

- Thèse présentée pour obtenir le grade de
- Docteur de l'Université de Strasbourg
- •
- Discipline : Sciences du Vivant
- Aspects moléculaires et cellulaires de la biologie
- par Paul Jung •

Réarrangements chromosomiques et génomique fonctionnelle chez la levure *Saccharomyces cerevisiae*. Génomique comparative des génomes mitochondriaux des levures hémiascomycètes

Soutenue publiquement le 17 décembre 2010 •

Membres du jury
Mme Corinne Clavé (Professeur, Université Victor Segalen Bordeaux)
Rapporteur externe
M. Marc Lemaire (Professeur, Université Claude Bernard Lyon)
Rapporteur externe
M. Ivan Tarassov (Directeur de recherche CNRS, Université de Strasbourg)
Examinateur
M. Joseph Schacherer (Maître de conférences, Université de Strasbourg)
Membre invité
M. Jacky de Montigny (Professeur, Université de Strasbourg)
Directeur de thèse

Le travail présenté dans ce mémoire a été réalisé au sein du laboratoire de Génétique Moléculaire, Génomique et Microbiologie UMR7156 UDS/CNRS. Je tiens à remercier Serge Potier et Jean-Luc Souciet de m'avoir accueilli au sein de leur laboratoire et de leur équipe.

Je tiens à remercier Madame Corinne Clavé (Université Victor Segalen, Bordeaux) et Messieurs Marc Lemaire (Université Claude Bernard, Lyon) et Ivan Tarassov (Université de Strasbourg) d'avoir accepté d'examiner mon travail.

Un grand merci à Jacky de Montigny de m'avoir encadré durant mon stage de Master et de m'avoir donné l'occasion de poursuivre ces travaux par cette thèse. Je voudrais notamment te remercier pour m'avoir donné l'occasion de m'initier à l'enseignement supérieur. Cela fut également un « plaisir » d'avoir été ton standardiste pendant ces trois années.

Je tiens particulièrement à remercier Joseph Schacherer pour m'avoir tant aidé lors de mes travaux et pour son écoute bienveillante. Merci aussi pour tes blagues et ta quasi-permanente bonne humeur. Lorsque tu as accepté le poste d'enseignant chercheur, Jacky m'avait dit que nous devrions bien nous entendre : je crois qu'il ne s'était pas trop trompé.

Merci à toutes les personnes du groupe Levure que j'ai eu la chance de côtoyer lors de ces trois années : Claudine, Laurence, Anne, Véronique, Marie-Laure, Catherine et Valérie. J'ai aussi une pensée spéciale pour les étudiants en master ou en thèse avec qui j'ai pu travaillé : Simo, Zlatyo, Élie, Samuyl et Fabien, sans oublier Cyrielle à qui je lègue ma très chère paillasse.

Je tiens aussi à remercier tous les doctorants que j'ai eu la chance de rencontrer : Nic, Sandro, Émilie M., Fahran, Mika, Audrey, Jessica, Marie, François et David. Merci également à Christelle pour sa gentillesse et sa chevelure flamboyante ainsi qu'à Sandrine pour sa bonne humeur et ses coups de gueule, sans oublier Anne-Marie pour son aide. Et merci à tous les membres du laboratoire que je n'ai pas cité. Merci à Christophe, mon binôme de fac, pour nos soirées squash et billard... je te promets que je te battrai un jour.

Enfin, je tiens plus que particulièrement à remercier deux anciens thésards du laboratoire. Émilie « maître », tu as été et resteras pour moi un modèle de par ta gentillesse, ta bonne humeur, ton abnégation au travail... Christian, tu as toujours été de bons conseils aussi bien professionnellement que personnellement, dans les moments de joie comme de doute.

Et voilà le moment de remercier mes parents, mes frères et toute ma famille pour m'avoir toujours demandé « tu en es où de tes études ? » ou « mais tu fais quoi exactement ? » et pour m'avoir soutenu durant ces huit années passées à l'université. Je ne pourrais finir sans remercier ma compagne Monia et notre fils Justin : merci pour tout l'amour et toute la joie que vous m'apportez. Je ne sais pas si j'aurais fini entier après les derniers jours de la rédaction de ce manuscrit sans vous.

TABLE DES MATIÈRES

IN	TRODUCTION	19
	···· · · · · · · · · · · · · · · · · ·	
1.	Historique de la plasticité des génomes	19
2.	Eléments de plasticité	22
	2.1. Les réarrangements chromosomiques	22
	2.1.1. Les duplications	22
	2.1.2. Les délétions	23
	2.1.3. Les inversions	23
	2.1.4. Les translocations	23
	2.1.5. Les insertions	23
	2.2. Les échanges de matériel génétique entre organismes	24
	2.2.1. Transfert horizontal	24
	2.2.2. Introgression	25
3.	Impacts des réarrangements chromosomiques	26
	3.1. Apparition et perte de nouvelles fonctions	26
	3.1.1. Devenir des gènes dupliqués	26
	3.1.1.1.La conservation de fonction	27
	3.1.1.2.La néofonctionnalisation	28
	3.1.1.3.La subfonctionnalisation	28
	3.1.1.4.La pseudogénisation	30
	3.1.2. Gènes « rebrassés » et apparition de gènes de fusion	30
	3.1.2.1.La permutation par duplication	30
	3.1.2.2.La fusion de gènes	31
	3.2. Adaptation à l'environnement	32
	3.3. Pathologies humaines et réarrangements chromosomiques	33
	3.3.1. Les aneuploïdies dans les tumeurs solides	33
	3.3.2. Les gènes de fusion et les leucémies	33
	3.3.2.1.Implication des gènes de fusion RAR α -PML et	

	RAR a-PLZF dans la leucémie promyélocytaire aiguë	34
	3.3.2.2.Implication du gène de fusion IgH-MYC dans la	
	leucémie de Burkitt	34
4.	Saccharomyces cerevisiae et l'évolution des génomes	
	des levures hémiascomycètes	35
	4.1. Saccharomyces cerevisiae : un organisme modèle pour étudier les	
	réarrangements chromosomiques	35
	4.2. Évolution des génomes des levures hémiascomycètes	37
M	ATÉRIELS ET MÉTHODES	41

1. Organismes et milieux de culture 41 1.1. Les organismes 41 41 1.1.1. Saccharomyces cerevisiae 1.1.2. Zygosaccharomyces rouxii, Pichia sorbitophila et Pichia farinosa 42 1.2. Les milieux de culture 43 1.3. Les conditions de culture 43 1.4. Conservation des souches 43 1.5. Mesure de la croissance cellulaire 43 2. Techniques génétiques de la levure 44 2.1. Obtention des cellules diploïdes 44 2.2. Obtention des cellules haploïdes 44 2.3. Dissection des tétrades 44 45 2.4. Analyse phénotypique 2.4.1. Détermination des phénotypes 45 45 2.4.2. Détermination du signe sexuel : Halo Mating Type Assay 46 3. Détermination de la morphologie des cellules et des colonies 4. Préparation des acides nucléiques 46 4.1. Préparation rapide d'ADN génomique 46 4.2. Préparation des ARN totaux de S. cerevisiae 46 46 4.2.1. Préparation des cellules 4.2.2. Préparation des ARN 47

5.	Analyse,	quantification et	purification des a	acides nucléiques	48
----	----------	-------------------	--------------------	-------------------	----

	5.1. Analyse des acides nucléiques : électrophorèse en gel d'agarose	48	
	5.2. Quantification des acides nucléiques		
	5.3. Purification de l'ADN	48	
	5.3.1. Précipitation alcoolique de l'ADN	48	
	5.3.2. Purification de l'ADN par chromatographie	48	
6.	Les PCR	49	
	6.1. Amplification d'un fragment d'ADN par PCR	49	
	6.1.1. Les différentes étapes de la PCR	49	
	6.1.2. Matériel	50	
	6.2. Amplification d'un fragment d'ARN par RT-PCR	50	
7.	Les RT-PCR quantitatives (qRT-PCR)	50	
	7.1. Réverse transcription des ARN totaux	51	
	7.2. Préparation de la gamme d'ADNc	52	
	7.2.1. Transcription in vitro	52	
	7.2.2. Formation de la gamme d'ADNc	52	
	7.3. La réaction de qPCR	52	
8.	Détermination de la croissance cellulaire	54	
	8.1. Méthode automatique	54	
	8.2. Méthode manuelle	54	
	8.2.1. Croissance cellulaire	54	
	8.2.2. Consommation spécifique de glucose	55	
	8.2.3. Production spécifique d'éthanol	56	
	8.2.4. Détermination du poids sec	58	
9.	Test d'activité enzymatique ATCase	58	
	9.1. Préparation de l'extrait cellulaire	58	
	9.2. Dosage des protéines	58	
	9.3. Test enzymatique	59	
10.	Transformation des cellules de levures	60	
	10.1. La méthode d'électroporation	60	
	10.1.1. Préparation des cellules compétentes	60	
	10.1.2. Électroporation	60	
	10.2. Transformation par l'acétate de lithium	60	
11.	Bioinformatique	61	
	11.1. Les banques de données		

11.2.	Annotation des génomes	61
11.3.	Construction des arbres phylogénétiques	61
11.4.	Traitement des données issues des hybridations sur puces à ADN	61
PARTIE	Ι	63
CHAPIT	RE 1 : PRÉSENTATION DU SYSTÈME DE SÉLECTION	63
Introducti	on	63
1. Le sys	stème de sélection	63
2. Les ou	utils de caractérisation	64
2.1. H	ybridations de type ADN-ADN par la méthode Southern	66
2.2. C	aryotypes électrophorétiques par électrophorèse en champs alternés	66
2.3. C	aractérisation fine des réarrangements par PCR	67
2.4. H	ybridations sur puces à ADN	67
3. Les ty	pes de réarrangements chromosomiques sélectionnés	67
3.1. L	es insertions de rétrotransposons	68
3.2. L	es délétions	69
3.3. L	es duplications	70
3.	3.1. Les duplications géniques	71
3.	3.2. Les duplications segmentales	71
3.	3.3. Les aneuploïdies et les translocations	72
3.4. L	es mécanismes d'apparition des réarrangements chromosomiques	74
3.5. L	es gènes de fusion associés aux réarrangements chromosomiques	75
4. La col	lection de révertants utilisée lors de cette étude	75
Conclusio	n	77

CHAPITRE 2 : IMPACTS DES RÉARRANGEMENTS CHROMOSOMIQUES 79

Introduction		
1.	Caractérisation de la physiologie cellulaire des révertants	79
	1.1. Comparaison de la morphologie des colonies	80

	1.1.1.	Analyse de la croissance sur milieu complet	81
	1.1.2.	Analyse de la croissance sur des milieux contenant	
		d'autres sources de carbone	84
	1.2. Analy	vse de la morphologie cellulaire des révertants	88
2.	Croissance	e cellulaire	90
	2.1. Analy	/se de la croissance cellulaire	90
	2.1.1.	Détermination des temps de génération des révertants	91
	2.1.2.	Détermination des temps de latence des révertants	93
	2.2. Déter	mination de la consommation spécifique de glucose dans les révertants	96
	2.3. Déter	mination de la production spécifique d'éthanol dans les révertants	97
3.	Analyse g	lobale de la transcription	100
	3.1. Les d	onnées analysées	100
	3.2. Expre	ession des gènes impliqués dans la voie de biosynthèse	
	des n	ucléotides pyrimidiques	101
	3.3. Rôle	de l'expression des gènes et de la ploïdie dans la réponse cellulaire	104
	3.3.1.	Les parties dupliquées et délétées	104
	3.3.2.	Rôle de la ploïdie dans la réponse cellulaire	106
	3.3.3.	Réponse cellulaire et stress	109
Co	nclusion		113

CHAPITRE 3 : IMPACTS DES GÈNES DE FUSION

117

Int	Introduction		
1.	Caractérisation de la fonctionnalité des domaines des protéines chimères	121	
	1.1. Le domaine codant l'ATCase	121	
	1.1.1. Détermination de l'activité ATCase	121	
	1.1.1.1.Les souches haploïdes		
	1.1.1.2.Les souches diploïdes		
	1.1.2. Expression des gènes de fusion		
	1.1.3. Localisation des protéines chimères	125	
	1.2. Le gène receveur		
	1.2.1. Analyse fonctionnelles des gènes receveurs	127	

1.2.1.1.Remplacement chromosomique par croisement	131
1.2.1.2.Remplacement du gène sauvage par un allèle mutant	
construit in vitro	133
1.2.1.3.Le gène de fusion YDJ1-ATCase	133
1.2.1.4.Le gène de fusion PBS2-ATCase	135
2. Caractérisation de la rétroinhibition des protéines chimères	136
2.1. Résultats	137
2.2. Caractérisation de l'absence de rétroinhibition	138
Conclusion	140

CONCLUSIONS ET PERSPECTIVES

143

PARTIE II

C	OMPARAI	SON DES	GÉNOMES	MITOCHONDRIAUX	DES	LEVURES
H	ÉMIASCO	MYCÈTES				147
т	1					1 47
Int	roduction					147
1.	Les génon	nes mitochond	riaux des levures	s du groupe CTG		154
	1.1. Organ	nisation des gér	nomes des levure	es du groupe CTG		155
	1.1.1.	Contenu en g	ènes			155
	1.1.2.	Les introns de	es gènes mitocho	ondriaux		160
	1.1.3.	La maturation	n des ARN			161
	1.2. Géno	mique compara	ative des génome	es mitochondriaux		
	des le	evures du grou	pe CTG			162
	1.2.1.	Aspects génér	raux			162
	1.2.2.	Code génétiq	ue utilisé			163
	1.2.3.	Analyse de la	synténie des gé	nomes mitochondriaux		164
	1.2.4.	Analyse phyl	ogénétique			165
	1.3. Conc	lusion				170
2.	Organisati	on du génome	mitochondrial d	e Zygosaccharomyces rouxi	i	171

175

197

2
3
3
3

RÉFÉRENCES BIBLIOGRAPHIQUES

ENSEIGNEMENT ET RECHERCHE

ANNEXES	211
Annexe 1 : The complete mitochondrial genome of the yeast Pichia sorbitophila	211
Annexe 2 : Complete mitochondrial genome sequence of the yeast Pichia farinosa	
and comparative analysis of closely related species	219
Annexe 3 : The Ty1 LTR-retrotransposon population in Saccharomyces cerevisiae ge	enome:
dynamics and sequence variations during mobility.	229

LISTE DES FIGURES

INTRODUCTION GÉNÉRALE

Figure I1	Représentation schématique d'une introgression	26
Figure I2	Le devenir des gènes dupliqués	27
Figure I3	Exemple de néofonctionnalisation pour les gènes EDN et ECP	29
Figure I4	Représentation schématique d'une permutation par duplication	31
Figure I5	Représentation schématique d'une fusion de gènes	32
Figure I6	Arbre phylogénétique des levures hémiascomycètes	38

MATÉRIEL ET MÉTHODES

Figure MM1	Cycle cellulaire de S. cerevisiae	42
Figure MM2	Schéma récapitulatif des étapes nécessaires pour l'obtention	
	d'une gamme d'ADNc utilisable pour les qRT-PCR en 2 étapes	53
Figure MM3	Exemple d'une courbe de croissance obtenue	
	par la méthode automatique	55

PARTIE I

CHAPITRE 1 : LE SYSTÈME DE SÉLECTION

Figure 1.1	Le gène URA2, la protéine Ura2p et la voie de biosynthèse	
	de l'UTP chez S. cerevisiae	65
Figure 1.2	Localisation des sites de restriction BamHI au niveau	
	du locus $ura2_{15-30-72}$	66
Figure 1.3	Réacticvation de l'ATCase par insertion de rétrotransposns Ty1	69

Figure 1.4	Détermination du site d'initiation de la transcription dans les	révertants
	possédant une insertion de rétrotransposon Ty1 en orientation sens	70
Figure 1.5	Exemple de délétion segmentale	71
Figure 1.6	Exemple de duplication génique	72
Figure 1.7	Les duplications caractérisées à l'état diploïde	73
Figure 1.8	Duplication génique par rétroposition	74

CHAPITRE 2 : IMPACTS DES RÉARRANGEMENTS CHROMOSOMIQUES

Figure 2.1	Morphologie des colonies sur un milieu complet	82
Figure 2.2	Morphologie des colonies du révertant B066/1 n	83
Figure 2.3	Morphologie des colonies sur des milieux contenant	
	diverses sources de carbone	86
Figure 2.4	Morphologie des révertants	89
Figure 2.5	Détermination des temps de génération des révertants	92
Figure 2.6	Détermination des paramètres de croissance à partir des résultats obten	us
	par le lecteur de microplaques TECAN INFINITE F200	94
Figure 2.7	Détermination des temps de latence des révertants	95
Figure 2.8	Exemple de la détermination du rendement de glucose	
	consommé Y pour la souche 2nU+	96
Figure 2.9	Consommation spécifique de glucose et production spécifique	
	d'éthanol des révertants	98
Figure 2.10	Exemples de coefficients de corrélation	101
Figure 2.11	Expression des gènes dupliqués, étant impliqués dans la	
	formation des chromosomes chimères	105
Figure 2.12	Exemple de processus cellulaires dont les gènes ont une	
	expression modifiée	108
Figure 2.13	Profil d'expression des gènes ESR dans les révertants	111

CHAPITRE 3 : IMPACTS DES GÈNES DE FUSION

Figure 3.1 La voie de biosynthèse des nucléotides pyrimidiques chez S. cerevisiae 118

Figure 3.2	Structure d'un gène chimère	119
Figure 3.3	Représentation des protéines de fusion	129
Figure 3.4	Représentation schématique de la stratégie pour réaliser	
	les tests phénotypiques associés aux gènes receveurs de la fusion	131
Figure 3.5	Représentation schématique du remplacement du chromosome	
	porteur de la copie intacte du gène receveur de la fusion par	
	croisement avec une souche EUROSCARF	132
Figure 3.6	Représentation schématique de l'inactivation du gène SNA3	
	par double recombinaison homologue	134

PARTIE II

COMPARAISON DES GÉNOMES MITOCHONDRIAUX DES LEVURES HÉMIASCOMYCÈTES

Figure 4.1	Arbre phylogénétique des levures hémiascomycètes	153
Figure 4.2	Représentation des réactions de PCR pour obtenir les génomes	
	mitochondriaux complets de P. sorbitophila et P. farinosa	155
Figure 4.3	Représentation des génomes mitochondriaux	
	de P. sorbitophila et P. farinosa	158
Figure 4.4	Cartes linéaires des génomes mitochondriaux de P. stipitis,	
	P. guilliermondii, C. tropicalis et L. elongisporus	159
Figure 4.5	Alignement multiple de la protéine Atp9p	164
Figure 4.6	Conservation de la synténie au sein des espèces du genre Pichia	165
Figure 4.7	Arbre phylogénétique des levures du groupe CTG	169
Figure 4.8	Carte linéaire du génome mitochondrial de Z. rouxii	172

LISTE DES TABLEAUX

MATÉRIEL ET MÉTHODES

Tableau MM1	Les différentes souches de S. cerevisiae utilisées	
	lors de cette étude	41

PARTIE I

CHAPITRE 1 : LE SYSTÈME DE SÉLECTION

5
5

CHAPITRE 2 : IMPACTS DES RÉARRANGEMENTS CHROMOSOMIQUES

Tableau 2.1	Liste des révertants utilisés et les réarrangements	
	chromosomiques associés	80
Tableau 2.2	Variation de la morphologie des colonies dans un milieu	
	complet YPD	83
Tableau 2.3	Tableau récapitulatif des morphologies des colonies	
	en fonction de la source de carbone utilisée	87
Tableau 2.4	Tableau récapitulatif des morphologies des cellules	
	de la collection de révertants	88
Tableau 2.5	Les temps de génération des révertants	93
Tableau 2.6	Consommation spécifique de glucose et production	
	spécifique d'éthanol	99
Tableau 2.7	Mesure des coefficients de corrélation inter-lames	102
Tableau 2.8	L'expression des gènes codant des protéines intervenant	

	dans la voie de biosynthèse des nucléotides pyrimidiques	103
Tableau 2.9	Nombre de gènes possédant une expression différente	
	dans un type de réarrangement chromosomique	106
Tableau 2.10	Gènes exprimés spécifiquement chez les révertants	
	diploïdes aneuploïdes	109
Tableau 2.11	Résistance à la rapamycine des révertants aneuploïdes	113

CHAPITRE 3 : IMPACTS DES GÈNES DE FUSION

Tableau 3.1	Les gènes chimères analysés	120
Tableau 3.2	Mesure des activités ATCase des souches haploïdes	122
Tableau 3.3	Mesure des activités ATCase des souches diploïdes	123
Tableau 3.4	Mesure de l'expression des gènes de fusion	125
Tableau 3.5	Phénotypes des gènes receveurs de la fusion	130
Tableau 3.6	Analyse fonctionnelle du gène chimère PBS2-ATCase	
	du révertant AWR58	136
Tableau 3.7	Analyse de la résistance au 5-FU de la collection de révertants	139
Tableau 3.8	Analyse de la transcription des gènes FUR1 et FUR4	
	dans les révertants	139

PARTIE II

COMPARAISON DES GÉNOMES MITOCHONDRIAUX DES LEVURES HÉMIASCOMYCÈTES

Tableau 4.1	Caractéristiques des génomes des mitochondries	
	des levures hémiascomycètes	152
Tableau 4.2	Les introns mitochondriaux et leur localisation	160
Tableau 4.3	Caractéristiques des génomes mitochondriaux annotés	163

LISTE DES ABREVIATIONS

aa	: acides aminés
Δ	: délétion
ADN	: acide désoxyribonucléique
ARN	: acide ribonucléique
dNTP	: 2'-désoxyribonucléotides
EDTA	: éthylène diamine tétraacétate
kb	: kilobase
min	: minute
π_{EtOh}	: production spécifique d'éthanol
pb	: paire(s) de base
Q_{Glu}	: consommation spécifique de glucose
qsp	: quantité suffisante pour
rpm	: rotations par minute
Tris	: tris-hydroxyméthyl-aminométhane

INTRODUCTION

Introduction

Au laboratoire de « Génétique Moléculaire, Génomique et Microbiologie », une des thématiques étudiées est l'évolution des génomes des organismes eucaryotes. Les axes majeurs de recherche s'intéressent au remodelage de l'ADN et s'orientent autour de l'apparition et de la disparition de gènes. Nous travaillons d'une part sur la levure *Saccharomyces cerevisiae* dans le but de déterminer les mécanismes qui conduisent aux réarrangements chromosomiques et leurs impacts sur la physiologie de la cellule. D'autre part, nous analysons également la variabilité des génomes d'espèces différentes de levures hémiascomycètes par le biais d'une analyse de génomique comparative. Lors de ce travail, nous nous sommes intéressés aux réarrangements chromosomiques chez *S. cerevisiae* et la réponse cellulaire suite à la mise en place de ces remaniements. Notre étude a également porté sur l'analyse des génomes des mitochondries d'un groupe particulier de levures ayant une position phylogénétique intéressante au sein des hémiascomycètes et comprenant notamment des espèces pathogènes pour l'Homme.

Dans cette introduction, après un bref retour historique des découvertes majeures qui ont mené à la description de la dynamique des génomes, nous aborderons les différents processus permettant cette plasticité. Nous traiterons également les conséquences des réarrangements chromosomiques sur la cellule avec notamment la création de nouveaux gènes. Enfin, nous détaillerons la génomique comparative des levures hémiascomycètes et traiterons l'utilisation de *S. cerevisiae* comme organisme modèle pour étudier les remaniements chromosomiques.

1. Historique de la plasticité des génomes

Chaque organisme présente un ensemble de caractéristiques qui est transmis à ses descendants. Cette transmission a été mise en évidence en 1866 suite aux travaux menés sur l'espèce *Pisum sativum* plus connue sous le nom de pois par Johan Gregor Mendel. Ses travaux furent envoyés à des botanistes dans de nombreux instituts mais n'eurent d'échos favorables qu'à partir de l'année 1900. En effet, à cette date, les travaux menés

indépendamment par d'autres biologistes comme Hugo de Vries, Carl Correns et Erich von Tschermark aboutirent à des conclusions similaires à celles décrites par Mendel 20 ans plus tôt. Ces auteurs citèrent d'ailleurs les travaux de Mendel et permirent ainsi que les règles de l'hérédité soient connues sous le nom de lois de Mendel. En 1902, Hugo de Vries postulait que la naissance d'une nouvelle espèce résultait de l'apparition d'une mutation (de Vries, 1902). Les travaux de génétique réalisés après la redécouverte des lois de Mendel notamment ceux de Thomas Morgan et ses collaborateurs sur la drosophile ont abouti à l'élaboration des premières cartes génétiques. Ces premières explorations de génomes animaux et végétaux ont donné une vision statique de celui-ci. Le fait qu'un certain nombre de fonctions essentielles devaient être conservées dans un organisme donné, laissait supposer que seules quelques modifications mineures pouvaient être acceptées. Il était alors admis que les modifications intervenaient majoritairement lors du brassage génétique en méiose et que les mutations ne jouaient qu'un faible rôle dans la variation du génome. De ce fait, les génomes ont été définis comme étant des structures stables qui ne pouvaient subir que très peu de modifications. L'étude des caractères héréditaires sera nommée génétique en 1906 par William Bateson. Trois ans plus tard, ces éléments héréditaires seront décrits comme des gènes par Wilhelm Johannsen. Le mot génome est quant à lui utilisé pour la première fois en 1920 dans un ouvrage rédigé en allemand par le botaniste Hans Winckler pour définir le lot de chromosomes et les gènes qu'ils contiennent (Winkler, 1920).

L'idée que les génomes sont des structures statiques fut remise en cause dans les années 1950 avec les travaux de Barbara McClintock et ses études sur les plants de maïs (McClintock, 1956). Ces résultats montraient l'existence d'éléments mobiles dans le génome appelés « gènes sauteurs ». Ces éléments correspondaient aux premiers éléments transposables qui furent mis en évidence dans une espèce. Ces éléments, aujourd'hui appelés éléments transposables ou transposons, furent ensuite découverts aussi bien chez les eucaryotes que chez les procaryotes. La génomique et l'étude des séquences complètes montrent aujourd'hui l'universalité de ces éléments transposables dans les génomes ainsi que leur l'importance quantitative. Dans des génomes comme ceux des plantes, le nombre de ces éléments peut représenter jusqu'à 65% du génome des espèces *Gossipyum* ou 70% du génome du maïs (Sanmiguel et Bennetzen, 1998 ; Hawkins et *al.*, 2006).

Dans la fin des années 1970 et l'avancée des techniques de biologie moléculaire, les scientifiques se sont de plus en plus intéressés aux génomes. L'année 1977 est une date

importante en biologie moléculaire car elle coïncide avec l'apparition des méthodes de séquençage de l'ADN dont notamment celle de Sanger (Sanger et *al.*, 1977). Cependant, il fallut attendre presque 20 années pour obtenir les premiers génomes entièrement séquencés. Les génomes des procaryotes *Haemophilus influenzae* et *Mycoplasma genitalium* ont été les premiers et ont été publiés en 1995 (Fleischmann et *al.*, 1995 ; Fraser et *al.*, 1995). Le premier génome eucaryote à être séquencé est celui de S. *cerevisiae* (Goffeau et *al.*, 1996). Depuis, le nombre de génomes entièrement séquencés ne cesse d'augmenter et au jour du 30 octobre 2010, 1485 génomes complets sont publiés et accessibles dans les bases de données.

La comparaison des génomes a débuté à la fin des années 1990 grâce notamment au consortium Génolevures qui a comparé des séquences partielles de levures hémiascomycètes (Souciet et al., 2000). Depuis, les analyses de génomique comparative se sont répandues dans le monde scientifique et des analyses phylogénétiques sont effectuées sur différents organismes. Ces analyses ont notamment permis de démontrer que les réarrangements chromosomiques sont ubiquitaires et sont à la base de l'apparition de nouvelles espèces. Par exemple chez Arabidospis thaliana, un processus de spéciation apparaît suite à une modification dans le nombre de chromosomes (Bomblies et Weigel, 2007). De la même manière, 29 inversions sont retrouvées entre les génomes de Drosophila melanogaster et Drosophila vakuba et sont à la base de la différence entre ces deux espèces (Ranz et al., 2007). La comparaison des génomes de des plantes Vitis vinifera, A. thaliana, Oryza sativa et Populus trichocarpa montre qu'un événement d'héxaploïdisation a eu lieu après la séparation des espèces monocotylédones des dicotylédones (Jaillon et al., 2007). De plus, des analyses des génomes d'individus appartenant à une même espèce sont également menées pour étudier les variations nucléotidique et chromosomique associées à des phénotypes. Ainsi, une analyse génomes de 1000 individus humains montre de nombreux réarrangements des chromosomiques et un polymorphisme nucléotidique important. Par exemple, dans les génomes de 60 individus appartenant à une même communauté ayant des ancêtres européens et vivant dans l'Utah, 259 duplications et 33% de divergence nucléotidique ont été mis en évidence à partir d'un séquençage à faible couverture génomique (The 1000 Genomes Project Consortium, 2010).

La génomique, mot énoncé pour la première fois lors d'un congrès par Tom Roederick en 1986, peut se décliner aujourd'hui en génomique structurale, fonctionnelle et comparative selon les aspects abordés. Elle a pour conséquence d'apporter un éclairage nouveau sur notre vision du génome en tant que structure, en tant que contenu d'informations et surtout en tant que dynamique, source d'innovations.

2. Éléments de plasticité

2.1. Les types de réarrangements chromosomiques

Les génomes actuels peuvent être considérés comme le résultat de l'accumulation d'événements qui se sont produits dans le passé. Les comparaisons de génomes d'espèces proches montrent qu'ils sont notamment le siège de grands remaniements chromosomiques telles des duplications, des délétions, des inversions, des insertions et des translocations. Ils peuvent être à l'origine de la mort cellulaire mais aussi la cause de maladies génétiques notamment chez l'Homme.

2.1.1. Les duplications

Les duplications sont définies comme étant l'événement à partir duquel un gène passe d'une copie à deux. Les possibilités par lesquelles une séquence d'ADN peut se dupliquer au sein d'un génome sont multiples et peuvent être classées selon la taille de la région doublée. On retrouve :

- La duplication génique : les éléments dupliqués peuvent être situés l'un à côté de l'autre. On appelle cet événement une duplication en tandem. Ces duplications peuvent aussi être dispersées au sein du génome.
- La duplication segmentale : une série de gènes contigus est dupliquée. Les deux fragments dupliqués peuvent être situés sur le même chromosome (duplication intrachromosomale) ou sur deux chromosomes distincts (duplication interchromosomale). Lors de duplications intrachromosomales, les deux copies peuvent être fusionnées en tandem.
- La duplication de chromosome : un chromosome entier est dupliqué au sein d'une cellule. Cette duplication est également appelée aneuploïdie (aberration chromosomique due à un nombre anormal de chromosomes). Une cause de la plupart des aneuploïdies est une non-disjonction des chromatides sœurs au cours de la mitose ou de la méiose. Cette non-disjonction correspond à un défaut du processus de ségrégation des chromosomes.

 La duplication totale du génome. Cette duplication est également appelée polyploïdisation. Chez les polyploïdes, il faut distinguer les autopolyploïdes, composés de chromosomes provenant d'une espèce unique, des allopolyploïdes qui possèdent des jeux de chromosomes provenant d'au moins deux espèces.

2.1.2. Les délétions

Les délétions sont des événements à partir desquels un segment d'ADN est perdu. Ces délétions peuvent survenir au niveau de séquences intergéniques et géniques. Pour les délétions géniques, en fonction de la taille de ce fragment délété, deux types de délétions sont décrits :

- Les délétions intragéniques où la séquence délétée est interne à un seul gène
- Les délétions segmentales où plusieurs gènes contigus sont supprimés

2.1.3. Les inversions

Contrairement aux événements de duplications et de délétions, les événements inversions n'engendrent pas des gains ou des pertes de matériel génétique et sont dits équilibrés. En effet, les inversions ne modifient pas le nombre de gènes, mais changent la structure d'un chromosome. Ces réarrangements permettent de changer l'orientation d'un ou plusieurs gènes contigus.

2.1.4. Les translocations

Les translocations sont définies comme étant un échange de fragments d'ADN entre deux chromosomes. Deux types de translocations sont décrites : les réciproques et les non réciproques. Les translocations réciproques sont des réarrangements équilibrés contrairement aux translocations non réciproques. Dans ce dernier cas, une duplication d'un segment de chromosome remplace la délétion d'un fragment d'un autre chromosome.

2.1.5. Les insertions

Les éléments transposables peuvent être définis comme des séquences d'ADN capables de se déplacer et se multiplier dans le génome de façon autonome. Deux types d'éléments sont décrits : les transposons à ADN et les rétrotransposons. Les transposons à ADN se déplacent dans le génome grâce à une transposase qui est une enzyme codée par cet élément mobile. La région codant la transposase est encadrée par de courtes séquences

inversement répétées. Le mode de fonctionnement est du type couper-coller ou copier-coller dans le cas où il y a eu réplication avant transposition.

Les rétrotransposons se déplacent selon un mode copier-coller et sont spécifiques des organismes eucaryotes. Différents types de rétrotransposons sont décrits. Nous ne les détaillerons pas davantage. Après transcription par l'ARN polymérase II, l'ARN messager est rétrotranscrit en ADN par la transcriptase réverse codée par le rétrotransposon. Cet ADN est alors inséré dans le génome grâce à une activité intégrase également codée par cet élément.

Ces insertions dans le génome se font aléatoirement sans nécessiter de séquences particulières. Elles peuvent avoir différents effets en fonction du lieu d'insertion de l'élément mobile :

- dans une région non codante : pas d'effet sur les gènes, que ce soit au niveau de leur transcription ou de leur traduction
- dans une phase codante : effet mutagène par interruption de la phase de lecture
- dans un intron : différentes possibilités. Cette insertion peut n'avoir aucun effet, stopper prématurément la transcription ou entraîner un épissage incorrect de l'ARNm conduisant à une protéine tronquée ou chimère
- dans une région promotrice : modification des séquences promotrices entraînant une un arrêt ou une modification de la transcription du gène

2.2. Les échanges de matériel génétique entre organismes

D'autres événements que les réarrangements chromosomiques sont également à la base des modifications des génomes. Le génome d'un organisme peut acquérir des séquences d'ADN appartenant à un autre génome.

2.2.1. Transfert horizontal

Le transfert horizontal est défini comme un gain d'une séquence d'ADN provenant d'un autre organisme. Ainsi, une espèce peut acquérir un fragment d'ADN appartenant à une deuxième espèce. Ces transferts horizontaux sont notamment retrouvés entre différentes bactéries isolées de patients atteints de maladies nosocomiales. Chez les procaryotes, trois mécanismes essentiels sont à la base des transferts horizontaux : la transduction, la conjugaison et la transformation bactériennes. Ces mécanismes ne seront pas décrits dans ce manuscrit.

Les transferts horizontaux sont également retrouvés chez les eucaryotes. Les fragments d'ADN exogènes peuvent pénétrer dans la cellule par endocytose ou suite à une endosymbiose (Kidwell, 1993). Ces segments d'ADN peuvent notamment provenir d'organismes procaryotes. Par exemple l'urochordé *Ciona intestinalis* possède un gène codant une cellulose synthase d'origine bactérienne (Nakashima et *al.*, 2004). Chez *S. cerevisiae*, le gène *URA1* provient également d'un transfert horizontal à partir de la bactérie *Lactococcus lactis* (Hall et *al.*, 2005). Les gènes *BIO3* et *BIO4*, dont les protéines encodées interviennent dans la voie de biosynthèse de la biotine, sont des autres exemples de transfert entre *S. cerevisiae* et des souches bactériennes. En fonction du pourcentage d'identité obtenu lors d'alignements de séquences, il apparaît que le gène *BIO3* pourrait provenir de *Yersinia enterocolitica* et *BIO4* de *Magnetospirillum magnetotacticum* (Hall et *al.*, 2005). De ce fait, ce mécanisme d'échange de matériel génétique permet l'apparition de nouvelles fonctions au sein d'un organisme.

2.2.2. Introgression

Un autre mécanisme permettant l'acquisition d'un fragment d'ADN est l'introgression. Suite à l'hybridation de deux espèces différentes A et B, les génomes de ces deux espèces vont constituer celui de la souche hybride (Figure II). Une série de croisements retours avec la souche A permettra un enrichissement du génome de l'hybride par celui de cette souche. Si un fragment d'ADN de la souche B donne un avantage prolifératif ou permet une adaptation à un nouvel environnement de l'hybride, il sera conservé (Figure II). Les événements d'introgression sont naturels. Par exemple, un échange d'une région de 40 kb contenant 14 gènes dont un est l'homologue de l'énolase I a été mis en évidence entre différentes espèces de la levure basidiomycète *Cryptococcus neoformans* par ce mécanisme d'introgression (Kavanaugh et *al.*, 2006). L'introgression est également utilisée dans l'industrie agroalimentaire pour la fabrication de plantes génétiquement modifiées. Par exemple le facteur de transcription codé par le gène *GhDREB* est transmis de l'espèce de coton *Gossypium hirsutum* à l'espèce de froment *Triticum aestivum* par introgression, lui conférant une plus grande tolérance vis-à-vis de la sécheresse (Gao et *al.*, 2009).



<u>Figure I1 :</u> Représentation schématique d'une introgression. Cet arbre indique quatre espèces différentes (A, B, C et D). Il montre notamment l'éloignement des espèces A et B. Un fragment de l'ADN de l'espèce B est transféré par introgression dans le génome de l'espèce A. Les barres bleues représentent deux chromosomes de l'espèce A. Les barres rouges représentent deux chromosomes de l'espèce B.

3. Impacts des réarrangements chromosomiques

3.1. Apparition et perte de nouvelles fonctions

3.1.1. Devenir des gènes dupliqués

Peu après l'événement de duplication, deux gènes dupliqués codent la même protéine. De ce fait, ces gènes sont redondants et l'un d'entre eux peut acquérir des mutations sans affecter la croissance cellulaire. Ces mutations peuvent survenir au niveau de la séquence promotrice ou au niveau de la séquence codante du gène. En fonction du type de mutations, le gène dupliqué va pouvoir acquérir une nouvelle fonction (néofonctionnalisation), partager la fonction initiale (subfonctionnalisation) ou alors dégénérer en un gène non fonctionnel (pseudogénisation) (Figure I2).



<u>Figure 12</u>: Le devenir des gènes dupliqués. Le gène initial code pour une protéine possédant deux domaines fonctionnels représentés en vert et orange. Les croix indiquent l'apparition de mutations ponctuelles.

3.1.1.1. La conservation de fonction

La conservation de la fonction de la copie dupliquée est probablement un phénomène qui nécessite un intérêt sélectif pour la cellule. En effet, la présence d'une copie supplémentaire peut présenter un avantage pour les gènes fortement exprimés comme les gènes ribosomiques. En effet, la synthèse protéique représente une des activités importantes de la cellule. La production d'une quantité supplémentaire de composants du ribosome peut donc être bénéfique pour la cellule. Ainsi, une mutation d'un gène ribosomique dupliqué résulte d'une baisse de la croissance. Chez *S. cerevisiae*, c'est notamment le cas pour les gènes paralogues *RPL20A* et *RPL20B* (Koszul et *al.*, 2004).

