO}_4$  120 mM, EDTA 20 mM), sorbitol 1,2 M et DTT 5 mM. Les cellules sont alors placées dans la glace puis reprises dans 1 ml de CPES. Un aliquote de 500  $\mu\text{l}$  de cellules auquel est ajouté 0,5 mg de zymolyase 100T est alors placé dans un bain-marie à 37°C pendant 3 min pour activer la zymolyase. Un volume de 500  $\mu\text{l}$  d'agarose BioRad 1,2% dans du CPE×1 est ajouté. Le mélange est vortexé et prélevé avec une pipette Pasteur puis coulé dans les puits du moule. Une fois pris en masse, les blocs sont démoulés, recouverts par 15 ml de CPE + zymolyase 100T (1,5 mg) puis incubés 2 h à 37°C. Le tampon est ensuite remplacé par 10 ml de tampon de lyse (Tris 10 mM pH8 ; EDTA 0,45 M ; SDS 1% ; Protéinase K 0,8 mg/ml) et les blocs sont incubés une nuit à 50°C. Le lendemain, deux lavages de 30 min à l'eau à 50°C sont réalisés suivis de deux lavages de 30 min à l'eau à température ambiante. Les blocs sont ensuite repris et conservés dans 10 ml d'EDTA 0,25 M à 4°C.

## 3.2. Extraction des ARN totaux

La préparation des ARN totaux a été effectuée en combinant le protocole d'extraction au phénol chaud et la purification sur colonnes RNeasy commercialisées par Qiagen (le protocole utilisé est décrit par le fournisseur). Toutes les solutions et tout le matériel utilisés sont précautionneusement préparés puis utilisés dans un contexte sans nucléases.

La première étape est de récupérer les cellules en milieu de phase exponentielle, à une absorbance comprise entre 0,35 et 0,45. Les cellules sont alors déposées sur un filtre (0,45 $\mu\text{m}$ ) et l'ensemble est congelé instantanément par trempage dans de l'azote liquide. Cela permet de conserver les ARN

intacts et de ne pas induire de stress pouvant changer le contenu en ARN de la cellule. Les filtres sont ensuite conservés à  $-20^{\circ}\text{C}$  jusqu'à l'extraction des ARN.

Les filtres portant les cellules sont placés dans un tube Eppendorf de 2 ml avec 750  $\mu\text{l}$  de tampon de lyse (TrisHCl pH 7,5 10 mM, EDTA 10 mM, SDS 0,5%) et vortexés pendant 3 min. Le même volume de phénol acide est ajouté, le mélange est vortexé à nouveau pendant 3 min avant d'être incubé 1 h à  $60^{\circ}\text{C}$ . Lors de cette incubation, les tubes sont vigoureusement mélangés toutes les 20 min. Afin d'engendrer un choc thermique, les tubes sont placés dans la glace pendant 10 min puis centrifugés 5 min à 12400 rpm. Les acides nucléiques, contenus dans la phase supérieure, sont transférés dans un tube Phase Lock Gel™ Heavy commercialisé par Eppendorf. Un volume de 750  $\mu\text{l}$  de chloroforme est ajouté et le tube est mélangé par inversion avant d'être centrifugé 5 min à 12400 rpm. Le surnageant est récupéré puis placé dans un nouveau tube et les acides nucléiques sont précipités par ajout de 75  $\mu\text{l}$  d'acétate de sodium 3 M et de 1,5 ml d'éthanol absolu puis incubés pendant une nuit à  $-20^{\circ}\text{C}$ .

Les tubes sont ensuite centrifugés 15 min à 12400 rpm. Le culot est repris dans 350  $\mu\text{l}$  d'éthanol 80% et 350  $\mu\text{l}$  de tampon RLT puis le mélange est placé sur une colonne RNeasy qui permet de purifier les ARN. Cette dernière est centrifugée pendant 15 s à 12400 rpm. Elle est ensuite lavée successivement par 700  $\mu\text{l}$  de tampon RW1 puis 500  $\mu\text{l}$  de tampon RTE. Après chaque lavage, la colonne est centrifugée 15 s à 12400 rpm. Elle est à nouveau rincée par l'ajout de 500  $\mu\text{l}$  de tampon RTE puis centrifugée 2 min. La colonne est placée dans un nouveau tube Eppendorf où les ARN sont élués par ajout de 50  $\mu\text{l}$  d'eau et centrifugés pendant 1 min à 12400 rpm. La composition des tampons RLT, RW1 et RTE n'est pas donnée par le fournisseur.

La non dégradation des ARN totaux est à contrôler par dépôt et migration sur gel d'agarose 1% (partie 4.2.1).

## **4. Purification, quantification et analyse des acides nucléiques**

### **4.1. Purification de l'ADN**

#### **4.1.1. Précipitation alcoolique de l'ADN**

En présence d'éthanol et d'un sel, les acides nucléiques précipitent. Cette propriété biochimique est utilisée pour purifier une solution d'ADN ou la concentrer.

À l'ADN extrait sont ajoutés 2,5 volumes d'éthanol 100% froid et 1/10 de volume d'acétate de sodium 3 M pH7. Le mélange est incubé à  $-20^{\circ}\text{C}$  pendant au moins 1 heure puis est centrifugé 15 min à 12400 rpm. Le culot d'ADN est ensuite lavé avec de l'éthanol 70% froid puis l'ADN est dissout dans le volume désiré d'eau distillée ou de tampon TE. Cette solution peut ensuite être conservée à  $-20^{\circ}\text{C}$ .

### **4.1.2. Purification de l'ADN à l'hexamine cobalt**

Cette purification sert à enlever les polysaccharides par précipitation. Elle a été utilisée pour tous les ADN génomiques qui ont été séquencés.

L'ADN extrait est mélangé à 1/10<sup>ème</sup> de volume de solution d'hexamine cobalt chlorure à 100 mM. L'ADN va ainsi précipiter. En cas de grandes quantités d'ADN, le surnageant peut être retiré sans centrifugation, ce qui facilitera la re-suspension dans le tampon d'échange. Dans le cas contraire, le mélange est centrifugé 5 min à 13000 rpm. Le culot est lavé avec 500 µl d'eau puis dissout dans un tampon d'échange (100 mM EDTA (pH = 8) ; 2 M Guanidium thiocyanate). Pour faciliter la dissolution le mélange peut être incubé sous agitation à 37°C. Un volume d'isopropanol est ajouté au mélange. Le tout est incubé 10 min à température ambiante puis centrifugé 5 min à 13000 rpm. Le culot est lavé avec de l'éthanol à 70%, séché puis dissout dans 200 à 50 µl d'eau MQ selon la taille du culot.

### **4.2. Purification des ARN par traitement à la DNase**

Les ARN extraits doivent être traités à la DNase afin d'enlever l'ADN résiduel qui fausserait la quantification des ARN dans une analyse de PCR quantitative.

Le kit utilisé est le DNase I amplification grade<sup>®</sup> d'Invitrogen. À une extraction d'ARN de 50 µl sont ajoutés 10 µl de tampon et 2 µl de DNase fournis par le kit. Le mélange est incubé 15 min à température ambiante. Puis 4 µl d'EDTA fournis par le kit sont additionnés puis mis à incuber 10 min à 65°C. La réaction est stoppée en mettant le mélange dans la glace.

Afin de contrôler que la purification ait fonctionné, une PCR contrôle est effectuée sur les ARN purifiés pour vérifier l'absence d'ADN résiduel.

### **4.3. Contrôle et dosage des acides nucléiques**

#### **4.3.1. Analyse des acides nucléiques par électrophorèse sur gel d'agarose**

Les molécules d'acides nucléiques peuvent être séparées en fonction de leur taille par électrophorèse sur gel d'agarose 1% dans du tampon TAE (Tris 4,84 g ; acétate de sodium 0,68 g ; acide acétique 1,35 ml ; EDTA 0,336 g ; H<sub>2</sub>O qsp 1 l) sous une tension de 30 V à 90 V. L'ajout de bromure d'éthidium (BET) dans le gel ou dans un bain après migration permet d'observer les molécules nucléaires sous rayons UV (Ultra Violet).

#### **4.3.2. Quantification des acides nucléiques**

La concentration d'une solution d'ADN, d'ADNc ou d'ARN peut être déterminée en mesurant leur absorbance à 260 nm. En effet, à 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

$\mu\text{g/ml}$ , 33  $\mu\text{g/ml}$  ou 40  $\mu\text{g/ml}$ . Dans cette étude, les mesures ont été réalisées à l'aide d'un spectrophotomètre de type Nanodrop ND1000.

Le rapport des mesures des longueurs d'ondes 260/280 permet d'estimer la pureté de l'extraction d'acide nucléique et donc l'absence de contaminants (protéines, phénol). Un rapport de 1,8 indique que l'extraction d'ADN est pure, et un rapport de 2 indique que l'extraction d'ARN est pure. Le rapport des mesures d'absorbance 260/230 permet de contrôler l'absence de contaminants absorbants à une longueur d'onde de 230 nm. Une valeur comprise entre 2 et 2,2 signifie que l'extraction est non contaminée.

### 4.3.3. Séparation des chromosomes par électrophorèse en champs alternés

L'électrophorèse en champs alternés (ou PFGE pour Pulsed Field Gel Electrophoresis) permet de séparer des fragments d'ADN de quelques kilobases (kb) à plus de 10000 kb ce qui va ainsi permettre d'observer les chromosomes de *S. cerevisiae* et *L. kluyveri* dont la taille varie de 250 à 2500 kb.

Le système utilisé est le CHEF (Contour-clamped Homogeneous Electric Fields) où les électrodes sont organisées en hexagone permettant de définir deux champs électriques séparés de  $120^\circ$  (Chu, 1989). Les fragments d'ADN sont soumis à ces deux champs électriques alternativement. Le principe de séparation est basé sur le fait qu'il existe une vitesse différentielle de réorientation des molécules lors du changement d'orientation du champ électrique. Ainsi les grandes molécules se réorientent moins rapidement que les petites et l'accumulation de ces retards va permettre la séparation des molécules en fin de migration. La migration s'effectue sur gel d'agarose 1% dans le tampon TBE 0,5x (Tris 5,4 g ; acide borique 2,75 g ; EDTA 0,5M 2 ml ; H<sub>2</sub>O qsp 1 l) à une température de 14°C. Les conditions utilisées sont les suivantes :

- pour *S. cerevisiae* : 22 h de migration à 6 V/cm avec un temps de commutation initial de 40 s et final de 120 s, l'angle entre les deux champs étant de  $120^\circ\text{C}$
- pour *L. kluyveri* : une migration en 3 phases : 1) 20 h de migration à 3 V/cm avec un temps de commutation de 480 s, 2) 20 h de migration à 4 V/cm avec un temps de commutation de 300s, 3) 12 h de migration à 6 V/cm avec un temps de commutation de 90 s, l'angle entre les deux champs étant de  $120^\circ\text{C}$  (Neuvéglise *et al.* 2000).

## 5. Amplification d'un fragment d'ADN par PCR

### 5.1. Principe

La PCR pour Polymerase Chain Reaction est une technique de biologie moléculaire qui permet l'amplification exponentielle et spécifique par une ADN polymérase thermostable d'un fragment spécifique d'ADN en utilisant des oligonucléotides amorces et des dNTP. Elle comprend trois étapes qui sont répétées pendant 30 à 35 cycles. Le nombre de cycle dépend de la quantité d'ADN matrice initiale. Les trois étapes sont :

- la dénaturation qui permet la séparation de l'ADN bicaténaire en ADN monocaténaire. La température classiquement utilisée est de 98°C pendant 15 s.
- l'hybridation des amorces sur les sites spécifiques de l'ADN. C'est l'étape clé pour déterminer la spécificité et le rendement de la réaction de PCR. En fonction de la séquence et de la taille des amorces, la température d'hybridation ( $T_m$ ) est différente. Elle est généralement comprise entre 50°C et 60°C. La durée de cette étape est de 30 s.
- l'élongation durant laquelle le brin complémentaire spécifique au brin matrice est synthétisé. Elle est réalisée à une température optimale pour l'enzyme utilisée (72°C pour la *Taq* DNA polymérase et la polymérase iProof). Lorsque la *Taq* DNA polymérase est utilisée, la durée d'élongation est d'1 min/kb, alors que si l'on utilise l'enzyme iProof, la durée d'élongation est de 30 s/kb.

Une étape de post-élongation de 10 min à 72°C est effectuée après la fin des cycles afin de terminer la synthèse de certains brins éventuellement inachevés. Nous utilisons au laboratoire un thermocycleur de type Eppendorf Mastercycler EP.

### 5.2. Milieux de PCR

Deux enzymes polymérases sont utilisées lors de ces travaux. La *Taq* polymérase pour l'amplification des fragments de taille inférieure à 5 kb, et la polymérase iProof pour les fragments supérieur à 5 kb ou pour des fragments qui doivent être amplifiés avec une haute fidélité (peu d'erreurs dans la séquence) car ces derniers seront utilisés dans les remplacements alléliques.

La *Taq* DNA polymérase commercialisée par MP - QBIogene est fournie avec un tampon x10 dont la constitution est : Tris-HCl pH 8,3 100 mM; KCl 500 mM; MgCl<sub>2</sub> 15 mM; gélatine 0,01%. La composition du milieu de réaction est : 10 µl de tampon 10x; 1 µl d'une solution des 4 dNTPs à 20 mM; 1 µl de chaque oligonucléotide à 20 µM; environ 100 ng d'ADN; 0,2 unité d'enzyme. Le mélange réactionnel est complété avec de l'eau pour avoir un volume final de 50 µl.

La polymérase iProof est commercialisée par Bio-rad. Elle est fournie avec le HF buffer concentré 5x et contenant 7,5 mM de  $MgCl_2$ . La composition d'un milieu réactionnel est : 10  $\mu$ l de HF Buffer ; 1  $\mu$ l d'une solution des 4 dNTPs ; 1  $\mu$ l de chaque oligonucléotide à 20  $\mu$ M environ 100 ng d'ADN ; 0,5  $\mu$ l d'enzyme. Le mélange réactionnel est complété à un volume total de 50  $\mu$ l.

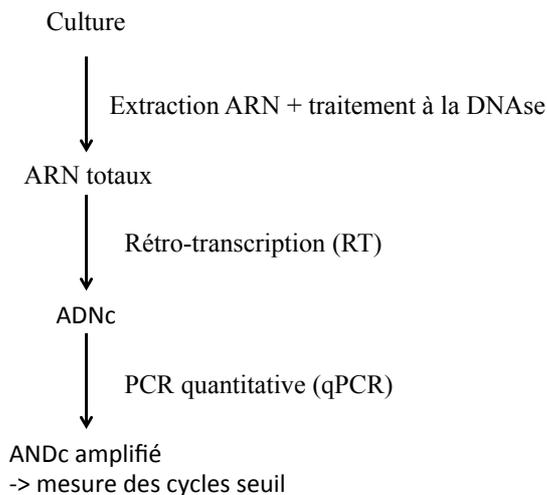
Les amorces utilisées dans le cadre du chapitre 1 et 2 et leurs séquences sont données dans les manuscrit respectifs des chapitres dans les sections « material and methods », les amorces utilisées dans le chapitre 2 sont quant à elles décrites dans l'annexe MM1.

## 6. Dosage d'ARN par RT-PCR quantitative

Les RT-PCR quantitatives ont été utilisées pour quantifier l'ARN présent dans une matrice initiale. La première étape est la rétro-transcription (RT) de l'ARN en ADN complémentaire (ADNc), et la deuxième est de réaliser la PCR quantitative. Ces deux étapes peuvent être réalisées en une fois, dans ce cas, la reverse-transcription et la PCR sont effectuées dans le même tube lors d'un cycle de RT-PCR classique. Pour les qRT-PCR en deux étapes, la reverse-transcription est réalisée avant la qPCR, et la quantification est faite au niveau des ADNc.

Deux types de qPCR existent : les qPCR absolues et les qPCR relatives. Dans le premier cas de figure, la matrice initiale est quantifiée grâce à une gamme étalon d'acide nucléique de référence, alors que dans le second, c'est la comparaison entre les résultats des ARN contrôles avec les ARN à tester qui permet de quantifier les ARN.

Nous avons choisi de faire des qRT-PCR relatives en deux étapes (figure MM4) afin de déterminer les quantités relatives des transcrits impliqués dans la résistance multiple de différentes souches et ce, dans différentes conditions.



**Figure MM4 :** Schéma récapitulatif des expériences à réaliser, en partant d'une culture cellulaire, pour quantifier des ARN.

### 6.1. Réverse transcription des ARN totaux par amorçage aléatoire

Cette étape va permettre de réverse transcrire l'ARN en ADN complémentaire (ADNc).

Pour réaliser ces expériences, le kit SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR d'Invitrogen a été utilisé. Ce kit est composé d'une enzyme permettant la réverse-transcription (RT), de la Rnase H d'*Escherichia coli* qui permettra de dégrader les ARN après la synthèse de l'ADNc et d'un mix réactionnel x2 (oligo(dT)20 2,5 µM, hexamères aléatoires 2,5 µg/ml, MgCl<sub>2</sub> 10 mM, dNTPs).

Pour réaliser la synthèse d'ADNc, le mélange réactionnel pour un volume de 20 µl est le suivant : 10 µl de mix réactionnel x2, 2 µl de l'enzyme RT, 1 µg d'ARN et de l'eau qsp 20 µl. Le mélange est placé 10 min à 25°C puis 30 min à 50°C. Il est ensuite mis à 85°C pendant 5 min avant d'être refroidi dans la glace, afin de terminer la réaction et d'inactiver l'enzyme. Pour éliminer les ARN résiduels, 2 unités de RNase H sont ajoutées au mélange qui est placé 20 min à 37°C.

### 6.2. Réaction de qPCR

Les réactions de qPCR ont été effectuées en utilisant le kit QuantiFast® SYBR® Green PCR kit de Qiagen. Il est composé du mélange réactionnel 2x QuantiFast SYBRGreen PCR Master Mix et d'eau RNase free. Le SYBR Green ® est un intercalant de l'ADN qui s'insère dans l'ADN bicaténaire toutes les 20 paires de bases environ. Grâce à ses propriétés de fluorescence, il va permettre de suivre la quantité de fragments de PCR formés à chaque cycle.

Chaque réaction est effectuée dans une microplaque de 96 puits avec un volume final de 25 µl. Une réaction est composée de 12,5 µl du mélange réactionnel, de 2,5 µl de chaque primer à 10 µM, de 2,5 µl d'eau RNase free et de 5 µl contenant au moins 100 ng d'ADNc. La plaque de 96 puits est alors vortexée à 900 rpm pendant 1 min, puis centrifugée quelques secondes afin que le liquide retombe au fond des puits. Les amorces utilisées sont listées dans le tableau MM4.

Les réactions de qPCR sont réalisées en triplicat technique dans un thermocycleur MyiQ Single-Color Real-Time PCR Detection System commercialisé par Biorad. Le programme de la réaction de qPCR à utiliser est le suivant :

- une étape d'activation de l'enzyme par la chaleur à 95°C durant 5 min
- un cycle en 2 étapes :
  - o une dénaturation de 10 s à 95°C
  - o une étape combinant l'hybridation des amorces et la polymérisation des nouveaux brins à 60°C durant 30 s.

Ce cycle est répété 40 fois.

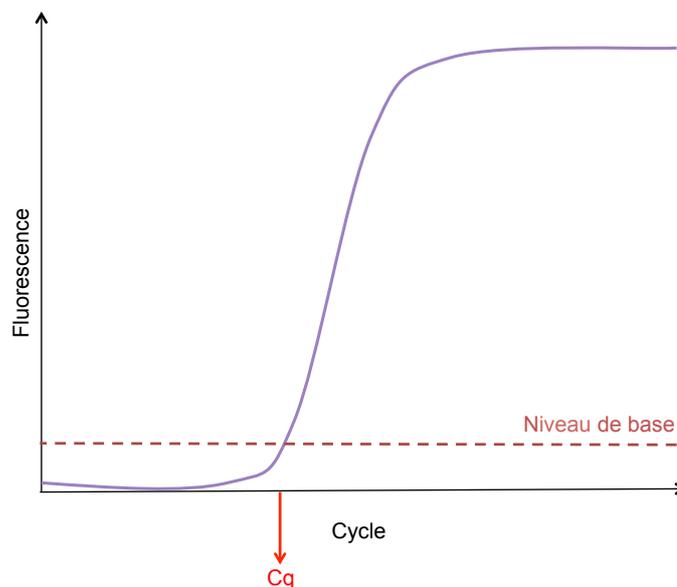
Gène	Séquences		Fonction
	Amorce 1	Amorce 2	
<i>TUB2</i>	TTACCGTTGCAGCCTTCTTT	AAACCTTGAGGAGCGACAGA	contrôle
<i>QCR9</i>	ATCTTTGCAGGTGCCTTTGT	GCAGCTATTCGAGCCTTGAC	contrôle
<i>PDR5</i>	AAAACCTCCACTCAATCGGCA	AAATTTTCGGAGTTGGGGTC	à tester

**Tableau MM4 :** Amorces utilisées pour la qPCR. La longueur des séquences amplifiées est comprise entre 200 et 300 pb.

### 6.3. Analyse des résultats de qPCR

Le cycle seuil de quantification Ct pour Threshold Cycle est la valeur clé utilisée pour l'analyse d'une qPCR. La valeur Ct représente le nombre de cycles de PCR nécessaires pour que le signal de fluorescence atteigne un niveau seuil (figure MM5). Le seuil est paramétré de manière à ce qu'il capture les données au cours de la phase exponentielle d'amplification. La valeur Ct est corrélée au nombre de copies de matrice présentes au départ. Une faible valeur de Ct reflète une forte quantité de matrice au départ, tandis qu'une valeur haute signifie que peu de matrice est présente au départ.

Les valeurs de Ct sont analysées grâce au logiciel REST2009 (Pfaffl *et al.* 2002). Les valeurs Ct des ARN contrôles (*TUB2* et *QCR9*), sont utilisées afin de normaliser les quantités des ARN à tester. Le but de ces analyses est d'estimer l'expression d'un même gène entre plusieurs souches, ou d'une même souche mais dans plusieurs conditions.



**Figure MM5 :** Profil de fluorescence d'une réaction de qPCR. Le niveau basal de fluorescence permettant de définir la valeur Ct est représenté en pointillés.

## 7. Séquençage

### 7.1. Séquençage des génomes par la technique Illumina HiSeq 2000

La stratégie utilisée pour le séquençage des génomes est le séquençage par Illumina HiSeq 2000. Cette méthode est basée sur le séquençage par synthèse (Mardis 2008). Pour construire la librairie, l'ADN est fragmenté par sonication. Des adaptateurs sont liés aux extrémités des fragments d'ADN. Les fragments de la taille désirée sont sélectionnés et amplifiés par PCR. Les brins sont dénaturés et chargés dans une cellule contenant des oligonucléotides complémentaires à ceux fixés sur les fragments d'ADN. La cellule est placée dans la cassette fluide du séquenceur. La molécule d'ADN à séquencer ainsi fixée est puis amplifiée par une étape de PCR bridge. Le brin reverse sera alors clivé et enlevé. Enfin, le brin restant (forward) est mis en présence des quatre types de nucléotides, chacun associé à un fluorophore différent et à un terminateur réversible. Le nucléotide à incorporer pour l'élongation du brin complémentaire est alors inséré. Un signal lumineux spécifique du nucléotide présent est alors émis et détecté. À la fin du processus de synthèse le brin reverse servira alors de matrice afin de générer les lectures dites paires. Une étape d'imagerie est effectuée après chaque cycle ce qui permet de détecter dans l'ordre les nucléotides incorporés. En associant une séquence d'index aux échantillons, il est possible de séquencer simultanément plusieurs échantillons. Cette pratique, appelée multiplexage, permet de réduire les coûts en divisant le nombre d'expériences de séquençage.

Cette technique a été utilisée afin de séquencer les ADN génomiques des souches de *L. kluveri*, des pools de résistants et des souches utilisées dans les expériences d'évolution. Les lectures paires obtenues font 100 pb. Les génomes ont été séquencés par multiplexage de 10 à 25 génomes selon les projets. La profondeur de séquençage obtenue dépend notamment du multiplexage, elle varie entre 50x et 150x. Les ADN de pools de ségréants résistants et les ADN génomiques des souches de *L. kluveri* ont été séquencés à l'EMBL (European Molecular Biology Laboratory) tandis que les génomes des souches évoluées ont été séquencés au BGI (Beijing Genomics Institute).

### 7.2. Séquençage des fragments PCR

Les produits PCR précipités sont placés dans des tubes PCR de 1,5 ml et dilués de manière à ce qu'il y ait une quantité de 200 ng dans 10 µl. Les amorces du couple ayant permis l'amplification du fragment sont ajoutées dans des tubes séparés à une concentration de 0,5 µM. Les séquençages des fragments PCR sont effectués au service séquençage d'eurofins par la méthode Sanger (<https://ecom.mwgdna.com/services/home.tcl>).

## 8. Techniques de transformation des levures

### 8.1. Transformation par électroporation

#### 8.1.1. Protocole pour *S. cerevisiae*

##### 8.1.1.1. Préparation des cellules électro-compétentes

Les cellules sont cultivées dans 40 ml de milieu complet jusqu'en fin de phase exponentielle (0,8 à 1,2 d'absorbance 600 nm) et sont placées 10 min dans la glace. Elles sont centrifugées 5 min à 5000 rpm à 4°C, puis lavées avec 20 ml de sorbitol 1 M. Elles sont à nouveau centrifugées 5 min à 6000 rpm à 4°C et rincées deux fois avec 20 ml puis 5 ml de sorbitol 1 M. Le culot est alors repris dans 200 µl de sorbitol 1 M.

##### 8.1.1.2. Électroporation

Un volume de 50 µl de cellules compétentes est déposé dans une cuve d'électroporation auxquels sont ajoutés 1 µg d'ADN nécessaire pour la transformation. Les cellules sont ensuite soumises à un choc électrique (1600 V, 10 µF, 600 W) pour laisser pénétrer les fragments d'ADN. Immédiatement après l'électroporation, 1 ml de milieu YPD Sorbitol 1 M est ajouté à la cuve. La solution est transférée dans un nouveau tube et les cellules sont régénérées environ 1 heure à 30°C. Suite à cette incubation, les cellules sont brièvement centrifugées puis reprises dans 200 µl de milieu YPD Sorbitol 1 M et 100 µl sont étalés sur boîte de milieu sélectif. Les boîtes sont incubées à 30°C et les colonies apparaissent après 3 à 5 jours de culture.

#### 8.1.2. Protocole pour *L. kluyveri* (Gojkovic *et al.* 2000)

##### 8.1.2.1. Préparation des cellules électro-compétentes

Les cellules sont cultivées dans 100 ml de milieu complet jusqu'en fin de phase exponentielle (0,8 à 1,5 d'absorbance 600 nm). La culture est alors centrifugée 10 min à 5000 g. Les cellules sont resuspendues dans 25 ml de tampon phosphate de potassium à 50 mM (pH de 7,5) contenant 25 mM de DTT (dithiothreitol). Les cellules sont incubées à 37°C pendant 30 min, puis lavées deux fois avec 50 ml de tampon STM froid (sucrose 270 mM, Tris-HCl (pH de 7,5) 10 mM et du MgCl<sub>2</sub> à 1 mM). Le tout est centrifugé à 5000 g pendant 10 min à 4°C. Une fois tout le surnageant soigneusement retiré, les cellules sont reprises dans 0,1 ml de tampon STM froid puis conservées dans la glace.

##### 8.1.2.2. Électroporation

Une quantité de 100 ng de plasmides concentrés dans 1 µl sont ajoutés à 50 µl de cellules en suspension. Les cellules sont ensuite soumises à un choc électrique (1600 V, 10 µF, 600 W) pour laisser pénétrer les fragments d'ADN. Immédiatement après l'électroporation, 1 ml de milieu YPD est ajouté à la cuve. La solution est centrifugée brièvement à 10 000 rpm quelques secondes. Le

surnageant est retiré et les cellules sont reprises dans 500 µl de milieu YPD et 100 µl sont étalés sur boîte de milieu sélectif. Les boîtes sont incubées à 30°C et les colonies apparaissent après 3 à 5 jours de culture.

## 8.2. Transformation par choc thermique avec acétate de lithium

Cette technique est utilisée pour la transformation des souches de *S. cerevisiae*.

### 8.2.1. Préparation des cellules compétentes

Environ  $10^7$  cellules en fin de phase exponentielle, dont la croissance a été réalisée en milieu YPD, sont centrifugées 10 min à 4000 rpm puis lavées avec 10 ml de tampon TE (TrisHCl pH 8 10 mM, EDTA 1 mM). Elles sont culotées pendant 10 min à 4000 rpm puis rincées avec 10 ml de solution LiOAc (LiOAc 0,1 M, TrisHCl pH 8 10 mM, EDTA 1mM) et resuspendues dans 1 ml de cette solution. 100 µl de ces cellules sont mises en contact avec 10 µl d'ADN transformant, 10 µl d'ADN entraîneur (ici de l'ADN de sperme de saumon dénaturé) et 1 ml de PEG 3350 40% stérilisé par filtration.

### 8.2.2. Choc thermique

Les cellules sont incubées 1 h à 30°C puis 40 min à 42°C. Elles sont récoltées par centrifugation de 2 min à 12400 rpm, resuspendues dans 100 µl de milieu YPD puis étalées sur milieu sélectif.

## 9. Bioinformatique

### 9.1. Les banques de données

Les banques de données ont permis de récupérer les séquences nucléiques et protéiques des gènes d'intérêt, de réaliser des recherches Blast, d'avoir une représentation d'une région chromosomique, et de chercher la fonction des gènes d'intérêt. Les banques utilisées lors de ce travail sont les suivantes :

- Saccharomyces Genome Database (SGD) : <http://www.yeastgenome.org/>
- Génolevures : <http://genolevures.org/>
- European Bioinformatic Institute (EBI) : <http://www.ebi.ac.uk/>
- National Center for Biotechnology Information (NCBI) : <http://www.ncbi.nlm.nih.gov/>
- SPGP Blast server : <http://www.moseslab.csb.utoronto.ca/spgp/>
- Yeast SNPs Browser : [http://gbrowse.princeton.edu/cgi-bin/gbrowse/yeast\\_strains\\_snps/](http://gbrowse.princeton.edu/cgi-bin/gbrowse/yeast_strains_snps/)

## **9.2. Analyse des données de séquençage de l'ADN génomique**

Les parties 9.2.1 à 9.2.4 ont été réalisées par les bioinformaticiens de l'équipe.

### **9.2.1. Nettoyage des séquences**

Les lectures issues du séquençage des ADN génomiques sont nettoyées grâce au logiciel FASTQ Quality Trimmer. Il va permettre de supprimer les nucléotides de mauvaise qualité à la fin des séquences et d'éliminer les lectures trop courtes (inférieure à 65 pb).

### **9.2.2. Alignement des séquences**

Les séquences sont alignées contre un génome de référence avec le logiciel BWA (Li and Durbin 2009).

Le logiciel BWA permet de traiter massivement les lectures générées par les outils de séquençage à haut débit et contrairement à d'autres algorithmes d'alignement, il permet de gérer de manière flexible les insertions et les délétions en intégrant des gaps et en autorisant les mésappariements.

### **9.2.3. Détermination des SNP**

Le logiciel BWA renvoie en sortie un fichier d'alignement qui constituera le fichier d'entrée pour le logiciel SAMtools (Li *et al.* 2009). Cet outil va permettre de générer un fichier tabulé qui présente les informations relatives à chaque position séquencée en 10 colonnes. Les colonnes nous informent sur : le chromosome, la position de la base, la base du génome de référence à cette position, le consensus des lectures alignées à cette position, la qualité du consensus, le score estimant la probabilité que cette position soit un SNP, la qualité maximale de l'alignement, les bases alignées ainsi que la qualité associée à chacune d'entre elles. Ce fichier nous informe donc sur les éventuelles positions polymorphes par rapport à la souche de référence. Nous avons ainsi pu les extraire.

### **9.2.4. Analyse des couvertures**

Dans le cadre du séquençage, on appelle couverture (ou profondeur) de séquençage le rapport entre la longueur de l'ensemble des séquences lues mises bout à bout et la taille du génome séquencé. Une grande profondeur de séquençage indique que chaque position du génome sera couverte par un grand nombre de lectures. Plus la couverture est élevée, plus la séquence consensus que l'on en dégagera sera fiable, sauf dans le cas particulier d'erreurs d'alignement dues à des séquences répétées ou de faible complexité. Afin de représenter la couverture le long des génomes séquencés, la couverture moyenne sur des fenêtres de 1 kb a été calculée.

### 9.2.5. Représentations graphiques des données de séquençage grâce au logiciel R

Les graphiques issus des données de séquençage ont été réalisés grâce au langage de programmation R qui propose de nombreuses fonctions permettant de produire des graphiques clairs. (<http://www.r-project.org/>). Les représentations de l'analyse des couvertures et des fréquences alléliques dans l'analyse des pools de ségréants ont été réalisées grâce à l'extension « ggplot2 » de R.

### 9.3. Alignement et assemblage des séquences issues du séquençage de fragments PCR

Des régions génomiques d'intérêt (*PDR5*, promoteur de *PDR5*) ainsi que les jonctions des chromosomes chimères ont été séquencées en plusieurs fois, leur taille étant trop grande pour qu'ils soient séquencés en une fois. Des fragments PCR de 700 pb recouvrant ces gènes ont donc été séquencés. Ils ont ensuite été alignés grâce aux logiciels Clustalw2 (<http://www.ebi.ac.uk/Tools/clustalw2/>) et Jalview (Clamp *et al.* 2004). L'assemblage entre les différentes séquences a été réalisé manuellement.

### 9.4. Construction d'arbres phylogénétiques

Le logiciel Seaview permet la construction d'arbres à partir des alignements générés par Clustalw2 (Gouy *et al.*, 2010). L'approche *Neighbor-joining* a été utilisée afin de regrouper les séquences sous forme d'arbre. La correction de la distance pour les séquences nucléiques est la méthode de *Jukes-Cantor*. En ce qui concerne les séquences protéiques, la correction des distances est réalisée par la méthode de *Poisson*. Les délétions sont ignorées lors de la comparaison de deux séquences entre elles grâce à l'option *pairwise gap removal*. La robustesse est indiquée au niveau de chaque nœud. Le nombre d'alignements permettant de la calculer a été fixé à 1000. La représentation choisie est celle du cladogramme où la longueur des branches est non significative.

### 9.5. La représentation graphique des données de résistance

Les données de résistance aux composés azolés chez les souches de *S. cerevisiae* et *L. kluyveri* ont été représentées par des graphiques de type « heat map ». Ces graphiques ont été construits grâce au logiciel MeV (Multi Experiment Viewer : <http://www.tm4.org/mev.html>). Le calcul de la corrélation de Pearson a permis la hiérarchisation par profils phénotypiques.

# Chapitre 1 :

## Les bases génétiques de la résistance à l'ammonium :

### un cas mendélien

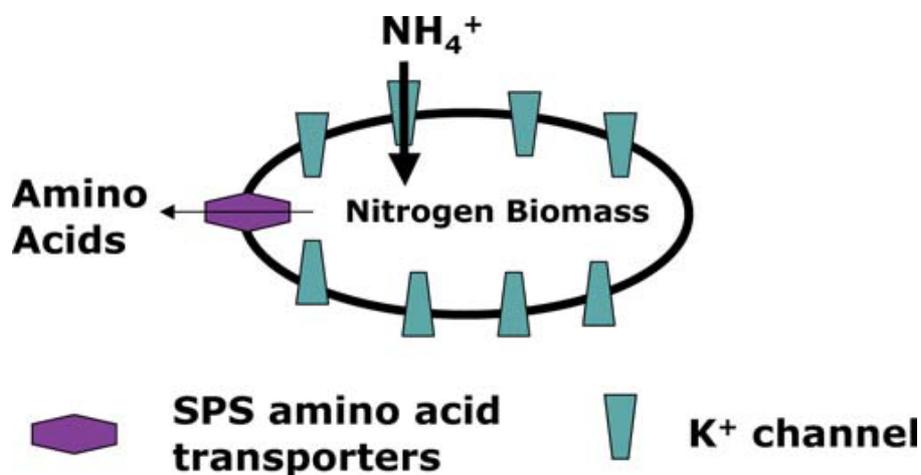
#### 1. Introduction

L'ammonium ( $\text{NH}_4^+$ ) est la source d'azote la plus utilisée chez les bactéries et les champignons. Chez les plantes et les animaux, l'excès d'ammonium est toxique pour la cellule (von Wiren and Merrick, 2004). En effet, lorsqu'une source d'azote entre dans la cellule, elle est souvent transformée en ammonium, avant d'être assimilée dans une voie métabolique. Il y a donc une balance entre le besoin d'ammonium et sa toxicité. Chez les plantes, il a été montré que l'ion ammonium pouvait entrer par les canaux potassiques lorsque l'ion potassium ( $\text{K}^+$ ) est limitant (Moroni et al., 1998). L'ion potassium, est l'ion le plus abondant dans la majorité des cellules. Sa concentration intracellulaire dépasse souvent 100 fois celle du milieu extérieur à la cellule (Rodríguez-Navarro, 2000).

Une étude a démontrée que l'ammonium est également toxique chez la levure *S. cerevisiae*, plus particulièrement chez les souches de laboratoires et ce, dans des conditions où la concentration en ions potassium est faible (Hess et al., 2006). Les auteurs suggèrent que l'ammonium pourrait entrer dans la cellule par les canaux potassiques comme chez les plantes. L'ammonium intracellulaire est utilisé pour générer de la biomasse mais, en cas de concentration trop élevée, il devient toxique. Une analyse transcriptomique réalisée à partir de cultures de *S. cerevisiae* en chemostats avec de l'ammonium comme seule source d'azote et en faible concentration de potassium (1,3 mM) a été réalisée afin de mesurer l'impact de la toxicité à l'ammonium sur l'expression des gènes. Dans ces conditions, peu de gènes se sont trouvés différenciellement exprimés. Cependant, les gènes codant les perméases de différents acides aminés sont surexprimés de 5 à 10 fois par rapport à des conditions de culture en concentration plus élevée de potassium (13 mM).

À partir de ces résultats, l'hypothèse émise pour expliquer la détoxification de l'azote en excès est la suivante : la cellule utiliserait l'azote en excès pour la synthèse de composés organiques qui seraient ensuite expulsés hors de la cellule. Pour détecter ces composés organiques dans le milieu extracellulaire, une analyse de spectrométrie de masse en tandem a été réalisée. Une forte excrétion de tous les acides aminés excepté l'arginine est détectée dans des conditions où l'ammonium est toxique. L'étude propose donc un modèle de « défense » des levures par une détoxification de l'ammonium en excès par l'utilisation de l'azote dans la production d'acides aminés qui seront exportés hors de la cellule (Figure 1.1).

Afin de tester si la toxicité et le mécanisme de détoxification de l'ammonium sont spécifiques à l'espèce ou encore au fond génétique de la souche utilisée (S288c), des cultures en chemostats en présence d'ammonium, en condition toxique ou non, ont été réalisées avec deux autres souches de *S. cerevisiae* (CEN.PK et  $\Sigma$ 1278b) ainsi qu'avec une souche de *Saccharomyces bayanus*. Dans des conditions de toxicité en ammonium, le nombre de cellules en phase stationnaire baisse de façon importante pour toutes les souches. De plus, la sécrétion d'acides aminés augmente très fortement dans toutes les souches testées. Ces résultats montrent donc que la toxicité à l'ammonium et sa détoxification via la synthèse et l'excrétion d'acides aminés ne sont pas des mécanismes spécifiques au fond génétique de la souche S288c, ni à l'espèce *S. cerevisiae*.



**Figure 1.1** : Modèle de la toxicité à l'ammonium et de la détoxification chez la levure *S. cerevisiae* (Hess et al., 2006). Si l'ammonium est présent en forte concentration dans le milieu, les ions  $\text{NH}_4^+$  pourront entrer dans la cellule par les canaux potassiques. La majorité de l'ammonium sera utilisée pour générer de la biomasse (si les autres nutriments sont également disponibles) mais le flux entrant non régulé peut aboutir à un excès d'ammonium dans la cellule qui deviendra, de par sa forte concentration, toxique. Afin de réduire sa concentration intracellulaire, l'azote sera utilisé dans la biosynthèse des acides aminés qui seront ensuite expulsés hors de la cellule via des perméases spécifiques aux acides aminés. L'azote qui sera fixé dans les acides aminés ne pourra alors plus entrer à nouveau dans la cellule.

L'ammonium peut être présent en forte concentration dans les milieux contenant de la matière organique en décomposition, comme par exemple les milieux pour la fermentation alcoolique. *S. cerevisiae* est largement utilisée ce type de fermentation et, selon les procédés employés à travers le monde, les souches sont adaptées à utiliser différents substrats dans diverses conditions. Dans l'étude précédente, il a été prouvé que la toxicité à l'ammonium n'était pas dépendante du fond génétique de la souche. Cependant, les deux souches de *S. cerevisiae* qui avaient été testées en plus de la souche de référence étaient des souches de laboratoire. Or, certains isolats utilisés dans la fermentation alcoolique ont pu développer une résistance à l'ammonium, notamment du fait de leur environnement. Dans l'article qui suit, la diversité naturelle présente au sein des souches de *S. cerevisiae* a été mise à

profit afin d'explorer la diversité de résistance à l'ammonium. Cette étude a permis de déterminer l'origine génétique de la résistance à l'ammonium chez l'espèce *S. cerevisiae*. Les principaux résultats de cette étude sont présentés ci-dessous.

La résistance d'une collection de 63 souches haploïdes *MATa* a tout d'abord été testée (Schacherer et al., 2009). Les souches ont été isolées d'origines écologiques (nature, vin, saké, boulangerie, malades infectés, laboratoire et distillerie) et géographiques (Europe, Asie, Amérique) variées. La souche K12, isolée à partir du saké, est la seule capable de pousser à de fortes concentrations en  $\text{NH}_4^+$  (456 mM). Elle a été croisée avec la souche de référence S288c qui est sensible au  $\text{NH}_4^+$ . Des tétrades issues du croisement ont été disséquées et la résistance à l'ammonium a été testée pour un certain nombre d'entre elles. Dans chacun des asques, on trouve 2 spores résistantes et deux spores sensibles. Cette ségrégation 2 :2 signifie que dans ce cas, un seul gène est impliqué dans la résistance. Une analyse de pool de ségrégants a révélé la région du génome impliquée dans la résistance chez la souche K12. Une analyse fonctionnelle des gènes de cette région a mis en évidence le gène *TRK1* codant un transporteur potassique ce qui en fait un bon candidat pour expliquer la résistance de la souche K12 au  $\text{NH}_4^+$ . L'implication du gène *TRK1* dans la résistance a été validée par un test d'hémizygotie réciproque. Sa séquence a été comparée à celles de 22 souches de *S. cerevisiae* d'origines écologiques variées dont la souche K1 également isolée du saké mais qui est contrairement à K12 sensible aux fortes concentrations de  $\text{NH}_4^+$ , mais également à 3 autres espèces d'Hémiascomycètes : *S. paradoxus*, *S. bayanus* et *S. mikatae*. Quatre changements d'acides aminés sont spécifiques à la souche K12. Au niveau du gène, deux de ces mutations ont lieu dans des régions hautement conservées ce qui suggère qu'une mutation dans celles-ci a une forte probabilité d'avoir un impact sur la conformation et/ou la fonction de la protéine. Un test de croissance sur « gradient » de potassium de la souche résistante K12, de la souche sensible FY4 et des mutants dans lesquels le gène *TRK1* a été délété, K12 *trk1Δ0* et FY4 *trk1Δ0* indique que la souche K12 pousse sur de faibles concentrations en  $\text{K}^+$  contrairement aux autres souches. Ces résultats suggèrent que le gène *TRK1* de la souche K12 coderait pour un transporteur ayant une meilleure affinité pour l'ion potassium.



## Genetic Basis of Ammonium Toxicity Resistance in a Sake Strain of Yeast: A Mendelian Case

Cyrielle Reisser,<sup>\*1</sup> Cynthia Dick,<sup>†1</sup> Leonid Kruglyak,<sup>\*§</sup> David Botstein,<sup>\*\*</sup> Joseph Schacherer,<sup>\*2,3</sup> and David C. Hess<sup>†,2,3</sup>

<sup>\*</sup>Department of Genetics, Genomics and Microbiology, University of Strasbourg, CNRS, UMR7156, Strasbourg, 67083 France, <sup>†</sup>Department of Biology, Santa Clara University, Santa Clara, California 95053, and <sup>‡</sup>Lewis-Sigler Institute for Integrative Genomics and Department of Ecology and Evolutionary Biology, <sup>§</sup>Howard Hughes Medical Institute, and <sup>\*\*</sup>Lewis-Sigler Institute for Integrative Genomics and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544

**ABSTRACT** High concentrations of ammonium at physiological concentrations of potassium are toxic for the standard laboratory strain of *Saccharomyces cerevisiae*. In the original description of this metabolic phenotype, we focused on the standard laboratory strains of *Saccharomyces*. In this study, we screened a large collection of *S. cerevisiae* natural isolates and identified one strain that is resistant to high concentrations of ammonium. This strain, K12, was isolated in sake breweries. When the K12 strain was crossed to the standard laboratory strain (FY4), the resulting tetrads displayed 2:2 segregation of the resistance phenotype, suggesting a single gene trait. Using a bulk segregant analysis strategy, we mapped this trait to a 150-kb region on chromosome X containing the *TRK1* gene. This gene encodes a transporter required for high-affinity potassium transport in *S. cerevisiae*. Data from reciprocal hemizyosity experiments with *TRK1* deletion strains in K12 and BY backgrounds, as well as analysis of the deletion of this gene in the K12 strain, demonstrate that the K12 allele of *TRK1* is responsible for ammonium toxicity resistance. Furthermore, we determined the minimal amount of potassium required for both the K12 and laboratory strain needed for growth. These results demonstrate that the gene encoded by the K12 allele of *TRK1* has a greater affinity for potassium than the standard allele of *TRK1* found in *Saccharomyces* strains. We hypothesize that this greater-affinity allele of the potassium transporter reduces the flux of ammonium into the yeast cells under conditions of ammonium toxicity. These findings further refine our understanding of ammonium toxicity in yeast and provide an example of using natural variation to understand cellular processes.

Single-celled organisms must deal with drastic environmental changes that result in several challenges to internal homeostasis. These external factors include nutrient starvation, temperature changes, osmolarity differences, pH differences, exposure to toxins, and changes in ionic

concentrations. The toxicity of ammonium is well established in plant and animal systems (von Wiren and Merrick 2004), but it was not until recently that ammonium toxicity was discovered in *S. cerevisiae* (Hess *et al.* 2006). One aspect that distinguishes ammonium toxicity from other stresses in yeast is that the transcriptional response to the stress is very limited and does not include the environmental stress response signature (Will *et al.* 2010). Furthermore, ammonium toxicity in *Saccharomyces* is proposed to involve the clash between the need for potassium uptake by the cells and the need to exclude excess ammonium ions. Yeast cells can detoxify excess ammonium by fixing the ammonium to carbon skeletons in the form of amino acids and excreting the amino acids into the extracellular space, similar to how mammals detoxify ammonium through the excretion of urea (Hess *et al.* 2006). All of the metabolic pathways involved in ammonium toxicity in yeast and how they interact are not fully characterized.

Copyright © 2013 Reisser *et al.*

doi: 10.1534/g3.113.005884

Manuscript received November 23, 2012; accepted for publication February 22, 2013

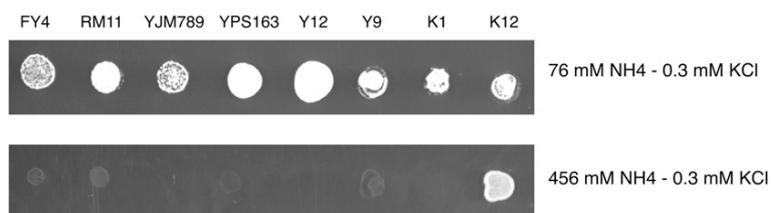
This is an open-access article distributed under the terms of the Creative Commons Attribution Unported License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supporting information is available online at <http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.113.005884/-/DC1>

<sup>1</sup>These authors contributed equally to this work.

<sup>2</sup>These authors contributed equally to this work.

<sup>3</sup>Corresponding authors: University of Strasbourg/CNRS28, rue Goethe, Strasbourg, 67083 France. E-mail: schacherer@unistra.fr; and hess.scu@gmail.com



**Figure 1** The K12 strain, resistant to high concentration of ammonium. Cells were grown on plate with high (456 mM  $\text{NH}_4$ ) and low (76 mM  $\text{NH}_4$ ) concentrations of ammonium with low (300  $\mu\text{M}$ ) potassium.

Hess *et al.* (2006) established that ammonium toxicity in yeast occurs at physiological levels of potassium and high ammonium. In this study we proposed that ammonium ions and potassium ions (both positively charged ions with nearly identical atomic radii) can both be transported by potassium transporters. At physiological levels of potassium, the high concentration of potassium transporters in the cell membrane creates a toxic “leak current” of ammonium into the cell. We observed that yeast cells can detoxify this excess ammonium by increasing the production of amino acids and excreting them into the media (Hess *et al.* 2006). However, if the ammonium concentration is sufficiently high under physiological concentrations of potassium, then the media is toxic to yeast. This link between ammonium toxicity and potassium transport is crucial to the findings in this article.

High levels of ammonium are found in decaying organic matter such as the substrates used in ethanol fermentations. Because fermentation methods and substrates vary across the world, we wanted to examine a collection of yeast strains adapted to these different fermentation processes and test for resistance to ammonium toxicity. Natural variation in yeast has been successfully used to study non-ammonium toxicity stress responses (Will *et al.* 2010), differential use of metabolites (Wenger *et al.* 2010), and the differential ability to restart stuck wine fermentations (Marullo *et al.* 2007).

We screened a collection of 63 *S. cerevisiae* strains for resistance to ammonium toxicity. We identified one strain, K12, previously used for sake fermentation, as being highly resistant to ammonium toxicity. Using bulk-segregant analysis, we identified a region of the genome linked to the casual mutation conferring resistance to ammonium toxicity. Sequence analysis of the region revealed one gene, *TRK1*, with multiple nonconservative mutations. *TRK1* encodes a subunit of the potassium ion transporter and based on its function was an excellent candidate for involvement in ammonium toxicity. Deletion of the *TRK1* gene in the K12 strain background creates a strain fully sensitive to ammonium toxicity. Furthermore, the K12 allele of *TRK1* displays greater affinity for potassium than the standard laboratory allele. We conclude that the K12-allele of *TRK1* confers resistance to ammonium toxicity and provides validation of the proposed mechanism of ammonium toxicity previously suggested. These results are of further interest because they represent rare cases of a single locus being responsible for a large phenotypic effect (resistance to ammonium toxicity) that evolved naturally.

## MATERIAL AND METHODS

### Strains

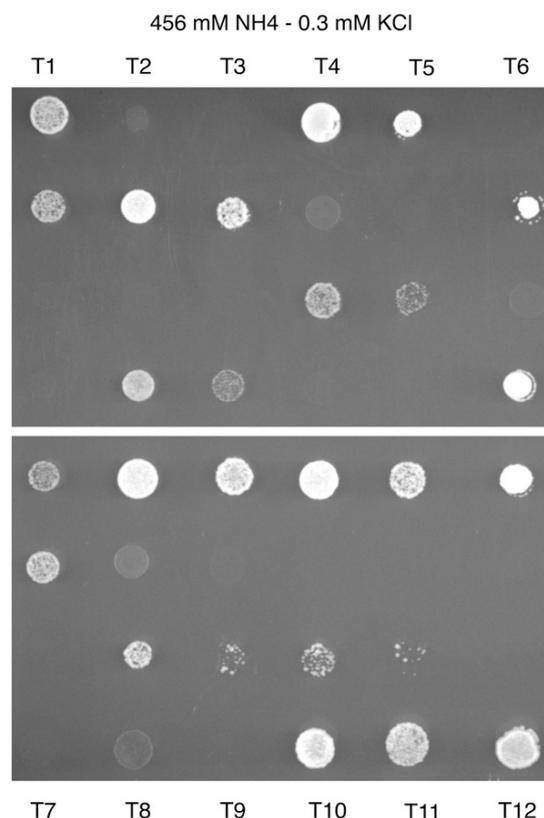
The *S. cerevisiae* strains used in this study are listed in Supporting Information, Table S1. FY4 (*MAT $\alpha$* ) was crossed with K12 (*MAT $\alpha$* ) on yeast peptone dextrose agar (YPD) (20 g/L bacto-peptones, 10 g/L yeast extract, 20 g/L glucose, 20 g/L agar) media. Diploid hybrids were sporulated for 2–5 d at 30° on 1% potassium acetate media. A total of 25 tetrads containing the four haploid meiotic products were dissected with a Singer instrument MSM.

### Media, growth conditions, and growth quantification

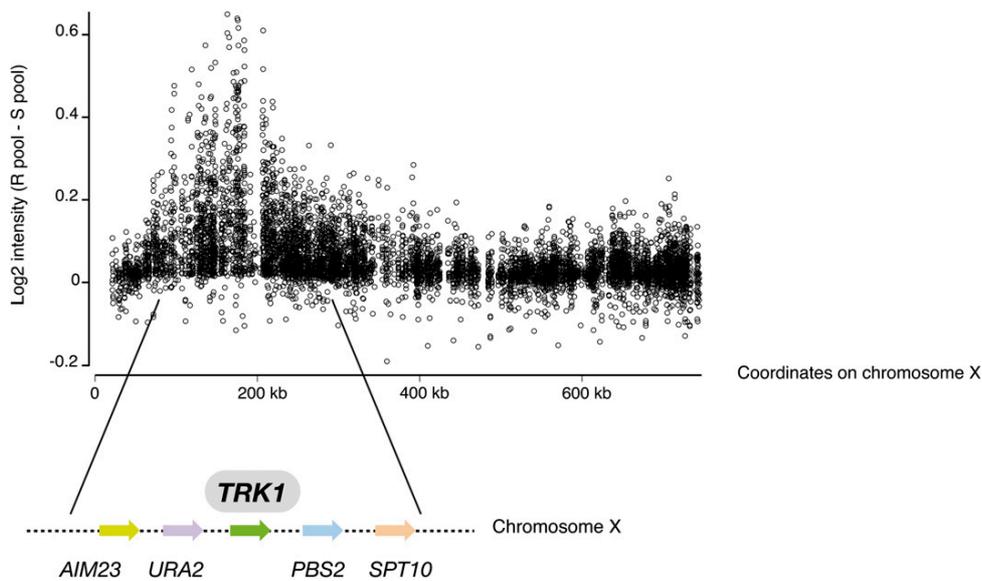
We adapted the phosphate limited medium from Hess *et al.* (2006) in this study. This base salt-mixed medium contained: 0.1 g/L calcium chloride dihydrate, 0.1 g/L sodium chloride, 0.5 g/L magnesium sulfate heptahydrate, 5 g/L ammonium sulfate, 0.05 g/L potassium chloride, 1 g/L sodium phosphate, and 2 g/L glucose, plus vitamins described previously. A total of 20 g/L of ultrapure agarose was added for the solid medium. Variations from this base medium are noted in the text and figure legends. The base ammonium toxicity medium contained 300  $\mu\text{M}$  potassium and 456 mM ammonium.

### Testing of the limiting concentration of potassium needed for growth

The minimum amount of potassium required to grow was tested at concentrations of 300, 150, 75, or 30  $\mu\text{M}$  KCl. Potassium plates contained 76 mM glutamate as the nitrogen source, 12 g/L ultrapure agarose, and salts, vitamins, and metals as described previously.



**Figure 2** Segregation of ammonium resistance in the cross between K12 and FY4. Segregants were grown on plate with high concentration of ammonium (456 mM  $\text{NH}_4$ ). All tetrads show a 2:2 segregation of growth and no growth.



**Figure 3** Bulk segregant analysis identifies *TRK1*. Genomic DNAs from pools of sensitive and resistant segregants were hybridized independently to Affymetrix yeast tiling arrays. Plotted here is the  $\log_2$  ratio of probe intensities (sensitive pool/resistant pool) along the chromosome X.

Ten-fold serial dilutions of K12, FY4, and two K12 *trk1Δ*-null allele strains were made from  $OD_{600} = 0.8-0.00008$ . Strains grew at 30° for 4 d before they were photographed.

#### Bulk segregant analysis

The ammonium resistant K12 strain was crossed with the FY4 reference strain, and the resulting diploid was sporulated. Segregants were scored for ammonium resistance on plates. In total, 30 sensitive and 30 resistant spores were selected and grown individually in 5 mL of YPD for 12–14 hr at 30°. Samples were pooled by phenotype (resistant and sensitive pools) and genomic DNA was extracted from 30 mL of YPD cultures using QIAGEN Genomic-Tips 100/G and Genomic DNA Buffers as per the manufacturer's instructions. DNA was digested with DNaseI, labeled, and hybridized to Affymetrix yeast tiling arrays as described in Gresham *et al.* (2006). In summary, the  $\log_2$  ratio of probe intensities (sensitive pool/ resistant pool) was plotted across every chromosome for each informative nucleotide to map the region involved in the resistance. The plots of each chromosome were scanned visually for local peaks in intensity.

#### Deletion construction

The disruption of the *TRK1* gene in the K12 strain was performed by homologous recombination. In summary, the *TRK1::KanMX* cassette was amplified from the corresponding deletion mutant (Winzeler *et al.* 1999). Approximately 400 bp up and downstream of the deleted *TRK1* gene were amplified with the cassette. The fragment was amplified using the primers TRK1.1 (5'-CATCAGAAATGTACG TAGGCA-3') and TRK1.2 (5'-CCATTAGCATCACTGATTCCA-3') and the iProof High-Fidelity DNA Polymerase from MP Biomedicals as described by the manufacturer. The resulting cassette was integrated into K12 by lithium acetate transformation and stable G418-resistant transformants were selected. Correct integration was confirmed by polymerase chain reaction.

#### Sequence analysis of the *TRK1* gene of different *S. cerevisiae* isolates

Protein sequences of the *TRK1* gene from various *S. cerevisiae* strains were retrieved from the SGD database and from the BLAST server of

the Fay Lab (<http://www.moseslab.csb.utoronto.ca/spgp/>). Because the sequence of the K12 strain was not available, we fully sequenced its *TRK1* gene. Sequence alignments were performed using CLUSTALW2 (Larkin *et al.* 2007) program with default parameters. The Neighbor-Joining tree was constructed using SplitsTree program (Huson and Bryant 2006).

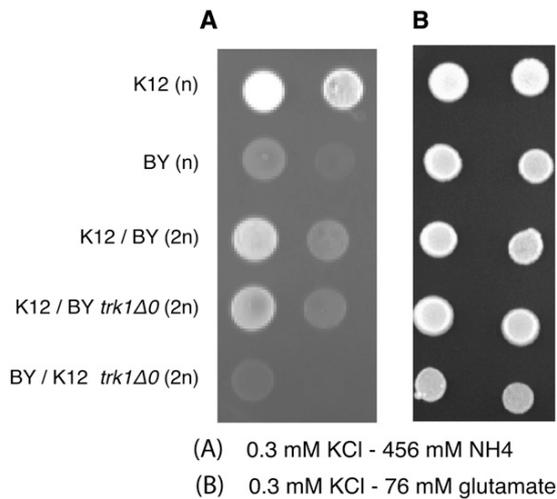
## RESULTS

#### Screening of natural isolates of *S. cerevisiae* for resistance to ammonium toxicity

At physiological concentrations of potassium high concentrations of ammonium are toxic to the several standard laboratory strains of yeast, including FY4. Low potassium and high ammonium conditions often are present in decomposing vegetation. Furthermore, several different types of grains, fruits, and vegetables are used as substrates for fermentation by *S. cerevisiae*. Depending on the specific fermentation process, these substrates are in different stages of decomposition. Thus, we hypothesized that certain fermentation strains have been exposed to the conditions of ammonium toxicity.

To test this hypothesis, we screened each strain in our collection for resistance to high concentration of ammonium. The collection comprises *S. cerevisiae* strains from various sources, including brewing, baking, wine, laboratory, as well as clinical strains (Schacherer *et al.* 2009) (Table S1). We placed each strain on plates containing low and high concentrations of ammonium (76 mM and 456 mM  $NH_4^+$ ; (Figure 1 and Figure S1). Of the 63 strains tested, we only identified one strain, K12, that was capable of growing in the presence of a high concentration of ammonium. K12 was isolated in sake breweries (Fay and Benavides 2005). Sake strains produce ethanol as well as organic acids and amino acids in the sake mash that contribute to sake aroma and taste (Akao *et al.* 2011).

To understand the degree of genetic complexity of ammonium resistance, we mated K12, the strain resistant to ammonium toxicity, to the FY4 laboratory strain, which is unable to grow under conditions of ammonium toxicity. Spores from 25 tetrads were tested for their ammonium toxicity phenotypes (Figure 2). Among the four spores from each tetrad, two spores had a growth phenotype and two spores had a no growth phenotype. This classic 2:2 segregation pattern



**Figure 4** Reciprocal hemizygosity with *trk1Δ0* in K12 and BY backgrounds. Cells were grown on plate with either 456  $\mu$ M ammonium as a nitrogen source (A) with 300  $\mu$ M potassium or 76  $\mu$ M glutamate as a nitrogen (B) source.

indicates that a single Mendelian locus is responsible for ammonium resistance of the K12 strain.

#### Dissection of the genetic origin of the resistance: *TRK1*, a candidate

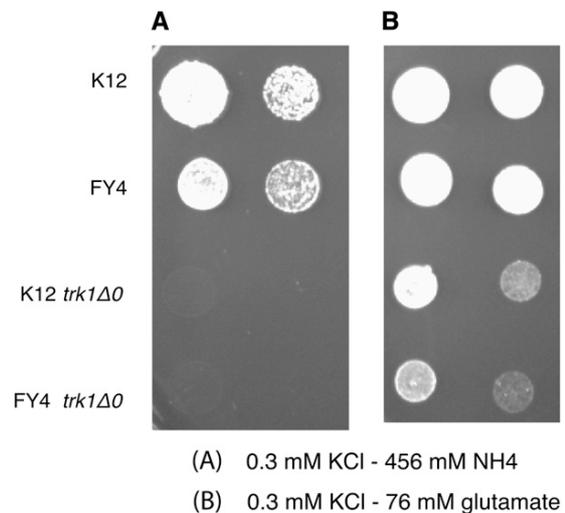
To determine the genomic location of the gene involved in ammonium resistance, we conducted a bulk segregant analysis (Dunham 2012). In summary, this strategy compares two pools of segregants coming from the K12/FY4 cross: one pool of spores with resistance to ammonium toxicity and a second group spores with sensitivity to ammonium toxicity. The DNA of each pool was genotyped for the thousands of single-nucleotide polymorphisms (SNPs) that differ between them by hybridization on a yeast tiling array. Comparison of these SNP genotypes revealed the region that is involved in the phenotypic variation. For Mendelian locus, these DNA pools should be equal mixtures of the K12 alleles and laboratory alleles across the whole genome except the region near the causal locus or loci. At these loci, the phenotypically “resistant” pool should consist only of the alleles from the K12 strain. We found one such region from the K12/FY4 cross.

Spores from the K12/FY4 diploid were screened for ammonium resistance and sensitivity. In total, 30 resistant spores were combined in a pool and 30 sensitive spores into another pool. Genomic DNA was extracted from each pool and then hybridized individually on a tiling array (based on the S288c reference genome). Based on previous results, we knew that the genetic divergence between K12 and FY4 is approximately 0.25%, representing a SNP density of 2.3 SNPs/kb (Schacherer *et al.* 2009). Because of this level of polymorphisms, we knew that the regions derived from the K12 strain would hybridize less efficiently to the array. After hybridization, we therefore plotted the  $\log_2$  ratios of probe intensities (sensitive pool/resistant pool; Figure 3). For most of the genome,  $\log_2$  ratios of probe intensities were around 0, showing that the allele frequency of K12 and FY4 is equivalent in both pools. Nevertheless, the bulk segregant analysis identified one linkage peak on chromosome X, which corresponds to less-efficient hybridization to the array of the resistant pool genomic DNA (Figure 3). The mapped region is approximately 150 kb in

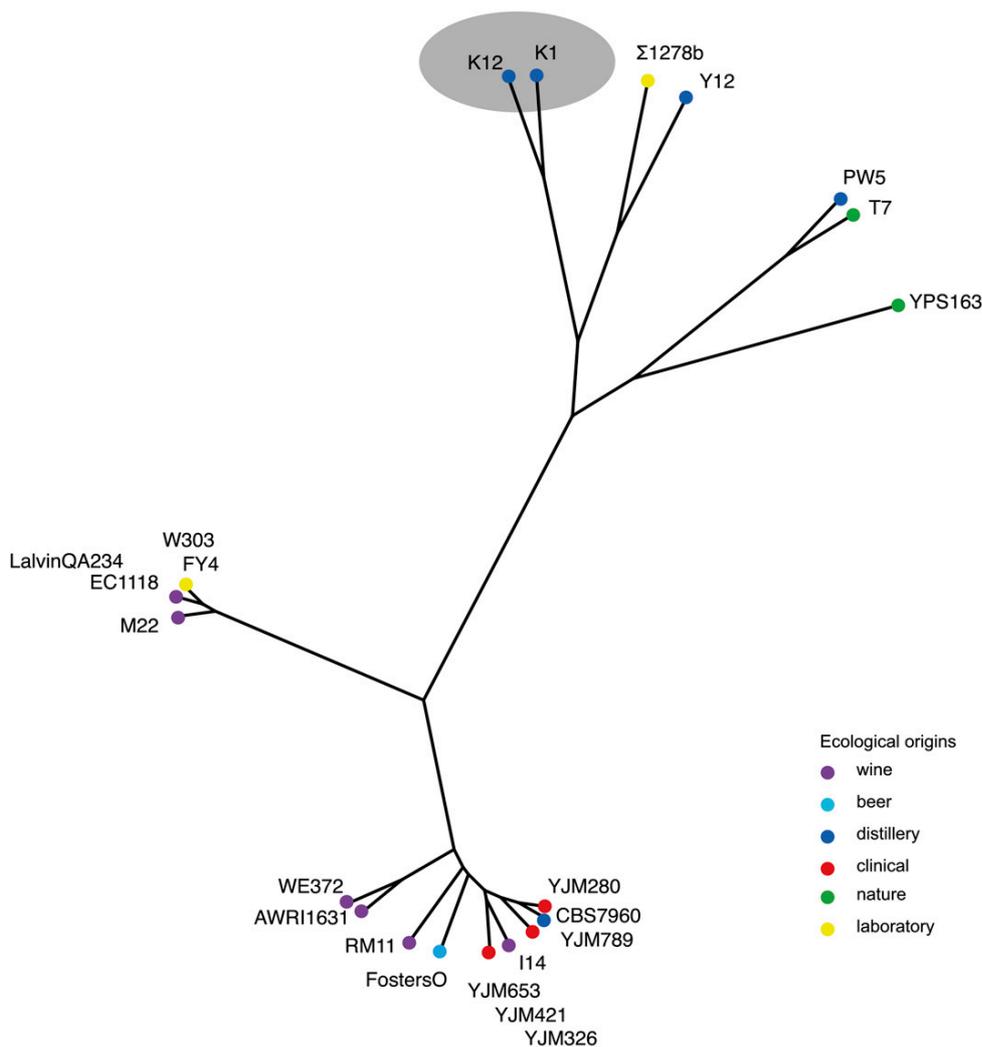
length and situated between open reading frames YJL171c and YJL095w. It encompasses more than 70 open reading frames. Interestingly, this region contains the *TRK1* gene (YJL139c). This gene is a good candidate for two reasons. First, *TRK1* is located in the middle of the linkage peak (Figure 3). Second, this gene encodes a transporter required for high-affinity potassium transport in *S. cerevisiae* (Gaber *et al.* 1988). Hess *et al.* (2006) hypothesized that ammonium toxicity in yeast occurs via an uncontrolled leak current of ammonium through the potassium channel, making *TRK1* an attractive candidate gene in which one might find allelic differences that cause, by themselves, resistance to ammonium toxicity.

#### Reciprocal hemizygosity experiments with the *trk1Δ0* strains

To demonstrate involvement of the *TRK1* allele of K12 for the phenotype, we created a deletion strain (*trk1Δ0*) in the K12 and FY4 genetic backgrounds as described in the Methods. Deletion of the high-affinity potassium transporter (*TRK1*) prevented these strains from growing at the low potassium levels used in the ammonium toxicity media (data not shown). These strains were nevertheless analyzed (see the section *Analysis of the trk1Δ0 strains for the ammonium toxicity phenotype and Discussion*), but another approach, reciprocal heterozygosity, was taken to test directly the relationship between the K12 *TRK1* allele and resistance to ammonium toxicity. To this end we generated three diploids. The first diploid was made simply by crossing K12 to BY, which we refer to as the control. The second diploid was generated by crossing the *trk1Δ0* in the BY background to K12. This strain only differed from the control in that the only functional allele of *TRK1* was the K12 allele. The third diploid was made by mating the *trk1Δ0* in the K12 background to BY, generating a strain that only differed from the control in that the only functional allele of *TRK1* only was the BY allele. These three diploid strains, as well as the haploid K12 and BY strains, were spotted on both ammonium toxicity medium and control medium (low potassium with glutamate as a nitrogen source; Figure 4). As previously shown, the K12 haploid strain grows well on ammonium toxicity medium whereas the BY haploid strain grows poorly. The K12/BY



**Figure 5** Phenotype of the K12 deletion mutant of *TRK1*. Cells were grown on plate with either 456  $\mu$ M ammonium as a nitrogen source (A) with 600  $\mu$ M potassium or 76  $\mu$ M glutamate as a nitrogen (B) source.



**Figure 6** Neighbor-joining tree of the *TRK1* genes of different *S. cerevisiae* isolates.

diploid strain and the K12/BY *trk1Δ0* diploid strain grow well on the ammonium toxicity medium. This finding shows that the K12 phenotype is dominant and that the BY *TRK1* allele is not required for the resistance phenotype. The important result is that the BY/K12 *trk1Δ0* diploid strain grows poorly on the ammonium toxicity medium, which shows that the presence of K12 *TRK1* allele is sufficient for the ammonium toxicity resistance.

#### Analysis of the *trk1Δ0* strains for the ammonium toxicity phenotype

Because *TRK1* is the high-affinity transporter for potassium, it was not surprising that *trk1Δ0* deletion strains, regardless of strain background, would not grow on our standard ammonium toxicity media: the potassium concentration was too low (300 μM) for the remaining systems for potassium assimilation. We therefore made plates with 600 μM potassium and either 76 μM glutamate as a nitrogen source or 450 μM ammonium as a nitrogen source. We spotted the K12 and FY4 wild-type strains along with the *trk1Δ0* in both strain backgrounds onto these plates (Figure 5). As expected, the wild-type strains of K12 and FY4 grew well on both plates as the potassium level is too high to induce ammonium toxicity. However, for the

*trk1Δ0* in both the K12 and BY backgrounds we observed the ammonium toxicity as predicted by our model of the role of *Trk1* in ammonium toxicity (see *Discussion*). Importantly, the *trk1Δ0* strains show equivalent ammonium toxicity in both strain backgrounds—the K12 strain no longer has resistance to ammonium toxicity in the absence of its unique *TRK1* allele (Figure 5). All these data allow us to conclude that *TRK1* is the single locus is responsible for ammonium resistance of the K12 strain.

#### Sequence analysis of *TRK1* gene reveals amino acid coding changes unique to the K12 strain

The relatedness of the *S. cerevisiae* strains was determined based on SNP mapping in all of these strains (Schacherer *et al.* 2009). This study showed the existence of a unique sake cluster that is distinct from wine and laboratory strains. Hence, we explored the genetic diversity of the *TRK1* gene and looked at the phylogeny based on the sequences of this gene. The protein sequences of the *TRK1* gene from 22 various *S. cerevisiae* strains, including K12, were retrieved from the different databases (see *Materials and Methods*). We also sequenced the *TRK1* gene of the K1 strain because K12 and K1 strains are closely related (Schacherer *et al.* 2009). DNA sequence variations

<b>K12</b>	1	MHFRRTMSRVPTLASLEIRYKKSFGHKFRDFIALCGHYFAPVKKYIFPFSFI AVHYFYTIS	60
<b>K1</b>		MHFRRTMSRVPTLASLEIRYKKSFGHKFRDFIALCGHYFAPVKKYIFPFSFI AVHYFYTIS	
<b>FY4</b>		MHFRRTMSRVPTLASLEIRYKKSFGHKFRDFIALCGHYFAPVKKYIFPFSFI AVHYFYTIS	
<b>Y12</b>		MHFRRTMSRVPTLASLEIRYKKSFGHKFRDFIALCGHYFAPVKKYIFPFSFI AVHYFYTIS	
<b>S. paradoxus</b>		MHIRGTMSRVPTLASLEVRKYKKSFGHKL RDFIALCGHYFSP LKKYVFPFSFI AVHYFYTIS	
<b>S. bayanus</b>		MHIRGTMSRVPTLASFEVRYKKSFGHKFRDFIALCGHYCSP IKKYIFPNFI AVHYFYTIV	
<b>S. mikatae</b>		MYFRRTMSRVPTLASLEVRKYKKTFGHKL RDFIALCGHYFAPVKKYVFPFSFI AVHYFYTIS	
...			
<b>K12</b>	61	LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVSCISTPIA	120
<b>K1</b>		LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVCCISTPIA	
<b>FY4</b>		LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVCCISTPIA	
<b>Y12</b>		LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVCCISTPIA	
<b>S. paradoxus</b>		LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVCCISTPIA	
<b>S. bayanus</b>		LTLITSILLYPVKNIRYIDALFLAAGAVTQGG LNTVDVNNLTLYQQI ILYI ICCISTPIA	
<b>S. mikatae</b>		LTLITSILLYPIKNTRYIDTLFLAAGAVTQGG LNTVDINNLSLYQQIVLYIVCCISTPIA	
...			
<b>K12</b>	541	EDERVQQSEPPSDSELKSHQQHQEKHQ LQQNLHRMYKTKSFDDNRSRAVPMERSRTIDMA	600
<b>K1</b>		EDERVQQSEPHSDSELKSHQQHQEKHQ LQQNLHRMYKTKSFDDNRSRAVPMERSRTIDMA	
<b>FY4</b>		EDERVQQSEPHSDSELKSHQQHQEKHQ LQQNLHRMYKTKSFDDNRSRAVPMERSRTIDMA	
<b>Y12</b>		EDERVQQSEPHSDSELKSHQQHQEKHQ LQQNLHRMYKTKSFDDNRSRAVPMERSRTIDMA	
<b>S. paradoxus</b>		EDERAQQPEPQFDSELK-----LHQHQ LQQNLHRMYKTKSFDDNRSRAVPMERSRTIDMA	
<b>S. bayanus</b>		EDDRMQQSEPPSGSELK LKQQ--QQHQ LQQNLHRMYKTKSFDDNRSK PVL MERSKTIDMA	
<b>S. mikatae</b>		EDERVQQSEPPSDSELKSHQQHQEQHQ FQQHLHRMYQT KSFDDNRSRAVPMQRSKTIDMA	
...			
<b>K12</b>	1141	NQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDHLEGMK LKRQART	1200
<b>K1</b>		NQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDHLEGMK LKRQART	
<b>FY4</b>		NQSF SRQFTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDHLEGMK LKRQART	
<b>Y12</b>		NQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHFDHLEGMK LKRQART	
<b>S. paradoxus</b>		NQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDHLEGMK LKRQART	
<b>S. bayanus</b>		NQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDHIEDL K LKRQART	
<b>S. mikatae</b>		DQSF SRQLTTLSKLVIIAM LIRGKNRGLPYSLDRAI I LPSDRLEHIDRLEGMK LKRQTRS	
...			
<b>K12</b>	1201	NTEDPITEHFKR SFIDVKHRWGAL KRKTTHSRNPKRSSTTL	1241
<b>K1</b>		NTEDPITEHFKR SFIDVKHRWGAL KRKTTHSRNPKRSSTTL	
<b>FY4</b>		NTEDPMTEHFKR SFIDVKHRWGAL KRKTTHSRNPKRSSTTL	
<b>Y12</b>		NTEDPMTEHFKR SFIDVKHRWGAL KRKTTHSRNPKRSSTTL	
<b>S. paradoxus</b>		NTEDPMTEHFKR SFIDVKHRWGAL KRRTTHSRNSKR SNSTTL	
<b>S. bayanus</b>		DTDDPMTEHLKRSISDAKHRWDELKHKRSLSRSSKRSTKTN	
<b>S. mikatae</b>		NTEDPMTEHFKR SFIDVKHRWGLKRKTTHSRNSKR NSTIV	

**Figure 7** Sequence alignment of *TRK1* genes of *S. cerevisiae* and other *Saccharomyces sensu stricto* isolates.

were examined, and a neighbor-joining tree was built (Figure 6). As observed at the genome level, the *TRK1* gene follows the same evolution with a close relationship of the strains isolated in sake breweries.

In addition, we compared more closely the protein sequences of K12, K1, Y12, and FY4 strains. Unlike the K12 strain, the K1, Y12, and FY4 strains are sensitive to high concentration of ammonium (Figure S1). Interestingly, comparison between these strains sequences revealed that four amino-acid changes are specific to the K12 strain: 1143 C → S, 551 H → P, 1190 E → G, and 1227 Q → K (Figure 7).

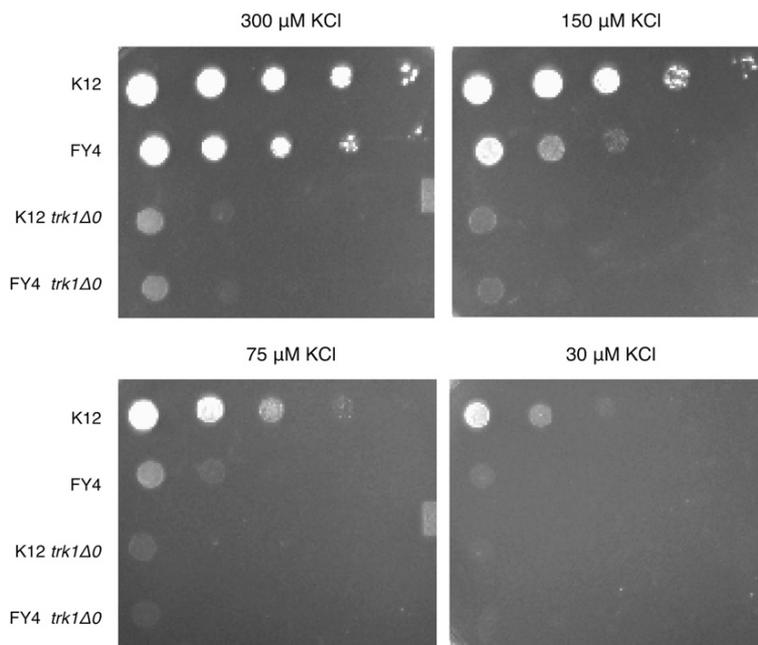
#### **K12 *TRK1* allele has two amino changes in highly conserved regions of the protein**

Two of the changes highlighted in Figure 6, C113S and E1190G, occur in highly conserved regions of the protein based on sequence conservation found in alignments with *Trk1* homologs from

*S. paradoxus*, *S. bayanus*, and *S. mikatae* (Figure 7). The C113S change occurs in a 104 amino acid stretch of the protein that had 68% identity among these species (including residue 113) and 87.5% strong similarity among these species. The E1190G change occurs in a 118 amino acid stretch of the protein that had 75% identity among these species (including residue 1190) and 92% strong similarity among these species. The E1190G amino acid change is of note because involves a charged amino acid changed to a glycine in the K12 strain. Such a charge change in a conserved region of an ion transporter could have a large effect on protein function.

#### **K12 *TRK1* allele displays greater affinity for potassium compared with the FY4 allele**

To test how the amino acid changes in the *TRK1* K12 allele affect protein function, we determined the limiting amount of potassium



**Figure 8** Phenotype of K12 and FY4 strains on limiting potassium media. Cells were grown on plate with different concentrations of potassium (30, 75, 150, and 300  $\mu\text{M}$  KCl).

for growth needed by the K12 and FY4 strains. To accomplish this, we prepared minimal media plates with glutamate as the nitrogen source and four different potassium concentrations ranging from 30  $\mu\text{M}$  to 300  $\mu\text{M}$ . Ten-fold dilutions of K12 and FY4 were plated on these media and grown at 30° for 4 d (Figure 8). As controls, the *trk1* $\Delta$ 0 mutant in both strain backgrounds also was plated to demonstrate how the null allele performs. These results show that FY4 has impaired growth at 150  $\mu\text{M}$  potassium and fails to grow at all at 30  $\mu\text{M}$  potassium. In contrast the K12 strain grows well at 150  $\mu\text{M}$  potassium and only displays impaired growth at 30  $\mu\text{M}$  potassium. The failure of either of the *trk1* $\Delta$ 0 mutants to grow on 150  $\mu\text{M}$  potassium confirms the importance of the *TRK1* gene to growth at limiting concentrations of potassium. Furthermore, these results support the conclusion that the potassium transporter encoded by the K12 allele of *TRK1* has greater affinity for potassium than the potassium transporter encoded by the FY4 allele of *TRK1*.

## DISCUSSION

### Support for the mechanism of ammonium toxicity

A model for ammonium toxicity was proposed that hypothesized ammonium leaking into the cells through potassium channels at physiological concentrations of potassium (Hess *et al.* 2006). The discovery that the K12 strain derives resistance to ammonium toxicity based on a highly divergent allele of the potassium transporter (*TRK1*) supports this link between ammonium and potassium in yeast. Furthermore, the data in this article are consistent with the direct transport of ammonium by the potassium transporter under conditions of ammonium toxicity as predicted (see the section *Model for ammonium toxicity resistance in K12*).

### Model for ammonium toxicity resistance in K12

We speculate that the dramatic alteration in primary amino acid sequence in the K12 *Trk1* protein (Figure 7) has the consequence of increasing the affinity for potassium as determined by our potassium

limitation experiments. Greater-affinity transporters are often associated with lower flux rates (Epstein and Hagen 1952; Epstein *et al.* 1963) because the transporter holds the cargo for longer during transport due to its greater affinity. If this is the case for the K12 allele of *TRK1*, that allele likely has a lower flux rate for both potassium and ammonium. With a lower ammonium leak rate in K12, these cells would have more time to detoxify the ammonium through excretion of amino acids. This would account nicely for their resistance to ammonium toxicity under these conditions. Of course, direct measurement of the affinity for potassium channels encoded by the different *TRK1* alleles is necessary to validate this model.

### Model for the behavior of the *trk1* $\Delta$ 0 strains

In the absence of the high-affinity, low-flux *Trk1* potassium transporter, *S. cerevisiae* must rely on low-affinity, high-flux systems for potassium transport (Gaber *et al.* 1998). As noted previously in this article, low-affinity transporter systems have a greater flux and therefore in the absence of the *Trk1* transporter both the potassium and ammonium flux into the cell will be greater. We believe this is why the *trk1* $\Delta$ 0 strains display the ammonium toxicity phenotype at greater concentrations of potassium compared with the wild-type strains. The crucial observation in Figure 4 is that the ammonium toxicity in the *trk1* $\Delta$ 0 strains is identical between the K12 and FY4 strain backgrounds. Thus, the removal of *Trk1* from the K12 background removes the resistance to ammonium toxicity. We do not observe any detectable resistance in the K12 background in the absence of *Trk1* suggesting that the K12 allele of *TRK1* is the sole determinant of resistance to ammonium toxicity.

## ACKNOWLEDGMENTS

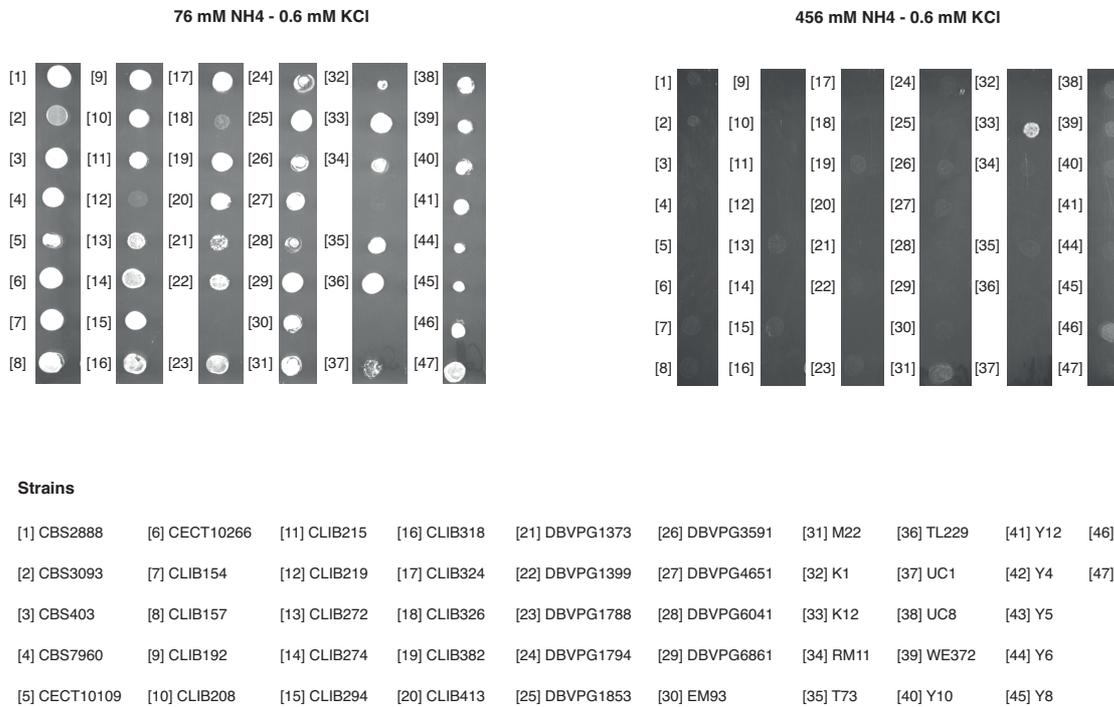
C.R. is supported by a grant from CNRS and Région Alsace, J.S. is supported by an Agence nationale de la recherche grant (2011-JSV6-004-01), D.C.H. and J.S. are supported by a National Institutes of Health grant (R01 GM101091-01), and D.C.H. is supported by a National Science Foundation grant (1122240).

## LITERATURE CITED

- Akao, T., I. Yashiro, A. Hosoyama, H. Kitagaki, H. Horikawa *et al.*, 2011 Whole-genome sequencing of sake yeast *Saccharomyces cerevisiae* Kyokai no. 7. *DNA Res.* 18: 423–434.
- Dunham, M. J., 2012 Two flavors of bulk segregant analysis in yeast quantitative trait loci (QTL). *Methods Mol. Biol.* 871: 41–54.
- Epstein, E., and C. E. Hagen, 1952 A kinetic study of the absorption of alkali cations by barley roots. *Plant Physiol.* 27: 457–474.
- Epstein, E., D. W. Rains, and O. E. Elzam, 1963 Resolution of dual mechanisms of potassium absorption by barley roots. *Proc. Natl. Acad. Sci. USA* 49: 684–692.
- Fay, J. C., and J. A. Benavides, 2005 Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet.* 1: e5.
- Gaber, R. F., C. A. Styles, and G. R. Fink, 1988 TRK1 encodes a plasma membrane protein required for high-affinity potassium transport in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8: 2848–2859.
- Gresham, D., D. M. Ruderfer, S. C. Pratt, J. Schacherer, M. J. Dunham *et al.*, 2006 Genome-Wide Detection of Polymorphisms at Nucleotide Resolution with a Single DNA Microarray. *Science* 311: 1932–1936.
- Hess, D. C., W. Lu, J. D. Rabinowitz, and D. Botstein, 2006 Ammonium toxicity and potassium limitation in yeast. *PLoS Biol.* 4: e351.
- Huson, D. H., and D. Bryant, 2006 Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23: 254–267.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan *et al.*, 2007 Clustal W and clustal X version 2.0. *Bioinformatics* 23: 2947–2948.
- Marullo, P., M. Aigle, M. Bely, I. Masneuf-Pomarède, P. Durrens *et al.*, 2007 Single QTL mapping and nucleotide-level resolution of a physiological trait in wine *Saccharomyces cerevisiae* strains. *FEMS Yeast Res.* 7: 941–952.
- Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, 2009 Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458: 342–345.
- von Wiren, N., and M. Merrick, 2004 Regulation and function of ammonium carriers in plants, yeast and bacteria. *Trends Curr Genet.* 9: 95–120.
- Wenger, J. W., K. Schwartz, and G. Sherlock, 2010 Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genet.* 6: e1000942.
- Will, J. L., H. S. Kim, J. Clarke, J. C. Painter, J. C. Fay *et al.*, 2010 Incipient balancing selection through adaptive loss of aquaporins in natural *Saccharomyces cerevisiae* populations. *PLoS Genet.* 6: e1000893.
- Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson *et al.*, 1999 Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* 285: 901–906.

Communicating editor: B. J. Andrews

## Supporting information



**Figure S1** Screening of 63 *S. cerevisiae* on high concentration of ammonium. Cells were grown on plate with high (456 mM NH<sub>4</sub>) and low (76 mM NH<sub>4</sub>) concentration of ammonium.

**Table S1:** Description of *S. cerevisiae* strains studied

Strains	Sources	Location
CBS2888	Soil South	Africa
CBS3093	Olive-mill wastes	Spain
CBS403	Ginger beer	West Africa
CBS7960	Cane-sugar syrup	Brazil
CECT10109	Prickly pear	Spain
CECT10266	Tanning liquor	Spain
CLIB154	Wine	Russia
CLIB157	Wine	Spain
CLIB192	Baker	France
CLIB208	Baker	China
CLIB215	Baker	New Zealand
CLIB219	Wine	Russia
CLIB272	Beer	United States
CLIB274	Baker	Czech republic
CLIB294	Distillery	France
CLIB318	Baker	Holland
CLIB324	Baker	Vietnam
CLIB326	Baker	Australia
CLIB382	Beer	Japan
CLIB413	Fermentation	China
DBVPG1373	Soil	Netherlands
DBVPG1399	Grape	Netherlands
DBVPG1788	Soil	Finland
DBVPG1794	Soil	Finland
DBVPG1853	White Tecc	Ethiopia
DBVPG3591	Cocoa beans	Unknown
DBVPG4651	Tuber Magnum	Italy
DBVPG6041	Faeces of Man	Unknown
DBVPG6861	Polluted stream water	Brazil
EM93	Rotting Fig	California
M22	Wine	Italy
K1	Sake	Japan
K12	Sake	Japan
RM11	Wine	California
T73	Wine	Spain
TL229	Cheese	France
UC1	Wine	France
UC8	Wine	South Africa
WE372	Wine	South Africa
Y10	Fermentation	Philippines
Y12	Palm Wine	Ivory Coast
Y4	Fruit	Indonesia
Y5	Bill wine	West Africa
Y6	Unknown	French Guiana
Y8	Vineyard	Turkey
Y9	Ragi	Indonesia

### 3. Discussion et conclusion

Ce travail constitue un très bon exemple d'exploration de la diversité intraspécifique qui réside au sein de l'espèce de *S. cerevisiae*. Dans cette étude, nous avons déterminé l'origine génétique de la résistance à une forte concentration d'ammonium de la souche K12. La résistance est due au gène *TRK1* codant un transporteur potassique de haute affinité. La séquence primaire du transporteur de la souche K12 possède deux changements d'acides aminés dans des régions hautement conservées. Ces modifications peuvent avoir un impact sur la fonction du transporteur. Un test de croissance sur gradient de potassium de la souche résistante K12, de la souche sensible FY4 et des mutants K12 *trk1Δ0* et FY4 *trk1Δ0* suggère que la souche K12 a une meilleure affinité au potassium grâce à la présence de cet allèle modifié du gène *TRK1*. Dans de futures études il serait intéressant de valider cette hypothèse en mesurant l'affinité pour le potassium de l'allèle *TRK1* de K12 et de FY4 par des analyses biochimiques. Si les souches possédant l'allèle du gène de K12 codent un transporteur avec une meilleure affinité pour le potassium que les souches possédant l'allèle de FY4, nous pourrions en déduire que notre hypothèse était correcte. Chez la souche K12, l'entrée de  $\text{NH}_4^+$  dans la cellule serait alors réduite par rapport à celui des autres souches.

Notre laboratoire a créé en 2013 le projet 1002 génomes de levures. Ce projet d'étude des populations de grande envergure vise à séquencer les génomes de 1002 isolats de *S. cerevisiae*. Il serait intéressant de cribler la résistance à l'ammonium pour toutes les souches de cette collection de la même manière que cela a été fait pour les 63 souches. Ainsi, il sera possible de voir si d'autres souches sont résistantes et si cela est dû à un mécanisme similaire que celui de la souche K12. De plus, en analysant la proximité génétique des éventuelles souches résistantes avec la souche K12, nous pourrions tester si cette particularité phénotypique pourrait provenir d'un ancêtre commun ou si au contraire elle est apparue de manière séparée suite à la pression de sélection exercée par l'environnement dans lequel elles vivent. Nous pourrions ainsi avoir une vue d'ensemble de la toxicité de l'ammonium au sein de l'espèce *S. cerevisiae*. Cette analyse pourra également être étendue à la collection des 45 souches de *L. kluyveri* tout en gardant à l'esprit que les souches de cette espèce sont toutes issues de l'environnement (principalement exsudats d'arbre et sol), ces niches écologiques contiennent probablement peu d'ammonium.

Ce travail constitue une preuve de principe pour l'analyse de pool de ségréants appliquée à la thématique du laboratoire : déterminer les origines génétiques d'une variabilité phénotypique (ici la résistance à l'ammonium) au sein d'une collection de levure. Dans ce cas, le phénotype est du à un seul locus. Dans le cadre de ce manuscrit, cette méthodologie a été appliquée à des traits complexes comme la résistance à des antifongiques azolés au sein de deux espèces de levures : *S. cerevisiae* et *L. kluyveri*. Ce projet et les résultats que nous avons obtenus sont présentés dans le chapitre 2 de ce manuscrit.

## Chapitre 2 : Analyse de la variabilité intraspécifique de la résistance aux composés azolés chez *S. cerevisiae* et *L. kluyveri*

### Introduction

#### 1. *S. cerevisiae*, un pathogène opportuniste

Comme nous l'avons déjà mentionné dans l'introduction générale du manuscrit, *S. cerevisiae* est à la fois une levure domestiquée avec entre autres des souches utilisées dans la production de boissons alcoolisées et à la fois une levure issue de l'environnement à l'image des 99 souches isolées de forêts primaires chinoises (Wang *et al.* 2012). Cette espèce peut également vivre dans l'appareil digestif de l'Homme de manière commensale. C'est notamment pour cette caractéristique qu'un traitement aux levures peut être préconisé après une prise d'antibiotiques. Cependant, de nombreuses infections à *S. cerevisiae* ont été signalées depuis les années 1980 (Enache-Angoulvant and Hennequin 2005). Les souches isolées à partir d'infections sont dites cliniques. Les infections décrites concernent différents organes et touchent des personnes dont le système immunitaire est déficient, elles sont dites immunodéprimées (de Llanos *et al.* 2006). Le nombre de personnes immunodéprimées s'est considérablement accru ces dernières décennies notamment du fait de l'expansion du VIH (Virus d'Immunodéficience Acquis), de l'utilisation des immunosuppresseurs, des chimiothérapies, des examens et soins plus invasifs et agressifs (Minamoto and Rosenberg 1997; Paya 1993). Certaines souches cliniques ont des caractéristiques semblables à d'autres levures pathogènes telles que la virulence, la capacité de pousser à une température corporelle (37°C à 42°C) et la capacité à former des hyphes qui sont plus invasifs que la forme de levure unicellulaire classique (Clemons *et al.* 1994; McCusker *et al.* 1994). Ces phénotypes communs entre les souches cliniques ont conduit à la question de leurs origines génétiques : est-ce que les souches cliniques ont un ancêtre commun récent ?

En 2007, la première souche clinique a été complètement séquencée (Wei *et al.* 2007). Cette souche, YJM789, a été isolée à partir du poumon d'un patient atteint du SIDA et souffrant d'une pneumonie. De nombreuses différences par rapport au génome de S288c ont été mises en évidence grâce à ce séquençage aussi bien au niveau nucléotidique que chromosomique. Au total, 59 361 SNP et 5 869 insertions-délétions ont été détectés, ainsi qu'une région inversée de 32,5 kb sur le chromosome XIV et une translocation entre le chromosome VI et X. Ces variations génétiques peuvent avoir un impact direct sur le phénotype comme par exemple le fort polymorphisme détecté dans le gène de résistance multiple *PDR5* qui pourrait conférer à la souche une meilleure résistance à un traitement antifongique. De plus, il a été mis en évidence 115 ORFs supplémentaires par rapport au

génomique de référence. Certains gènes comme YJM-GNAT pourraient même provenir d'un transfert horizontal, et dans ce cas précis probablement depuis une bactérie. Ce gène, tout comme d'autres membres de la famille GNAT, pourrait être à l'origine d'une résistance à des antibiotiques aminoglycosides.

L'étude de la variabilité génétique de 63 souches de *S. cerevisiae* par hybridation sur puces à ADN a montré que les souches cliniques n'ont pas un ancêtre commun proche mais que les infections proviennent d'événements isolés durant lesquels des souches issues de l'environnement ou de la vigne ont pu coloniser des tissus humains (Schacherer *et al.* 2009). Les phénotypes communs observés entre les souches cliniques sont donc probablement dus à une convergence évolutive.

## 2. Les antifongiques

Les antifongiques sont des molécules à usage thérapeutique utilisées pour traiter les infections fongiques où pour les prévenir chez des patients immunodéprimés. Ils sont nettement moins nombreux que les antibiotiques utilisés contre les infections bactériennes. Cette différence s'explique par le fait que les mycètes sont des organismes eucaryotes tout comme les organismes hôtes. Il est nécessaire que la cible de l'antifongique soit spécifique du pathogène et épargne les cellules de l'organisme hôte. Le traitement (classe d'antifongique et voie d'administration) dépendra du pathogène et de l'organe infecté. Il existe 8 classes d'antifongiques dont 7 sont utilisées en médecine : les polyènes, les analogues de la pyrimidine, les azolés, les candines, les allylamines, les thiocarbamates et les morpholines (Sanglard *et al.* 2009). Les antifongiques ont pour cible l'ergostérol, sa voie de biosynthèse, la voie de biosynthèse du 1,3- $\beta$ -glucan ou encore la synthèse de l'ADN et de l'ARN. L'ergostérol est un composant majeur de la membrane des mycètes. Il joue le même rôle que le cholestérol dans les membranes des cellules animales. Il fluidifie la membrane et lui permet de garder son intégrité. Sa présence spécifique dans les cellules des mycètes et son absence dans les cellules humaines en font une cible idéale pour les antifongiques.

### 2.1. Les polyènes

Les polyènes appartiennent à une classe de composés naturels de nature amphipatique. Ils se lient à l'ergostérol présent dans la membrane des mycètes. Ils vont créer des pores dans la membrane, ce qui va permettre le passage des petites molécules et va avoir pour conséquence la mort cellulaire. Les polyènes sont donc une classe d'antifongiques fongicides (Masiá Canuto and Gutiérrez Rodero 2002). Il y a deux principaux composés dans cette classe : l'amphotéricine B et la nystatine. L'amphotéricine B a été le 1<sup>er</sup> antifongique commercialisé et ce, depuis plus de 50 ans (Gallis *et al.* 1990). Il est l'antifongique le plus prescrit en cas d'infections graves et systémiques pour lesquelles il faut agir rapidement.

## 2.2. Les analogues de la pyrimidine

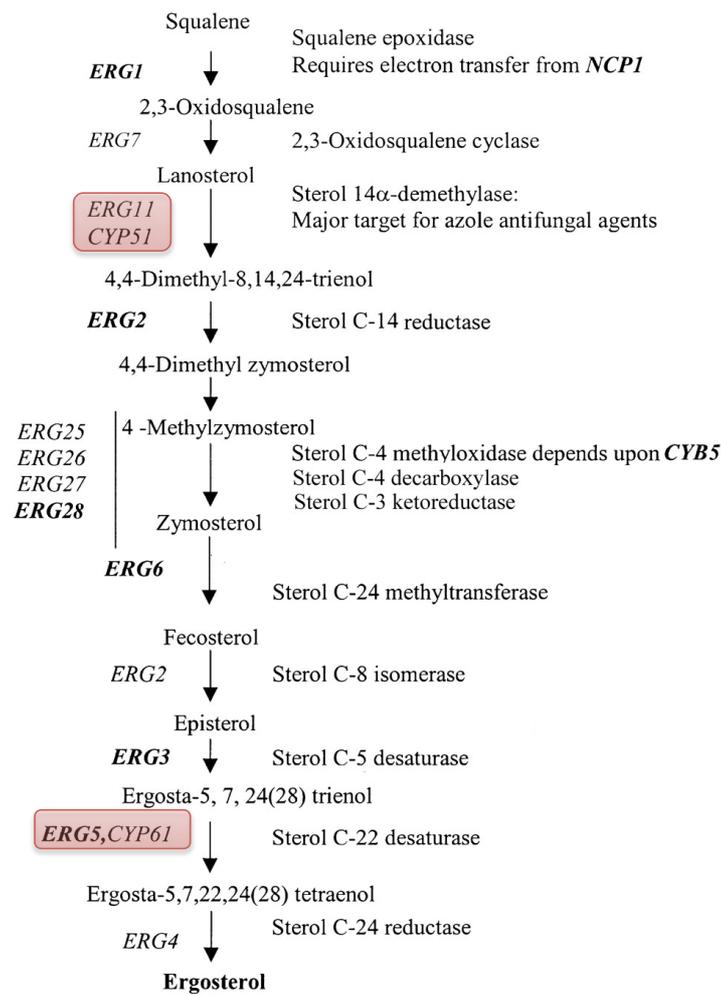
Un seul antifongique est présent dans cette classe : la 5-fluorocytosine. Une fois dans la cellule fongique, elle va être transformée par la cytosine désaminase en 5-fluorouracile qui sera incorporée dans l'ADN et l'ARN, ce qui va inhiber la division cellulaire. La 5-fluorouracile est fongistatique (Polak and Scholer 1975). Elle est rarement utilisée seule mais plutôt en combinaison avec d'autres antifongiques.