Cette conservation de fonction peut cependant être accompagnée de mutations dans la région promotrice d'un des deux gènes dupliqués pouvant engendrer des taux de transcription différents. Par exemple chez la drosophile, les gènes paralogues FRQ1 et FRQ2 qui codent des fréquenines (homologues aux récepteurs du calcium des cellules neuronales codés par le gène NCS chez les Mammifères), ont une séquence peptidique conservée mais une transcription modifiée. En effet, les transcrits de FRQ1 sont deux à trois fois plus abondants à la fin de l'embryogenèse par rapport à ceux de FRQ2. Cette différence d'expression provient

de mutations dans la région promotrice modifiant notamment les sites de fixation de facteurs de transcription (Sanchez-Gracia et *al.*, 2010).

3.1.1.2.La néofonctionnalisation

Une vue classique de l'évolution des gènes suite à leur duplication, présentée par Ohno (1970), est la suivante : un gène conserve la fonction initiale alors que le second peut acquérir des mutations au niveau de sa séquence codante qui peuvent engendrer l'apparition de nouvelles fonctions (Figure 12). Ce phénomène est décrit chez les primates pour deux gènes appartenant à la superfamille des ribonucléases, *EDN* (Eosinophyl-derived neurotoxin) et *ECP* (Eosinophil cationin protein). Les protéines codées par ces deux gènes sont retrouvées au niveau des granules des leucocytes éosinophiles. Le gène *EDN* code une protéine jouant un rôle antiviral du fait de son activité ribonucléolytique sur l'ARN génomique des rétrovirus. La protéine codée par *ECP* a une faible activité ribonucléasique et agit comme toxine pour différentes bactéries et parasites. Les Hominidés et les singes de l'ancien monde (Catarhiniens) comme le macaque ou le babouin possèdent ces deux gènes (Zhang et *al.*, 1998). Par contre, les singes du nouveau monde (Plathyriniens) comme le tamarin ne possèdent qu'un gène homologue à *EDN* et *ECP*. Ce gène est appelé *EDN* car il code une protéine n'ayant aucune activité antibactérienne ou antiparasitaire. Son activité RNase est beaucoup plus faible que celle observée pour les Cathariniens.

Un événement de duplication en tandem du gène *EDN* ancestral serait survenu après la divergence des Plathyriniens et des Cathariniens (Figure I3). Le gène ancestral ne possédait pas l'activité antibactérienne ou antiparasitaire, ce qui laisse supposer que cette fonction a été acquise par le gène *ECP* après la duplication (Zhang et *al.*, 1998).

3.1.1.3.La subfonctionnalisation

Le processus de subfonctionnalisation, aussi connu sous le terme de duplication – dégénération – complémentation DDC (Force est *al.*, 1999), consiste à un partage des fonctions entre deux gènes dupliqués. Les deux copies dupliquées d'un gène possédant initialement deux fonctions porteront chacune un domaine fonctionnel, l'autre dégénérant par l'acquisition de mutations délétères (Lynch et Force, 2000). Un exemple de subfonctionnalisation a été mis en évidence lors de la comparaison des levures *Kluyveromyces lactis* et *S. cerevisiae* (Hittinger et Carroll, 2007). Comme nous le verrons dans cette introduction (dans la partie «*Saccharomyces cerevisiae* et l'évolution des génomes des

levures hémiascomycètes»), *S. cerevisiae* diverge d'une espèce ayant subi une duplication totale du génome alors que ce n'est pas le cas pour *K. lactis*. Chez *S. cerevisiae*, les gènes *GAL1* et *GAL3* codent des protéines intervenant dans la voie de dégradation du galactose : *GAL1* code la galactokinase qui est l'enzyme intervenant dans la première étape du catabolisme de ce sucre, alors que *GAL3* code pour un co-inducteur de cette voie. Chez la levure *Kluyveromyces lactis*, l'orthologue du gène *GAL1* code une protéine portant les deux fonctions. Ces résultats suggèrent que suite à la duplication, les deux copies du gène ont divergé chez *S. cerevisiae* pour porter chacune une des deux fonctions initiales.





3.1.1.4. La pseudogénisation

La néofonctionnalisation et la subfonctionnalisation sont des mécanismes permettant de fixer les copies dupliquées dans la descendance. Néanmoins, dans la plupart des cas, une des deux copies acquiert des mutations délétères pour le gène qui ne permettent pas de conserver cette copie active (Lynch et Conery, 2000). On parle alors de pseudogénisation.

Chez les Mammifères, la famille multigénique des récepteurs olfactifs possède jusqu'à mille gènes majoritairement regroupés dans différentes parties du génome. Néanmoins, tous ces gènes ne sont pas fonctionnels car une partie d'entre eux dégénèrent en pseudogènes. De ce fait seule une proportion de cette famille de gène code pour des protéines actives. Ainsi, environ 70% sont des pseudogènes chez l'Homme alors que seulement 20% ont subi ce processus chez la souris (pour revue, Kratz et *al.*, 2002 ; Hasin-Brumshtein et *al.*, 2009).

3.1.2. Gènes « rebrassés » et apparition de gènes de fusion

3.1.2.1. La Permutation par duplication

Par le biais de la néofonctionnalisation, les duplications engendrent des gains de fonctions avec l'acquisition de gènes codant des nouvelles protéines. Au niveau de l'évolution des protéines, les séquences codantes peuvent acquérir des modifications entraînant l'émergence de nouveaux repliements. Un exemple de ce type de brassage intragénique est la permutation circulaire dans laquelle le gène code une protéine qui conserve la fonction initiale mais subit un changement de topologie. Cette permutation circulaire est caractérisée par un changement des positions C-terminale et N-terminale par rapport à la protéine sauvage, en conservant la même séquence protéique (Vogel et Morea, 2006).

La permutation par duplication est un exemple qui illustre ce processus. Cet événement consiste en la duplication d'un gène puis la fusion en tandem des copies dupliquées. Il peut alors y avoir des délétions partielles au niveau des parties N et C-terminales. On obtient ainsi un variant de la protéine initiale ne changeant que par le positionnement des codons d'initiation et de terminaison de la traduction (Figure I4). Peisajovich et *al.* (2006) ont montré une permutation par duplication dans le gène codant les ADN-méthyl-transférases chez *Escherichia coli*.


<u>Figure I4 :</u> Représentation schématique d'une permutation par duplication. Les lettres A, B et C représentent trois régions de la protéine.

3.1.2.2. La fusion de gènes

Lors de la permutation par duplication, deux copies dupliquées sont fusionnées en tandem. Les différents types de réarrangements chromosomiques (duplication, délétion, insertion, inversion et translocation) peuvent aussi conduire à la fusion de deux gènes distincts. Cette fusion peut être réalisée entre deux séquences codantes ou entre une séquence codante et une région promotrice (Figure I5). Les gènes de fusion (ou gènes chimères) ainsi formés peuvent coder des protéines multifonctionnelles possédant des domaines protéiques appartenant initialement à deux protéines différentes.

Dans la majorité des cas de protéines multifonctionnelles, les fonctions portées par ces protéines interviennent dans une même voie. Par exemple chez les Mammifères, le gène *CAD* code une protéine possédant trois domaines fonctionnels (CPSase, ATCase et DHOase) qui interviennent dans les trois premières étapes de la voie de biosynthèse des nucléotides pyrimidiques. Chez *S. cerevisiae*, les gènes *HIS4* et *HIS7* dont les protéines interviennent dans la voie de biosynthèse de l'histidine proviennent aussi de la fusion de gènes. La protéine His7p possède des domaines protéiques décrits pour les gènes *HisH* et *HisF* de *E. coli*, alors que His4p a ceux décrits pour *HisI*, *HisE* et *HisD* (Fani et *al.*, 1994 ; Fani et *al.*, 2007).



<u>Figure 15 :</u> Représentation schématique d'une fusion de gènes. Les flèches représentent le site d'initiation de la transcription. Les rectangles représentent les séquences codantes des gènes 1 et 2.

A) Fusion de la séquence codante du gène 2 avec la région promotrice du gène 1.

B) Fusion des séquences codantes des gènes 1 et 2.

3.2. Adaptation à l'environnement

Les réarrangements chromosomiques apparaissant dans le génome d'un individu d'une espèce donnée peuvent lui conférer un avantage sélectif dans certaines conditions environnementales. Il a été montré que des souches de *Pseudomonas aeruginosa* isolées de poumons de patients atteints de mucoviscidose présentent des inversions chromosomiques de grande taille (pouvant aller jusqu'à 4 Mb) par rapport à la souche standard PAO (Kresse et *al.*, 2003). Ces inversions peuvent être accompagnées de l'insertion d'éléments génétiques comme le transposon Tn6061 qui confère une résistance de ces bactéries à différentes drogues (Coyne et *al.*, 2010).

De la même manière, des souches de *Candida albicans*, une levure pathogène pour l'Homme, isolées de patients atteints possèdent des aneuploïdies au sein de leur génome en comparaison avec une souche de référence (Selmecki et *al.*, 2006 ; Selmecki et *al.*, 2009). Ces souches sont résistantes vis-à-vis du fluconazole, un antifongique, alors que la souche sauvage est sensible indiquant que la présence de chromosome surnuméraire confère cette résistance.

3.3. Pathologies humaines et réarrangements chromosomiques

3.3.1. Les aneuploïdies dans les tumeurs solides

La présence d'aneuploïdie dans des cellules est connue depuis le début du XX^{ème} siècle grâce aux travaux de Theodor Boveri en 1902 puis 1914 sur des embryons d'oursins (pour revue, Balmain, 2001). Des aneuploïdies sont présentes dans 90% des cellules cancéreuses chez l'Homme. Cependant, leur rôle dans l'initiation ou la progression tumorale n'est pas encore défini (Marx et *al.*, 2002 ; Schvartzman et *al.*, 2010). En effet, l'apparition de chromosome supplémentaire peut provenir d'une instabilité chromosomique en méiose et en mitose. Une diminution de l'expression de gènes codant des protéines mitotiques comme *BUBR1, MAD2*, ou *BUB3* peut expliquer une telle instabilité et être l'élément déclencheur de l'apparition d'aneuploïdie (Michel et *al.*, 2001 ; Babu et *al.*, 2003 ; Baker et *al.*, 2004).

La présence d'aneuploïdie engendre un changement dans le nombre de copies de gènes par rapport à une cellule saine, ce qui peut altérer la régulation de certains gènes suppresseurs de tumeurs ou d'oncogènes comme *MYC* et *ERBB2* et ainsi induire la tumorisation (Alitalo et *al.*, 1983 ; Slamon et *al.*, 1989). Néanmoins, de récentes études ont démontré que les aneuploïdies, en plus de leur implication dans la progression tumorale, pouvaient également inhiber cette prolifération en fonction du contexte cellulaire (Weaver et *al.*, 2007). La présence de chromosome surnuméraire peut inhiber la tumorisation en surexprimant certains gènes comme *ETS2* dans des cas de trisomie 21 (Sussan et *al.*, 2007).

3.3.2. Les gènes de fusion et les leucémies

Les aneuploïdies ne sont pas les seuls réarrangements chromosomiques retrouvés dans des cellules cancéreuses. En effet, des événements d'amplification génique, de translocation et de polyploïdie sont aussi retrouvés dans ces cellules (Lengauer et *al.*, 1998 ; Albertson et *al.*, 2003). Les translocations réciproques, contrairement aux autres remaniements cités, n'engendrent pas de modifications dans le nombre de gènes ou de chromosomes mais un changement dans la structure du génome. Ce type de translocation est accompagné de la

formation de gènes de fusion à la base du processus de tumorisation. Nous allons voir ici deux exemple de leucémies en fonction des gènes de fusion observés : la leucémie promyélocytaire aiguë et la leucémie de Burkitt.

3.3.2.1. Implication des gènes de fusion *RARα-PML et RARα-PLZF* dans la leucémie promyélocytaire aiguë

La leucémie promyélocytaite aiguë résulte de translocations variées où les gènes *PML* ou *PLZF* sont fusionnés en phase avec la séquence codante du gène codant le récepteur de l'acide rétinoïque (*RARa*) (Koken et *al.*, 1997 ; Zhou et *al.*, 2006). Sous sa forme sauvage, ce récepteur est un facteur de transcription qui se dimérise et réprime de nombreux gènes. Cependant, cette répression est levée par de faibles concentrations d'acide rétinoïque (10^{-8} M). Lors de la fusion *RARa-PML* ou *RARa-PLZF*, la même concentration de cet acide ne permet pas le détachement de ce facteur de transcription des régions promotrices de ces gènes cibles. Parmi eux, on retrouve notamment des gènes impliqués dans la différenciation cellulaire. Le maintien de l'expression de ces gènes induit ainsi la tumorisation.

3.3.2.2. Implication du gène de fusion IgH-MYC dans la leucémie de Burkitt

La leucémie aiguë de Burkitt est une pathologie liée aux lymphocytes B, qui est causée, dans 75% des cas par une translocation entre les chromosomes 8 et 14. Cette translocation entraîne la fusion du proto-oncogène *MYC* avec les séquences promotrices du gène codant les chaînes lourdes des immunoglobulines IgH (Albertson, 2006). Ce proto-oncogène joue un rôle important dans la multiplication et la mort cellulaires dans une cellule saine. Les lymphocytes B produisant en permanence des anticorps, le gène *MYC* sera continuellement actif dans ces cellules, entraînant ainsi la multiplication cellulaire. La leucémie aiguë de Burkitt est une pathologie n'affectant que les lymphocytes B car si la fusion IgH-*MYC* intervenait dans un autre type cellulaire, elle n'aurait aucun effet : les gènes codant pour les anticorps ne sont exprimés que dans ce type de lymphocyte.

4. Saccharomyces cerevisiae et l'évolution des génomes des levures hémiascomycètes

4.1. *Saccharomyces cerevisiae* : un organisme modèle pour étudier les réarrangements chromosomiques

La levure hémiascomycète *S. cerevisiae* est un organisme modèle eucaryote. C'est notamment le premier organisme eucaryote dont le génome a été entièrement séquencé. Ce génome, d'une taille de 12 Mb, est constitué d'environ 6000 gènes (Goffeau et *al.*, 1996).

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 rapidement grâce à des techniques classiques de biologie moléculaire. Une collection de délétions systématiques de tous les gènes de *S. cerevisiae* a été construite dans le but de réaliser une analyse fonctionnelle globale de l'ensemble du génome. Tous les gènes de la levure ont été délétés et remplacés par le gène KanMX entraînant une résistance vis-à-vis de 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. Il s'agit de la collection EUROSCARF (Brachmann et al., 1998).

S. cerevisiae est également un outil particulièrement adapté pour étudier les réarrangements chromosomiques. Cet organisme eucaryote possède un cycle haplodiplobiontique et les souches de laboratoire sont hétérothalliques. Ainsi, les analyses peuvent être menées aussi bien en contexte haploïde que diploïde. À partir des caractéristiques de *S. cerevisiae*, différents systèmes de sélection naturels de réarrangements chromosomiques ont été mis au point. Ces cribles induisent des pressions de sélection, soit en limitant des substrats de culture, soit en utilisant des analogues toxiques. Plusieurs types de ces cribles ont été utilisés comme par exemple :

Le système CAN1 URA3 (Chen et Kolodner, 1999). Dans ce crible de sélection, le gène HXT13, localisé à environ 7,5 kb du gène CAN1, a été remplacé par le gène URA3. Le gène CAN1 code une perméase à l'arginine permettant l'entrée de cet acide aminé dans la cellule mais aussi de la canavanine, son analogue toxique. Le gène URA3 code pour l'orotidine-5'-phosphate décarboxylase qui convertit l'OMP en UMP mais également le 5-FOMP (provenant de la métabolisation du 5-FOA)

en 5-FUMP pouvant conduire à la mort de la cellule. Ainsi, une souche sauvage est sensible à la canavanine et au 5-FOA. Ce crible positif repose sur la sélection de révertants résistants à ces deux composés suite à la perte des gènes *URA3* et *CAN1*. Les modifications observées à l'origine de ces pertes sont des remaniements chromosomiques comme des délétions, des translocations ou des événements plus complexes comme la fusion de chromosomes.

- Le système de sélection *RPL20A* (Koszul et *al.*, 2004). Une mutation dans le gène *RPL20A*, qui code une protéine ribosomale, entraîne un retard de croissance. La surexpression du gène paralogue *RPL20B*, possédant 99% d'identité de séquence avec *RPL20A*, permet une restauration de la croissance. Le crible utilisé repose sur l'utilisation d'une souche $\Delta rpl20A$ dans laquelle le gène *RPL20A* est délété. Les souches sélectionnées à partir de ce crible présentent à nouveau une croissance normale. L'analyse de ces révertants montre que la restauration du phénotype sauvage se fait suite à un réarrangement chromosomique permettant la duplication du gène *RPL20B*.
- La sélection par réponse à une pression de sélection (Dunham et al., 2002). Contrairement aux autres systèmes de sélection, ce crible ne repose pas sur l'activité de gènes marqueurs mais sur une pression de sélection. Une souche de S. cerevisiae est cultivée en chemostat en conditions limitantes de glucose. Après 100 à 500 générations dans ces conditions, des clones issus de cette culture sont analysés par hybridation sur puces à ADN de type CGH. Il a été montré que ces cellules pouvaient posséder plusieurs types de réarrangements chromosomiques comme des délétions, des translocations ou des duplications et notamment des aneuploïdies.

Au laboratoire, un système de sélection a également été mis au point (Exinger et Lacroute, 1979 ; Welcker et *al.*, 2000). Tout comme les différents cribles précédemment décrits, il permet la sélection de souches porteuses de remaniements chromosomiques spontanés *in vivo*. Il repose sur le principe de sélection positive à partir d'un gène marqueur (*URA2*) et présente l'avantage de permettre l'obtention de différents types de réarrangements comme les duplications, les translocations, les délétions et les insertions de rétrotransposons Ty1 dans un contexte haploïde et diploïde. Il sera décrit dans le chapitre I.

4.2. Évolution des génomes des levures hémiascomycètes

Les levures hémiascomyèctes ont servi de précurseurs dans les études de génomique comparative. À la fin des années 1990, le consortium Génolevures a posé les bases de la génomique comparative des levures hémiascomycètes en séquençant les génomes de 13 levures hémiascomycètes avec une faible couverture génomique de 0,4X et en réalisant leur annotation (Souciet et al., 2000). Lors de cette dernière décennie, ce consortium a enrichi les bases de données de génomes en effectuant le séquençage et l'annotation complète des levures Candida glabrata, Debaryomyces hansenii, Kluyveromyces lactis, Kluyveromyces thermotolerans, Saccharomyces kluyverii, Yarrowia lipolytica et Zygossacharomyces rouxii (Dujon et al., 2004 ; Souciet et al., 2009). D'autres laboratoires ont séquencé d'autres espèces de levures hémiascomycètes. Par exemple, le génome d'Ashbya gossypii a été séquencé par le Biozentrum de Bâle (Dietrich et al., 2004) alors que ceux de levures pathogènes pour l'Homme comme Candida albicans et Candida parapsilosis l'ont été par le Broad Institute (Butler et al., 2009). À ce jour, plus de trente espèces de levures appartenant à l'ordre des Hémiascomycètes sont séquencées (Figure I6). Ces espèces sont classées dans des groupes comme les Saccharomycetaceae, les Dipodascaceae et le groupe CTG (Figure I6) (Dujon, 2010).

Après le séquençage et l'assemblage du génome de *S. cerevisiae* en 1996 (Goffeau et *al.*, 1996), la mise en évidence de blocs de duplication a permis de proposer que ce génome avait évolué suite à une duplication totale du génome (Wolfe et Shields, 1997). Les analyses génomiques et la comparaison de différentes espèces séquencées ont confirmé l'évolution de levures comme *S. cerevisiae* et *C. glabrata* à partir d'un ancêtre commun dont le génome a été dupliqué (Kellis et *al.*, 2004). La séparation des espèces pré-duplication et post-duplication se fait entre les genres *Saccharomyces* et *Zygosaccharomyces* (Figure I6). De plus, ces analyses comparatives ont permis de mettre en évidence d'autres réarrangements chromosomiques entre les levures hémiascomycètes. Par exemple, plus de 300 inversions et translocations sont retrouvées lors des comparaisons des génomes de *S. cerevisiae* et *A. gossypii* (Dietrich et *al.*, 2004).



<u>Figure 16 :</u> Arbre phylogénétique des levures hémiascomycètes (modifié à partir de Dujon, 2010). Trois groupes majeurs sont définis pour ces levures : les *Saccharomycetaceae* (soulignés en bleu), les *Dipodasceae* (violet) et le groupe CTG (orange). Les genres allant des *Saccharomyces sensu stricto* aux *Vanderwaltozyma* divergent d'un ancêtre commun ayant eu une duplication totale de son génome (WGD).

Les réarrangements chromosomiques sont des événements essentiels dans l'évolution moléculaire des génomes. Ils permettent entre autres d'acquérir de nouvelles fonctions. Pour analyser ces remaniements, des cribles de sélection peuvent être utilisés chez S. cerevisiae dont notamment ceux exposés plus haut. Au laboratoire, le système de sélection mis au point repose sur un retour à la prototrophie vis-à-vis de l'uracile d'un mutant du gène URA2. Durant ces 15 dernières années, ce crible a permis d'isoler plus de 400 souches révertantes prototrophes. La caractérisation de ces révertants a démontré la présence de réarrangements chromosomiques variés. Ces souches possèdent également des gènes chimères où une partie du gène URA2 est fusionnée à un gène receveur. Suite à l'analyse de ces révertants, une collection de souches haploïdes et diploïdes, porteuses de différents remaniements et possédant différents types de gènes de fusion, a été constituée. Cette collection a servi pour étudier d'une part les impacts des réarrangements chromosomiques sur la physiologie cellulaire et de comprendre la réponse d'une cellule suite à l'apparition d'un tel événement. Nous avons notamment mis en évidence que la ploïdie jouait un rôle prépondérant dans cette réponse. D'autre part, cette collection a servi pour comprendre les impacts des gènes de fusion et pour déterminer le rôle des différentes parties de ces gènes. Ces résultats seront développés dans les trois premiers chapitres.

Dans la seconde partie, nous exposerons une analyse comparative des génomes mitochondriaux des levures hémiascomycètes et particulièrement ceux appartenant au groupe CTG (Figure I6). Notre laboratoire faisant partie du consortium Génolevures, j'ai participé à l'annotation des génomes de *Pichia sorbitophila* et *Arxula adenivorans*. Suite à cette annotation, il a été montré que 60% des gènes de *P. sorbitophila* ont deux versions alléliques dans le génome et sont hétérozygotes. De ce fait, cette espèce est décrite comme provenant de l'hybridation de deux espèces distinctes (Dujon, 2010). Une des deux espèces apparentée à *P. sorbitophila* est supposée être *Pichia farinosa* car ces deux levures ont été définies comme appartenant à la même espèce par des approches biochimiques (Kurtzman et Fell, 1998). Le génome de *P. farinosa* a été séquencé par notre laboratoire pour le comparer à celui de *P. sorbitophila*. Dans notre étude présentée ici, nous avons assemblé et annoté les séquences mitochondriales de *P. sorbitophila* et de *P. farinosa*. Dans le but de comparer ces espèces, nous avons utilisé les séquences mitochondriales annotées des espèces proches de *P. sorbitophila* comme *C. albicans, C. parapsilosis* et *D. hansenii* (Anderson et *al.*, 2001 ; Nosek et *al.*, 2004 ; Sacerdot et *al.*, 2008) et d'autres non-annotées comme *Pichia*

guilliermondii, *Candida tropicalis*, *Lodderomyces elongisporus et Pichia stipitis* (Butler et *al.*, 2009 ; Jeffries et *al.*, 2007). Nous avons annoté ces séquences et nous les avons utilisées pour réaliser une analyse phylogénétique des levures du groupe CTG (Figure I6). Cette étude montre clairement que les levures *P. farinosa* et *P. sorbitophila* sont phylogénétiquement éloignées et peuvent être définies comme des espèces distinctes. Ces résultats seront exposés dans le quatrième chapitre.

MATÉRIEL ET MÉTHODES

Matériel et Méthodes

1. Organismes et milieux de culture

1.1. Les organismes

1.1.1. Saccharomyces cerevisiae

La levure *Saccharomyces cerevisiae* est un champignon hémiascomycète hétérothallique. Son cycle cellulaire est haplodiplobiontique, c'est-à-dire qu'il peut se maintenir aussi bien sous forme haploïde que diploïde grâce à sa multiplication végétative (Figure MM1). En conditions défavorables (carence en carbone ou en azote), les cellules diploïdes vont sporuler et donner naissance à des asques contenant quatre spores non ordonnées. Lorsque ces spores sont en conditions plus favorables, elles vont germer et donner deux cellules de signe a et deux de signe α . Deux cellules de signes opposés peuvent ensuite fusionner pour générer une cellule diploïde. Ces étapes peuvent être reproduites en laboratoire (Figure MM1).

Les caractéristiques des différentes souches de *S. cerevisiae* utilisées lors de ce travail sont répertoriées dans le Tableau MM1. Ces cellules sont toutes isogéniques à la souche de laboratoire FL100. L'utilisation de souches de la collection EUROSCARF, isogéniques à la souche de référence S288c, a également été nécessaire pour mener notre projet à terme.

Souches	Contexte génétique	Génotype	Phénotype
aUT	FL100	MATa ura2 _{15,30,72} trp1Δ63	Ura ⁻ Trp ⁻
aUHT	FL100	MATa ura2 _{15,30,72} his3Δ200 trp1Δ63	Ura ⁻ His ⁻ Trp ⁻
αUL	FL100	MAT α ura $2_{15,30,72}$ leu $2\Delta 0$	Ura ⁻ Leu ⁻
EUROSCARF	BY4742	MAT α his3 $\Delta 1$ leu2 $\Delta 0$ lys3 $\Delta 0$ ura3 $\Delta 0$	His Leu Lys Ura

Tableau MM1 : Les différentes souches de S. cerevisiae utilisées lors de cette étude.

1.1.2. Zygosacharomyces rouxii, Pichia sorbitophila et Pichia farinosa

Ces trois levures sont des champignons hémiascomycètes osmotolérantes. Les souches *Zygossacharomyces rouxii* (CBS 732) et *Pichia sorbitophila* (CBS 7064) sont des cellules diploïdes alors que *Pichia farinosa* (CBS 185) est haploïde.



Figure MM1 : Cycle cellulaire de S. cerevisiae.

1.2. Les milieux de culture

Les milieux sont stérilisés par autoclavage à 110°C pendant 20 minutes (0,5 bar de pression). La composition des milieux utilisés est la suivante :

- milieu complet YPD : 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 présporulation GNA : glucose 50 g/l ; bacto beef extract 11,25 g/l ; bactopeptone 18,75 g/l ; yeast extract 10 g/l
- milieu de sporulation AcK : acétate de potassium 10 g/l

Tous ces milieux peuvent être solidifiés par l'ajout de 2% d'agar. Le milieu YNB peut être supplémenté par l'ajout d'acides aminés (50 μ g/l) en fonction des auxotrophies présentes. Le glucose comme source de carbone en milieu complet peut être remplacé par du galactose (20 g/l), du sorbitol (20 g/l), de l'acide acétique (20 g/l), du glycérol (3%) ou de l'éthanol (2%).

1.3. Les 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*, *P. sorbitophila* et *P. farinosa*, alors qu'elle est de 25 °C pour *Z. rouxii*.

1.4. Conservation des souches

Les souches de levure sont conservées à -80 °C dans 1,5 ml de milieu complet contenant 10% de glycérol.

1.5. Mesure de la croissance cellulaire

La concentration cellulaire d'une culture de S. cerevisiae peut être déterminée de deux manières :

 En milieu liquide, par mesure de l'absorbance de la culture à 600 nm par spectrophotomètre (ici Spectronic[®] 20 Genesys[™]).

Une unité de densité optique, pour S. cerevisiae équivaut à :

- $7x10^7$ cellules haploïdes/ml
- 4x10⁷ cellules diploïdes/ml

 En milieu solide, par étalement sur milieu complet gélosé d'une concentration connue de culture, sachant que chaque cellule viable mènera à l'apparition d'une colonie.

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 α 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.

Une autre méthode pour obtenir des diploïdes est l'utilisation de la complémentation allélique. L'utilisation du micromanipulateur est alors remplacée par un étalement sur milieu sélectif.

2.2. Obtention de cellules haploïdes

Des cellules diploïdes en phase exponentielle de croissance sont étalées sur milieu GNA et incubés pendant une nuit à 30 °C. Une anse de la culture est alors transférée sur milieu AcK et incubée à 30°C pendant 3 à 4 jours.

2.3. Dissection de tétrades

Elle permet d'analyser les phénotypes des spores issues de la méiose. Une anse de milieu de sporulation est mise en suspension dans 200 µl d'eau stérile à laquelle sont ajoutés 30 unités de zymolyase. On incube le mélange à 30 °C pendant 35 min.

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.

2.4. Analyse phénotypique

2.4.1. Détermination des phénotypes

Elle se fait en comparant les capacités de croissance des spores sur différents milieux sélectifs selon deux méthodes :

- le repiquage : chaque colonie est prélevée à l'aide d'un cure-dent ou d'une anse puis repiquée sur un autre milieu
- le test en gouttes : chaque colonie est mise en suspension dans de l'eau stérile. Une goutte est alors déposée sur différents milieux sélectifs. Cette technique permet de tester la croissance pour les spores étudiés sur des substrats différents ou sur des gammes de concentration de substrats.

2.4.2. Détermination du signe sexuel : Halo Mating Type Assay

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 α 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 µl de ces dilutions sont étalés sur des boîtes de milieu complet (il faut une boîte par souche). Ces dernières sont alors incubées à 30°C pendant 30 min.

Les cellules 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 cellules présentant un halo sur la boîte où la souche DBY7730 est étalée sont de signe α alors que celles ayant un halo sur la seconde boîte sont de signe a.

3. Détermination de la morphologie des cellules et des colonies

Les cellules en phase exponentielle de croissance sont prélevées et placées sous un microscope de type LEICA DM4000B, avec un grossissement 1000X.

Les souches sont isolées sur différents milieux et leur croissance est effectuée pendant 6 jours à 30°C. Les clones sont alors photographiés à l'aide d'un macroscope Leica Z16 APO A, avec un grossissement 10X.

4. Préparation des acides nucléiques

4.1. Préparation rapide d'ADN génomique (Hoffman et Winston, 1987)

Des cellules en phase stationnaire, cultivées dans 15 ml de milieu YPD ou YP galactose, sont récoltées par centrifugation de 5 min à 3000 rpm. Elles sont ensuite lavées avec 1 ml d'eau stérile, transférées dans un microtube Eppendorf puis reprises dans 250 µl de tampon de broyage (TrisHCl pH 8 10 mM ; EDTA 1 mM ; NaCl 100 mM, Triton 2% ; SDS 1%) auxquels 1 ml de billes de verre et 250 µl de phénol-chloroforme sont ajoutés. Ce mélange est vortexé vigoureusement pendant 3 min puis centrifugé à 11000 rpm durant 5 min. Afin de se débarrasser des protéines et autres débris cellulaires, seule la phase aqueuse du mélange est transférée dans un nouveau tube et lavée avec un volume de phénol-chloroforme. Les acides nucléiques sont alors précipités par l'ajout de 600 µl d'éthanol absolu et 20 µl d'acétate d'ammonium 4 M, puis placés pendant 1 heure à -20°C. Le mélange est centrifugé à 12400 rpm pendant 15 min, puis le culot est séché et repris dans 400 µl de tampon TE (TrisHCl pH 8 10 mM; EDTA 1 mM). Afin d'enlever les ARN de la préparation d'acides nucléiques, 5 µl de RNase A à 10 mg/ml sont ajoutés et le mélange est placé à 37°C pendant 10 min. L'ADN est alors précipité par l'ajout d'1 ml d'éthanol absolu et 10 µl d'acétate de sodium 3M et placé à -20°C pendant au moins 1 heure. Après une centrifugation de 15 min à 12400 rpm, le culot est lavé par 1 ml d'éthanol 80 % puis centrifugé pendant 5 min toujours à la même vitesse. Le culot est ensuite séché puis repris dans 50 µl de tampon TE ou d'eau déminéralisée.

4.2. Préparation des ARN totaux de S. cerevisiae

4.2.1. Préparation des cellules

Cinq ml de cellules en phase exponentielle de croissance (la densité optique à 600 nm est comprise entre 0,32 et 0,35 au spectrophotomètre Spectronic® 20 GenesysTM), cultivées

en milieu YNB, sont placés sur un filtre grâce à une pompe à vide (ce filtre bloque les organismes dont la taille est supérieure à 0,45 μ m). Ce filtre est alors congelé dans de l'azote liquide puis placé à -80°C jusqu'à la préparation des ARN pour ne pas induire de stress aux cellules lors de la récolte des cellules par centrifugations.

4.2.2. Préparation des ARN

La préparation des ARN totaux de *S. cerevisiae* a été effectuée en combinant le protocole d'extraction au phénol chaud et les 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.

Les filtres portant les cellules sont placés dans un tube Eppendorf de 2 ml en contact avec 750 µl 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°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 GelTMHeavy commercialisé par Eppendorf. 750 µl de chloroforme sont ajoutés 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 pendant une nuit à -20 °C en présence de 75 µl d'acétate de sodium 3 M et de 1,5 ml d'éthanol absolu.

Les tubes sont ensuite centrifugés 15 min à 12400 rpm. Le culot est repris dans 350 μ l d'éthanol 80% et 350 μ 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 sec à 12400 rpm. Elle est ensuite lavée successivement par 700 μ l de tampon RW1 puis 500 μ l de tampon RTE. Après chaque lavage, la colonne est centrifugée 15 sec à 12400 rpm. Elle est à nouveau rincée par l'ajout de 500 μ l de tampon RTE puis centrifugée 2 min. La colonne est placée dans un nouveau tube Eppendorf où les ARN seront élués en présence de 50 μ l d'eau grâce à une centrifugation de 1 min à 12400 rpm. La composition des tampons RLT, RW1 et RTE n'est pas donnée par le fournisseur.

5. Analyse, quantification et purification des acides nucléiques

5.1. Analyse des acides nucléiques : électrophorèse en 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 TAEx1 (Tris 4,84 g ; acétate de sodium 0,68 g ; acide acétique 1,35 ml ; EDTA 0,336 g ; H₂0 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 UV.

5.2. 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 μ g/ml, 33 μ g/ml ou 40 μ g/ml. Dans cette étude, les mesures ont été réalisées à l'aide d'un spectrophotomètre de type Nanodrop ND1000.

5.3. Purification de l'ADN

5.3.1. Précipitation alcoolique de l'ADN

En présence d'éthanol et d'un sel, les acides nucléiques précipitent. Cette propriété permet de purifier ou de concentrer une solution d'ADN.

On ajoute 2,5 volumes d'éthanol 100 % conservé à -20 °C, 1/10 de volume d'acétate de sodium 3 M et 1 μ l de glycogène. Le mélange est incubé à -20 °C pendant au moins 1 heure puis centrifugé pendant 15 min à 12400 rpm. Le culot est lavé avec de l'éthanol 80 %. Il est ensuite séché puis repris dans le volume désiré d'eau ou de tampon TE (TrisHCl pH 8 10 mM ; EDTA 1 mM).

5.3.2. Purification de l'ADN par chromatographie

Cette méthode est basée sur la technique de chromatographie par gel-filtration (ou exclusion). Elle consiste en une séparation différentielle des molécules en fonction de leur taille sur une résine de Sépacryl. Cette technique a été utilisée pour séparer des fragments de PCR des autres molécules présentes dans le milieu réactionnel (dNTPs, sel, amorces,...). Ces molécules, de taille plus petite, vont être retenues sur la colonne alors que l'ADN amplifié va être élué rapidement par centrifugation.

La colonne est remise en suspension, placée dans un microtube puis centrifugée à 2300 rpm pendant 1 min pour enlever l'excès de tampon. Après avoir placé la colonne dans un nouveau tube, 25 à 50 µl de l'échantillon à purifier sont déposés sur la surface de la résine puis centrifugés 2 min à 2300 rpm.

6. Les PCR

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

6.1.1. Les différentes étapes de la PCR

La PCR est une technique de biologie moléculaire qui permet l'amplification exponentielle et spécifique d'un fragment d'ADN en utilisant des oligonucléotides amorces, des dNTPs et une ADN polymérase thermostable. 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 :

- <u>la dénaturation</u> permet la séparation de l'ADN bicaténaire en ADN monocaténaire. La température classiquement utilisée est de 94°C pendant 30 s.
- <u>l'hybridation</u> 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 (Tm) est différente. Elle est généralement comprise entre 50°C et 60°C. La durée de cette étape varie de 30 s à 1 min.
- <u>l'élongation</u> où 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). Si l'on utilise la *Taq* DNA polymérase, la durée d'élongation est d'1 min/kb.

Le nombre de cycle dépend de la quantité de matrice initiale et il faut généralement de 25 à 35 cycles de PCR pour obtenir 1 mg d'ADN à partir de quelques picogrammes d'ADN. 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.

6.1.2. Matériel

Les réactions de PCR sont effectuées dans des tubes de 200 µl de volume ; le volume réactionnel est de 50 µl. Elles se font dans un thermocycleur de type Mastercycler[™] Gradient d'Eppendorf. L'enzyme généralement utilisée est la *Taq* DNA polymérase commercialisée par MP Biomedicals. Elle est fournie avec son tampon x10 (TrisHCl pH 9 10 mM, KCl 50 mM, MgCl2 1,5 mM, Triton X-100 0,1%, BSA 0,2 mg/ml) ainsi qu'une solution contenant les 4 dNTPs à une concentration de 10 mM.

La composition du milieu réactionnel pour un volume de 50 μ l est de : 5 μ l de tampon x10, 0,5 μ l de la solution contenant les 4 dNTPs, 10 pmoles de chaque amorce, environ 100 ng d'ADN total, 0,2 unités d'enzyme et de l'eau bidistillée qsp 50 μ l.

6.2. Amplification d'un fragment d'ARN par RT-PCR

La RT-PCR permet l'amplification spécifique et exponentielle d'un fragment d'ARN. La réaction est une PCR classique précédée d'une étape de réverse-transcription de l'ARN en ADN complémentaire. Pour réaliser les RT-PCR nous avons utilisé le kit SuperscriptTM III One-Step RT-PCR System with Platinium® *Taq* DNA Polymerase de Invitrogen. Ce kit est composé d'un mix réactionnel x2 (0,4 mM de chaque dNTPs, MgSO₄ 3,2 mM) et d'un mix d'enzymes contenant une réverse-transcriptase et une *Taq* DNA polymérase.

La réaction de RT-PCR commence par une réverse-transcription en plaçant le mélange réactionnel à 50°C pendant 30 min. Il s'en suit ensuite 40 cycles de PCR classique, avec une température d'élongation de 68°C qui est la température optimale pour la *Taq* DNA polymérase utilisée. Une étape d'élongation finale de 5 min à 68°C est également nécessaire.