## 2.3. Les composés azolés

Ce sont les antifongiques les plus utilisés. Les azoles inhibent la voie de biosynthèse de l'ergostérol agissant sur des enzymes clé de cette voie. Certains azoles ciblent et bloquent l'enzyme lanosterol 14- $\alpha$ -déméthylase du cytochrome P450 codée par le gène *ERG11*. Cette enzyme existe aussi chez les cellules des mammifères, le composé azolé est donc capable de bloquer également les enzymes de l'organisme hôte. Toutefois, l'affinité des antifongiques est bien plus grande pour l'enzyme fongique, ce qui permet de ne pas bloquer les enzymes d'origines humaines (Koltin and Hitchcock 1997). Cependant, à fortes doses, ces composés peuvent avoir un effet toxique sur les cellules de mammifères. Cette toxicité est moins élevée avec l'utilisation du fluconazole, de l'itraconazole et du voriconazole car leur affinité pour le cytochrome P450 fongique est supérieure (Sheehan *et al.* 1999). Des azoles ont également pour cible la C-22 stérol désaturase, une autre enzyme du cytochrome P450, codée par le gène *ERG5* (Skaggs *et al.* 1996). Le blocage de la voie de biosynthèse de l'ergostérol va avoir pour conséquence l'accumulation des précurseurs en concentrations toxiques. Ces précurseurs qui sont méthylés vont s'intégrer dans la membrane à la place de l'ergostérol ce qui va conduire à une altération de la structure et de la fonction de la membrane. Il existe deux classes de composés azolés qui se distinguent par le nombre d'atomes d'azote présents dans le cycle azoté : les imidazoles (2 N) qui ont été découverts en premier et les triazoles (3 N).

Les imidazoles comprennent notamment le kétoconazole, le miconazole et la clotrimazole. Ils permettent de traiter les infections locales des muqueuses et de la peau. Le kétoconazole a été le 1<sup>er</sup> composé azolé commercialisé dans les années 80. Les premiers composés imidazolés comme le kétoconazole et le miconazole ont des mécanismes d'action complexe avec plusieurs cibles.

Les triazoles incluent le fluconazole, le voriconazole, l'itraconazole et le posaconazole. Le fluconazole et l'itraconazole sont des triazoles de première génération. Leur spectre d'action est plus étendu que les imidazoles. Le voriconazole et posaconazole sont appelés les azolés de seconde génération. Ils ont une activité accrue sur les pathogènes émergents (Maertens 2004). Le fluconazole est moins toxique, présente moins de problèmes liés à l'interaction avec d'autres médicaments et a un meilleur profil pharmacocinétique que le kétoconazole mais son spectre d'action est limité aux levures (Maertens 2004).



**Figure 2.1 :** La voie de biosynthèse de l'ergostérol chez les levures et les étapes d'inhibition des composés azolés (schéma provenant de Parveen *et al.* 2004). Les gènes ciblés par les composés azolés sont surlignés en rouge.

## 2.4. Les candines

C'est la catégorie la plus récente d'antifongiques, ils ont pour cible un composant majeur de la paroi cellulaire : le  $\beta$ -1,3 glucane. Trois composés sont classés dans cette catégorie, la caspofongine, la micafongine et l'anidulafongine. Les candines sont utilisées contre les infections à *Candida* et à *Aspergillus* (Sanglard *et al.* 2009).

## 2.5. Les allylamines, thiocarbamates et morpholines

Ils inhibent la voie de biosynthèse de l'ergostérol à différentes étapes. Les allylamines et thiocarbamate ciblent la squalene epoxidase (Erg1) et les morpholines inhibent les enzymes Erg2 et Erg24. Ils sont utilisés contre les agents pathogènes d'origine tropicale et contre les infections dermatophytes (Sanglard *et al.* 2009).

## 2.6. Quels traitements pour *S. cerevisiae* ?

Les traitements utilisés en cas d'infection à *S. cerevisiae* sont : l'amphotéricine B (AmB), le fluconazole (FC), le kétoconazole (KC) et des combinaisons miconazole (MC) + 5-FC, Am B + 5-FC, FC + Am B, AmB + KC et IC + 5-FC (Enache-Angoulvant and Hennequin 2005). Mise à part l'amphotéricine B qui est utilisée en cas d'infection grave et étendue, les composés azolés sont les antifongiques les plus prescrits pour les infections à *S. cerevisiae*. C'est pourquoi, dans le cadre de ce travail, l'étude de la résistance s'est très vite focalisée sur les antifongiques azolés et notamment le fluconazole et le kétoconazole qui sont les principaux antifongiques de la classe des triazoles et des imidazoles. Des souches résistantes sont apparues à cause de l'utilisation massive de certains antifongiques lors des traitements à long terme ou lors d'infections répétées. De plus, la majorité des antifongiques ayant une action fongistatique plutôt que fongicide, statistiquement plus de cellules ont la possibilité de devenir résistantes.

## 3. Les mécanismes de résistance aux composés azolés chez les levures

Le plus grand nombre d'études portent sur la levure *Candida albicans* qui est la levure pathogène opportuniste causant le plus grand nombre d'infections à travers le monde (Lamping *et al.* 2010). C'est pourquoi, la majorité des cas de résistance rapportés concerne cette espèce. Différents mécanismes de résistance existent : la modification de la cible de l'antifongique, la surexpression de la cible, l'altération de la voie de la biosynthèse des stérols, le blocage de l'entrée de l'antifongique dans la cellule, la formation de biofilms et l'expulsion de l'antifongique par des transporteurs.

### 3.1. Modification de la cible

Chez *Candida albicans*, de nombreuses études ont détecté des mutations non synonymes dans la séquence codante du gène *ERG11* de souches cliniques résistantes à des composés azolés. Le gène *ERG11* code pour l'enzyme lanosterol 14-alpha-demethylase de la voie de biosynthèse de l'ergostérol. Cette enzyme est la cible principale des antifongiques azolés. Une mutation non synonyme dans sa séquence codante va conduire à la modification de la séquence protéique. Certaines de ces modifications peuvent aboutir à une baisse de l'affinité de l'antifongique pour l'enzyme, ce qui entraînera une résistance au composé.

De nouvelles mutations sont régulièrement mises en évidence (Marichal *et al.* 1999; Xiang *et al.* 2013; Ying *et al.* 2013). Ce fort polymorphisme du gène *ERG11* suggère que la protéine Erg11 est hautement permissive aux changements structuraux. Plusieurs mutations peuvent apparaître chez une même souche, ce qui rend difficile l'estimation de l'impact de chaque mutation sur le phénotype de résistance. Pour estimer la contribution de chacune des mutations, des expressions hétérologues de gènes codant pour des enzymes modifiées par mutagenèse dirigée sont utilisées chez *S. cerevisiae* (Sanglard *et al.* 1998). Chez cette espèce, des mutations non synonymes dans la séquence codante de *ERG11* conduisent également à une augmentation de la résistance au fluconazole (Anderson *et al.* 2003).

### 3.2. Surexpression de la cible

La surexpression de la cible a pour conséquence d'augmenter la quantité d'enzymes présentes dans la cellule. Une partie de la cible sera fixée par les antifongiques mais une autre partie restera disponible du fait de l'augmentation du nombre de protéines et pourra ainsi assurer sa fonction dans la machinerie de la cellule. Chez *C. albicans*, il a été prouvé que la surexpression de *ERG11* entraînait une résistance aux composés azolés. Cette surexpression est provoquée par la réduction de la quantité d'ergostérol mais peut également survenir de manière constitutive suite à des mutations dans les gènes codant des facteurs de transcription (Henry *et al.* 2000; Heilmann *et al.* 2010; Perea *et al.* 2001).

### 3.3. L'altération de la voie de biosynthèse des stérols

Lorsque l'enzyme codée par *ERG11* est bloquée par un antifongique, en plus d'une accumulation de lanostérol dans la cellule due à sa non-utilisation, l'enzyme codée par *ERG3* transforme le stérol 14 $\alpha$ -méthylé en un composé intermédiaire toxique, le 14-méthyl-3,6-diol. Des mutations dans le gène *ERG3* qui permettent d'inactiver ou de tronquer la protéine ont été détectées chez *S. cerevisiae* et *Candida dubliniensis*. L'absence d'enzyme fonctionnelle dans ce cas va empêcher la production du composé toxique (Watson *et al.* 1989; Pinjon *et al.* 2003).

### 3.4. L'entrée de l'antifongique dans la cellule

La composition de la membrane de la cellule peut subir une modification afin de limiter l'entrée des antifongiques dans la cellule, les empêchant ainsi de se fixer à leur cible. Une étude a montré que chez *C. albicans*, la résistance au miconazole peut être due à la baisse du nombre de lipides, de phospholipides et du contenu en stérols dans la membrane cellulaire. La modification du contenu et de l'organisation de la membrane a donc pour conséquence de limiter l'entrée de l'antifongique dans la cellule (Sharma and Khuller 1996).

### 3.5. Les mécanismes de Résistance Multiples aux Drogues (MDR)

Les mécanismes de MDR permettent une résistance multiple aux composés toxiques. Les molécules toxiques comprennent entre autres les antifongiques, les antibiotiques, les herbicides et les anticancéreux. Ces mécanismes sont à l'origine de problèmes de santé publique comme par exemple la résistance des organismes pathogènes à un traitement antifongique et la résistance des cellules tumorales à un traitement anticancéreux. Ils incluent la formation de biofilm et l'expulsion du composé toxique par des transporteurs (Mah and O'Toole 2001; Sanglard *et al.* 2009).

#### 3.5.1. La formation de biofilms

Les biofilms sont des communautés de micro-organismes hautement structurées. Par le biofilm, les micro-organismes adhèrent entre eux mais également à une surface grâce à la matrice extracellulaire qui les encapsule (Ramage *et al.* 2012). Ce mode de vie est très courant, en effet près de 80% des micro-organismes vivent en biofilm plutôt qu'en utilisant un mode de vie planctonique (Donlan 2002). Les biofilms peuvent engendrer des problèmes de santé car bien qu'ils puissent se former sur des surfaces naturelles, ils peuvent également apparaître sur des surfaces synthétiques comme par exemple les cathéters utilisés en médecine. *C. albicans* a une grande capacité à former des biofilms ce qui n'est pas le cas de *S. cerevisiae* pour qui, le développement d'un biofilm s'arrête lors des premiers stades de sa formation. Cet arrêt prématuré est notamment dû à une production très limitée de matrice extracellulaire (Chandra *et al.* 2001). Les biofilms ont la propriété de résister aux composés toxiques (Ramage *et al.* 2002). Plusieurs raisons ont été avancées pour expliquer cette résistance multiple. La résistance peut être due à la forte densité des cellules présentes dans le biofilm. La matrice extracellulaire peut jouer un rôle de barrière mécanique et limiter la pénétration de certains composés toxiques dans le biofilm. La présence d'un flux de liquide au sein du biofilm peut également jouer un rôle d'évacuation des composés toxiques. D'autres causes de résistance sont dues à l'état des cellules (cycle cellulaire, expression des gènes, mutations, ...). Les biofilms peuvent contenir des cellules dites « persistantes » qui sont hautement tolérantes à des antifongiques. Elles sont phénotypiquement différentes de la souche sauvage, mais leur génome n'est pas muté. Les mécanismes conduisant à leur résistance accrue sont encore inconnus (Ramage *et al.* 2012). La perception des différents stress (pH, température, pression osmotique, ...) via les récepteurs va induire des réponses de la part de la cellule comme la formation du biofilm.

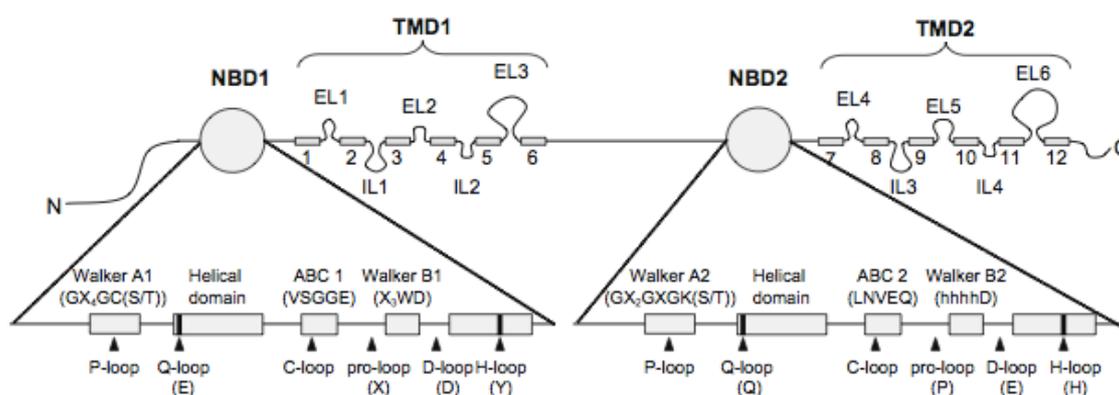
#### 3.5.2. L'expulsion de l'antifongique par des transporteurs

Les transporteurs de type ABC (ATP-Binding Cassette) et les pompes MFS (Major Facilitator Superfamily) sont les deux types de pompes à efflux impliquées dans la résistance multiple aux antifongiques. Les transporteurs de type ABC utilisent l'hydrolyse de l'ATP pour expulser le composé tandis que les pompes MFS utilisent le gradient électrochimique de la membrane. Les transporteurs

ABC sont présents chez tous les organismes vivants. Trois des cinq classes de transporteurs ABC (PDR pour Pleiotropic Drug Resistance, MDR pour MultiDrug Resistance, MRP pour Pleiotropic drug Résistance-associated Protein) sont impliquées dans la résistance aux antifongiques (Lamping *et al.* 2010). Il existe 29 gènes codant pour des transporteurs ABC chez *S. cerevisiae*. Les principales pompes impliquées dans la résistance aux composés azolés sont les transporteurs de type PDR. C'est la surexpression des gènes codant ces transporteurs qui conduit à une résistance.

### 3.5.2.1. Les transporteurs PDR

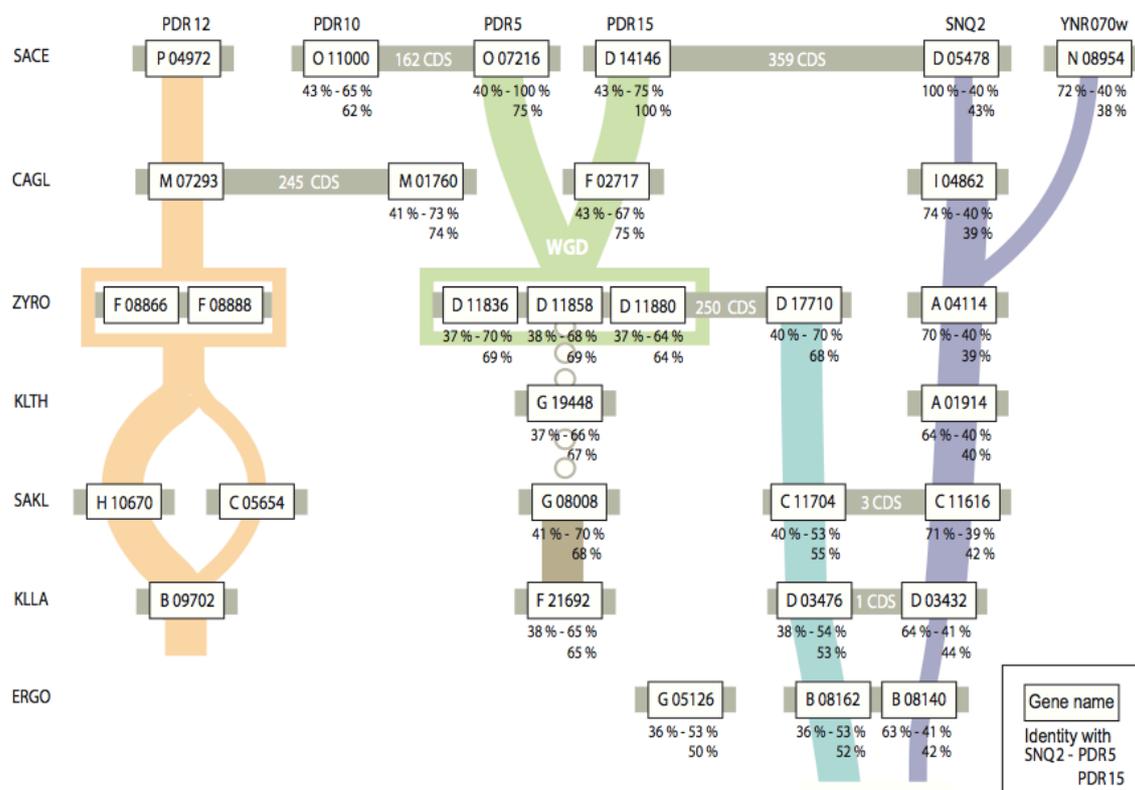
Les transporteurs PDR sont formés de deux domaines transmembranaires (TMD pour Transmembrane Domain) alternés avec deux domaines liant l'ATP (NBD pour Nucleotide Binding Domain) dans l'ordre suivant : NBD-TMD-NBD-TMD (Figure 2.2). Le domaine TMD contient des hélices  $\alpha$  dans 12 segments transmembranaires et le domaine NBD contient une alternance de feuilles  $\beta$  et d'hélices  $\alpha$  de manière à former un repliement de Rossman. Les 12 domaines transmembranaires sont reliés entre eux par 6 boucles externes (ECL1-6) et 4 boucles internes (ICL1-4). Les domaines NBD contiennent quand à eux 7 motifs conservés : (1) le motif Walker A ou P-loop GxxGxGKS/T (x est un acide aminé); (2) Le motif Walker B hhhD où h est un acide aminé aliphatique; (3) la boucle C au motif LSGGQQ/R/KQR qui est la signature des transporteurs ABC; (4) la boucle Q; (5) la boucle pro; (6) la boucle H; (7) et la boucle D (Lamping *et al.* 2010). Les transporteurs PDR jouent un rôle dans la détoxification des cellules, dans le transport des stérols, dans le quorum sensing des cellules, dans le vieillissement et ont un rôle potentiel dans le transport des ions (Prasad and Goffeau 2012).



**Figure 2.2 :** Schéma de la topologie d'un transporteur de type PDR avec ses principales caractéristiques. N = N-terminal ; C = C-terminal ; NBD = nucleotide binding domain ; TMD = transmembrane domain ; EL = extracellular loop ; IL = intracellular loop. Les segments transmembranaires sont annotés de 1 à 12. Les acides aminés conservés sont mis entre parenthèse avec X qui représente un acide aminé et h un acide aminé aliphatique (Lamping *et al.* 2010).

Chez *S. cerevisiae*, il y a 6 transporteurs PDR *sensu stricto* Pdr5, Pdr10, Pdr12, Pdr15, Snq2 et YNR070wp (Decottignies and Goffeau 1997). De plus, elle contient 3 transporteurs PDR *sensu lato* qui n'expulsent pas d'antifongiques et qui n'ont pas la substitution K/C dans le motif Walker A: Aus1 et Pdr11 qui importent des stérols et YOL075c dont la fonction reste encore inconnue.

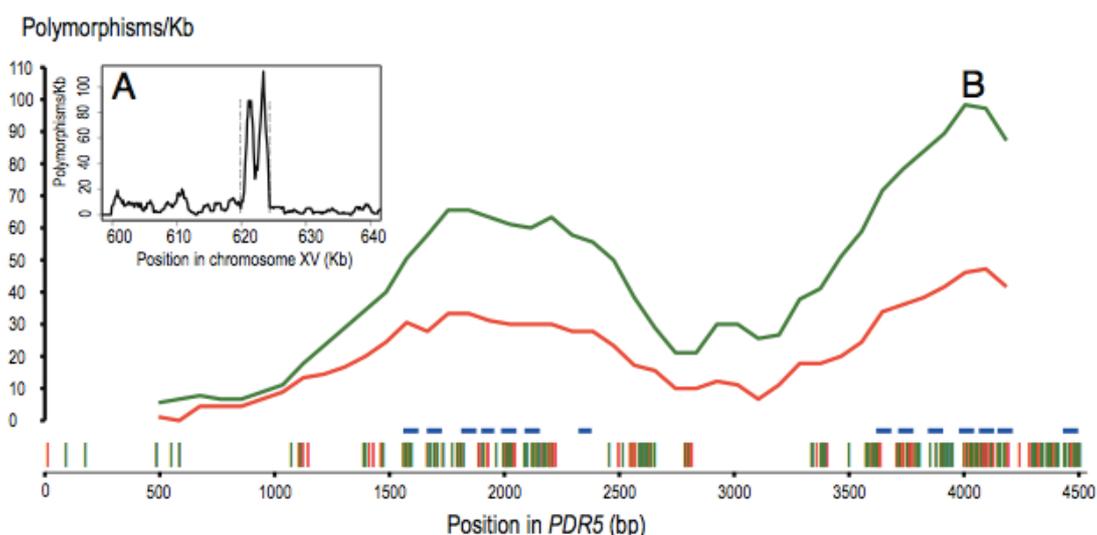
Une analyse phylogénétique combinée à une analyse des gènes adjacents des gènes PDR au sein de 9 espèces d'hémiascomycètes a permis de retracer leur histoire évolutive (figure 2.3). Les gènes *PDR5* et *PDR15* ainsi que *SNQ2* et *YNR070W* (également appelé *PDR18*) sont des gènes paralogues apparus lors de la duplication complète du génome qui a eu lieu au cours de l'histoire évolutive des Hémiascomycètes (figure 2.3) (Seret *et al.* 2009).



**Figure 2.3 :** Relations entre les gènes *PDR sensu stricto* basées sur l'analyse des gènes adjacents (Seret *et al.* 2009). Chaque boîte représente un gène *PDR*. Les lignes grises relient les gènes situés sur le même chromosome. Lorsqu'ils sont non adjacents, le nombre de séquences codantes entre deux gènes *PDR* est écrit en blanc. Des voisins communs sont représentés par des lignes de connexion colorées. Une large bande signifie que plus d'un voisin est partagé tandis qu'une fine bande indique un seul gène voisin. Les lignes discontinues indiquent des liens uniquement prouvés par analyse phylogénétique. Les pourcentages d'identité par rapport aux gènes *PDR5*, *PDR15* et *SNQ2* sont écrits en dessous des gènes. Les lettres WGD indiquent l'événement de la duplication totale du génome.

Le transporteur Pdr5p est capable d'expulser une centaine de composés hydrophobes hors de la cellule (Ernst *et al.* 2008), dont le kétoconazole et le fluconazole. De plus, 316 composés parmi lesquels des hormones (progestérone), des anesthésiants (dyclonine) et un certain nombre

d'antifongiques (la 5-fluorocytosine, kétoconazole) ont été testés sur une collection de mutants de délétion de *S. cerevisiae* (Hillenmeyer *et al.* 2008). La souche portant une délétion du gène *PDR5* présente une sensibilité à plus de 20% des composés testés. Ce pourcentage élevé permet de le considérer comme un gène de résistance multiple. Les facteurs de transcription *PDR1* et *PDR3* font également partie de cette catégorie. Le gène *PDR5* est également intéressant car son gène orthologue *CDR1* de *C. albicans* est impliqué dans la résistance à des traitements antifongiques (Chen *et al.* 2010).



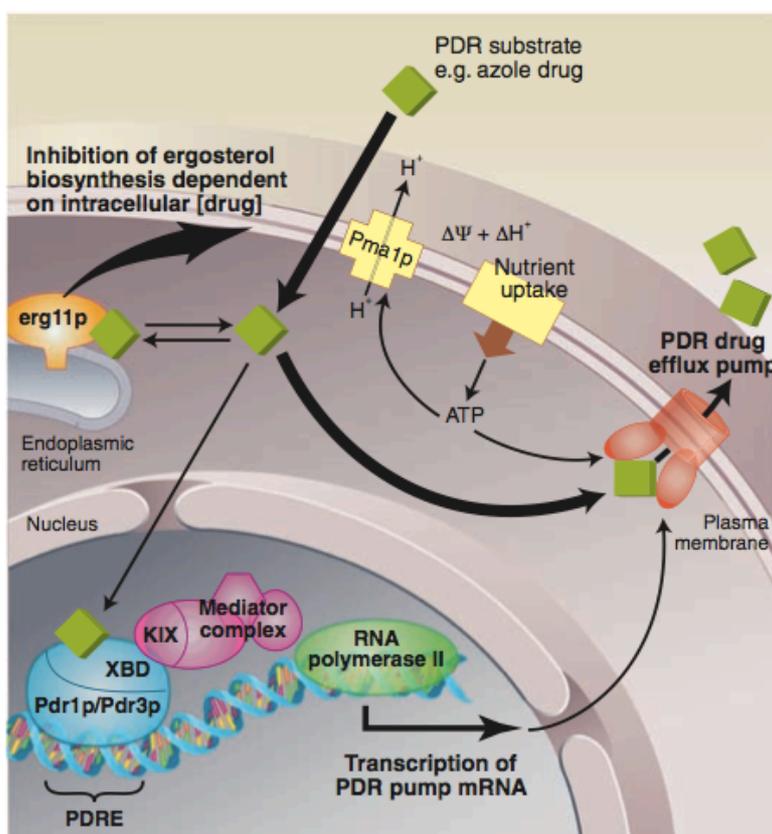
**Figure 2.4 :** Distribution du polymorphisme entre les souches YJM789 et S288c du gène *PDR5* (Wei *et al.* 2007). A) Distribution du polymorphisme au niveau du chromosome XV entre les coordonnées 600 à 640 kb. Les lignes pointillées indiquent les positions start et stop de la phase codante. (B) La distribution des mutations synonymes et non synonymes dans la séquence codante de *PDR5*. Les lignes rouges indiquent le polymorphisme des substitutions non synonyme et les vertes celui des synonymes. Les barres horizontales bleues représentent les domaines transmembranaires et les barres verticales du bas les sites de substitution.

Le gène *PDR5* peut présenter un très fort polymorphisme comme c'est le cas chez la souche clinique YJM789. Lors du séquençage de la souche en 2007, 22 ORFs hautement polymorphes avaient été détectées dont notamment le gène *PDR5*. Les données indiquent la présence d'environ 250 SNPs et une divergence en acides aminés de 5,3% avec S288c alors que la divergence moyenne de l'ensemble du génome est de 0,43% (figure 2.4). Les SNP sont localisés au niveau de deux régions et plus précisément dans celles codant les 2 domaines transmembranaires TMS. Ces régions jouent probablement un rôle sur la reconnaissance des substrats transportés (Tutulan-Cunita *et al.* 2005). Les auteurs n'ont pas présenté d'hypothèse quant à ce fort polymorphisme localisé. L'analyse des séquences prouve que ce polymorphisme ne provient pas d'une recombinaison entre *PDR5* et son gène paralogue *PDR15*. La souche YJM789 est par ailleurs sensible au cycloheximide qui est habituellement exporté par le transporteur Pdr5p. Une hypothèse émise par les auteurs pour expliquer cette sensibilité est que la version allélique de *PDR5* chez YJM789 coderait pour un transporteur ne

permettant peut être pas l'export du cycloheximide comme chez la souche de référence (Winzeler *et al.* 1998).

### 3.5.2.2. Les facteurs de transcription de type PDR

De nombreux facteurs de transcription tels que *PDR1*, *PDR3*, *YRR1*, *STB5*, *RDR1* et *YAP1* permettent de réguler les mécanismes PDR chez la levure *S. cerevisiae* que ce soit par surexpression ou répression de la transcription de gènes impliqués dans la résistance multiple. Les facteurs de transcription les plus décrits dans la littérature, notamment pour leur capacité à surexprimer le gène *PDR5*, sont *PDR1* et *PDR3*. Ils appartiennent à la famille des « zinc cluster proteins » qui se caractérisent par la présence du motif liant l'ADN  $Zn_2Cys_6$  (Akache and Turcotte 2002). Les antifongiques activent les facteurs de transcription de type PDR en se fixant sur leur domaine XBD (Xenobiotic Binding Domain), ce qui va induire le recrutement de l'ARN polymérase (Figure 2.5) (Monk and Goffeau, 2008).



**Figure 2.5 :** La voie PDR classique chez la levure (Monk and Goffeau 2008). L'antifongique va se lier aux domaines XBD des facteurs de transcription Pdr1 et Pdr3. Ils vont alors recruter l'ARN polymérase II via le complexe médiateur afin de transcrire les gènes codant des transporteurs de résistance multiple. L'augmentation de la transcription va permettre la traduction d'un plus grand nombre de protéines qui vont s'insérer dans la membrane et expulser en masse le composé toxique hors de la cellule.

Le gène *PDR3* est le paralogue de *PDR1* issu de la duplication totale du génome. Tous deux régulent l'expression des gènes possédant des éléments *PDRE* (Pleiotropic Drug Response Elements) au niveau de leur promoteur. La séquence consensus des éléments PDRE est 5'-TCCGCGGA-3'. L'impact de ces facteurs de transcription ne se limite pas au mécanisme de résistance multiple mais touche aussi la composition de la paroi cellulaire, la réponse au stress et la composition lipidique de la membrane (Nawrocki *et al.* 2001). *PDR1* et *PDR3* ont des cibles communes notamment les gènes codant les transporteurs *PDR5*, *SNQ2*, *PDR10* et *PDR15* (MacPherson *et al.* 2006).

Certaines mutations ponctuelles dans la séquence de *PDR1* et *PDR3* vont avoir pour conséquence la surexpression de gènes codant les transporteurs PDR dont *PDR5*, *SNQ2* et *YOR1* ce qui va induire la résistance de la levure aux composés toxiques (Nourani *et al.* 1997; Carvajal *et al.* 1997).

En conclusion, les gènes *PDR* sont des gènes clés de la résistance aux antifongiques. Ils sont impliqués dans la détoxification à de centaines de composés. Une mutation dans leur séquence ainsi que leur surexpression peuvent avoir des conséquences directes sur la résistance à ces composés. C'est pourquoi, nous avons décidé d'explorer dans un premier temps la variabilité intra-spécifique des gènes *PDR* au sein des souches de *S. cerevisiae* et *L. kluyveri*. Dans une première étape, nous avons analysé la séquence de ces gènes afin d'établir le degré de polymorphisme.

Dans un second temps, nous avons testé la variabilité intra-spécifique de résistance à des antifongiques azolés au sein des souches de *S. cerevisiae* et *L. kluyveri*. Un des objectifs était de voir si une corrélation pouvait être mise en évidence entre le polymorphisme des gènes *PDR* et la variabilité de résistance.

Enfin, nous avons voulu déterminer les origines génétiques de la diversité de résistance observée au fluconazole et au kétoconazole grâce à une stratégie d'analyse de pool de ségrégants.

## Résultats

### 1. Étude de la variabilité des gènes de résistance multiple

Comme illustré précédemment, les variations de séquences des gènes *PDR* entre souches d'une même espèce peuvent être à l'origine des différences de résistance observées, comme c'est le cas chez les levures du genre *Candida*. C'est pourquoi la variabilité de gènes impliqués dans la résistance multiple a été analysée au sein d'une collection de 63 souches de *S. cerevisiae* et de 28 souches de *L. kluyveri* d'origines écologiques et géographiques variées (voir Matériels et Méthodes). Les gènes sélectionnés sont 1) *PDR12*, *PDR10*, *PDR5*, *PDR15*, *SNQ2* et *PDR18* codant les transporteurs PDR *sensu stricto* de *S. cerevisiae* et 2) les facteurs de transcription codés par *PDR1* et *PDR3* fortement impliqués dans la régulation des gènes de résistance multiple. Chez *L. kluyveri*, ce sont les orthologues de ces gènes qui ont été analysés (tableau 2.1 et figure 2.3) (Seret *et al.* 2009).

Catégorie	Nom standard	Nom systématique gènes levures	
		<i>S. cerevisiae</i>	<i>L. kluyveri</i>
Transporteur PDR <i>sensu stricto</i>	<i>PDR12</i>	<i>SACE0P04972</i>	<i>SAKL0H10670</i>
			<i>SAKL0C05654</i>
	<i>PDR10</i>	<i>SACE0O11000</i>	
	<i>PDR5</i>	<i>SACE0O07216</i>	<i>SAKL0G08008</i>
	<i>PDR15</i>	<i>SACE0D14146</i>	<i>SAKL0C11704</i>
	<i>SNQ2</i>	<i>SACE0D05478</i>	<i>SAKL0C11616</i>
Facteurs de transcription PDR	<i>PDR18</i>	<i>SACE0N08954</i>	
	<i>PDR1</i>	<i>SACE0G05566</i>	
	<i>PDR3</i>	<i>SACE0B02442</i>	<i>SAKL0D01100</i>

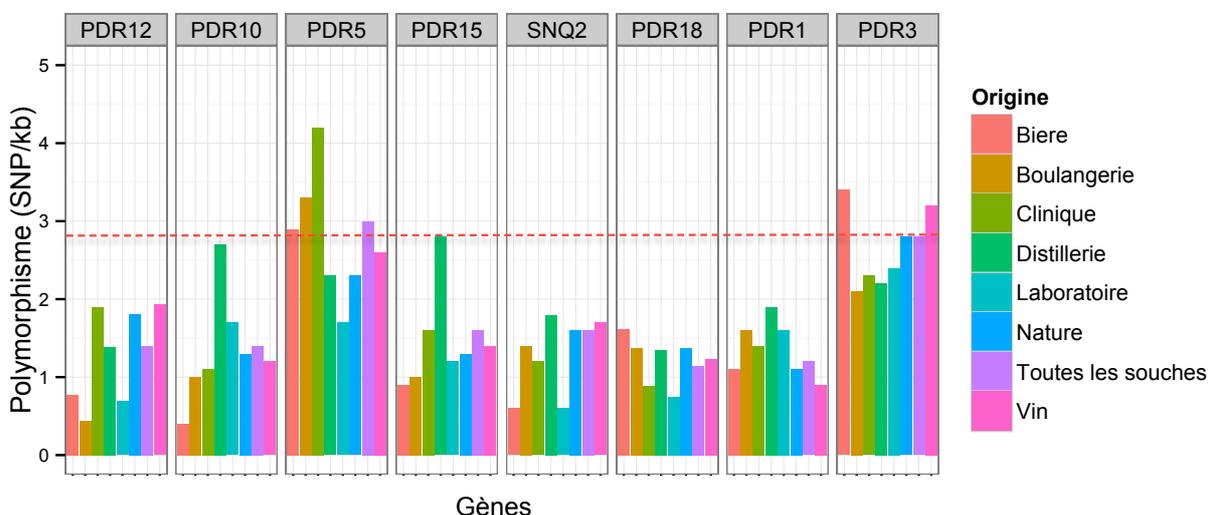
**Tableau 2.1** : Liste des gènes *PDR* sélectionnés pour l'analyse de la variabilité génétique.

#### 1.1. Analyse des gènes *PDR* chez *S. cerevisiae*

Grâce aux cartes génomiques disponibles pour les 63 souches de notre collection (Schacherer *et al.* 2009), il a été possible d'avoir une mesure du polymorphisme des gènes connus comme intervenant dans la résistance multiple aux antifongiques. Chez *S. cerevisiae* les souches les plus proches génétiquement sont celles provenant d'une même origine écologique. C'est pourquoi nous avons regroupé les souches par origine écologique et nous avons calculé la moyenne de polymorphisme en SNP/kb par gène et par catégorie écologique (figure 2.6). Dans ce cadre, notre volonté a été de savoir

s'il y avait une corrélation entre l'origine des souches et la variabilité nucléotidique observée. D'après les cartes génétiques, la moyenne de polymorphisme sur l'ensemble des génomes est de 2,8 SNP/kb. La comparaison de ce polymorphisme moyen avec ceux que nous avons déterminés pour les différents gènes étudiés nous a permis de tirer plusieurs conclusions.

Les gènes *PDR12*, *PDR10*, *PDR15* (mis à part le gène *PDR15* des souches de distillerie), *SNQ2*, *PDR18* et *PDR1* présentent un polymorphisme inférieur à la moyenne et ce, quelle que soit l'origine des souches. Ces gènes intervenant dans la résistance multiple sont donc conservés par rapport à l'ensemble du génome. Les gènes *PDR5* et *PDR3* présentent une moyenne de polymorphisme pour toutes les souches égale ou supérieure à la moyenne de polymorphisme présent sur l'ensemble du génome : il est de 2,8 SNP/kb pour *PDR3* et 3 SNP/kb pour *PDR5*. La variabilité génétique du gène *PDR5* des souches cliniques est encore plus élevée avec une moyenne de 4,2 SNP/kb.



**Figure 2.6 :** Polymorphisme en SNP/kb de 8 gènes *PDR* au sein de la collection de *S. cerevisiae*. Les 63 souches sont regroupées par origine écologique. La ligne rouge en pointillés représente la moyenne de polymorphisme sur l'ensemble du génome au sein des souches de *S. cerevisiae*.

Dans la suite de notre étude, nous nous sommes plus particulièrement focalisés sur ce polymorphisme plus important du gène *PDR5* chez les souches cliniques. Ce gène, d'une taille de 4535 pb, codant un transporteur PDR, est situé sur le chromosome XV entre les coordonnées 619841 et 624376.

### 1.1.1. Analyse du gène *PDR5*

Afin de savoir si toutes les souches cliniques ont un gène *PDR5* polymorphe, nous avons utilisé les données de puces à ADN disponibles sur le serveur : [http://gbrowse.princeton.edu/cgi-bin/gbrowse/yeast\\_strains\\_snps/](http://gbrowse.princeton.edu/cgi-bin/gbrowse/yeast_strains_snps/).

L'analyse a permis de mettre en évidence une disparité du polymorphisme entre les 14 souches cliniques de notre collection (figure 2.7). En effet, il varie de 0 SNP/kb pour la souche YJM436 à 6,2

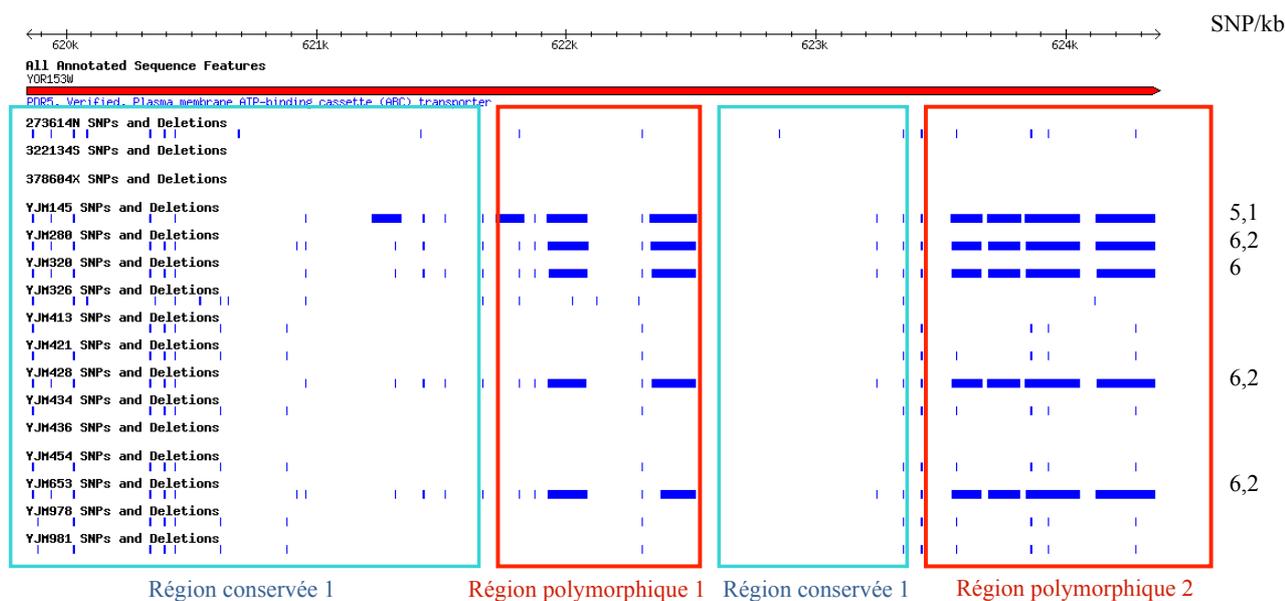
SNP/kb pour les souches YJM280, YJM428 et YJM653. Sur les 14 souches cliniques, 5 présentent un polymorphisme particulièrement important pour le gène *PDR5* (YJM145, YJM280, YJM320, YJM428 et YJM653). Ces 5 souches ont des origines quelque peu différentes. La souche YJM653 a été isolée à partir d'un lavage broncho-alvéolaire à l'hôpital de Stanford (Californie, USA). Les souches YJM428 et YJM280 ont été isolées à partir de liquide d'ascite. La souche YJM320 a été isolée à partir d'une hémoculture à l'hôpital de San Jose (Californie, USA). La souche YJM145 est une souche haploïde isolée à partir de la souche clinique YJM789 dont le génome a déjà été séquencé. YJM145 possède donc la même version allélique de *PDR5* que la souche YJM789 dont nous avons parlé dans l'introduction de ce chapitre. Elle provient d'un patient en phase SIDA (syndrome d'immunodéficience acquise) atteint d'une pneumonie.

#### 1.1.1.1. Analyse de la distribution du polymorphisme du gène *PDR5* des souches cliniques

La question est de savoir si les SNP sont distribués de manière uniforme sur l'ensemble du gène ou s'ils sont regroupés dans certaines régions. Pour y répondre nous avons utilisé les données de polymorphisme issues des expériences d'hybridation sur puces à ADN. L'analyse montre très clairement que les mutations sont principalement localisées au niveau de deux régions bien distinctes (figure 2.7). La première région hautement polymorphe de 661 paires de base est située entre les coordonnées 621843 et 622504 du chromosome XV. La deuxième région hautement polymorphe de 956 paires de base est située entre les coordonnées 623420 et 624376. Ces deux régions représentent un peu plus de 35% (1617 paires de base sur 4535 paires de base) de la totalité du gène.

Puis, nous avons focalisé notre attention sur une possible corrélation entre ces deux régions hautement polymorphes et les domaines protéiques de Pdr5p. Comme mentionné précédemment, la protéine Pdr5p est composée de deux domaines transmembranaires (TMD) alternés avec deux domaines liant l'ATP (NBD) dans l'ordre suivant : NBD1-TMD1-NBD2-TMD2 (figure 2.2). Alors que la région polymorphe 1 est à cheval entre les deux domaines TMD1 et NBD2, la région polymorphe 2 est principalement concentrée sur le domaine TMD2. Il n'y a donc pas de corrélation évidente entre régions polymorphes et domaines protéiques.

La distribution du polymorphisme de YJM280, YJM320, YJM428 et YJM653 suit en très grande partie celle de YJM145 donc de YJM789. Afin de savoir si la version du gène *PDR5* est bien la même pour les 5 souches cliniques nous avons procédé au séquençage de *PDR5*.



**Figure 2.7 :** Représentation du polymorphisme du gène *PDR5* pour les souches cliniques. L'axe des abscisses représente la localisation sur le chromosome XV entre les coordonnées 61900 et 624300. Les traits bleus symbolisent les SNP. Une bande bleue représente de nombreux SNP proches les uns des autres.

### 1.1.1.2. Analyse des séquences du gène *PDR5*

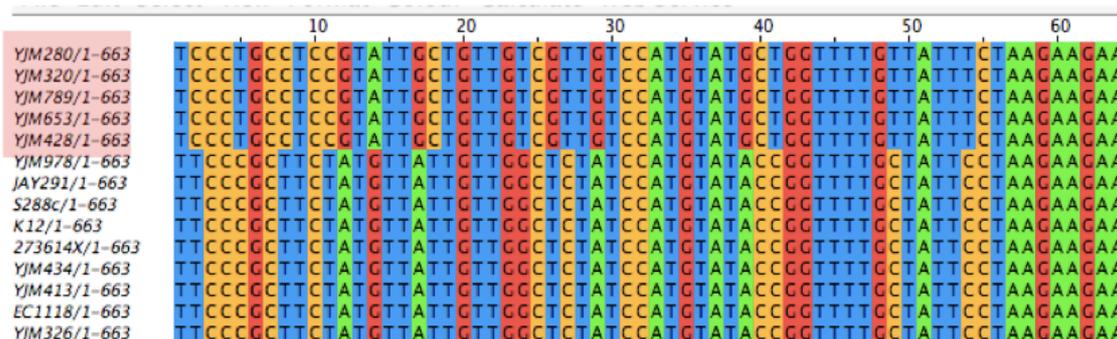
Nous avons procédé au séquençage du gène *PDR5* de 11 souches, dont les 5 souches cliniques, et nous les avons comparé à la séquence du gène des 5 autres souches déjà séquencées (Tableau 2.2). À la vue de la taille importante du gène (4535 pb), le séquençage du gène *PDR5* a été effectué en 7 fragments de 1000 pb qui sont chevauchants.

Afin de pouvoir comparer les séquences entre elles, nous avons procédé à des alignements en utilisant le logiciel ClustalW2. Les alignements nous ont permis de déterminer la nature des changements entre les séquences. Deux types d'alignements ont été réalisés : les alignements de séquences nucléotidiques d'une part et protéiques d'autre part.

La comparaison des séquences nucléotidiques, illustrée dans la figure 2.8, montre très clairement la présence de deux versions du gène dans la collection des souches étudiées. La majorité des souches ont une séquence similaire à celle de la souche de référence S288c. Par contre, les 5 souches cliniques (YJM145, YJM280, YJM320, YJM428 et YJM653) ont une version différente du gène *PDR5*. Il est intéressant de noter que cette version est, à quelques SNP près, identique pour les 5 souches. Le patron de mutation, encore appelé haplotype, est donc conservé entre les 5 souches cliniques. La question qui en découle logiquement concerne l'origine de cette version du gène *PDR5* présente chez les 5 souches cliniques. Alors qu'elles ont des origines différentes, elles partagent la même version du gène *PDR5*.

Catégorie de souches	Nom et origine des souches	Origine écologique	Origine géographique
Souches séquencées (séquence disponible sur SGD)	S288c (laboratoire)	laboratoire	/
	YJM789 (clinique)	clinique	Etats-Unis
	RM11 (vigne)	vigne	Californie
	JAY291 (distillerie)	distillerie	/
Autre espèce séquencée	<i>Saccharomyces paradoxus</i> (outgroup)		
Souches cliniques hautement polymorphes	YJM145	clinique	Etats-Unis
	YJM280	clinique	Etats-Unis
	YJM320	clinique	Etats-Unis
	YJM428	clinique	Etats-Unis
	YJM653	clinique	Etats-Unis
Autres souches	K12	saké	Japon
	YJM326	clinique	Etats-Unis
	YJM413	clinique	Etats-Unis
	273614X	clinique	Royaume Uni
	YJM434	clinique	Europe
	YJM978	clinique	Italie

**Tableau 2.2 :** Souches dont le gène *PDR5* a été séquencé.



**Figure 2.8 :** Exemple d'alignement de la région hautement polymorphe 1 sur Jalview. Les souches cliniques polymorphes pour *PDR5* ont été regroupées lors du processus d'alignement sur ClustalW (cadre rouge).

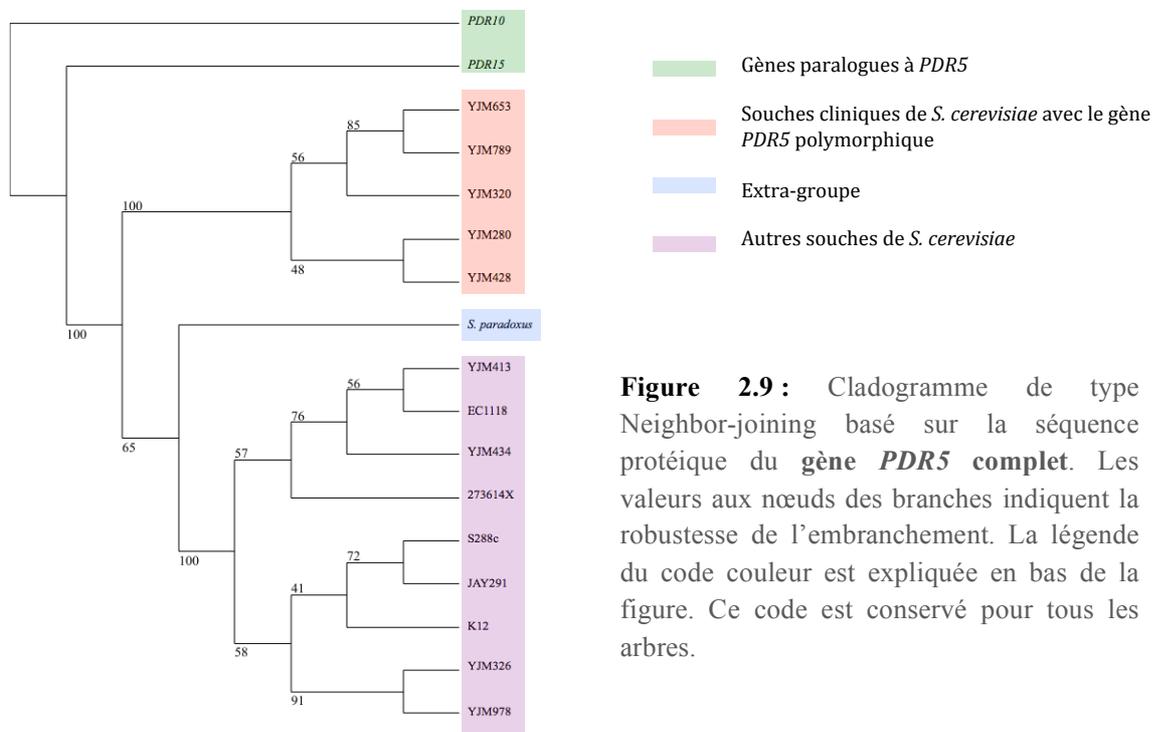
### 1.1.1.3. Hypothèses sur l'origine de la version allélique de *PDR5* des 5 souches cliniques

Trois hypothèses permettraient d'expliquer l'apparition de l'allèle présent chez les 5 souches cliniques. La première hypothèse serait l'accumulation de mutations de manière indépendante au cours du temps. La seconde hypothèse permettant d'expliquer la présence des deux régions polymorphes serait un événement de recombinaison entre le gène *PDR5* et un des gènes paralogues à celui-ci, *PDR10* ou *PDR15*. Enfin, la troisième hypothèse consisterait en une introgression c'est-à-dire un mouvement du gène *PDR5* depuis une autre espèce par hybridation.

Afin de tester ces différentes hypothèses, nous avons procédé à des comparaisons de séquences complémentaires. Pour vérifier l'hypothèse d'un événement de recombinaison, nous avons comparé les séquences des gènes *PDR5* avec les séquences des gènes *PDR10* et *PDR15* de la souche de référence S288c. Les gènes *PDR10* et *PDR15* sont des paralogues de *PDR5*, localisés respectivement sur les chromosomes XV et IV. Comme *PDR5*, ils codent des transporteurs intervenant dans la résistance multiple. De plus, les facteurs de transcription Pdr1p et Pdr3p sont communs à ces trois gènes et participent à la régulation de leur expression. Afin de tester l'hypothèse d'une introgression, nous avons comparé les séquences obtenues avec la séquence du gène *PDR5* de l'espèce *Saccharomyces paradoxus* qui est la plus proche de *S. cerevisiae*.

Avec l'ensemble de ces séquences, nous avons construit des arbres phylogénétiques dans le but de retracer l'évolution et les liens généalogiques au sein d'une espèce et entre les espèces. Pour être phylogénétique, l'arbre doit notamment être enraciné par l'extra-groupe qui est représenté dans cette étude par *S. paradoxus*. Pour l'analyse du gène *PDR5*, les « caractères » sont les distances entre les séquences. La distance est calculée par le nombre de nucléotides ou d'acides aminés différents entre deux séquences, divisé par le nombre total de sites. Des arbres basés sur les alignements de séquences nucléotidiques et protéiques ont été construits. Nous allons nous focaliser sur les arbres basés sur les séquences protéiques car ce sont elles qui ont une répercussion directe sur la protéine.

#### 1.1.1.4. Analyse de la séquence protéique de Pdr5p



**Figure 2.9 :** Cladogramme de type Neighbor-joining basé sur la séquence protéique du gène *PDR5* complet. Les valeurs aux nœuds des branches indiquent la robustesse de l'embranchement. La légende du code couleur est expliquée en bas de la figure. Ce code est conservé pour tous les arbres.

Pour l'arbre construit à partir des séquences complètes de *PDR5* (figure 2.9), nous observons que les branches des gènes *PDR10* et *PDR15* sont les premières à se détacher du taxon commun. Le taxon commun est constitué de deux clades : un premier contenant les 5 souches cliniques dont le gène *PDR5* est polymorphe et un deuxième dans lequel les autres souches analysées de *S. cerevisiae* sont retrouvées. L'embranchement de *S. paradoxus* est situé entre ces deux clades. La robustesse de l'embranchement séparant les deux clades est de 100.

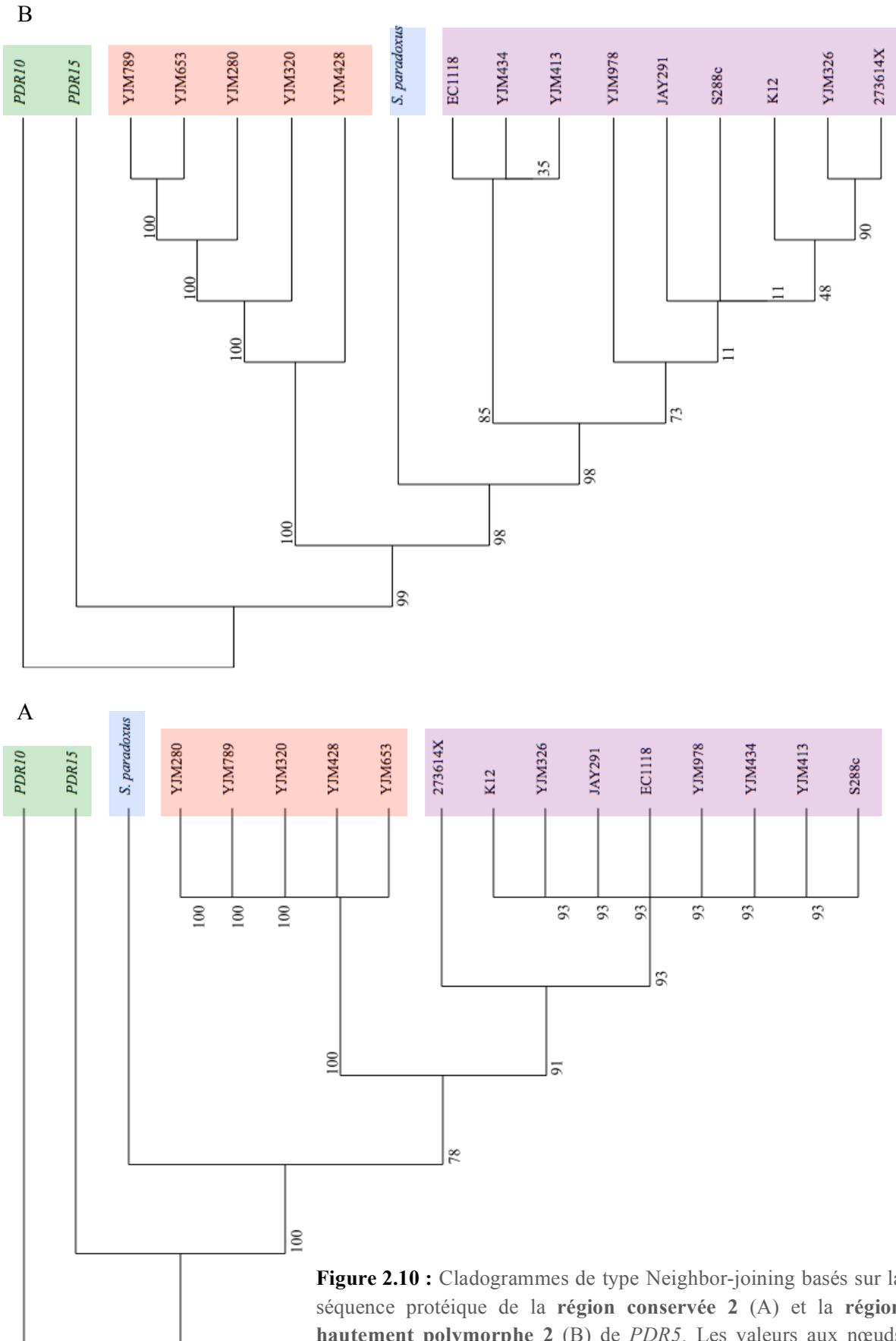
#### 1.1.1.5. Analyse par région de la séquence protéique de Pdr5p

Nous avons ensuite analysé séparément les séquences protéiques des régions conservées et celles des régions hautement polymorphes.

Pour les deux arbres basés sur les régions conservées 1 et 2 (Fig. 2.10A), il y a un changement majeur par rapport à l'arbre précédent. En effet, l'embranchement de *S. paradoxus* se sépare du taxon commun juste après les embranchements des deux gènes paralogues. La distance évolutive au niveau des régions conservées est plus grande entre les deux espèces du genre *Saccharomyces* qu'entre toutes les souches de *S. cerevisiae*.

Pour les arbres basés sur les 2 régions hautement polymorphes (Fig. 2.10B), l'embranchement de *S. paradoxus* est de nouveau situé entre les deux clades des souches de *S. cerevisiae*. La robustesse de l'embranchement des 5 souches cliniques est proche ou égale à 100 selon l'arbre. L'arbre construit sur les séquences entières de *PDR5* reflète majoritairement l'arbre des 2 régions hautement polymorphes.

L'ensemble de ces arbres permet de tirer des conclusions claires quant aux différentes hypothèses. Du fait de l'absence des gènes *PDR10* et *PDR15* dans le taxon commun aux souches polymorphes, l'hypothèse d'une recombinaison entre le gène *PDR5* et ses gènes paralogues est à rejeter. De plus, la position du gène *PDR5* de *S. paradoxus* montre l'importance de la divergence des souches cliniques comme YJM789 avec S288c. L'accumulation de mutations indépendantes semble peu probable dans ce cas. Finalement le gène *PDR5* de *S. paradoxus* n'étant pas dans le taxon des 5 souches cliniques, il n'y a pas eu d'introgression à partir de cette espèce. Toutefois l'hypothèse d'une introgression à partir d'une autre espèce de levure plus distante phylogénétiquement reste plus que probable.



### 1.1.1.6. Recherche sur l'origine des régions hautement polymorphes

Les régions hautement polymorphes 1 et 2 ont fait l'objet d'une recherche de type Blast sur des banques de données (Génolevures, SGD et GenBank). Nous n'avons pas retrouvé de séquence plus proche que celle de S288c. L'origine de la version du gène *PDR5* des 5 souches cliniques reste à ce jour inconnue. L'introgession a pu se faire à partir d'une espèce dont le génome n'a pas encore été séquencé. Lors des infections à *S. cerevisiae*, des copathogènes sont souvent détectés. Cette cohabitation avec d'autres organismes pourrait donner lieu à des échanges de matériel génétique par transfert horizontal. Ces organismes peuvent donc constituer une piste de recherche quant à l'origine de la version de *PDR5* des souches cliniques. Malheureusement, dans la plupart des cas, ces informations sont manquantes dans le dossier des patients, le pathogène n'est pas identifié, ou encore le copathogène n'est pas isolé afin d'approfondir les recherches.

En conclusion, la version allélique présente chez les 5 souches cliniques (YJM145, YJM280, YJM320, YJM428 et YJM653) a vraisemblablement apporté un avantage sélectif permettant son maintien au sein de l'espèce. Les changements nucléotidiques quelles que soient leurs origines aboutissent vraisemblablement à un gain de fonction au niveau du transporteur. Une des pressions environnementales communes à ces souches est la présence d'antifongiques. Cette observation est d'autant plus intéressante dans la mesure où le gène *PDR5* est un des principaux gènes intervenant dans la résistance multiple aux antifongiques.

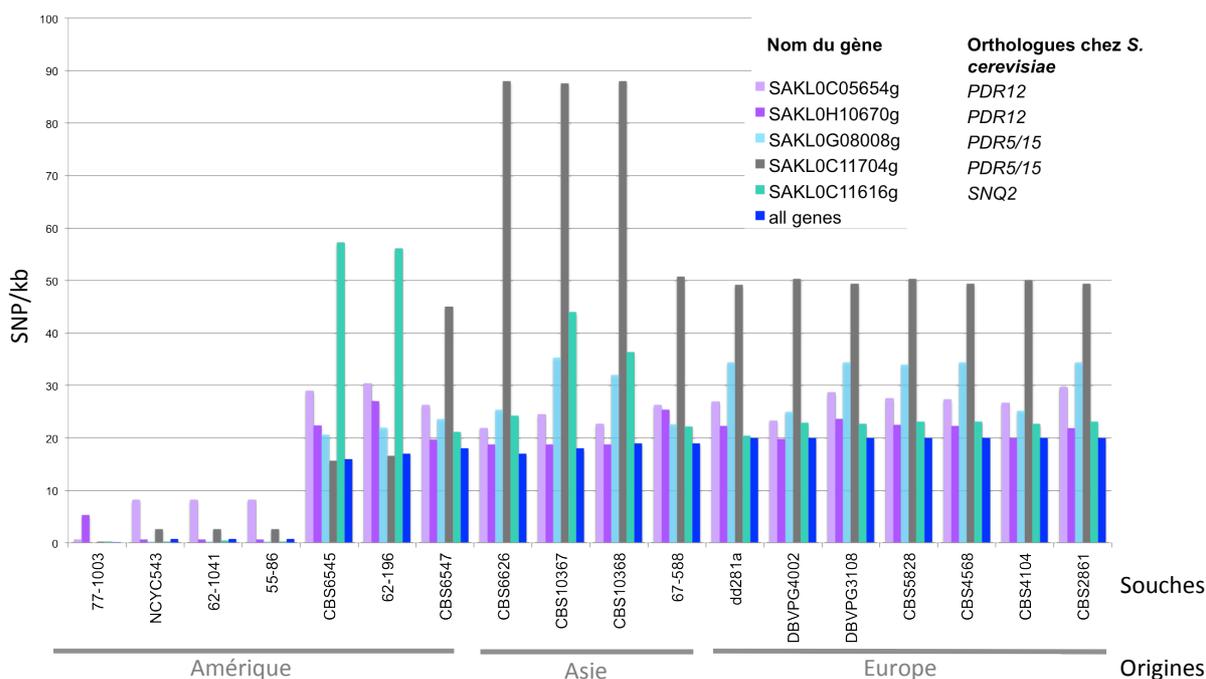
## 1.2. Analyse des gènes *PDR* chez *L. kluyveri*

Grâce au reséquençage des génomes des souches de *L. kluyveri* qui a été réalisé au sein de l'équipe, il a été possible de déterminer le polymorphisme des gènes de résistance multiple directement par comparaison de séquence. L'étude s'est basée sur 18 souches séquencées. Brièvement, les séquences des gènes d'intérêts ont été recherchées par Blast contre les contigs issus de l'assemblage des génomes. Un assemblage manuel a été réalisé lorsque les souches présentaient plusieurs contigs par gène. L'alignement des gènes grâce au logiciel Jalview (Waterhouse *et al.* 2009) a ensuite permis le dénombrement des SNPs par gène et par souche. La traduction des séquences nucléotidiques en séquence protéique a de plus permis d'identifier les mutations synonymes des non synonymes.

Comme indiqué dans l'introduction générale, les souches les plus proches génétiquement sont celles provenant d'une même origine géographique. Pour analyser la variabilité des gènes *PDR*, les souches ont donc été regroupées par origine géographique (figure 2. 11).

Contrairement aux gènes *PDR* de *S. cerevisiae* qui apparaissent dans l'ensemble très conservés par rapport au reste du génome, les souches de *L. kluyveri* présentent au contraire des gènes *PDR* plus polymorphe que l'ensemble du génome. Les souches d'origine américaine se démarquent par leur

faible polymorphisme par rapport à la souche de référence à l'exception du gène *SAKL0CO5654*. Les gènes présentant le plus fort polymorphisme sont les gènes *SAKLC11704* (orthologue de *PDR15*) et *SAKL0C11616* (orthologue de *SNQ2*). Le polymorphisme est corrélé à l'origine géographique des souches et donc aux groupes basés sur la distance génétique que nous avons définis dans l'introduction. Le polymorphisme le plus élevé est celui du gène *SAKLC11704* chez les souches d'Asie pour lequel il avoisine les 90 SNP/kb.



**Figure 2.11 :** Polymorphisme en SNP/kb des gènes impliqués dans la résistance multiple chez *L. kluyveri*.

En conclusion de l'analyse des gènes *PDR* chez les deux levures, les gènes impliqués dans la résistance multiple chez *S. cerevisiae* sont en majorité plus conservés que l'ensemble du génome mis à part le gène *PDR5*. Il a été mis en évidence une deuxième version de ce gène chez des souches cliniques. Ces souches pourtant éloignées phylogénétiquement possèdent les mêmes mutations localisées au niveau du gène *PDR5*. Cette version allélique proviendrait probablement d'une introgression d'origine encore inconnue.

Les gènes de résistance multiple chez *L. kluyveri* sont au contraire plus polymorphes que les autres phases codantes du génome. Est-ce que cette grande plasticité des gènes *PDR* chez *L. kluyveri* aurait pour conséquence une plus grande diversité de résistance ? C'est à cette question que nous allons répondre dans la suite de ce travail.

## 2. Étude de la résistance aux antifongiques au sein des deux espèces

La résistance des souches de deux collections de *S. cerevisiae* et *L. kluyveri* a été analysée afin de cartographier à plus long terme les régions intervenant dans la résistance grâce à des analyses de liaison. Après des tests préliminaires, nous avons choisi de nous focaliser sur les antifongiques de type azolés pour plusieurs raisons. Ce sont les antifongiques actuellement les plus utilisés, ils sont prescrits en cas d'infection à *S. cerevisiae* dont en particulier le fluconazole, et la variabilité de résistance au sein des deux espèces de levures est grande. Parmi les antifongiques azolés, nous en avons sélectionné 5 : le kétoconazole, le clotrimazole et la miconazole appartenant à la classe des imidazoles et le fluconazole et l'itraconazole faisant partie de la classe des triazoles (figure 2.12). Deux concentrations ont été choisies et testées par antifongique. Elles correspondent à celles pour lesquelles la plus grande variation phénotypique est observée au sein des espèces.

### Les antifongiques azolés

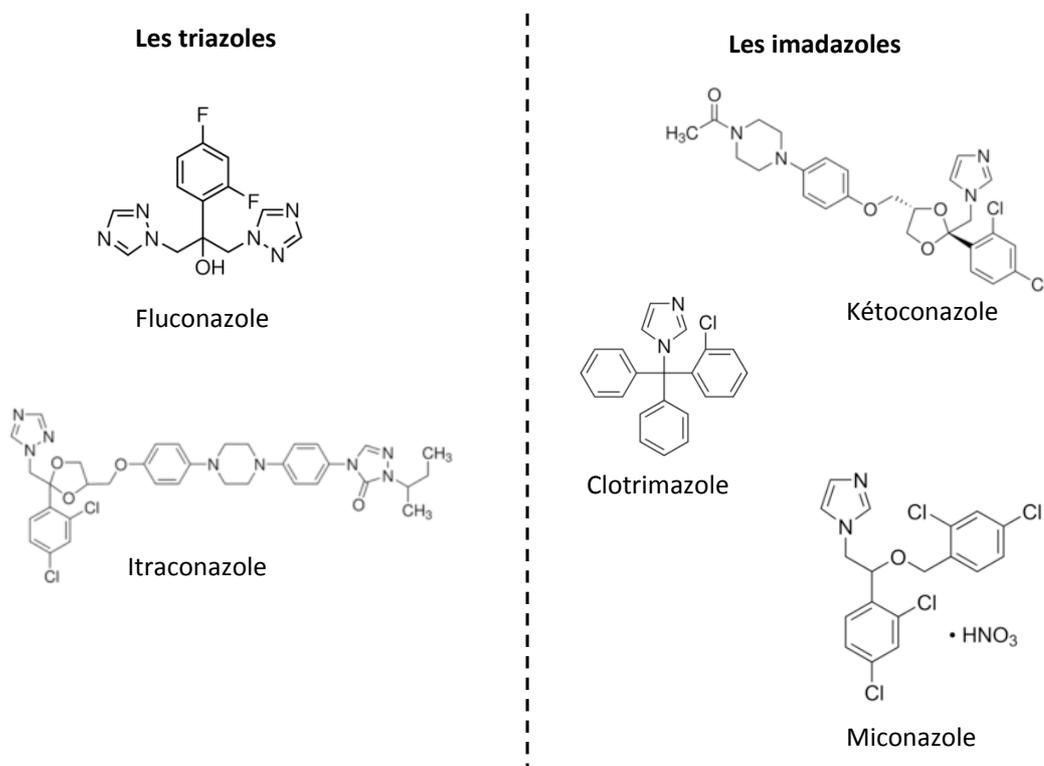


Figure 2.12 : Les antifongiques azolés utilisés dans le criblage des collections.

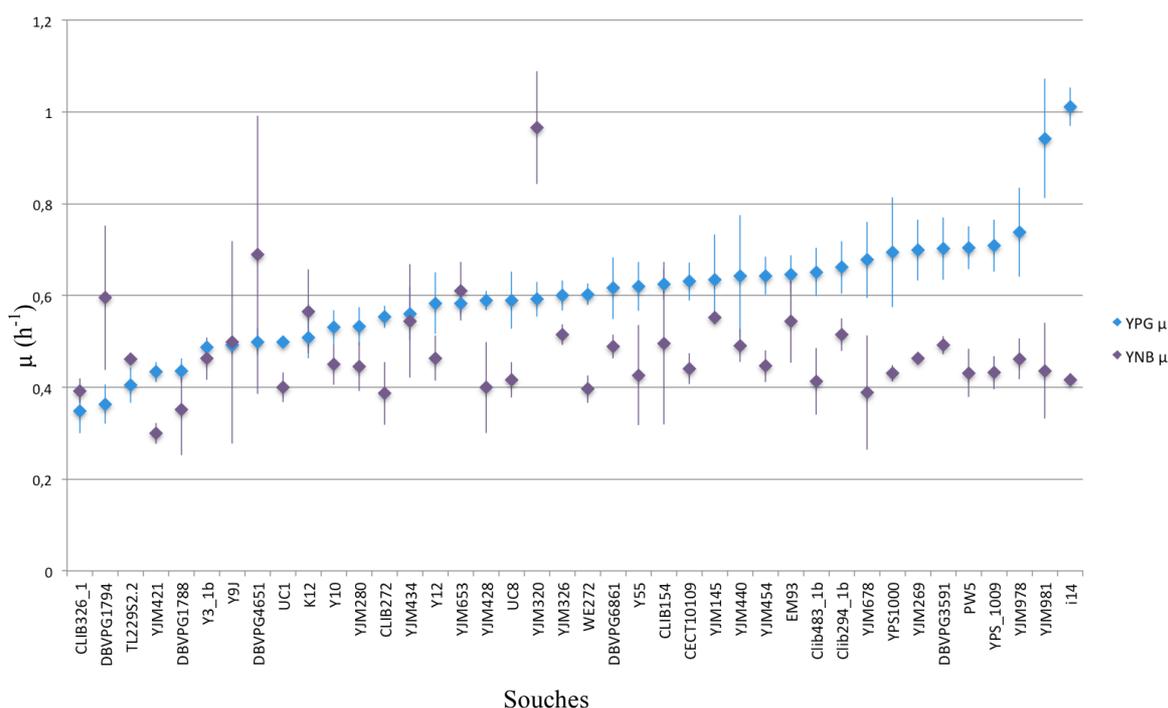
### 2.1. Stratégie de phénotypage par microcultures

Afin de tester la résistance des souches des 2 collections dans 11 conditions différentes (milieu YPG sans antifongique, milieux YPG avec 5 antifongiques différents chacun testé dans deux concentrations) et ce en triplicat, une stratégie de phénotypage à haut-débit a été mise au point. Nous avons choisi d'utiliser la stratégie de suivi de microcultures qui a pour avantage d'être une approche

très sensible, reproductible et qui permet de mesurer avec précision plusieurs caractères quantitatifs comme la vitesse spécifique de croissance, la biomasse et la phase de latence.

### 2.1.1. Choix du milieu de référence

Les deux milieux de croissance YPG (milieu complet) et YNB (milieu minimum) ont été envisagés comme milieux de référence pour ces expériences de microcultures. Afin de choisir celui qui donnera les résultats les plus reproductibles, nous avons testé la croissance en triplicat d'une quarantaine de souches de *S. cerevisiae* en YPG et en YNB (figure 2.13). Pour analyser la croissance, nous avons calculé les vitesses spécifiques de croissance  $\mu$  pour chaque culture. La moyenne des  $\mu$  sur YPG ( $0,60 \text{ h}^{-1}$ ) est plus élevée que la moyenne des  $\mu$  sur YNB ( $0,48 \text{ h}^{-1}$ ), ce qui signifie que les souches poussent en moyenne plus vite de 20% sur YPG que sur YNB. La moyenne des écarts type des  $\mu$  en milieu YPG est de  $0,052$  et en YNB elle est de  $0,070$ . La croissance en milieu riche est donc plus reproductible que celle en milieu minimum. Nous avons donc sélectionné le milieu YPG comme milieu de référence pour les microcultures. Il sera supplémenté des antifongiques choisis dans les différentes conditions.



**Figure 2.13** : Comparaison des vitesses spécifiques de croissance  $\mu$  ( $\text{h}^{-1}$ ) de 40 *S. cerevisiae* sur YPG et sur YNB.

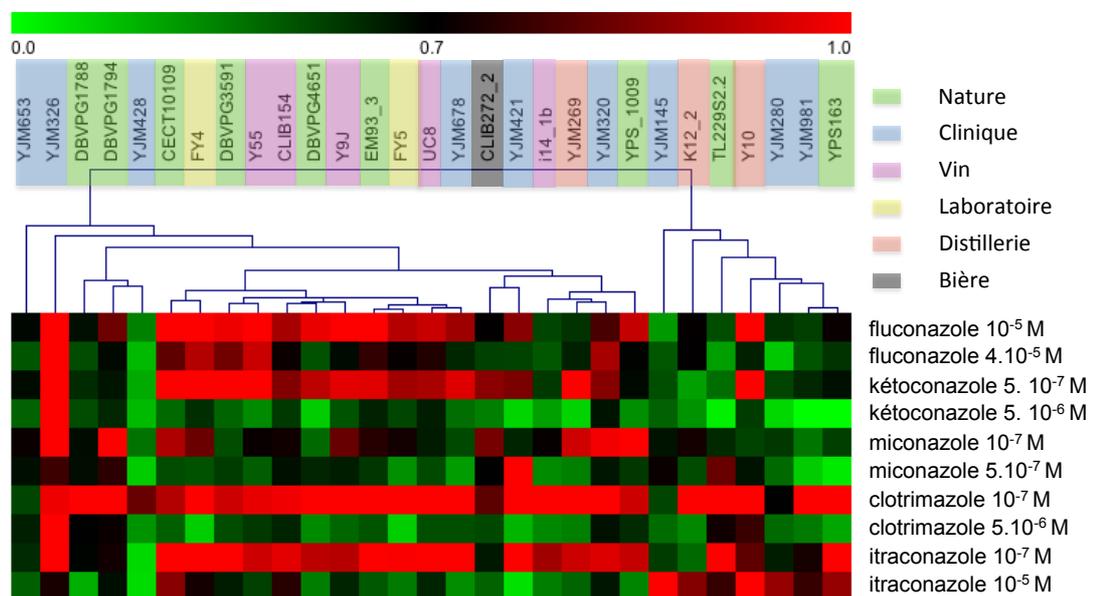
### 2.1.2. Estimation de la résistance

Afin de mesurer la résistance, nous avons réalisé les cultures des souches en milieu complet et en milieu complet supplémenté d'antifongiques, puis nous avons comparé les vitesses spécifiques de croissances ( $\mu$ ). C'est le rapport  $\mu_{\text{YPG}} / \mu_{\text{Antifongique}}$  qui va nous permettre d'estimer la résistance. Si une

souche pousse aussi bien avec que sans l'antifongique ajouté, le rapport obtenu sera proche ou égal à 1. Dans ce cas, la souche sera dite résistante. À l'inverse, plus une souche sera sensible, plus le rapport sera alors proche de 0.

## 2.2. Variation intraspécifique de la résistance aux composés azolés chez *S. cerevisiae*

Un premier criblage des souches de la collection, nous a permis de sélectionner 28 souches haploïdes sur des critères de reproductibilités. En effet, pour une partie des souches de *S. cerevisiae* les cultures en milieu liquide ne sont pas homogènes. Les souches de brasserie sont notamment caractérisées par leur capacité à floculer ce qui leur permet de sédimenter à la fin des processus de fermentation et de pouvoir être facilement séparée de la bière (Gibson *et al.* 2007). Cette propriété n'est pas compatible avec les microcultures car la présence d'agrégats fausse la lecture de l'absorbance et les données mesurées sont inexploitable. Les 28 souches ont été cultivées sur YPG et YPG supplémenté d'antifongiques. Les résistances à chaque antifongique et pour chaque souche sont présentées dans la figure 2.14. Les souches ont été classées selon leur profil phénotypique par un regroupement hiérarchique en utilisant une corrélation de Pearson.



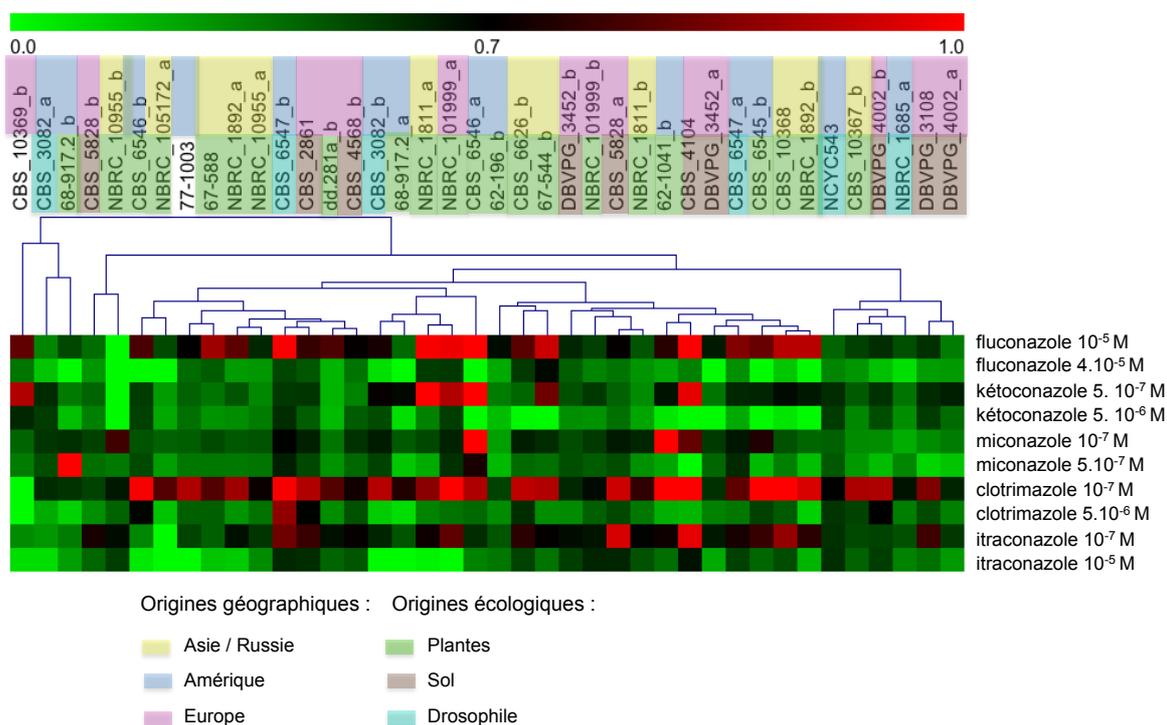
**Figure 2.14 :** Résistance aux composés azolés de 28 souches de *S. cerevisiae*. La résistance est déterminée par le rapport obtenu entre la vitesse spécifique de croissance sans antifongique et celle avec antifongique. La résistance est symbolisée par un gradient de couleur rouge (souches résistantes) à vert (souches sensibles). Les souches sont regroupées et hiérarchisées par profil de résistance grâce au calcul de la corrélation de Pearson. L'origine écologique des souches est indiquée par le surlignage en couleur des noms de souches.

Une variabilité de la résistance est observée au sein de la collection de souches de *S. cerevisiae*. Nous pouvons observer que les souches provenant d'un même environnement écologique ne sont pas regroupées. Les souches ayant les profils de résistance aux composés azolés les plus proches ne sont donc pas les souches les plus proches génétiquement. Certaines souches sont extrêmement sensibles en général aux composés azolés comme les souches cliniques YJM428 et YJM280 d'autres au contraire sont résistantes comme les souches cliniques YJM326 et YJM320. Ces souches au profil phénotypique extrême seront retenues pour être utilisées dans des analyses de groupe de ségrégants pour déterminer les origines génétiques de ces variations phénotypiques.

### **2.3. Variation intraspécifique de la résistance aux composés azolés chez *L. kluyveri***

Au total, la résistance de 40 souches de *L. kluyveri* aux mêmes composés azolés a été testée (figure 2.15). D'une manière générale, les souches de *L. kluyveri* sont plus sensibles que les souches de *S. cerevisiae* pour les mêmes concentrations d'antifongiques. De plus, la variabilité de résistance observée au sein de la collection est moins importante. En effet, pour une condition donnée, la majorité des souches sera sensible comme par exemple pour la croissance sur miconazole ( $10^{-7}$  et  $5.10^{-7}$  M). Pour d'autres conditions au contraire, telle que la croissance en présence de clotrimazole à  $10^{-7}$  M, la majorité des souches sera résistante. Le temps de génération moyen est d'environ 1h10 sur YPG, ce qui correspond à une vitesse spécifique de croissance moyenne de  $0,59 \text{ h}^{-1}$ . L'analyse des profils phénotypiques ne conduit pas à un regroupement des souches en fonction de leur origine écologique, ni de leur origine géographique. Les souches génétiquement proches n'ont donc pas forcément le même profil de résistance vis à vis des composés azolés.

En conclusion, la variabilité de résistance mesurée au sein des deux espèces, n'est ni corrélée à l'origine géographique, ni corrélée à l'origine écologique des souches. De plus chez *S. cerevisiae*, les 5 souches cliniques ayant une version allélique de *PDR5* polymorphe ne sont pas regroupées au sein d'un même cluster.



**Figure 2.15 :** Résistance aux composés azolés de 40 souches de *L. kluyveri*. La résistance est déterminée par le rapport entre la vitesse spécifique de croissance sans antifongique et celle avec antifongique. La résistance est symbolisée par un gradient de couleur rouge (souches résistantes) à vert (souches sensibles). Les souches sont regroupées et hiérarchisées par profil de résistance grâce au calcul de la corrélation de Pearson. L'origine écologique et géographique des souches est indiquée par le surlignage en couleur des noms de souches.

### 3. Détermination de l'origine génétique de la résistance par analyse de pool de ségrégants

Une stratégie d'analyse de groupe de ségrégants a été choisie afin de déterminer les origines génétiques des variations de résistance observées au sein des collections de *S. cerevisiae* et de *L. kluyveri*.

#### 3.1. Design expérimental

##### 3.1.1. Choix des antifongiques

Nous avons choisi de réaliser cette analyse, à la fois pour le fluconazole et pour le kétoconazole, d'une part à cause de leur utilisation massive dans les traitements médicaux contre les infections et, d'autre part, car chacun appartient à une classe différente de composés azolés, les imidazoles pour le kétoconazole et les triazoles pour le fluconazole.

### 3.1.2. Choix des parents

Le choix des parents est déterminant puisque c'est le fond génétique des souches qui déterminera les variations phénotypiques observées. La sélection des souches parentales s'est donc basée sur leur profil de résistance aux composés azolés. D'autres critères ont été fixés pour la sélection des parents : les parents doivent pouvoir se croiser, permettre l'obtention de nombreux descendants viables et conduire à la formation de descendants possédant des points de recombinaison. Ce dernier point est important pour que la cartographie de QTL soit possible. Il est également nécessaire que la divergence génétique entre les parents soit suffisante pour permettre une répartition des sites polymorphes sur l'ensemble du génome afin de distinguer efficacement les génomes des parents.

#### 3.1.2.1. *S. cerevisiae*, des souches parentales aux phénotypes extrêmes

Chez *S. cerevisiae* les souches sélectionnées sont YJM326 (*MATa*) qui est résistante à  $10^{-5}$  et  $4.10^{-5}$  M de fluconazole, et à  $5.10^{-7}$  et  $5.10^{-6}$  M de kétoconazole, ainsi que la souche YJM280 (*MAT $\alpha$* ) qui est sensible à ces mêmes concentrations. Ces souches sont d'origine clinique, c'est pourquoi pour des raisons de clarté, ce couple sera appelé **Sace\_clinique**. Le fait d'utiliser des souches cliniques est un atout supplémentaire, car les souches ont été soumises à des stress de type antifongique et elles ont été en contact avec d'autres micro-organismes, ce qui permet potentiellement l'échange d'informations génétiques par le biais de transferts horizontaux.

La particularité de ces deux souches est que la sensibilité pour YJM280 ou la résistance pour YJM326 ne se limite pas au fluconazole et kétoconazole mais s'étend aussi au miconazole, clotrimazole et itraconazole (figure 2.14).

#### 3.1.2.2. *L. kluyveri*, une sélection de parents aux phénotypes intermédiaires

La collection de *L. kluyveri*, constituée au début de ce travail, marque les débuts de l'analyse de la variation intraspécifique de cette espèce. La diversité génétique des souches de la collection qui s'étend jusqu'à 2,73% est largement supérieure à celle observée chez *S. cerevisiae* qui atteint un maximum de 0,5%. Au début du projet d'analyse de pool de ségréants, notre hypothèse était que les souches devaient avoir une divergence génétique inférieure à 1% pour qu'il n'y ait pas une baisse de viabilité de la descendance. En d'autres termes, nous pensions que les souches trop divergentes génétiquement avaient une descendance dont la viabilité était inférieure à 90%. Des croisements basés sur le profil phénotypique des souches mais surtout sur la divergence génétique entre elles ont donc été testés (annexe 2.1). Plusieurs problèmes sont apparus lors des différentes tentatives de croisements : 1) tous les croisements n'ont pas abouti à la formation de zygotes 2) tous les zygotes isolés n'étaient pas viables et 3) tous les zygotes viables ne sporulaient pas dans les conditions utilisées au laboratoire.

### **Le couple NBRC1811a et NBRC10955b**

Le croisement entre les souches NBRC1811a (*MATa*) et NBRC10955b (*MATa*) a abouti à la formation d'un zygote viable qui présentait une efficacité de sporulation élevée. De plus, à partir des tétrades de nombreux descendants viables ont pu être obtenus. Entre les deux souches, 90 387 sites polymorphes sont dénombrés ce qui correspond à une divergence génétique de 0,8%. Il y aura donc 1 SNP tous les 125 pb. Ce couple sera annoté **Lakl\_Japon** en référence à leur origine japonaise. Elles sont issues de feuilles en décomposition.

### **Le couple CBS 5828a et dd281a\_b**

Un autre couple de souches a été sélectionné chez l'espèce *L. kluyveri*. Il s'agit des souches CBS 5828a (*MATa*) et dd281a\_b (*MATa*) pour lesquelles la divergence génétique est moins élevée que le couple Lakl\_Japon. La divergence génétique entre ces souches est de 0,25% avec la détection de 28 475 sites polymorphes. Il y aura donc 1 SNP tous les 400 pb. Ce couple sera annoté **Lakl\_Europe** en référence à leur origine européenne.

### **3.1.3. Génération des ségrégants**

Pour obtenir une collection d'environ 500 descendants, pour chacun des 3 couples deux méthodologies sont possibles :

- 1) l'obtention de ségrégants de manière ordonnée où tous les ségrégants viables issus d'un même asque seront isolés par microdissection
- 2) l'obtention des ségrégants en vrac par lyse mécanique des tétrades.

Les descendants du couple Sace\_clinique ont été générés par dissection de tétrades. Par contre, la technique classique de microdissection des asques de *S. cerevisiae* n'est pas utilisable chez *L. kluyveri*. Cela est notamment dû à l'inefficacité des enzymes utilisées classiquement pour digérer la paroi de l'asque (zymolyase, lyticase et autres). Ce n'est que récemment qu'une méthodologie de dissection a été mise au point au laboratoire. Les descendants des couples Lakl ont donc été obtenus par la méthode des spores en vrac.

### **3.1.4. Phénotypage des ségrégants**

Au total, près de 810 ségrégants ont été isolés et phénotypés pour le couple Sace\_clinique, 550 pour le couple Lakl\_Japon et 950 pour Lakl\_Europe. Nous avons testé pour chacun d'eux, la résistance à  $4.10^{-5}$  M de fluconazole et  $5.10^{-6}$  M de kétoconazole par des croissances en microcultures. La vitesse spécifique de croissance de chaque ségrégant, dans chaque condition a été déterminée en duplicats. Nous avons choisi de ne constituer que des groupes de souches résistantes car ils sont suffisants pour déterminer les origines génétiques de la résistance.

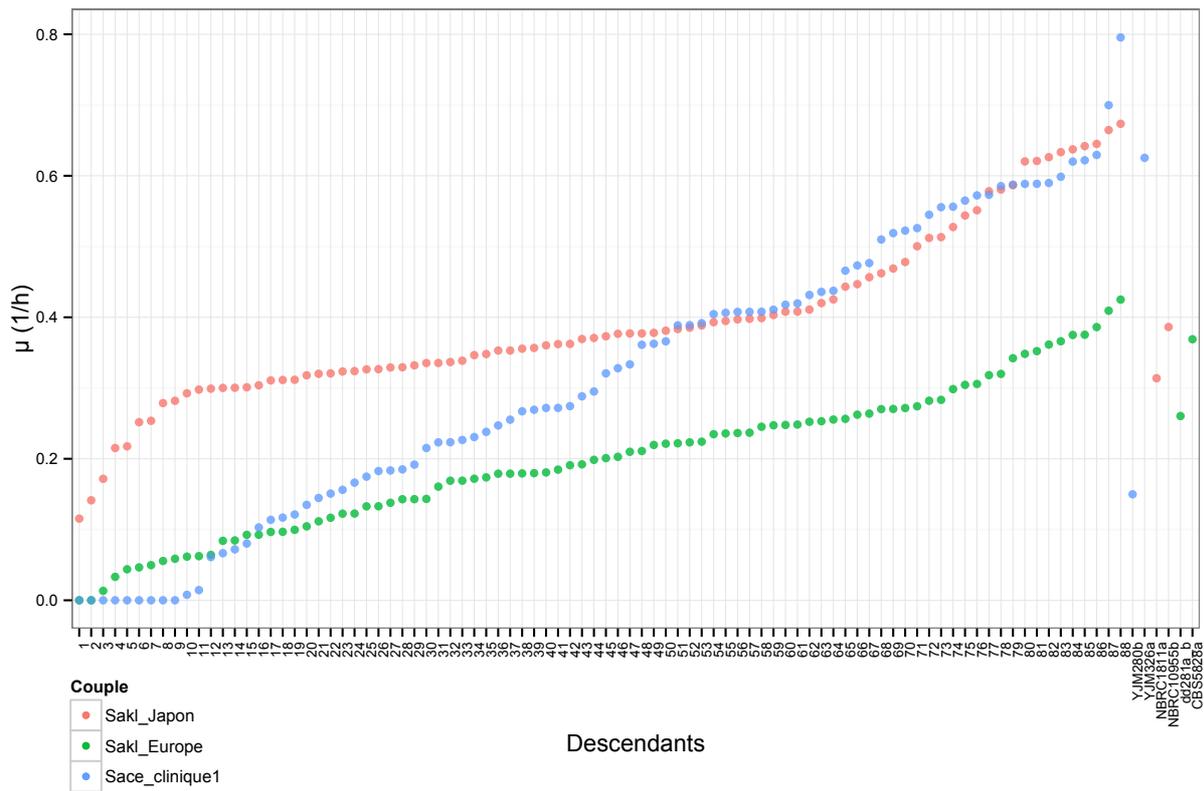
Chez *S. cerevisiae*, les souches parentales présentent chacune un phénotype extrême. La résistance et sensibilité des descendants ont été définies par rapport aux vitesses spécifiques des parents. Les ségrégants ayant une vitesse spécifique de croissance supérieure ou égale à la souche parentale résistante seront résistants et à l'inverse, les ségrégants possédant une vitesse spécifique inférieure à la souche parentale sensible seront sensibles.

Chez *L. kluyveri*, les souches parentales choisies possèdent un phénotype de résistance intermédiaire. Les parents ne peuvent pas conséquemment servir de référence pour la résistance. Nous avons sélectionnés au sein de la descendance, 10% des ségrégants les plus résistants pour le couple Lakl\_Japon et 5% des plus résistants pour le couple Lakl\_Europe. Un pourcentage différent a été choisi entre les deux couples car chronologiquement le couple Lakl Japon a été réalisé, séquencé et analysé avant le couple Lakl\_Europe. C'est donc dans un souci d'amélioration de la précision des résultats que la sélection des résistants a été plus stricte dans la 2<sup>ème</sup> étude. Il a donc fallu générer et analyser plus de ségrégants pour le couple Lakl\_Europe. La vitesse spécifique de croissance des parents sur milieu complet n'étant pas la même pour les deux souches parentales, c'est le rapport des vitesses spécifique de croissance sans et avec antifongique qui a servi de base de comparaison entre les souches. Pour chaque groupe environ 50 ségrégants ont été sélectionnés.

#### **3.1.4.1. La résistance des ségrégants au fluconazole**

Au total, pour les différents croisements, la croissance des 2310 ségrégants a été suivie par microcultures en milieu riche supplémenté de fluconazole à  $4 \cdot 10^{-5}$  M. Les cultures ont été réalisées en duplicas. La vitesse spécifique de croissance moyenne par ségrégant a ensuite été déterminée.

Sur la figure 2.16, les vitesses spécifiques de croissance sont représentées pour un échantillon de 88 ségrégants pour les 3 couples de levures, ainsi que la vitesse spécifique de croissance des souches parentales. Les ségrégants ont été classés par ordre croissant de vitesse spécifique de croissance. Conformément à ce qui a été vu précédemment, les souches parentales Sace\_clinique ont des vitesses spécifiques extrêmes et les souches de *L. kluyveri* ont des phénotypes intermédiaires. Les vitesses spécifiques de croissance des ségrégants des trois expériences forment un trait continu, c'est à dire qu'une multitude de phénotypes intermédiaires sont mesurés. Ces résultats signifient que la résistance au fluconazole est un trait complexe chez *S. cerevisiae* et chez *L. kluyveri*.

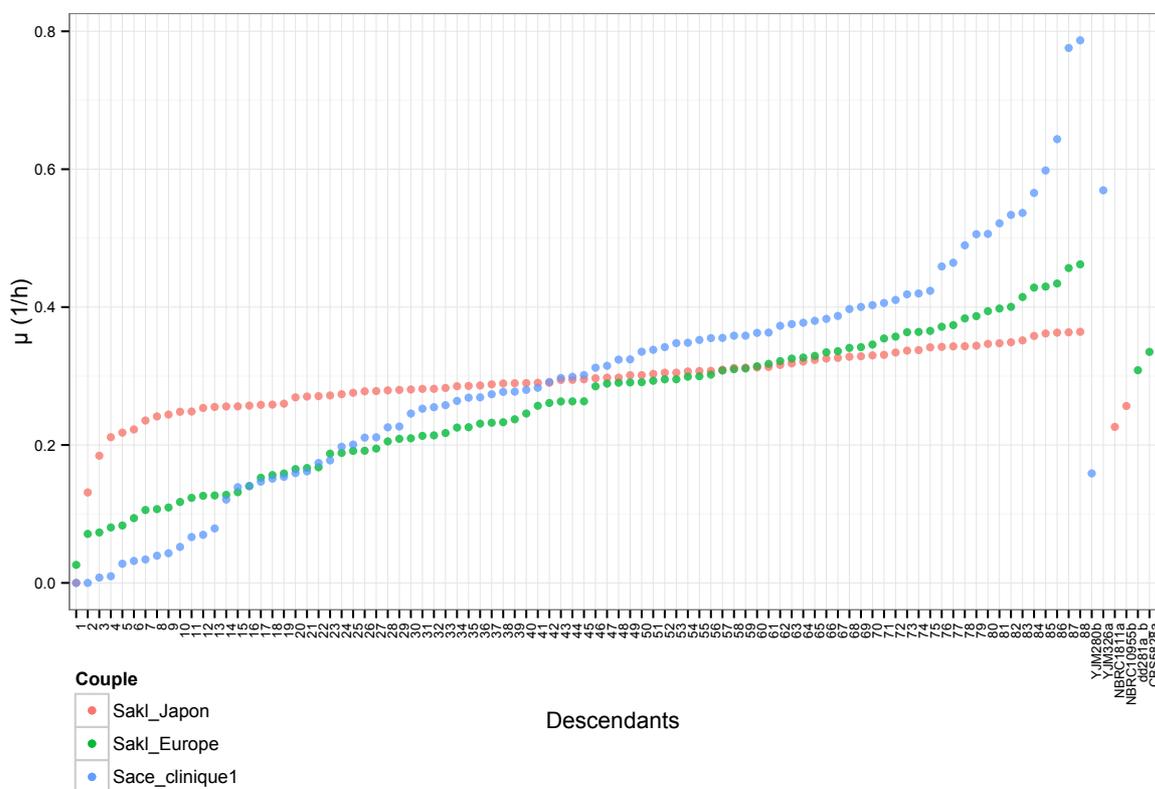


**Figure 2.16 :** Vitesse spécifique de croissance de 88 ségrégants par couple sur milieu complet supplémenté de  $4.10^{-5}$  M de fluconazole. Les descendants issus de chaque couple sont symbolisés par une couleur différente.

### 3.1.4.2. La résistance des ségrégants au kétoconazole

La même stratégie de phénotypage a été utilisée pour étudier la résistance à  $5.10^{-6}$  M de kétoconazole. Les ségrégants de *S. cerevisiae* ont une vitesse spécifique de croissance qui varie de 0 à  $0,79 \text{ h}^{-1}$  (figure 2.17). Pour *L. kluyveri*, ce paramètre varie de 0 à  $0,36 \text{ h}^{-1}$  pour les descendants Saki\_Japon, et de 0 à  $0,46 \text{ h}^{-1}$  pour ceux du couple Laki\_Europe (figure 2.17). Les descendants de tous les couples ont une vitesse spécifique continue, plusieurs gènes sont donc impliqués dans la résistance au kétoconazole dans les deux espèces.

Une partie des ségrégants de *S. cerevisiae* présentent une forte résistance au fluconazole et/ou au kétoconazole. Ce phénomène est peut-être du à l'origine clinique des souches parentales. Les souches de *L. kluyveri* sont quant à elles isolées de niches écologiques naturelles dans lesquelles elles ne sont probablement jamais en contact avec des antifongiques azolés.



**Figure 2.17 :** Vitesse spécifique de croissance de 88 ségrégants par couple sur YPG supplémenté de  $5.10^{-6}$  M de kétoconazole. Les descendants issus de chaque couple sont symbolisés par une couleur différente.

### 3.2. Analyse des séquences des pools de ségrégants du couple Sace\_clinique

Les ADN extraits des groupes de souches résistantes ont été séquencés par une stratégie de séquençage à haut débit, la technique Illumina HiSeq 2000 avec des lectures paires de 100 pb (voir la partie matériels et méthodes). Dans le but de vérifier que les lectures obtenues couvrent l'ensemble du génome, nous avons analysé la couverture des séquençages.

Le séquençage du pool Sace\_clinique en présence de fluconazole a été réalisé deux fois du fait d'une mauvaise couverture du génome. Le pourcentage du génome couvert au minimum 5x par les deux séquençages réunis est de 82,21%. Le séquençage de l'ADN du pool de souches résistantes sur kétoconazole a conduit à de meilleurs résultats puisque 95,62% du génome est couvert au minimum 5x.

### 3.2.1. Stratégie d'analyse du séquençage

Un schéma explicatif de la stratégie globale est donné dans la figure 2.18. Les lectures issues du séquençage des pools de ségréants sont alignées contre le génome de référence S288c à l'aide du logiciel BWA (Li and Durbin 2009). Pour optimiser l'alignement, nous autorisons un maximum de 8 mésappariements et 2 ouvertures de gap par lecture de 100 pb. Ces paramètres ont pour objectif d'éviter un mauvais alignement qui conduirait par la suite à la détection de polymorphismes n'existant pas réellement, appelés faux positifs.

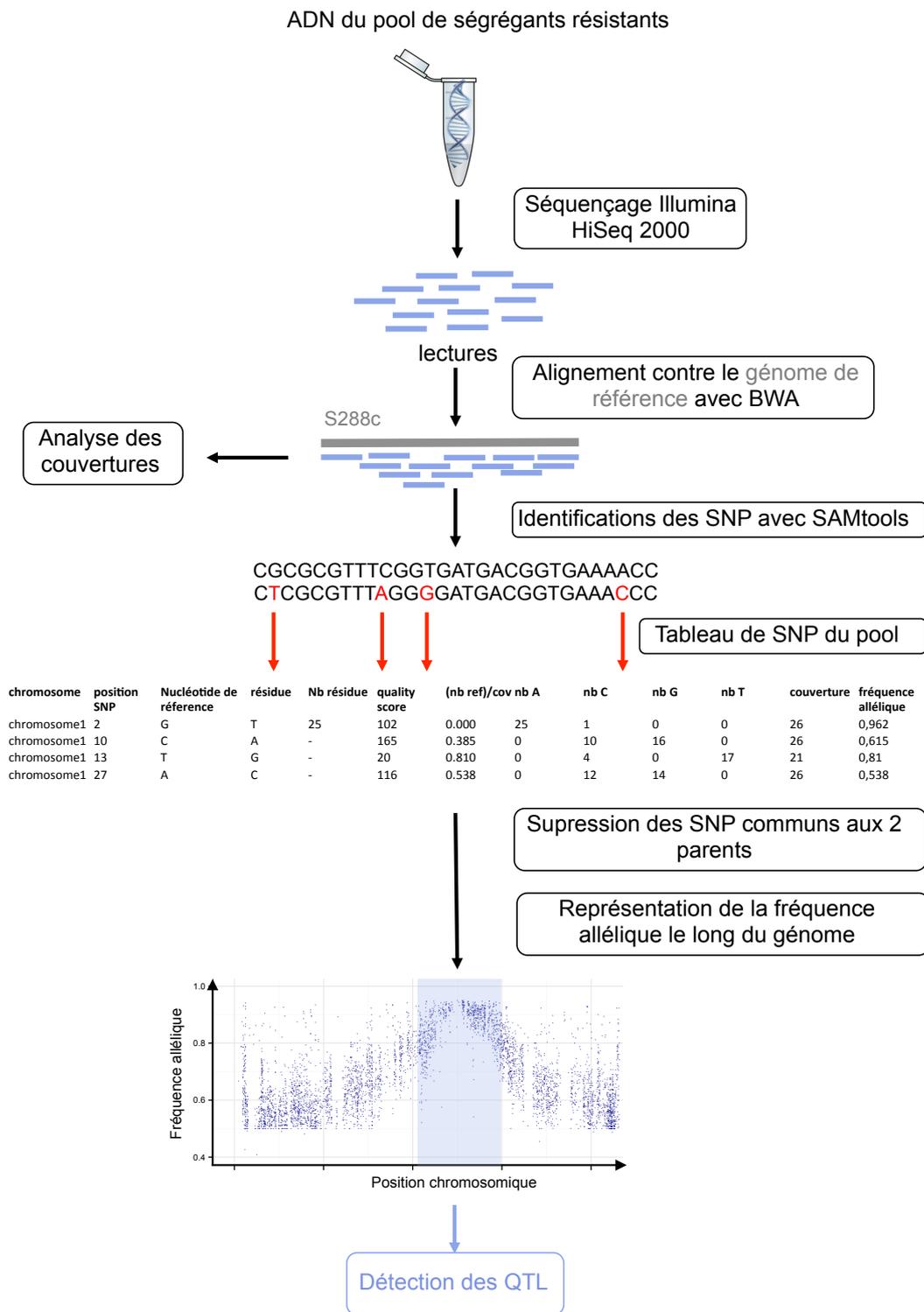
Afin de vérifier que les lectures couvrent l'ensemble du génome, nous avons analysé la couverture tout au long du génome. Pour réaliser cette analyse, la moyenne sur des fenêtres de 1 kb a été calculée, puis représentée graphiquement.