La composition du milieu réactionnel pour un volume de 25 μ l est de : 12,5 μ l de mix réactionnel x2, 100 ng d'ARN, 10 pmoles de chaque amorce, 0,5 μ l du mix d'enzymes et de l'eau qsp 25 μ l.

7. Les RT-PCR quantitatives (qRT-PCR)

Les PCR et RT-PCR quantitatives peuvent être utilisées pour quantifier une quantité d'ADN ou d'ARN matrice initiale. 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, les résultats sont comparés deux à deux et un ratio de concentration peut être utilisé pour analyser les résultats.

Dans notre cas, les qRT-PCR ont été utilisées pour déterminer la quantité de transcrits. Les amorces ont été choisies de part et d'autres des jonctions des gènes de fusion et permettent l'amplification d'un fragment d'intérêt d'environ 200 pb. Chaque couple d'amorces ayant des efficacités d'hybridation différentes et la taille de chaque fragment d'intérêt n'étant pas exactement identique, les qRT-PCR absolues ont été choisies.

En fonction du nombre d'étapes, différentes qRT-PCR peuvent être appliquées. On distingue la qRT-PCR en une étape et celle en deux étapes. Lors des qRT-PCR en une étape, la réverse-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 réverse-transcription est réalisée avant la qPCR, et la quantification est faite au niveau des ADNc.

Nous avons choisi de faire des qRT-PCR absolues en deux étapes. Pour les réaliser, la réverse-transcription des ARN totaux a été effectuée ainsi que des gammes d'ADNc, pour déterminer exactement la quantité de matrice initiale.

7.1. Réverse transcription des ARN totaux

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, MgCl₂ 10 mM, dNTPs).

Pour réaliser la synthèse d'ADNc, le mélange réactionnel pour un volume de 20 μ l est le suivant : 10 μ l de mix réactionnel x2, 2 μ l de l'enzyme RT, 1 μ g d'ARN et de l'eau qsp 20 μ l. Pour que la réverse-transcription se produise, 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.

7.2. Préparation de la gamme d'ADNc

Chaque transcrit étudié ayant une taille et une séquence nucléotidique différentes, une gamme d'ADNc par transcrit est nécessaire. Une PCR, bornée par les amorces qui seront utilisées pour les qPCR, est effectuée et les fragments d'ADN obtenus sont purifiés. Le promoteur de l'ARN polymérase T7 et une queue polyA sont ensuite greffés à ces fragments grâce à une série de deux PCR. Ces nouveaux fragments sont ensuite utilisés pour faire une transcription *in vitro* puis une réverse-transcription.

7.2.1. Transcription in vitro

Pour réaliser ces expériences, le kit T7 RibomaxTM Express Large Scale RNA Production System de Promega a été utilisé. L'enzyme utilisée dans ce kit est l'ARN polymérase T7. Lors de cette réaction, 1 μ g de matrice, 2 μ l d'enzyme, 10 μ l de tampon x2 (contenant notamment les rNTPs) et de l'eau qsp 20 μ l sont placés à 37°C pendant 30 min.

Un traitement à la DNase I commercialisée par Invitrogen est nécessaire afin d'éliminer les ADN de la réaction. Les produits issus de la transcription *in vitro* sont mis en contact avec environ 200 unités de DNase I et placés pendant 15 min à 37°C.

7.2.2. Formation de la gamme d'ADNc

Le même protocole que celui décrit précédemment est utilisé en utilisant les amorces $oligo(dT)_{20}$ pour former les ADNc (cf. paragraphe 7.1 – Figure MM2).

7.3. La réaction de qPCR

Les réactions de qPCR ont été effectuées en utilisant le kit MESA GREEN qPCR MasterMix Plus for SYBRTM Assa w/ fluorescein de Eurogentec. Il est composé d'un tampon x2 qui comporte des dNTPs, 4 mM de MgCl2, une Meteor Taq DNA polymerase, du SYBR® Green I et de la fluorescéine. Le SYBR® Green I 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. Ainsi, le nombre de molécules d'ADNc de départ est quantifié en se basant sur les gammes étalons d'ADNc.



<u>Figure MM2</u>: Schéma récapitulatif des étapes nécessaires à l'obtention d'une gamme d'ADNc utilisable pour les qRT-PCR en 2 étapes.

Les qPCR sont réalisées dans in lightcycler MyiQ Single-Color Real-Time PCR Detection System commercialisé par BIORAD. Chaque réaction est effectuée dans une microplaque de 96 puits avec un volume final de 25 μ l. Chaque réaction est composée de 12,5 μ l de tampon x2, de 5 pmoles de chaque amorce, d'une quantité x d'ADNc et de l'eau qsp 25 μ l. La quantité d'ADNc des gammes étalons varie de 1 ng à 10⁻⁵ ng, alors que pour déterminer précisément la quantité d'un ADNc particulier, 100 ng et 10 ng d'ADNc totaux sont utilisés.

Les étapes de la qPCR sont, comme pour la PCR, un enchaînement de cycles répétant les étapes dénaturation-hybridation-élongation. Le traitement des résultats est réalisé grâce au MyiQ System Software.

8. Détermination des paramètres de croissance cellulaire

8.1. Méthode automatique

Un clone isolé est mis en préculture dans 10 ml de milieu minimum et incubé pendant environ 24 h. La lecture de l'absorbance à 600 nm permet de calculer la dilution nécessaire pour établir une nouvelle culture cellulaire où la densité optique est exactement égale à 0,02.

Cent cinquante µl de cette culture sont alors prélevés et placés dans une microplaque de 96 puits qui est installée dans un lecteur de microplaques de type TECAN INFINITE F200. Ce lecteur permet la croissance cellulaire en agitant la plaque à environ 120 rpm et à 30 °C. Pour chaque puits, une mesure d'absorbance est réalisée toutes les 10 minutes. Les résultats obtenus sont traités à l'aide d'un logiciel développé au laboratoire par Benoît Kammerer. Ce logiciel détermine trois paramètres : le temps de génération, la vitesse spécifique de croissance et le temps de la phase de latence pendant lequel la cellule s'adapte à son environnement (Figure MM3). Un autre paramètre de la croissance cellulaire peut être défini à partir de ces courbes de croissance : la biomasse formée (Figure MM3). Ce dernier n'est pas utilisé dans notre étude.

8.2. Méthode manuelle

Lors de cette étude, le milieu utilisé est le milieu YNB contenant du glucose à une concentration de 5 g/l, concentration à laquelle les dosages enzymatiques sont plus précis.

8.2.1. Croissance cellulaire

Comme pour la méthode automatique, un clone isolé est mis en préculture dans 10 ml de milieu minimum et incubé pendant environ 24 h. Cette préculture est utilisée pour ensemencer précisément 50 ml de ce même milieu YNB. Le suivi de croissance cellulaire se fait après 12 h d'incubation à 30 °C et à 130 rpm. À ce temps, puis toutes les 2 h, des tubes aliquots sont prélevés afin de mesurer l'absorbance de la culture à 600 nm et ils sont également utilisés pour déterminer la consommation spécifique du glucose et la production d'éthanol. Le temps spécifique de croissance μ est déterminé en phase exponentielle de croissance et le temps de génération g est déterminé selon l'équation $g = \ln 2/\mu$.



<u>Figure MM3 :</u> Exemple d'une courbe de croissance obtenue par la méthode automatique. Ce graphique représente les absorbances en coordonnées semi-logarithmiques (axe des ordonnées) en fonction du temps (axe des abscisses). Divers paramètres quantitatifs sont obtenus à partir de cette méthode : le temps de la phase de latence (1), le temps de génération et la vitesse spécifique de croissance (2), ainsi que la biomasse formée (3).

8.2.2. Consommation spécifique de glucose

Le dosage du glucose en solution est réalisé par une double réaction enzymatique en utilisant le kit D-Glucose de R-BIOPHARM. La première réaction est la phosphorylation du glucose en glucose-6-phosphate (G6P) en présence de l'hexokinase (HK) et d'ATP.

En présence de l'enzyme glucose-6-phosphate-déshydrogénase (G6P-DH), le G6P est oxydé par le NADP⁺ en gluconate-6-phosphate. Le cofacteur est lui réduit en NADPH + H^+ .



En présence d'un large excès de NADP⁺, les quantités de glucose et de cofacteurs réduits sont stoechiométriquement égaux. Le dosage du glucose repose donc sur la formation de NADPH + H^+ , produit dosable par spectrophotométrie à 340 nm.

Chaque solution aliquote est diluée au 1/100 et 50 µl de celle-ci sont mis en solution dans 950 µl d'eau et 500 µl de solution 1 (tampon triéthanolamine, pH 7,6, 110 mg de NADP, 260 mg d'ATP). Une première lecture d'absorbance à 340 nm est effectuée après 5 min à température ambiante. Afin que les réactions enzymatiques soient réalisées, 10 µl de solution 2 (320 U de HK, 160 U de G6P-DH) sont ajoutés. Après 15 min à température ambiante, une seconde lecture à 340 nm est effectuée. En parallèle des échantillons, un dosage sans glucose et un autre avec 25 µg de solution de glucose sont effectués, respectivement pour faire le zéro d'absorbance et pour normaliser les valeurs de concentration de chaque échantillon.

Pour déterminer la concentration de glucose dans chaque échantillon, la formule utilisée est : $c = (V \times PM \times dA) \div (E \times d \times v \times 1000)$

avec : c = concentration de glucose en g/l V = volume final de la réaction (1510 µl) PM = poids moléculaire du glucose (180,16 g/mol) dA = différence d'absorbance avant et après les réactions enzymatiques E = coefficient d'extinction molaire du NADPH + H+ (6,3 l/mmol/cm) d = distance du trajet optique (1 cm)v = volume de l'échantillon (50 µl)

soit, $c = 0,8636 \times dA$

Pour obtenir la consommation spécifique de glucose Q_{Glu} , on divise le temps spécifique de croissance μ par la concentration c ainsi obtenue.

8.2.3. Production spécifique d'éthanol

Comme pour le glucose, le dosage de l'éthanol repose sur une double réaction enzymatique en utilisant le kit Ethanol de R-BIOPHARM. La première est l'oxydation d'éthanol en acétaldéhyde par une enzyme de type alcool déshydrogénase (ADH) en présence de NAD⁺.

La seconde réaction est l'oxydation d'acétaldéhyde en acide acétique par l'aldéhyde déshydrogénase (Al-DH).

Acétaldéhyde + NAD^+ \longrightarrow Acide acétique + $NADH + H^+$

En présence d'un large excès de NAD⁺, les quantités d'éthanol et de NADH + H⁺ sont stoechiométriquement égales. Le dosage d'éthanol repose donc sur la formation de NADH + H⁺, composé dosable à 340 nm.

Chaque aliquot est dilué au 1/100 et 50 μ l de celle-ci sont ajoutés à 1,5 ml de solution 1 (tampon diphosphate de potassium pH 9). Une première lecture de l'absorbance à 340 nm est réalisée après 3 min à température ambiante. La seconde réaction enzymatique débute par l'ajout de 25 μ l de solution 2 (possédant environ 100 unités d'ADH, 0,4 unités d'Al-DH et approximativement 2 mg de NAD⁺). Une deuxième prise d'absorbance à 340 nm est réalisée après 5 min à température ambiante. En parallèle des échantillons, un dosage sans éthanol et un autre avec 3 μ g de solution d'éthanol sont effectués, respectivement pour faire le zéro d'absorbance et pour normaliser les valeurs de concentration de chaque échantillon.

Pour déterminer la concentration d'éthanol dans chaque échantillon, la formule utilisée est : $c = (V \times PM \times dA) \div (E \times d \times v \times 1000 \times 2)$

avec : c = concentration de glucose en g/l V = volume final de la réaction (1575 µl) PM = poids moléculaire de l'éthanol (46,07 g/mol) dA = différence d'absorbance avant et après les réactions enzymatiques E = coefficient d'extinction molaire du NADH + H+ (6,3 l/mmol/cm) d = distance du trajet optique (1 cm)v = volume de l'échantillon (50 µl)

soit, $c = 0,1152 \times dA$

Pour obtenir la production spécifique d'éthanol π_{EtOh} , le temps spécifique de croissance μ est divisé par la concentration c ainsi obtenue.

8.2.4. Détermination du poids sec

Un clone isolé est mis en suspension dans 200 ml de milieu YNB. Les cellules sont arrêtées en phase exponentielle de croissance ($0.3 < DO_{600nm} < 0.5$). 180 ml de culture sont centrifugés 5 min à 7000 rpm puis le culot cellulaire est lavé avec 20 ml d'eau. Le culot est repris dans 1 ml d'eau puis placé dans un tube Eppendorf de 1,5 ml. Les cellules sont centrifugées 5 min à 13200 rpm puis séchées pendant 2 jours à 65°C. Le poids sec des cellules de chaque souche en phase exponentielle de croissance peut alors être mesuré.

9. Test d'activité enzymatique ATCase

Ce test repose sur la capacité qu'a l'ATCase à transformer l'aspartate en carbamylaspartate (ou acide uréidosuccinique USA) en présence de carbamylphosphate.



9.1. Préparation de l'extrait cellulaire

10 ml de cellules en phase exponentielle de croissance, cultivées en milieu minimum ou en milieu complet, sont centrifugés 10 min à 7000 rpm puis les cellules sont reprises dans 300 μ l de tampon phosphate de sodium pH 7,5 100 mM. L'équivalent de 100 μ l de billes de verre est ajouté au mélange qui est vortexé pendant 3 min puis centrifugé 5 min à 12400 rpm. Le surnageant contenant les protéines est récupéré et est stocké à -20°C jusqu'à son utilisation lors du test enzymatique.

9.2. Dosage des protéines

Les protéines ont été dosées par la technique de Bradford (Bradford, 1976). De 5 à 20 μ l d'extrait cellulaire sont mélangés à 200 μ l de réactif et d'eau qsp 1 ml. Parallèlement, une gamme de BSA de 2,5 μ g à 20 μ g est réalisée. Après 5 min à température ambiante, les absorbances sont lues à 595 nm.

9.3. Test enzymatique (Prescott et Jones, 1969)

De 0 à 20 µl d'extrait cellulaire est ajouté à un mélange réactionnel contenant 5 µmoles d'aspartate, 5 µmoles de carbamylphosphate et de l'eau qsp 1 ml. Parallèlement, le même test est réalisé sans aspartate ainsi qu'une gamme de 0 à 135 nmoles d'USA. Les tubes sont incubés dans un bain-marie à 28°C pour une durée de 30 min à 1 h et la réaction est stoppée par l'ajout d'1 ml de milieu de coloration. Ce dernier est composé de 2 volumes de solutions d'antipyrine (5 mg/ml dans de l'acide sulfurique 50%) et d'un volume de diacétylmonoxime (8 mg/ml dans de l'acide acétique 5%). En présence d'un acide fort, l'antipyrine se lie à l'USA pour donner un composé de couleur jaune-orangée ayant un maximum d'absorbance à 466 nm. Les tubes sont bouchés à l'aide d'une bille de verre pour éviter l'évaporation, placés une nuit à 4°C à l'obscurité puis incubés 75 min à 60°C. Ils sont ensuite remis à température ambiante et leur absorbance est lue à 466 nm.

Afin de déterminer l'activité enzymatique de l'ATCase, il faut déterminer l'absorbance de l'USA en utilisant la relation :

 A_{466} USA = (ES - E) - (S - ESO)

ES = Valeur du tube essai E = Valeur du tube essai sans aspartate S = Valeur du tube sans extrait cellulaire ESO = Valeur du tube sans extrait cellulaire sans aspartate

Les valeurs de DO_{466nm} sont reportées sur la gamme étalon d'USA et permettent de déterminer la quantité d'USA formée pendant le temps de réaction en présence de l'extrait cellulaire et de déterminer l'activité spécifique de l'ATCase :

AS : Quantité d'USA formé (nmoles) par min et par mg de protéines totales

10. Transformation des cellules de levures

10.1. La méthode d'électroporation

10.1.1. Préparation des cellules compétentes

Les cellules sont cultivées dans 40 ml de milieu complet jusqu'en fin de phase exponentielle 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 puis 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.

10.1.2. Électroporation

Cinquante μ l de cellules compétentes sont déposés dans une cuve d'électroporation auxquels sont ajoutés 1 μ g d'ADN nécessaire pour la transformation. Les cellules sont ensuite soumises à un choc électrique (1600 V, 10 μ F, 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 45 min à 30°C. Suite à cette incubation, les cellules sont brièvement centrifugées puis reprises dans 200 μ l de milieu YPD Sorbitol 1 M et 100 μ l sont étalés sur boîte de milieu sélectif.

10.2. Transformation chimique par l'acétate de lithium

Environ 10^7 cellules, 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 (comme par exemple de l'ADN de sperme de saumon dénaturé) et 1 ml de PEG 3350 40%. 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.

11.Bioinformatique

11.1. Les banques de données

Les banques de données utilisées lors de ce travail sont :

- Saccharomyces Genome Database (SGD) <u>http://www.yeastgenome.org/</u>
- Génolevures <u>http://cbi.labri.fr/Genolevures/</u>
- Genbank <u>http://www.ncbi.nlm.nih.gov/</u>

11.2. Annotation des génomes

Les gènes codant les protéines, ainsi que ceux codant les ARN ribosomiques sont annotés à l'aide des programmes BLASTX et BLASTN (Altschul et *al.*, 1997). Les gènes codant les ARN de transfert sont annotés grâce aux programmes tRNAscan-SE (Lowe and Eddy, 1997) et ARWEN (Laslett and Canbäck, 2008).

11.3. Construction des arbres phylogénétiques

L'analyse et les comparaisons des séquences nucléotidiques sont réalisées à l'aide du logiciel DNA StriderTM 1.3. Les alignements des séquences protéiques et des nucléotides sont réalisés respectivement à l'aide des logiciels MUSCLE (Edgar, 2004) et TRANalign, disponible sur le package EMBOSS (<u>http://bips.u-strasbg.fr/EMBOSS/</u>). Les séquences de nucléotides correspondant aux phases ouvertes de lecture des gènes sont concaténées puis la construction des arbres phylogénétiques est réalisée à l'aide du programme SEAVIEW (Gouy et *al.*, 2010).

11.4. Traitement des données issues des hybridations sur puces à ADN

Les expériences d'hybridations des ADNc sur des puces à ADN ont été réalisées à l'ENS de Paris. Les résultats obtenus ont été analysés en utilisant les programmes Java TreeView (Saldanha, 2004) et MeV (Saeed et *al.*, 2003).

Partie 1

Chapitre I

Présentation du système de sélection

Chapitre II

Impacts des réarrangements chromosomiques

<u>Chapitre III</u> Impacts des gènes de fusion

Conclusions et perspectives
CHAPITRE I

PRÉSENTATION DU SYSTÈME DE SÉLECTION

Présentation du système de sélection

Introduction

Un crible de sélection basé sur un allèle mutant du gène *URA2* chez *S. cerevisiae* a été utilisé au laboratoire et a permis ces 15 dernières années d'isoler et de caractériser plus de 400 souches de *S. cerevisiae* porteuses de remaniements chromosomiques variés comme des insertions, des délétions, et des duplications. Ces réarrangements peuvent engendrer la formation de gènes chimères où le gène *URA2* est fusionné à un autre gène. Dans ce travail, nous avons constitué une collection de souches en retenant un ensemble de révertants haploïdes et diploïdes, ayant des réarrangements chromosomiques variés, dans le but de déterminer les conséquences de ces différents événements génomiques sur la réponse cellulaire. Nous avons également choisi ces souches en fonction du gène de fusion qu'elles contiennent afin de déterminer les propriétés de ces gènes et la conséquence de leur présence sur la physiologie de la cellule. Dans ce premier chapitre, nous allons d'abord rappeler le principe de ce crible, décrire la stratégie générale d'identification des réarrangements chromosomiques et présenter succinctement les types de remaniements isolés. Enfin, nous décrierons les différentes souches que nous avons retenues et présenteront les raisons qui nous ont conduit à les choisir pour la suite du travail.

1. Le système de sélection

Le système de sélection utilisé au laboratoire repose sur un allèle mutant du gène *URA2* chez la souche *S. cerevisiae* FL100 (Exinger et Lacroute, 1979 ; Welcker et *al.*, 2000). Situé sur le chromosome X, ce gène code une protéine multifonctionnelle qui comporte deux domaines fonctionnels, un domaine glutamine amido-transférase - carbamylphosphate synthétase (GATase - CPSase) et un domaine aspartate transcarbamylase (ATCase), ainsi q'un domaine cryptique DHOase-like. Cette protéine intervient dans les deux premières étapes de la voie de biosynthèse des nucléotides pyrimidiques (Figure 1.1). L'allèle muté $ura2_{15-30-72}$ contient trois mutations dans les parties codant les domaines GATase et CPSase : deux mutations non-sens et une délétion d'une guanine engendrant un décalage du cadre de lecture (respectivement aux positions 262, 916 et 2972 par rapport à la position de l'ATG initiateur). Ces mutations n'interfèrent pas avec la transcription du gène *URA2*, mais conduisent à un arrêt prématuré de la traduction de la protéine Ura2p. Le gène muté engendre ainsi une auxotrophie pour l'uracile résultant d'une interruption de la voie de biosynthèse des nucléotides pyrimidiques. Néanmoins, ce phénotype d'auxotrophie est uniquement dû à la perte de fonction du domaine ATCase car les activités GATase - CPSase sont complémentées par les protéines Cpa1p et Cpa2p intervenant dans la voie de biosynthèse de l'arginine (Figure 1.1). Une réactivation du domaine ATCase est donc nécessaire et suffisante pour un retour à la prototrophie vis-à-vis de l'uracile.

La réversion d'une mutation ponctuelle intervient une fois pour un million de cellules générées (fréquence de l'ordre de 10^{-6}). Pour un retour à la prototrophie, la réversion des trois mutations du gène *ura2*₁₅₋₃₀₋₇₂ est possible, mais la fréquence de cette triple réversion est trop faible (10^{-18}) comparée au nombre de cellules utilisées dans la sélection (10^{12} cellules environ). Les souches protorophes, appelées révertants, sont sélectionnées avec une fréquence d'apparition de l'ordre de 10^{-10} ou 10^{-11} .

2. Les outils de caractérisation

Ce système de sélection a permis d'isoler ces dernières années plus de 400 révertants dans les contextes haploïdes et diploïdes. La caractérisation des remaniements chromosomiques a été réalisée sur chacun de ses révertants grâce à plusieurs approches de biologie moléculaire qui sont résumées ci-dessous.



<u>Figure 1.1</u>: Le gène *URA2*, la protéine Ura2p et la voie de biosynthèse de l'UTP chez *S. cerevisiae*. Le gène *URA2* code une protéine multifonctionnelle intervenant dans les deux premières étapes de la voie de biosynthèse des nucléotides pyrimidiques. Les positions +262 et +916 correspondent aux mutations non-sens.

La position +2972 correspond à position de la délétion d'une guanine.

2.1. Hybridations de type ADN-ADN par la méthode de Southern (Southern, 1975)

Au niveau du gène URA2, des sites de coupures des enzymes de restriction BamHI, EcoRI ou BglII peuvent être utilisées conduisant à des profils de restriction facilement analysables. Le profil obtenu suite à la coupure par BamHI est celui qui est systématiquement analysé pour une première identification des révertants (Figure 1.2). Ce profil est constitué d'une bande unique de 6,8 kb lorsque l'on hybride une sonde nucléotidique correspondant au domaine ATCase (Roelants et al., 1995). Pour les révertants, des profils de restriction BamHI variables ont été mis en évidence, tous différents du profil unique de 6,8 kb, certains constitués d'une bande supplémentaire correspondant à la région codant l'ATCase. Ce profil suggère un événement de duplication. Les autres profils obtenus sont compatibles avec une insertion de rétrotransposon Ty1 ou avec des délétions (Roelants et al., 1995) L'analyse de ces différents profils est la première étape pour caractériser les différents réarrangements chromosomiques.





2.2. Caryotypes électrophorétiques par électrophorèses en champs alternés

La caractérisation du chromosome porteur du réarrangement chromosomique a été déterminée en réalisant un caryotype électrophorétique grâce à une électrophorèse en champs alternés qui permet de séparer les chromosomes en fonction de leur taille, suivie de l'hybridation de type ADN-ADN d'une sonde correspondant à la région codant l'ATCase. Les profils obtenus montrent des modifications dans le nombre ou la taille des chromosomes. Dans le cas de duplications, la région codant l'ATCase peut être retrouvée au niveau du chromosome X mais également sur d'autres chromosomes démontrant des duplications

interchromosomiques et intrachromosomiques (Schacherer et *al.*, 2005 ; Schacherer et *al.*, 2007). Cette approche est essentiellement utilisée pour préciser la localisation des duplications.

2.3. Caractérisation fine des réarrangements par PCR

Les insertions de rétrotransposons Ty1 ont été cartographiées par des réactions de PCR en utilisant une amorce spécifique de l'ATCase et une amorce située dans la séquence LTR du rétrotransposon Ty1. De manière identique, des PCR ont été utilisées pour tester les événements de délétion. Des marches chromosomiques ont été effectuées en utilisant des oligonucléotides spécifiques des gènes situés en amont sur le chromosome X et d'autres spécifiques de la région ATCase.

2.4. Hybridations sur puces à ADN

L'hybridation comparative des ADN des révertants et des souches sauvages sur puces à ADN Affymetrix a permis une caractérisation plus précise des réarrangements chromosomiques de type duplication et délétion en déterminant les bornes des segments dupliqués ou délétés et en mettant en évidence les gènes fusionnés à la région codant l'ATCase (Schacherer et *al.*, 2007). Des réactions de PCR et de séquençage ont complété ces expériences d'hybridations pour vérifier la structure des gènes de fusion néoformés et déterminer leur jonction au niveau nucléotidique.

3. Les types de réarrangements chromosomiques sélectionnés

Ce crible génétique a été utilisé pour isoler et caractériser des réarrangements chromosomiques au sein des souches de *S. cerevisiae* FL100 et S288c. La caractérisation des réarrangements chromosomiques à la base de la réactivation de l'ATCase chez FL100 a été réalisée en combinant les divers outils moléculaires que nous venons de résumer. Des événements d'insertions de rétrotransposons de type Ty1, des délétions ainsi que des duplications ont été caractérisés. (Roelants et *al.*, 1995 ; Welcker et *al.*, 2000 ; Schacherer et *al.*, 2004 ; Schacherer et *al.*, 2005 ; Schacherer et *al.*, 2007 ; Tourrette et *al.*, 2007).

Pour les révertants de la souche S288c, des réarrangements chromosomiques de type délétion et duplication ont été isolés, mais aucune insertion de rétrotransposons n'a été décrite dans les conditions standards (Fritsch et *al.*, 2009).

3.1. Les insertions de rétrotransposons

Les insertions de rétrotransposons de type Ty1 ont été isolées à partir de cellules haploïdes uniquement. Cette différence entre haploïdes et diploïdes résulte d'une forte transcription des rétrotransposons Ty1 dans le premier contexte alors qu'elle est fortement diminuée dans le second limitant ainsi la réverse transcription de l'ARN messager en ADN puis son insertion dans le génome (Paquin et Williamson, 1984). Ces insertions représentent la classe majeure des réarrangements chromosomiques obtenus à partir de notre système de sélection dans la souche FL100. En effet, 66% des révertants sélectionnés possèdent des insertions de rétrotransposons de type Ty1 (Roelants et *al.*, 1995). En revanche, dans la souche S288c, aucune insertion de rétrotransposons n'a pu être sélectionnée dans des conditions similaires (Fritsch et al, 2009).

Le taux de rétroposition augmente lorsque la température est inférieure à la température optimale de croissance (Paquin et Williamson, 1984). Cette propriété a été vérifiée pour notre crible. La sélection de révertants à 25°C a conduit à sélectionner plus d'insertions de Ty1 dans la souche FL100. De même, en utilisant la souche S288c dans ces conditions de température, des insertions de rétrotransposons Ty1 réactivant la région codant l'ATCase ont pu être sélectionner (Fritsch et *al.*, 2009).

Ces éléments mobiles sont retrouvés insérés au sein du gène *URA2* en aval des 3 mutations, en orientation inverse par rapport à *URA2* (orientation antisens) (Figure 1.3A). Cette insertion permet le retour à la prototrophie vis-à-vis de l'uracile car la séquence LTR des rétrotransposons Ty1, appelée δ , initie la transcription de la région ATCase et entraîne ainsi la synthèse de cette enzyme (Roelants et *al.*, 1997).

Dans le contexte S288c, des révertants sélectionnés à partir de l'allèle *ura2*₁₅₋₃₀, ne contenant que les deux mutations non-sens, ont permis d'isoler des insertions de rétrotransposons Ty1. Curieusement dans cette souche, les rétrotransposons insérés peuvent être retrouvés dans la même orientation que le gène *URA2* (orientation sens) en aval de la dernière mutation (Figure 1.3B). Nous avons réalisé une marche chromosomique par RT-PCR en utilisant une amorce dans la région ATCase et une autre au niveau de différentes régions de cet élément mobile. Une amplification est obtenue lorsque l'amorce est située dans le LTR ou au niveau de la jonction entre la région ATCase et le rétrotransposon. Par contre, la RT-PCR effectuée avec une amorce située dans la partie TYB montre une absence

d'amplification (Figure 1.4). Ces résultats indiquent que dans les révertants possédant une insertion de rétrotransposons en orientation sens, l'initiation de la transcription de la région codant l'ATCase se fait dans la partie LTR des rétrotransposons (Bleykasten-Grosshans et *al.*, soumis). Comme mentionné précédemment, la séquence LTR permet également l'initiation de la transcription dans les révertants possédant une insertion de rétrotransposons Ty1 en orientation antisens (Roelants et *al.*, 1997). La mise en commun de nos résultats avec ceux de Roelants et *al.* (1997) permet d'affirmer que la transcription de l'ATCase dans ces révertants débute dans les régions LTR des rétrotransposons indépendamment de leur orientation.



Figure 1.3 : Réacticvation de l'ATCase par insertion de rétrotransposns Ty1.

- A) Insertion en orientation antisens.
- B) Insertion en orientation sens

La flèche indique le point d'initiation de la transcription de la région ATCase.

3.2. Les délétions

Les événements de délétion ont été sélectionnés dans 17 % des révertants, qu'ils soient haploïdes ou diploïdes. La taille des segments délétés est variable et conduit à l'obtention de deux classes de délétion (Welcker et *al.*, 2000 ; Tourrette et *al.*, 2007). La première est décrite comme génique car les délétions sont internes au gène *URA2* et englobent au moins les 3 mutations. La région ATCase conservée, reste sous la dépendance du promoteur *URA2*. Le second type d'événement est une délétion segmentale où la partie délétée comprend les trois

mutations du gène $ura2_{15-30-72}$ ainsi que d'autres gènes situés en amont sur le chromosome X. Ainsi, dans ces souches, la partie ATCase peut être fusionnée en phase avec une séquence codante ou à un promoteur d'un autre gène (Figure 1.5).



<u>Figure 1.4</u> : Détermination du site d'initiation de la transcription dans les révertants possédant une insertion de rétrotransposon Ty1 en orientation sens.

A) RT-PCR avec une amorce située au niveau de la jonction rétrotransposon-ATCase

B) RT-PCR avec une amorce située au niveau de la région LTR

C) RT-PCR avec une amorce située au niveau de la partie TYB

La ligne en pointillée représente une absence de fragment de PCR alors que les lignes pleines indiquent des fragments amplifiés.

3.3. Les duplications

Ces réarrangements ont été obtenus dans 17% des révertants. En fonction de la ploïdie de la souche $ura2_{15-30-72}$, différents types de duplications ont été caractérisés : des duplications géniques ont été sélectionnées à l'état haploïde alors que des aneuploïdies et des translocations non réciproques l'ont été en contexte diploïde (Schacherer et *al.*, 2004; Schacherer et *al.*, 2007). En revanche, des duplications segmentales ont été retrouvées dans les deux contextes. Au niveau de toutes ces duplications, la région ATCase est dupliquée et fusionnée avec un gène accepteur, pouvant être situé sur le chromosome X ou non.



<u>Figure 1.5</u>: Exemple de délétion segmentale. Une partie du gène muté $ura2_{15-30-72}$ est délétée ainsi que d'autres gènes situés en amont sur le chromosome X. Dans cet exemple, la région ATCase est fusionnée avec le gène *PBS2* situé au locus YJL128c.

3.3.1. Les duplications géniques

Dans une duplication génique, seule la région ATCase est dupliquée. La caractérisation par électrophorèse en champs alternés et hybridation des chromosomes de ces révertants a démontré que le segment dupliqué pouvait s'insérer au niveau de tous les chromosomes (Schacherer et *al.*, 2004). La caractérisation par PCR a permis de mettre en évidence dans chaque cas, la fusion d'un rétrotransposon Ty1 au niveau de la séquence TYA avec la partie ATCase du gène *URA2* (Figure 1.6).

3.3.2. Les duplications segmentales

Pour les duplications segmentales, la région ATCase ainsi que plusieurs gènes contigus situés en aval du gène *URA2* sur le chromosome X sont dupliqués et fusionnés en tandem (Figure 1.7B). La taille de la région dupliquée s'étend de 6 à 90 kb selon les révertants (Schacherer et *al.*, 2005 ; Schacherer et *al.*, 2007).



<u>Figure 1.6</u>: Exemple de duplication génique. La partie ATCase dupliquée est fusionnée avec la région TYA des rétrotransposons Ty1. Ce gène de fusion peut alors s'insérer sur les chromosomes. Les lettres A, B, C, D, E, F représentent schématiquement une série de gènes contigus situés sur le chromosome IV.

3.3.3. Les aneuploïdies et les translocations

La réactivation de l'ATCase peut également survenir suite à des translocations réciproques et non réciproques entre 2 chromosomes. Au niveau des translocations réciproques, un échange de fragments de chromosomes s'opère par exemple entre le chromosome X et le chromosome II (Tourrette et al., 2007). Dans le cas d'aneuploïdies et de translocations non réciproques, une partie du chromosome X, s'étendant de l'ATCase jusqu'au télomère, est dupliquée et fusionnée avec un autre chromosome pour former un chromosome chimère (Schacherer et al., 2007). Pour les aneuploïdies, une partie d'un autre chromosome comprenant notamment un télomère et un centromère est également dupliquée ; la fusion des deux parties de chromosomes dupliquées conduit à l'obtention d'un chromosome surnuméraire (Figure 1.7C). Dans le cas des translocations non réciproques, le chromosome accepteur de la partie dupliquée du chromosome X est délété de plusieurs gènes et d'un télomère (Figure 1.7D).



Figure 1.7 : Les duplications caractérisées à l'état diploïde.

A) Les chromosomes X et XIII dans une souche sauvage

- B) Exemple de duplication segmentale
- C) Exemple d'aneuploïdie

D) Exemple de translocation non réciproque

L'encadré bleu représente la duplication de gènes contigus de l'ATCase sur le chromosome X. Chr : chromosome.

3.4. Les mécanismes d'apparition des réarrangements chromosomiques

Les rétrotransposons Ty1 sont des éléments mobiles dans le génome grâce au mécanisme de rétroposition (Figure 1.8). Après la transciption de ces éléments, leur ARNm est transporté dans le cytoplasme où les séquences TYA et TYA-TYB sont traduites. Le partie TYA code pour une protéine de la nucléocapside, dont l'assemblage forme les particules pseudo-virales (VLPs), alors que la partie TYA-TYB code pour une protéine possédant plusieurs activités dont la réverse-transcriptase et l'intégrase. Après son export dans le cytoplasme, l'ARNm de ces éléments mobiles est empaqueté au sein des VLPs où il est rétrotranscrit en ADN avant d'être inséré au sein du génome.



<u>Figure 1.8</u>: Duplication génique par rétroposition. Le fond gris délimite le noyau. L'ADN double brin est représenté par une flèche épaisse et son ARN par un rectangle mince.

Les travaux de Schacherer et *al.* (2004) ont montré que l'ATCase est fusionnée au niveau de la séquence TYA du rétrotransposon Ty1. De plus, une queue polyA est retrouvée au niveau de la région 3'UTR de l'ATCase. Cette structure suggère que le gène $ura2_{15-30-72}$ est transcrit. L'ARNm de ce gène est alors exporté hors du noyau et empaqueté dans des VLPs. Dans ces particules, ces messagers peuvent être rétrotranscrits et fusionnés avec la séquence

du rétrotransposon selon un mécanisme qui n'est pas totalement élucidé. À l'issue de ce mécanisme de rétroposition, l'ADN de fusion est importé dans le noyau et inséré au niveau de l'ADN (Schacherer et *al.*, 2004) (Figure 1.8).

La caractérisation des délétions, des duplications segmentales, des aneuploïdies et des translocations non réciproques montre dans tous les cas que la région codant l'ATCase est fusionnée avec un gène accepteur. Les gènes de fusion ainsi obtenus ont été caractérisés au niveau nucléotidique. La jonction entre le gène receveur et la région ATCase est constituée d'une séquence de taille variable pouvant aller jusqu'à 17 nucléotides initialement présente au niveau du gène récepteur et de la région codant l'ATCase. La présence de ces microhomologies indique que le mécanisme de Non-Homologous End Joining (NHEJ) est privilégié par rapport à celui de recombinaison homologue pour la mise en place de ces événements. Les travaux de Fritsch et *al.* (2009) ont permis de confirmer que le mécanisme de NHEJ était effectivement à la base de la formation de ces remaniements.

3.5. Les gènes de fusion associés aux réarrangements chromosomiques

Le retour à la prototrophie vis-à-vis de l'uracile est possible grâce aux divers réarrangements chromosomiques décrits ici, mais nécessite comme nous venons de le voir la formation de gènes chimères pour réactiver l'ATCase. Ces gènes résultent de la fusion de la région ATCase en phase avec une séquence codante d'un autre gène ou au niveau de son promoteur. Les protéines chimères codées par ces gènes possèdent toutes un domaine ATCase actif dans leur partie C-terminale alors que la partie N-terminale est constituée d'une région codée par le gène receveur dont la taille est variable.