À partir du fichier de sortie du logiciel BWA, le logiciel SAMtools a permis de générer une séquence consensus pour le génome séquencé (Li *et al.* 2009). Le fichier tabulé ainsi créé présente des informations relatives à chaque position séquencée, ce qui va permettre d'établir une liste des positions polymorphes entre le génome de référence et le génome séquencé.

Les SNP identiques identifiés chez les deux parents par rapport au génome de référence ont été retirés de la liste des SNP. En effet, ils ne permettent pas de discriminer les génomes des deux parents et ne sont donc pas informatifs.

Nous avons ensuite comparé le tableau des SNP identifiés pour le pool, aux SNP des génomes parentaux et ce, toujours par rapport à la souche de référence. À chaque position polymorphe, le nombre de lectures contenant le SNP par rapport au nombre de lectures totales va permettre de déterminer le pourcentage de ségréants ayant le génome du parent porteur du SNP. En procédant ainsi, pour toutes les positions polymorphes la fréquence allélique au sein du pool pourra être calculée tout le long du génome. Les sites polymorphes présents dans des régions qui ne sont pas impliquées dans la résistance au composé testé auront une ségrégation aléatoire, ce qui signifie que 50% des descendants auront hérités du génome d'un parent, et les 50% restant celui de l'autre parent. Si au contraire, la région dans laquelle se situe le SNP est liée au phénotype d'intérêt, il y aura une variation de la fréquence allélique. Les régions génomiques d'un parent conférant une résistance à l'antifongique seront surreprésentées dans la descendance résistante. Ce seront donc les variations de la fréquence allélique qui vont permettre d'identifier les QTL.

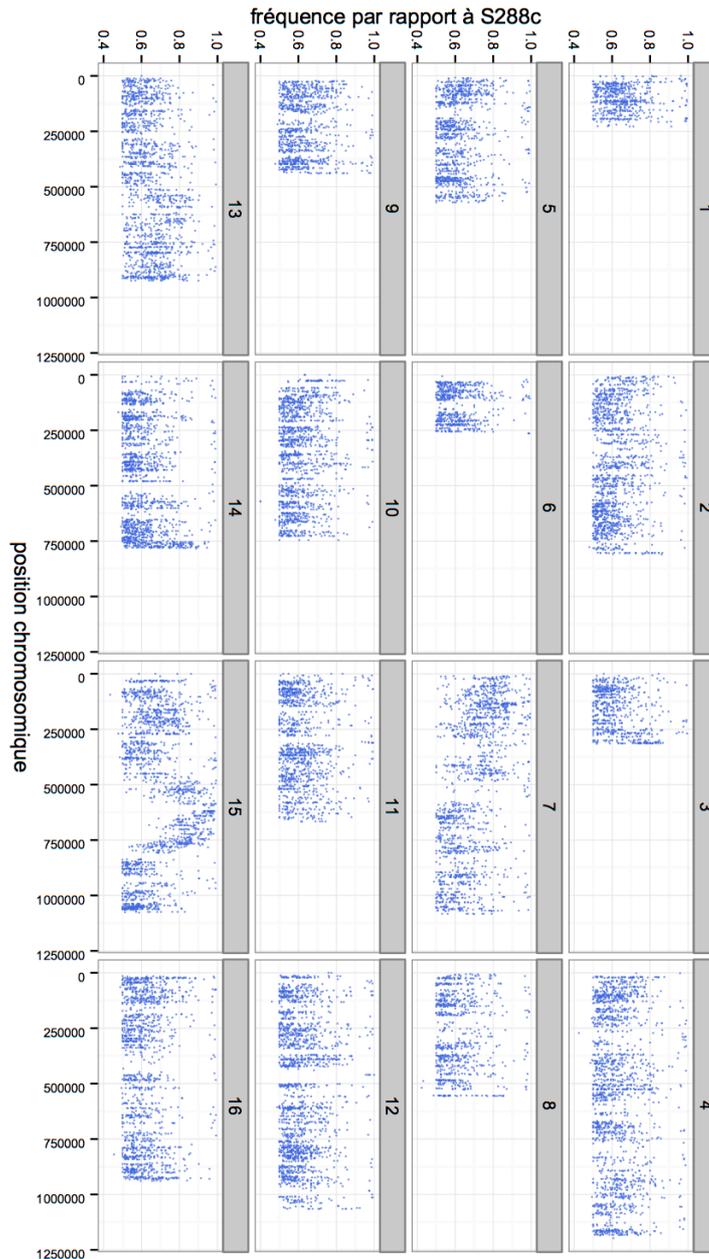
Le travail de nettoyage des données, alignements des lectures et détection des SNP a été réalisé par les collègues de l'équipe en charge des aspects bioinformatiques. Le travail que j'ai réalisé a débuté à partir de la constitution des tableaux de SNP pour les pools.



**Figure 2.18 :** Schéma de la stratégie globale employée, débutant par le séquençage de l'ADN du pool jusqu'à la détection des QTL.

### 3.2.2. Résultats du pool de résistance au fluconazole

La fréquence allélique des SNP détectés a été représentée graphiquement chromosome par chromosome (figure 2.19). La qualité des séquences n'a malheureusement pas permis l'identification de plusieurs QTL et ce, malgré le séquençage à deux reprises de l'ADN du pool. Un seul QTL majeur se distingue sans ambiguïté au niveau du chromosome XV. À son niveau, la fréquence allélique atteint de valeurs proches de 1, ce qui signifie que cette région provient pratiquement exclusivement d'un seul des parents.



**Figure 2.19 :** Analyse de la fréquence allélique du pool de résistance au fluconazole Sace\_clinique. Sur ce graphique, la fréquence allélique est représentée sur l'axe des ordonnées en fonction de la position nucléotidique du génome de référence.

### 3.2.2.1. Identification d'un pic au niveau du chromosome XV

Cette carte de la fréquence allélique a permis de délimiter un QTL qui s'étend des coordonnées 585 606 à 661 057 et couvre une région d'environ 75 kb sur le chromosome XV. L'analyse des séquences nous a révélé que cette région du génome est principalement héritée du parent YJM326 chez les ségrégants résistants. Ces résultats sont conformes à ce qui était attendu puisque la souche YJM326 est résistante aux composés azolés.

### 3.2.2.2. Contenu génétique du pic

La région d'environ 75 kb contient 54 éléments génétiques (séquences codantes d'ADN, éléments répétés, tRNA, séquence de répllication autonome) dont 39 séquences codantes. La liste de ces gènes est donnée en annexe 2.2. L'analyse fonctionnelle des gènes présents sous le pic indique la présence de 3 gènes *PDR* : le gène *PDR5* codant un transporteur de résistance multiple et deux gènes (*PDR2* et *YRMI*) codant des facteurs de transcription. Ces trois gènes sont des candidats potentiels de QTG (Quantitative Trait Gene) pouvant expliquer la résistance de YJM326. Le facteur de transcription Pdr2 active entre autres le gène *FLR1* qui code un transporteur permettant d'exporter le fluconazole, et le gène *SNQ2* codant un transporteur de résistance multiple. Le facteur de transcription Yrm1 active la transcription de multiples gènes dont celle du gène *SNQ2* (Cui *et al.* 1998). Les fonctions des autres gènes de la liste ne paraissent pas être de bons candidats qui pourraient être liés à la résistance au fluconazole. Malgré la présence de ces gènes candidats, nous n'avons pas voulu écarter les autres et négliger ainsi une éventuelle piste. Nous avons donc entrepris l'analyse du polymorphisme de tous les gènes présents dans le QTL.

### Analyse du polymorphisme des gènes

Dans cette région, seuls 32 SNP ont été détectés par l'analyse des séquences. Quelle est la répartition de ces SNP ? Tous les gènes sont-ils polymorphes ? Quelle est la proportion de mutations silencieuses ? Nous avons tenté de répondre à toutes ces questions.

Pour cela, nous nous sommes focalisés dans un premier temps sur le polymorphisme contenu dans les séquences codantes. Les génomes de souches YJM280 et YJM326 ont été séquencés par le laboratoire de Justin Fay localisé à l'université de Washington. Ces données font parties d'un projet plus global de séquençage d'une trentaine de souches de *S. cerevisiae*. Elles ne sont pas publiées mais sont disponibles sur internet. Un serveur de recherche par Blast est également disponible à l'adresse : <http://www.moseslab.csb.utoronto.ca/spgp/>. Grâce à l'utilisation de ce serveur, nous avons récupéré les séquences de YJM280 et YJM326 des gènes présents dans le QTL.

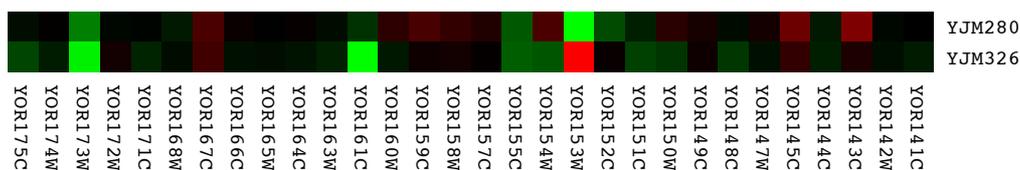
Après alignement et assemblage, les séquences des 30 gènes ont pu être comparées. Un total de 493 SNP a été détecté entre les deux souches parentales. Ce chiffre est plus de 10 fois supérieur aux 32 SNP que nous avons détecté avec nos séquences. Il faut cependant préciser que nous avons uniquement compté les SNP entre les parents et la souche de référence, et qui en plus n'étaient pas

communs entre les parents. La traduction des séquences nucléiques en séquences protéiques nous a permis de discriminer les mutations synonymes des non synonymes. Au total, 16 gènes possèdent des mutations non synonymes. Les trois gènes *PDR* (*PDR5*, *PDR2* et *YRR1*) sont très polymorphes avec respectivement 313, 24 et 17 SNP dont 94, 11 et 5 sont des mutations non synonymes. Dans cette région, les gènes *PDR* sont plus polymorphes que les autres gènes. À lui seul, *PDR5* possède 63,5% des SNP détectées dans les 30 gènes.

### Analyse de l'expression des gènes

Les mutations non synonymes peuvent avoir des conséquences sur la conformation et l'activité des protéines. Les mutations situées dans les régions non codantes peuvent également avoir un impact, notamment sur l'expression des gènes. Dans une première étape, nous avons analysé si les gènes présents au niveau du QTL sont exprimés différemment entre la souche YJM326 et la souche YJM280.

Nous avons accès aux données d'un projet d'analyse d'expressions de l'ensemble des gènes au sein de la collection des 63 souches de *S. cerevisiae*, dont les souches YJM326 et YJM280 font partie. Ces données ont été générées dans le cadre d'un autre projet. Brièvement les ARN totaux des différentes souches cultivées en milieu minimum et arrivées en phase exponentielle de croissance, avaient été extraits. Ces ARNm ont été hybridés sur des puces à ADN de type Agilent. Les expressions des gènes du QTL des souches parentales YJM280 et YJM326 par rapport à S288c ont été analysées (figure 2.20). L'expression du gène *PDR5* (*YOR153W*) est clairement modifiée. Il est sous-exprimé chez la souche YJM280 mais surexprimé chez la souche YJM326.



**Figure 2.20 :** Expression des gènes du QTL par rapport à la souche de référence S288c. Le gradient de couleur vert – rouge indique le niveau d'expression : le vert signifie que le gène est sous-exprimé par rapport à la souche de référence et le rouge que le gène est surexprimé.

En conclusion, en plus d'avoir un lien direct ou indirect avec le fluconazole, les gènes *PDR* *PDR5*, *PDR2* et *YRR1* ont des mutations non synonymes qui peuvent avoir des répercussions directes sur la protéine et à fortiori sur la résistance au fluconazole. De plus, le gène *PDR5* est surexprimé chez YJM326 sur un milieu minimum sans antifongique. Le gène *PDR5* capable d'exporter le fluconazole paraît être le meilleur gène candidat pour expliquer la résistance au fluconazole. Un dernier argument en faveur du gène *PDR5* en tant que gène responsable de la résistance, concerne sa localisation au centre du pic. Les loci directement impliqués dans les analyses de pool de ségréants se situent souvent à cette localisation car c'est là où la variation de la fréquence allélique est la plus élevée.

Nous avons vu au début de ce chapitre que 5 souches cliniques dont YJM280 possédaient une version du gène *PDR5* différente de celle des autres souches de la collection. Aux vues de la sensibilité des 5 souches cliniques aux composés azolés, cette version génique ne confère pas une résistance à ces composés (du moins à elle seule). Nous pouvons poser l'hypothèse qu'entre les deux versions alléliques de *PDR5* des parents, celle de YJM326 est la plus efficace pour exporter le fluconazole.

### 3.2.2.3. Analyse de l'expression du gène *PDR5*

Comme nous avons pu le voir dans l'introduction, la surexpression d'un transporteur de résistance multiple est un mécanisme de résistance commun chez les levures pathogènes. Les données d'expression ont montré que *PDR5* était surexprimé chez YJM326 par rapport à la souche de référence en milieu minimum. La surexpression semble donc être constitutive puisqu'il n'y avait pas de fluconazole dans le milieu. Nous avons voulu confirmer que le gène *PDR5* était toujours surexprimé dans les conditions dans lesquelles nous avons testé la résistance chez les ségrégants.

Une analyse de RT-qPCR (reverse-transcription PCR quantitative) a été réalisée afin de confirmer la surexpression de *PDR5* chez la souche YJM326 en milieu complet et en milieu complet supplémenté de  $4.10^{-5}$  M de fluconazole. Les gènes dits de ménage comme *TUB2* (codant la  $\beta$ -tubuline, un des constituant des microtubules) et *QCR9* (codant la sous unité 9 de ubiquinol cytochrome-c reductase) ont été utilisés afin de normaliser l'expression de *PDR5*. L'expression du *PDR5* de YJM326 a été comparée à celle de la souche YJM280 utilisée comme contrôle. Les résultats de l'expression sont donnés dans le tableau 2.3. Le gène *PDR5* est surexprimé chez YJM326 par rapport à YJM280 dans les deux conditions testées. La résistance de la souche YJM326 et de ses descendants est donc probablement, due en partie à la surexpression du gène *PDR5*.

Gène	Type	Milieu	Expression	Erreur standard	Intervalle de confiance à 95%	P(H1)	Résultat
<i>TUB2</i>	Référence	YPG	1,096				
<i>QCR9</i>	Référence	YPG	0,913				
<i>PDR5</i>	Test	YPG	482,706	355,989 - 597,686	330,156 - 686,044	0,019	Surexpression
<i>TUB2</i>	Référence	$4.10^{-5}$ M FC	1,275				
<i>QCR9</i>	Référence	$4.10^{-5}$ M FC	0,785				
<i>PDR5</i>	Test	$4.10^{-5}$ M FC	27,159	22,629 - 34,520	20,067 - 37,404	0,012	Surexpression

**Tableau 2.3 :** Mesure de l'expression du gène *PDR5* chez YJM326 par rapport à la souche YJM280 en YPG et YPG supplémenté de fluconazole. Un test statistique par technique de randomisation permet de tester l'hypothèse selon laquelle la différence d'expression de *PDR5* entre les deux souches parentales est due ou non au hasard. Dans notre cas, la valeur P indique que le gène *PDR5* est significativement surexprimé chez YJM326 par rapport à YJM280 dans les deux conditions de culture.

La surexpression de *PDR5* peut être due à une régulation en *cis* ou en *trans*. La régulation de la transcription en *cis* est due à des éléments situés au niveau du promoteur qui vont recruter les facteurs de transcription et la polymérase. La régulation *trans* concerne quant à elle des éléments qui ne sont pas situés au niveau du gène lui-même comme par exemple les facteurs de transcription. La surexpression de *PDR5* permet d'améliorer la résistance, et puisque seule cette région est mise en évidence par l'analyse de pool de ségrégants, nous pouvons donc émettre l'hypothèse que la surexpression est due à un élément génétique situé au niveau de la séquence promotrice (Carvajal *et al.* 1997; Sanglard *et al.* 2009).

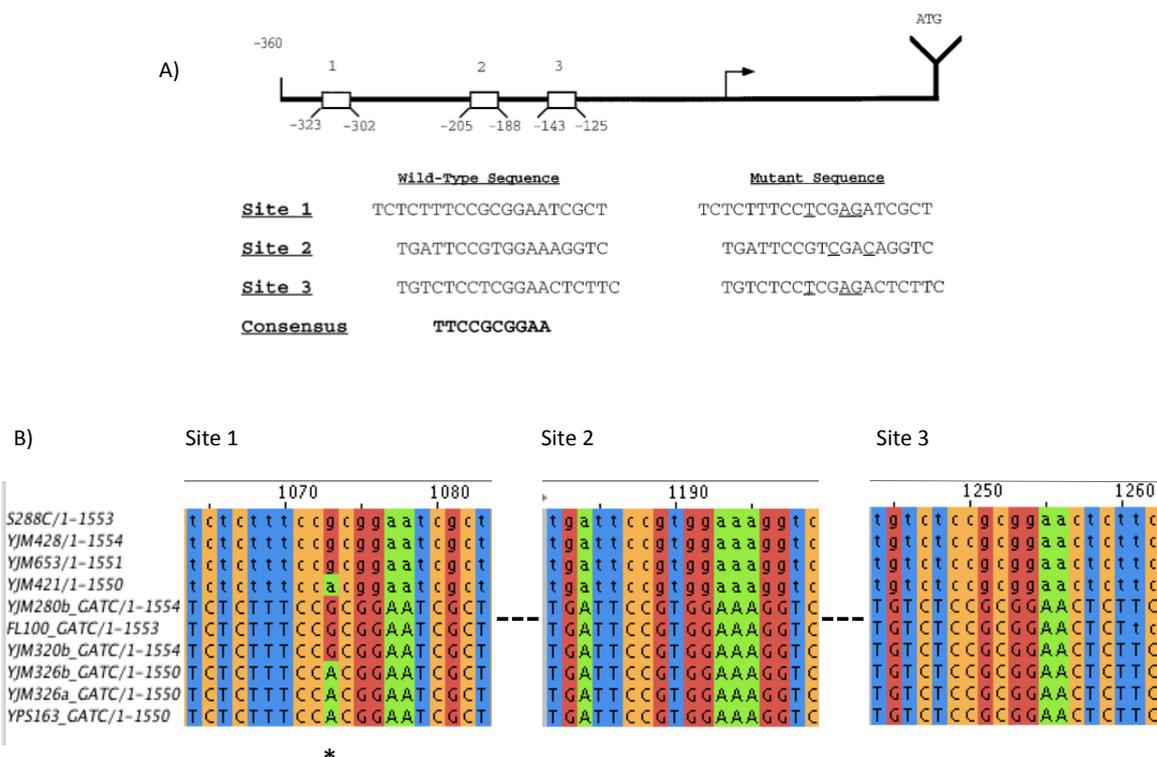
#### 3.2.2.4. La séquence promotrice du gène *PDR5*

Pour aller plus loin dans l'analyse, la séquence intergénique entre *YOR152c* et *PDR5* de différentes souches de *S. cerevisiae* a été séquencée, ou récupérée lorsque cela était possible dans des banques de données.

Le but de cette analyse est de savoir s'il y a un lien entre la séquence promotrice et la surexpression de *PDR5*. Pour cela, nous avons regardé d'une part les séquences PDRE (Pleiotropic Drug Response Elements) recrutant les facteurs de transcription de type PDR et d'autre part la séquence de la région intergénique dans sa globalité.

Il y a trois séquences de type PDRE qui ont été mises en évidence au niveau du promoteur de *PDR5*. Elles permettent de recruter les facteurs de transcription Pdr1 et Pdr3 (Katzmann *et al.* 1996). Elles sont annotées de 1 à 3, en allant de la séquence la plus éloignée du codon start à la plus proche (figure 2.21). Les séquences PDRE ont été analysées dans les différentes souches afin de savoir si une corrélation est observée entre la surexpression de *PDR5* et la présence d'une mutation dans ces séquences. À la 4<sup>ème</sup> position de la séquence PDRE 1, il y a une substitution d'une Adénine vers une Guanine pour les souches YJM326, YPS163, YJM421, YPS1009, PW5 et YPS128.

Les séquences de la région intergénique ont été alignées et assemblées afin de permettre une comparaison entre elles. Un arbre de type Neighbor-joining a été réalisé à partir de l'alignement de toutes les séquences (figure 2.22). La séquence de YJM326 est effectivement très polymorphe par rapport à celle de S288c et de YJM280. Au total, 66 SNP différencient leur séquence. Afin de contrôler si cette séquence promotrice conduit effectivement à la surexpression du gène *PDR5*, il faudrait réaliser un remplacement allélique en insérant la séquence promotrice de YJM280 à la place de celle de YJM326. Si l'expression du gène *PDR5* baisse, la séquence promotrice de YJM326 est effectivement à l'origine de la surexpression du gène *PDR5*. La résistance au fluconazole de la souche YJM326 dont le promoteur aura été remplacé sera également contrôlée afin de confirmer l'impact de la surexpression sur la résistance.

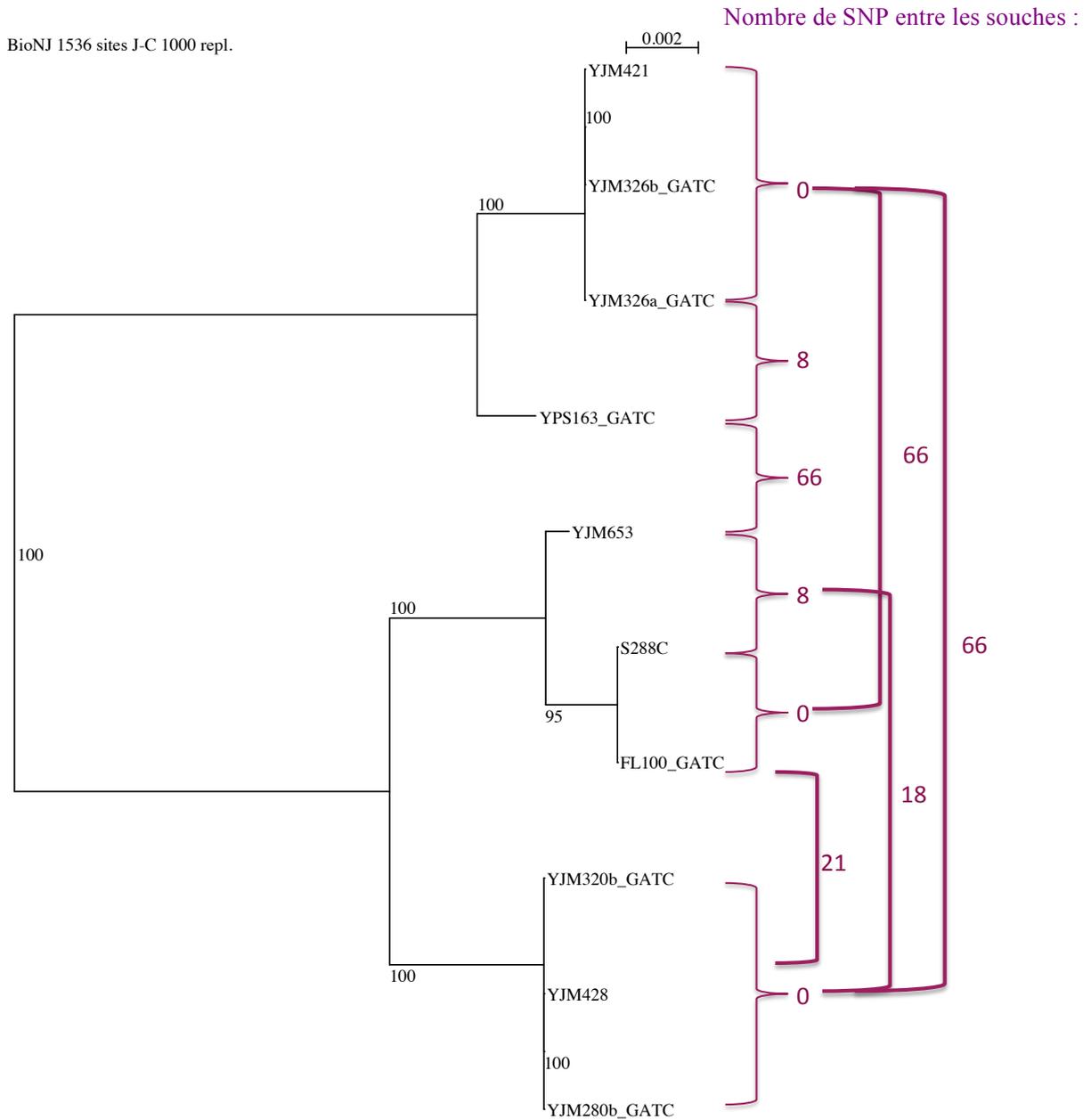


**Figure 2.21 :** Analyse des séquences PDRE dans le promoteur de *PDR5*. A) Localisation des séquences PDRE et leur séquence consensus (Katzmann *et al.* 1996). B) Alignement des séquences PDRE chez 10 souches de *S. cerevisiae*. Les séquences en lettres minuscules sont celles récupérées sur les bases de données et les séquences écrites en majuscules sont celles séquencées lors de ma thèse. L'astérisque symbolise la présence d'un SNP dans une séquence PDRE. Les couleurs symbolisent les nucléotides.

En conclusion de cette partie, cette analyse de pool de ségrégants en présence de fluconazole basée sur le couple Sace\_clinique a révélé un QTL majeur situé sur le chromosome 15. Dans ce QTL, le gène candidat le plus probable pour expliquer la résistance au fluconazole est *PDR5*, du fait de 1) sa fonction, 2) sa position centrale sous le pic de la variation de la fréquence allélique, 3) son fort polymorphisme et 4) sa surexpression dans les conditions de l'expérience. Pour confirmer l'implication du gène et de son promoteur, il faudrait procéder au remplacement allélique du gène *PDR5* d'une part et de son promoteur d'autre part. Comme expliqué précédemment, deux souches YJM326 mutées doivent être construites, la 1<sup>ère</sup> pour laquelle le gène *PDR5* est remplacé par l'allèle de YJM280 et la 2<sup>ème</sup> pour laquelle le promoteur est remplacé par celui de YJM280. La résistance au fluconazole des deux souches mutées sera alors comparée à celle de la souche YJM326. Si la résistance baisse chez les souches mutées, l'implication des allèles dans la résistance sera ainsi confirmée.

Nous pouvons également proposer les expériences complémentaires dans lesquelles le gène *PDR5* et son promoteur chez la souche YJM280 seront remplacés par les allèles de la souche YJM326.

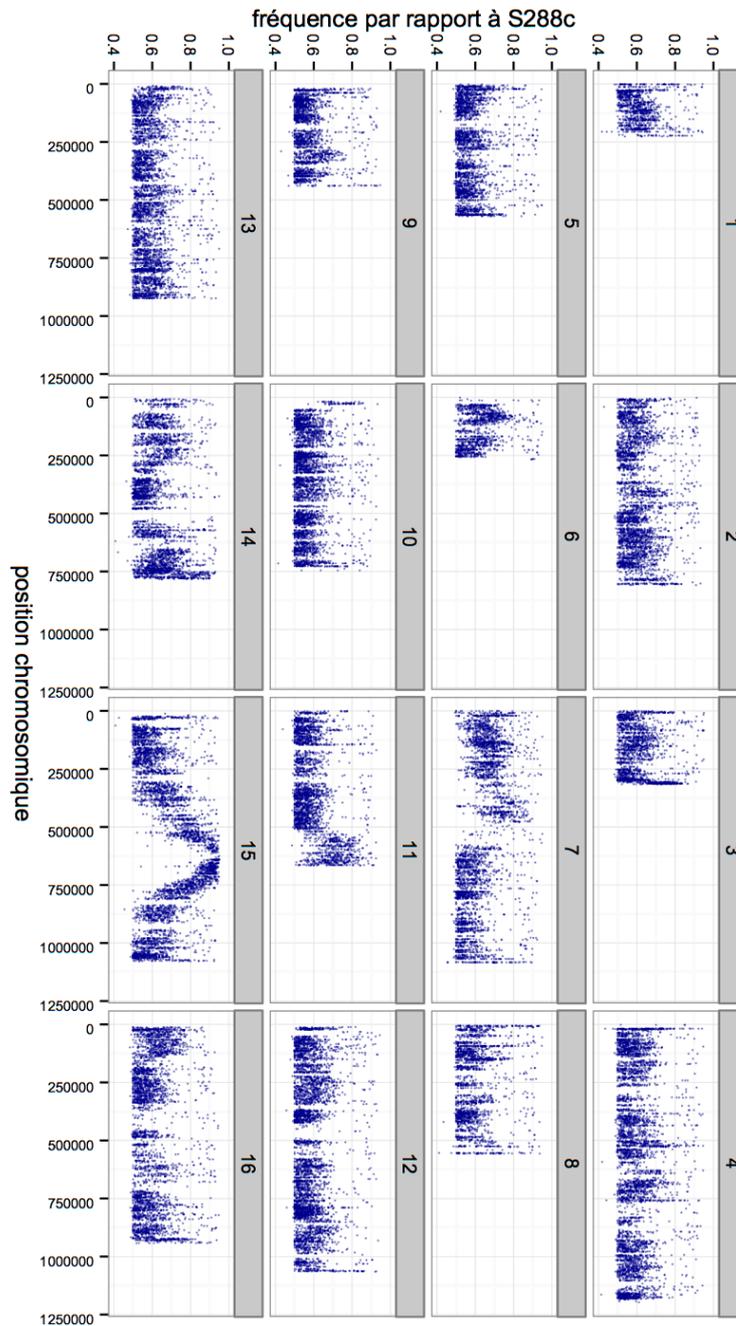
L'analyse de la résistance au fluconazole permettra de déterminer si les versions alléliques de YJM326 sont suffisantes pour améliorer la résistance de la souche sensible YJM280.



**Figure 2.22** : Cladogramme de type Neighbor-joining basé sur la séquence nucléique de la région intergénique *YOR152c* et *PDR5*. Les valeurs aux nœuds des branches indiquent la robustesse de l'embranchement. Le nombre de SNP entre les clades est indiqué en mauve.

### 3.2.3. Résultats du pool de résistance au kétoconazole

Les lectures issues du séquençage du pool de ségréants résistant au kétoconazole Sace\_clinique ont été traitées et analysées en suivant la même stratégie que le pool de ségréants résistants au fluconazole (figure 2.18). La fréquence allélique des SNP détectés a été représentée graphiquement le long des 16 chromosomes (figure 2.23).



**Figure 2.23** : Représentation de la fréquence allélique le long du génome de référence dans le pool de résistance Sace\_clinique au kétoconazole.

### 3.2.3.1. Identification des QTL

L'analyse de la variation des fréquences alléliques a mis en évidence deux QTL, l'un sur le chromosome XI et l'autre sur le chromosome XV.

#### 3.2.3.1.1. Détection d'un pic sur le chromosome XI

Un pic est détecté à la fin du chromosome XI. Il s'étend des coordonnées 549 080 à 612 517 et 39 gènes y sont localisés. L'analyse fonctionnelle ne nous permet pas de prédire qu'un gène pourrait être un bon candidat pour le QTL. Pour trouver le locus ou les loci impliqués dans la résistance, l'une des possibilités expérimentales consisterait à réaliser des délétions par petites régions (en évitant de déléter les gènes essentiels), et réaliser le test d'hémizygocité réciproque, comme mentionné dans l'introduction.

Devant l'ampleur de la tâche à accomplir et la difficulté d'obtenir des mutants de délétions pour ces nombreuses régions, nous avons décidé de ne pas continuer d'avantage l'analyse de ce QTL.

#### 3.2.3.1.2. Détection d'un pic sur le chromosome XV

Un pic est visible au niveau du chromosome XV. Il s'étend des coordonnées 583 247 à 653 139 et comprend 34 gènes (annexe 2.2). Le gène *PDR5* est encore une fois situé au centre du pic, à l'endroit où la variation de la fréquence allélique est la plus forte (supérieure à 0,9). Une analyse de PCR quantitative a été réalisée sur les souches parentales en milieu complet supplémenté de kétoconazole. Les résultats obtenus prouvent que le gène *PDR5* de YJM326 est surexprimé par rapport à celui de YJM280.

Les conclusions dégagées sont donc les mêmes que celles de l'analyse de pool de ségréants sur fluconazole. Il faudrait réaliser le remplacement allélique du gène *PDR5* d'une part et de sa région promotrice d'autre part, puis tester la résistance au fluconazole des mutants afin de prouver définitivement leur implication dans la résistance au kétoconazole.

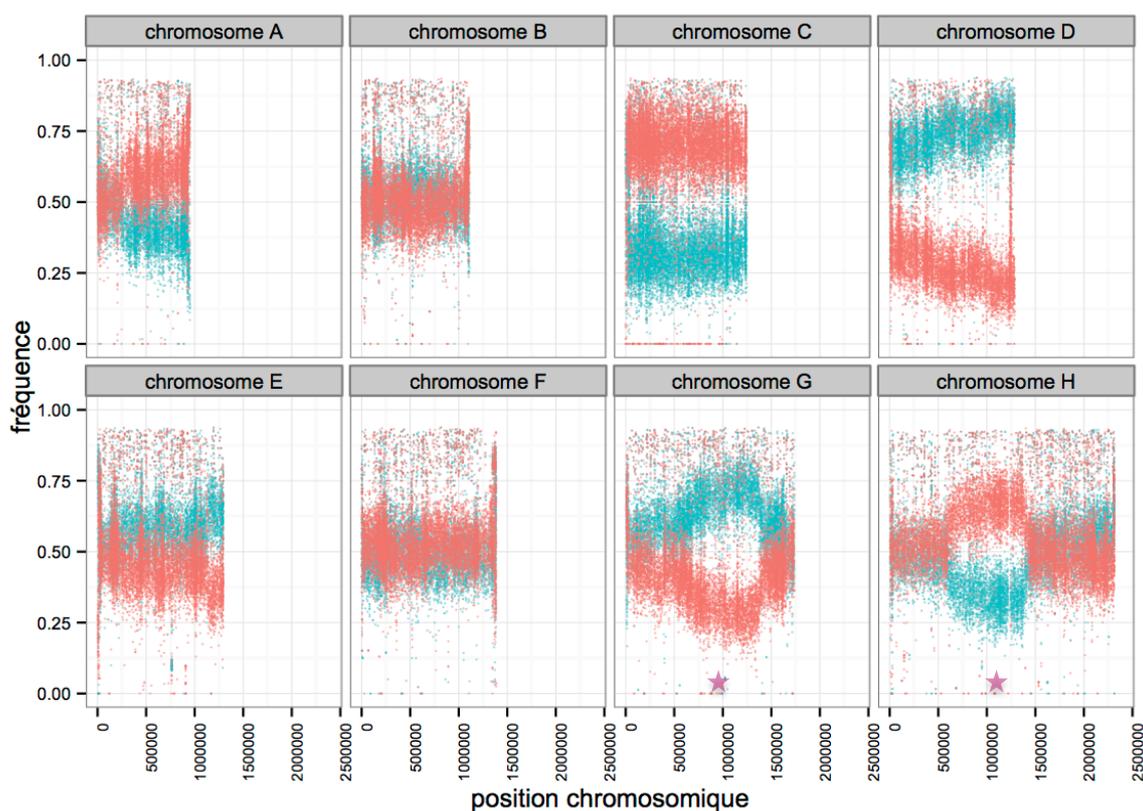
### 3.3. Analyse des pools des ségréants du couple de *L. kluyveri* Lakl\_Japon

Pour *L. kluyveri*, nous nous sommes également intéressés à la résistance au fluconazole et au kétoconazole par des analyses de pool de ségréants. Pour cela, nous nous sommes focalisés sur deux couples de souches : Lakl\_Japon (souches NBRC1811a et NBRC10955b) et Lakl\_Europe (souches CBS5828a et dd281a\_b).

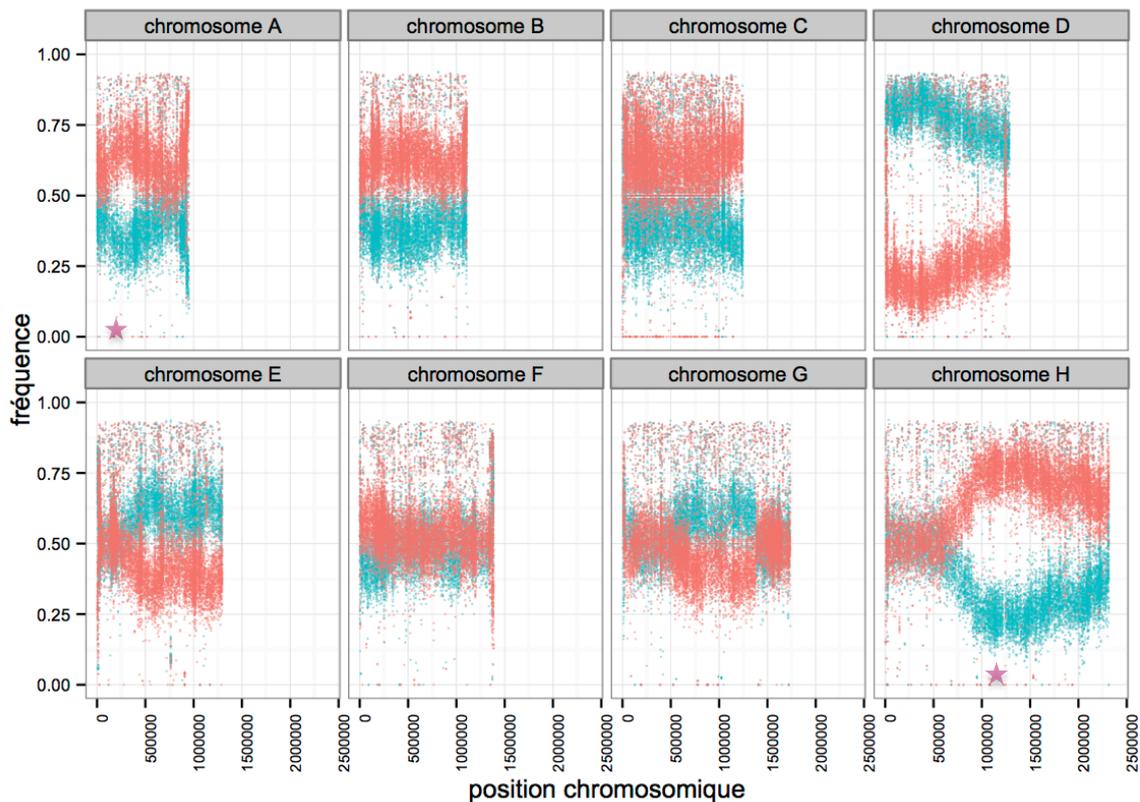
Les données de séquençage des pools pour *L. kluyveri* étaient de meilleure qualité que ceux de *S. cerevisiae*. Les lectures issues du séquençage ont été alignées contre les génomes parentaux. Près de 84000 sites polymorphes ont été détectés, représentant 1 site polymorphe toutes les 130 pb.

### 3.3.1. Résultats du pool de résistance au fluconazole et au kétoconazole

Les données de séquençage nous ont donc permis de déterminer la variation de fréquence allélique des deux versions parentales le long du génome. Ces fréquences sont représentées dans la figure 2.24 pour le pool de ségréants résistants au fluconazole et dans la figure 2.25 pour le pool de ségréants résistants au kétoconazole. Contrairement aux résultats que nous avons obtenus pour *S. cerevisiae*, nous pouvons observer que de nombreuses régions du génome présentent une déviation de la fréquence allélique. Ce biais dans la variation peut même concerner des chromosomes entiers. Une des hypothèses expliquant ces variations de fréquence atypiques le long du génome pourrait être la présence d'incompatibilités génétiques entre les souches parentales entraînant les déviations observées. Afin de vérifier cette hypothèse, il faudrait générer, séquencer et déterminer la variation de la fréquence allélique d'un pool de ségréants choisis aléatoirement dans la descendance.



**Figure 2.24 :** Fréquence allélique de NBRC10955b (orange) et de NBRC1811a (bleu) le long du génome pour le pool de ségréants résistants au fluconazole. Les étoiles roses symbolisent la présence de QTL.



**Figure 2.25 :** Fréquence allélique de NBRC10955b (orange) et de NBRC1811a (bleu) le long du génome pour le pool de ségrégants résistants au kétoconazole. Les étoiles roses symbolisent la présence de QTL.

Cependant, certains QTL peuvent tout de même être définis. Dans le cadre de la résistance au fluconazole, nous avons pu identifier une région sur le chromosome G et un autre sur le chromosome H (figure 2.24). Pour la résistance au kétoconazole, deux autres régions sont mises en évidence : l'une sur le chromosome A et l'autre sur le chromosome G (figure 2.25). Néanmoins, ces régions ont une taille comprise entre 0,5 et 1 Mb. Il serait donc nécessaire d'affiner la cartographie afin de pouvoir déterminer des gènes candidats. La détection de grande région a pu récemment être expliquée lors de l'étude de la recombinaison chez *L. kluyveri*. En effet, il a été observé que le taux de recombinaison est 2 à 3 fois inférieur chez *L. kluyveri* par rapport à *S. cerevisiae*. Ce faible taux de recombinaison va avoir un impact sur la cartographie et plus particulièrement sur la taille des QTL détectés.

### 3.3.2. Conclusion de l'analyse pour le couple Sakl\_Japon et perspectives

Ainsi, notre première analyse de pools de ségrégants n'a malheureusement pas permis de mettre en évidence des QTL conduisant à l'identification de gènes impliqués dans la variation de la résistance au fluconazole et au kétoconazole.

C'est pourquoi nous avons décidé de réaliser une nouvelle analyse de pool de ségrégants en changeant plusieurs paramètres. Dans un premier temps, nous avons choisi des isolats dont la

divergence génétique était moins importante que précédemment afin d'éviter la présence d'incompatibilités génétiques. Dans un deuxième temps, nous avons également généré et analysé un plus grand nombre de ségrégants. Et enfin, nous avons généré un contrôle composé d'un groupe de descendants choisis aléatoirement afin de déterminer la variation de la fréquence allélique sans sélection au fluconazole ou kétoconazole.

### 3.4. Analyse des séquences des pools de ségrégants issus du couple de *L. kluyveri* Laki\_Europe

Nous avons ainsi réalisé une deuxième analyse de pool de ségrégants impliquant les isolats CBS5828a et dd281a\_b. Dans ce cadre, nous avons généré, séquencé et déterminé la variation de la fréquence allélique pour un groupe de ségrégants résistants au fluconazole, résistants au kétoconazole et un groupe de ségrégants choisis aléatoirement, appelé groupe contrôle.

#### 3.4.1. Résultats du séquençage pour le groupe contrôle

Au total, un pool de 50 ségrégants choisis aléatoirement a été séquencé. La variation de la fréquence allélique de la version de CBS5828a est représentée dans la figure 2.26. Comme précédemment, nous pouvons voir une déviation de la fréquence allélique le long de chromosomes. Il est possible d'observer une surreprésentation du génome de CBS5828a dans la descendance pour quasiment l'ensemble du génome à l'exception du bras gauche du chromosome C et de l'extrémité droite du chromosome H.

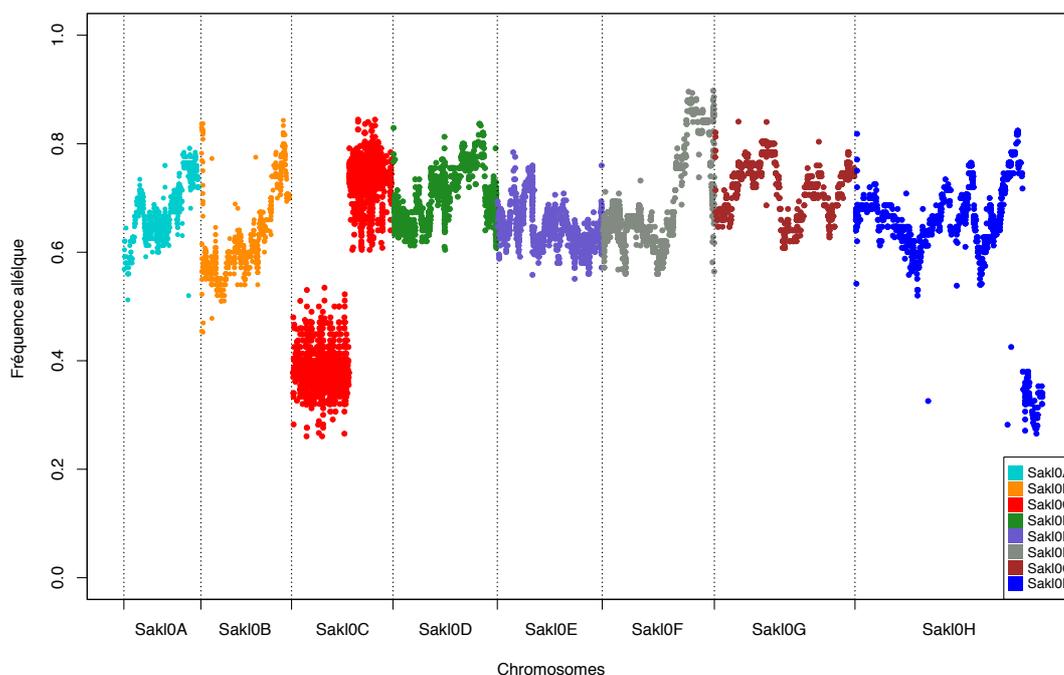


Figure 2.26 : Fréquence allélique de CBS5828a au niveau du groupe contrôle.

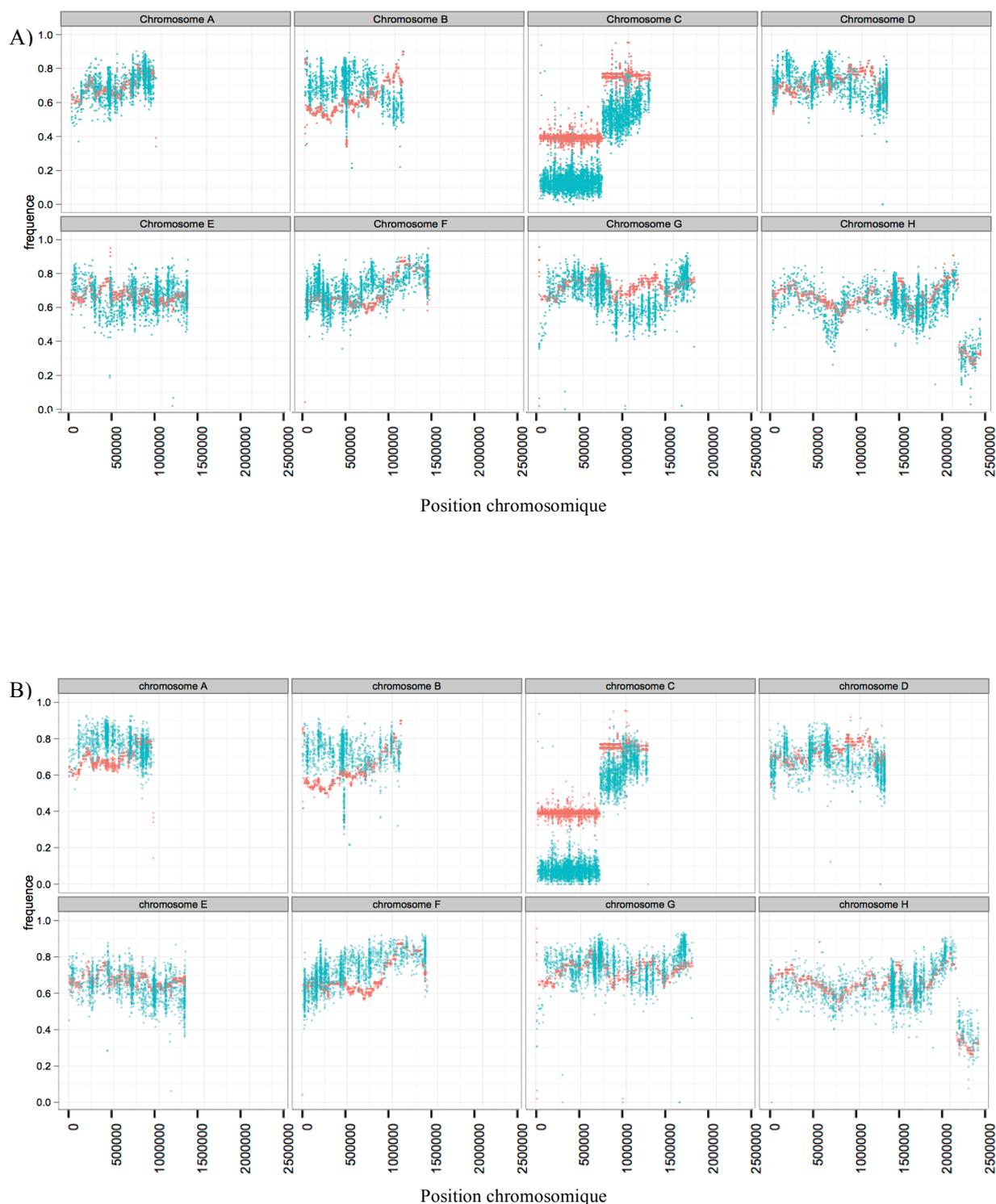
À ce jour, il est difficile de déterminer l'origine de cette variation des fréquences alléliques atypique dans la descendance. Il est évidemment que des incompatibilités génétiques pourraient expliquer tout ou partie de cette observation.

### **3.4.2. Comparaisons entre les groupes de ségréants résistants et le contrôle**

Afin de pouvoir identifier des QTL impliqués dans la résistance au fluconazole ou kétoconazole, nous avons décidé de prendre en compte la variation allélique observée dans le groupe contrôle. Dans la figure 2.27, les fréquences alléliques déterminées pour les groupes résistants et le groupe contrôle ont été représentées. Nous avons donc essayé de déterminer des régions pour lesquelles nous pouvions observer une différence significative de la fréquence allélique entre les pools de ségréants et le pool contrôle. Malheureusement, aucune région n'a pu être identifiée. Par conséquent, aucun QTL n'a pu être mis en évidence.