4. Collection de révertants utilisée lors de cette étude

Nous avons retenu une collection de révertants sur différents critères dans le but de comprendre les impacts des réarrangements chromosomiques et des gènes de fusion sur la cellule. Pour ce travail, nous avons choisi une collection de révertants ayant divers types de remaniements : des délétions segmentales, des duplications géniques et segmentales, ainsi que des aneuploïdies et des translocations non réciproques. De plus, nous avons également pris en compte les caractéristiques des gènes accepteurs de la fusion dans le choix des souches. Ces gènes codent des protéines qui sont essentielles ou non pour la survie cellulaire et dont les fonctions sont variées comme le montre le Tableau 1.1.

le gène	lation des v70p.	ıts thiols. s oxydatifs.		'n.	0 avec des 82p.	0 avec des 82p.	ion du 2- lors de la		/acuoles.	ribosomique	capside des oosons.	/oies de lue.	dans la lophase.		aploïde (n) ou chromosomes
Fonction de la protéine codée par accepteur	Protéine chaperone impliquée dans la régul fonctions des protéines Hsp90p et Hsp	Péroxyredoxine spécifique des groupemen Participe à la protection contre les dommage.	Pompe à proton H ⁺ -ATPase.	Facteur d'élongation de la traductio	Chaperone cytoplasmique de la famille Hsp9 fonctions identiques à la protéine Hsp	Chaperone cytoplasmique de la famille Hsp9 fonctions identiques à la protéine Hsp	Enolase II, enzyme qui catalyse la convers phosphoglycérate en phosphoénolpyruvate glycolyse.	Fonction inconnue.	Protéine membraniare localisée dans les v	Protéine faisant partie de la petite sous-unité 40S.	Protéine intervenant dans la formation de la particules pseudo-virales des rétrotransp	MAP kinase kinase intervenant dans les v signalisation suite à un choc osmotiq	Sous-unité du complexe RENT, impliqué régulation du nucléole et la sortie de la té	Protéine de fonction inconnue	mosomiques et ont été isolées dans les contextes ha eurs de la partie ATCase sont localisés sur différents ons variées.
Genre de gène	Non essentiel	Non essentiel	Essentiel	Essentiel	Non essentiel	Non essentiel	Essentiel	Non essentiel	Non essentiel	Non essentiel	Non essentiel	Non essentiel	Essentiel	Non essentiel	éarrangements chro nt. Les gènes accept protéines de fonctic
Gène receveur	YDJ1	AHP1	PMA1	YEF3	HSC82	HSC82	EN02	protéine hypothétique	SNA3	RPS22A	ТуА	PBS2	NET1	MTC1	issues de divers r oïde et inversemer gènes codent des
Locus accepteur (chromosome)	YNL064c (XIV)	YLR109w (XII)	YGL008c (VII)	YLR249w (XII)	YMR186w (XIII)	YMR186w (XIII)	YHR174w (VIII)	YJL133c-A (X)	YJL151c (X)	YJL190c (X)	Ty1	YJL128c (X)	YJL076w (X)	YJL123c (X)	és. Ces souches sont i le et construits en dipl e locus accepteur. Ces
Réarrangement chromosomique	Aneuploïdie	Aneuploïdie	Aneuploïdie	Aneuploïdie	Aneuploïdie	Translocation non réciproque	Translocation non réciproque	Duplication segmentale	Duplication segmentale	Duplication segmentale	Duplication génique	Délétion	Délétion	Délétion	on des révertants analys- ont été isolés en haploïd varenthèse dans la colonn
Ploïdie	2n et n	2n	2n et n	2n et n	2n et n	2n	2n	2n et n	n et 2n	n et 2n	с	n et 2n	2n	2n	<u>1</u> : Collectic n). Certains qués entre p
Souche	B038	B199	B259	B284	B066/1	B132	B326	B286	Rev27	Rev46	Rev25	AWR58	B283/1	BR15/2	<u>Tableau 1.</u> diploïde (2 et sont indi

Dans le but de comparer les conséquences d'un même réarrangement dans les deux contextes de ploïdie, nous avons construit des souches haploïdes et diploïdes portant ces remaniements pour tous les révertants où cela a été possible. Ainsi, les révertants haploïdes ont été croisés avec la souche α UL de génotype *ura2*_{15,30,72} *leu2* Δ 0, isogénique à la souche de référence FL100, pour obtenir une cellule diploïde hétérozygote pour le gène de fusion. Inversement, les souches haploïdes ont été obtenues par sporulation des révertants diploïdes puis par isolement des spores Ura+ (Tableau 1.1). Cependant, six révertants ne sont présents que dans un contexte : dans les cellules diploïdes qui portent des translocations non réciproques ou des délétions segmentales des gènes essentiels ont été délétés sur le chromosome porteur du gène de fusion. De ce fait, de telles cellules ne peuvent pas être isolées à l'état haploïde.

Malgré plusieurs essais et modifications du protocole, nous n'avons pas réussi à faire sporuler le révertant aneuploïde B199 expliquant que seul le contexte diploïde a pu être étudié pour cette souche. (Tableau 1.1). Du fait de la forte diminution de la transcription des rétrotransposons dans le contexte diploïde, le révertant Rev25 possédant une duplication génique n'est disponible qu'à l'état haploïde.

Conclusion

Le système de sélection basé sur l'allèle mutant $ura2_{15-30-72}$ a permis de sélectionner de nombreux révertants à partir de la souche de laboratoire FL100. Les réarrangements chromosomiques à la base de la réactivation de l'ATCase permettent l'apport au minimum d'une région promotrice entraînant la formation de gènes de fusion. Ces remaniements sont variés et dépendent de l'état de ploïdie. De plus, la nature des gènes accepteurs de la fusion est également importante car la fonction de la protéine codée par ces gènes peut être perdue : la fusion de la région ATCase dans un gène essentiel peut être létale à l'état haploïde alors qu'elle est viable dans un contexte diploïde.

L'objectif de ce projet a été de déterminer de manière précise l'influence des réarrangements chromosomiques sélectionnés notamment les aneuploïdies et les translocations non-réciproques ainsi que celle des gènes de fusion sur le fonctionnement

cellulaire. Cette approche a également eu pour optique d'apporter des éléments de réponse sur la présence de manière quasi-systématique de remaniements chromosomiques et de gènes chimères dans les cellules cancéreuses. Pour cela, le travail a été orienté dans deux axes distincts qui seront présentés dans les deux chapitres suivants. Le premier axe a consisté à définir les conséquences des différents remaniements sur la croissance cellulaire et l'expression de l'ensemble des gènes. Le second a reposé sur la détermination des propriétés des gènes chimères impliqués dans le retour à la prototrophie.

CHAPITRE II

IMPACTS DES RÉARRANGEMENTS CHROMOSOMIQUES

Impacts de réarrangements chromosomiques

Introduction

Les révertants qui constituent notre collection possèdent divers remaniements chromosomiques et sont disponibles pour la plupart, dans les deux contextes haploïde et diploïde (Tableau 2.1). Cette collection de souches a été utilisée dans le but de déterminer les impacts de ces divers réarrangements sur la morphologie des colonies et des cellules, mais aussi pour analyser leurs influences sur la croissance cellulaire. Nous avons également déterminé dans ces souches, l'influence de ces réarrangements sur le transcriptome. Le fait de posséder un même réarrangement dans les deux contextes de ploïdie permet également de déterminer le rôle de la ploïdie dans la réponse cellulaire suite à l'apparition de tels événements.

1. Caractérisation de la physiologie cellulaire des révertants

Des morphologies de colonies variables ont été observées lors de l'analyse des trois souches de laboratoire de *S*. *cerevisiae* S288c, SK1 et $\sum 1278b$ (Granek et Magwene, 2010). Diverses morphologies sont également observées pour une même souche en fonction de différentes sources de carbone et de leur concentration, ou en fonction de la ploïdie (Granek et Magwene, 2010). Depuis la fin des années 1950, il est admis que le taux de ploïdie influence proportionnellement le volume cellulaire chez *S. cerevisiae* (Mortimer, 1958). Cette augmentation de la taille de la cellule est issue d'un dérèglement du cycle cellulaire, notamment au niveau des points de contrôle entre les phases G1 et S. Des analyses plus récentes ont permis de mettre en évidence 17 gènes régulés de manière différente en fonction de la ploïdie. Parmi ces gènes, la baisse de l'expression des cyclines Cln1p et Pc11p explique le décalage du passage du point START du cycle cellulaire, et donc le maintien prolongé dans la phase G0 à la base de l'augmentation du volume intracellulaire (Galitski et *al.*, 1999). Un autre gène dont la transcription est réprimée en fonction de la ploïdie est le gène *FLO11* qui code une protéine de surface à la base de la croissance invasive et pseudo-hyphale (Galitski et *al.*, 1999; Gancedo, 2001). Plus le taux de ploïdie augmente, plus la transcription du gène

Souche	Ploïdie	Type de réarrangement	Souche	Ploïdie	Type de réarrangement
2nU+	2n	Sauvage	FL100	n	Sauvage
B038	2n	Aneuploïdie	B038 n	n	Aneuploïdie
B199	2n	Aneuploïdie	B259 n	n	Aneuploïdie
B259	2n	Aneuploïdie	B284 n	n	Aneuploïdie
B284	2n	Aneuploïdie	B066/1 n	n	Aneuploïdie
B066/1	2n	Aneuploïdie	B286 n	n	Duplication segmentale
B132	2n	Translocation non réciproque	Rev27	n	Duplication segmentale
B326	2n	Translocation non réciproque	Rev46	n	Duplication segmentale
B286	2n	Duplication segmentale	AWR58	n	Délétion
Rev27 2n	2n	Duplication segmentale	Rev25	n	Duplication génique
Rev46 2n	2n	Duplication segmentale			
AWR58 2n	2n	Délétion			
B283/1	2n	Délétion			
BR15/2	2n	Délétion			

FLO11 est faible permettant aux cellules de *S. cerevisiae* d'avoir une croissance non-invasive (Galitski et *al.*, 1999).

<u>Tableau 2.1</u>: Liste des révertants utilisés et les réarrangements chromosomiques associés. Ces souches sont isogéniques aux cellules sauvages diploïde 2nU+ et haploïde FL100, dans les contextes diploïde et haploïde respectivement.

1.1. Comparaison de la morphologie des colonies

Dans cette partie, nous avons analysé la morphologie des colonies des révertants, ainsi que des cellules sauvages nous servant de référence, sur des milieux gélosés contenant différentes sources de carbone (glucose, galactose, sorbitol, glycérol, éthanol, acide acétique) et nous les avons observées au macroscope Leica Z16 APO à un grossissement X10. La croissance de tous les révertants de la collection sur ces milieux ne montre aucune différence de prolifération notable avec les souches sauvages. La conservation de la croissance sur les milieux contenant du glycérol ou de l'éthanol, composés non-fermentables, montre que les révertants n'ont pas perdu la capacité de respirer.

1.1.1. Analyse de la croissance sur milieu complet

Le milieu complet YPD (2 % de glucose) est classiquement utilisé pour cultiver des souches de *S. cerevisiae*. Sur ce milieu, les cellules sauvages haploïde (FL100) et diploïde (2nU+) montrent deux phénotypes distincts : les colonies des cellules FL100 ont une forme irrégulière alors que celles de 2nU+ ont une forme circulaire et bombée (Figure 2.1 ; Tableau 2.2). De manière identique, des différences sont retrouvées au niveau des révertants haploïde et diploïde (Tableau 2.2). Par exemple, les révertants B284 et B284 n, qui possèdent le même chromosome surnuméraire, présentent des phénotypes différents : la colonie haploïde est circulaire, possédant un pic et un liseré plus sombre (forme en sombrero) alors que la seconde est également circulaire mais possède une texture « peau d'orange ». Les morphologies des révertants B038 et B038 n sont également différentes : à l'état diploïde, ce sont de petites colonies circulaires lisses alors qu'en contexte haploïde, les clones sont de plus grosse taille avec un aspect rugueux (Figure 2.1). Ces résultats indiquent que la morphologie des colonies des révertants est influencée par la ploïdie.

Lorsque l'on compare les révertants dans le même contexte de ploïdie, des variations morphologiques sont retrouvées les distinguant des cellules sauvages. Par exemple, à l'état haploïde, certaines colonies ont des formes circulaires avec un aspect lisse (AWR58) ou rugueux (B038 n) alors que la souche FL100 a une forme irrégulière (Tableau 2.2). Ces variations morphologiques entre les différents révertants montrent qu'aucun patron phénotypique ne peut être attribué pour un type de réarrangements chromosomiques (Tableau 2.2).

Un cas particulier est le révertant B066/1 n, qui est une souche issue de la méiose du révertant diploïde B066/1. L'isolement des colonies pour cette souche haploïde permet de mettre en évidence deux types de morphologies. Certaines sont circulaires et lisses alors que d'autres sont plus grosses et granuleuses (Figure 2.2). L'analyse phénotypique de ces deux types de colonies montre que les colonies lisses sont Ura⁻ alors que les colonies granuleuses sont Ura⁺ ce qui suggère une perte du chromosome surnuméraire lors des divisions cellulaires. Une analyse par électrophorèse en champs alternés de ces clones permettrait de mettre en évidence une telle perte. Il est important à noter ici que cette morphologie granuleuse est également retrouvée dans un milieu minimum YNB liquide. Dans ce milieu liquide, cette souche flocule fortement ce qui a rendu impossible la quantification de sa croissance cellulaire.



<u>Figure 2.1 :</u> Morphologie des colonies sur un milieu complet. Cette observation a été réalisée au macroscope Leica Z16 APO (grossissement X10). Colonne gauche : souches diploïdes Colonne droite : souches haploïdes

Morphologie	Souches haploïdes	Souches diploïdes
Circulaire		
bombée		2nU+ - B199 - B326 -
		Rev27 2n - B283/1
peau d'orange	B284 n	
sombrero		B284
lisse	AWR58 - B286 n -	B038 - B259 - B066/1 - B132 -
	Rev27 - Rev46	Rev46 2n - B286 - AWR58 2n
rugueux	B038 n - B259 n	
Irrégulière	FL100 - Rev25	BR15/2

<u>Tableau 2.2</u>: Variation de la morphologie des colonies dans un milieu complet YPD. Les lignes grises représentent le phénotype associé au contour des colonies.



<u>Figure 2.2</u>: Morphologie des colonies du révertant B066/1 n. Cette observation a été réalisée au macroscope Leica Z16 APO (grossissement X10). Deux phénotypes distincts sont retrouvés chez ce révertant (indiqués par des flèches), correspondant respectivement à de petites colonies circulaires et lisses Ura⁻, et à des colonies de taille supérieure et rugueuse Ura⁺.

1.1.2. Analyse de la croissance sur des milieux contenant d'autres sources de carbone

En présence de glucose, les souches de la collection de révertants présentent différentes morphologies de colonies. Cette variation phénotypique est encore plus marquée quand les cellules sont en présence d'autres sources de carbone comme le galactose, l'éthanol, le sorbitol, le glycérol ou l'acide acétique (Figure 2.3). Dans cette partie, nous nous sommes intéressés aux formes et textures des colonies des révertants. Les résultats obtenus sont décrits dans le tableau 2.3.

La présence d'autres sources de carbone dans le milieu conduit à des résultats similaires à ceux obtenus avec du glucose. Nous observons une différence de morphologie marquée entre les souches haploïdes et diploïdes (Figure 2.3 ; Tableau 2.3). De plus, des morphologies différentes sont observées pour un même type de remaniement dans le même contexte de ploïdie. Par exemple, les colonies des révertants aneuploïdes B038 et B066/1 en présence de glycérol ont des phénotypes différents : les colonies de la souche B038 ont une forme irrégulière avec des contours épais alors que les colonies de B066/1 sont circulaires et lisses (Figure 2.3 ; Tableau 2.3). Comme précédemment, ces résultats indiquent qu'aucun patron morphologique ne peut être défini pour un même réarrangement chromosomique (Tableau 2.3).

À partir de cette analyse, il apparaît clair que les remaniements chromosomiques influencent les morphologies des colonies. De plus, des phénotypes différents sont retrouvés en fonction de la ploïdie montrant que ce contexte génétique est important. Ces résultats sont comparables à ceux obtenus par Granek et Magwene (2010) pour différentes souches de laboratoire. Enfin, nous avons pu mettre en évidence qu'aucun patron de morphologie des colonies ne peut être défini pour un même type de réarrangement chromosomique suggérant que ce sont plutôt des gènes impliqués dans le remaniement et leur nature qui sont responsables de cette modification de morphologie .

•>		Glycérol	Galactose	Sorbitol	Ethanol	Acetate
A)	2nU+	۲				. O
	B199	0,0		• •, ⁴	•	
	B038					
	B259	۲			•	
	B284	200 B		6	୍ଚ୍ଚ୍ଚ	
	B066/1					
	B132		а в ф.		•	
	B326					÷.
	Rev27 2n	~		•	°.	
	Rev46 2n	9 ⁹ ⁹	*			
	B286			6	\$	o
	AWR58 2n	. 3 .			· • • •	
	B283/1			40 - 1200a 400 - 121 1 400 - 121 1		
	BR15-2	00		••••		



<u>Figure 2.3 :</u> Morphologie des colonies sur des milieux contenant diverses sources de carbone. Cette observation est réalisée au macroscope Leica Z16 APO avec un grossissement X10.

- A) Révertants diploïdes
- B) Révertants haploïdes

Souches	Glycérol			Galactose			Sorbitol		Éthanol				Acétate				
	EP	EPI	EC	CL	Ι	CL	СР	EP	CP	CEB	CG	CL	EG	CP	EP	CPL	EG
Diploïdes																	
2nU+	+				+			+			+				+		
B199			+		+			+				+			+		
B038		+				+			+			+				+	
B259	+					+		+					+			+	
B284	+				+			+					+			+	
B066/1				+		+			+			+				+	
B132				+	+				+			+				+	
B326				+	+				+					+		+	
Rev27 2n				+		+			+			+				+	
Rev46 2n				+		+			+			+				+	
B286			+			+		+					+				+
AWR58 2n				+		+				+		+				+	
B283/1		+				+		+			+					+	
BR15/2				+	+				+					+		+	
Haploïdes																	
FL100				+			+		+					+		+	
B038 n				+		+			+					+		+	
B259 n	+					+			+			+				+	
B284 n	+					+			+					+		+	
B066/1 n	+					+		+					+			+	
Rev27				+		+			+					+		+	
Rev46				+		+			+					+		+	
B286 n				+	+				+			+				+	
AWR58				+		+			+					+		+	
Rev25				+	+				+			+				+	

<u>Tableau 2.3 :</u> Tableau récapitulatif des morphologies des colonies en fonction de la source de carbone utilisée. Les souches sauvages sont surlignées en gris, les aneuploïdies en jaune, les translocations non réciproques en orange, les duplications segmentales en vert, la duplication génique en blanc et les délétions en bleu. En fonction de la morphologie des colonies, différentes classes de colonies ont été identifiées :

CEB : colonies bombées avec des contours épais périphériques

- CG : colonies granuleuses
- CL : colonies lisses

CP : petites colonies

CPL : petites colonies lisses

EC : colonies avec des contours épais concentriques

EG : colonies avec des contours épais granuleux

EP : colonies avec des contours épais périphériques

EPI : colonies avec des contours épais périphériques irréguliers

I : colonies irrégulières

+ : présence d'une colonie avec cette morphologie

1.2. Analyse de la morphologie cellulaire des révertants

L'observation au microscope optique (grossissement x1000) des révertants de notre collection a été réalisée en utilisant le rouge neutre, un colorant des vacuoles qui a donné un meilleur contraste lors de nos observations. Celles-ci montrent des variations morphologiques importantes entre les cellules. En effet, à l'état haploïde, la souche sauvage FL100 a une forme ellipsoïdale caractéristique de *S. cerevisiae*, alors que les révertants B038 n et B284 n ont une forme plus allongée. De plus, des agrégats de cellules sont présents chez les souches B038 n, B066/1 n et AWR58, alors que de tels aspects ne sont pas retrouvés chez FL100. De la même manière, les révertants diploïdes ont une forme beaucoup plus allongée que la cellule sauvage 2nU+ et peuvent également former des agrégats, même s'ils sont moins marqués qu'en contexte haploïde (Figure 2.4 ; Tableau 2.4).

Morphologie	Souches haploïdes	Souches diploïdes
Ellipsoïdale	FL100 - B066/1 - Rev27 - Rev25	2nU+ - B199 - B066/1 - B326 -
		Rev27 2n
Cellules allongées	B259 n - B284 n	B038 - B259 - B284 - B132 -
		B286
Cellules plus		
grosses	B038 n - Rev46 - B286 n	Rev46 2n - AWR58 2n - B283/1 -
Formation		BR15/2
d'agrégats	B038 n - B286 n - AWR58	AWR58 2n - BR15/2

<u>Tableau 2.4</u>: Tableau récapitulatif des morphologies des cellules de la collection de révertants. Les morphologies sont classées en fonction de la taille et de la forme des cellules ainsi qu'en fonction de la formation d'agrégats.

La présence de cellules ayant des formes plus allongées que les cellules sauvages suggère un accroissement du volume cellulaire. Une augmentation du volume cellulaire a été observée pour des souches de *S. cerevisiae* possédant des chromosomes surnuméraires ou des translocations non réciproques dans plusieurs études (Torres et *al.*, 2007; Nikitin et *al.*, 2008). Les auteurs expliquent cet accroissement par un décalage des passages entre les phases G1/S ou G2/M qui induisent des retards de croissance (Torres et *al.*, 2007; Nikitin et *al.*, 2008). Nous avons analysé la croissance cellulaire pour l'ensemble de notre collection de souches afin de déterminer si le cycle cellulaire était lié à cette variabilité des morphologies des cellules.



<u>Figure 2.4</u>: Morphologie des révertants. Cette observation a été réalisée au microscope optique LEICA DM4000B avec un grossissement X1000 après une coloration au rouge neutre.

Colonnes gauches : Souches diploïdes.

Colonnes droites : Souches haploïdes.

2. Croissance cellulaire

S. cerevisiae est un organisme fermentaire qui consomme des glucides comme le glucose pour produire de l'éthanol. En présence d'oxygène, les cellules de *S. cerevisiae* favorisent la fermentation alcoolique plutôt que la respiration, malgré une production moindre d'énergie : c'est l'effet Crabtree. Dans cette partie nous avons déterminé les temps de génération pour chaque révertant et la durée de la phase de latence. Nous nous sommes également intéressés à la consommation spécifique de glucose et la production spécifique d'éthanol des souches de notre collection.

2.1. Analyse de la croissance cellulaire

Le temps de génération (g) et la phase de latence (L) sont deux des paramètres quantitatifs les plus utilisés en microbiologie. Deux approches ont été utilisées pour les déterminer. La première méthode est semi-automatique et repose sur des mesures d'absorbance de cultures réalisées dans une microplaque de 96 puits par un lecteur de type TECAN INFINITE F200. Chaque puit de cette plaque, contenant 150 μ l de milieu minimum, est faiblement ensemencé (DO₆₀₀ = 0,002). La plaque est agitée à 30°C à 120 rpm pendant 36 heures et une mesure de l'absorbance est réalisée toutes les 10 minutes. Dans le petit volume de culture utilisé, de nombreux révertants floculent. Ce phénomène de floculation perturbe la mesure d'absorbance rendant les résultats des temps de génération impossibles à exploiter de façon reproductible. Cette approche permet néanmoins d'estimer le temps de la phase de latence, c'est-à-dire le temps que met une cellule pour s'adapter à son environnement et débuter sa croissance.

La seconde approche est manuelle. Les cultures sont réalisées dans un volume de 50 ml dans un milieu minimum. Un échantillon de ces cultures est prélevé toutes les 2 heures et sert à mesurer l'absorbance à 600 nm et à déterminer les consommations et productions spécifiques de glucose et d'éthanol. Cette approche a été utilisée pour calculer les vitesses spécifiques de croissance et les temps de génération pour tous les révertants. Seules les souches B066/1 n et Rev46 2n n'ont pas pu être testées du fait de leur forte floculation et de leur très lente croissance.

2.1.1. Détermination des temps de génération des révertants

Nous avons déterminé le temps de génération et la vitesse spécifique de croissance par la méthode manuelle pour l'ensemble des souches de la collection en prenant une valeur d'absorbance toutes les 2 heures. Les résultats obtenus sont la moyenne de 3 cultures indépendantes. La souche de référence diploïde 2nU+ se divise environ toutes les 90 minutes alors que ce temps est d'environ 150 minutes pour FL100. Un test de Student effectué sur les valeurs obtenues a été utilisé pour déterminer quelles valeurs sont statistiquement différentes de celles des souches sauvages, c'est-à-dire les résultats pour lesquels nous sommes certains qu'ils sont différents de ceux de FL100 et 2nU+.

Cette analyse montre que la vitesse de prolifération des révertants diploïdes est inférieure à celle de la souche sauvage 2nU+ excepté pour les révertants B199, B038 et B283/1. Cette augmentation du temps de génération est deux fois supérieure chez des souches comme B326 ou B066/1 (Figure 2.5).

À l'état haploïde, la différence des temps de génération est nettement moins marquée qu'en contexte diploïde. En effet, seuls les révertants Rev46, B286 n et AWR58 montrent une vitesse de croissance nettement inférieure à FL100 (Figure 2.5).

Nous avons calculé les ratios du temps de génération des révertants par rapport à celui de la souche sauvage de référence pour tous les révertants. Ces ratios (g révertant 2n / g 2nU+ et g révertant n / g FL100) indiquent une baisse de la vitesse de croissance si le rapport est supérieur à 1 ou une hausse si ce rapport est inférieur à 1.

En contexte diploïde, hormis la souche B199, tous ces rapports sont largement supérieurs à 1 montrant clairement une baisse de la vitesse de croissance. Pour les haploïdes, ces rapports sont plus proches de 1 et suggèrent une faible variation de la croissance cellulaire dans ce contexte. La comparaison des 2 contextes montre que ce rapport est inférieur dans les révertants haploïdes comparés aux diploïdes et ce pour le même remaniement chromosomique (Tableau 2.5). Par exemple, ce ratio est de 2,06 pour la souche B259 alors qu'il n'est que de 1,15 pour B259 n (Tableau 2.5). Ceci montre que la prolifération des cellules haploïdes est moins altérée par les réarrangements chromosomiques et notamment les aneuploïdies, que celle des souches diploïdes.





Figure 2.5 : Détermination des temps de génération des révertants.

- A) Les souches diploïdes
- B) Les souches haploïdes

Les astérisques indiquent les temps de génération qui sont statistiquement comparables à ceux des souches sauvages.

Souches	Temps de génération g (h)	Écart- type	g révertant/ g 2nU+	Souches	Temps de génération g (h)	Écart- type	g révertant/ g FL100
2n U+	1,52	0,255	1	FL100	2,41	0,094	1
B199	1,23	0,206	0,81				
B038	2,09	0,135	1,375	B038 n	2,86	0,061	1,19
B259	3,13	0,152	2,06	B259 n	2,77	0,09	1,15
B284	3,22	0,42	2,12	B284 n	2,53	0,11	1,05
B066/1	3,75	0,16	2,47	B066/1 n	nd	nd	nd
B132	2,8	0,04	1,84				
B326	3,57	0,447	2,35				
Rev27 2n	2,74	0,106	1,8	Rev27	2,36	0,205	0,98
Rev46 2n	nd	nd	nd	Rev46	3,85	0,278	1,6
B286	2,86	0,22	1,88	B286 n	4,36	0,28	1,81
AWR58 2n	2,02	0,023	1,33	AWR58	3,27	0,115	1,36
B283/1	2,35	0,441	1,55				
BR15/2	2,69	0,25	1,77				
				Rev25	2,57	0,373	1,07

<u>Tableau 2.5</u> : Les temps de génération des révertants. La moyenne des temps de génération (g) est calculée à partir de 3 cultures indépendantes.

Le temps de génération g est indiqué en heure.

Partie gauche du tableau : Souches diploïdes.

Partie droite du tableau : Souches haploïdes.

Les ratios des temps de génération des révertants par rapport à ceux des souches sauvages (g révertant / g 2nU+ et g révertant / g FL100 dans les contextes diploïde et haploïde respectivement) indiquent une baisse (> 1) ou une hausse (< 1) de la vitesse de croissance.

2.1.2. Détermination des temps de latence des révertants

Nous avons déterminé les temps de latence des souches de notre collection à partir des résultats obtenus par le lecteur de microplaque de type TECAN INFINITE F200. Ils résultent de la moyenne de 6 cultures réalisées indépendamment. Ce temps est défini comme étant la période durant laquelle une cellule s'adapte à son environnement pour débuter sa croissance (Figure 2.6). Les résultats obtenus montrent de grands écart-types pour une même souche résultant de la non-reproductibilité des calculs de temps de génération. Malgré ces grandes variances, les temps de latence des révertants Rev27 2n, Rev46 2n et AWR58 2n sont largement supérieurs à celui de la souche sauvage 2nU+ ce qui démontre une adaptation plus longue de ces révertants à un nouvel environnement (Figure 2.7). En contexte haploïde, seul le révertant Rev27 a un temps de latence supérieur à FL100 alors que les autres semblent débuter leur croissance plus rapidement (Figure 2.7).



<u>Figure 2.6</u>: Détermination des paramètres de croissance à partir des résultats obtenus par le lecteur de microplaques TECAN INFINITE F200. Ici, nous avons représenté l'exemple de la souche sauvage diploïde 2nU+. Chaque point correspond à la mesure de l'absorbance d'une culture en fonction du temps de croissance. La vitesse spécifique de croissance est déterminée dans la phase exponentielle de croissance. Le temps de latence correspond à l'intersection entre la droite correspondant à l'ensemencement initial (droite verte) et celle déterminée pour calculer le temps de génération (droite rouge).


Figure 2.7 : Détermination des temps de latence des révertants.

- A) Souches diploïdes
- B) Souches haploïdes

2.2. Détermination de la consommation spécifique de glucose dans les révertants

Nous avons dosé les quantités de glucose présentes pour chaque prélèvement effectué pour la détermination du temps de génération dans le but de déterminer la consommation spécifique de glucose. Le milieu de culture utilisé est un milieu minimum liquide contenant initialement du glucose à une concentration de 5 g/l. Le dosage a été réalisé par une méthode enzymatique indirecte qui consiste à déterminer la quantité de cofacteurs NADPH + H^+ produits. En présence, d'héxokinase et de glucose-6-phosphate déshydrogénase, le glucose passe sous une forme glucose-6-phosphate puis gluconate-6-phosphate en réduisant le cofacteur NADP en NADPH + H^+ .

En fonction du glucose consommé, nous avons calculé le rendement de consommation Y qui correspond à la quantité consommée par unité d'absorbance (Figure 2.8). Pour comparer les consommations de glucose et s'affranchir des temps de génération variables, la vitesse spécifique de croissance μ est divisée par le rendement Y. Les cellules ayant des formes variables (paragraphe ci-dessus), le nombre de cellules ne sera pas le même d'une souche à l'autre pour une absorbance donnée. De ce fait, pour s'affranchir de cette variabilité, le poids sec de chaque cellule a été calculé. Le rapport μ /Y est ainsi divisé par le poids sec pour obtenir la consommation spécifique de glucose Q_{Glu}. L'unité de cette consommation spécifique de glucose est définie comme étant le nombre de mmoles de glucose consommées par heure et par gramme de cellules (mmol.g⁻¹.h⁻¹).



<u>Figure 2.8</u>: Exemple de la détermination du rendement de glucose consommé Y pour la souche 2nU+. Les quantités de glucose consommé sont représentées en abscisse et l'absorbance des cultures en ordonnée.

Les résultats obtenus montrent qu'à l'état diploïde, seul le révertant B199 a une consommation spécifique de glucose supérieure à celle de la cellule sauvage (c'est la souche qui pousse plus vite). Dans les souches B038, B259 et AWR58 2n, la consommation est comparable à celle de la souche sauvage (Figure 2.9). Par contre, toutes les autres souches diploïdes ont une consommation de glucose moindre. Il est notamment intéressant de noter que les révertants consommant le glucose le plus rapidement sont ceux dont la vitesse de croissance est la plus rapide. Ceci suggère que la consommation de glucose peut être limitante pour une prolifération optimale de ces révertants.

Dans le contexte haploïde, ces différences sont moins marquées et seuls les révertants B259 n, Rev46 et B286 n consomment le glucose plus lentement que la souche FL100 (Figure 2.9). Ce résultat est différent de celui obtenu par Torres et *al* (2007) pour des souches disomiques de *S. cerevisiae*, possédant des chromosomes entiers additionnels, dans lesquelles on observe une augmentation de la consommation de glucose. Les auteurs de cette étude ont avancé l'hypothèse selon laquelle cette augmentation de la consommation de glucose serait nécessaire afin de former plus d'ATP pour assurer la réplication, la transcription des copies de gènes supplémentaires ou encore pour réguler d'éventuels systèmes de contrôle du taux de protéines (Torres et *al.*, 2007).

2.3. Détermination de la production spécifique d'éthanol dans les révertants

De la même manière que la consommation de glucose Q_{Glu} , nous avons calculé la production spécifique d'éthanol π_{EtOh} . Ce dosage enzymatique indirect repose sur la quantité de NADH + H⁺ formé. En présence d'une alcool déshydrogénase et d'une aldéhyde déshydrogénase, l'éthanol est transformé en acétaldéhyde puis en acide acétique en réduisant le cofacteur NADH + H⁺. Comme pour la consommation de glucose, la production spécifique d'éthanol a été obtenue en divisant la vitesse spécifique de croissance par le rendement de production d'éthanol Y puis normalisée en prenant en compte le poids sec de chaque souche. Des variations dans la production d'éthanol sont observées entre les révertants et les souches sauvages (Figure 2.9). L'analyse des 20 révertants de notre collection montre que 14 d'entreeux présentent une production d'éthanol similaire à FL100 ou 2nU+ (Figure 2.9).





- A) Souches diploïdes
- B) Souches haploïdes

Le ratio de la quantité d'éthanol formé à partir d'une molécule de glucose (Q_{Glu}/π_{EtOh}) présenté dans le tableau 2.6 permet de déterminer la quantité d'éthanol produit à partir du même nombre de molécules de glucose consommées. Plus ce rapport est élevé, moins la production d'éthanol est importante. La comparaison de ces ratios montre des différences liés à la ploïdie. À l'état diploïde, ce ratio est identique pour de nombreux révertants par rapport à la souche 2nU+ ce qui indique que la molécule de glucose consommée permet la formation du même nombre de molécules d'éthanol. Par contre, dans les souches comme B038, B286 ou B283/1, on observe un rapport supérieur à celui de 2nU+ ce qui montre une production d'éthanol moindre à partir de la même quantité de glucose consommée (Tableau 2.6).

De la même manière, dans le contexte haploïde, des hausses de ce ratio sont observées dans toutes les souches à l'exception de 2 révertants. Cette propriété indique que la quantité d'éthanol produit à partir d'une molécule de glucose est moins importante que celle dans la souche FL100 (Tableau 2.6). Ces résultats, combinés avec ceux de la consommation spécifique de glucose, montrent que la fermentation est moins efficace et suggèrent qu'elle peut être partiellement remplacée par le métabolisme respiratoire afin de former plus d'ATP.

Souches	\mathbf{Q}_{Glu}	π_{EtOh}	Q_{glu}/π_{EtOh}	Souches	\mathbf{Q}_{Glu}	π_{EtOh}	Q_{glu}/π_{EtOh}
2n U+	21,86	33,4	0,65	FL100	13,94	28,44	0,49
B199	26	39,41	0,66				
B038	19,71	24,33	0,81	B038 n	11,79	16,08	0,73
B259	16,64	26,72	0,62	B259 n	6,38	8,2	0,78
B284	12,55	18,57	0,68	B284 n	16,94	25,96	0,65
B066/1	10,66	16,86	0,63	B066/1 n	nd	nd	nd
B132	8,43	14,63	0,58				
B326	11,37	14,46	0,79				
Rev27 2n	12,6	18,19	0,69	Rev27	16,31	34,32	0,47
Rev46 2n	nd	nd	nd	Rev46	8,6	12,11	0,71
B286	14,09	18,58	0,76	B286 n	6,85	10,44	0,66
AWR58 2n	16,64	24,31	0,68	AWR58	13,28	16,64	0,8
B283/1	16,19	21,41	0,76				
BR15/2	10,23	13,83	0,74				
				Rev25	9,91	23,44	0,42

<u>Tableau 2.6</u> : Consommation spécifique de glucose et production spécifique d'éthanol. Chaque résultat est la moyenne de 3 expériences indépendantes.

 Q_{Ghu} : Consommation spécifique de glucose (mmol.g⁻¹.h⁻¹)

 π_{EtOh} : Production spécifique d'éthanol (mmol.g⁻¹.h⁻¹)

Le rapport Q_{glu}/π_{EtOh} est calculé pour déterminer la quantité d'éthanol formé à partir d'une molécule de glucose. Ce ratio est inversement proportionnel à la quantité d'éthanol formé : plus le rapport Q_{glu}/π_{EtOh} est élevé, moins il y a d'éthanol formé.

3. Analyse globale de la transcription

L'analyse globale de la transcription permet d'obtenir le profil d'expression de la totalité des gènes du génome. Cette analyse a été effectuée par l'hybridation compétitive des ADNc des révertants et des souches sauvages sur des puces à ADN de type Agilent 8x 15K, en collaboration avec l'École Normale Supérieure de Paris. Pour réaliser ces expériences, nous avons préparé et purifié les ARN totaux des souches sauvages et des révertants à partir de cellules en milieu de phase exponentielle de croissance. Les ADNc synthétisés à partir de ces ARN sont marqués avec 2 fluorophores : la cyanine 3 pour les ADNc des souches sauvages et la cyanine 5 pour ceux des révertants. L'intensité de la fluorescence mesurée provient de l'hybridation des ADNc et correspond à la quantité d'ADNc fixé sur les sondes : plus la quantité d'ADNc est abondante, plus cette intensité est élevée. Pour comparer les transcrits entre les souches sauvages et les révertants, le rapport des intensités des cyanines 3 et 5 est calculé et les résultats transformés en valeurs logarithmiques. Les valeurs supérieures à 0 représentent des gènes dont la transcription est augmentée par rapport à la souche sauvage alors que celles inférieures à 0 correspondent à des gènes dont la transcription est diminuée.

3.1. Les données analysées

Les puces utilisées lors de l'analyse transcriptomique sont constituées pour chaque gène de sondes de 60 oligonucléotides. Chaque sonde est placée 2 fois sur la puce. Toutes les hybridations ont été effectuées 2 fois, lors d'expériences réalisées indépendamment. De ce fait, nous possédons 4 valeurs d'hybridations par transcrit. Afin de déterminer si ces valeurs sont reproductibles, nous avons calculé deux types de coefficients de corrélation. Chaque sonde étant présente en 2 exemplaires sur chaque puce, le premier coefficient de corrélation est déterminé en fonction des valeurs des hybridations intra-lames. Le second est calculé en fonction des résultats obtenus entre les différentes hybridations (inter-lames). Nous avons réalisé des matrices de points pour chaque type de coefficient (Figure 2.10). Des données reproductibles doivent s'aligner sur une droite de coefficient directeur égal à 1 et posséder un coefficient de corrélation (R²) supérieur à 0,5. Lorsque les coefficients intra-lames sont reproductibles, nous calculons ceux inter-lames et conservons les résultats ayant un coefficient de corrélation supérieur à 0,5. Ainsi, les données transcriptomiques de 8 révertants diploïdes et 7 haploïdes ont été conservées pour être analysées (Tableau 2.7). Les 15 révertants retenus portent des réarrangements variés (duplications géniques, segmentales, aneuploïdies et délétions).

Dans la suite, nous ferons une synthèse de l'ensemble des résultats que nous avons analysés. Nous nous intéresserons d'abord aux gènes directement impliqués dans la voie de biosynthèse des nucléotides pyrimidiques. Nous détaillerons ensuite les résultats concernant les gènes qui ont été dupliqués ou délétés lors du remaniement chromosomique. Enfin, nous analyserons les autres gènes non impliqués dans la structure du remaniement, mais qui montrent une modification de leur expression.



<u>Figure 2.10</u>: Exemples de coefficients de corrélation. Dans cet exemple, nous avons représenté les coefficients de corrélation interlames (R^2) de deux révertants. Chaque point correspond à la valeur transcriptomique d'un gène. L'axe des abscisses représente les valeurs des hybridations de la puce 1 et alors que celles de la puce 2 sont indiquées sur l'axe des ordonnées.

A) Révertant B199 retenu pour les analyses des données

B) Révertant B283/1 non retenu pour ls analyses de données

3.2. Expression des gènes impliqués dans la voie de biosynthèse des nucléotides pyrimidiques

Le système de sélection ayant permis l'obtention des révertants est basé sur la réactivation du gène *URA2*. Ce gène code la protéine Ura2p qui intervient dans les deux premières étapes de la voie de biosynthèse des nucléotides pyrimidiques en transformant l'aspartate en carbamylphosphate (étape 1) puis en acide uréidosuccinique (étape 2). Dans nos révertants, seule la seconde étape est réalisée par la protéine chimère contenant notamment la région ATCase, la première étape étant assurée par les protéines Cpa1p et Cpa2p formant un complexe qui intervient dans la formation de carbamylphosphate dans la voie de biosynthèse de l'arginine (Messenguy et *al.*, 1983).

Révertants	Type de réarrangement	Coefficient de corrélation inter-lames
Diploïdes		
B199	Aneuploïdie	0,8492
B038	Aneuploïdie	0,7247
B259	Aneuploïdie	0,7177
B284	Aneuploïdie	0,4266
B066/1	Aneuploïdie	0,5447
B132	Translocation non réciproque	0,4204
B326	Translocation non réciproque	0,4211
Rev27 2n	Duplication segmentale	0,4571
Rev46 2n	Duplication segmentale	0,6635
B286	Duplication segmentale	0,739
AWR58 2n	Délétion	0,6471
B283/1	Délétion	0,2526
BR15/2	Délétion	0,7281
Haploïdes		
B038 n	Aneuploïdie	0,8425
B259 n	Aneuploïdie	0,7675
B284 n	Aneuploïdie	0,7847
B066/1 n	Aneuploïdie	0,3806
Rev27	Duplication segmentale	0,7716
Rev46	Duplication segmentale	0,1665
B286 n	Duplication segmentale	0,8147
AWR58	Délétion	0,5082
Rev25	Duplication génique	0,5263

<u>Tableau 2.7</u> : Mesure des coefficients de corrélation inter-lames. Les résultats des souches surlignées en orange ne sont pas prises en compte pour cette analyse transcriptomique.

Les puces Agilent utilisées dans cette expérience d'hybridation compétitive contiennent une sonde correspondant au gène *URA2* qui est située au niveau de la région ATCase. De ce fait, les ADNc correspondant aux gènes chimères s'hybrident sur ces sondes et les valeurs d'expression associées au gène *URA2* sont équivalentes à celles du gène de fusion néoformé. Les résultats montrent que les gènes chimères ont une transcription augmentée dans toutes les cellules haploïdes (sauf AWR58) et les diploïdes aneuploïdes. Pour les autres révertants, seul le gène de fusion du révertant B286 est surexprimé alors que les gènes chimères des autres souches sont sous-exprimés (Tableau 2.8).

Gène	B199	B038	B259	B066	Rev46 2n	B286	AWR58 2n	BR15-2
URA1	0.676	1.392	1.02875	1.2075	-0.477	1.677	-0.7115	0
URA2	1.6895	1.7135	1.868	0	-1.242	3.8875	-0.6825	-1.1735
URA3	0	0	-0.29525	0	0	1.013	-1.7495	-0.7025
URA4	0.31525	0.90725	0.41325	0.293	0	1.053	-0.565	0
URA5	0	0	0	0	0	-0.378	0	0
URA6	-0.53375	-0.6305	0	0	-0.6545	-0.3505	-0.4775	0
URA7	0	-0.2815	-0.3695	0	-0.498	-1.0665	0	-0.656
URA8	-0.29025	-0.30625	1.71575	-0.81775	0.2485	0	-0.5575	-0.401
URA10	0	0	0.58325	0	0	0.9475	-0.3065	0.323
CPA1	1.15125	1.53375	0.97125	0.68325	1.1365	1.4655	0.6065	1.619
CPA2	0	1.97275	3.09225	0.83275	2.004	2.175	0.377	1.07

Gène	B038 n	B259 n	B284 n	Rev27	B286 n	AWR58	Rev25
URA1	2.265	2.025	3.79	1.175	2.35	1.15	2.77
URA2	4.165	4.3	3.18	0.55	5.425	-2.215	1.62
URA3	0.215	0.91	2.645	0	0.705	0	1.92
URA4	0.815	1.76	2.94	0.915	1	0	2.315
URA5	1.045	0.83	0.875	0.77	0.685	0	0.9
URA6	0.72	0.26	0	-0.775	-0.63	0	-1.185
URA7	1.075	1.635	1.72	0.78	0	0	0.93
URA8	-1.675	-1.16	-1.405	0	-1.485	0.475	-0.695
URA10	-0.6	-0.76	-0.505	-0.37	0.945	0.68	-0.645
CPA1	1.745	2.185	2.195	1.37	2.69	1.025	1.555
CPA2	0.455	0.47	1.66	2.99	1.8	1.25	2.57

<u>Tableau 2.8</u> : L'expression des gènes codant des protéines intervenant dans la voie de biosynthèse des nucléotides pyrimidiques. Les gènes surexprimés sont représentés en rouge, orange ou jaune en fonction de leur niveau de transcription. Les gènes sous-exprimés sont représentés en vert clair et foncé.

A) Révertants diploïdes

B) Révertants haploïdes

L'expression de ces gènes *CPA1* et *CPA2*, dont les protéines encodées sont supposées fournir le carbamylphosphate pour la catalyse de l'acide uréidosuccinique par l'ATCase, est augmentée par rapport aux cellules sauvages. Cette augmentation de la transcription peut être conforme au fait que les protéines Cpa1p et Cpa2p codés par ces gènes permettent la synthèse du carabamylphosphate pour les voies de synthèse de l'arginine mais dans ces souches également pour les nucléotides pyrimidiques.

Parmi les autres gènes de la voie de biosynthèse des nucléotides pyrimidiques, les gènes codant les enzymes intervenant dans la voie de biosynthèse endogène permettant la formation d'UMP sont majoritairement surexprimés chez les révertants haploïdes (*URA1*,

URA3, URA4, URA5). Chez les diploïdes aneuploïdes, seuls les gènes impliqués dans la formation d'acide orotique (*URA1, URA4*) ont une transcription augmentée (Tableau 2.8). Ces résultats montrent que seules les premières étapes de la voie de biosynthèse des nucléotides pyrimidiques sont régulées de façon identique dans tous les révertants.

3.3. Rôle de l'expression des gènes et de la ploïdie dans la réponse cellulaire

3.3.1. Les parties dupliquées et délétées

Les réarrangements chromosomiques étudiés ici possèdent un nombre de gènes dupliqués variables pouvant aller jusqu'à 300 gènes pour les souches B066/1 et B284. Ces gènes ont été mis en évidence par hybridation génomique compétitive sur puce à ADN (Schacherer et *al.*, 2007).

L'analyse de l'expression de ces gènes au niveau des divers révertants montre que ces gènes sont légèrement surexprimés chez les révertants B199, B259 et B066/1 possédant une aneuploïdie (Figure 2.11). Par contre, chez les autres souches, ces gènes peuvent avoir une expression augmentée ou diminuée, voire n'avoir aucun changement d'expression par rapport aux souches sauvages (Figure 2.11). On observe en effet sur cette figure, dans les zones bleues et orange qui correspondant aux gènes impliqués dans le remaniement, des variations d'expression d'un gène à l'autre, Ces résultats indiquent clairement une non-corrélation entre le nombre de copies de gènes et leur expression au sein de notre collection de révertants.

Pour les gènes non impliqués dans le remaniement, des variations d'expression sont également mises en évidence (Figure 2.11). À l'état diploïde, un peu moins de la moitié des gènes ont une expression différente comparée à la souche sauvage, alors qu'ils représentent deux tiers des gènes à l'état haploïde. Dans l'ensemble, environ la moitié des gènes dont l'expression est différente par rapport à la souche de référence a une expression augmentée et l'autre moitié a une expression diminuée.

Afin d'analyser davantage les impacts de cette modification de l'expression des gènes au sein des révertants, nous avons regroupé les gènes régulés positivement d'une part et négativement d'autre part pour déterminer quelles étaient les fonctions de la cellule qui ont été affectées par les réarrangements chromosomiques. Ce travail a été poursuivi en regroupant les gènes appartenant à une même fonction ou à un ensemble de fonctions. Les groupes que nous avons ainsi formés ont été appelés clusters.



<u>Figure 2.11</u>: Expression des gènes dupliqués étant impliqués dans la formation des chromosomes chimères. Pour chaque souche, les 2 chromosomes impliqués dans la formation des chromosomes chimères surnuméraires sont représentés. Le cadre bleu indique la région dupliquée du chromosome X alors que le cadre jaune correspond à la région dupliquée des chromosomes receveurs.

- A) Révertants diploïdes
- B) Révertants haploïdes

3.3.2. Rôle de la ploïdie dans la réponse cellulaire

Dans cette partie, nous avons regroupé dans des clusters des gènes dont l'expression est augmentée ou diminuée possédant des fonctions communes. Pour réaliser cette « clusterisation », nous avons conservé les valeurs transcriptomiques inférieures à - 0,5 et supérieures à + 0,5. Le logiciel MeV permet la formation de clusters en réalisant des tests statistiques comme le test de Student (Margolin et *al.*, 2005). Nous avons employé ce logiciel en utilisant un test de Student avec des valeurs statistiques seuils différentes (p-value = 0,05 et 0,01). Ainsi, deux clusters de 1200 et 510 gènes ont été générés chez les révertants diploïdes. Ces nombres étant très élevés, nous avons testé puis choisi une méthode alternative. À partir des valeurs transcriptomiques, nous avons regroupé les gènes dont l'expression est augmentée et ceux où elle est diminuée en fonction du type de réarrangement chromosomique et en fonction de la ploïdie (Tableau 2.9). Ces regroupements de gènes servent alors pour définir les fonctions cellulaires associées à ces clusters.

Clusters	Gènes surexprimés	Gènes sous-exprimés
Aneuploïdes 2n	189	128
Délétions 2n	412	475
Duplications segmentales 2n	405	415
Aneuploïdies n	666	747
Duplications segmentales n	945	930

<u>Tableau 2.9</u>: Nombre de gènes montrant une expression différente pour un type de réarrangement chromosomique. Cette clusterisation a été effectuée en distinguant les souches diploïdes (2n) des souches haploïdes (n).

Le logiciel GoTermFinder (<u>http://go.princeton.edu/cgi-bin/GOTermFinder</u>) a été utilisé pour déterminer les fonctions et les processus dans lesquels interviennent les protéines codées par les gènes retenus (Boyle et *al.*, 2004). La figure 2.12 présente quelques exemples des différents résultats obtenus pour les révertants. Nous avons classé les différents gènes retenus par grand groupe de fonction. Dans cette figure, chaque ligne représente un gène ayant une expression différente par rapport à la souche de référence. Par exemple, des gènes intervenant dans la formation du ribosome ou dans la maturation des ARN sont sous-exprimés dans les révertants diploïdes (Figure 2.12).

L'analyse de l'ensemble du transcriptome montre que dans les révertants diploïdes, les protéines codées par les gènes surexprimés sont impliqués dans les complexes mitochondriaux, notamment au niveau de la chaîne respiratoire. Il y a également une expression modifiée de nombreux gènes impliqués dans le cycle de Krebs ou dans la glycolyse. D'autres protéines codées par des gènes surexprimés interviennent dans plusieurs voies métaboliques notamment dans l'utilisation de l'éthanol ou dans la biosynthèse d'acides aminés, dont notamment ceux à radical aromatique. Pour les protéines codées par les gènes sous-exprimés, on retrouve essentiellement des protéines impliquées dans la formation du ribosome et dans le transport d'ions métalliques (Figure 2.12).

À l'inverse, les gènes impliqués dans la synthèse du ribosome sont surexprimés dans les souches révertants haploïdes comme d'ailleurs les gènes impliqués dans l'initiation de la traduction (Figure 2.12). Cette observation va dans le sens d'une traduction accrue chez les haploïdes. De manière similaire aux révertants diploïdes, les gènes intervenant dans la chaîne respiratoire sont également surexprimés en haploïde.

À côté des gènes dont l'expression est modifiée dans tous les révertants, chaque souche possède un jeu de gènes spécifiques dont l'expression est également modifiée. Les produits de ces gènes participent également au métabolisme énergétique ou à la traduction. Par exemple, les révertants diploïdes B038, B199 et B259 ont une sous-expression de plusieurs dizaines de gènes intervenant dans la formation du ribosome renforçant l'hypothèse selon laquelle la traduction est différente en fonction des contextes de ploïdie (Tableau 2.10).

Chaque révertant possède également des gènes spécifiquement régulés dont les fonctions ne peuvent pas être regroupées dans les différentes fonctions décrites dans le tableau 2.10. Par exemple, nous n'avons pas réussi à faire sporuler le révertant diploïde B199 pour obtenir une version haploïde malgré plusieurs essais et modification de protocole. Le fait que des gènes intervenant dans la méiose soient spécifiquement sous-exprimés dans cette souche peut expliquer en partie l'incapacité de sporulation de ce révertant.



<u>Figure 2.12</u>: Exemples de processus cellulaires dont les gènes ont une expression modifiée. Les gènes sous-exprimés sont représentés en vert et les gènes surexprimés en rouge. Les processus cellulaires ont été déterminés à partir du programme GoTermFinder. Chaque ligne représente l'expression modifiée d'un gène par rapport aux souches de référence et chaque colonne indique un révertant.

A) Révertants diploïdes

B) Révertants haploïdes

C)

	B199	B038	B259	B066/1
Gènes surexprimés				
Transport d'éléctrons	oui	/	oui	oui
Synthèse d'ATP	oui	/	/	/
Processus Red/ox	oui	/	oui	oui
Précurseurs nucléotidiques	/	/	oui	oui
Métabolisme des sucres	oui	oui	/	oui
Métabolisme des vitamines	oui	oui	/	/
Métabolisme de l'alcool	/	oui	oui	/
Métabolisme des cétones	/	oui	oui	oui
Gènes sous-exprimés				
Ribosome	oui	oui	oui	/
Biosynthèse des macromolécules	oui	oui	oui	oui

<u>Tableau 2.10 :</u> Gènes exprimés spécifiquement chez les révertants diploïdes aneuploïdes. Ces gènes sont regroupés selon leur fonction. Des fonctions peuvent être communes à plusieurs souches.

Lors de notre analyse des ratios de consommation spécifique de glucose par rapport à la production d'éthanol (Q_{Glu}/π_{EtOh}), nous avons postulé que le métabolisme fermentaire pouvait être partiellement remplacé par la respiration. L'analyse globale de la transcription est consistante avec cette hypothèse car des gènes intervenant dans la glycolyse, le cycle de Krebs ou la phosphorylation oxydative sont surexprimés chez tous les révertants. Ce changement de métabolisme suggère une production accrue d'ATP.

3.3.3. Réponse cellulaire et stress

Lorsque des cellules de *S. cerevisiae* sont soumises à un stress comme un choc thermique ou un composé toxique dans le milieu, environ 900 gènes ont une expression modifiée par rapport à des conditions de croissance standard. Ces gènes ont été définis comme étant des éléments de réponse au stress et ont été appelés ESR (Gasch et *al.*, 2000). Parmi les 900 gènes ESR répertoriés chez *S. cerevisiae*, chaque révertant que nous avons analysé possède au moins 150 de ces gènes qui ont une expression différente des souches sauvages, ce qui montre que nos souches répondent également à un stress. Cette observation est à mettre en parallèle avec les travaux de Torres et *al.* (2007) qui ont montré que des souches disomiques de *S. cerevisiae*, dans le contexte haploïde, avaient une signature d'expression comparable à celle obtenue lors d'un stress. Ils l'ont d'ailleurs, qualifié de stress aneuploïdique (Williams et Amon, 2009). Pour chaque souche, nous avons donc analysé les valeurs transcriptomiques correspondantes aux 900 gènes ESR puis nous avons effectué une

série de tests statistiques. Ces tests de Student indiquent que 354 gènes ont statistiquement une expression différente des souches sauvages (Figure 2.13).

Le révertant AWR58 est particulier car peu de gènes ESR ont une expression modifiée dans le transcriptome. Cette propriété peut être liée à la fonction du gène receveur de la fusion. En effet, le partie ATCase du gène *URA2* est fusionnée avec le gène *PBS2* qui code une protéine kinase impliquée dans la voie de signalisation HOG (Schüller et *al.*, 1994). La phosphorylation de la protéine Hog1p par pbs2p permet l'activation des facteurs de transcription Mns2p et Mns4p à la base de la régulation des gènes ESR (Gasch, 2007). Au niveau de la protéine chimère Pbs2p-ATCase, les domaines kinases sont délétés ce qui peut induire une non phosphorylation de Hog1p et ainsi bloquer cette voie de signalisation (voir chapitre 3). Cette hypothèse peut ainsi expliquer pourquoi peu de gènes ESR ont une expression différente chez le révertant AWR58 par rapport au sauvage contrairement aux autres révertants (Figure 2.13).

L'analyse des gènes ESR montre des résultats variables en fonction de la ploïdie. En effet, les gènes impliqués dans le processus de catabolisme des protéines vacuolaires sont surexprimés dans le contexte diploïde et sous-exprimés en haploïde, quelque soit le type de remaniement. Par contre, les gènes impliqués dans la biogenèse du ribosome et la machinerie traductionnelle sont surexprimés dans les révertants haploïdes et sous-exprimés chez les diploïdes (Figure 2.13). La sous-expression de ces gènes dans les révertants diploïdes suggère une baisse de la traduction dans ce contexte de ploïdie. La baisse de traduction est une des hypothèses fondées par Torres et ses collègues (2008) suite à l'apparition d'une aneuploïdie. Lors de ce travail de revue, les auteurs ont repris les analyses transcriptomiques de plusieurs études sur des souches aneuploïdes de S. cerevisiae et ont mis en évidence une corrélation entre le nombre de copies de gènes et leur expression. Ils ont également suggéré que la réponse cellulaire s'effectuait au niveau protéique plutôt qu'au niveau des transcrits. Cependant, une augmentation de la quantité de protéines peut changer la balance protéique et ainsi perturber la croissance cellulaire. Afin d'équilibrer cette balance, ces mêmes auteurs ont posé 2 hypothèses : une baisse de la traduction ou alors la dégradation de l'excès de protéines synthétisées (Torres et al., 2008).



<u>Figure 2.13</u>: Profil d'expression des gènes ESR dans les révertants. Les gènes sous-exprimés sont représentés en vert et les gènes surexprimés en rouge. Chaque ligne correspond aux données transcriptomiques correspondant à un gène.

La balance protéique peut se définir comme une quantité de protéines équilibrée nécessaire pour un développement cellulaire optimal. Par exemple, une surexpression de l'une ou l'autre des histones ou des tubulines entraîne un déséquilibre qui peut mener à des baisses de la croissance ou à la mort cellulaire *chez S. cerevisiae*. L'excès de protéines peut alors être régulé par une dégradation des protéines additionnelles (Gunjan et Verreault, 2003 ; Lacefield et *al.*, 2006).

L'analyse des gènes ESR montre clairement que dans les révertants diploïdes, le taux de traduction diminue contrairement aux souches haploïdes. Par contre, pour les cellules haploïdes, l'augmentation de l'expression des gènes intervenant dans l'activité ou la régulation du protéasome suggère une dégradation accrue de l'excès éventuel de protéines synthétisées. Ces résultats montrent que le contexte de ploïdie influence la réponse cellulaire et que la balance protéique est équilibrée selon deux mécanismes différents : les souches diploïdes diminuent la traduction alors que les haploïdes dégradent l'excès de protéines synthétisées.

L'augmentation de la traduction dans ces cellules a été testée en utilisant indirectement la rapamycine, un inhibiteur de la machinerie traductionnelle. Cette molécule inhibe la voie de signalisation TOR qui mène à une transcription moindre des ARN ribosomiques et donc à une baisse de la traduction (pour revue, Crespo et Hall, 2002). Les révertants diploïdes comme d'ailleurs la souche 2nU+ sont résistantes à des concentrations de 10 nM de rapamycine montrant que la voie de signalisation cellulaire initiée par la protéine TOR et ainsi la traduction n'a pas d'influence sur ces souches (Tableau 2.11). Par contre, à l'état haploïde, les révertants B284 n et B038 n ont une plus grande sensibilité vis-à-vis de cette drogue par rapport à la souche de référence FL100 (Tableau 2.11). L'absence de croissance de ces deux souches montre que la voie de signalisation TOR inhibe avec une meilleure efficacité leur machinerie traductionnelle. Ces résultats suggèrent que la traduction est plus facilement inhibée dans ces deux révertants ce qui est consistant avec l'idée que la traduction est augmentée chez les souches haploïdes. En effet, si une augmentation de la machinerie traductionnelle est nécessaire à la survie cellulaire des révertants, l'introduction de composés inhibiteurs aura un impact plus rapide et plus important sur cette survie. Cette approche n'est qu'une étude préliminaire pour déterminer cette augmentation. En effet, au moins deux mille gènes ont une transcription modifiée dans les révertants dont des gènes pouvant intervenir dans la voie de signalisation TOR. De telles différences peuvent expliquer la résistance du

révertant B259 n (Tableau 2.11). Une analyse plus approfondie de ces gènes semble donc nécessaire pour comprendre comment une cellule s'adapte suite à l'apparition d'un réarrangement chromosomique.

	YPD	YPD + Ra	pamycine
	30°C	10nM	25nM
Diploïdes			
2nU+	++	++	-
B199	++	+	-
B038	++	++	-
B259	++	++	-
B066/1	++	++	
Haploïdes			
FL100	++	++	-
B038n	++	-	-
B284n	++	+	-
B259n	++	++	-

<u>Tableau 2.11</u> : Résistance à la rapamycine des révertants aneuploïdes. Ces analyses ont été effectuées à partir de deux dilutions d'une suspension de cellules différentes de cellules. ++ : croissance des deux dilutions

+ : : croissance d'une seule dilution

- : aucune croissance

Conclusion

L'observation macroscopique des colonies des souches de notre collection montre des variations morphologiques en fonction de la ploïdie et de la source de carbone présente dans le milieu. La comparaison de ces phénotypes indique également qu'il n'y a pas de patron de morphologie pour un type de réarrangement chromosomique.

L'analyse microscopique des cellules des différents révertants et des souches sauvages démontre aussi des phénotypes variables : certaines cellules sont plus allongées avec un volume cellulaire plus élevé que les cellules de références avec des formes moins caractéristiques de *S. cerevisiae*, alors que d'autres peuvent s'ordonner en agrégats. Ces différences morphologiques, notamment l'augmentation du volume cellulaire, peuvent

s'expliquer par un cycle cellulaire allongé avec notamment un délai au niveau de la transition des phases G1 et S, ce qui peut accroître le temps de croissance des révertants (Giaver et *al.*, 2002 ; Torres et *al.*, 2007).

Lorsque l'on compare les temps de génération des révertants diploïdes, seules les souches B199, B038 et B283/1 ont un temps de génération comparable à la souche de référence 2nU+. Contrairement aux souches diploïdes, les révertants haploïdes ont un temps de génération proche de celui de FL100, excepté pour les souches Rev46, B286 n et AWR58. Ces propriétés suggèrent que les réarrangements chromosomiques ont moins d'impact sur la croissance cellulaire dans le contexte haploïde que dans le contexte diploïde.

Les consommations spécifiques de glucose (Q_{Glu}) des révertants diploïdes et haploïdes sont inférieures à celles des souches sans remaniements, excepté pour la souche B199. De manière identique, la production d'éthanol (π_{EtOh}) est diminuée chez la majorité des révertants. Le calcul du ratio Q_{Glu}/π_{EtOh} permet d'estimer la quantité d'éthanol produite par molécule de glucose consommée. Ce rapport, plus élevé pour de nombreux révertants haploïdes et diploïdes en comparaison à leurs souches de référence, indique une production moins importante d'éthanol. Ceci peut être le résultat d'un remplacement partiel du métabolisme fermentaire par la respiration. Un résultat similaire a été démontré chez des souris trisomiques dans lesquelles une fermentation hétérologue est supposée du fait de leur production d'éthanol et d'acide lactique (Williams et *al.*, 2008).

Parmi la collection de 22 révertants, 15 ont été utilisés pour une analyse globale de la transcription. Cette étude montre que le nombre de copies de gènes et leur expression ne sont pas corrélés. En effet, de nombreux gènes dupliqués ne sont pas surexprimés. Récemment, Rossi et ses collègues (2010) ont apporté les mêmes conclusions lors de leurs études sur des souches de *S. cerevisiae* aneuploïdes obtenues à partir d'un autre système de sélection. Par contre, cette corrélation entre nombre de copies de gènes et expression a été montrée dans d'autres cellules aneuploïdes de *S. cerevisiae* (Hughes et *al.*, 2000 ; Torres et *al.*, 2007).

L'observation des gènes dérégulés sur l'ensemble du génome montre qu'un peu moins de la moitié des gènes le sont à l'état diploïde, alors que dans le contexte haploïde cette proportion est de deux tiers. Dans chaque cas, 50 % des gènes ont une expression augmentée et 50 % une expression diminuée. Cette expression différente de nombreux gènes montre que

la réponse cellulaire s'effectue au niveau de tout le génome et pas seulement au niveau des régions dupliquées. De telles expressions différentielles sur la totalité du génome sont aussi retrouvées au niveau de cellules cancéreuses (Upender et *al.*, 2004).

L'analyse des données transcriptomiques a été poursuivie en regroupant les gènes dont la transcription est augmentée ou diminuée. L'étude des fonctions et des processus dans lesquels interviennent ces gènes a été réalisée à l'aide du programme Gotermfinder (Boyle et *al.*, 2004). Que ce soit en contexte haploïde ou diploïde, les gènes surexprimés codent des protéines intervenant dans le métabolisme énergétique, notamment au niveau du cycle de Krebs et de la phosphorylation oxydative. Ceci suppose une production accrue d'ATP comme nous le mentionnions après les calculs des ratios Q_{Glu}/π_{EtOh} . Cet excès d'ATP peut être utilisé pour la réplication, la transcription ou la traduction des gènes additionnels. Par contre, cet excès d'énergie peut aussi perturber la cellule comme cela est décrit pour les souches disomiques de *S. cerevisiae*. Dans ce cas de figure, on parle alors de stress aneuploïdique (Williams et Amon., 2009).

Plus de 900 gènes sont décrits comme des éléments de réponse au stress ESR chez *S. cerevisiae* (Gasch et *al.*, 2000). Un test de Student sur ces 900 gènes dans nos révertants montre que 354 gènes ont une expression différente chez ces souches par rapport aux cellules de référence. Cette étude montre clairement que la ploïdie est un paramètre important dans la réponse cellulaire. Par exemple, de nombreux gènes impliqués dans la machinerie traductionnelle sont surexprimés chez les révertants haploïdes et sous-exprimés chez les diploïdes.

Les travaux de Gygi et *al.* (1999) et Ghaemmaghami et *al.* (2003) ont montré que les taux de transcits et de protéines sont corrélés. Si on applique ce principe, il devrait y avoir une quantité augmentée de molécules traduites dans les cellules possédant un réarrangement chromosomique. Cette augmentation peut induire un déséquilibre de la balance protéique. Celle-ci est définie comme un équilibre de protéines requis pour la croissance cellulaire. Deux hypothèses ont été émises pour comprendre comment une cellule ayant un chromosome surnuméraire est capable d'équilibrer la balance protéique : (i) une baisse de la traduction ou (ii) une dégradation de l'excédant de protéines (Torres et *al.*, 2008). Dans notre étude, le contexte de ploïdie influence clairement le type de réponse suite à un réarrangement chromosomique. À l'état diploïde, la balance protéique est probablement équilibrée par une

baisse de la traduction. Par contre, chez les révertants haploïdes, les gènes impliqués dans le mécanisme de traduction sont surexprimés de même que des gènes intervenant dans la régulation ou l'activité du protéasome. Ceci suppose une augmentation de la traduction et de la dégradation des protéines par le protéasome pour équilibrer la balance protéique. La hausse de la traduction à l'état haploïde est notamment confirmée par la sensibilité des souches aneuploïdes à la rapamycine, un inhibiteur de la traduction.

Les modifications métaboliques, notamment au niveau de la chaîne respiratoire, peuvent également être liées au gène URA2 et à la voie de biosynthèse des nucléotides pyrimidiques. En effet, l'utilisation de l'antimycine A, un composé bloquant le complexe III de la chaîne respiratoire, inhibe la voie de biosynthèse des pyrimidines (Khutornenko et al., 2010). Nous pouvons ainsi supposer qu'une modification dans cette voie de biosynthèse peut induire des changements dans le métabolisme respiratoire. Le système de sélection utilisé nécessite une réactivation de la région ATCase du gène URA2 car les fonctions GATase et CPSase sont complémentées par les protéines Cpa1p et Cpa2p qui interviennent dans la voie de biosynthèse de l'arginine (Messenguy et al., 1983). Les données transcriptomiques indiquent que ces deux gènes sont surexprimés pour produire plus de carbamylphosphate, ainsi que ceux impliqués dans les réactions à partir de l'acide uréidosuccinique. Les gènes CPA1 et CPA2 peuvent être régulés par 2 mécanismes : une rétroinhibition en présence d'arginine (Lacroute et al., 1965) ou un contrôle général de la biosynthèse de plusieurs acides aminés (Messenguy, 1979). Les cultures cellulaires à la base de ces données ont été réalisées dans un milieu sans arginine permettant de réfuter la première hypothèse. Par contre, la seconde est possible et même renforcée parce que de nombreux gènes intervenant dans les voies de biosynthèse d'acides aminés sont surexprimés.

CHAPITRE III

IMPACTS DES GÈNES DE FUSION

Impacts des gènes de fusion

Introduction

La fusion de gènes est un événement clé dans l'évolution des protéines. Elle peut mener à une perte de fonctions selon la jonction entre deux gènes. Par exemple, dans la leucémie myélocytaire aiguë, le gène *AML1* est le récepteur de la fusion avec le gène *ETO*. Dans ce cas de figure, la jonction se fait au niveau de la séquence correspondant à la partie N-terminale de la protéine Aml1p. De ce fait, la fonction de facteur de transcription portée par cette protéine est perdue et le corépresseur Eto2p se retrouve au niveau des gènes cibles de Am1p, ce qui induit la tumorisation (pour revue, Martens et Stunnenberg, 2010). La fusion peut aussi conduire à l'émergence de protéines multifonctionnelles comme celles codées par les gènes *CAD* chez les Mammifères qui portent trois domaines assurant les premières étapes de la voie de biosynthèse des nucléotides pyrimidique. Chez *S. cerevisiae*, le gène *ARO1* catalyse cinq des sept étapes de la voie de biosynthèse du chorismate, un précurseur des acides aminés aromatiques (Duncan et *al.*, 1987). La formation de tels gènes peut être considérée comme bénéfique pour la cellule car elle permet entre autre de diminuer le nombre de facteurs régulant leur expression (Marcotte et *al.*, 1999 ; Doolittle, 1999).

Un autre exemple de protéine multifonctionnelle chez *S. cerevisiae*, est la protéine codée par le gène *URA2* avec ses domaines fonctionnels GATase-CPSase et ATCase, séparés d'un domaine cryptique DHOase-like (Souciet et *al.*, 1989) (Figure 3.1). L'aspartate-transcarbamylase (ATCase) est une enzyme qui est un modèle d'étude chez la bactérie *Escherichia coli* du fait de sa régulation allostérique (Gerhart et Schachman, 1968). Cette enzyme est activée par l'ATP et rétroinhibée par les nucléotides pyrimidiques (Figure 3.1). Elle est formée de six chaînes catalytiques et de six chaînes régulatrices.



Figure 3.1 : La voie de biosynthèse des nucléotides pyrimidiques chez S. cerevisiae.

Dans notre collection de révertants, la réactivation de l'activité ATCase conduit à la formation de gènes de fusion associés aux réarrangements chromosomiques (Roelants et *al.*, 1995 ; Welcker et *al.*, 2000 ; Tourrette et *al.*, 2007). Les gènes receveurs de la fusion ont été caractérisés et la jonction entre ces derniers et la partie ATCase a été définie au nucléotide près (Figure 3.2). Les gènes accepteurs codent pour des protéines possédant des fonctions variées, il peut s'agir de protéines ribosomiques, de facteurs de transcription ou de traduction, de protéines chaperones, etc... (Tableau 3.1). La taille de la séquence correspondant à l'ATCase est variable dans les différents gènes chimères. Il en est de même pour la région provenant du gène accepteur : certaines fusions se font au niveau des séquences promotrices alors que d'autres sont en phase avec les séquences codantes dont les tailles sont également variables. Ces différences de structure soulèvent deux questions : (i) pour un gène de fusion, l'activité enzymatique de la partie ATCase est-elle influencée par la taille de la séquence présente et par la partie de la protéine acceptrice résiduelle ? (ii) les fonctions des protéines réceptrices sont-elles majoritairement altérées ou maintenues ?



YEF3 5' -----AGATCTTGACCAACAAGCTTGGAAAGA--- 3' URA2

<u>Figure 3.2</u>: Structure d'un gène chimère. Dans cet exemple, les gènes *YEF3* et *URA2* sont fusionnés. On retrouve dans les deux séquences une microhomologie de séquence de 17 pb au niveau de laquelle la fusion est réalisée (soulignée en jaune). Le gène de fusion code une protéine qui contient le domaine ATCase de la protéine Ura2p et une partie du gène *YEF3*.

la e Fonction de la protéine codée par le gène accepteur aa)	Protéine chaperone impliquée dans la régulation des fonctions des protéines Hsp90p et Hsp70p.	Péroxyredoxine spécifique des groupements thiols. Participe à la protection contre les dommages oxydatifs.	Pompe à proton H^+ -ATPase .	Facteur d'élongation de la traduction.	Chaperone cytoplasmique de la famille Hsp90 avec des fonctions identiques à la protéine Hsp82p.	Chaperone cytoplasmique de la famille Hsp90 avec des fonctions identiques à la protéine Hsp82p.	Enolase II, enzyme qui catalyse la conversion du 2- phosphoglycérate en phosphoénolpyruvate lors de la glycolyse.	Fonction inconnue.	Protéine membraniare localisée dans les vacuoles.	Protéine faisant partie de la petite sous-unité ribosomique 40S.	Protéine intervenant dans la formation de la capside des particules pseudo-virales des rétrotransposons.	MAP kinase kinase intervenant dans les voies de signalisation suite à un choc osmotique.	Sous-unité du complexe RENT, impliqué dans la régulation du nucléole et la sortie de la télophase.	Protéine de fonction inconnue
Taille de protéine chimère (1152	859	1292	1484	1137	1137	1300	402	1129	1265)	ί	1452	547	1061
Coordonnées Ura2p (aa)	1062-2214 (2214)	1465-2214 (2214)	1352-2214 (2214)	11110-2214 (2214)	1566-2214 (2214)	1566-2214 (2214)	1350-2214 (2214)	1827-2214 (2214)	1130-2214 (2214)	999-2214 (2214)	1594-2214 (2214)	1119-2214 (2214)	1667-2214 (2214)	1183-2214 (2214)
Coordonnées ORF acceptrice (aa)	fusion avec le promoteur	1-110 (117)	1-430 (918)	1-380 (1044)	1-489 (705)	1-489 (705)	1-436 (437)	1-15 (74)	1-45 (133)	1-50 (131)	2	1-357 (668)	fusion avec le promoteur	1-30 (478)
Souch e	B038 (n et 2 n)	B199 (2n)	B259 (n et 2 n)	B284 (n et 2 n)	B066/1 (n et 2n)	B132 (2n)	B326 (2n)	te B286 (n et 2 n)	Rev27 (n et 2n)	Rev46 (n et 2 n)	Rev25 (n)	AWR58 (n et 2 n)	B283/1 (2n)	BR15/2 (2n)
Gène chimèr e	<i>YDJI</i> -ATCase	<i>AHPI</i> -ATCase	PMA1-ATCase	<i>YEF3</i> -ATCase	<i>HSC82</i> -ATCase	HSC82-ATCsae	<i>ENO2</i> -ATCa s e	YJL133c-A-ATCas	<i>SNA3</i> -ATCase	RPS22A-ATCase	<i>TyA</i> -ATCase	PBS2-ATCase	NETI-ATC as e	MTCI-ATCase

haploïde (n) ou di ploïde (2n). La fusion peut se produire au niveau du promoteur ou en phase avec la protéine réceptrice. Les coordonnées sont exprimées en nombre d'acides aminés (aa) présents au sein de la protéine chimère. La valeur entre parenthèse indique la taille de la Tableau 3.1: Les gènes chimères analysés. Les souches portent divers réarrangements chromosomiques et ont été isolées dans les contextes protéine sauvage. La protéine Ura2p est rétroinhibée par les produits terminaux de la voie de biosynthèse des nucléotides pyrimidiques, notamment l'UTP qui peut se fixer au niveau de deux sites (Figure 3.1). Ces sites sont distincts, l'un se situe au niveau de la fin de la région CPSase et l'autre au sein de l'ATCase (Jaquet et *al.*, 1995). L'analogue toxique de l'uracile, le 5-fluorouracile (5-FU), peut être utilisé pour tester le maintien ou la perte de rétroinhibition. Dans une cellule sauvage, l'UTP néosynthétisé inhibe l'activité de la protéine Ura2p, le 5-FU présent dans le milieu est incorporé dans la cellule puis métabolisé en 5-FUMP qui entre dans la voie de biosynthèse *de novo* des nucléotides pyrimidiques. Cette propriété conduit à la sensibilité d'une souche sauvage au 5-FU à une concentration de 10^{-6} M (Jund et Lacroute, 1970). En revanche, des souches qui possèdent des mutations dans le gène *URA2* conduisant à une perte de rétroinhibition mais à un maintien des fonctions enzymatiques sont résistantes jusqu'à 10^{-2} M (Jaquet et *al.*, 1993). L'analyse de la résistance au 5-FU des révertants est intéressante car le phénotype obtenu pourra nous renseigner sur la perte ou non de la rétroinhibition par l'UTP de la protéine chimère.

1. Caractérisation de la fonctionnalité des domaines des protéines chimères

1.1. Le domaine codant l'ATCase

1.1.1. Détermination de l'activité ATCase

L'ATCase est une enzyme permettant la formation d'acide uréidosuccinique (CAA) à partir d'acide aspartique comme donneur de carbone. Dans notre analyse, les activités enzymatiques ont été déterminées à partir de la méthode colorimétrique mise au point par Prescott et Jones en 1969. Cette approche repose sur une propriété du CAA : celui-ci comporte un groupement uréide qui se condense avec l'antipyrine en présence d'un acide fort pour donner un produit de couleur jaune-orangée, dont le pic d'absorbance est à 466 nm (Prescott et Jones, 1969).