### **3.4.3. Conclusion des pools de ségréants issus du couple Lakl\_Europe**

Dans cette deuxième étude, nous avons voulu optimiser certaines étapes de l'analyse de pool de ségréants. Malheureusement, cette optimisation n'a pas permis d'identifier des QTL impliqués dans la résistance au fluconazole ou kétoconazole. Un des problèmes majeurs qui a été rencontré est la présence d'un biais de la fréquence allélique au niveau d'un groupe de descendants choisis aléatoirement. À ce jour, il est difficile d'en déterminer l'origine. Cependant, il est indéniable que ce biais limite la détection des QTL.



**Figure 2.27 :** Analyse des séquences des pools de ségréants SakI\_Europe. Comparaison des fréquences alléliques du groupe contrôle (couleur orange) et des groupes de ségréants résistants (couleur bleu) au fluconazole (A) et kétoconazole (B) contre le génome du parent CBS5828a.

## Conclusions générales et perspectives

L'analyse des séquences des gènes *PDR* a permis de constater que ces gènes étaient fortement conservés chez *S. cerevisiae*. Au contraire, ces gènes présentent une grande variabilité chez *L. kluyveri*. Chez *S. cerevisiae* deux versions du gène codant le transporteur Pdr5 ont été trouvées dont l'une est spécifique aux souches cliniques. L'analyse des séquences et la construction d'arbres de type neighbor-joining nous indiquent que l'origine la plus probable pour expliquer la présence de cette version du gène *PDR5* est une introgression.

Nous avons également pu tester la résistance à différents composés azolés grâce à un suivi de la croissance en microcultures pour des isolats de *S. cerevisiae* et *L. kluyveri*. Les résultats mettent en évidence une variabilité de résistance au sein des deux collections mais plus particulièrement entre isolats de *S. cerevisiae*. Certaines souches de *S. cerevisiae* présentent des résistances multiples aux composés testés.

Au vu de cette diversité de résistance, nous avons voulu déterminer des origines génétiques de la résistance au fluconazole et au kétoconazole. Pour cela, nous avons réalisé des analyses de pool de ségrégants.

Chez *S. cerevisiae*, les analyses de liaison nous ont permis d'identifier une région génomique commune potentiellement impliquée dans la résistance au fluconazole et au kétoconazole. Cette région est située sur le chromosome XV et contient le gène *PDR5* qui code un transporteur de résistances multiples. L'analyse des séquences a révélé de nombreuses mutations, spécifiques à la souche résistante, dans la séquence codante du gène *PDR5* mais également au niveau de son promoteur. Une analyse de PCR quantitative a montré que le gène *PDR5* est surexprimé chez la souche parentale résistante, YJM326.

Par la suite, il serait intéressant de confirmer définitivement l'implication de ce gène dans la résistance. Pour cela, il est envisagé de faire un remplacement allélique ou un test d'hémizygotie réciproque. Il semble évident que la surexpression du gène *PDR5* joue un rôle dans la résistance de la souche YJM326, c'est pourquoi il serait intéressant et indispensable de prendre en considération la séquence promotrice du gène *PDR5* dans ces expériences.

Une perspective plus générale serait d'utiliser une analyse de type iQTL qui permettrait de détecter les QTL avec une plus grande sensibilité et ce avec une meilleure résolution (Parts *et al.* 2011). En effet, la résistance au fluconazole et au kétoconazole est un trait polygénique pourtant nous n'avons mis en évidence que deux QTL. La méthode iQTL dérive de l'analyse de pool de ségrégants. Elle permet de réaliser plusieurs générations de ségrégants grâce à un système de sélection basé sur des auxotrophies. De plus, les ségrégants présentant le phénotype d'intérêt sont sélectionnés en masse.

Le fait de prendre un plus grand nombre de ségrégants et qui proviennent d'une génération F6 ou davantage, permet d'augmenter le nombre de points de recombinaison ce qui permettra de délimiter les QTL avec plus de précision.

Chez *L. kluyveri*, les analyses de liaison n'ont malheureusement pas pu permettre de déterminer des QTL. De nombreuses améliorations peuvent cependant être apportées. Tout d'abord, il est possible depuis peu d'estimer la viabilité des descendants issus d'un croisement de deux souches de *L. kluyveri* grâce à la nouvelle technique de dissection mise au point au laboratoire. Il serait judicieux de choisir des souches pour lesquelles la viabilité est importante (supérieure à 90%). L'objectif serait d'éviter la présence d'incompatibilités génétiques ou la présence de réarrangements chromosomiques. Ces derniers induisent un biais dans la fréquence allélique. En évitant ces événements, il serait possible d'avoir une fréquence allélique de 0,5 le long du génome pour chacune des versions parentales dans le cas d'un groupe contrôle. Ceci constituerait un bon point de départ pour une analyse de pool de ségrégants.

# Chapitre 3

## Expériences d'évolution en présence de fluconazole de souches de *S. cerevisiae* et *L. kluyveri*

### 1. Introduction

Dans les deux premiers chapitres de ce manuscrit, nous avons étudié la diversité de résistance présente au sein de deux collections de levures. La diversité de résistance mesurée entre les souches est le fruit d'une histoire évolutive longue de près de 11 000 générations. Nous connaissons la résistance des souches de *S. cerevisiae* et *L. kluyveri* aux composés azolés grâce à l'analyse de la croissance que nous avons réalisé. Nous avons notamment vu que la résistance peut être associée à un polymorphisme nucléotidique, comme chez la souche clinique YJM326 pour laquelle nous avons pu identifier l'origine principale de la résistance au fluconazole et au kétoconazole au niveau du gène *PDR5* et de sa région promotrice.

Parallèlement à cette évolution d'une dizaine de milliers d'années, nous nous sommes demandés quels seraient les mécanismes d'évolution impliqués à court terme, si une souche devait faire face à un nouvel environnement ? Dans l'optique d'étudier les processus adaptatifs chez *S. cerevisiae* et *S. kluyveri*, nous avons mis au point et réalisé des expériences dites d'évolution adaptative dans le cadre de notre phénotype d'intérêt : la résistance au fluconazole.

Pour réaliser cette expérience d'évolution adaptative, nous avons choisi d'utiliser deux souches sensibles au fluconazole : la souche clinique YJM280 de *S. cerevisiae* isolée aux États-Unis, et la souche NBRC10955 de *L. kluyveri* isolée d'une feuille en décomposition au Japon. Ces souches ont été cultivées et repiquées toutes les 48 heures sur un milieu riche (YPG) supplémenté de fluconazole à  $4 \cdot 10^{-5}$  M durant 200 générations (environ 30 jours) (figure 3.1). Un réplica biologique a été réalisé pour les deux souches. Les vitesses spécifiques de croissance en présence de fluconazole de la souche de départ (souche ancestrale) et celles des souches isolées à différents temps de l'expérience (souches évoluées) ont été comparées. Dès la première semaine de l'expérience, la vitesse spécifique de croissance des souches évoluées a été supérieure à celle de la souche ancestrale. Grâce au séquençage complet du génome de la souche de départ et des souches évoluées, nous avons pu identifier les variations génétiques à l'origine de la résistance au fluconazole. L'adaptation s'est réalisée pour les deux espèces par la formation de chromosomes surnuméraires.

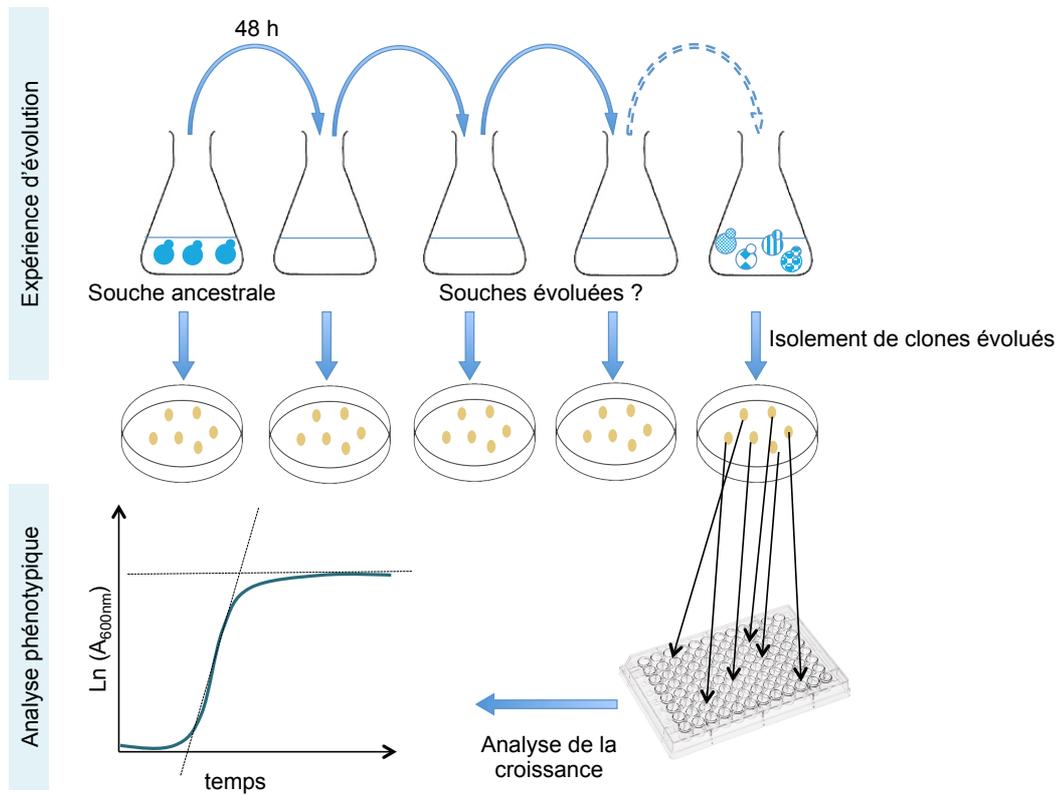
Pour les souches évoluées de *S. cerevisiae*, les chromosomes IV et VIII sont dupliqués. Des gènes de résistance multiples sont présents sur ces deux chromosomes dont les gènes *PDR5* et *SNQ2* codant tous les deux des transporteurs de résistance multiple. Par un effet de dosage génique, les gènes *PDR5* et *SNQ2* seront transcrits en plus grande quantité comparé à ceux de la souche ancestrale, ce qui

permettra la synthèse d'un plus grand nombre de transporteurs. Pour confirmer cette hypothèse, une des deux copies de ces gènes a été/va être délétée chez les souches évoluées. Les expériences de délétions sont actuellement en cours.

Chez *S. kluyveri*, le séquençage complet des génomes des souches évoluées a révélé la présence d'une aneuploïdie du chromosome B et la présence de chromosomes chimères surnuméraires BG et GH. Le chromosome chimère BG est constitué d'une partie du chromosome B et d'une partie du chromosome G, tandis que le chromosome GH est constitué d'une partie du chromosome G et H. Chacun des chromosomes chimères est présent dans des souches évoluées issues de deux réplicas indépendants. L'analyse des séquences nous a révélé la présence d'un élément transposable au niveau des jonctions des deux chromosomes chimères. Ces résultats suggèrent que les éléments transposables sont à l'origine de la formation des chromosomes chimères. Des gènes impliqués dans les mécanismes de résistance sont présents sur ces chromosomes surnuméraires (chromosome dupliqué et chromosomes chimères). Nous retrouvons l'homologue du gène *PDR5* mais également l'homologue du gène *ERG11*. Le gène *ERG11* est la cible du fluconazole et, toujours par un effet de dosage génique, il pourrait participer à la résistance. La délétion d'une copie de *PDR5* et *ERG11* chez les souches évoluées a permis de tester la résistance de souches ayant les chromosomes surnuméraires mais ne possédant qu'une copie des gènes d'intérêt. La délétion d'une copie du gène *PDR5* ne fait pas diminuer la vitesse spécifique de croissance de la souche évoluée sur fluconazole. Par contre, la délétion d'une copie du gène *ERG11* entraîne une diminution de la vitesse spécifique de croissance sur fluconazole.

En conclusion, ce travail a mis en évidence un mécanisme d'adaptation commun aux espèces *S. cerevisiae* et *L. kluyveri* face à un stress de type antifongique : la formation de chromosomes surnuméraires.

Les principaux résultats relatifs à cette étude sont présentés sous forme de manuscrit (voir partie suivante). Les expériences de délétions, actuellement en cours, seront intégrées au manuscrit une fois réalisées.



**Figure 3.1 :** Expérience d'évolution et stratégie d'analyse de la résistance.



## **Evolving mechanisms leading to the fluconazole resistance are conserved within the Hemiascomycetes yeasts**

Cyrielle Reisser<sup>1</sup>, Anne Friedrich<sup>1</sup>, Jacky de Montigny<sup>1</sup> and Joseph Schacherer<sup>1</sup>

Department of Genetics, Genomics and Microbiology,  
University of Strasbourg / CNRS, UMR7156  
Strasbourg, France

**Keywords:** yeasts ; evolving experiments ; chimeric chromosomes ; aneuploidies

**Running title:** Evolution on fluconazole in two yeast species

**Punchline:** the formation of extra chromosomes is a common and efficient way to evolve rapidly in the presence of fluconazole in the *S. cerevisiae* and *S. kluyveri* species.

## Abstract

The adaptive evolution experiments provide a powerful tool to study the genetics bases of evolution. They allow us to observe an evolution in action, in a controlled environment. Fluconazole is the antifungal the most widely used in the world. The adaptation of yeast to fluconazole leads to the acquisition of resistance and raises public health problem. The molecular mechanisms of resistance to fluconazole are known but the processes of adaptation remain unclear.

To have a better understanding of the mechanisms of adaptation that occurred in yeasts, we performed evolution experiments with two genetically distant Hemiascomycetes yeasts : *Saccharomyces cerevisiae* and *Saccharomyces kluyveri*. We evolved two replicate lines of yeast per species under fluconazole stress for 200 generations. The sequencing of the genome of six evolved strains reveals that the resistance is acquired by the formation of extra chromosomes by aneuploidy (*S. cerevisiae* and *S. kluyveri*) or by the formation of chimeric chromosomes (*S. kluyveri*). Transposable elements as repeated sequence are implied in the formation of chimeric chromosomes in the *S. kluyveri* strain. Finally we found that the molecular bases of the resistance to fluconazole of the evolved strains reside in the two copies (due to the extrachromosomes) of key genes.

This study, reveals a common mechanism of adaptation to fluconazole between two divergent Hemiascomycetes yeasts. This supports the importance that have aneuploidies and chromosomal rearrangements in the evolution process.

## Introduction

The adaptive evolution experiments provide a powerful tool to study the genetics bases of evolution. They allow us to observe an evolution in action, in a controlled environment. By this way, we can restart the experiment several times and compare the mechanisms of adaptation that occurred between the different replicates. In addition, we can test the existing hypotheses on the evolution mechanisms (Kawecki et al. 2012).

Microorganisms offer many advantages to be used in this type of experiences. They are easy to propagate and enumerate, they have a short generation time, they have low nutritive needs, they grow in small spaces, they can be stored at -80°C that allow the direct comparison between the ancestral and the evolved strain, and finally they can reproduce asexually (Elena and Lenski 2003). Usually, the protocol of evolution experiments follows the same overall pattern. A clone (the ancestral strain), whose genome and phenotype are characterized, is growing up in a new environment, in a continuous, or in a discontinuous culture, during tens, hundreds or thousands of generations. These experimental designs maintain a genetic diversity within the population (Barrick and Lenski 2013). The accumulation of genetic changes acquired during the experiment, leads to a fitness advantage of the evolved strains.

The yeasts are widely used in the evolving experiment because in addition to the advantages of microorganisms, they share some cellular processes with all the eukaryotic cells. In this study, we used two model yeast species: *Saccharomyces cerevisiae* and *Saccharomyces kluyveri* (recently renamed *Lachancea kluyveri*). We used these yeasts because their molecular and genomic data as well as the techniques for their precise genetic analyses and manipulation are abundant. They both belong to the Hemiascomycetes clade, however their lineages have separated before that the whole genome duplication event of the Hemiascomycetes occurred, 1.7 billion of generations ago. Therefore, *S. cerevisiae* is a post-duplicated yeast with a set of 16 chromosomes, and *S. kluyveri* a protoploid yeast with a set of 8 chromosomes. Despite the distance that separates both species, they share similar life cycle. By using these two genetically distant yeasts, we wish to explore the general adaptation processes involved in evolution.

The adaptation and resistance of yeasts to antifungal treatments is a public health problem. In our study, we focus on the adaptation to the fluconazole, which is the most widely prescribed antifungal to treat yeast infections. Like the other azole drugs, the fluconazole inhibits the ergosterol biosynthesis by linking the lanosterol 14- $\alpha$ -demethylase enzyme of the cytochrome P450 (Lamb et al., 1999). The fluconazole success is due to its efficiency to treat yeast infections, and because it causes few side effects (Maertens, 2004). Several mechanisms of resistance to fluconazole have been identified until today: 1) decreased membrane permeability, 2) decreased drug concentration, 3) target site alteration, 4) up-regulation of target enzyme and 5) development of bypass pathways (Kanafani and Perfect, 2008). The molecular bases of the resistance are known but the processes of adaptation

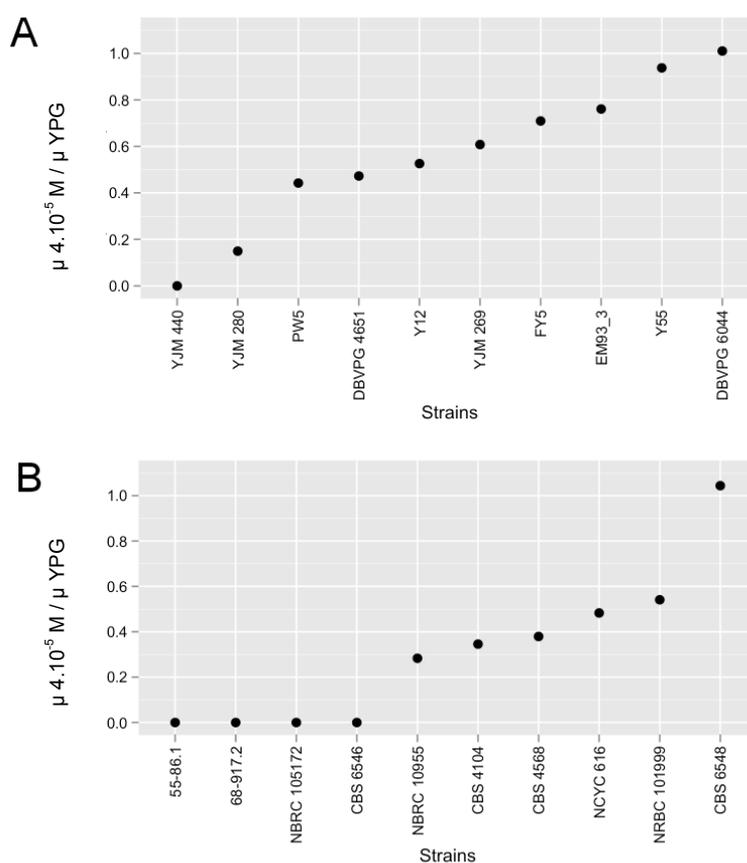
remain unclear. Yet, understanding how a strain becomes resistant is an essential step to better adapt the treatment of yeast infections, or to develop new antifungals.

The fluconazole sensitivity is highly variable within the two species. To get a better understanding of the evolutionary processes that have occurred during the history of *S. cerevisiae* and *S. kluyveri* species we performed an evolving experiment of two sensitive strains. We exposed the two strains to fluconazole for 200 generations. We compared the growth rate of the evolved strains to the starting strain (ancestral strain) and characterised them by whole genome sequencing. The sequencing of the genome of six evolved strains (three from each species) reveals that the adaptation occurred by the formation of extrachromosomes. This common mechanism for these two genetically distant yeasts permit to have in one-step of evolution a higher fluconazole resistance. In the *S. kluyveri* evolved strains, two chimerical chromosomes were formed in two independent lineages. At the junction site we found transposable elements. By a gene dosage effect, some genes involved in the fluconazole resistance and located on the extrachromosomes induce the resistance.

## Results

### Selection of the strains to be evolved / Diversity in the fluconazole resistance in the yeast isolates

In order to select strains that we will use in the evolving experiment, we have assessed the diversity of fluconazole resistance in two yeast species: *S. cerevisiae* and *S. kluyveri*. The two strains will be distinguished for the rest of the article by the names Sace for the *S. cerevisiae* strain and Sakl for the *S. kluyveri* strain. In each collection, the resistance level at  $4 \cdot 10^{-5}$  M of fluconazole to 10 strains from different backgrounds was determined by growth in microcultivation in rich medium during 48 hours with and without drug (Figure 1). For the both species, the resistance is very variable because some strains aren't able to growth, and the other growth as well as without the drug. The *S. cerevisiae* strain YJM280 and the *S. kluyveri* strain NBRC10955 are sensitive to this concentration of fluconazole but they are still able to grow very slowly. That's why we select them to perform an evolving experiment on fluconazole.



**Figure 1. Resistance of *S. cerevisiae* (A) and *S. kluyveri* (B) strains to fluconazole.** The resistance is determined by the ratio between the growth rate in rich medium with drug and the growth rate in rich medium.

### Generation of evolved strains

To study the evolution of both haploid strains YJM280 (*S. cerevisiae*) and NBRC10955 (*S. kluyveri*) in an environment containing fluconazole, we carried out an evolving experiment in serial cultures. Briefly, each 2 days the same quantity of cells from the on-going culture was used to inoculate a 50 mL fresh medium supplemented of  $4 \cdot 10^{-5}$  M of fluconazole (see materials and methods). Two biological replicates R1 and R2 were achieved for each species. The experience was performed over 200 generations (approximately 1 month). Measuring the absorbance at 600 nm at the end of each 48 hours of culture allowed us to observe a better growth along the experiment until a maximum is reached (Figure S1).

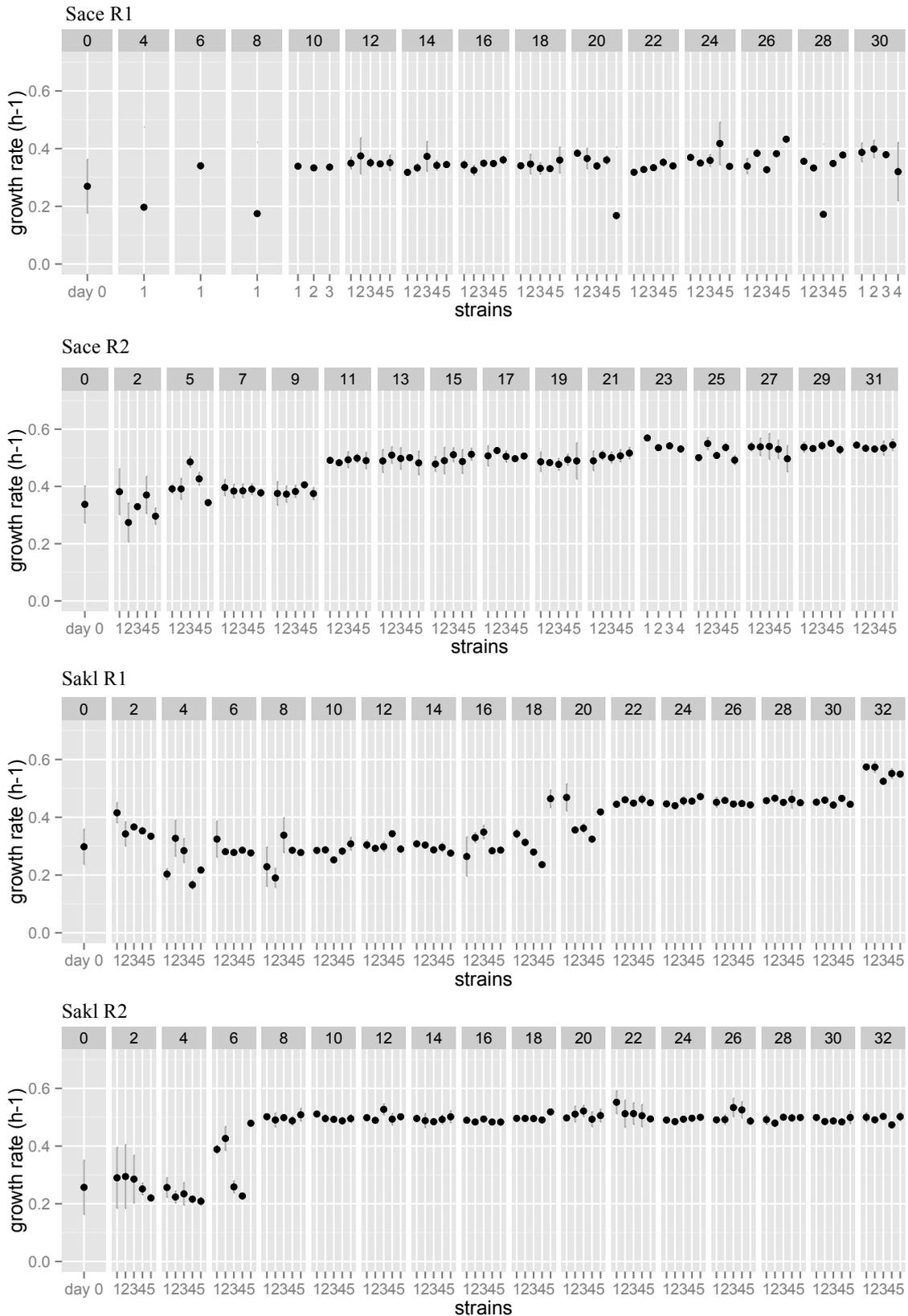
In less than one week the number of generations per 48 h between the subcultures of Sace R1, Sace R2, Sakl R1 and Sakl R2 has increased and reached a plateau between 12 and 13 generations. These results suggest that the 4 cultures became resistant. We wonder if there are subpopulations in the cultures/batch? Therefore, we analysed the resistance at the clonal level. Even if the number of generation per 48 h has evolved rapidly, we have analysed clones isolated during all the experiment. For this, 5 individual clones of each replicates were selected and subjected to precise phenotyping to detect any resistant cells.

### Appearance and evolution of the fluconazole resistance in the 4 evolved lineages

To know if there are subpopulations in the evolved cultures, microcultivations for 5 clones per subculture were carried out (Figure 2). The growth of each clone was performed in duplicate and followed by the reading of the absorbance at 600 nm every 10 minutes during 48 hours. The determination of the growth rate on fluconazole permits to see the adaptation of the growth on fluconazole within the 4 lineages.

The evolution of the growth rate within the two replicates of *S. cerevisiae* is different. The growth rate of the Sace R1 clones does not exceed  $0.4 \text{ h}^{-1}$  after one month of experiment so we decided to don't extend further the analyzes for it (Figure 2). In the Sace R2 experience, it seems to coexist clones with different resistance levels until day 7 (Figure 2A). At day 11, the growth rate on fluconazole enhance abruptly to  $0.49 \text{ h}^{-1}$  and then slower until the end at day 31. The growth rate has nearly doubled in one month of evolution.

The *S. kluyveri* experiments follow the same scheme (Figure 2C and 2D). The culture becomes suddenly resistant to fluconazole in one stage for Sakl R2 or in two stages for Sakl R1. The two replicates show an improved resistance and in each experiment one population derived from an evolved strain seems to take the advantage. To know if the resistance is specific to fluconazole or if the developed mechanism allows the resistance for other drugs we decided to test the resistance to other azoles compounds.



**Figure 2. Resistance to fluconazole of the evolved clones.** Growth rate of 5 clones per 2 days of experiment were analysed by microcultivation. Each graph represents one evolving experiment.

### **The acquired resistance is specific to fluconazole**

To analyse the specificity of the acquired fluconazole resistance, the growth rates on 5 azoles drugs for the same selected clones were tested by microcultivation (Figure 3). In addition to the fluconazole resistance we tested another triazole antifungal, the itraconazole and we tested three imidazoles: the ketoconazole, the clotrimazole and the mitraconazole. All these 5 drugs stop the ergosterol biosynthesis by linking to the lanosterol 14- $\alpha$ -demethylase enzyme (Sanglard et al., 2009). So if the growth rates to these compounds are the same in the evolved strains than the ancestral strains, it means that the mechanism of resistance is totally specific to fluconazole.

As we seen previously, the growth rate on fluconazole of the Sace R2 population is doubled at the end of the experiment and this phenomenon affects all the tested clones (Figure 3). The growth rates to other drugs do not followed the same trend and the values are specific to each clone. The resistance of Sace R2 evolved strains is specific to fluconazole.

In the Sakl R1 experiment, the evolved strains have in addition to the fluconazole resistance a higher growth rate on ketoconazole than the ancestral strain. But the final growth rate at day 32 is nevertheless higher on the fluconazole than ketoconazole. The growth rates to other drugs remain low even at the end of the experiment. In the Sakl R2 strains a gap is apparent between the fluconazole growth rate and the other drugs growth rates. So the resistance is completely specific to fluconazole.

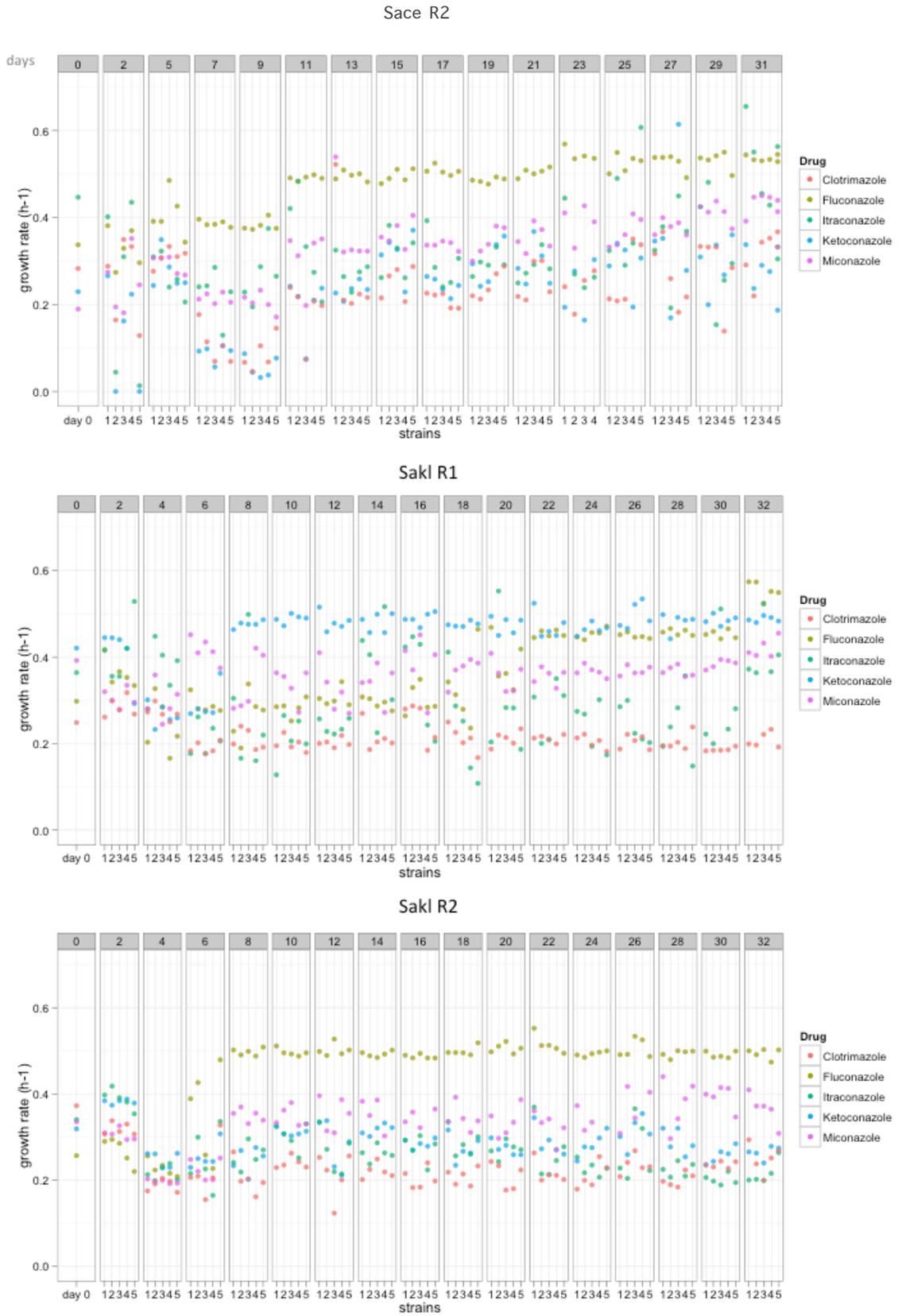
To conclude, the adaptation for the Sace R2 and Sakl R2 evolved lineages is specific to the fluconazole. We wonder what are the mechanisms at the origin of the adaptation to fluconazole in these 3 lineages. To see if large-scale rearrangements are involved, we analysed the karyotype of isolated clones by PFGE (Pulsed Field Gel Electrophoresis).

### **PFGE reveals modifications in the karyotype of 2 evolved strains**

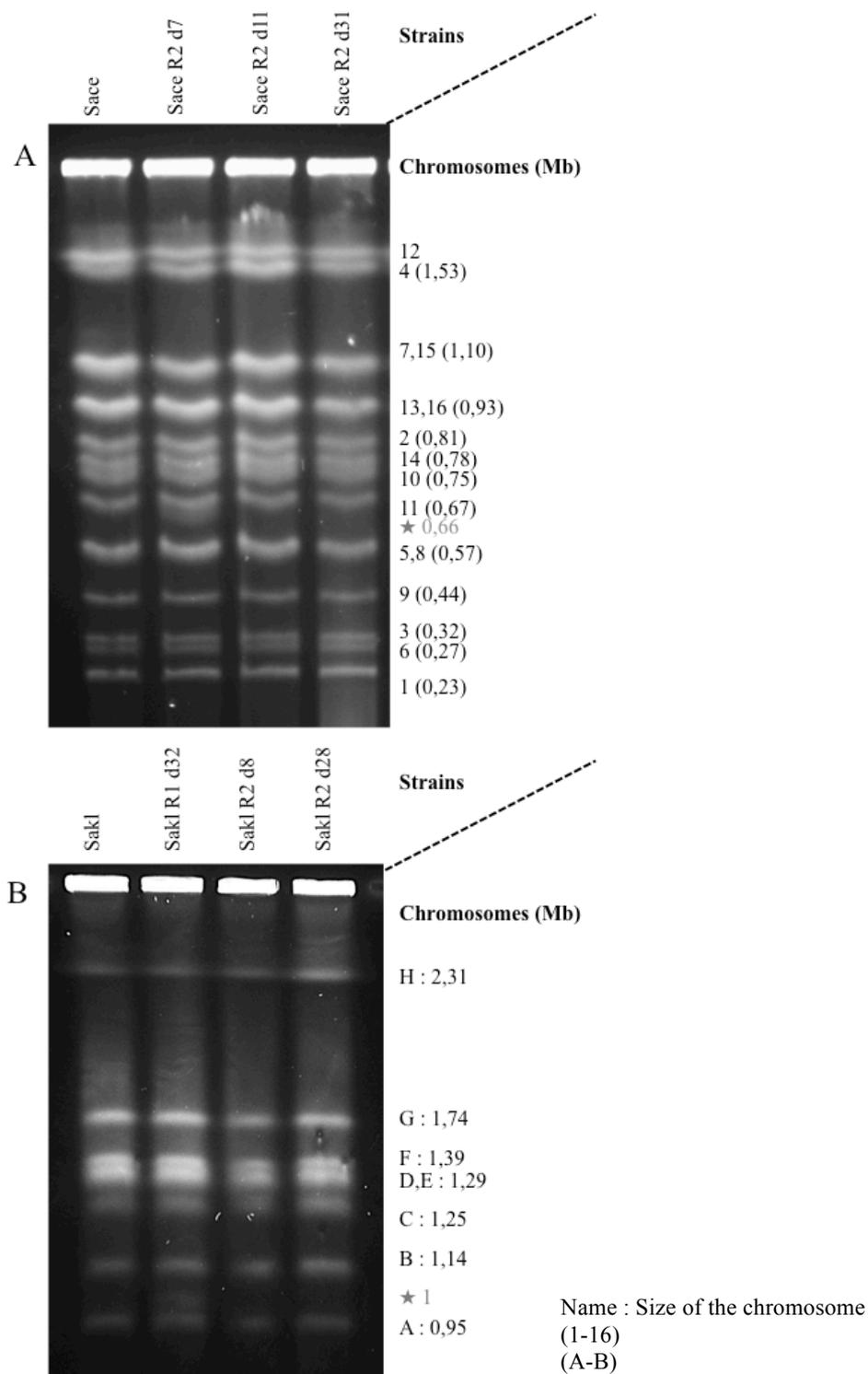
To perform the karyotype analyses, 3 clones per species were selected. They were chosen at important steps of the evolution: when the resistance has improved suddenly and at the end of the experience. For Sace R2 the evolved strain Sace R2d7 ("d7" for day 7), Sace R2d11 and Sace R2d31 were selected. For *S. kluyveri* species the R1d32, R2d8 and R2d28 clones were taken. On the previous figures 2 and 3, the selected clones for sequencing were labelled as clone 1 of the concerned day.

The karyotypes of the Sace R2d11 and Sace R2d31 strains show a similar profile than the ancestral strain (Figure 4A), whereas the clone Sace R2d7 shows an additional band of 0.66 Mb. The fact that the rest of the karyotype is conserved indicates that a supernumerary chromosome probably causes this additional band.

The karyotype of the Sakl R1d32 has an extra band around 1 Mb. The strains Sakl R2d8 and R2d28 have exactly the same one than the ancestral strain. To discover the origin of the additional bands, and see if there are other genomic variations, we sequenced the genome of the 6 evolved strains in addition to the ancestral strains.



**Figure 3. Resistance to azoles drugs of evolved clones.** Growth rate on clotrimazole (orange), fluconazole (green khaki), itraconazole (green), ketoconazole (blue) and miconazole (mallow) were measured by microcultivation. Each graph represents one evolving experiment.



**Figure 4. Large-scale chromosomal rearrangement occurred in 2 evolved strains.** PFGE of the six evolved strains of *S. cerevisiae* (A) and *S. kluyveri* (B) and the ancestral strains. The chromosomes are shown on the right side. The stars represents a band which is not present in the ancestral strain. (A) The strain R1D7 shows a new band of size approximately 0.66 Mb. (B) The strain R1D32 shows a new band of size approximately 1 Mb.

### Whole Genome Sequencing (WGS) of 6 evolved strains reveals extrachromosomes

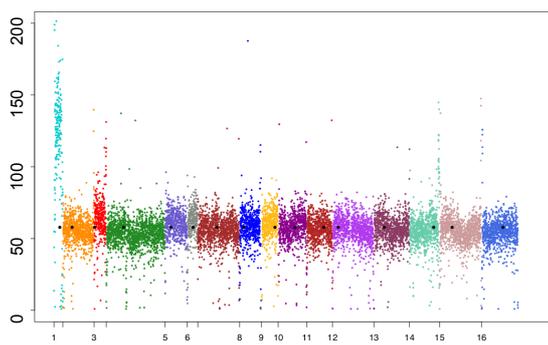
We subjected the genomic DNA of the ancestral strains and of the 6 evolved strains to deep sequencing using the Illumina HiSeq 2000 sequencing technique. To detect the copy number variations, the read depth of the sequenced strains was measured all along the genome. Therefore, the average of reads per base pair was estimated per 1 kb windows. The coverage representations along the ancestral genomes are illustrated in Figure 5.

For the *S. cerevisiae* strains, the mean coverage is 57x. The ancestral strain presents a doubling of coverage for the chromosome 1, which indicates a pre-existing aneuploidy (Figure 5a). The strain Sace R2d7 has an anomaly of the coverage of the chromosome 7 right arm. This anomaly is probably due to a sequencing problem. The high coverage of the chromosome 8 (around 80x) of the Sace R2d7, Sace R2d11 and Sace R2d31 strains reveals an additional aneuploidy. This aneuploidy appeared rapidly during the experiment and it may be the cause of the first increase of fluconazole resistance. In the evolved strains Sace R2d11 and Sace R2d31 another aneuploidy is detected; that of the chromosome 4 (coverage around 76x). These aneuploidies from the chromosome 4 and 8 may play a role in the fluconazole resistance. The acquire advantage permit to maintain them until the end of the experiment.

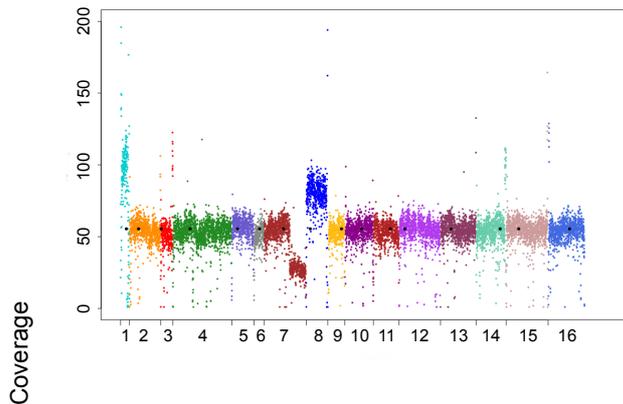
For the *S. kluyveri* strains, the mean coverage is 59x. The homogeneous coverage along the ancestral strain genome indicates its euploid state (Figure 5b). Local variations in the coverage reveal two segmental duplications for the Sakl R1d32 evolved strain. The first one is the duplication of a segment of the chromosome B ranging to coordinates 0 to 336395 bp and contains the centromere. The second is that of the chromosome G between the coordinates 0 and 688530 bp. The fact that the duplicated regions contain 2 telomeres and one centromere suggest that a chimeric extrachromosome appeared in this evolved strain. The same configuration is present in the Sakl R2 evolved strains. Two segmental duplications are observed; one of the chromosome G (0 to 688530) and the other of the chromosome H (394847 to end) (Figure 4b). The new chromosomes will be called BG and GH because of the chromosomes that compose them. Interestingly, the supernumerary chromosomes BG and GH contain exactly the same chromosome G portion. Therefore, we think that it plays probably a major role in the fluconazole resistance. The fact that two chimeric chromosomes appear in two independent experiments indicates that this mechanism is an efficient way to improve rapidly the resistance in the *S. kluyveri* species. In addition to this, the evolved strains R2d8 present an aneuploidy of the chromosome B and the strain R2d28 of the chromosomes B and C. In the R2\_d8 strain an anomaly in the coverage of the chromosome C is detect. A previous study shows a high GC content on the chromosome C left arm (Payen et al., 2009). Other unpublished results obtained in our laboratory suggest that high GC content may cause coverage anomalies.

A)

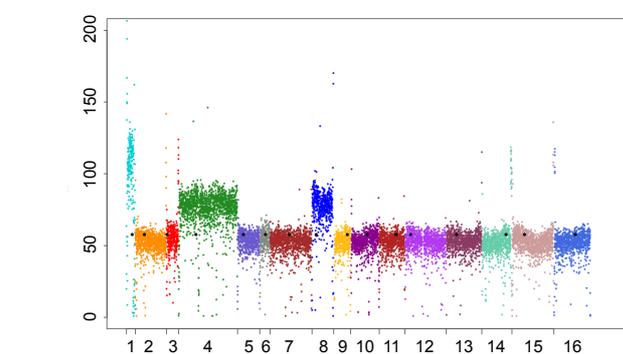
Sace



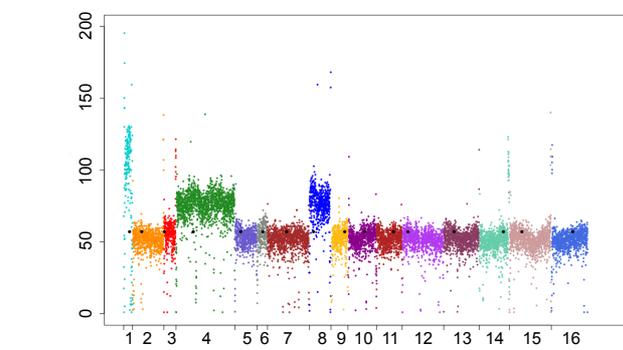
Sace R2d7



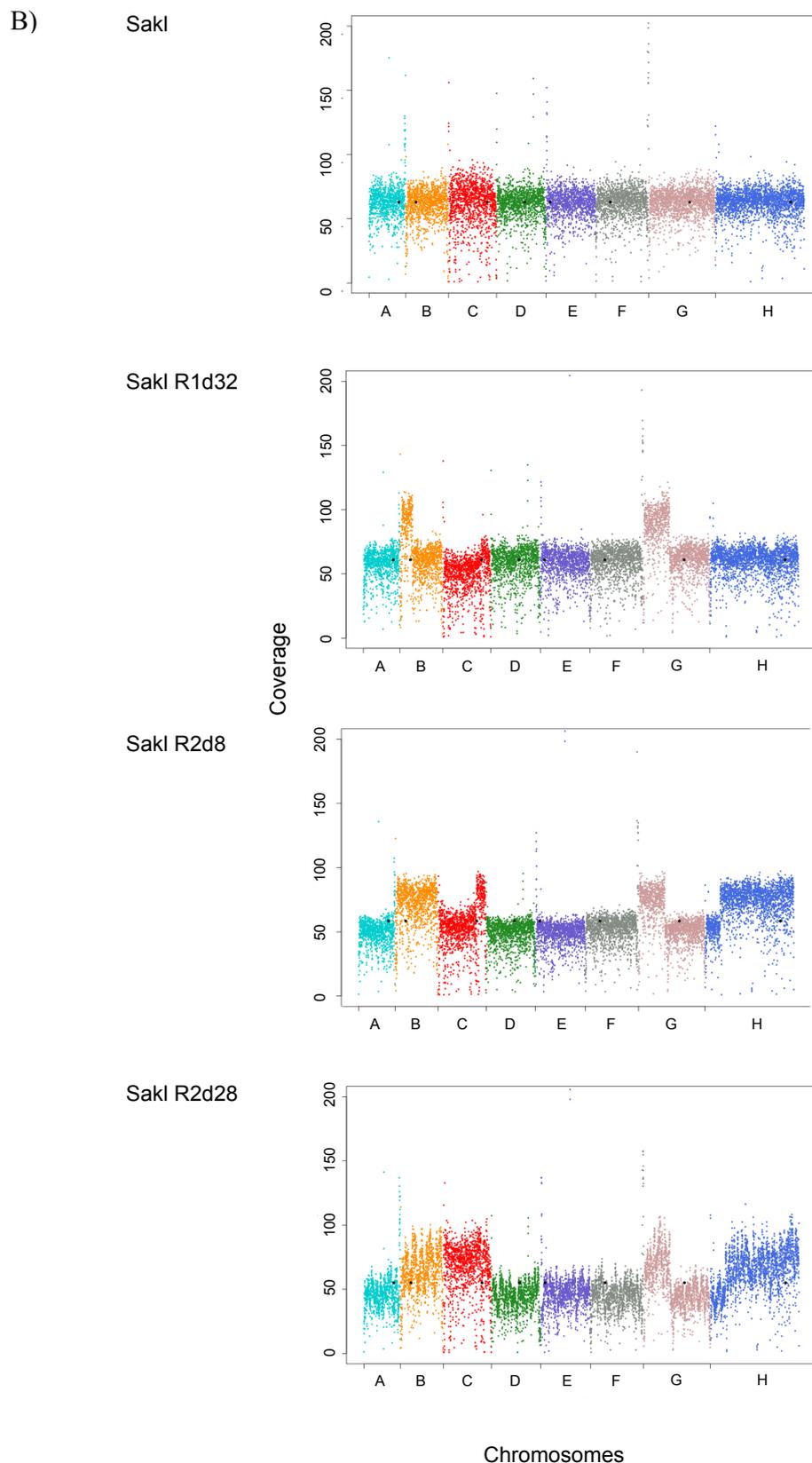
Sace R2d11



Sace R2d31



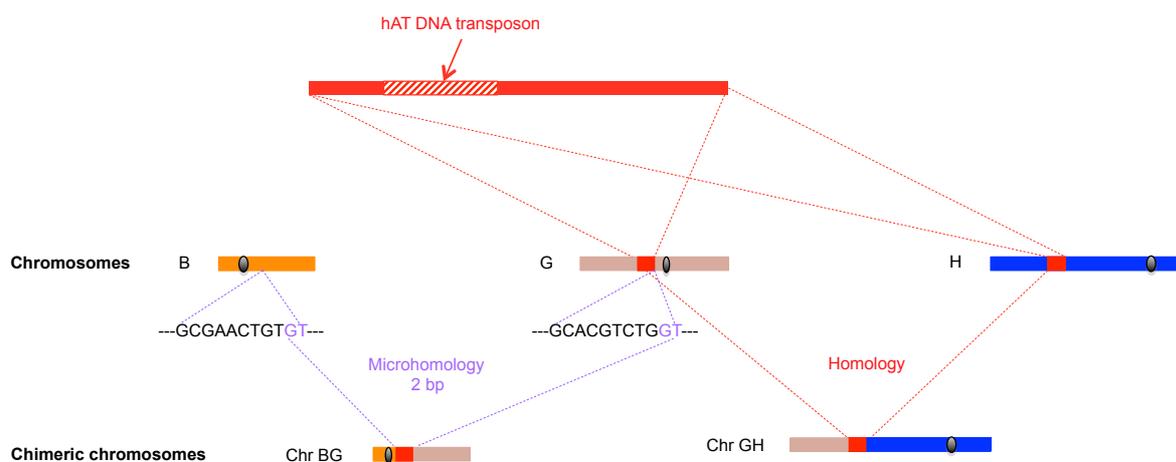
Chromosomes



**Figure 5. Copy number variations are identified in the evolved strains.** From the WGS data, a coverage analysis on 3 *S. cerevisiae* (A) and 3 *S. kluyveri* (B) evolved strains and the 2 ancestral strains were performed. Small black circles along each plot represent the centromeres.

### Transposable elements are present at the junctions of the *S. kluyveri* chimeric chromosomes

To confirm the presence of chimeric chromosomes and to understand how they were formed, we finemapped the junction sites of the putative rearranged chromosomes. We designed primers at 1 kb of the possible breakpoint according to the coverage analysis and use these primers to find out the junction site of the chromosomal fragments. The size of the generated amplicon was approximately 3300 bp higher than expected for both junction BG and GH. The sequencing of the junction reveals an additional sequence of 3366 bp containing a hypothetical ORF (Open Reading Frame) of 853 bp. This ORF has 51.5% of identity with a hAT-like DNA transposon of *Lachancea thermotolerans* (Figure S2). To know if the same transposon is also present into the chromosome B, G and/or H at the break point, we sequenced the regions located at the chromosomal break point. The sequencing shows the presence of the hAT DNA transposon into the chromosome G and H. The hAT transposon is absent at these locations in the reference strain of the species CBS3082. The bordering regions of 929 pb and 1584 pb of the the transposon hAT are the same between the chromosome G and H (Figure 6).



**Figure 6. Scheme of the chimeric chromosomes formation.** The homology of sequence of a 3366 bp region between the chromosomes G and H allowed the formation of the chromosome GH by homologous recombination. A microhomology of 2 bp between the chromosomes B and G is detected and has probably allowed the formation of the chromosome BG.