Les activités enzymatiques ont été mesurées pour l'ensemble de la collection de révertants ainsi que pour les souches sauvages FL100 (haploïde) et 2nU+ (diploïde). Ces activités ont été déterminées à partir d'extraits cellulaires préparés à partir de cellules en phase exponentielle de croissance dans des milieux minimum YNB et complet YPD. L'activité est définie comme étant la quantité d'acide uréidosuccinique formé par minute par milligrammes de protéines totales contenues dans l'extrait cellulaire (nmol/min/mg).

1.1.1.1. Les souches haploïdes

Les activités enzymatiques de l'ATCase mesurées dans les révertants haploïdes cultivés en milieu complet sont toutes inférieures à celle de la souche FL100 (Tableau 3.2). Pour les révertants B286 n, Rev27 et AWR58, les activités n'ont pas pu être mesurées car les quantités de CAA produites étaient trop faibles pour être détectées.

Lorsque ces mêmes tests sont réalisés à partir de cellules ayant poussé en milieu minimum sans uracile, les activités ATCase ne sont pas différentes de celle de la souche sauvage comme le démontre un test de Student. Ces activités sont par contre majoritairement supérieures à celles réalisées en milieu complet (Tableau 3.2). Par exemple, l'activité enzymatique du révertant B259 n est augmentée d'un facteur 20 (de 1,28 à 25,67 nmol/min/mg en milieu minimum).

Souche	Gène chimère	YPD	Écart-type	YNB	Écart-type	Ratio YNB/YPD
FL100	URA2 wt	15,8	0,85	10,67	0,94	0,67
B038 n	YDJ1-ATCase	3,78	0,87	15,33	6,24	4,05
B259 n	PMA1-ATCase	1,28	0,13	25,67	13,77	20,05
B284 n	YEF3-ATCase	1,18	1,2	13,33	17,46	11,3
B066/1 n	HSC82-ATCase	10,12	2,23	10,33	8,97	1,02
B286 n	YJL133c-A-ATCase	nd	nd	18	8,16	nd
Rev27	SNA3-ATCase	nd	nd	5,67	2,62	nd
Rev46	RPS22A-ATCase	5,39	1,25	12,33	4,78	2,29
AWR58	PBS2-ATCase	nd	nd	6,97	8,18	nd
aRev25	<i>TyA</i> -ATCase	2,27	0,41	4,33	5,44	1,91

<u>Tableau 3.2</u>: Mesure des activités ATCase des souches haploïdes. Les activités enzymatiques sont définies comme étant le nombre de nmoles de CAA formées par minute par mg de protéines totales contenues dans l'extrait cellulaire (nmol/min/mg). Ces extraits cellulaires sont obtenus à partir de cellules en phase exponentielle de croissance dont la croissance a été effectuée dans un milieu complet YPD et en milieu minimum YNB. L'écart-type représente la variation des valeurs autour de la moyenne calculée. Le ratio YNB/YPD est le rapport de la moyenne des activités ATCase mesurées à en milieu minimum YNB et en milieu complet YPD.

1.1.1.2. Les souches diploïdes

De la même manière que pour les révertants haploïdes, l'activité ATCase de la souche sauvage 2nU+ et des révertants diploïdes a été déterminée. Elle est variable, quelque soit le milieu de culture utilisé (Tableau 3.3). Dans un milieu minimum YNB, les révertants B199,

B038 et B326 ont des activités ATCase largement supérieures à celle de la souche sauvage 2nU+ (respectivement 139, 67 et 37 par rapport à 2,67 nmol/min/mg).

La comparaison des activités ATCase des révertants montre globalement que cette activité est supérieure pour les souches cultivées en milieu minimum (Tableau 3.3). Ce résultat n'est pas surprenant dans la mesure où l'ATCase est nécessaire pour les cellules cultivées en milieu minimum. En milieu complet, l'uracile présent dans le milieu peut être utilisé par la voie de sauvetage pour être converti en UMP (Figure 3.1).

Souche	Gène chimère	YPD	Écart- type	YNB	Écart- type	Ratio YNB/YPD
2nU+	URA2 wt	4,37	1,04	2,67	4,62	0,61
B199	AHP1-ATCase	4,53	0,07	139	22,23	30,68
B038	YDJ1-ATCase	8,06	2,61	67	22,55	8,31
B259	PMA1-ATCase	2,23	0,92	4,67	3,51	2,09
B284	YEF3-ATCase	1,26	0,24	4,67	6,6	3,71
B066/1	HSC82-ATCase	13,3	1,08	8,33	8,02	0,63
B132	HSC82-ATCase	8,32	1,2	9,67	8,34	1,16
B326	ENO2-ATCase	5,04	1,73	37	9,9	7,34
B286	YJL133c-A-ATCase	3,7	1,41	13,5	3,95	3,65
Rev27 2n	SNA3-ATCase	nd	nd	2,67	1,89	nd
Rev46 2n	RPS22A-ATCase	2,5	1,63	9,33	6,65	3,73
BR15/2	MCT1-ATCase	2,66	1,58	3,33	4,71	1,25
B283/1	NET1-ATCase	3,34	2,18	3,67	5,15	1,1
AWR58 2n	PBS2-ATCase	nd	nd	13	2,16	nd

<u>Tableau 3.3</u>: Mesure des activités ATCase des souches diploïdes. Les activités enzymatiques sont définies comme étant le nombre de nmoles de CAA formées par minute par mg de protéines totales contenues dans l'extrait cellulaire (nmol/min/mg). Ces extraits cellulaires sont obtenus à partir de cellules en phase exponentielle de croissance dont la croissance a été effectuée dans un milieu complet YPD et en milieu minimum YNB. L'écart-type représente la variation des valeurs autour de la moyenne calculée. Le ratio YNB/YPD est le rapport de la moyenne des activités ATCase mesurées à en milieu minimum YNB et en milieu complet YPD.

La variabilité observée au niveau de ces activités ATCase peut résulter d'une modification de la conformation de l'enzyme. Elle peut également s'expliquer par une localisation modifiée des protéines chimères comparée à la localisation de la protéine Ura2p dans une cellule de *S. cerevisiae*. Une autre hypothèse pour expliquer ces différences pourrait être liée avec la quantité de protéines chimères disponibles pour former le CAA.

1.1.2. Expression des gènes de fusion

Lors de notre analyse du transcriptome présentée dans le chapitre 2, la sonde oligonucléotidique correspondant au gène *URA2* présente sur la puce à ADN se situe au niveau de la région codant ATCase. De ce fait, les valeurs obtenues pour déterminer la quantité de transcrits du gène *URA2* dans les révertants correspondent à celles des gènes de fusion.

Même si les résultats de cette analyse globale ne permettent pas de quantifier de manière absolue la transcription des gènes, ils permettent d'avoir une idée de l'expression des gènes chimères par rapport au gène URA2 des souches sauvages. Les valeurs transcriptomiques inférieures à 0 indiquent des gènes de fusion sous-exprimés par rapport à URA2 et celles supérieures à 0 représentent des gènes surexprimés (Tableau 3.4). Ces données ont été complétées et affinées par des réactions de qRT-PCR absolues en deux étapes, par la quantification d'un intermédiaire ADNc obtenu à partir des transcrits. Ces réactions ont été effectuées pour le gène URA2 des souches sauvages FL100 et 2nU+ et pour les gènes chimères. Pour ces derniers, les oligonucléotides utilisés ont permis d'amplifier une région recouvrant la jonction entre le gène receveur et la région codant l'ATCase. Le rapport de l'expression des gènes de fusion par rapport au gène sauvage permet de distinguer les gènes surexprimés (valeurs > 1) de ceux sous-exprimés (valeurs < 1) (Tableau 3.4). Ces résultats de qRT-PCR sont comparables aux résultats obtenus par l'analyse globale de la transcription et permettent une quantification précise des taux de transcrits. Par exemple, le gène de fusion PBS2-ATCase à l'état diploïde (AWR58 2n) à une valeur en transcriptome de -1,17 et le ratio de qRT-PCR est égal à 0,125 (Tableau 3.4), ce qui démontre dans chacun des cas une sousexpression de ce gène chimère par rapport au gène URA2 sauvage. En plus, le rapport de qRT-PCR permet de démontrer que le gène URA2 est 8 fois plus exprimé que le gène chimère PBS2-ATCase.

Les résultats obtenus par les réactions de qRT-PCR confirment ceux issus des analyses transcriptomiques : la majorité des gènes de fusion sont surexprimés par rapport au gène *URA2* sauvage. La corrélation des résultats pour les deux méthodes permet aussi de confirmer les valeurs de transcriptomiques obtenues par l'hybridation des ARN sur puces à ADN sur lesquels nous nous sommes beaucoup appuyés dans le chapitre précédent.

Gène chimère	Souches diploïdes	Valeur transcriptomique du gène <i>URA2</i>	qRT-PCR fusion/ <i>URA2</i>	Souches haploïdes	Valeur transcriptomique du gène <i>URA2</i>	qRT-PCR fusion/ <i>URA2</i>
AHP1-ATCase	B199	1,69				
YDJ1-ATCase	B038	1,71		B038 n	4,16	
PMA1-ATCase	B259	1,87	170	B259 n	4,3	1260
YEF3-ATCase	B284	/		B284 n	3,18	
HSC82-ATCase	B066/1	0	0,011	B066/1 n	2,07	2,2
HSC82-ATCase	B132	0,37	0,06			
ENO2-ATCase	B326	2,24				
YJL133c-A-ATCase	B286	3,89	220	B286 n	5,42	1625
SNA3-ATCase	Rev27 2n	-0,524		Rev27	0,55	
RPS22A-ATCase	Rev46 2n	-1,24		Rev46	1,91	
MCT1-ATCase	BR15/2	-0,68				
NET1-ATCase	B283/1	/				
PBS2-ATCase	AWR58 2n	-1,17	0,125	AWR58	-2,21	0,09
TYA-ATCase				Rev25	1,62	

<u>Tableau 3.4</u>: Mesure de l'expression des gènes de fusion. Les valeurs transcriptomiques représentent les données obtenues par l'analyse globale de la transcription (chapitre 2). Les colonnes qRT-PCR fusion/*URA2* correspondent aux ratios des valeurs obtenues par cette méthode de PCR quantitative des gènes de fusion par rapport au gène *URA2* sauvage. Les gènes surexprimés sont représentés en rouge, et ceux sous-exprimés en vert.

En appliquant le principe que les nombres de transcrits et de protéines sont corrélés comme affirmé par Gygi et *al.*, 1999 ; Ghaemmaghami et *al.*, 2003, nous pouvons suggérer que la quantité de protéines chimères est variable mais suffisante pour un retour à la prototrophie.

1.1.3. Localisation des protéines chimères

La majorité des protéines sont monofonctionnelles et leur fonction nécessite une localisation cellulaire précise comme les facteurs de transcription au niveau nucléaire ou les facteurs d'initiation de la traduction au niveau cytoplasmique. Par contre, la localisation de la protéine codée par le gène *URA2* est mal définie : des études ont montré que cette protéine était nucléaire, d'autres l'ont montrée dans le cytoplasme, alors que les dernières analyses la décrivent comme étant cytosolique et associée aux membranes (Nagy et *al.*, 1982 ; Benoist et *al.*, 2000). Les travaux de Vorisek et ses collègues (2002) indiquent que la protéine Ura2p peut être localisée au niveau de la membrane nucléaire. Une localisation périnucléaire peut être considérée comme logique car la protéine permet la synthèse des nucléotides pyrimidiques nécessaires à la synthèse des acides nucléiques, effectuée dans le noyau : une

proximité nucléaire permettrait un apport plus rapide de ces précurseurs nucléotidiques dans le noyau.

Les gènes chimères des révertants sont composés, outre une partie du gène *URA2*, de gènes receveurs dont les fonctions sont variées et dont les localisations cellulaires sont différentes (Tableau 3.1). Leurs positions au sein de la cellule sont diverses : on retrouve des protéines cytoplasmiques comme celles codées par les gènes *PBS2* ou *RPS22A*, d'autres sont nucléolaires comme celle codée par le gène *NET1* ou membranaires comme c'est le cas des protéines codées par les gènes *PMA1* ou *SNA3*. La localisation cellulaire de la partie ATCase ou de la protéine codée par le gène receveur de la fusion peut être modifiée suite à la formation de la protéine chimère.

1.2. Le gène receveur

La réactivation de l'ATCase est la condition *sine qua none* à un retour à la prototrophie. De ce fait, les protéines chimères possèdent une région ATCase active, mais le maintien de la fonctionnalité des gènes receveurs n'est pas connu. Ces gènes codent des protéines de fonctions variées, essentielles ou non pour la cellule. Différents cas de figures peuvent être considérés en fonction de la taille des protéines codées par ces gènes (Tableau 3.1 ; Figure 3.3) :

- La totalité ou presque de la séquence codante du gène receveur est délétée lors de la fusion. Nous retrouvons d'une part les gènes de fusion *YDJ1*-ATCase et *NET1*-ATCase où la région codant l'ATCase est fusionnée avec les séquences promotrices de ces gènes (0% de la séquence codante est présente dans le gène de fusion). D'autre part, il y a également les gènes chimères possédant moins de 20% de la séquence codante du gène accepteur comme par exemple le gène de fusion *MTC1*-ATCase.
- La séquence de la protéine codée par le gène receveur est partielle et s'étend de 20 à 40% de la protéine totale. C'est le cas des gènes de fusion SNA3-ATCase, RPS22A-ATCase, YEF3-ATCase et YJL133c-A-ATCase.
- La séquence varie de 40 à 60%. On retrouve ce cas de figure au niveau des gènes chimères *AHP1*-ATCase, *PMA1*-ATCase et *PBS2*-ATCase.

 Plus de 60% de la séquence codante des gènes receveurs reste dans la fusion. Pour les révertants B066/1, B066/1 n et B132, 69% de la séquence codante du gène *HSC82* est fusionnée à l'ATCase et 99% du gène *ENO2* dans la souche B326.

Plus la taille de la protéine acceptrice est importante au niveau de la molécule chimère, plus la fonctionnalité peut être supposée conservée. Cependant, seule une analyse fonctionnelle de la protéine chimère et la mise en évidence de la conservation de fonction est nécessaire pour valider ou réfuter cette hypothèse.

1.2.1. Analyse fonctionnelle des gènes receveurs

Les gènes receveurs de la fusion codent des protéines essentielles ou non pour la prolifération cellulaire. Des mutations dans sept des gènes non essentiels engendrent des phénotypes qui peuvent être utilisés pour déterminer la conservation de la fonctionnalité des gènes receveurs de la fusion (Tableau 3.5). Par exemple, le gène *PBS2* muté confère une sensibilité accrue vis-à-vis de la polymymixine B, un antibiotique perméabilisant la membrane plasmique (Boguslawski et Polazzi, 1987).

Pour réaliser ces tests de fonctionnalité, seul le gène de fusion doit être présent dans la cellule car une copie sauvage du gène récepteur permet le maintien de la fonction. Le révertant AWR58 isolé en contexte haploïde possède un remaniement de type délétion où le gène *PBS2* est fusionné à la région codant l'ATCase. Ce révertant haploïde peut être analysé directement car le gène *PBS2* impliqué dans la fusion avec la région codant l'ATCase est présent uniquement sous forme d'un gène de fusion. Par contre, dans les cas de duplications en contexte haploïde, une copie intacte des gènes receveurs de la fusion est toujours présente (Figure 3.4). *In vivo*, le maintien de la fonction du gène receveur est assuré dans ces révertants par la copie intacte du gène. De ce fait, la partie fusionnée ne participe pas obligatoirement à la fonction.




Figure 3.3 : Représentation schématique des protéines de fusion.

Les nombres correspondent aux coordonnées des protéines en acides aminés.

Bleu : domaine actif GATase-CPSase

Vert : domaine cryptique DHOase-like

Rouge : domaine actif ATCase

Noir : site de fixation de l'UTP pour la rétroinhibition de la protéine Ura2p

Orange : protéines codées par les gènes receveurs de la fusion

Jaune : domaines fonctionnels des protéines acceptrices

NLS : domaine de localisation nucléaire bipartite

Exemple de phénotype associé au gène accepte u r	Sensibilité à 37 °C	Sensibilité à un stress oxidati f	Léthal	Léthal	Sensibilité au molybdate	Sensibilité au molybdate	Léthal	/	Accumulation de glycogène	Sensibilité au nickel	1	Léthal	Sensibilité à la polymyxine B	Baisse de la croissance	
Fonction de la protéine codée par le gène accepteur	Protéine chaperone impliquée dans la régulation des fonctions des protéines Hsp90p et Hsp70p	Péroxyredoxine spécifique des groupements thiols Participe à la protection contre les dommages oxydatifs	Pompe à proton H+-ATPase	Facteur d'élongation de la traduction	Chaperone cytoplasmique de la famille Hsp90 avec des fonctions identiques à la protéine Hsp82p	Chaperone cytoplasmique de la famille Hsp90 avec des fonctions identiques à la protéine Hsp82p	Enolase II, enzyme qui catalyse la conversion du 2- phosphoglycérate en phosphoénolpyruvate lors de la glycolyse	Fonction inconnue	Protéine membraniare localisée dans les vacuoles	Protéine faisant partie de la petite sous-unité ribosomique 40S	Protéine intervenant dans la formation de la capside des particules pseudo-virales des rétrotransposons	MAP kinase kinase intervenant dans les voies de signalisation suite à un choc osmotique	Sous-unité du complexe RENT, impliqué dans la régulation du nucléole et la sortie de la télophase	Protéine de fonction inconnue	
Type de gène	NE	NE	Щ	Е	NE	NE	Щ	NE	NE	NE	NE	NE	Щ	NE	
Gène chimère	YDJI-ATCase	AHPI-ATCase	PMAI-ATCase	YEF3-ATCase	<i>HSC</i> 82-ATCase	HSC82-ATCsae	EN02-ATCase	YJL133c-A-ATCase	SNA3-ATCase	RPS22A-ATCase	TYA-ATCase	PBS2-ATCase	NETI-ATCase	MTCI-ATCase	
Souche	B038 (n et 2n)	B199 (2n)	B259 (n et 2n)	B284 (n et 2n)	B066/1 (n et 2n)	B132 (2n)	B326 (2n)	B286 (n et 2n)	Rev27 (n et 2n)	Rev46 (n et 2n)	Rev25 (n)	AWR58 (n et 2n)	B283/1 (2n)	BR15/2 (2n)	

on. Les phénotypes indiqués sont ceux observés en cas	NE : gène non-essentiel.
ypes des gènes receveurs de la fusion. Le	es correspondants. E : gène essentiel ; NE :
<u>Tableau 3.5 :</u> Phénoty	d'inactivation des gènes

.

Il est cependant intéressant de tester la fonctionnalité de la partie restante du gène receveur dans la fusion. En effet, l'événement de fusion peut être considéré comme une création d'un nouveau gène à deux domaines dont l'un est obligatoirement fonctionnel (ATCase) alors que rien n'est certain pour l'autre. Dans cette optique, nous avons donc entrepris de tester cette éventuelle fonctionnalité. Pour ce faire, l'inactivation de la copie intacte du gène receveur est nécessaire. Deux approches différentes peuvent être utilisées pour ces inactivations de gènes, l'une reposant sur un remplacement chromosomique par croisement, l'autre par un remplacement du gène sauvage par un allèle mutant construit *in vitro*.



<u>Figure 3.4</u>: Représentation schématique de la stratégie pour réaliser les tests phénotypiques associés aux gènes receveurs de la fusion. Dans cet exemple, le gène receveur est *AHP1*. Le gène *ura2* représente l'allèle muté *uar2*₁₅₋₃₀₋₇₂. $\Delta AHP1$ représente l'inactivation de ce gène.

1.2.1.1. Remplacement chromosomique par croisement

Pour obtenir des souches portant un allèle non fonctionnel du gène impliqué dans le gène de fusion, nous avons utilisé la collection de souches EUROSCARF. Cette dernière est composée d'un ensemble de plus de 20.000 souches de *S. cerevisiae* S288c qui portent chacune une délétion pour un gène donné. Ces délétions ont été réalisées en remplaçant le gène sauvage par un gène muté interrompu par la cassette KanMX conférant une résistance

vis-à-vis de la généticine à ces cellules (Brachmann et *al.*, 1998). Cette collection est composée de souches haploïdes pour les gènes non-essentiels et de souches diploïdes hétérozygotes pour les gènes essentiels. Dans le but d'isoler des souches haploïdes avec le réarrangement chromosomique associé et la copie intacte du gène receveur inactivée, nous avons croisé le révertant avec une souche EUROSCARF. Après méiose du diploïde obtenu, nous avons analysé les spores et conservé celles ayant un phénotype Ura⁺ KanR. En effet, de tels clones conservent le gène de fusion ainsi que le réarrangement chromosomique associé et possèdent également la copie du gène receveur avec la cassette KanMX intégrée (Figure 3.5).



<u>Figure 3.5</u>: Représentation schématique du remplacement du chromosome porteur de la copie intacte du gène receveur de la fusion par croisement avec une souche EUROSCARF. Dans cet exemple, nous avons représenté le remplacement du chromosome XIV chez le révertant B038 n. Les génotypes de chaque souche sont indiqués à chaque étape de ce remplacement chromosomique.

Cette technique ne peut être utilisée que pour les révertants ayant un chromosome surnuméraire. En effet, les réarrangements intrachromosomiques comme les duplications segmentales ne pourront pas être utilisées avec cette méthode : le chromosome X comportant le gène de fusion possède également l'allèle intact du gène receveur et un remplacement de ce chromosome par celui ayant la cassette KanMX engendrerait une perte du gène chimère.

1.2.1.2. Remplacement du gène sauvage par un allèle mutant construit in vitro

La seconde méthode que nous avons envisagée menant à la délétion d'une copie intacte du gène receveur est aussi basée sur l'utilisation de la collection de souches EUROSCARF (Brachman et *al.*, 1998). Cette approche peut être utilisée pour tous les révertants, aussi bien ceux porteurs de remaniements interchromosomiques comme les aneuploïdies, que ceux ayant des remaniements intrachromosomiques comme les duplications segmentales. Pour un gène donné, l'allèle interrompu par la cassette KanMX présent dans la souche correspondante de la collection EUROSCARF est amplifiée par PCR en utilisant des amorces situées à 100 pb en amont et en aval de celle-ci. Les fragments d'ADN obtenus sont alors utilisés pour remplacer la copie intacte du gène receveur dans le révertant par l'allèle contenant la cassette KanMX suite à une intégration par double recombinaison homologue (Figure 3.6). Néanmoins, après plusieurs tentatives et l'utilisation de deux méthodes de transformation différente (électroporation et transformation chimique par l'acétate de lithium), aucun transformant n'a pu être obtenu. De ce fait, nous n'avons pas isolé de souches possédant les copies intactes des gènes receveurs interrompus par la cassette KanMX pour tester la fonctionnalité des gènes accepteurs de la fusion.

Malgré l'absence de transformants issus de l'interruption de gènes par double recombinaison homologue, deux gènes de fusion ont été analysés : l'un obtenu suite à un remplacement chromosomique et l'autre par analyse directe de phénotypes associés.

1.2.1.3. Le gène de fusion YDJ1-ATCase

Ce gène chimère est présent dans le révertant diploïde B038 et haploïde B038 n. La jonction entre *YDJ1* et la région codant l'ATCase est réalisée au niveau du promoteur du gène receveur. De ce fait, la fonction de *YDJ1* est perdue dans le gène de fusion. Nous avons néanmoins vérifié cette perte de fonction par un remplacement chromosomique par croisement ce qui a surtout permis de valider cette approche. La protéine Ydj1p est une

protéine chaperone impliquée dans la régulation des fonctions des autres protéines chaperones Hsp70p et Hsp90. Une souche ayant cette protéine mutée est thermosensible (Tableau 3.5) (Caplan et *al.*, 1992). Pour effectuer ce remplacement chromosomique, une souche EUROSCARF, de génotype *leu2⁻ lys2⁻ his3⁻ ura3⁻ YDJ1*::KanMX, est croisée à deux reprises avec des souches isogéniques à FL100 de génotype approprié pour isoler une souche de génotype *ura2⁻ ura3⁺ YDJ1*::KanMX (Figure 3.5). Cette dernière est alors croisée avec le révertant B038 n. Après méiose, les spores Ura⁺ résistantes à la généticine (génotype *ura2+ YDJ1*::KanMX conservant le gène de fusion) sont isolées et utilisées pour déterminer le phénotype de cette souche pour le gène *YDJ1* (Figure 3.5). Cette souche haploïde n'est pas viable à 37°C ce qui confirme bien la non fonctionnalité de *YDJ1* au sein du gène de fusion. Cette analyse a surtout permis de valider l'approche de remplacement chromosomique.

Cette approche a été tentée sur deux autres révertants. Elle n'a pas abouti du fait d'une absence de sporulation des diploïdes résultant du croisement de la souche EUROSCARF avec les révertants correspondants.



Figure 3.6 : Représentation schématique de l'inactivation du gène *SNA3* par double recombinaison homologue.

1.2.1.4.Le gène de fusion PBS2-ATCase

Le gène chimère *PBS2*-ATCase a été isolé en contexte haploïde (révertant AWR58) et résulte d'un événement de délétion où la région codant l'ATCase est fusionnée au gène *PBS2*. La protéine Pbs2p est une MAP kinase kinase qui intervient dans des voies de signalisation cellulaire suite à un choc osmotique en phosphorylant la protéine Hog1p (Schüller *et al.*, 1994). Cette dernière active de nombreux facteurs de transcription comme les protéines Mns2p et Mns4p qui sont impliquées dans la réponse cellulaire suite à un stress (Gasch, 2007).

Le gène *PBS2* muté entraîne une baisse de résistance à un choc osmotique et une sensibilité accrue vis-à-vis de la polymyxine B (Tableau 3.5) (Boguslawski et Polazzi, 1987). La résistance au choc osmotique a été testée sur des milieux complets supplémentés de concentrations croissantes de NaCl variant de 0 à 0,5 M. De manière similaire, la sensibilité à la polymyxine B a été testée en supplémentant des milieux complets avec des concentrations variables allant de 0 à 288 mM de cet antibiotique. Pour ces deux phénotypes, nous avons utilisé quatre dilutions différentes d'une suspension de cellules. Les résultats montrent que le révertant AWR58 est plus sensible à un choc osmotique et à la polymyxine B que la souche de référence FL100 (Tableau 3.6). Ces sensibilités prouvent que la partie de la protéine Pbs2p de la protéine chimère Pbs2p-ATCase n'est plus fonctionnelle.

La protéine Pbs2p possède un domaine kinase entre les acides aminés 360 et 623 qui permet son activité de phosphorylation (Figure 3.3). Lors de la fusion avec la région ATCase de la protéine Ura2p, seuls les 357 premiers résidus de Pbs2p sont conservés. Le domaine kinase est donc perdu, ce qui explique sa perte de fonctionnalité au sein de la protéine chimère.

Souches			Polymyxine	e B (mM)		
	0	7,2	36	72	144	288
FL100	++++	++++	++++	++++	++	
AWR58	++++	+++-	+++-	++		

Souches			NaCl (M)		
	0	0,05	0,01	0,25	0,5
FL100	++++	++++	++++	++++	++++
AWR58	++++	++++	++++	++	

<u>Tableau 3.6 :</u> Analyse fonctionnelle du gène chimère *PBS2*-ATCase du révertant AWR58. Ces tests ont été réalisés avec quatre dilutions différentes d'une suspension de cellules.

++++ : croissance des 4 dilutions

+++- : croissance de 3 dilutions

++-- : croissance de 2 dilutions

---- : croissance d'aucune dilution

Lors de cette étude, nous n'avons pas pu inactiver la copie active du gène receveur par intégration de la cassette KanMX. De plus, nous n'avons pu effectuer de remplacement chromosomique que pour une souche du fait de la faible sporulation des révertants croisés avec une souche EUROSCARF. De ce fait, nous avons testé la fonctionnalité uniquement de deux gènes receveurs de la fusion. Ainsi, il nous reste cinq gènes receveurs non-essentiels à analyser dont deux plus particulièrement du fait de la longueur de séquence conservée dans le gène de fusion.

2. Caractérisation de la rétroinhibition des protéines chimères

Les activités CPSase et ATCase sont rétroinhibées par la fixation de l'UTP sur deux sites distincts de la protéine Ura2p (Jaquet et *al.*, 1995). Le premier est localisé à la fin du domaine CPSase alors que le second est dans la partie ATCase (Figure 3.3). Les travaux d'Antonelli et *al.* (1998) montrent que le site localisé dans la région CPSase est majoritairement responsable de la baisse de l'activité ATCase lors de la rétroinhibition. La taille de la partie du gène *URA2* impliquée dans la fusion est variable dans les différents révertants. Ainsi, on retrouve au niveau de la partie C-terminale de la protéine de fusion entre 17 % et 55 % de la protéine Ura2p. De ce fait, des révertants n'ont pas le site de rétroinhibition de la région CPSase (révertants B199 et B326), d'autres l'ont en partie

seulement (B284 et AWR58) et d'autres enfin le possèdent entièrement comme les révertants B038 et Rev46 par exemple (Figure 3.3).

La rétroinhibition peut être testée en utilisant le 5-fluoro-uracile (5-FU), un analogue toxique de l'uracile. Ce composé entre dans la cellule par la perméase à l'uracile et peut être métabolisé en 5-FUMP, 5-FUDP puis 5-FUTP et 5-FCTP (Figure 3.1). Ces nucléotides pyrimidiques fluorés sont toxiques pour la cellule. Si la protéine Ura2p est rétroinhibée par l'UTP, la voie de biosynthèse sera ralentie et les taux d'UMP et de nucléotides pyrimidiques formés seront moindres. De ce fait, le 5-FU qui sera notamment métabolisé en 5-FUTP conduira à une toxicité pour la cellule. Par contre, une souche possédant une protéine Ura2p ayant perdu la rétroinhibition par l'UTP est plus résistante vis-à-vis du 5-FU. En effet, les intermédiaires métaboliques comme l'UMP seront continuellement synthétisés et entreront en compétition avec les composés fluorés ce qui diluera l'effet du 5-FU et de ses dérivés. Un mutant de rétroinhibition a donc un phénotype de résistance vis-à-vis de cet analogue toxique (Jund et Lacroute, 1970).

2.1. Résultats

Les souches sauvages FL100 et 2nU+ sont sensibles au 5-FU car la protéine Ura2p sauvage est rétroninhibée : elles ne se développent qu'à des faibles concentrations de 5-FU de l'ordre de 10^{-6} M (Jund et Lacroute, 1970). Par contre, un mutant de rétroinhibition est résistant à des concentrations pouvant aller jusqu'à 10^{-2} M (Jund et Lacroute, 1970, Jaquet et *al.*, 1993). La résistance au 5-FU des souches révertantes a été testée en milieux minima additionnés de concentrations croissantes de ce composé (Tableau 3.7). Les résultats montrent pour les souches de la collection des variabilités de résistance au 5-FU allant de 10^{-2} M à 5. 10^{-5} M (Tableau 3.7).

Ces variations de résistances peuvent s'expliquer par une perte de rétroinhibition mais aussi par un import moindre de 5-FU dans ces cellules. Une autre hypothèse, qui reposerait sur une résistance causée par une baisse de l'activité ATCase de la protéine chimère et de la formation de moins d'UTP pour sa rétroinhibition, peut être écartée. En effet, l'activité ATCase est similaire pour la majorité des révertants par rapport aux cellules sauvages, ce qui permet d'exclure cette hypothèse.

2.2. Caractérisation de l'absence de rétroinhibition

Comme nous venons de le voir ci-dessus, deux sites de fixation de l'UTP ont été mis en évidence au niveau de la protéine Ura2p, celui présent au niveau de la partie CPSase permettant d'assurer seul l'essentiel de la rétroinhibition de l'activité ATCase (Jaquet et *al.*, 1995 ; Antonelli et *al.*, 1998). Les révertants dans lesquels ce site de rétroinhibition n'est plus présent au niveau de leur protéine chimère ont une résistance au 5-FU jusqu'à des concentrations de 10^{-2} M, confirmant l'absence de rétroinhibition de la région CPSase comme les révertants Rev27 et AWR58, les résistances sont variables allant de 10^{-2} M à 10^{-5} M. Chez les révertants dans lesquels les protéines chimères possèdent les deux sites de fixation de l'UTP (les souches B038, B038 n, Rev46 et Rev46 2n), cette résistance est plus faible et va de 10^{-4} M à 5.10^{-5} M, sauf pour la souche Rev46 qui résiste à 10^{-2} M de 5-FU. La sensibilité à une concentration à 5.10^{-5} M est comparable à celle des souches sauvages montrant que la rétroinhibition est vraisemblablement maintenue dans la protéine de fusion. Une résistance à 10^{-4} M suggère une rétroinhibition de la protéine chimère mais qui serait moins efficace que celle de la protéine native.

Les travaux menés par Jund et Lacroute (1970) ont montré que des souches possédant des mutations dans les gènes FUR4 et FUR1 étaient résistantes au 5-FU à des concentrations allant de 10⁻⁴ M à 10⁻³ M. Les protéines codées par ces deux gènes interviennent dans la voie de sauvetage de l'uracile : le gène FUR4 code la perméase à l'uracile alors que FUR1 code l'UPRTase catalysant la formation d'UMP à partir d'uracile exogène (Figure 3.1). Les résultats de transcriptomique présentés dans le chapitre précédent montre que l'expression du gène FUR4 est diminuée dans la majorité des révertants par rapport aux souches sauvages. La transcription de FUR1 est quant à elle inchangée dans ces souches. Il n'est donc pas exclu que la résistance au 5-FU des révertants B038 et B038 n à 10⁻⁴ M soit due à une baisse de l'import de cet analogue toxique dans la cellule (Tableau 3.8).

Le cas particulier de la résistance au 5-FU à une concentration de 10^{-2} M du gène de fusion *RPS22A*-ATCase à l'état diploïde, malgré la présence des deux sites de rétroinhibition, pourrait être liée à une éventuelle non accessibilité de ces sites de fixation de l'UTP dans la protéine de fusion.

Gène chimère	Site CPSase	Souches diploïdes	Résistance au 5-FU (M)	Souches haploïdes	Résistance au 5-FU (M)
URA2 wt	Entier	2nU+	1.10-6	FL100	1.10 ⁻⁶
AHP1-ATCase	Absent	B199	1.10^{-2}		
YDJ1-ATCase	Entier	B038	1.10 ⁻⁴	B038 n	1.10 ⁻⁴
PMA1-ATCase	Absent	B259	1.10^{-2}	B259 n	1.10^{-2}
YEF3-ATCase	Partiel	B284	1.10 ⁻²	B284 n	1.10 ⁻²
HSC82-ATCase	Absent	B066/1	1.10^{-2}	B066/1 n	1.10^{-2}
HSC82-ATCase	Absent	B132	1.10^{-2}		
ENO2-ATCase	Absent	B326	1.10^{-2}		
YJL133c-A- ATCase	Absent	B286	1.10 ⁻²	B286 n	1.10 ⁻²
SNA3-ATCase	Partiel	Rev27 2n	1.10 ⁻⁵	Rev27	1.10 ⁻⁵
RPS22A-ATCase	Entier	Rev46 2n	5.10 ⁻⁵	Rev46	1.10 ⁻²
MCT1-ATCase	Partiel	BR15/2	1.10^{-2}		
NET1-ATCase	Absent	B283/1	1.10 ⁻²		
PBS2-ATCase	Partiel	AWR58 2n	5.10-4	AWR58	1.10 ⁻³
TYA-ATCase	Entier			Rev25	1.10 ⁻⁵

<u>Tableau 3.7</u>: Analyse de la résistance au 5-FU de la collection de révertants. Les concentrations de 5-FU sont exprimées en molaire. Le dégradé de couleur correspond aux concentrations de résistance au 5-FU variables des souches : le rouge montre une souche résistante à 10^{-2} M alors que le jaune montre une souche résistante à 10^{-6} M.

Souches	FUR4	FUR1	Souches	FUR4	FUR1
B199	-1,068	0			
B038	-1,98	0	B038 n	-1,34	0
B259	0	0	B259 n	-1,96	0,52
B066/1	-0,123	0	B284 n	-2	1,055
B286	-1,467	0	B286 n	-1,935	0
Rev46 2n	-1,17	1,32	Rev27	-1,67	0
AWR58 2n	0	0	AWR58	-1,06	0
BR15/2	-0,53	0	Rev25	-2,695	0,715

<u>Tableau 3.8</u>: Analyse de la transcription des gènes *FUR1* et *FUR4* dans les révertants. Les valeurs indiquées proviennent de l'analyse globale de la transcription. Les données négatives représentent une transcription diminuée, alors que les valeurs positives représentent une transcription augmentée par rapport à la souche sauvage.

Conclusion

Le système de sélection basé sur l'allèle mutant du gène *URA2* a permis d'obtenir des souches avec différents réarrangements chromosomiques contenant un gène chimère dans lequel la région codant l'ATCase est fusionnée à un autre gène. Cette fusion peut se faire au niveau du promoteur ou en phase avec la séquence codante d'un gène qui peut être essentiel ou non pour la croissance. Les gènes receveurs codent des protéines de fonctions variées. Tous les gènes chimères obtenus suite au crible puis analysés dans notre étude ont un point commun : ils sont fusionnés avec une région de taille variable du gène *URA2* codant au minimum pour la partie ATCase.

L'activité enzymatique de l'ATCase, mesurée dans les souches qui portent ce gène chimère, est variable en fonction des différents gènes de fusion et des conditions de culture utilisées. Les valeurs d'activités sont augmentées lorsque les souches sont cultivées sur un milieu minimum en absence d'uracile. Les différences d'activités observées peuvent être liées à différents facteurs comme le nombre de protéines chimères présentes, l'accessibilité du site catalytique suite à un changement conformationnel de la protéine chimère par rapport à la protéine sauvage ou encore par une localisation cellulaire modifiée.

L'analyse globale de la transcription montre que ces gènes de fusion sont principalement surexprimés par rapport au gène *URA2* présent dans une cellule sauvage. Cette augmentation de la transcription est probablement nécessaire à la réactivation de l'ATCase car une modification de la structure de la protéine chimère ou de sa localisation peut entraîner une baisse de l'activité de l'enzyme. En effet, comme nous l'avons mentionné précédemment, l'ATCase est une enzyme allostérique qui contient six sous-unités catalytiques et six sousunités régulatrices. Le repliement ou la cohésion de ces sous-unités au sein de la protéine chimère peuvent être altérés du fait de la présence d'autres domaines peptidiques dans la partie N-terminale de la protéine chimère. De la même manière, une mauvaise localisation de l'ATCase dans la cellule peut également perturber son activité car l'enzyme peut être moins accessible aux différents substrats de la catalyse enzymatique. Chez *S. cerevisiae*, les études de localisation placent l'ATCase dans le cytoplasme et dans les membranes, notamment périnucléaires (Vorisek et *al.*, 2002). Des gènes récepteurs de la fusion codent des protéines possédant des localisations cellulaires spécifiques comme nucléaire ou membranaire. Par exemple, la protéine Hsc82p possède un domaine protéique NLS (Nuclear Localisation Site) qui place cette protéine dans le noyau de la cellule. En fonction de la localisation de la protéine acceptrice et du maintien de domaines protéiques, l'ATCase peut ainsi changer de localisation. Par exemple, les gènes de fusion *PMA1*-ATCase et *SNA3*-ATCase (révertants B259 et Rev27) codent des protéines qui possèdent de nombreux domaines transmembranaires, ce qui peut placer l'ATCase au niveau des membranes vacuolaires ou mitochondriales.