The insertion of the hAT DNA transposon into the chromosome G at the position 688530 interrupts the gene *SAKLOG08096g*. This gene encodes a putative protein of unknown function. The sequence of chromosome B corresponding to the junction is similar to the reference strain with no transposable element. The alignments of the junction BG with the chromosome B and G reveals a microhomology of 2 nucleotides (GT) that is located at the position 336406 - 336407 of the chromosome B.

The transposon copy present into the chromosome H is located at 394836 pb in an intergenic region between the gene *SAKLOH04224g* similar to *ADE6* gene of *S. cerevisiae* and the gene *SAKLOH04246g* similar to *ERG25* of *S. cerevisiae*. Both transposons copies in chromosomes G and H have nearly the same sequence with 97.3 % of identity except an additional repeated sequence of 25 bp (5' TGAATCGGTTTGTGAATCGGTTTCGT 3') in the chromosome H copy.

The presence of transposable elements at the break junctions and in the junctions suggests their involvement in the chimeric chromosomes formation.

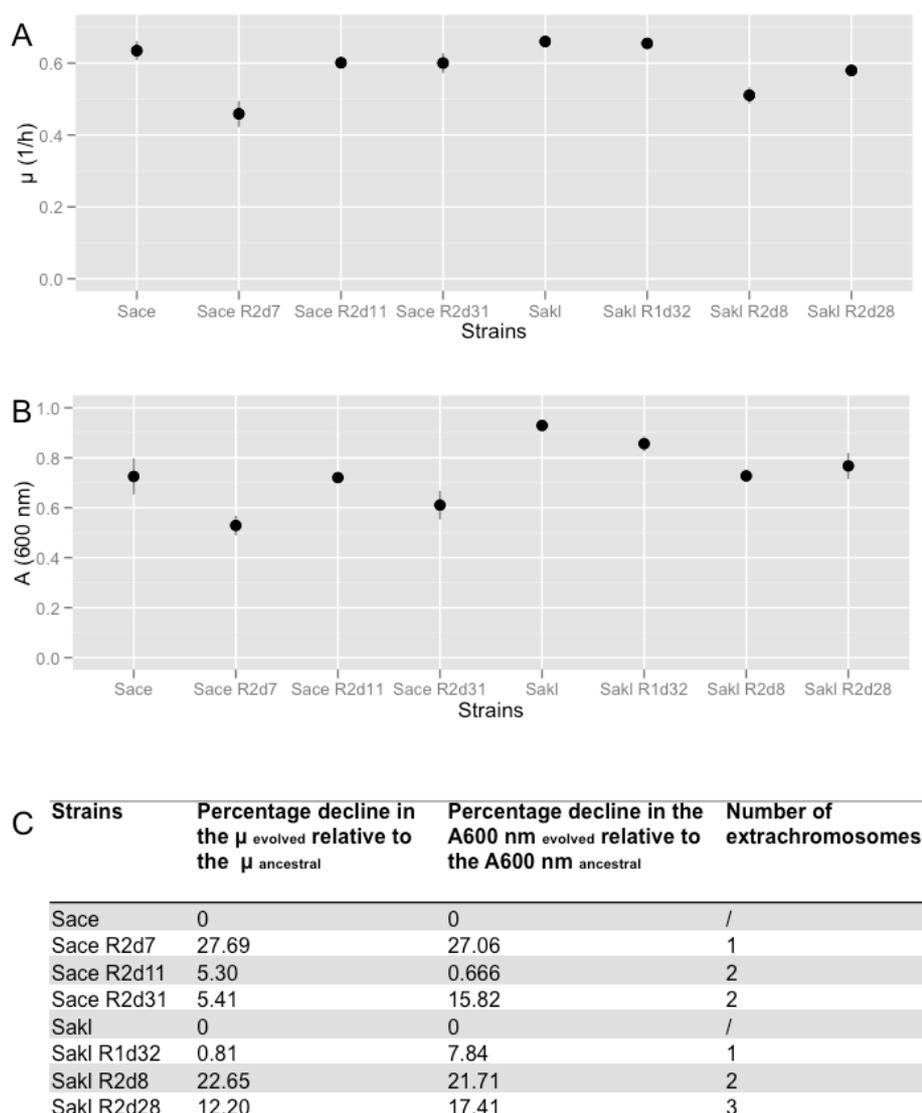
#### **Cost associated with extrachromosomes**

The extrachromosomes allow a better growth on fluconazole so a better resistance, but what is about growth rate on the rich medium? Maintaining the extrachromosomes might have a cost for the cell.

We hypothesis that the one, two or three additional chromosomes of the evolved strains contribute to fluconazole resistance. A previous study reveals that aneuploidies in yeast at the contrary to aneuploidies in cancer cells gives a proliferative disadvantage with a lower growth rate, a lower biomass and a increase in glucose uptake (Torres et al., 2007). To estimate the cost of the supernumerary chromosomes, microcultivations of 48 hours in YPD of the ancestral and the evolved strains were done. Then, growth rate and the finale absorbance at 600 nm were measured (Figure 7). The absorbance (A) at 600 nm permits to measure the biomass generated during the culture.

For the both species, the growth rate and the  $A_{600}$  nm of the ancestral strain are higher than those of evolved strains (Figure 7 A and B). But the differences are small. The supernumerary chromosomes gained during the evolving experiment confer an advantage in presence of fluconazole but not in non-selective environment.

We compared the growth rate and the Absorbance 600 nm with the number of extrachromosomes but there is no correlation (Figure 7 C).



**Figure 7. Aneuploidy cost.** Growth rate of the ancestral and evolved strains on YPG (A). Estimation of the decrease of the evolved growth rate relative the ancestral growth rate (B).

### The resistance origin in the *S. kluyveri* evolved strains

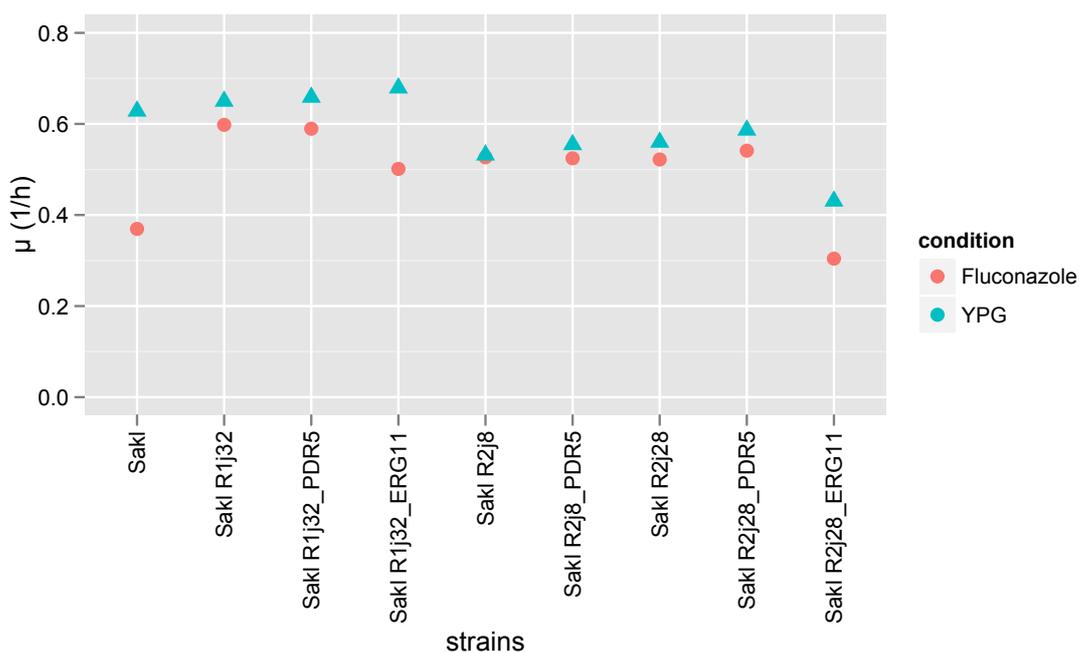
The same part of the chromosome G is involved independently two times in the adaptation of the two biological replicas. That suggests its strong involvement in the fluconazole resistance. We hypothesis that the resistance is mainly due to the two copies of one or several genes on this chromosomal region, so we choose to focus on it. To highlight the candidate(s) gene(s) a functional analysis was performed. The research results in two candidate genes: *SAKL0G08008g* homolog to the *PDR5* gene of *S. cerevisiae* and

*SAKL0G06336g* homolog to *ERG11* gene. *PDR5* encodes a transporter ABC (ATP Binding Cassette) able to export tens of compounds including the fluconazole (Ernst et al., 2008). *ERG11* encodes lanosterol 14- $\alpha$ -demethylase, an enzyme that catalyzes the C-14 demethylation of lanosterol to form 4,4'-dimethyl cholesta-8,14,24-triene-3- $\beta$ -ol during the ergosterol biosynthesis (Karst and Lacroute, 1977). This enzyme is a target of the azole drugs including fluconazole (Lupetti et al., 2002). In *Candida* species, it can be at the origin of resistance to antifungal by mutations into the coding sequence (Marichal et al., 1999) or by its upregulation (Henry et al., 2000).

To test the involvement of these genes, we delete one copy of the candidate gene at least in each evolved replicat R1 and R2. The fluconazole susceptibility of the mutants was measured by microcultivation. The growth rate was calculated in presence or absence of fluconazole.

The *PDR5* mutants (Sakl R1j32\_*PDR5*, Sakl R2j8\_*PDR5* and Sakl R2j28\_*PDR5*) grow on YPG as well as the evolved strain (figure 8). So the deletion has no impact on the growth on YPD. There is no difference between the growth in presence or absence of fluconazole. So the second copy of the *PDR5* gene plays no roles in the increase of the fluconazole resistance.

The *ERG11* mutants of each replicate have different profiles. The Sakl R1d32\_*ERG11* grows on YPD as well as the evolved strain (figure 8). By comparison to the YPG culture, there is a decrease of the growth rate of 26.1 % on fluconazole. That means that the additional copy of *ERG11* has an impact on the fluconazole resistance. In the replicate 2, the growth of Sakl R2d28 on YPD reveals a growth defect compared to the wild evolved strain, which means that the deletion of one copy of the *ERG11* genes results in a loss of fitness in the strains of the replicate 2. The growth on fluconazole falls of 29.4 % compared to the growth rate on YPD. The *ERG11* gene plays also a role in the resistance to fluconazole in the replicate 2.



**Figure 8. Resistance on fluconazole analyses of the mutants.** Growth rate on YPD and YPD supplemented by fluconazole of the evolved and mutant strains.

#### The resistance origin in the *S. cerevisiae* evolved strains

During the *S. cerevisiae* experience two aneuploidies appeared. A functional analysis of the gene presents on the duplicated chromosomes 4 and 8 permit to select the genes that can explain the resistance of the evolved strains. We select genes involved in drug resistance such as the *MDR* genes (Multi Drug Resistance) or *PDR* (Pleiotropic Drug Resistance) because the azole resistance in *S. kluyveri* remain unexplored and no candidate gene has to be put aside. We chose 4 genes for further analyses: *PDR15* and *SNQ2* on the chromosome 4, *YHK8* and *SSZ1* on the chromosome 8. *PDR15* and *SNQ2* are two PDR transporters able to export antifungals and xenobiotics out of the cell. They are nine known members in *S. cerevisiae*. *PDR15* is the paralogue of the so known *PDR5* that appeared after the Whole Genome Duplication (WGD) of the Hemiascomycetous lineage (Seret et al., 2009). *SNQ2* is involved in drug response, but also in cation resistance and in quorum sensing for yeast populations growing in liquid culture (Hlaváček et al., 2009; Miyahara et al., 1996). *YHK8* code an antiporter of the Major Facilitator Superfamily. It is upregulated in strain sensitive to azoles drugs (Barker et al., 2003). *SSZ1* is an Hsp70 protein involved in PDR resistance via sequential activation of *PDR1* and *PDR5* (Hallstrom et al., 1998).

We will delete one copy of these genes in the three evolved strains by replacing it with the *hph* gene (Hygromycin resistance) and then, perform the microcultivation phenotyping.

## Discussion

Despite the importance of studying the acquisition of resistance in yeasts, relatively few direct experimental data describing the underlying processes of adaptation to antifungals have been done. By combining adaptive evolving experiments with high throughput sequencing technologies, we have analysed the evolution of sensitive strains of two Hemiascomycetes yeasts: *S. cerevisiae* and *S. kluyveri*.

The evolving experiments that we performed in this work lead to the rapid adaptation of two sensitive strains to the growth on fluconazole. The evolved strains growth 1.5 at 2 times faster than the ancestral stains under fluconazole stress. Despite the evolutionary distance between the two yeasts, a common mechanism of adaptation has been identified: the formation of extrachromosomes. The supernumerary chromosomes are due to chromosomal duplication in *S. cerevisiae* and *S. kluyveri* species and chimerical chromosomes in *S. kluyveri* species.

The fact that two chimeric chromosomes were formed in independent experiments underlines the importance of this adaptation mechanism in the *S. kluyveri* species. Their formations are probably explained by two different mechanisms. The chimeric chromosome BG of the evolved strain Sakl R1d32 appeared probably per duplication of the entire chromosomes B and G and homologous recombination thanks to the repeated sequences of the transposons. Transposons are already known for their implication in chromosomal rearrangements (Mieczkowski et al., 2006). So even if they aren't functional, transposons allow to evolve in a short scale of time. For the chromosome GH of the evolved strains Sakl R2d8 and Sakl R2d28, there is a microhomology of 2 nucleotides between the chromosome G and H. The appearance of this chromosome is probably due to a BIR (Break-Induce Replication) mechanism (Kraus et al., 2001). Because the transposon content is specific to each strain, it would be interesting to perform the same experiment with other strains of *S. kluyveri* that are lacking the hAT transposon at these positions. By this way we can estimate the importance and frequency of the chimeric chromosomes formation in this species.

Previous studies have shown the multiples side effects of aneuploidies (cells possessing a chromosome number that is either more or less than the normal number). Some effects are common to several aneuploidies and other are specific. For example, the “fluffy” phenotype of the colonies in the *S. cerevisiae* species is shared by disomie of chromosome I, VIII, XIV and IX (Tan et al., 2013). Aneuploidies have also an impact on the transcriptome by approximately doubling the expression of the genes on the duplicated region, they exhibit increased glucose uptake, and they have a delay in the G1 phase of the cell cycle (Torres et al., 2007). Another study based on the transcriptome analysis of 16 disomes of *S. cerevisiae* shows that a part of the transcriptional changes caused by aneuploidy are related to slow growth and/or stress (Sheltzer et al., 2012). Aneuploidies can also confer a fitness advantage under diverse conditions as the extreme temperatures and pH, high salt, and presence of

antifungals (Pavelka et al., 2010). In our study the aneuploidies of the evolved strains confer a better growth rate on YPD in presence of fluconazole.

Cases of azoles drug resistance due to aneuploidies have already been reported in the fungal pathogens species *Cryptococcus neoformans*, *Candida glabrata* and *Candida albicans* (vanden Bossche et al., 1992; Poláková et al., 2009; Selmecki et al., 2006; Sionov et al., 2009). In addition to these data on these pathogenic yeasts, our study has permitted to see the formation of additional chromosomes in real time in two non-pathogenic Hemiascomycetes. This suggests that the formation of extrachromosomes is not dependent of the life style of the yeast. So all together, these articles suggest that the extrachromosomes are a conserved mechanism among the yeasts that allows a fast adaptation to an antifungal stress. The most studying yeasts in evolving experiment are *S. cerevisiae* and *C. albicans*. As we introduce in this study, it would be wise to test the adaptation to fluconazole of other species as *L. thermotolerans* or *Yarrowia lipolytica*. Other mechanisms of adaptation can be found by this way.

Keeping a supernumerary chromosome has a cost for the cell and we estimate it by the measurement and comparison of the growth rate and the finale absorbance of the culture on rich medium of ancestral and evolved strains. Several studies have proved the genomic instability of aneuploid cells as well by a gain or loss of chromosome but also by defective DNA damage repair (Jin Zhu et al., 2012; Sheltzer et al., 2011). So the formation of extrachromosome is probably the first step to adapt in a new and stressfull environment. If the stress persists, we can imagine than alternative solutions less costly and more stable, will be established as it had been observed during an extended experiment in high temperature (Yona et al., 2012).

## Materials and methods

### Yeast strains and media

The strains used are listed in table 1. The *Saccharomyces cerevisiae* haploid strain used is YJM280  $\Delta HO::KanMX$ . The wild strain was isolated from an American AIDS patient. The *Saccharomyces kluyveri* haploid strain NBRC10955 has been isolated from a decaying leaf of *Rhizophora mucronata* in Japan. Both of these strains are referred as the ancestral strains. Cells were cultured in rich medium (YPD: 1% yeast extract, 2% bactopectone, 2% glucose) supplemented with  $4 \cdot 10^{-5}$  M of fluconazole. The microcultures experiments were performed in YPD supplemented by different azoles drugs in the following concentrations:  $4 \cdot 10^{-5}$  M of fluconazole,  $10^{-5}$  M of itraconazole,  $5 \cdot 10^{-6}$  M of ketoconazole,  $5 \cdot 10^{-6}$  M of clotrimazole,  $5 \cdot 10^{-7}$  M of miconazole.

Specie	Strain	Ecologic origin	Geographic origin
<i>S. kluyveri</i>	CBS 4104	soil	Netherlands (Wageningen)
	CBS 4568	soil	Sweden
	55-86.1	Salix exudate	California, USA
	68-917.2	old Acer flux and insect frass (maple)	Canada
	CBS 6546	frass on Populus sp	USA, California
	NRBC 10955	Decaying leaf, Rhizophora mucronata	Japan
	NRBC 105172	Moss	Thailand
	NRBC 101999	Exudate of <i>Ulmus</i>	Spain, Conde Park in Cordoba
	NCYC 616	NRRL YB-4288-7,FX, slime flux of willow	USA
	CBS 6548	Drosophila pseudoobscura (fruit fly)	USA, California, Gualala Creek
	YJM 440	Clinical	United States
	YJM 280	Clinical. peritoneal fluid	Unites States
	PW5	palm wine	Nigeria
	DBVPG 4651	Tuber Magnatum	Italy
<i>S. cerevisiae</i>	Y12	Palm Wine	Ivory Coast
	YJM 269	Apple juice	Unknown
	FY5	Laboratory	Unknown
	EM93_3	Rotting Fig	California
	Y55	Wine	France
	DBVPG 6044	Unknown	West Africa

**Table 1.** List of strains used.

### Evolution experiments

All experiments were carried out by serial dilution. The 50 mL yeast cultures were grown at 30°C at 120 rpm to stationary phase. From these cultures,  $4 \cdot 10^4$  cells were inoculated in a new 50 mL cultures every 48 h until approximately 200 generations, when significant phenotype change was detected. Every 15 generations, 5 clones were picked up and subjected to high-throughput phenotyping using microcultivation approach.

### Microcultures analysis

The clones were subjected to precise phenotyping in YPD medium using a high-resolution microcultivation approach with microplate reader Tecan infinity 2000. One 24 h pre-cultivation in YPD were followed by a 48-h cultivation in YPD with  $4 \cdot 10^{-5}$  M of fluconazole. Incubation was kept at  $30^{\circ}\text{C} \pm 0.5^{\circ}\text{C}$ . The plates were subjected to shaking at 300 rpm. Strains were run in duplicate on separate plates. Readings of absorbance (600nm) were taken every 10 min, which allows determining growth curves. The growth rate, and the population doubling time were extracted from high-density growth curves and  $\log_2$  transformed. The mitotic proliferation rate (population doubling time), lag (population adaptation time) and efficiency (total change in population density) were extracted from high density growth curves and  $\log_2$  transformed as previously described (Warringer et al., 2011). The resistance to fluconazole of the strain collections of *S. cerevisiae* and *S. kluyveri* was calculated by the ratio between the growth rate on YPD and the growth rate on YPD supplemented by fluconazole.

### Pulsed field gel electrophoresis

A total of  $1-2 \cdot 10^8$  yeasts cells were used for plug preparation as described in Carle et al., 1985. PFGE was performed using a CHEF DR-III system (Bio-Rad) in a 1% agarose gel in 0.5 X TBE buffer and run at  $14^{\circ}\text{C}$ . The programme for *S. cerevisiae* is for 24 h at 6 V/cm with an initial switching time of 60 seconds, a final of 120 seconds, and an angle of  $120^{\circ}$ . The electrophoresis for *S. kluyveri* followed these conditions: 20 h, 480 s pulse time at 3 V/cm; 20 h, 300 s pulse time at 4 V/cm; and 12 h, 90 s pulse time at 6 V/cm (Neuvéglise et al., 2000).

### Whole Genome Sequencing and coverage analysis

For the DNA extraction, yeast cell cultures were grown overnight at  $30^{\circ}\text{C}$  in 20 mL of YPD medium up to the early stationary phase ( $6-8 \cdot 10^8$  cells). Total genomic DNAs were subsequently extracted using the QIAGEN Genomic-tip 100/G in line with the manufacturer's instructions.

The ancestral strains and the selected evolved clones were sequenced with the Illumina HighSeq 2000 technique. We used paired-end libraries of 500 pb, 100 pb/read and 50X coverage. The reads were aligned using BWA (<http://bio-bwa.sourceforge.net/bwa.shtml>) against the reference genome (S288c for *S. cerevisiae* and CBS 3082 for *S. kluyveri*). The average number of reads at each bp of 1 kb allowed us to estimate the coverage along the genome.

### Chromosome walking and sequence analyses

Chromosome walking was used to sequence the junctions of the chimeric chromosomes and in the breakpoints in the original chromosomes. The fragments were amplified using the Taq polymerase from MP Biomedicals as described by the manufacturer. The primers are listed in the table 2.

Sequences were aligned and assembled using Jalview (Waterhouse et al., 2009). Junctions sequence analyses were performed by BlastN and tBlastX on the Génolevures website

(<http://genolevures.org/blast.html>). Pairwise local alignments between the junctions and the DNA transposon KLTH0H03344t were performed using EMBOSS matcher on European Bioinformatics Institute website ([http://www.ebi.ac.uk/Tools/psa/emboss\\_matcher/nucleotide.html](http://www.ebi.ac.uk/Tools/psa/emboss_matcher/nucleotide.html)).

Target	Strain	Step of the sequenci	Primers	Sequence
junction B	day 0		SAKLB_1kbjct_F	CTAATATGGGGTGTGTCCTA
			SAKLB_1kbjct_R	AAATTGCTTGACTTTCTCCC
junction G	day 0	step 1	SAKLG_1kbjct_F	GAAATAGTTCATTGATGAGC
			SAKLG_1kbjct_R	ACGGTTCAATGATCTTCCAT
		step 2	SAKLG32_A_F	ACCTAGTCTGCGTGCGTCTG
		step 3	SAKLG32_B_F	GGATGGACAAATCCAAATTC
		step 4	SAKLG_C_F	TGGAGTCCCATGTTGCGATA
		step 5	SAKLG_D_F	AGGCGTAATCCACAACGGTT
junction H	day 0	step 6	SAKLG_E_F	TGAATCGGTTTAACTCATTG
		step 1	SAKLHj0_AF	CATTCCTGGTTGTGAAAACT
			SAKLHj0_AR	TCTAAACAAATTCITTTCCAT
		step 2	SAKLHj0_BF	TACTGCGTATTACTTTTCAG
			SAKLHj0_BR	GTACAGATATAAATGTACG
		step 3	SAKLHj0_CR	GGCTGTCTTGCACTGCTTGT
junction BG	1d32		SAKLHj0_CF	GATATGCTTGTAGAAAAAC
		step 1	SAKL_500pbjct_F	TTTCTTTTCGGACTCCTGTG
			SAKLG_1kbjct_F	GAAATAGTTCATTGATGAGC
		step 2	SAKLB_A_F	ATATGCTACGTGAGGGCTAC
			SAKLG32A_F	ACCTAGTCTGCGTGCGTCTG
		step 3	SAKLB_B_F	TCTTGCAACTGCTTGTCTG
junction GH	2d8, 2d28		SAKLG32B_F	GGATGGACAAATCCAAATTC
		step 1	SAKLG_1kbjct_F	GAAATAGTTCATTGATGAGC
			SAKLOH_1kbj28_R	TGTCCTATAATCATCTTAAA
		step 2	G28A_F	AAGGGCTAACCATTAGATTT
			H28A_R	ACGTCAAACCAAGTGATAGT
		step 3	H8B_R	TCCGATATGACGGCTAAAGC
			H28B_R	TGATGATGAGCAAGAAGCA
		step 4	H28C_R	TCAACACTTCGACATCCAAG
			H8C_R	CTGTTTCAAGAGTGTCTAC
			G8C_F	TATCCCAACCCATAAGACG
		step 5	G8D_F	GGATTTGAAGCCACTGAGCT
			H8D_R	ACTATCACGTTTTACTTATA
step 6	G8E_F	CGGTCGCTCATACCTGGCTA		
	H8E_R	GAACCGTTGTGGATTACGCC		

**Table 2.** Primers used for the sequencing of the junctions BG and GH.

### Gene deletion

All the fragments were amplified using the iProof High-Fidelity DNA Polymerase from MP Biomedicals as described by the manufacturer. The disruptions of the *PDR5* and *ERG11* gene in the evolved NBRC10955 strains and the *PDR15*, *SNQ2*, *PDR13* and *YHK8* genes in the evolved strains YJM280 were performed by homologous recombination. The *ERG11::KanMX* and the *PDR5::KanMX* cassettes were amplified by fusion PCR. The primers at the junction of two fragments have an overlapping sequence of 20 pb with the other fragment. They are listed in table 3. First, the three fragments were amplified separately: 1) *KanMx* gene 2) approximately 200 bp up and 3) downstream of the deleted *ERG11* gene. The fragments were purified by gel extraction using the kit QiAquick from Qiagen and following the manufacturer's protocol. A first PCR without primers was performed

with an equimolar quantity of the purified fragments with a melting temperature of 60 degrees. 5  $\mu$ L of this first PCR were used as template for the second classical PCR. The resulting cassette was integrated into the yeast by electroporation transformation and stable G418-resistant transformants were selected (Gojković et al., 2004). Correct integration was confirmed by PCR.

Deletion	Primer	Matrix	Sequence
SAKLG: 508500 to 594200	G_508500'_F	NRBC10955b	agcactaaaaaacagatt
	G_508719_R	NRBC10955b	GTCGACCTGCAGCGTACGatgggtacaatccatatct
	G_594000'_F	NRBC10955b	CGAGCTCGAATTCATCGATatatggtaatatgtgactct
	G_594200'_R	NRBC10955b	cgactgcaattacaatt
SAKLG: 594000 to 594467	Kan500'_F	BY $\Delta$ PDR5::KanMX	agatatggattgtaaccaatCGTACGCTGCAGGTCGAC
	Kan500'_R	BY $\Delta$ PDR5::KanMX	agagtcacatattaccatatATCGATGAATTCGAGCTCG
	G_594000'_F	NRBC10955b	atatggtaatatgtgactct
	G_594467'_R	NRBC10955b	GTCGACCTGCAGCGTACGaagcatacaatctcagtcgg
ERG11	G_687400'_F	NRBC10955b	CGAGCTCGAATTCATCGATatcgtgccactttgaaata
	G_687900'_R	NRBC10955b	accagaaataccttggaac
	Kan600'_F	BY $\Delta$ PDR5::KanMX	ccgactgagattgtatgcttCGTACGCTGCAGGTCGAC
	Kan600'_R	BY $\Delta$ PDR5::KanMX	tattttcaaagtggcacgatATCGATGAATTCGAGCTCG
ERG11	ERG11_530330F	NRBC10955b	ataagcaccagcagaaggag
	ERG11_530540R	NRBC10955b	GTCGACCTGCAGCGTACGagatgtgttggtgagcattt
	ERG11_531760F	NRBC10955b	CGAGCTCGAATTCATCGATgtagaagaccattggaacac
	ERG11_531960R	NRBC10955b	gaaagattcaaacatgtct
	KanMx_erg11F	BY $\Delta$ PDR5::KanMX	aaatgctaccaacacatctCGTACGCTGCAGGTCGA
KanMx_erg11R	BY $\Delta$ PDR5::KanMX	gtgtccaatggtcttctacATCGATGAATTCGAGCTCG	

**Table 3.** Primers used for the gene deletions by *KanMx* and *HphMx*.

## Acknowledgements

We thank the BGI (Beijing Genomics Institute) for the sequencing of the evolved strains. This work was supported by the ANR blanc GB-3G. C.R. is supported by a grant from CNRS and Région Alsace.

## References

- Barker, K.S., Pearson, M.M., and Rogers, P.D. (2003). Identification of genes differentially expressed in association with reduced azole susceptibility in *Saccharomyces cerevisiae*. *J. Antimicrob. Chemother.* *51*, 1131–1140.
- Vanden Bossche, H., Marichal, P., Odds, F.C., Le Jeune, L., and Coene, M.C. (1992). Characterization of an azole-resistant *Candida glabrata* isolate. *Antimicrob. Agents Chemother.* *36*, 2602–2610.
- Ernst, R., Kueppers, P., Klein, C.M., Schwarzmüller, T., Kuchler, K., and Schmitt, L. (2008). A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 5069–5074.
- Gojković, Z., Knecht, W., Zameitat, E., Warneboldt, J., Coutelis, J.-B., Pynyaha, Y., Neuveglise, C., Møller, K., Löffler, M., and Piskur, J. (2004). Horizontal gene transfer promoted evolution of the ability to propagate under anaerobic conditions in yeasts. *Mol. Genet. Genomics MGG* *271*, 387–393.
- Hallstrom, T.C., Katzmann, D.J., Torres, R.J., Sharp, W.J., and Moye-Rowley, W.S. (1998). Regulation of transcription factor Pdr1p function by an Hsp70 protein in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* *18*, 1147–1155.
- Henry, K.W., Nickels, J.T., and Edlind, T.D. (2000). Upregulation of ERG genes in *Candida* species by azoles and other sterol biosynthesis inhibitors. *Antimicrob. Agents Chemother.* *44*, 2693–2700.
- Hlaváček, O., Kucerová, H., Harant, K., Palková, Z., and Váchová, L. (2009). Putative role for ABC multidrug exporters in yeast quorum sensing. *FEBS Lett.* *583*, 1107–1113.
- Jin Zhu, Pavelka, N., Bradford, W.D., Rancati, G., and Rong Li (2012). Karyotypic Determinants of Chromosome Instability in Aneuploid Budding Yeast. *PLoS Genet.* *8*, 1–44.
- Kanafani, Z.A., and Perfect, J.R. (2008). Resistance to Antifungal Agents: Mechanisms and Clinical Impact. *Clin. Infect. Dis.* *46*, 120–128.
- Karst, F., and Lacroute, F. (1977). Ergosterol biosynthesis in *Saccharomyces cerevisiae*: mutants deficient in the early steps of the pathway. *Mol. Gen. Genet. MGG* *154*, 269–277.
- Kraus, E., Leung, W.Y., and Haber, J.E. (2001). Break-induced replication: A review and an example in budding yeast. *Proc. Natl. Acad. Sci. U. S. A.* *98*, 8255–8262.
- Lamb, D., Kelly, D., and Kelly, S. (1999). Molecular aspects of azole antifungal action and resistance. *Drug Resist. Updat.* *2*, 390–402.
- Lupetti, A., Danesi, R., Campa, M., Del Tacca, M., and Kelly, S. (2002). Molecular basis of resistance to azole antifungals. *Trends Mol. Med.* *8*, 76–81.

Maertens, J.A. (2004). History of the development of azole derivatives. *Clin. Microbiol. Infect.* *10*, 1–10.

Marichal, P., Koymans, L., Willemsens, S., Bellens, D., Verhasselt, P., Luyten, W., Borgers, M., Ramaekers, F.C., Odds, F.C., and Bossche, H.V. (1999). Contribution of mutations in the cytochrome P450 14 $\alpha$ -demethylase (Erg11p, Cyp51p) to azole resistance in *Candida albicans*. *Microbiol. Read. Engl.* *145 (Pt 10)*, 2701–2713.

Mieczkowski, P.A., Lemoine, F.J., and Petes, T.D. (2006). Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair* *5*, 1010–1020.

Miyahara, K., Mizunuma, M., Hirata, D., Tsuchiya, E., and Miyakawa, T. (1996). The involvement of the *Saccharomyces cerevisiae* multidrug resistance transporters Pdr5p and Snq2p in cation resistance. *FEBS Lett.* *399*, 317–320.

Neuvéglise, C., Bon, E., Léplinge, A., Wincker, P., Artiguenave, F., Gaillardin, C., and Casarégola, S. (2000). Genomic exploration of the hemiascomycetous yeasts: 9. *Saccharomyces kluyveri*. *FEBS Lett.* *487*, 56–60.

Pavelka, N., Rancati, G., Zhu, J., Bradford, W.D., Saraf, A., Florens, L., Sanderson, B.W., Hattem, G.L., and Li, R. (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* *468*, 321–325.

Payen, C., Fischer, G., Marck, C., Proux, C., Sherman, D.J., Coppée, J.-Y., Johnston, M., Dujon, B., and Neuvéglise, C. (2009). Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome Res.* *19*, 1710–1721.

Poláková, S., Blume, C., Zárate, J.A., Mentel, M., Jørck-Ramberg, D., Stenderup, J., and Piskur, J. (2009). Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. *Proc. Natl. Acad. Sci. U. S. A.* *106*, 2688–2693.

Sanglard, D., Coste, A., and Ferrari, S. (2009). Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. *FEMS Yeast Res.* *9*, 1029–1050.

Selmecki, A., Forche, A., and Berman, J. (2006). Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. *Science* *313*, 367–370.

Seret, M.-L., Diffels, J.F., Goffeau, A., and Baret, P.V. (2009). Combined phylogeny and neighborhood analysis of the evolution of the ABC transporters conferring multiple drug resistance in hemiascomycete yeasts. *BMC Genomics* *10*, 459.

Sheltzer, J.M., Blank, H.M., Pfau, S.J., Tange, Y., George, B.M., Humpton, T.J., Brito, I.L., Hiraoka, Y., Niwa, O., and Amon, A. (2011). Aneuploidy Drives Genomic Instability in Yeast. *Science* *333*, 1026–1030.

Sheltzer, J.M., Torres, E.M., Dunham, M.J., and Amon, A. (2012). Transcriptional consequences of aneuploidy. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 12644–12649.

Sionov, E., Chang, Y.C., Garraffo, H.M., and Kwon-Chung, K.J. (2009). Heteroresistance to Fluconazole in *Cryptococcus neoformans* Is Intrinsic and Associated with Virulence. *Antimicrob. Agents Chemother.* *53*, 2804–2815.

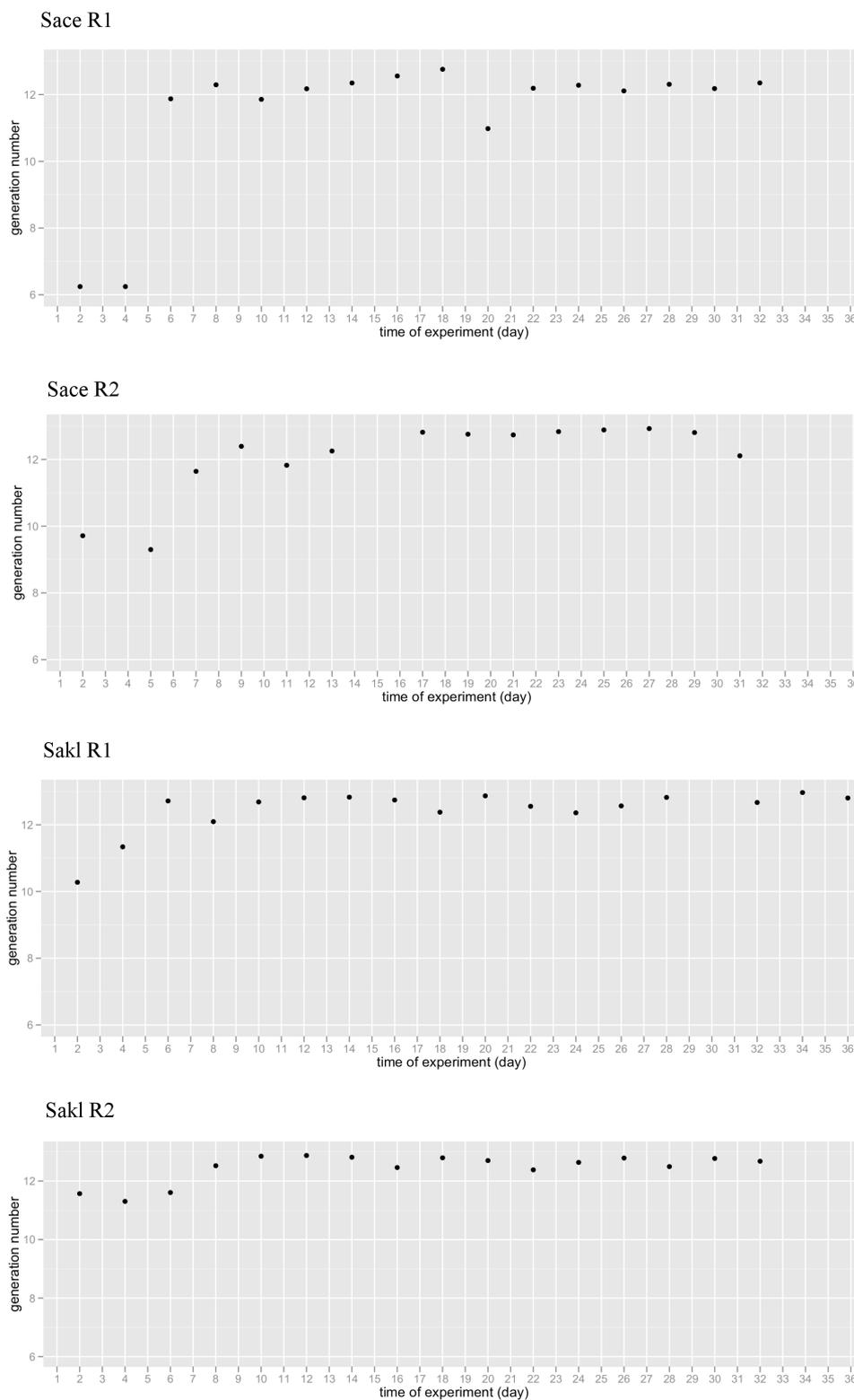
Tan, Z., Hays, M., Cromie, G.A., Jeffery, E.W., Scott, A.C., Ahyong, V., Sirt, A., Skupin, A., and Dudley, A.M. (2013). Aneuploidy underlies a multicellular phenotypic switch. *Proc. Natl. Acad. Sci. U. S. A.* *110*, 12367–12372.

Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* *317*, 916–924.

Warringer, J., Zörgö, E., Cubillos, F.A., Zia, A., Gjuvslund, A., Simpson, J.T., Forsmark, A., Durbin, R., Omholt, S.W., Louis, E.J., et al. (2011). Trait variation in yeast is defined by population history. *PLoS Genet.* *7*, e1002111.

Yona, A.H., Manor, Y.S., Herbst, R.H., Romano, G.H., Mitchell, A., Kupiec, M., Pilpel, Y., and Dahan, O. (2012). Chromosomal duplication is a transient evolutionary solution to stress. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 21010–21015.

## Supplemental data



**Figure S1. Number of generation of the global population per subcultures. Each graph represents one evolving experiment.**

```

# Length: 656
# Identity: 338/656 (51.5%)
# Similarity: 338/656 (51.5%)
# Gaps: 24/656 ( 3.7%)
# Score: 262
#
#
#=====
hAT_NRBC10955 16 GAAAATGAGTCTACAGGAAATTTGTTCGAGGCATATCATCAAAATCCATTC 65
KLTH0H03344t 31 GAAGGAGAGTCCACTGGGAACATGGCCAAGCATGTCCGTGCCTTGCATCC 80
hAT_NRBC10955 66 TGCACAGTACAAGTCTTACGAAACGGCAAAGCTCGAAGGAGTCAACACTGG 115
KLTH0H03344t 81 AGCGGCTTGAAGGCCAAAGAGGGGAAAGTTTGA AAAACGCCCCCGCGAA 130
hAT_NRBC10955 116 ATATGCTTGCTAGAAAAACGAGATCCCTACCGTTGTCCGATTCCTTGATA 165
KLTH0H03344t 131 ACGTCACTTTCAGAAGGCTTGAATCGCTGCGGGTGTCCAGTGGGTTGCA 180
hAT_NRBC10955 166 GCTGAGATCA--AGCATAATCGTGGGGCTTACCTAACTATCATGCTTGTT 213
KLTH0H03344t 181 GACGAGCATAGGAGGAAAATAGGAGAGCT--CGAAACACTAATGCTGGTT 228
hAT_NRBC10955 214 ACAGAGGCGTTACTGCCTTTCGCGCCGTGGAATACACGAGTTGG---AA 260
KLTH0H03344t 229 GTCGAGCATGCCTGCCCTCAGTCTTGTTCAGTCCGCGTCTGGGCCAA 278
hAT_NRBC10955 261 ACTGCTTAGCA-ATGCTATGCCTGGTTTTTCTTTGATCAGGTCCAGA AAC 309
KLTH0H03344t 279 GTGGAATGACAGATGCTCTCCGA---ATTCTAGTATCAAGTCAAGCACA 324
hAT_NRBC10955 310 ACGATTCTGGAAAAC TGAACAATACTCAGGCTATTTTCGGGAGG-CTC 358
KLTH0H03344t 325 AC-CTAGTAAAGAAGTTGCAC TTGTATCAGAAAACATGAACAGGTCAC 373

hAT_NRBC10955 359 TCTTTCAGCTCTTGAAAACTCT-ACTTTCCTGAACATCCA ACTAGATAT 407
KLTH0H03344t 374 TGCCTTGAAC TTGAGGAAC TCTCAATTTGCT-AATATCCAATTTGACAT 422
hAT_NRBC10955 408 ATGGTCTGGTGGCAACGGTTCGCTCATACTGGCTATTTCTGCCTCATTCC 457
KLTH0H03344t 423 ATGGATTTCCCGGCGGGGAGTCA TTTTGGGCGTCATGGTATCTTTTCG 472
hAT_NRBC10955 458 CGCCAAACATTTCTTGATGAACACAGAATGTCTGCTGTAGGGGACG--TTA 505
KLTH0H03344t 473 CACCAAACCTTTTGAATAAAA--AGCATTT-TGCAGCAGGCCAAAAC TCC 518
hAT_NRBC10955 506 GAA--TTCTTTTAAACAACGATGGCGAGCCTGTGAATTGCCATCTACTAG 553
KLTH0H03344t 519 GAAGTCCCTTTTGAACAAC TCCGGTTGTGCTCAAAGCACCCACCTGATTG 568
hAT_NRBC10955 554 ATTTTCGTAGACTTTAGCGACAAAAAGCACACAGGCGAAAAGATTTGTGTG 603
KLTH0H03344t 569 ATTTTCGTGACTTTGGGACTGAGAAGATTACGGACAAGCACTTATGTGAG 618
hAT_NRBC10955 604 AAAGTTATGAAGTTCTAAACGAGTTTGGGATTTCTGCACAAAGCCCGGAG 653
KLTH0H03344t 619 ACCCTTTTACTGGTTTTAGAAAAGTATGAGCTCAAAGACATGACTGCGAC 668

hAT_NRBC10955 654 TATCAC 659
KLTH0H03344t 669 GATAAC 674

```

**Figure S2. Pairwise DNA sequences alignment between the junction site and the transposon of *L. thermotolerans* *KLTH0H03344t*.** The alignment was performed with emboss matcher on the European Bioinformatics Institute website ([www.ebi.ac.uk/](http://www.ebi.ac.uk/)).



### 3. Discussion et perspectives

#### 3.1. L'expérience d'évolution

Les expériences d'évolution que nous avons mené ont permis l'adaptation des souches de *S. cerevisiae* et de *L. kluyveri* au fluconazole. Cette adaptation s'est faite par une augmentation de la vitesse spécifique de croissance. Nous émettons l'hypothèse que cette augmentation est due à un mécanisme d'adaptation qui est apparu chez un clone puis s'est fixé au sein de la population. En effet, quand nous avons mesuré la vitesse spécifique de croissance sur fluconazole de clones sélectionnés au hasard, tous les clones sont devenus résistants seulement 2 à 4 jours après l'apparition du 1er clone résistant.

#### 3.2. Le mécanisme d'adaptation

Le séquençage des souches évoluées sélectionnées à différents temps de l'expérience a mis en évidence un mécanisme d'adaptation commun entre la souche de *S. cerevisiae* et la souche de *L. kluyveri*. Il s'agit de la formation de chromosomes surnuméraires provenant de chromosomes dupliqués (*S. cerevisiae* et *L. kluyveri*) ou de chromosomes dupliqués recombinaisonnels qui forment un chromosome chimère (*L. kluyveri*).

Ces résultats sont en cohérence avec des études récentes qui ont montré l'impact des aneuploïdies sur le phénotype. Chez *S. cerevisiae*, il a été démontré que les aneuploïdies ont un impact sur la morphologie des colonies, la croissance à différentes températures et la résistance à différents stress (sel, antifongique, médicament anticancéreux) (Tan *et al.* 2013; Pavelka *et al.* 2010).

Cependant, maintenir un nombre anormal de chromosomes a un coût pour une cellule (Torres *et al.* 2007). Dans notre analyse, nous avons pu observer une diminution de la vitesse spécifique de croissance et de la biomasse en milieu riche. Une perspective envisageable serait de continuer l'expérience d'évolution plus longtemps. Il est probable que les chromosomes surnuméraires soient perdus au profit de variations génétiques plus localisées (mutation, duplication d'un gène) à l'image de ce qui avait été obtenu dans l'évolution d'une souche de *S. cerevisiae* à pH et température élevé(e) (Yona *et al.* 2012). Ainsi nous pourrions comparer les mécanismes compensatoires chez les deux espèces.

La formation des chromosomes chimères BG et GH formés chez *L. kluyveri* est probablement liée à la présence des éléments transposables au niveau des jonctions. Mais qu'en serait-il pour une souche qui ne possède pas d'éléments transposables dans ces régions du génome et qui serait soumise à la croissance sous fluconazole ? Est-ce que cette adaptation par la formation des chromosomes chimères est un mécanisme courant chez *L. kluyveri* ? Pour y répondre il serait intéressant de faire évoluer dans les mêmes conditions une autre souche de *L. kluyveri* ne possédant pas d'éléments transposables au niveau des sites de jonction.

### 3.3. Les origines de la résistance

Le mécanisme d'adaptation qui a permis aux souches d'évoluer a été mis en évidence, mais il faut cependant encore déterminer les origines génétiques de la résistance. Nous avons émis l'hypothèse que par un effet de dosage génique, certains gènes présents sur les chromosomes surnuméraires seraient responsables de la résistance. Nous avons établi une liste des gènes candidats grâce à une analyse fonctionnelle (tableau 3.1).

Espèce	Souche évoluée	Gène candidat	Fonction	Localisation
<i>L. kluyveri</i>	Sak1_R1d32	<i>PDR5</i>	Code un transporteur de type PDR.	chromosome G
	Sak1_R2d8 Sak1_R2d28	<i>ERG11</i>	Code l'enzyme lanosterol 14-alpha-demethylase qui intervient dans la biosynthèse de l'ergosétol.	chromosome G
<i>S. cerevisiae</i>	Sace_R2d11	<i>PDR15</i>	Code un transporteur de type PDR.	chromosome 4
	Sace_R2d31	<i>SNQ2</i>	Code un transporteur de type PDR.	chromosome 4
	Sace_R2d7 Sace_R2d11 Sace_R2d31	<i>PDR13</i>	Code une protéine Hsp70. Sa surexpression permet la celle de <i>PDR5</i> et <i>YOR1</i> via Pdr1.	chromosome 8
		<i>YHK8</i>	Code un présumé antiporteur de la famille DHA1 (transporteurs MDR). Il est surexprimé dans les cellules ayant une sensibilité réduite aux composés azolés.	chromosome 8

**Tableau 3.1 :** Liste des gènes candidats pour expliquer la résistance des souches évoluées au fluconazole. L'implication des gènes écrits en vert n'a pas encore pu être testée (perspectives).

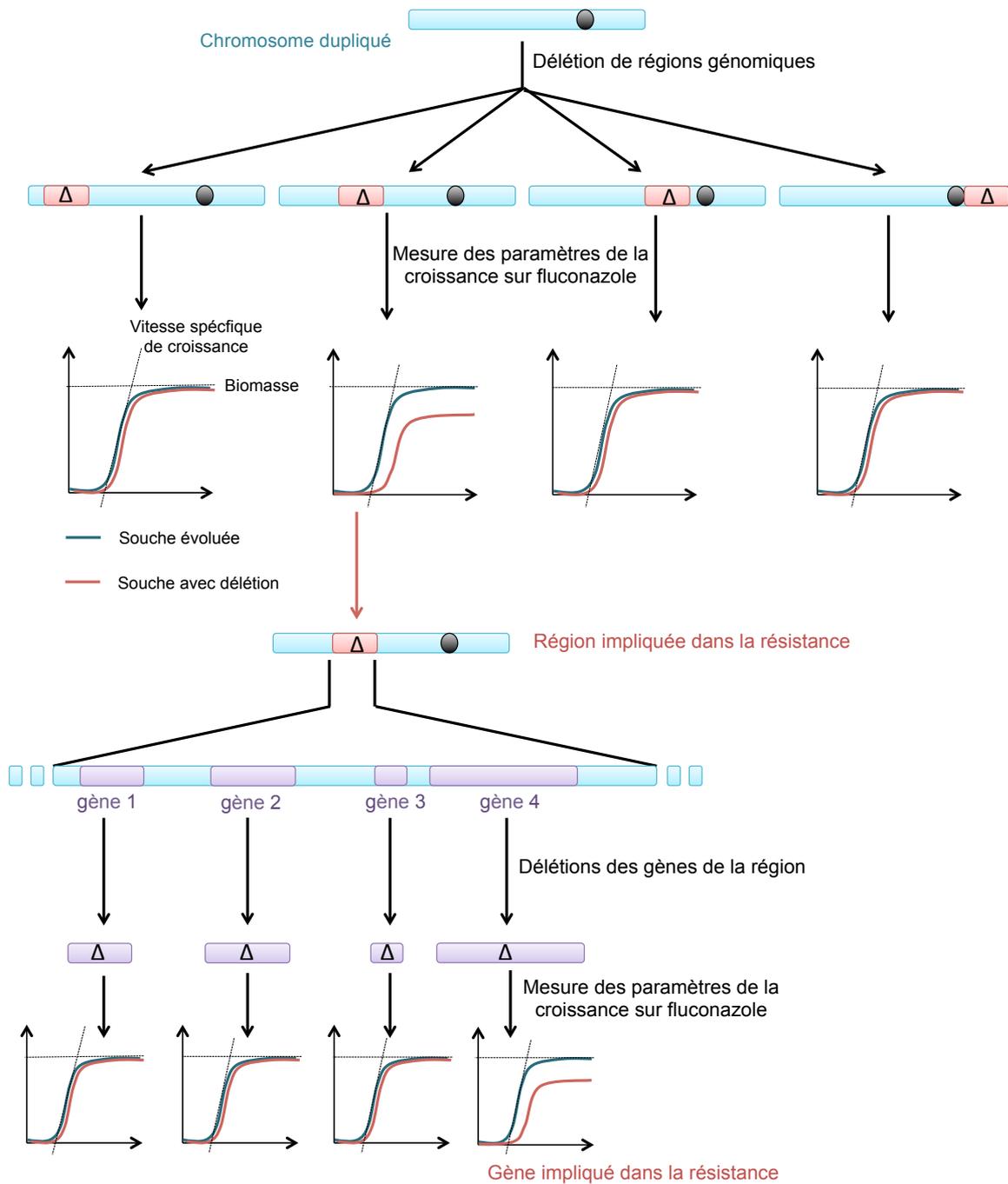
Chez *S. cerevisiae*, nous n'avons malheureusement pas encore pu effectuer les délétions de ces gènes. Il faudra ensuite tester la résistance de ces mutants au fluconazole. Ces expériences font parties des perspectives à court terme de ce projet.