La partie C-terminale des protéines chimères est composée du domaine ATCase actif, ce qui est essentiel au retour à la prototrophie. Par contre, la partie N-terminale correspond à la protéine acceptrice et le maintien de sa fonctionnalité n'est pas obligatoire, la fonction du gène receveur étant assuré par la copie intacte. Il est cependant intéressant de tester la fonctionnalité de la partie restante du gène receveur dans la fusion. En effet, l'événement de fusion peut être considéré comme une création d'un nouveau gène à deux domaines dont l'un est obligatoirement fonctionnel (ATCase) alors que rien n'est certain pour l'autre. L'analyse de la fonctionnalité de ces protéines a été testée pour deux gènes de fusion : YDJ1-ATCase (souches B038 et B038 n) et PBS2-ATCase (révertants AWR58 et AWR58 2n). Dans le premier exemple où la partie ATCase est fusionnée au niveau du promoteur du gène YDJ1, la perte de fonction est évidente et a été vérifiée dans une souche transformée par un remplacement chromosomique par croisement. Cette analyse a surtout permis de valider la technique de remplacement de chromosome. Dans le second exemple où la région codant l'ATCase est fusionnée en phase avec la séquence codante du gène PBS2, la sensibilité accrue à la polymyxine B et à la pression osmotique démontre que la fonction kinase de la protéine Pbs2p est perdue. Cette perte s'explique aisément par l'absence du domaine kinase au niveau de la protéine chimère correspondante.

Comme nous l'avons déjà mentionnée, la taille de la partie C-terminale de la protéine chimère est variable. En fonction de cette taille et du maintien de domaines protéiques, nous avons défini un révertant pour lequel l'activité du gène receveur est supposée conservée et trois qui devraient être testés. Le révertant B326 possède le gène de fusion *ENO2*-ATCase où 99% de *ENO2* est encore présent. Il est vraisemblable que la fonction phosphopyruvate hydratase de la protéine Eno2p soit maintenue. Il faudra cependant le tester. Le gène chimère *HSC82*-ATCase possède 69% du gène receveur. Les domaines peptidiques importants de la protéine correspondante sont encore présents. Au niveau du gène de fusion *PMA1*-ATCase,

les motifs protéiques de la protéine native Pma1p sont conservés mais des domaines transmembranaires sont délétés ce qui peut modifier la conformation de la protéine. De la même manière, pour le gène de fusion *SNA3*-ATCase, les motifs protéiques de la protéine Sna3p sont conservés au niveau de la protéine chimère mais des domaines transmembranaires sont délétés. Il serait donc intéressant de tester le maintien ou la perte fonctionnalité dans les trois révertants porteurs de ces gènes de fusion.

Au niveau de l'activité enzymatique de l'ATCase, un équilibre entre l'expression des gènes de fusion, la conformation de la protéine chimère et leur localisation semble donc être à la base de la restauration de cette activité et de sa variabilité. Au stade de notre travail, aucunes réponses expérimentales définitives ne sont pour l'instant apportées pour caractériser l'emplacement cellulaire de la protéine de fusion et la conformation des protéines chimères.

La conformation de la protéine Ura2p sauvage permet de fixer l'UTP au niveau de deux sites de rétroinhibition distincts : un dans l'ATCase et l'autre au niveau de la fin du domaine CPSase qui permet la rétroinhibition de l'ATCase (Jaquet et al., 1995 ; Antonelli et al., 1998). Le site ATCase est présent au niveau de toutes les protéines chimères mais celui de la CPSase peut être présent totalement, partiellement ou absent. La rétroinhibition est testée en utilisant le 5-FU, un analogue toxique de l'uracile. Une cellule ayant la protéine Ura2p sauvage est sensible jusqu'à une concentration de 10⁻⁶ M de ce composé alors qu'un mutant de rétroinhibition l'est à 10⁻² M (Jund et Lacroute, 1970 ; Jaquet et al., 1993). Les résultats que nous avons obtenus montrent une variabilité de résistance au 5-FU des révertants allant de 10⁻ ² M à 10⁻⁵ M. Les protéines chimères possédant seulement le site de fixation de l'UTP dans la région codant l'ATCase sont résistantes à 10⁻² M ce qui montre une absence de rétroinhibition. Dans les protéines de fusion possédant les deux sites de fixation de l'UTP. cette résistance est accrue et peut aller jusqu'à 10⁻⁵ M montrant que la rétroinhibition est présente au niveau de ces protéines chimères. Par contre, d'autres protéines de fusion possédant ces deux sites de fixation de l'UTP sont résistantes à 10⁻⁴ M de 5-FU. Cette augmentation de résistance peut provenir d'une baisse de l'import de 5-FU dans la cellule. L'analyse globale de la transcription que nous avons exposée dans le chapitre 2 montre une diminution de l'expression du gène FUR4 codant la perméase à l'uracile, ce qui est consistant avec l'hypothèse d'une baisse de l'import de 5-FU.

CONCLUSIONS ET PERSPECTIVES

Conclusions et perspectives

Dans cette première partie de notre travail, nous nous sommes attachés à comprendre les impacts que peuvent engendrer des remaniements chromosomiques et la présence gènes de fusion sur la physiologie de la cellule. Nous présenterons ici en guise de conclusion un résumé des principaux résultats et proposerons des approches complémentaires qui permettraient de compléter les travaux que nous venons de présenter dans les trois premiers chapitres.

Les réarrangements chromosomiques apparaissent naturellement dans des génomes comme celui de S. cerevisiae avec une fréquence d'apparition de l'ordre d'un événement toutes les 10¹⁰ à 10¹² cellules générées. Afin d'isoler des souches de S. cerevisiae possédant de tels remaniements, différents cribles de sélection ont été mis au point dont notamment celui utilisé au sein de notre laboratoire. Ce dernier repose sur l'utilisation d'un allèle mutant du gène URA2 contenant deux mutations non-sens et une délétion d'un nucléotide engendrant un décalage du cadre de lecture lors de la traduction. La présence des trois mutations dans l'allèle $ura2_{15-30-72}$ engendre la formation d'une protéine Ura2p tronquée et non active, induisant une auxotrophie vis-à-vis de l'uracile. Cette auxotrophie résulte de l'absence de l'activité enzymatique ATCase car les deux autres activités GATase et CPSase portées par la protéine sauvage sont complémentées par les protéines Cpa1p et Cpa2p intervenant dans la voie de biosynthèse de l'arginine. Un retour à la prototrophie est possible grâce à la restauration de l'activité ATCase. Cette propriété a permis de sélectionner de nombreux révertants possédant différents types de réarrangements chromosomiques où la partie ATCase est fusionnée à un autre gène, soit en phase avec la séquence codante, soit au niveau de son promoteur. Les types de remaniements varient en fonction de la ploïdie de la souche. À l'état haploïde, des insertions de rétrotransposons Ty1 sont retrouvées de même que des délétions ou des duplications géniques et segmentales. Par contre, dans un contexte diploïde, seules des délétions ou des duplications segmentales ont été obtenues, ces dernières pouvant entraîner des événements plus complexes comme la formation de chromosomes chimères surnuméraires (aneuploïdie) ou des translocations non réciproques.

Lors de ce travail, une collection de révertants haploïdes et diploïdes ayant différents types de remaniements chromosomiques et des gènes de fusion de structures variables a été utilisée. Cette collection nous a permis d'analyser puis de comprendre les modifications moléculaires et cellulaires engendrées par les différents réarrangements chromosomiques ainsi que l'implication des protéines de fusion dans la survie cellulaire.

La comparaison morphologique des colonies montre des variations entre les révertants et les souches sauvages en fonction de la ploïdie et de la source de carbone présente dans le milieu. Ces résultats indiquent également qu'il n'y a pas une morphologie particulière qui correspond à un type de réarrangement. Sur le plan microscopique, les cellules ayant des remaniements chromosomiques ont généralement des formes plus allongées, peuvent s'agréger entre elles et ont un volume cellulaire plus important suggérant que le cycle cellulaire est modifié suite à la présence d'un remaniement. L'analyse de la croissance confirme cette hypothèse pour les révertants diploïdes qui ont un temps de génération majoritairement plus long que les cellules sauvages. Par contre, cette différence est moins marquée en contexte haploïde. Pour déterminer si les divers réarrangements chromosomiques influencent de la même manière le cycle cellulaire de *S. cerevisiae*, des analyses de cytométrie de flux sont envisagées. Cette technique permettra d'affiner nos résultats et de préciser dans quelle phase du cycle cellulaire, les révertants sont bloquées. Ils permettront également de tester l'influence de la ploïdie sur cet arrêt.

L'analyse globale de la transcription de nos révertants par rapport aux souches sauvages renforce l'idée que la ploïdie influence la réponse cellulaire. Par exemple, les gènes intervenant dans la biogenèse du ribosome et la machinerie traductionnelle ont une transcription diminuée dans les révertants diploïdes et augmentée dans les souches haploïdes. De manière identique, la transcription des gènes impliqués dans l'activité du protéasome est augmentée chez les révertants haploïdes. Ces résultats indiquent que l'équilibre protéique nécessaire à la prolifération cellulaire est maintenu en baissant la traduction chez les souches diploïdes. À l'état haploïde, cet équilibre est maintenu en dégradant l'excès de protéines synthétisées. Des analyses complémentaires sont envisagées afin de confirmer ces hypothèses comme par exemple l'utilisation de la bélactosine A, une drogue inhibant l'activité protéolytique (Asai et *al.*, 2004). Une baisse de la viabilité ou de la croissance cellulaire des révertants haploïdes uniquement permettrait de montrer que la protéolyse est essentielle au

maintien de la balance protéique chez les souches haploïdes. De plus, une étude protéomique permettrait de mettre en évidence les corrélations que nous avons parfois supposées entre les gènes dont la transcription est modifiée et leur traduction. Cette approche permettrait surtout une analyse différentielle des protéomes des révertant à partir de gels de protéines en deux dimensions. Une autre méthode envisageable est l'utilisation de la spectrométrie de masse quantitative à haute résolution. Une telle technique est déjà utilisée pour déterminer les variations des taux de protéines et leur corrélation avec les gènes dont le nombre de copies est variable (CNV) au sein de cellules cancéreuses (Geiger et *al.*, 2010). Chez des souches de *S. cerevisiae* aneuploïdes, il a été récemment rapporté par des études d'analyse globale de la transcription couplées à de la spectrométrie de masse quantitative que le nombre de gènes CNV, leur expression et leur taux de protéines étaient parfaitement corrélés (Li et *al.*, 2010, communication en congrès).

Même si la taille de la partie ATCase fusionnée est variable entre les gènes chimères, cette région est active, car c'est la condition sine qua none pour un retour à la prototrophie. Néanmoins, cette activité enzymatique est variable en fonction du milieu et est plus élevée quand il n'y a pas d'uracile extracellulaire. Cette variation peut être due à différents paramètres liés à l'expression des gènes codant ces protéines chimères, la localisation cellulaire ou encore la conformation de ces protéines. Les données issues de l'analyse transcriptomique montrent une expression variable des gènes de fusion dans les révertants par rapport au gène URA2 dans une souche sauvage. Les taux d'ARNm et de protéines étant supposés corrélés chez S. cerevisiae (Ghaemmaghami et al., 2003 ; Gygi et al., 1999), ces résultats suggèrent que la quantité de protéines chimères est également variable. Dans certains cas comme celui du révertant B286, l'augmentation de l'expression n'est pas corrélée avec une hausse de l'activité ATCase. Ces résultats indiquent que le nombre de protéines n'est pas le seul paramètre influençant l'activité de cette enzyme. Afin de comprendre les autres paramètres, une approche bioinformatique utilisant des logiciels comme MODELLER peut être utilisée pour définir une conformation des protéines (Eswar et al., 2008). Ce logiciel permet de prédire une structure protéique à partir de séquences nucléotidiques.

La localisation cellulaire des protéines chimères peut être déterminée selon deux méthodes : l'immuno-marquage sur des cellules entières et l'utilisation de gènes de fusion où la GFP est greffée à un gène chimère. Dans le premier cas, des anticorps primaires (I) sont utilisés pour reconnaître le domaine ATCase et des anticorps secondaires (II) pour détecter les chaînes lourdes des immunoglobulines I. Ces anticorps II peuvent être couplés à des fluorophores qui seront utilisés pour déterminer la localisation cellulaire de l'ATCase. Pour réaliser une telle approche, il faudrait obtenir des anticorps contre l'ATCase, les purifier et vérifier la spécificité de ces immunoglobulines. Dans le cas de l'utilisation des gènes de fusion, il faut greffer le gène codant la protéine GFP (ou une protéine équivalente) au niveau de la partie ATCase. La microscopie par épifluorescence peut alors être utilisée pour localiser les protéines chimères dans la cellule.

Au niveau des gènes de fusion où la jonction se fait en phase avec le cadre de lecture, les protéines chimères ont une partie C-terminale constante qui correspond à la région ATCase active. Par contre, la partie N-terminale est variable et est codée par le gène receveur de la fusion. Ces derniers codent des protéines de fonctions diverses, pouvant être essentielles ou non. La fonctionnalité de deux gènes receveurs non-essentiels sur sept a été testée expérimentalement. Une des souches analysée a été obtenue suite à un remplacement chromosomique par croisement. La seconde a été utilisée directement pour les analyses phénotypiques. Ainsi, la fonctionnalité de cinq gènes receveurs de la fusion reste à être analysée dont deux particulièrement intéressants du fait de la longueur de la séquence présente dans le gène chimère. Pour réaliser ces tests, une approche alternative consisterait à cloner la séquence du gène de fusion dans un plasmide. Ce plasmide pourrait alors être introduit dans une souche EUROSCARF qui servirait pour les analyses phénotypiques.

À plus long terme, la comparaison des souches ayant des réarrangements chromosomiques suite à diverses pressions de sélection semble nécessaire pour s'affranchir des contraintes de sélection. Au laboratoire, un autre crible de sélection est utilisé pour la souche *S. cerevisiae* S288c. Dans cette souche, les deux gènes *URA3* et *TRP1* sont insérés au niveau du chromosome I. La protéine Ura3p permet la métabolisation de l'acide 5-fluoroorotique (5-FOA) qui est toxique pour la cellule. De manière similaire, le gène *TRP1* actif rend une souche sensible vis-à-vis de l'acide fluoroanthranilique (FAA). Les sélections dans un milieu contenant le 5-FOA et le FAA ont été réalisées et les révertants obtenus analysés. Ces souches possèdent une délétion des gènes *URA3* et *TRP1*, mais également d'autres réarrangements chromosomiques et notamment des translocations et des gains de chromosomes. Une analyse de ces révertants permettrait d'avoir des résultats dans des contextes génétiques différents et avec des pressions de sélections variables.

Partie 2

Comparaison des génomes mitochondriaux des levures hémiascomycètes

Comparaison des génomes mitochondriaux des levures hémiascomycètes

Introduction

Dans le début des années 1950, Ephrussi et Slonimski ont isolé des mutants de Saccharomyces cerevisiae appelés « petites » notamment en raison de la taille de la colonie sur milieu gélosé contenant du glucose en quantité limitante et du glycérol (Ephrussi et Slonimski, 1955). L'étude génétique de ces mutants a montré que le mode de transmission du caractère « petite » n'était pas mendélien mais cytoplasmique et était contrôlé par un élément qui fut appelé rho. Ces mutants sont incapables de respirer et utilisent uniquement la fermentation pour se développer. L'analyse par microscopie électronique des mutants « petites » montrait que les mitochondries avait un aspect aspécifique par rapport à des souches « grandes », dépourvues de cristaux typiques dans leur membrane interne (Yotsuyanagi, 1962). Dans les années 1960, les travaux menés par Margit et Sylvan Nass ont montré que les mitochondries sont des organelles possédant un génome indépendant du génome nucléaire (Nass et Nass, 1963). Il a été confirmé que le facteur rho correspondait à l'ADN mitochondrial et que les mutants « petites » avait un ADN mitochondrial muté souvent porteur de grands remaniements. La mise en évidence de cet ADN a aussi permis à Lynn Margulis de poser l'hypoythèse de l'acquisition des mitochondries suite à une endosymbiose d'une bactérie (Margulis, 1970). Des études plus récentes ont montré qu'une α-protéobactérie proche du groupe des *Rickettsiae* était à la base de cet événement (Yang et al., 1985; Gray, 1998).

La taille du génome mitochondrial est très variable chez les eucaryotes : il est constitué d'environ de 367.000 pb chez *Arabidopsis thaliana* alors qu'il est inférieur à 6.000 pb chez *Plasmodium falciparum* (Unseld et *al.*, 1997 ; Conway et *al.*, 2000). La topologie de l'ADN mitochondrial est également variable car il peut être sous une forme linéaire ou circulaire (Williamson, 2002). Le contenu en gènes est aussi variable et comprend cinq gènes

chez *P. falciparum* et environ cent gènes chez le protozoaire *Reclinomonas americana* (Lang et *al.*, 1997). De plus, la taille des génomes et le nombre de gènes codant des protéines ne sont pas corrélés : le génome mitochondrial humain, dont la taille est de 16 569 pb, contient 13 gènes codant des protéines alors que celui de *A. thaliana*, qui est 22 fois plus grand, ne code que 33 gènes. En comparaison, celui de *S. cerevisiae*, d'une taille de 85.779 pb, possède 8 gènes codant des protéines (Foury et *al.*, 1998).

L'analyse de la séquence du génome mitochondrial de S. cerevisiae a débuté à la fin des années 1970 avec le séquençage du gène ATP9 (Macino et Tzagoloff, 1979 ; Hensgens et al., 1979). Cependant, il a fallu attendre 20 années et les travaux de Foury et al. (1998) pour obtenir la séquence complète de ce génome et son annotation. Actuellement, plusieurs génomes mitochondriaux de levures ascomycètes sont connus. Ils possèdent un ensemble de gènes constants entre les différentes espèces étudiées. On retrouve les gènes codant pour trois sous-unités du complexe cytochrome c oxydase (COX1, COX2, COX3), l'apocytochrome b (COB) et trois sous-unités de l'ATP synthase (ATP6, ATP8, ATP9 – excepté chez Podospora anserina où le gène ATP9 est absent). De plus, ils contiennent les gènes codant pour les ARN ribosomiques 15S et 23S ainsi que ceux codant pour un jeu complet d'ARNt capables de décoder tous les codons des gènes mitochondriaux. Des gènes comme ceux codants sept sousunités du complexe NADH déshydrogénase sont présents spécifiquement chez certaines espèces. En effet, ces gènes ne sont retrouvés que chez les levures hémiascomycètes aérobie obligatoire. De même, les gènes VAR1 et RPM1 codant la protéine ribosomique Var1p et la RNaseP ne sont présents que dans quelques espèces de levures ascomycètes (Tableau 4.1). Malgré cet ensemble de gènes conservés et les quelques exceptions que nous venons de citer, la taille de ces génomes est aussi très variable puisqu'elle s'étend de 20.000 pb chez S. pombe à plus de 100.000 pb chez P. anserina (Lang et al., 1985 ; Cummings et al., 1990).

À la fin des années 1990, le consortium Génolevures a posé les bases de la génomique comparative des levures hémiascomycètes en séquençant les génomes de 13 levures hémiascomycètes avec une faible couverture génomique de 0,4X et en réalisant leur annotation (Souciet et *al.*, 2000). Lors de cette dernière décennie, ce consortium a enrichi les bases de données de génomes en effectuant le séquençage et l'annotation complète des levures *Candida glabrata, Debaryomyces hansenii, Kluyveromyces lactis, Kluyveromyces thermotolerans, Saccharomyces kluyverii, Yarrowia lipolytica et Zygossacharomyces rouxii (Dujon et <i>al.*, 2004; Souciet et *al.*, 2009). Actuellement plus de 30 génomes de levures

hémiascomycètes sont disponibles (Figure 4.1). Après le séquençage et l'assemblage du génome de *S. cerevisiae* en 1996 (Goffeau et *al.*, 1996), la mise en évidence de blocs de duplication a permis de proposer que ce génome avait évolué suite à une duplication totale du génome (Wolfe et Shields, 1997). Les analyses génomiques et la comparaison des différentes espèces séquencées montrent que des levures comme *S. cerevisiae* ou *Candida glabrata* ont évolué différemment à partir d'un ancêtre commun dont le génome a été dupliqué (Kellis et *al.*, 2004). La séparation des espèces pré-duplication et post-duplication se fait entre les genres *Saccharomyces* et *Zygosaccharomyces* (Figure 4.1).

Les levures Hémiascomycètes sont divisées en trois groupes majeurs : les *Saccharomycetaceae*, les *Dipodasceae* et le groupe CTG (Figure 4.1). Les levures du groupe CTG ont la particularité de décoder les codons CUG comme la sérine au lieu de la leucine (Ohama et *al.* 1993 ; Santos et Tuite 1995). Certaines espèces de ce groupe appartiennent au genre des *Candida*, comme *Candida albicans* et *Candida parapsilosis* et sont des pathogènes humains. D'autres espèces appartenant à ce groupe, comme celles du genre *Pichia*, sont également retrouvées comme agents infectieux chez des personnes immunodéprimées. On retrouve notamment les champignons *Pichia farinosa* et *Pichia anomala* (Bakir et *al.*, 2004 ; Adler et *al.*, 2007).

Les espèces du genre *Pichia* représentent plus de 91 espèces au sein des levures hémiascomycètes et ont été regroupées grâce à leur possibilité d'assimilation du nitrate comme source d'azote, leur faible fermentation de sucre et leur bourgeonnement multilatéral (Kurtzman et Fell, 1998). Parmi ces espèces, on trouve *Pichia pastoris* qui est une levure qui présente un intérêt fondamental. Ses capacités d'utilisation du méthanol ont permis de mettre en évidence et de comprendre le métabolisme de ce composé (Couderc et Baratti, 1980). D'autre part, cette espèce a également un intérêt appliqué ; elle est utilisée en biotechnologies pour la production de protéines comme par exemple des immunoglobulies IgG (Li et *al.*, 2006). L'espèce *Pichia stipitis* est utilisée pour la formation de bioéthanol à partir de la fermentation du xylose provenant de lignocellulose des plantes (Jeffries et *al.*, 2007). D'autres levures sont également intéressantes du fait de leurs propriétés d'osmotolérance comme *Pichia sorbitophila* et *Pichia farinosa* qui ont été isolées respectivement comme un contaminant d'une solution de sorbitol 70% et d'une bière jordanienne (Rodriques de Miranda et *al.*, 1980). Ces deux levures ont été décrites par des approches biochimiques comme étant deux souches d'une même espèce (Kurtzman et Fell, 1998).

Lors du programme Génolevures I, les génomes de 13 levures hémiascomycètes ont été partiellement séquencés et annotés. L'analyse phylogénétique de ces levures, basée sur les séquences d'ADN ribosomiques 18S et 26S, ont montré que P. sorbitophila était proche de Candida tropicalis, une levure pathogène pour l'Homme et de Debaryomyces hansenii, une levure halotolérante (Souciet et al., 2000). L'analyse de la conservation de l'ordre des gènes dans le génome nucléaire des espèces séquencées et annotées montre que le pourcentage de synténie baisse en fonction de l'éloignement phylogénétique avec S. cerevisiae à l'exception de P. sorbitophila. En effet, ce pourcentage est de 59% chez cette dernière alors qu'il est de 16% chez D. hansenii et de 18% chez C. tropicalis (Llorente et al., 2000). En plus de son caractère osmotolérant, la position phylogénétique intéressante et la conservation de la synténie de P. sorbitophila avec S. cerevisiae sont les raisons pour lesquelles le génome de P. sorbitophila a été entièrement séquencé et annoté dans le cadre du programme Génolevures 3. J'ai participé à cette annotation qui a consisté dans un premier temps à déterminer les codons d'initiation et de terminaison des phases codantes des gènes du génome nucléaire de P. sorbitophila. Ces phases ouvertes de lecture sont identifiées et classées grâce au pourcentage de similarité avec les protéines de S. cerevisiae, de D. hansenii ou Y. lipolytica. Lors de cette première partie de l'annotation, nous avons aussi cherché d'éventuels introns au niveau des gènes. Dans un second temps, nous avons regroupé ces gènes en familles, définies par les taux de similarité et également par les fonctions putatives des de ces protéines. Les deux étapes d'annotation ont duré environ deux mois.

L'annotation du génome de *P. sorbiophila* et son analyse montre que 60% des gènes ont deux versions alléliques différentes dans le génome et sont hétérozygotes. Ces résultats suggèrent que le diploïde *P. sorbitophila* peut provenir de l'hybridation de deux espèces différentes (Dujon, 2010). À partir des données biochimiques déterminées par Kurtzman et Fell (1998), les levures *P. sorbitophila* et *P. farinosa* sont définies comme étant deux souches d'une même espèce. De ce fait, nous pouvons penser que *P. farinosa* est l'un des deux parents de *P. sorbitophila*. Afin de déterminer la parenté de ces deux levures, la souche *P. farinosa* a été séquencée par notre laboratoire, en utilisant une méthode à haut débit de type SOLEXA, mais aucun assemblage satisfaisant de son génome nucléaire n'a pu être réalisé. En effet, ce type de séquençage a conduit à l'obtention de fragments d'ADN de 40 pb. Un tel type de séquençage est recommandé lorsque l'on veut assembler un génome en prenant un génome de référence. Par contre, il ne l'est pas pour faire du séquençage *de novo*. En prenant le génome

de *P. sorbitophila* comme référence, seuls 0,2% des fragments d'ADN séquencés s'alignent sur ce génome ce qui explique l'absence d'assemblage satisfaisant de *P. farinosa*. Néanmoins, nous avons réalisé avec succès un assemblage *de novo* du génome mitochondrial de cette levure.

Dans cette partie du travail, nous avons caractérisé les séquences correspondant aux génomes mitochondriaux des levures P. sorbitophila et de P. farinosa suite à leur séquençage. Dans le but de déterminer la parenté de ces deux organismes à partir des génomes mitochondriaux, nous avons envisagé de réaliser une analyse phylogénétique des levures du groupe CTG. Les génomes des espèces Candida albicans, Candida parapsilosis et Debaryomyces hansenii sont connus et annotés et ont été utilisés lors de cette analyse (Anderson et al., 2001; Nosek et al., 2004; Sacerdot et al., 2008). De plus, nous avons récupéré dans la base de données GeneBank des séquences correspondant à l'ADN mitochondrial des levures Pichia guilliermondii, Candida tropicalis et Lodderomyces elongisporus appartenant au groupe CTG (Butler et al., 2009). La séquence du génome mitochondrial de Pichia stipitis nous a été donnée par M. Thomas Jeffries (Jeffries et al., 2007). L'analyse phylogénétique étant basée sur les séquences des gènes codant les protéines, nous avons d'abord caractérisé les gènes présents dans les génomes des six levures P. sorbitophila, P. farinosa, P. guilliermondii, P. stipitis, C. tropicalis et L. elongisporus. Nous avons ainsi montré que tous ces génomes possèdent le même ensemble de gènes, mais que la synténie n'est pas conservée entre ces espèces. Nous avons ensuite utilisé les séquences de ces gènes pour réaliser l'analyse phylogénétique. Cette analyse montre clairement que les levures du groupe CTG sont divisées en deux clades : le premier regroupant les espèces diploïdes du genre Candida et le second comprenant les espèces haploïdes du genre Pichia. Nous avons également démontré que l'emplacement de P. sorbitophila dans cette phylogénie était distant de celui de *P. farinosa* montrant que ces deux levures appartiennent à des espèces différentes.

Une autre levure hémiascomycète intéressante du fait de ses capacités halotolérantes et osmotolérantes est *Zygosaccharomyces rouxii*. Le génome de cette espèce a été entièrement séquencé et annoté lors du programme Génolevures 3. L'analyse génomique comparative de du génome de cette espèce avec d'autres levures a montré que *Z. rouxii* est une espèce préduplication proche des espèces post-duplication (Souciet et *al.*, 2009) (Figure 4.1). Dans cette partie, nous avons caractérisé un contig correspondant à l'ADN mitochondrial de cette espèce. Nous avons ensuite procédé à son annotation : le même ensemble de gène est retrouvé chez *Z*. *rouxii* par rapport aux autres levures hémiascomycètes. Cependant, nous n'avons pas pu déterminer la topologie circulaire de ce génome. Une topologie linéaire de cet ADN serait le second cas observé chez les levures appartenant au groupe des *Saccharomycetaceae* après *Hanseniaspora uvarum* (Pramateftaki et *al.*, 2006).

Espèces	Topologie	Taille (nt)	% GC	COB	ATP	NADH	VAR 1	ORF introniques	ARNt	ARNr	RNase P
				CUX 1, 2, 3	6, 7, 8	dêshydrogênase					
Saccharomyces cerevisiae	C	85779	17	4	б			11	24	2	-
Saccharomyces servazzii	С	30782	22	4	б		1	4	23	7	-
Saccharomyces castelli	С	25753	20	4	б		-	1	23	7	-
Candida (torpulosis) glabrata	С	20063	17	4	ю		1	3	23	7	-
Vanderwaltozyma polyspora	С	21684	15	4	б		-1		23	NA	NA
Kluyveromyces thermotolerans	C	23584	24	4	3			3	24	2	-
Kluyveromyces lactis	C	40291	26	4	3		—	1	22	7	-
Ashbya gossypii	C	23564	18	4	3				23	7	
Hanseniaspora uvarum	Γ	18844	29	4	3			1	23	7	
Pichia canadensis	C	27694	18	4	3	L	1	2	25	7	
Debaryomyces hansenii	С	29462	27	4	б	L		4	25	7	
Candida metapsilosis	L	24152	25	4	б	L		1	24	7	
Candida neerlandica	C	32141	24	4	3	L		4	24	7	
Candida orthopsilosis	C	22528	25	4	3	L		1	24	7	
Candida parapsilosis	L	32745	23	4	б	L		9	24	7	
Candida albicans	С	40420	32	4	б	L		NA	30	7	
Yarrowia lipolytica	С	47916	22	4	б	L		10	27	7	
Candida zemplinina	С	23114	21	4	б		-1	7	25	7	
Scizosaccharomyces pombe	С	19431	30	4	б			3	25	7	-
Scizosaccharomyces octosporus	С	44227	24	4	б			9	24	7	-
Scizosaccharomyces japonicus	С	80059	20	4	3			2	25	2	
0 - 1								1	4		
<u>1 abreau 4.1 :</u> C	aracteristiques (les genomes de	S muchor	idries des levures	s nemiascon	iyceles. La laille est in	nquee en m	icleondes. Les nomo	tes de		
chaque colonne	correspondent a	au nombre de g	senes prese	nts dans chaque	categorie.						

C : topologie circulaire de l'ADN mitochondrial L : topologie linéaire de l'ADN mitochondrial

NA : non indiqué



<u>Figure 4.1</u>: Arbre phylogénétique des levures hémiascomycètes (modifié à partir de Dujon, 2010). Trois groupes majeurs sont définis: les *Saccharomycetaceae* (soulignés en bleu), les *Dipodasceae* (violet) et le groupe CTG (orange). Les clades allant des *Saccharomyces sensu stricto* aux *Vanderwaltozyma* divergent d'un ancêtre commun ayant eu une duplication totale de son génome (WGD).

1. Les génomes mitochondriaux des levures du groupe CTG

La souche de *P. sorbitophila* CBS 7064 a été séquencée au Génoscope avec la technique de Whole Shotgun Sequencing. Les fragments d'ADN obtenus ont été assemblés grâce au programme ARACHNE (Jaffe et *al.*, 2003). Suite à cet assemblage, 15 contigs ont été classés comme étant non nucléaires. Nous avons recherché parmi ces contigs ceux qui étaient susceptibles de contenir des séquences mitochondriales en réalisant des alignements de séquences sur des génomes mitochondriaux connus. Ainsi, trois contigs ont pu être identifiés comme correspondant à de l'ADN mitochondrial. Nous avons pu les assembler manuellement car ils possédaient des séquences de plus de 100 pb qui étaient chevauchantes. Nous avons vérifié cet assemblage expérimentalement par des réactions de PCR et séquençages des fragments obtenus (Figure 4.2). De plus, des amplifications réalisées en choisissant des amorces de part et d'autre du contig puis le séquençage des fragments obtenus nous permettent d'affirmer que cette molécule d'ADN est circulaire (Figure 4.2).

Nous avons également séquencé la souche de *P. farinosa* CBS 185 grâce à une méthode à haut débit de type SOLEXA à la plateforme séquençage de l'IGBMC de Strasbourg. Le génome mitochondrial de cette levure a été assemblé à partir du programme SOAP2 (Li et *al.*, 2009) en prenant des génomes mitochondriaux annotés comme référence. Malgré une bonne qualité d'assemblage, quinze zones d'environ 100 nucléotides n'étaient pas suffisamment couvertes et comportaient une série de nucléotides non-identifiés N. Des réactions de PCR et de séquençages complémentaires ont été nécessaires pour avoir un génome mitochondrial entier (Figure 4.2). De la même manière, nous avons démontré que ce génome est circulaire comme celui de *P. sorbitophila*.

Nous avons également récupéré dans la base de données GeneBank des contigs correspondant à l'ADN mitochondrial des levures du groupe CTG *P. guilliermondii*, *C. tropicalis* et *L. elongisporus* (Butler et *al.*, 2009). La séquence mitochondriale de *P. stipitis* nous a été donnée par M. Thomas Jeffries (Jeffries et *al.*, 2007).

Nous avons annoté manuellement les séquences mitochondriales de toutes ces levures en utilisant les programmes BLASTX et BLASTN pour les gènes codant respectivement des protéines et des ARN ribosomiques (Altschul et *al.*, 1997). Les gènes codant les ARN de transfert ont été identifiés en combinant les programmes BLASTN, tRNAscan-SE et ARWEN (Altschul et *al.*, 1997, Lowe et Eddy, 1997; Laslett et Canbäck, 2008).



<u>Figure 4.2</u> : Représentation des réactions de PCR pour obtenir les génomes mitochondriaux complets de *P. sorbitophila* et *P. farinosa*. Les flèches rouges sont des exemples d'amorces utiliséees pour les PCR. Les fragments d'intérêts ainsi obtenus sont ensuite séquencés pour compléter les génomes mitochondriaux.

A) P. sorbitophila

B) P. farinosa

1.1. Organisation des génomes mitochondriaux des levures du groupe CTG

1.1.1. Contenu en gènes

L'annotation des génomes de *P. sorbitophila* et de *P. farinosa* montre que ces deux espèces possèdent quatorze gènes codant des protéines. Ces gènes codent trois sous-unités du complexe cytochrome c oxydase (*COX1*, *COX2*, *COX3*), l'apocytochrome b (*COB*), trois sous-unités de l'ATP synthase (*ATP6*, *ATP8*, *ATP9*), ainsi que sept sous-unités du complexe NADH déshydrogénase (*NAD1*, *NAD2*, *NAD3*, *NAD4*, *NAD5*, *NAD6* et *NAD4L*). En plus de

ces gènes, ceux codant les sous-unités ribosomiques 15S et 23S sont également présents de même qu'un ensemble de gènes codant des ARNt capables de charger tous les acides aminés nécessaires pour la traduction des protéines au sein de la mitochondrie (Figure 4.3). Le même ensemble de gènes est retrouvé chez les levures hémiascomycètes aérobies obligatoires dont les génomes mitochondriaux sont entièrement séquencés et annotés comme ceux de *Yarrowia lipolytica* et *C. albicans* (Kerscher et *al.*, 2001 ; Anderson et *al.*, 2001).

De la même manière, nous avons effectué les annotations des séquences correspondant à l'ADN mitochondrial des levures *P. stipitis*, *P. guilliermondii*, *C. tropicalis* et *L. elongisporus*. Ces annotations indiquent que le même ensemble de gènes est présent chez ces espèces, aussi bien ceux codant des protéines, que ceux transcrits en ARN ribosomiques et en ARN de transfert (Figure 4.4). Nous n'avons pas effectué des analyses expérimentales supplémentaires pour vérifier la topologie de ces ADN. De ce fait, les génomes mitochondriaux de ces quatre levures sont considérés ici comme incomplets. Par exemple, les contigs de *P. stipitis* et de *P. guilliermondii* ont à leurs extrémités une séquence qui correspond respectivement aux gènes *LSU* et *COX1* (Figure 4.4). L'absence des ARNt^{Asn} et ARNt^{Cys} chez *C. tropicalis* pourrait éventuellement être due au fait que les séquences ne sont pas complètes.

À partir de l'annotation de ces six séquences mitochondriales, nous pouvons affirmer que tous les gènes de *P. sorbitophila*, *P. stipitis* et *P. guilliermondii* sont transcrits à partir d'un brin d'ADN. Les gènes présents dans les génomes de *P. farinosa*, *C. tropicalis* et *L. elongisporus* le sont à partir des deux brins d'ADN (Figures 4.3 et 4.4).

Lors de l'annotation du génome de *P. sorbitophila*, nous avons remarqué que tous les gènes étaient transcrits à partir du même brin d'ADN. Par contre, une phase ouverte de lecture est située sur le brin opposé (Figure 4.3). L'analyse de ce gène potentiel montre qu'il code un motif LAGLIDADG caractéristique des introns du groupe I. Nous avons aussi remarqué que plusieurs codons stops étaient présents au niveau de la phase ouverte de lecture. Le fait que cette séquence soit sur le brin non transcrit, ne soit pas fusionnée à un exon et possède plusieurs codons stops prématurés dans la phase de lecture suggère que cet intron est une trace d'un déplacement des introns au sein du génome mitochondrial de cette levure.





<u>Figure 4.3</u> : Représentation des génomes mitochondriaux de *P. sorbitophila* et *P. farinosa*. Le cercle externe indique la localisation des gènes. Le cercle interne représente les ARNm putatifs. Les ARNt sont représentés par le code à une lettre.

Rectangle bleu foncé : gènes codant les protéines

Rectangle bleu clair : introns codant une endonucléase

Rectangle blanc : introns non codant

Rectangle vert : gènes codant les ARN ribosomiques

Rectangle rouge : gènes codant les ARNt

Rectangle jaune : gènes hypothétique

Rectangle orange : ARN messagers putatifs transcrits à partir du brin +

Rectangle gris : ARN messagers putatifs transcrits à partir du brin -

Triangle : sites d'initiation de la transcription

Barre : sites de clivage des ARN

P. stipitis



Figure 4.4 : Cartes linéaires des génomes mitochondriaux de *P. stipitis*, *P. guilliermondii*, *C. tropicalis* et *L. elongisporus*.