Chez les souches évoluées de *L. kluyveri* nous nous sommes focalisés sur la partie présente 2x du chromosome G car elle a été impliquée dans les formations indépendantes des chromosomes chimères BG et GH. Nos gènes candidats basés sur cette analyse étaient *PDR5* et *ERG11* (tableau 3.1). Le gène *ERG11* explique une partie de la résistance. Une perspective à court terme serait de trouver d'autres gènes impliqués dans la résistance. Pour cela, il faudrait déterminer une liste des gènes candidats des autres régions dupliquées du génome (chromosomes B, C, une région du H), en se basant sur leur fonction. Leur implication dans la résistance sera vérifiée par délétion de leur phase codante.

Une autre stratégie possible pour déterminer les origines de la résistance serait de procéder à des délétions de régions, contenant plusieurs gènes, puis de tester la croissance de ces mutants sur fluconazole (figure 3.2). Le mutant de délétion pour lequel un paramètre de la croissance est modifié, possède la région impliquée dans la résistance. Pour analyser cette région, il sera nécessaire de déléter un à un les gènes qui y sont localisés. L'analyse de la résistance des souches délétées pour les gènes candidats, par rapport à la souche évoluée, permettra ainsi de déterminer les gènes à l'origine de la

résistance au fluconazole. Nous avons commencé à tester cette technique sur la région du chromosome G. L'inconvénient de cette méthode est qu'elle nécessite de réaliser un grand nombre de délétions.

En conclusion, les expériences d'évolution adaptatives sur fluconazole chez *S. cerevisiae* et *L. kluyveri* nous auront permis de déterminer un mécanisme d'adaptation commun chez les deux espèces. À court terme il faudra cependant préciser l'origine génétique de la résistance grâce aux expériences de délétions des gènes candidats.



**Figure 3.2 :** Stratégie des délétions par région des chromosomes surnuméraires. Cette technique a pour but de trouver les régions génomiques, puis les gènes impliqués dans la résistance au fluconazole.

## Références bibliographiques

- Aa, E., J. P. Townsend, R. I. Adams, K. M. Nielsen, and J. W. Taylor, 2006 Population structure and gene evolution in *Saccharomyces cerevisiae*. *Fems Yeast Research* 6: 702–715.
- Akache, B., and B. Turcotte, 2002 New regulators of drug sensitivity in the family of yeast zinc cluster proteins. *Journal of Biological Chemistry* 277: 21254–21260.
- Anderson, J. B., C. Sirjusingh, A. B. Parsons, C. Boone, C. Wickens *et al.*, 2003 Mode of selection and experimental evolution of antifungal drug resistance in *Saccharomyces cerevisiae*. *Genetics* 163: 1287–1298.
- Argueso, J. L., M. F. Carazzolle, P. A. Mieczkowski, F. M. Duarte, O. V. C. Netto *et al.*, 2009 Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome research* 19: 2258–2270.
- Bleykasten-Grosshans, C., and C. Neuveglise, 2011 Transposable elements in yeasts. *Comptes Rendus Biologies* 334: 679–686.
- Borneman, A. R., A. H. Forgan, I. S. Pretorius, and P. J. Chambers, 2008 Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *Fems Yeast Research* 8: 1185–1195.
- Borneman, A. R., B. A. Desany, D. Riches, J. P. Affourtit, A. H. Forgan *et al.*, 2011 Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS genetics* 7: e1001287.
- Botstein, D., and G. R. Fink, 2011 Yeast: An Experimental Organism for 21st Century Biology. *Genetics* 189: 695–704.
- Brauer, M. J., C. M. Christianson, D. A. Pai, and M. J. Dunham, 2006 Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae*. *Genetics* 173: 1813–1816.
- Byrne, K. P., and K. H. Wolfe, 2005 The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome research* 15: 1456–1461.
- Carle, G. F., and M. V. Olson, 1985 An electrophoretic karyotype for yeast. *Proceedings of the National Academy of Sciences of the United States of America* 82: 3756–3760.
- Carreto, L., M. F. Eiriz, A. C. Gomes, P. M. Pereira, D. Schuller *et al.*, 2008 Comparative genomics of wild type yeast strains unveils important genome diversity. *BMC genomics* 9: 524.
- Carvajal, E., H. B. vandenHazel, A. CybularzKolaczowska, E. Balzi, and A. Goffeau, 1997 Molecular and phenotypic characterization of yeast *PDR1* mutants that show hyperactive transcription of various ABC multidrug transporter genes. *Molecular & General Genetics* 256: 406–415.
- Cavaliere, D., P. E. McGovern, D. L. Hartl, R. Mortimer, and M. Polsinelli, 2003 Evidence for *S. cerevisiae* fermentation in ancient wine. *Journal of Molecular Evolution* 57: S226–S232.
- Chandra, J., D. M. Kuhn, P. K. Mukherjee, L. L. Hoyer, T. McCormick *et al.*, 2001 Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *Journal of Bacteriology* 183: 5385–5394.

- Chen, L. M., Y. H. Xu, C. L. Zhou, J. Zhao, C. Y. Li *et al.*, 2010 Overexpression of *CDR1* and *CDR2* genes plays an important role in fluconazole resistance in *Candida albicans* with G487T and T916C mutations. *Journal of International Medical Research* 38: 536–545.
- Chu, G., 1989 Pulsed field electrophoresis in contour-clamped homogeneous electric fields for the resolution of DNA by size or topology. *Electrophoresis* 10: 290–295.
- Clamp, M., J. Cuff, S. M. Searle, and G. J. Barton, 2004 The Jalview Java alignment editor. *Bioinformatics (Oxford, England)* 20: 426–427.
- Clemons, K. V., J. H. McCusker, R. W. Davis, and D. A. Stevens, 1994 Comparative pathogenesis of clinical and nonclinical isolates of *Saccharomyces cerevisiae*. *The Journal of infectious diseases* 169: 859–867.
- Cui, Z., T. Shiraki, D. Hirata, and T. Miyakawa, 1998 Yeast gene *YRR1*, which is required for resistance to 4-nitroquinoline N-oxide, mediates transcriptional activation of the multidrug resistance transporter gene *SNQ2*. *Molecular microbiology* 29: 1307–1315.
- Dawes, I. W., and I. D. Hardie, 1974 Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether. *Molecular & general genetics: MGG* 131: 281–289.
- Decottignies, A., and A. Goffeau, 1997 Complete inventory of the yeast ABC proteins. *Nature genetics* 15: 137–145.
- Deutschbauer, A. M., and R. W. Davis, 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature genetics* 37: 1333–1340.
- Diao, L., and K. C. Chen, 2012 Local ancestry corrects for population structure in *Saccharomyces cerevisiae* genome-wide association studies. *Genetics* 192: 1503–1512.
- Diezmann, S., and F. S. Dietrich, 2011 Oxidative stress survival in a clinical *Saccharomyces cerevisiae* isolate is influenced by a major quantitative trait nucleotide. *Genetics* 188: 709–U308.
- Doniger, S. W., H. S. Kim, D. Swain, D. Corcuera, M. Williams *et al.*, 2008 A Catalog of neutral and deleterious polymorphism in yeast. *PLoS Genetics* 4: e1000183.
- Donlan, R. M., 2002 Biofilms: microbial life on surfaces. *Emerging infectious diseases* 8: 881–890.
- Dujon, B., 2010 Yeast evolutionary genomics. *Nature Reviews Genetics* 11: 512–524.
- Dujon, B., 2006 Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends in genetics: TIG* 22: 375–387.
- Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola *et al.*, 2004 Genome evolution in yeasts. *Nature* 430: 35–44.
- Dunham, M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown *et al.*, 2002 Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* 99: 16144–16149.
- Enache-Angoulvant, A., and C. Hennequin, 2005 Invasive *Saccharomyces* infection: a comprehensive review. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 41: 1559–1568.

- Ernst, R., P. Kueppers, C. M. Klein, T. Schwarzmueller, K. Kuchler *et al.*, 2008 A mutation of the H-loop selectively affects rhodamine transport by the yeast multidrug ABC transporter Pdr5. *Proceedings of the National Academy of Sciences of the United States of America* 105: 5069–5074.
- Fay, J. C., and J. A. Benavides, 2005 Evidence for domesticated and wild populations of *Saccharomyces cerevisiae*. *PLoS Genet* 1: e5.
- Fisher, R. A., 1918 The correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*. 52: 399-433.
- Foss, E. J., D. Radulovic, S. A. Shaffer, D. M. Ruderfer, A. Bedalov *et al.*, 2007 Genetic basis of proteome variation in yeast. *Nature genetics* 39: 1369–1375.
- Freese, E. B., M. I. Chu, and E. Freese, 1982 Initiation of yeast sporulation of partial carbon, nitrogen, or phosphate deprivation. *Journal of bacteriology* 149: 840–851.
- Friedrich A, Jung PP, Reisser C, Schacherer J. 2014 Population genomics of a preduplication yeast species. In preparation.
- Fujimura, H., 1991 Transformation of the yeast *Saccharomyces kluyveri* by *Saccharomyces cerevisiae*-based plasmids. *FEMS microbiology letters* 66: 149–152.
- Gallis, H., R. Drew, and W. Pickard, 1990 Amphotericin-B - 30 Years of clinical-experience. *Reviews of Infectious Diseases* 12: 308–329.
- Gerke, J., K. Lorenz, and B. Cohen, 2009 Genetic interactions between transcription factors cause natural variation in yeast. *Science (New York, N.Y.)* 323: 498–501.
- Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002 Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418: 387–391.
- Gibson, B. R., S. J. Lawrence, J. P. R. Leclaire, C. D. Powell, and K. A. Smart, 2007 Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiology Reviews* 31: 535–569.
- Goffeau, A., B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon *et al.*, 1996 Life with 6000 genes. *Science* 274: 546–567.
- Gojkovic, Z., K. Jahnke, K. D. Schnackerz, and J. Piskur, 2000 PYD2 encodes 5,6-dihydropyrimidine amidohydrolase, which participates in a novel fungal catabolic pathway. *Journal of molecular biology* 295: 1073–1087.
- Gordon, J. L., K. P. Byrne, and K. H. Wolfe, 2011 Mechanisms of Chromosome Number Evolution in Yeast. *PLoS Genetics* 7.
- Gouy, M., S. Guindon, and O. Gascuel, 2010 SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular biology and evolution* 27: 221–224.
- Heilmann, C. J., S. Schneider, K. S. Barker, P. D. Rogers, and J. Morschhauser, 2010 An A643T mutation in the transcription factor Upc2p causes constitutive *ERG11* upregulation and increased fluconazole resistance in *Candida albicans*. *Antimicrobial Agents and Chemotherapy* 54: 353–359.

- Helm, E., B. Nohr, and R. Thorne, 1953 The measurement of yeast flocculence and its significance in brewing. *Wallerstein Lab Commun* 315–325.
- Henry, K. W., J. T. Nickels, and T. D. Edlind, 2000 Upregulation of *ERG* genes in *Candida* species by azoles and other sterol biosynthesis inhibitors. *Antimicrobial agents and chemotherapy* 44: 2693–2700.
- Herskowitz, I., 1988 Life cycle of the budding yeast *Saccharomyces cerevisiae*. *Microbiological reviews* 52: 536–553.
- Hess, D. C., W. Lu, J. D. Rabinowitz, and D. Botstein, 2006 Ammonium toxicity and potassium limitation in yeast. *PLoS Biol* 4: e351.
- Hillenmeyer, M. E., E. Fung, J. Wildenhain, S. E. Pierce, S. Hoon *et al.*, 2008 The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science (New York, N.Y.)* 320: 362–365.
- Hoffman, C., and F. Winston, 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57: 267–272.
- Hu, X. H., M. H. Wang, T. Tan, J. R. Li, H. Yang *et al.*, 2007 Genetic dissection of ethanol tolerance in the budding yeast *Saccharomyces cerevisiae*. *Genetics* 175: 1479–1487.
- Johnson, L. J., V. Koufopanou, M. R. Goddard, R. Hetherington, S. M. Schäfer *et al.*, 2004 Population genetics of the wild yeast *Saccharomyces paradoxus*. *Genetics* 166: 43–52.
- Jung, P. P., A. Friedrich, C. Reisser, J. Hou, and J. Schacherer, 2012 Mitochondrial genome evolution in a single protoploid yeast species. *G3: Genes|Genomes|Genetics* 2: 1103–1111.
- Jung, P. P., A. Friedrich, A. Sigwalt and J. Schacherer, 2014 Phenotypic variation within *Saccharomyces kluyveri* yeast species. In preparation.
- Katou, T., M. Namise, H. Kitagaki, T. Akao, and H. Shimoi, 2009 QTL mapping of sake brewing characteristics of yeast. *Journal of bioscience and bioengineering* 107: 383–393.
- Katzmann, D. J., T. C. Hallstrom, Y. Mahe, and W. S. MoyeRowley, 1996 Multiple Pdr1p/Pdr3p binding sites are essential for normal expression of the ATP binding cassette transporter protein-encoding gene *PDR5*. *Journal of Biological Chemistry* 271: 23049–23054.
- Keeling, P. J., G. Burger, D. G. Durnford, B. F. Lang, R. W. Lee *et al.*, 2005 The tree of eukaryotes. *Trends in Ecology & Evolution* 20: 670–676.
- Kellis, M., B. W. Birren, and E. S. Lander, 2004 Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* 428: 617–624.
- Koltin, Y., and C. A. Hitchcock, 1997 The search for new triazole antifungal agents. *Current opinion in chemical biology* 1: 176–182.
- Kurtzman, C. P., 2003 Phylogenetic circumscription of *Saccharomyces*, *Kluyveromyces* and other members of the *Saccharomycetaceae*, and the proposal of the new genera *Lachancea*, *Nakaseomyces*, *Naumovia*, *Vanderwaltozyma* and *Zygorulasporea*. *FEMS Yeast Research* 4: 233–245.

- Kvitek, D. J., J. L. Will, and A. P. Gasch, 2008 Variations in stress sensitivity and genomic expression in diverse *S. cerevisiae* isolates. *PLoS genetics* 4: e1000223.
- Lamping, E., P. V. Baret, A. R. Holmes, B. C. Monk, A. Goffeau *et al.*, 2010 Fungal PDR transporters: Phylogeny, topology, motifs and function. *Fungal genetics and biology: FG & B* 47: 127–142.
- Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)* 25: 1754–1760.
- Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map format and SAMtools. *Bioinformatics (Oxford, England)* 25: 2078–2079.
- Liti, G., D. M. Carter, A. M. Moses, J. Warringer, L. Parts *et al.*, 2009 Population genomics of domestic and wild yeasts. *Nature* 458: 337–341.
- De Llanos, R., A. Querol, J. Pemán, M. Gobernado, and M. T. Fernández-Espinar, 2006 Food and probiotic strains from the *Saccharomyces cerevisiae* species as a possible origin of human systemic infections. *International Journal of Food Microbiology* 110: 286–290.
- Mackay, T. F. C., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. *Nature reviews. Genetics* 10: 565–577.
- MacPherson, S., M. Larochelle, and B. Turcotte, 2006 A fungal family of transcriptional regulators: the zinc cluster proteins. *Microbiology and molecular biology reviews: MMBR* 70: 583–604.
- Maertens, J. A., 2004 History of the development of azole derivatives. *Clinical Microbiology and Infection* 10: 1–10.
- Mah, T. F., and G. A. O’Toole, 2001 Mechanisms of biofilm resistance to antimicrobial agents. *Trends in microbiology* 9: 34–39.
- Mancera, E., R. Bourgon, A. Brozzi, W. Huber, and L. M. Steinmetz, 2008 High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Nature* 454: 479–485.
- Mardis, E. R., 2008 The impact of next-generation sequencing technology on genetics. *Trends in genetics: TIG* 24: 133–141.
- Margueron, R., P. Trojer, and D. Reinberg, 2005 The key to development: interpreting the histone code? *Current opinion in genetics & development* 15: 163–176.
- Marichal, P., L. Koymans, S. Willemsens, D. Bellens, P. Verhasselt *et al.*, 1999 Contribution of mutations in the cytochrome P450 14alpha-demethylase (Erg11p, Cyp51p) to azole resistance in *Candida albicans*. *Microbiology (Reading, England)* 145 ( Pt 10): 2701–2713.
- Masiá Canuto, M., and F. Gutiérrez Rodero, 2002 Antifungal drug resistance to azoles and polyenes. *The Lancet infectious diseases* 2: 550–563.
- McCullough, J., and I. Herskowitz, 1979 Mating Pheromones of *Saccharomyces kluyveri*: Pheromone Interactions Between *Saccharomyces kluyveri* and *Saccharomyces cerevisiae*. *Journal of Bacteriology* 138: 146–154.
- McCusker, J. H., K. V. Clemons, D. A. Stevens, and R. W. Davis, 1994 *Saccharomyces cerevisiae* virulence phenotype as determined with CD-1 mice is associated with the ability to grow at 42 degrees C and form pseudohyphae. *Infection and immunity* 62: 5447–5455.

- McGovern, P. E., J. Zhang, J. Tang, Z. Zhang, G. R. Hall *et al.*, 2004 Fermented beverages of pre- and proto-historic China. *Proceedings of the National Academy of Sciences of the United States of America* 101: 17593–17598.
- Minamoto, G. Y., and A. S. Rosenberg, 1997 Fungal infections in patients with acquired immunodeficiency syndrome. *Medical Clinics of North America* 81: 381–409.
- Monk, B. C., and A. Goffeau, 2008 Outwitting multidrug resistance to antifungals. *Science (New York, N.Y.)* 321: 367–369.
- Moroni, A., L. Bardella, and G. Thiel, 1998 The impermeant ion methylammonium blocks K<sup>+</sup> and NH<sub>4</sub><sup>+</sup> currents through KAT1 channel differently: Evidence for ion interaction in channel permeation. *Journal of Membrane Biology* 163: 25–35.
- Mortimer, R. K., and J. R. Johnston, 1986 Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113: 35–43.
- Nagarajan, M., J.-B. Veyrieras, M. de Dieuleveult, H. Bottin, S. Fehrmann *et al.*, 2010 Natural single-nucleosome epi-polymorphisms in yeast. *PLoS genetics* 6: e1000913.
- Nawrocki, A., S. J. Fey, A. Goffeau, P. Roepstorff, and P. M. Larsen, 2001 The effects of transcription regulating genes *PDR1*, *pdr1–3* and *PDR3* in pleiotropic drug resistance. *Proteomics* 1: 1022–1032.
- Neuvéglise, C., E. Bon, A. Léplinge, P. Wincker, F. Artiguenave *et al.*, 2000 Genomic exploration of the hemiascomycetous yeasts: 9. *Saccharomyces kluyveri*. *FEBS letters* 487: 56–60.
- Nourani, A., D. Papajova, A. Delahodde, C. Jacq, and J. Subik, 1997 Clustered amino acid substitutions in the yeast transcription regulator Pdr3p increase pleiotropic drug resistance and identify a new central regulatory domain. *Molecular & general genetics: MGG* 256: 397–405.
- Novo, M., F. Bigey, E. Beyne, V. Galeote, F. Gavory *et al.*, 2009 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: 16333–16338.
- Ohya, Y., J. Sese, M. Yukawa, F. Sano, Y. Nakatani *et al.*, 2005 High-dimensional and large-scale phenotyping of yeast mutants. *Proceedings of the National Academy of Sciences of the United States of America* 102: 19015–19020.
- Parts, L., F. A. Cubillos, J. Warringer, K. Jain, F. Salinas *et al.*, 2011 Revealing the genetic structure of a trait by sequencing a population under selection. *Genome research* 21: 1131–1138.
- Parveen, M., M. K. Hasan, J. Takahashi, Y. Murata, E. Kitagawa *et al.*, 2004 Response of *Saccharomyces cerevisiae* to a monoterpene: evaluation of antifungal potential by DNA microarray analysis. *Journal of Antimicrobial Chemotherapy* 54: 46–55.
- Pavelka, N., G. Rancati, J. Zhu, W. D. Bradford, A. Saraf *et al.*, 2010 Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468: 321–325.
- Paya, C., 1993 Fungal-Infections in solid-organ transplantation. *Clinical Infectious Diseases* 16: 677–688.
- Payen, C., G. Fischer, C. Marck, C. Proux, D. J. Sherman *et al.*, 2009 Unusual composition of a yeast chromosome arm is associated with its delayed replication. *Genome research* 19: 1710–1721.

- Perea, S., J. L. López-Ribot, W. R. Kirkpatrick, R. K. McAtee, R. A. Santillán *et al.*, 2001 Prevalence of molecular mechanisms of resistance to azole antifungal agents in *Candida albicans* strains displaying high-level fluconazole resistance isolated from human immunodeficiency virus-infected patients. *Antimicrobial agents and chemotherapy* 45: 2676–2684.
- Perlstein, E. O., D. M. Ruderfer, D. C. Roberts, S. L. Schreiber, and L. Kruglyak, 2007 Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nature genetics* 39: 496–502.
- Pfaffl, M. W., G. W. Horgan, and L. Dempfle, 2002 Relative expression software tool (REST) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. *Nucleic acids research* 30: e36.
- Pinjon, E., G. P. Moran, C. J. Jackson, S. L. Kelly, D. Sanglard *et al.*, 2003 Molecular mechanisms of itraconazole resistance in *Candida dubliniensis*. *Antimicrobial agents and chemotherapy* 47: 2424–2437.
- Polak, A., and H. Scholer, 1975 Mode of Action of 5-Fluorocytosine and Mechanisms of Resistance. *Chemotherapy* 21: 113–130.
- Prasad, R., and A. Goffeau, 2012 Yeast ATP-Binding Cassette Transporters Conferring Multidrug Resistance. *Annual Review of Microbiology*, edited by S. Gottesman, C. S. Harwood, and O. Schneewind. Annual Reviews, Palo Alto. Vol 66: 39–63.
- Ralser, M., H. Kuhl, M. Ralser, M. Werber, H. Lehrach *et al.*, 2012 The *Saccharomyces cerevisiae* W303-K6001 cross-platform genome sequence: insights into ancestry and physiology of a laboratory mutt. *Open biology* 2: 120093.
- Ramage, G., R. Rajendran, L. Sherry, and C. Williams, 2012 Fungal biofilm resistance. *International journal of microbiology* 2012.
- Ramage, G., K. VandeWalle, S. P. Bachmann, B. L. Wickes, and J. L. Lopez-Ribot, 2002 In vitro pharmacodynamic properties of three antifungal agents against preformed *Candida albicans* biofilms determined by time-kill studies. *Antimicrobial Agents and Chemotherapy* 46: 3634–3636.
- Rodríguez-Navarro, A., 2000 Potassium transport in fungi and plants. *Biochimica et biophysica acta* 1469: 1–30.
- Sanglard, D., A. Coste, and S. Ferrari, 2009 Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. *FEMS yeast research* 9: 1029–1050.
- Sanglard, D., F. Ischer, L. Koymans, and J. Bille, 1998 Amino acid substitutions in the cytochrome P-450 lanosterol 14 $\alpha$ -demethylase (CYP51A1) from azole-resistant *Candida albicans* clinical isolates contribute to resistance to azole antifungal agents. *Antimicrobial agents and chemotherapy* 42: 241–253.
- Schacherer, J., D. M. Ruderfer, D. Gresham, K. Dolinski, D. Botstein *et al.*, 2007 Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS one* 2: e322.
- Schacherer, J., J. A. Shapiro, D. M. Ruderfer, and L. Kruglyak, 2009 Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* 458: 342–345.

- Schacherer, J., Y. Tourrette, S. Potier, J.-L. Souciet, and J. de Montigny, 2007 Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. *DNA repair* 6: 1441–1452.
- Segrè, A. V., A. W. Murray, and J.-Y. Leu, 2006 High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS biology* 4: e256.
- Selmecki, A. M., K. Dulmage, L. E. Cowen, J. B. Anderson, and J. Berman, 2009 Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *Plos Genetics* 5: e1000705.
- Seret, M.-L., J. F. Diffels, A. Goffeau, and P. V. Baret, 2009 Combined phylogeny and neighborhood analysis of the evolution of the ABC transporters conferring multiple drug resistance in hemiascomycete yeasts. *BMC genomics* 10: 459.
- Sharma, S., and G. K. Khuller, 1996 Changes in the cellular composition of *Candida albicans* resistant to miconazole. *Indian journal of biochemistry & biophysics* 33: 420–424.
- Sheehan, D. J., C. A. Hitchcock, and C. M. Sibley, 1999 Current and emerging azole antifungal agents. *Clinical microbiology reviews* 12: 40–79.
- Skaggs, B. A., J. F. Alexander, C. A. Pierson, K. S. Schweitzer, K. T. Chun *et al.*, 1996 Cloning and characterization of the *Saccharomyces cerevisiae* C-22 sterol desaturase gene, encoding a second cytochrome P-450 involved in ergosterol biosynthesis. *Gene* 169: 105–109.
- Sniegowski, P. D., P. G. Dombrowski, and E. Fingerhahn, 2002 *Saccharomyces cerevisiae* and *Saccharomyces paradoxus* coexist in a natural woodland site in North America and display different levels of reproductive isolation from European conspecifics. *FEMS yeast research* 1: 299–306.
- Souciet, J.-L., B. Dujon, C. Gaillardin, M. Johnston, P. V. Baret *et al.*, 2009 Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Research* 19: 1696–1709.
- Steinmetz, L. M., H. Sinha, D. R. Richards, J. I. Spiegelman, P. J. Oefner *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416: 326–330.
- Steyer, D., C. Ambroset, C. Brion, P. Claudel, P. Delobel *et al.*, 2012 QTL mapping of the production of wine aroma compounds by yeast. *BMC genomics* 13: 573.
- Tan, Z., M. Hays, G. A. Cromie, E. W. Jeffery, A. C. Scott *et al.*, 2013 Aneuploidy underlies a multicellular phenotypic switch. *Proceedings of the National Academy of Sciences of the United States of America* 110: 12367–12372.
- Torres, E. M., T. Sokolsky, C. M. Tucker, L. Y. Chan, M. Boselli *et al.*, 2007 Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science (New York, N.Y.)* 317: 916–924.
- Tourrette, Y., J. Schacherer, E. Fritsch, S. Potier, J.-L. Souciet *et al.*, 2007 Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence of ploidy. *Molecular microbiology* 64: 382–395.
- Tutulan-Cunita, A. C., M. Mikoshi, M. Mizunuma, D. Hirata, and T. Miyakawa, 2005 Mutational analysis of the yeast multidrug resistance ABC transporter Pdr5p with altered drug specificity. *Genes to cells: devoted to molecular & cellular mechanisms* 10: 409–420.
- Van Overbeek, M., M. Sasaki, S. Tischfield, J.-A. Sampson, N. D. Socci *et al.*, 2012 Revised genome assembly for the *S. cerevisiae* SK1 strain.

- Verstrepen, K. J., G. Derdelinckx, H. Verachtert, and F. R. Delvaux, 2003 Yeast flocculation: what brewers should know. *Applied Microbiology and Biotechnology* 61: 197–205.
- Wang, Q.-M., W.-Q. Liu, G. Liti, S.-A. Wang, and F.-Y. Bai, 2012 Surprisingly diverged populations of *Saccharomyces cerevisiae* in natural environments remote from human activity. *Molecular Ecology* 21: 5404–5417.
- Warringer, J., E. Zörgö, F. A. Cubillos, A. Zia, A. Gjuvsland *et al.*, 2011 Trait variation in yeast is defined by population history. *PLoS genetics* 7: e1002111.
- Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton, 2009 Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25: 1189–1191.
- Watson, P. F., M. E. Rose, S. W. Ellis, H. England, and S. L. Kelly, 1989 Defective sterol C5-6 desaturation and azole resistance: A new hypothesis for the mode of action of azole antifungals. *Biochemical and Biophysical Research Communications* 164: 1170–1175.
- Wei, W., J. H. McCusker, R. W. Hyman, T. Jones, Y. Ning *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proceedings of the National Academy of Sciences of the United States of America* 104: 12825–12830.
- Wenger, J. W., K. Schwartz, and G. Sherlock, 2010 Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genetics* 6: 1–17.
- Winzler, E. A., D. R. Richards, A. R. Conway, A. L. Goldstein, S. Kalman *et al.*, 1998 Direct allelic variation scanning of the yeast genome. *Science (New York, N.Y.)* 281: 1194–1197.
- Von Wiren, N., and M. Merrick, 2004 Regulation and function of ammonium carriers in bacteria, fungi, and plants. *Molecular Mechanisms Controlling Transmembrane Transport* 9: 95–120.
- Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387: 708–713.
- Xiang, M.-J., J.-Y. Liu, P.-H. Ni, S. Wang, C. Shi *et al.*, 2013 Erg11 mutations associated with azole resistance in clinical isolates of *Candida albicans*. *Fems Yeast Research* 13: 386–393.
- Ying, Y., Y. Zhao, X. Hu, Z. Cai, X. Liu *et al.*, 2013 In vitro fluconazole susceptibility of 1,903 clinical isolates of *Candida albicans* and the identification of *ERG11* mutations. *Microbial Drug Resistance* 19: 266–273.
- Yona, A. H., Y. S. Manor, R. H. Herbst, G. H. Romano, A. Mitchell *et al.*, 2012 Chromosomal duplication is a transient evolutionary solution to stress. *Proceedings of the National Academy of Sciences of the United States of America* 109: 21010–21015.
- Yvert, G., S. Ohnuki, S. Nogami, Y. Imanaga, S. Fehrmann *et al.*, 2013 Single-cell phenomics reveals intra-species variation of phenotypic noise in yeast. *BMC systems biology* 7: 54.
- Zheng, D.-Q., P.-M. Wang, J. Chen, K. Zhang, T.-Z. Liu *et al.*, 2012 Genome sequencing and genetic breeding of a bioethanol *Saccharomyces cerevisiae* strain YJS329. *BMC genomics* 13: 479.



## Annexes

Projet	Région génomique / gène amplifié(e)	Nom de l'amorce	Séquence de l'amorce (orientation de 5' en 3')
Chapitre 2 : séquençage de la région promotrice de <i>PDR5</i>	Région promotrice de <i>PDR5</i>	pPDR5_1F	gcaaaaataaggccataaaa
		pPDR5_1R	aaggcgagcaatgcgaac
		pPDR5_2F	ttcacacaatcaggagtggaa
		pPDR5_2R	agtagaagaagacgaggagga
Chapitre 2 : séquençage de <i>PDR5</i>	chrXV : 619801 pb	PDR5_1F	aagtttctgtatccgctcgt
	chrXV : 620632 pb	PDR5_1R	caacctcgctactgtaaccaa
	chrXV : 620532 pb	PDR5_2F	ttaagaaacatttctgtggtg
	chrXV : 621356 pb	PDR5_2R	taactgcatcatgtagctga
	chrXV : 621114 pb	PDR5_3F	ctgagagaacctgaacaaa
	chrXV : 621919 pb	PDR5_3R	ccaacgtaggatcttcttctt
	chrXV : 621831 pb	PDR5_4F	cagaagctatggtcccgtt
	chrXV : 622620 pb	PDR5_4R	ataacacctatggtaacctt
	chrXV : 622519 pb	PDR5_5F	ttgggttaaacaggctactt
	chrXV : 623350 pb	PDR5_5R	cgaaccttcttttggttaattc
	chrXV : 623266 pb	PDR5_6F	cgaagtttgagggaattctga
	chrXV : 624090 pb	PDR5_6R	agcaattcgtaatcagcgca
	chrXV : 623510 pb	PDR5_7F	ttcttcaaagcaggaaacctcg
	chrXV : 624682 pb	PDR5_7R	tgcgtcctttcttttaaccg

**Annexe MM1** : Liste des amorces utilisées dans le chapitre 2.

67-588	NRBC181	2	NRBC1057	NRBC1095	NRBC1019	5	99	CBS10369	CBS8828	dd281a	CBS2861	CBS4568	2	8	DRVPC345	DRVPC310	55-86_1	62-1041	CBS3082a	77-1003	NCYC543	62-196	CBS6545	CBS6546	CBS6547	CBS6626	NRBC1892	CBS10367	CBS10368	CBS4104	68917-2	2	DRVPC400					
0.45/0.76/1.																																						
NRBC181	70																																					
NRBC1057	0.67/1.26/1.	0.76/1.39/1.																																				
2	88	82																																				
NRBC1095	0.69/1.13/1.	0.80/1.22/1.	0.95/1.57/1.																																			
5	63	54	66																																			
NRBC1019	0.82/1.12/1.	0.81/1.18/1.	0.98/1.60/1.	0.94/1.46/1.																																		
36	64	66	36	0.06/0.24/3.																																		
99	57	51	68/1.16/1.	0.98/1.60/1.	0.52/0.95/1.	0.52/0.94/1.																																
CBS10369	0.82/1.23/1.	0.90/1.31/1.	1.06/1.68/1.	1.03/1.56/1.	0.52/0.95/1.	0.52/0.94/1.																																
64	43	64	56	74																																		
CBS5828	50	46	59	52	83	82																																
dd281a	0.83/1.24/1.	0.91/1.33/1.	1.07/1.72/1.	1.04/1.61/1.	0.54/0.95/1.	0.54/0.93/1.	0.25/0.70/2.																															
73	61	55	76	73	78																																	
CBS2861	0.83/1.22/1.	0.90/1.28/1.	1.06/1.67/1.	1.03/1.54/1.	0.52/0.91/1.	0.52/0.91/1.	0.25/0.65/2.	0.26/0.71/2.																														
74	58	50	75	76	63	74																																
CBS4568	0.82/1.18/1.	0.90/1.27/1.	1.05/1.65/1.	1.03/1.54/1.	0.52/0.88/1.	0.52/0.87/1.	0.24/0.64/2.	0.25/0.63/2.	0.08/0.39/4.																													
92	42	58	49	71	69	52																																
DRVPC345	0.83/1.24/1.	0.90/1.28/1.	1.05/1.66/1.	1.03/1.53/1.	0.52/0.90/1.	0.52/0.90/1.	0.24/0.64/2.	0.26/0.72/2.	0.06/0.15/2.	0.08/0.38/4.																												
2	47	42	57	49	73	72																																
DRVPC310	0.82/1.19/1.	0.90/1.30/1.	1.06/1.68/1.	1.03/1.56/1.	0.52/0.90/1.	0.52/0.90/1.	0.24/0.66/2.	0.24/0.61/2.	0.08/0.40/5.																													
8	57	53	73	50	75	86																																
55-86_1	0.2	0.3	0.6	0.5	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1			
62-1041	0.2	0.3	0.8	0.3	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1			
CBS3082a	2.56/2.60/1.	2.54/2.61/1.	2.72/2.92/1.	2.73/2.79/1.	2.56/2.63/1.	2.56/2.63/1.	2.56/2.61/1.	2.56/2.57/1.	2.58/2.68/1.	2.57/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.	2.58/2.65/1.		
0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.	0.01/0.02/1.				
77-1003	2.57/2.62/1.	2.55/2.63/1.	2.73/2.94/1.	2.73/2.80/1.	2.57/2.62/1.	2.57/2.62/1.	2.57/2.58/1.	2.57/2.58/1.	2.58/2.66/1.	2.57/2.58/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.	2.58/2.66/1.			
0.01/0.06/8.	0.01/0.06/8.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.	0.01/0.05/3.				
NCYC543	2.15/2.59/1.	2.15/2.61/1.	2.35/3.00/1.	2.33/2.85/1.	2.17/2.65/1.	2.17/2.65/1.	2.17/2.65/1.	2.17/2.65/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.	2.18/2.66/1.				
15	21	28	22	22	22	22																																
62-196	1.98/2.43/1.	1.98/2.47/1.	2.19/2.84/1.	2.16/2.69/1.	2.00/2.51/1.	2.00/2.48/1.	2.00/2.49/1.	2.00/2.49/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.	2.01/2.51/1.			
14	14	14	14	14	14	14																																
CBS6545	2.98/2.46/1.	1.91/2.86/1.	2.16/2.72/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.	2.01/2.53/1.			
25	25	25	25	25	25	25																																
CBS6546	1.73/1.96/1.	1.74/1.98/1.	1.97/2.41/1.	1.92/2.26/1.	1.78/2.03/1.	1.78/2.03/1.	1.78/2.03/1.	1.78/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.	1.79/2.03/1.			
17	17	17	17	17	17	17																																
CBS6547	1.66/1.63/0.	1.65/1.63/0.	1.91/2.13/1.	1.85/1.95/1.	1.71/1.74/1.	1.71/1.74/1.	1.71/1.74/1.	1.71/1.74/1.	1.73/1.78/1.	1.73/1.78/1.	1.73/1.																											

Gène	Nom systématique	Fonction	QTL fluconazole	QTL kétoconazole
SIA1	YOR137C	Protein of unassigned function involved in activation of the Pma1p plasma membrane H <sup>+</sup> -ATPase by glucose	Non	Oui
RUP1	YOR138C	Protein involved in regulation of Rsp5p, which is an essential HECT ubiquitin ligase; has a WW domain consensus motif of PPPSY (residues 131-135) that mediates binding of Rsp5p to Ubp2p; contains an UBA domain (1, 2)	Oui	Oui
SFL1	YOR140W	Transcriptional repressor and activator; involved in repression of flocculation-related genes, and activation of stress responsive genes; negatively regulated by cAMP-dependent protein kinase A subunit Tpk2p	Oui	Oui
ARP8	YOR141C	Nuclear actin-related protein involved in chromatin remodeling, component of chromatin-remodeling enzyme complexes; has mRNA binding activity	Oui	Oui
LSC1	YOR142W	Alpha subunit of succinyl-CoA ligase, which is a mitochondrial enzyme of the TCA cycle that catalyzes the nucleotide-dependent conversion of succinyl-CoA to succinate; phosphorylated	Oui	Oui
SUF5	tG(CCC)O	Glycine tRNA (tRNA-Gly), predicted by tRNAscan-SE analysis; can mutate to suppress +1 frameshift mutations in glycine codons	Oui	Oui
YORWdelta1 2		Ty1 LTR	Oui	Oui
YOR142W-B		Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes	Oui	Oui
YOR142W-A		Retrotransposon TYA Gag gene co-transcribed with TYB Pol; translated as TYA or TYA-TYB polyprotein; Gag is a nucleocapsid protein that is the structural constituent of virus-like particles (VLPs); similar to retroviral Gag	Oui	Oui
YORWTy1-2		Long Terminal Repeat (LTR) retrotransposon of the Copia (Pseudoviridae) group; contains co-transcribed genes TYA Gag and TYB Pol, encoding proteins involved in structure and function of virus-like particles, flanked by two direct repeats	Oui	Oui
YORWdelta1 3		Ty1 LTR	Oui	Oui
YORWdelta1 2		Ty1 LTR (1, 2 and see Summary Paragraph)	Oui	Oui
YORWdelta1 4		Ty1 LTR	Oui	Oui
YOR142W-B		Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes	Oui	Oui
YORWTy1-2		Long Terminal Repeat (LTR) retrotransposon of the Copia (Pseudoviridae) group; contains co-transcribed genes TYA Gag and TYB Pol, encoding proteins involved in structure and function of virus-like particles, flanked by two direct repeats	Oui	Oui
THI80	YOR143C	Thiamine pyrophosphokinase, phosphorylates thiamine to produce the coenzyme thiamine pyrophosphate (thiamine diphosphate)	Oui	Oui
ELG1	YOR144C	Subunit of an alternative replication factor C complex important for DNA replication and genome integrity; suppresses spontaneous DNA damage; involved in homologous recombination-mediated repair and telomere homeostasis	Oui	Oui
ARS1502		Autonomously Replicating Sequence	Oui	Oui
PNO1	YOR145C	Essential nucleolar protein required for pre-18S rRNA processing, interacts with Dim1p, an 18S rRNA dimethyltransferase, and also with Nob1p, which is involved in proteasome biogenesis; contains a KH domain	Oui	Oui
YOR146W	YOR146W	Dubious open reading frame unlikely to encode a protein, based on available experimental and comparative sequence data; open reading frame overlaps the verified gene PNO1/YOR145C	Oui	Oui
MDM32	YOR147W	Mitochondrial inner membrane protein with similarity to Mdm31p, required for normal mitochondrial morphology and inheritance; interacts genetically with MMM1, MDM10, MDM12, and MDM34	Oui	Oui
SPP2	YOR148C	Essential protein that promotes the first step of splicing and is required for the final stages of spliceosome maturation; interacts with Prp2p, which may release Spp2p from the spliceosome following the first cleavage reaction	Oui	Oui
SMP3	YOR149C	Alpha 1,2-mannosyltransferase involved in glycosyl phosphatidyl inositol (GPI) biosynthesis; required for addition of the fourth, side branching mannose to the GPI core structure	Oui	Oui
MRPL23	YOR150W	Mitochondrial ribosomal protein of the large subunit	Oui	Oui
RPB2	YOR151C	RNA polymerase II second largest subunit B150, part of central core; similar to bacterial beta subunit	Oui	Oui
ARS1518		Putative replication origin, identified in multiple array studies, not yet confirmed by plasmid-based assay	Oui	Oui
YOR152C		Putative protein of unknown function; YOR152C is not an essential gene	Oui	Oui
PDR5	YOR153W	Plasma membrane ATP-binding cassette (ABC) transporter, multidrug transporter actively regulated by Pdr1p; also involved in steroid transport, cation resistance, and cellular detoxification during exponential growth	Oui	Oui
SLP1	YOR154W	Integral membrane protein of unknown function; member of the SUN-like family of proteins; genetic interactions suggest a role in folding of ER membrane proteins	Oui	Oui

Gène	Nom systématique	Fonction	QTL fluconazole	QTL kétoconazole
SIA1	YOR137C	Protein of unassigned function involved in activation of the Pma1p plasma membrane H <sup>+</sup> -ATPase by glucose	Non	Oui
RUP1	YOR138C	Protein involved in regulation of Rsp5p, which is an essential HECT ubiquitin ligase; has a WW domain consensus motif of PPPSY (residues 131-135) that mediates binding of Rsp5p to Ubp2p; contains an UBA domain (1, 2)	Oui	Oui
SFL1	YOR140W	Transcriptional repressor and activator; involved in repression of flocculation-related genes, and activation of stress responsive genes; negatively regulated by cAMP-dependent protein kinase A subunit Tpk2p	Oui	Oui
ARP8	YOR141C	Nuclear actin-related protein involved in chromatin remodeling, component of chromatin-remodeling enzyme complexes; has mRNA binding activity	Oui	Oui
LSC1	YOR142W	Alpha subunit of succinyl-CoA ligase, which is a mitochondrial enzyme of the TCA cycle that catalyzes the nucleotide-dependent conversion of succinyl-CoA to succinate; phosphorylated	Oui	Oui
SUF5	tG(CCC)O	Glycine tRNA (tRNA-Gly), predicted by tRNAscan-SE analysis; can mutate to suppress +1 frameshift mutations in glycine codons	Oui	Oui
YORWdelta1 2		Ty1 LTR	Oui	Oui
YOR142W-B		Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes	Oui	Oui
YOR142W-A		Retrotransposon TYA Gag gene co-transcribed with TYB Pol; translated as TYA or TYA-TYB polyprotein; Gag is a nucleocapsid protein that is the structural constituent of virus-like particles (VLPs); similar to retroviral Gag	Oui	Oui
YORWTy1-2		Long Terminal Repeat (LTR) retrotransposon of the Copia (Pseudoviridae) group; contains co-transcribed genes TYA Gag and TYB Pol, encoding proteins involved in structure and function of virus-like particles, flanked by two direct repeats	Oui	Oui
YORWdelta1 3		Ty1 LTR	Oui	Oui
YORWdelta1 2		Ty1 LTR (1, 2 and see Summary Paragraph)	Oui	Oui
YORWdelta1 4		Ty1 LTR	Oui	Oui
YOR142W-B		Retrotransposon TYA Gag and TYB Pol genes; transcribed/translated as one unit; polyprotein is processed to make a nucleocapsid-like protein (Gag), reverse transcriptase (RT), protease (PR), and integrase (IN); similar to retroviral genes	Oui	Oui
YORWTy1-2		Long Terminal Repeat (LTR) retrotransposon of the Copia (Pseudoviridae) group; contains co-transcribed genes TYA Gag and TYB Pol, encoding proteins involved in structure and function of virus-like particles, flanked by two direct repeats	Oui	Oui
THI80	YOR143C	Thiamine pyrophosphokinase, phosphorylates thiamine to produce the coenzyme thiamine pyrophosphate (thiamine diphosphate)	Oui	Oui
ELG1	YOR144C	Subunit of an alternative replication factor C complex important for DNA replication and genome integrity; suppresses spontaneous DNA damage; involved in homologous recombination-mediated repair and telomere homeostasis	Oui	Oui
ARS1502		Autonomously Replicating Sequence	Oui	Oui
PNO1	YOR145C	Essential nucleolar protein required for pre-18S rRNA processing, interacts with Dim1p, an 18S rRNA dimethyltransferase, and also with Nob1p, which is involved in proteasome biogenesis; contains a KH domain	Oui	Oui
YOR146W	YOR146W	Dubious open reading frame unlikely to encode a protein, based on available experimental and comparative sequence data; open reading frame overlaps the verified gene PNO1/YOR145C	Oui	Oui
MDM32	YOR147W	Mitochondrial inner membrane protein with similarity to Mdm31p, required for normal mitochondrial morphology and inheritance; interacts genetically with MMM1, MDM10, MDM12, and MDM34	Oui	Oui
SPP2	YOR148C	Essential protein that promotes the first step of splicing and is required for the final stages of spliceosome maturation; interacts with Prp2p, which may release Spp2p from the spliceosome following the first cleavage reaction	Oui	Oui
SMP3	YOR149C	Alpha 1,2-mannosyltransferase involved in glycosyl phosphatidyl inositol (GPI) biosynthesis; required for addition of the fourth, side branching mannose to the GPI core structure	Oui	Oui
MRPL23	YOR150W	Mitochondrial ribosomal protein of the large subunit	Oui	Oui
RPB2	YOR151C	RNA polymerase II second largest subunit B150, part of central core; similar to bacterial beta subunit	Oui	Oui
ARS1518		Putative replication origin, identified in multiple array studies, not yet confirmed by plasmid-based assay	Oui	Oui
YOR152C		Putative protein of unknown function; YOR152C is not an essential gene	Oui	Oui
<b>PDR5</b>	YOR153W	Plasma membrane ATP-binding cassette (ABC) transporter, multidrug transporter actively regulated by Pdr1p; also involved in steroid transport, cation resistance, and cellular detoxification during exponential growth	Oui	Oui
SLP1	YOR154W	Integral membrane protein of unknown function; member of the SUN-like family of proteins; genetic interactions suggest a role in folding of ER membrane proteins	Oui	Oui

**Annexe 2.2 :** Liste des gènes présents dans le QTL du chromosome XV des analyses de pool de ségréants du couple Sace\_clinique au fluconazole et au kétoconazole.

## Publications et communications

### Publications réalisées durant la thèse :

Jung, P. P., A. Friedrich, C. Reisser, J. Hou, and J. Schacherer, 2012 Mitochondrial genome evolution in a single protoploid yeast species. *G3: Genes|Genomes|Genetics* 2: 1103–1111.

Reisser, C., C. Dick, L. Kruglyak, D. Botstein, J. Schacherer *et al.*, 2013 Genetic basis of ammonium toxicity resistance in a sake strain of yeast: a mendelian case. *G3: Genes|Genomes|Genetics* 3: 733–740.

### Publications en préparation :

Reisser C, Friedrich A, de Montigny J, Schacherer J. Evolving mechanisms leading to the fluconazole resistance are conserved within the Hemiascomycetes yeasts.

Friedrich A, Jung PP, Reisser C, Schacherer J. Population genomics of a preduplication yeast species.

### Présentation orale :

Séminaire de microbiologie de Strasbourg, April 2012, Strasbourg, France: « Analyse de la résistance multiple aux antifongiques au sein de deux espèces de levures ».

### Posters :

Doctoral school day, December 2012, Strasbourg, France: « Intra-specific variability of multidrug resistance in yeast populations ».

EMBO congress, October 2011, Sant Feliu, Spain: « Intra-specific variability of multidrug resistance in yeast populations ».

## Activité d'enseignement

09/2010 à 06/2012 : 2 ans de monitorat en biologie animale U.E. biologie cellulaire 2 : techniques d'histologie et de cytologie.





# Analyse de la variabilité intraspécifique chez les levures : résistance à l'ammonium et aux composés azolés

Dans toutes les espèces, les mutations et les réarrangements chromosomiques constituent des moteurs de l'évolution des génomes. Ils génèrent une diversité génétique à l'origine de la variation phénotypique observée entre les individus d'une même espèce. Cette variation est particulièrement importante chez les levures. Elles constituent donc d'excellents modèles pour déterminer les origines génétiques de la variation intra-spécifique.

C'est dans ce contexte que ce travail s'est focalisé sur l'étude de la variation de résistance à l'ammonium et aux antifongiques azolés chez deux espèces de levures : *Saccharomyces cerevisiae* et *Lachancea kluyveri*.

L'analyse des origines génétiques de la résistance à ces composés a mis en évidence que les variations génétiques pouvaient avoir lieu à plusieurs niveaux : séquence codante pour la résistance à l'ammonium et séquence régulatrice pour la résistance aux antifongiques. De plus, la réalisation d'expériences d'évolution adaptative a permis de mettre en évidence que l'adaptation à un nouvel environnement se faisait par dosage génique via l'acquisition d'un chromosome supplémentaire chez les espèces étudiées.

**Mots-clés :** *S. cerevisiae*, *L. kluyveri*, antifongiques, ammonium, analyses de pool de ségréants, QTL, génomique, expériences d'évolution.

In all species, mutations and chromosomal rearrangements are drivers of genomes evolution. These processes generate the genetic diversity at the origin of the phenotypic variations observed between the individuals of the same species. This variation is essential for their adaptation to a new environment. The yeasts are isolated from various ecological and geographical niches and show an important phenotypic variation. According to these characteristics, they are excellent model organisms to determine the genetic origins of the observed phenotypic variation.

In this context, the study focused on the variation of resistance to ammonium and azole antifungals within two yeast species: *Saccharomyces cerevisiae* and *Lachancea kluyveri*.

The analyses of the genetic origin of the resistance to these compounds show that this genetic variation could occur at several levels: coding sequence for resistance to ammonium and regulatory sequence for resistance to antifungal agents. In addition, evolving experiments have showed that the adaptation to a new environment was done by gene dosage, through the acquisition of extrachromosomes in both species studied.

**Keywords:** *S. cerevisiae*, *L. kluyveri*, antifungals, ammonium, bulk segregants analyses, QTL, genomic, evolving experiments.

Discipline : Sciences du Vivant

Spécialité : Aspects moléculaires et cellulaires de la biologie

Laboratoire de Génétique Moléculaire, Génétique et Microbiologie

UMR7156 UDS/CNRS

28 rue Goethe 67083 Strasbourg Cedex.