Rectangle jaune : gènes codant les protéines, transcrits à partir du brin -

Rectangle rouge : gènes codant les protéines, transcrits à partir du brin +

Rectangle vert : gènes codant les ARN ribosomiques

Rectangle bleu : gènes codant les ARNt, transcrits à partir du brin +

Rectangle violet : gènes codant les ARNt, transcrits à partir du brin -

Flèche : sites d'initiation de la transcription

Étoile : sites de clivage des ARN

1.1.2. Les introns des gènes mitochondriaux

Les introns mitochondriaux sont autocatalytiques et sont classés en deux groupes (groupe I et groupe II) en fonction de leur mode d'épissage (Lang et *al.*, 2007). Les introns du groupe I codent pour des maturases ou des endonucléases qui sont classées en quatre catégories en fonction de leur structure secondaire et de leurs motifs protéiques. On retrouve des endonucléases avec le motif LAGLIDADG, celles avec une boîte His-Cys ou encore les endonucléases avec les domaines GIY-YIG et celles possédant un domaine HNH. En revanche, les introns du groupe II codent pour une réverse-transcriptase qui leur est nécessaire pour leur déplacement dans le génome (Lang et *al.*, 2007). Ces deux types d'introns peuvent être fusionnés en phase avec un exon, ou traduits indépendamment (Haugen et *al.*, 2005).

Le nombre d'introns dans les séquences mitochondriales que nous avons annotées est variable : aucun chez *L. elongisporus*, 5 chez *P. farinosa* et 16 chez *P. sorbitophila* (Tableau 4.2). Ces introns sont répartis entre les gènes *COX1*, *COB*, *NAD5* et *LSU*. Le nombre d'introns codant des protéines est également variable car les 4 introns présents chez *P. stipitis* sont codants alors qu'il y en 11 chez *P. sorbitophila* (Tableau 4.2).

Sur les 11 introns codants de *P. sorbitophila*, 10 appartiennent aux introns du groupe I et ont au moins un motif LAGLIDADG. Le dernier est un intron du groupe II et représente le seul intron de ce groupe dans les génomes annotés ici. Les quatre introns de *P. stipitis* et les trois introns de *C. tropicalis* font partie du groupe I avec un motif LAGLIDADG. Deux des quatre introns codants de *P. farinosa* possèdent ce même motif alors que les deux restants ont les boîtes caractéristiques GIY-YIG.

Espèces	COX1	СОВ	NAD5	LSU
P. sorbitophila	7 (6)	5 (3)	2 (2)	2 (0)
P. farinosa	2 (2)	3 (2)	0(0)	0(0)
P. stipitis	2 (2)	2 (2)	0(0)	0(0)
P. guilliermondii	?	2(1)	0(0)	1(0)
C. tropicalis	3 (3)	0(0)	0(0)	0(0)
L. elongisporus	0(0)	0(0)	0(0)	0(0)

<u>Tableau 4.2 :</u> Les introns mitochondriaux et leur localisation. Les nombres entre parenthèse représentent les introns codants.
Chez *P. guilliermondii*, deux introns codants sont retrouvés au niveau du gène *COB*, dont un possède le motif GIY-YIG. Aucun intron n'est retouvé au niveau des gènes *NAD5* et *LSU*. Le génome de *P. guilliermondii* est incomplet et une partie de *COX1* est placée de chaque côté du contig. De ce fait, dans la séquence annotée, la région centrale de ce gène est manquante ainsi que les potentiels introns fréquemment présents (Figure 4.4).

1.1.3. La maturation des ARN

Au niveau du génome mitochondrial de *S. cerevisiae*, les motifs TATAAGTAA et AATAATATTCTT sont décrits comme étant les séquences d'initiation de la transcription et de maturation des ARN. Des séquences similaires sont retrouvées dans les génomes mitochondriaux d'autres levures comme *K. lactis* et *Y. lipolytica* (Schäfer, 2005). Une analyse *in silico* des séquences mitochondriales des levures étudiées ici montre que le motif TATAAGAA est retrouvé plusieurs fois. Ces séquences sont situées en amont des gènes, par exemple avant les gènes *COX3* et ARNt^{Leu} chez *P. sorbitophila*. De la même manière dans l'ADN mitochondrial de *P. farinosa*, le motif de clivage des ARN ATAATATT est retrouvé en aval des gènes comme *COX2*, *COX3*, *NAD1* et *LSU* (Figures 4.3 et 4.4).

Un autre type de maturation des ARN est décrite chez les champignons filamenteux *Neurospora crassa* et *Aspergillus nidulans*. Il est décrit comme étant une maturation par ponctuation des ARNt. En fonction de leur emplacement dans le génome, la libération des ARNt par des RNases permet la maturation des ARNm (Schäfer, 2005). Au niveau des génomes mitochondriaux étudiés dans ce travail, les ARNt sont majoritairement regroupés. Cependant, certains sont dispersés et peuvent encadrer des gènes comme les gènes *COX2* ou *NAD4* chez *P. sorbitophila* et *L. elongisporus*. Ceci suppose qu'une maturation par ponctuation peut également exister dans ces levures (Figures 4.3 et 4.4).

En combinant les deux modes de maturation, des ARNm monocistroniques et polycistroniques sont prédits *in silico*. Les paires de gènes *ATP8-ATP6*, *NAD1-NAD6*, *NAD2-NAD3 et NAD4L-NAD5* gènes co-transcrits pourraient également être co-traduits du fait de la courte région intergénique entre deux gènes (Figures 4.3 et 4.4).

1.2. Génomique comparative des génomes mitochondriaux des levures du groupe CTG

1.2.1. Aspects généraux

Les génomes mitochondriaux des levures hémiascomycètes sont majoritairement circulaires même si certaines espèces peuvent avoir des ADN linéaires comme *Candida parapsilosis* (Tableau 4.1) (Kovac et *al.*, 1984). Nous avons montré par des réactions PCR que les génomes de *P. sorbitophila* et *P. farinosa* sont clairement des molécules circulaires (Figure 4.3). Nous n'avons pas testé la topologie des génomes de *P. stipitis*, *P. guilliermondii*, *C. tropicalis* et *L. elongisporus*. L'annotation des contigs mitochondriaux de *P. stipitis* et *P. guilliermondii* a montré qu'une partie des gènes *LSU* et *COX1* respectivement était présente de part et d'autre de ces séquences (Figure 4.4). De ce fait, nous pouvons supposer que les génomes mitochondriaux de ces deux levures sont circulaires. Une analyse par des réactions de PCR au niveau de ces gènes permettrait de le confirmer.

La taille des séquences mitochondriales des six espèces analysées est variable. En effet, la taille du génome de *P. sorbitophila* est de 39.107 pb, celle de *C. tropicalis* est de 50.304 pb, alors qu'elle n'est que de 23.890 pb chez *P. guilliermondii*. Chez *P. farinosa*, *P. stipitis* et *L. elongisporus*, cette taille est comparable et est proche de 30.000 pb (Tableau 4.3). Malgré ces variations, ces génomes possèdent le même ensemble de gènes codant des protéines, deux gènes codant les grande et petite sous-unités ribosomiques et un jeu complet d'ARNt (Figures 4.3 et 4.4).

La différence de taille de ces génomes peut s'expliquer par la présence d'introns ou de longues régions intergéniques. La comparaison des proportions de régions intergéniques chez ces différents génomes montre que la taille de l'ADN mitochondrial de *C. tropicalis* est due à des régions non codantes qui représentent 49 % de la séquence totale (Tableau 4.3). Chez *P. sorbitophila*, 16 introns sont retrouvés et permettent d'expliquer une taille proche de 40.000 pb (Figure 4.3). De plus, ce génome a la spécificité de posséder un intron non lié à un exon qui permet également d'expliquer cette variation de taille par rapport aux autres génomes de levures étudiées dans ce travail.

Espèces	Souche	Ploïdie	Taille (nt)	% GC	Nombre de gènes	Nombre d'introns	% région intergénique
P. sorbitophila	CBS 7064	2n	39107	23	41	16	8
P. farinosa	CBS 185	n	32065	27	43	5	20
P. stipitis	CBS 6054	n	30649	25	40	4	16
P. guilliermondii	ATCC 6260	n	23890	25	41	2*	16
C. tropicalis	MYA-3404	2n	50304	37	46	3	49
L. elongisporus	NRLL YB-4239	2n	35601	29	41	0	42

<u>Tableau 4.3</u>: Caractéristiques des génomes mitochondriaux annotés. Les espèces analysée dans cette étude sont haploïde (n) et diploïde (2n). Le nombre de gène représente les gènes codant pour les protéines, les ARN ribosomiques et les ARN de transfert.

* : le nombre d'introns de *P. guilliermondii* est peut être sous-estimé du fait que le gène *COX1* est incomplet.

1.2.2. Code génétique utilisé

Le code génétique mitochondrial des levures hémiascomycètes diffère du code universel car le codon UGA, défini comme étant un codon stop dans le code universel, est décodé par un ARNt incorporant un résidu tryptophane (Knight et *al.*, 2001). De plus, il a été décrit chez les levures *P. canadensis* et *C. parapsilosis* que les codons AUA et CUN codent l'isoleucine et la leucine au lieu de la méthionine et la thréonine comme chez *S. cerevisiae* (Sekito et *al.*, 1995 ; Nosek et *al.*, 2004). Nous avons réalisé des alignements multiples de protéines codées par les gènes du génome mitochondrial de diverses levures hémiascomycètes qui permettent de mettre en évidence des positions où l'isoleucine et la leucine sont conservées (Figure 4.5). Chez les espèces du genre *Pichia* analysées ici (*P. sorbitophila*, *P. farinosa*, *P. stipitis* et *P. guilliermondii*) ainsi que chez *D. hansenii*, ces positions correspondent aux codons AUA et CUN respectivement, montrant que probablement la même réassignation se produit (Figure 4.5). Cette propriété est également retrouvée pour les espèces *C. tropicalis* et *L. elongisporus* et pourrait donc être une spécificité des levures du groupe CTG.

ATP9_PI SO	MOLVLAAKYMO	344 <mark>M</mark> ATL	.GL GGAA	NGI AL	VEVAL	NGTTRNPSI	RATI FPQ4	MLGFA	TAEA	60
ATP9_DEHA	MQLALAAKY	GAS <mark>M</mark> ATL	.GLOGAA	GI AL	VFVAL	NGTSRNPSL	RATLEPOP	LGFAL	AEA	60
ATP9_CAPA	MQI ALAAKYI O	GAS <mark>I</mark> ATL	.GL GGAA	GI AL	.VFVAL	NGTSRNPSL	.RSTLFPQ4	LGFAL	SEA	60
ATP9_KLTH	MOLVLAAKY	GAGISTI	GLLGAG	GI AI	VFAAL	NG/SRNPSL	.KDTLFPFA	LGFAL	SEA	60
ATP9_SACE	MOLVLAAKY	GAGISTI	GLLGAG	GI AI	VFAAL	NG/SRNPSI	KDTVFPMA	LGFAL	SEA	60
ATP9_YALI	MOLVLAGKY	GAGLASI	GLVGAG	GI AI	VFAAL	NGVSRNPAL	KGQLFTYS	LGFAL	SEA	60
	:..**	*.::::	** **.	***:	**.**	**.:***::	: :*. :	****	:**	
		_				-		_		
ATP9_PI SO	CGLFCLMMSFL	LI YAV	76							
ATP9_DEHA	CGLFCLMWSFLLLYAV 76									
ATP9_CAPA	CGLFCLM SFLLLYAV 76									
ATP9_KLTH	TGLFCLM SFLLLYGV 76									
ATP9_SACE	TGLFCLM/SFLLLFG/ 76									
ATP9_YALI	TGLFALM AFL	LLYAV.	76							
_	***.**	*::.*								

<u>Figure 4.5</u>: Alignement multiple de la protéine Atp9p. Un exemple d'alignements de protéines issues de *P. sorbitophila* (PISO), *D. hansenii* (DEHA), *C. parapsilosis* (CAPA), *K. thermotolerans* (KLTH), *S. cerevisiae* (SACE) et *Y. lipolytica* (YALI). Les boîtes bleues et oranges représentent des positions où l'isoleucine et la leucine sont conservées et qui correspondent respectivement aux codons AUA et CUN chez *P. sorbitophila*.

1.2.3. Analyse de la synténie des génomes mitochondriaux

Nous avons également analysé l'organisation des gènes au niveau de ces génomes mitochondriaux. Cette analyse montre la conservation de larges blocs de synténie comme celui s'étendant du gène *LSU* jusqu'au gène *NAD1* (Figure 4.6). Par contre, ces blocs sont plus dispersés chez les souches *C. tropicalis, L. elongisporus* ou encore *C. parapsilosis* (Figure 4.6).

Les souches *P. sorbitophila* et *P. farinosa* ont été décrites par des approches biochimiques comme étant la même espèce (Kurtzman et Fell, 1998). On parle alors d'espèces synonymes. L'analyse de la synténie que nous avons effectuée montre que des blocs de gènes sont conservés entre ces deux espèces mais que certains subissent également des remaniements comme l'inversion de la région allant du gène *ATP6* au gène codant l'ARNt^{Leu} (Figure 4.6). La comparaison de l'ordre des gènes mitochondriaux des différentes levures étudiées ici mène aux mêmes conclusions que celles que nous avons obtenues entre *P. sorbitophila* et *P. farinosa* : des blocs de synténie sont conservés mais des remaniements sont également présents (Figure 4.6). Ces mêmes conclusions entre différentes espèces et entre *P. sorbitophila* et *P. farinosa* nous laisse suggérer que ces deux levures pourraient être des espèces distinctes.



D. hansenii P. sorbitophila P. farinosa P. stipitis P. guilliermondii L. elongisporus C. parapsilosis C. tropicalis

<u>Figure 4.6</u> : Conservation de la synténie au sein des espèces du genre *Pichia*. De larges blocs de synténie (barres épaisses) sont conservés entre les différentes espèces du genre *Pichia* et avec *D. hansenii*.

1.2.4. Analyse phylogénétique

L'analyse de la synténie entre *P. sorbitophila* et *P. farinosa* nous a permis de supposer que ces deux levures appartiennent à des espèces distinctes. Pour vérifier cette hypothèse, une analyse phylogénétique portant sur les gènes mitochondriaux a été entreprise. Lors de cette analyse, nous avons aussi utilisé les génomes mitochondriaux d'autres levures du groupe CTG comme *C. albicans*, *C. parapsilosis* et *D. hansenii* (Anderson et *al.*, 2001 ; Nosek et *al.*, 2004 ; Sacerdot et *al.*, 2008).

Nous avons construit un arbre phylogénétique basé sur la concaténation des séquences des gènes COX2, COX3, ATP6, ATP8, ATP9, NAD1, NAD2, NAD3, NAD4, NAD4L, NAD5 et

NAD6, l'ensemble représentant 3242 acides aminés (Figure 4.7). Trois méthodes de calcul ont été employées pour cette analyse : le Neighbor-Joining (NJ), le maximum de similarité (Maximum Likelihood ML) et une approche Bayesian. Dans chaque cas, la robustesse de l'analyse est réalisée pour 100 réplications indépendantes (bootstrap = 100). Pour les études de similarités par les approches Bayesian et ML, les séquences des gènes de *Candida metapsilosis, Candida orthopsilosis* et *Candida neerlandica* ont également été utilisées (Kosa et *al.*, 2006 ; Valach et *al.*, 2008). Cet apport a permis de montrer la fidélité de nos arbres phylogénétiques par rapport à ce qui a été proposé lors d'autres études.

Les arbres phylogénétiques construits dans les trois cas de figure mènent aux mêmes conclusions. Premièrement, cette étude démontre clairement la présence de deux clades au sein du groupe CTG : le premier contient les espèces diploïdes du genre *Candida* alors que le second comprend les souches du genre *Pichia* qui sont majoritairement haploïdes (Figure 4.7). Il est intéressant de noter ici que le sous-groupe *Pichia* est également divisé. En effet, les trois arbres phylogénétiques placent *P. sorbitophila* à la base de ce sous-groupe avec une forte robustesse. Deuxièmement, cette analyse montre que les levures *P. farinosa*, *P. stipitis*, *P. guilliermondii* et *D. hansenii* sont proches les unes des autres et qu'elles sont plus éloignées de *P. sorbitophila* (Figure 4.7). Du fait de leur éloignement phylogénétique, *P. sorbitophila* et *P. farinosa* peuvent être définies comme deux espèces distinctes.

Cette analyse montre également que les espèces *P. farinosa* et *P. stipitis* sont très proches l'une de l'autre avec une robustesse de presque 100% avec les trois méthodes de calcul (Figure 4.7). La levure *P. stipitis* est utilisée en biotechnologie pour fermenter le xylose issu des plantes et pour former le bioéthanol (Jeffries et *al.*, 2007). D'autres levures proches de *P. stipitis* peuvent également fermenter ce sucre, mais avec une efficacité moindre (Jeffries, 2006). Une meilleure connaissance des génomes de ces organismes, et notamment *P. farinosa*, pourraient apporter des éléments pour comprendre la forte efficacité de fermentation du xylose chez *P. stipitis* par rapport aux autres levures appartenant au groupe CTG.







<u>Figure 4.7</u> : Arbres phylogénétiques des levures du groupe CTG. Ces arbres ont été construits à partir de la concaténation des séquences des gènes *COX2*, *COX3*, *ATP6*, *ATP8*, *ATP9*, *NAD1*, *NAD2*, *NAD3*, *NAD4*, *NAD4L*, *NAD5* et *NAD6*

A) Neighbor-Joining

B) Maximum Likelihood

C) Approche Bayesian

Les nombres à chaque embranchement correspondent à la robustesse et à la sûreté de chaque nœud.

1.3. Conclusion

L'analyse comparative présentée ici permet de mieux comprendre les génomes des mitochondries des levures du groupe CTG et leur évolution. Malgré une taille variable, ces ADN mitochondriaux possèdent les gènes codant pour les 14 mêmes protéines. Le nombre d'introns et la longueur des régions intergéniques expliquent les variations de tailles de ces génomes. L'augmentation de la taille du génome de *P. sorbitophila* peut également résulter de la présence d'une phase codante située sur le brin non transcrit de cet ADN. Cette séquence définie comme étant un intron du groupe I suggère qu'elle résulte de la mobilité de ces éléments. Contrairement aux levures du groupe des *Saccharomycetaceae*, les gènes codant pour la protéine ribosomique Var1p et la RNAseP (*RPM1*) sont absents au niveau de tous les génomes mitochondriaux des levures du groupe CTG analysés et annotés jusqu'ici.

L'analyse phylogénétique réalisée lors de cette étude montre clairement que le groupe CTG est divisé en 2 sous-groupes distincts : le premier contient majoritairement les espèces haploïdes du genre *Pichia* alors que le second est composé des espèces diploïdes du genre *Candida*. Cette observation est consistante avec des études basées sur l'analyse du génome nucléaire qui ont montré une distinction phylogénétique entre des levures haploïdes et diploïdes appartenant au groupe CTG (Fitzpatrick et *al.*, 2006 ; Butler et *al.*, 2009).

Enfin, cette étude indique que *P. sorbitophila* et *P. farinosa*, sont des espèces distantes d'un point de vue phylogénétique et que, de ce fait, elles ne sont pas des espèces synonymes. En effet, en dépit de six blocs de synténie conservés, la distance évolutive importante entre *P. sorbitophila* et *P. farinosa* montre que ces deux levures sont des espèces différentes.

Ces résultats ont mené à la publication de deux articles qui sont placés dans la partie annexe de ce manuscrit. Le premier a été publié en septembre 2009 dans la revue *FEMS Yeast Research* et représente le séquençage et l'annotation du génome mitochondrial de *P. sorbitophila* (Jung et *al.*, 2009). Le second est actuellement en ligne et sous-presse pour la revue *Current Genetics* et est le résultat du séquençage et de l'annotation de l'ADN mitochondrial de *P. farinosa* et reprend toute la partie sur l'analyse phylogénétique des levures du groupe CTG (Jung et *al.*, 2010).

Récemment, des études taxonomiques ont permis d'approfondir la classification des espèces du genre *Pichia* et des sous-genres comme *Meyerozyma* ou *Millerozyma* sont apparus

170

(Kurtzman et Suzuki, 2010). Ces études sont basées sur une approche phylogénétique à partir des séquences des gènes codant les grandes et petites sous-unités ribosomiques. Notre analyse phylogénétique des génomes mitochondriaux a montré une forte similarité de résultats avec les résultats obtenus à partir des génomes nucléaires des levures du groupe CTG (Fitzpatrick et *al.*, 2006 ; Butler et *al.*, 2009). De ce fait, une étude des génomes mitochondriaux des levures du genre *Pichia* permettrait d'avoir un meilleur aperçu de la phylogénie de ces espèces. De plus, il est admis que ces espèces sont également présentes au niveau du groupe des *Saccharomycetaceae* (par exemple, *Pichia anomala*) et du groupe des *Dipodasceae* (*P. pastoris*). Une espèce particulièrement intéressante est *Pichia canadensis* qui est à notre connaissance la seule espèce hémiascomycète contenant les gènes codant les sept sous-unités du complexe NADH déshydrogénase et le gène *VAR1* (Sekito et *al.*, 1995). Une meilleure dans la compréhension de l'évolution des génomes mitochondriaux des levures hémiascomycètes.

2. Organisation du génome mitochondrial de Zygosaccharomyces rouxii

Suite au séquençage du génome de Z. rouxii lors du programme Génolevures 3, nous avons assemblé et identifié un contig de 39.851 pb correspondant à de l'ADN mitochondrial grâce à des comparaisons de séquences avec des ADN mitochondriaux d'autres levures hémiascomycètes. Comme nous l'avons effectué pour les génomes mitochondriaux de *P. sorbitophila* et *P. farinosa*, nous avons réalisé des réactions de PCR en prenant des amorces de part et d'autre du contig afin de déterminer la topologie de la molécule d'ADN. Aucun résultat positif n'ayant été obtenu, nous avons effectué d'autres réactions de PCR pour vérifier la qualité de l'assemblage de l'ADN mitochondrial. À nouveau, aucun fragment de PCR n'a été amplifié suggérant un mauvais assemblage du contig mitochondrial. De ce fait, nous ne pouvons pas conclure quant à la topologie de l'ADN mitochondrial de *Z. rouxii*.

Nous avons tout de même réalisé une annotation complète de ce contig. Comme pour l'annotation des génomes de *P. sorbitophila* et de *P. farinosa*, les gènes codant des protéines et les ARN ribosomiques ont été identifiés en utilisant les outils de type BLAST. Les ARNt l'ont été en combinant les logiciels tRNAscan-SE et ARWEN (Lowe et Eddy, 1997; Laslett et Canbäck, 2008).

2.1. Annotation du génome mitochondrial de Z. rouxii

L'annotation de ce segment d'ADN permet d'affirmer que le génome mitochondrial de *Z. rouxii* possède le même ensemble de gènes que les autres levures hémiascomycètes aérobie facultative : un ensemble d'ARNt capables de charger tous les acides aminés (sauf l'ARNt^{Glu}), les deux ARN ribosomiques (*SSU* et *LSU*) et les gènes *ATP6*, *ATP8*, *ATP9*, *COX1*, *COX2*, *COX3* et *COB*. Il contient également le gène *VAR1* codant pour la protéine ribosomique Var1p. Tous ces gènes sont transcrits à partir du brin - d'ADN à l'exception des gènes *COX3* et ARNt^{Val} qui le sont via le brin + (Figure 4.8).

L'annotation de ce contig montre que la séquence correspondant au gène *COX2* est située d'un côté du contig et est orientée de telle manière qu'il serait transcrit à partir du brin-. À l'autre extrémité du contig, une partie de ce gène est également présente mais en en orientation inverse. L'analyse de la séquence codante suggère que cette partie de gène serait transcrite à partir du brin + (Figure 4.8). Cette contradiction révèle clairement que le contig ainsi formé résulte d'un mauvais assemblage.



Z. rouxii

<u>Figure 4.8</u> : Carte linéaire du génome mitochondrial de *Z. rouxii*. Rectangle jaune : gènes codant les protéines, transcrits à partir du brin – Rectangle rouge : gènes codant les protéines, transcrits à partir du brin + Rectangle vert : gènes codant les ARN ribosomiques Rectangle violet : gènes codant les ARNt, transcrits à partir du brin – Rectangle bleu : gènes codant les ARNt, transcrits à partir du brin + Flèche : sites d'initiation de la transcription Étoile ; sites de clivage des ARN

2.2. Les introns

Le génome mitochondrial de Z. rouxii possède huit introns répartis entre les gènes COX1 et COB. Parmi ces introns, 7 codent pour des endonucléases et font partie des introns du groupe I : 6 d'entre eux ont au moins un motif LAGLIDADG alors que le dernier possède la boîte GIY-YIG. Un autre intron est également présent dans le gène codant la grande sousunité ribososmique (LSU). Cet intron a la particularité d'avoir les mêmes bornes que celles décrites chez S. cerevisiae et K. thermotolerans (Jacquier et Dujon, 1983). Par contre, cet intron ne code aucune endonucléase.

2.3. La maturation des ARN

Les gènes encodés par l'ADN mitochondrial de *Z. rouxii* sont transcrits à partir des deux brins d'ADN. Des séquences d'initiation et de clivage des ARN sont également retrouvées sur ces deux brins (Figure 4.8). La même séquence consensus TATAAGTAA que celle décrite chez *S. cerevisiae* est retrouvée en amont du gène codant l'ARNt^{Val} sur le brin +. Sur le brin -, ce même consensus est retrouvé trois fois en amont des gènes *COX1*, *ATP9* et ARNt^{Leu}. De plus, le consensus de clivage de l'ARN AATAATATTCTT est retrouvé deux fois sur le brin + et deux fois sur le brin -. En prenant en compte ces séquences consensus et la position des ARNt, nous avons pu déterminer de manière putative des transcrits monocistroniques et polycistroniques contenant notamment la paire de gènes *ATP6-ATP8* (Figure 4.8).

2.4. Topologie de l'ADN mitochondrial

Nous avons effectué plus d'une centaine de réactions de PCR sur l'ADN de Z. rouxii en prenant des amorces et des enzymes différentes et dans des conditions de réactions variables. Nous ne décrirons pas les différentes approches qui ont été réalisées. Une série de PCR internes au gène *COX2* ont été effectuées et leur séquençage montre une séquence identique à celle du contig prouvant que l'ADN matrice utilisé pour ces PCR contient l'ADN mitochondrial. Par contre, les autres PCR n'ont pas permis d'amplifier un fragment d'ADN ou ont conduit dans quelques cas à des amplifications aspécifiques. Cette absence de résultat positif peut donc provenir d'un mauvais assemblage des fragments d'ADN correspondant au génome mitochondrial comme déjà mentionné (paragraphe « Annotation du génome mitochondrial de Z. rouxii »). Néanmoins, un génome non circulaire pourrait aussi expliquer cette absence de résultats de PCR. En effet, des formes linéaires des génomes mitochondriaux sont retrouvées dans des espèces comme *C. parapsilosis* (Kovac et *al.*, 1984). Pour ces génomes linéaires, des séquences répétées en tandem ou inversées sont retrouvées aux extrémités (Nosek et Tomaska, 2003). De telles séquences inversées et répétées sont retrouvées au niveau du contig mitochondrial de *Z. rouxii*, notamment une partie du gène *COX2* qui est retrouvée de chaque côté de cette séquence (Figure 4.8).

D'autres génomes mitochondriaux sont dispersés en plusieurs molécules comme celui du mésozoaire *Dicyema* qui possède entre 100 et 1000 molécules d'ADN différentes (Watanabe et *al.*, 1999). La présence d'un génome morcelé en plusieurs entités pourrait expliquer la présence de régions inversées, mais également celle des grandes régions intergéniques (Figure 4.8). Ainsi, trois entités nucléotidiques pourraient composer le génome mitochondrial de *Z. rouxii* : le premier allant du gène *COX2* jusqu'au gène *COX3*, le deuxième comprenant *VAR1*, et le troisième s'étendant du gène *ATP9* jusqu'à la seconde région du gène *COX2* (Figure 4.8).

Toutes ces hypothèses concernant la topologie de l'ADN mitochondrial de Z. *rouxii* nécessitent des travaux expérimentaux supplémentaires. Pour répondre à cette question, il faudrait extraire l'ADN mitochondrial de manière spécifique sans contamination avec de l'ADN nucléaire. Nous pourrions envisager une préparation d'ADN total suivi d'une séparation en fonction de la densité en gradient de chlorure de césium. Si cette ADN mitochondrial a une densité différente de l'ADN nucléaire, il pourra être récupéré, purifié et utilisé pour la caractérisation de sa topologie. Cette caractérisation peut être effectuée en réalisant des électrophorèses en champs alternés par exemple. Si l'ADN est linéaire, une bande devrait apparaître sur le caryotype électrophorétique.

L'approfondissement de l'analyse du génome de Z. rouxii serait notamment intéressant car cet ADN serait le second génome mitochondrial linéaire caractérisé chez les levures hémiascomycètes appartenant au groupe des *Saccharomycetaceae* après celui de *H. uvarum* (Pramateftaki et *al.*, 2006). Son étude permettrait de comprendre les relations entre les topologies circulaire et linéaire de ces génomes. Elle permettrait également de déterminer les impacts de ces topologies sur la mitochondrie et son fonctionnement.

REFERENCES BIBLIOGRAPHIQUES

BIBLIOGRAPHIE

Adler A, Hidalgo-Grass C, Boekhout T, Theelen B, Sionov E & Polacheck I (2007) *Pichia farinosa* bloodstream infection in a lymphoma patient. *J. Clin. Microbiol* **45**: 3456-3458

Albertson DG (2006) Gene amplification in cancer. Trends Genet 22: 447-455

Albertson DG, Collins C, McCormick F & Gray JW (2003) Chromosome aberrations in solid tumors. *Nat. Genet* **34:** 369-376

Alitalo K, Schwab M, Lin CC, Varmus HE & Bishop JM (1983) Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma. *Proc. Natl. Acad. Sci. U.S.A* **80:** 1707-1711

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402

Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T & Kohn LM (2001) Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. *J. Bacteriol* **183**: 865-872

Antonelli R, Estevez L & Denis-Duphil M (1998) Carbamyl-phosphate synthetase domain of the yeast multifunctional protein Ura2 is necessary for aspartate transcarbamylase inhibition by UTP. *FEBS Lett* **422**: 170-174

Asai A, Tsujita T, Sharma SV, Yamashita Y, Akinaga S, Funakoshi M, Kobayashi H & Mizukami T (2004) A new structural class of proteasome inhibitors identified by microbial screening using yeast-based assay. *Biochem. Pharmacol* **67**: 227-234

Babu JR, Jeganathan KB, Baker DJ, Wu X, Kang-Decker N & van Deursen JM (2003) Rae1 is an essential mitotic checkpoint regulator that cooperates with Bub3 to prevent chromosome missegregation. *J. Cell Biol* **160**: 341-353

Baker DJ, Jeganathan KB, Cameron JD, Thompson M, Juneja S, Kopecka A, Kumar R, Jenkins RB, de Groen PC, Roche P & van Deursen JM (2004) BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nat. Genet* **36**: 744-749

Bakir M, Cerikcioğlu N, Tirtir A, Berrak S, Ozek E & Canpolat C (2004) *Pichia anomala* fungaemia in immunocompromised children. *Mycoses* **47**: 231-235

Balmain A (2001) Cancer genetics: from Boveri and Mendel to microarrays. *Nat. Rev. Cancer* 1: 77-82

Benoist P, Feau P, Pliss A, Vorisek J, Antonelli R, Raska I & Denis-Duphil M (2000) The yeast Ura2 protein that catalyses the first two steps of pyrimidines biosynthesis accumulates not in the nucleus but in the cytoplasm, as shown by immunocytochemistry and Ura2-green fluorescent protein mapping. *Yeast* **16**: 1299-1312

Boguslawski G & Polazzi JO (1987) Complete nucleotide sequence of a gene conferring polymyxin B resistance on yeast: similarity of the predicted polypeptide to protein kinases. *Proc. Natl. Acad. Sci. U.S.A* **84:** 5848-5852

Boyle EI, Weng S, Gollub J, Jin H, Botstein D, Cherry JM & Sherlock G (2004) GO::TermFinder--open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20**: 3710-3715

Brachmann CB, Davies A, Cost GJ, Caputo E, Li J, Hieter P & Boeke JD (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115-132

Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* **72:** 248-254

Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJP, Brunke S, Costanzo MC, Fitzpatrick DA, de Groot PWJ, Harris D, Hoyer LL et al. (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. *Nature* **459**: 657-662

Caplan AJ, Cyr DM & Douglas MG (1992) YDJ1p facilitates polypeptide translocation across different intracellular membranes by a conserved mechanism. *Cell* **71**: 1143-1155

Chen C & Kolodner RD (1999) Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat. Genet* **23**: 81-85

Conway DJ, Fanello C, Lloyd JM, Al-Joubori BM, Baloch AH, Somanath SD, Roper C, Oduola AM, Mulder B, Povoa MM, Singh B & Thomas AW (2000) Origin of *Plasmodium falciparum* malaria is traced by mitochondrial DNA. *Mol. Biochem. Parasitol* **111:** 163-171

Couderc R & Baratti J (1980) Oxidation of methanol by the yeast *Pichia pastoris* : purification and properties of the alcohol oxidase. *Agric. Biol. Chem.* **44**: 2279-2289

Coyne S, Courvalin P & Galimand M (2010) Acquisition of multidrug resistance transposon Tn6061 and IS6100-mediated large chromosomal inversions in *Pseudomonas aeruginosa* clinical isolates. *Microbiology (Reading, Engl.)* **156:** 1448-1458

Crespo JL & Hall MN (2002) Elucidating TOR signaling and rapamycin action: lessons from *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev* **66**: 579-591, table of contents

Cummings DJ, McNally KL, Domenico JM & Matsuura ET (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr. Genet* **17**: 375-402

DE Vries H (1902) THE ORIGIN OF SPECIES BY MUTATION. Science 15: 721-729

Dietrich FS, Voegeli S, Brachat S, Lerch A, Gates K, Steiner S, Mohr C, Pöhlmann R, Luedi P, Choi S, Wing RA, Flavier A, Gaffney TD & Philippsen P (2004) The *Ashbya gossypii* genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. *Science* **304**: 304-307

Doolittle RF (1999) Do you dig my groove? Nat. Genet 23: 6-8

Dujon B (2010) Yeast evolutionary genomics. Nat. Rev. Genet 11: 512-524

Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich J, Beyne E et al. (2004) Genome evolution in yeasts. *Nature* **430**: 35-44

Duncan K, Edwards RM & Coggins JR (1987) The pentafunctional arom enzyme of *Saccharomyces cerevisiae* is a mosaic of monofunctional domains. *Biochem. J* **246:** 375-386

Dunham MJ, Badrane H, Ferea T, Adams J, Brown PO, Rosenzweig F & Botstein D (2002) Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A* **99:** 16144-16149

Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME & McVean GA (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**: 1061-1073

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113

EPHRUSSI B & SLONIMSKI PP (1955) Subcellular units involved in the synthesis of respiratory enzymes in yeast. *Nature* **176**: 1207-1208

Eswar N, Eramian D, Webb B, Shen M & Sali A (2008) Protein structure modeling with MODELLER. *Methods Mol. Biol* **426**: 145-159

Exinger F & Lacroute F (1979) Genetic evidence for the creation of a reinitiation site by mutation inside the yeast *URA2* gene. *Mol. Gen. Genet* **173**: 109-113

Fani R, Liò P, Chiarelli I & Bazzicalupo M (1994) The evolution of the histidine biosynthetic genes in prokaryotes: a common ancestor for the hisA and hisF genes. *J. Mol. Evol* **38**: 489-495

Fani R, Brilli M, Fondi M & Lió P (2007) The role of gene fusions in the evolution of metabolic pathways: the histidine biosynthesis case. *BMC Evol. Biol* **7 Suppl 2:** S4

Fitzpatrick DA, Logue ME, Stajich JE & Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. *BMC Evol. Biol* **6**: 99

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA & Merrick JM (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd. Science* **269**: 496-512

Force A, Lynch M, Pickett FB, Amores A, Yan YL & Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545

Foury F, Roganti T, Lecrenier N & Purnelle B (1998) The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* **440**: 325-331

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman RD, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterback TR, Saudek DM, Phillips CA, Merrick JM et al. (1995) The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403

Fritsch ES, Schacherer J, Bleykasten-Grosshans C, Souciet J, Potier S & de Montigny J (2009) Influence of genetic background on the occurrence of chromosomal rearrangements in *Saccharomyces cerevisiae*. *BMC Genomics* **10**: 99

Galitski T, Saldanha AJ, Styles CA, Lander ES & Fink GR (1999) Ploidy regulation of gene expression. *Science* **285**: 251-254

Gancedo JM (2001) Control of pseudohyphae formation in *Saccharomyces cerevisiae*. *FEMS Microbiol*. *Rev* 25: 107-123

Gao S, Chen M, Xia L, Xiu H, Xu Z, Li L, Zhao C, Cheng X & Ma Y (2009) A cotton (*Gossypium hirsutum*) DRE-binding transcription factor gene, GhDREB, confers enhanced tolerance to drought, high salt, and freezing stresses in transgenic wheat. *Plant Cell Rep* 28: 301-311

180

Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, Storz G, Botstein D & Brown PO (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**: 4241-4257

Gasch AP (2007) Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast* 24: 961-976

Geiger T, Cox J & Mann M (2010) Proteomic changes resulting from gene copy number variations in cancer cells. *PLoS Genet* **6** Available at: http://www.ncbi.nlm.nih.gov/pubmed/20824076 [Accédé Novembre 1, 2010]

Gerhart JC & Schachman HK (1968) Allosteric interactions in aspartate transcarbamylase. II. Evidence for different conformational states of the protein in the presence and absence of specific ligands. *Biochemistry* **7:** 538-552

Ghaemmaghami S, Huh W, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK & Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425**: 737-741

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, Arkin AP, Astromoff A, El-Bakkoury M, Bangham R, Benito R, Brachat S, Campanaro S, Curtiss M, Davis K, Deutschbauer A et al. (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**: 387-391

Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H & Oliver SG (1996) Life with 6000 genes. *Science* **274**: 546, 563-567

Gouy M, Guindon S & Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol* **27**: 221-224

Granek JA & Magwene PM (2010) Environmental and genetic determinants of colony morphology in yeast. *PLoS Genet* 6: e1000823

Gunjan A & Verreault A (2003) A Rad53 kinase-dependent surveillance mechanism that regulates histone protein levels in *S. cerevisiae. Cell* **115:** 537-549

Gygi SP, Rochon Y, Franza BR & Aebersold R (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol* **19:** 1720-1730

Hall C, Brachat S & Dietrich FS (2005) Contribution of horizontal gene transfer to the evolution of *Saccharomyces cerevisiae*. *Eukaryotic Cell* **4:** 1102-1115

Hasin-Brumshtein Y, Lancet D & Olender T (2009) Human olfaction: from genomic variation to phenotypic diversity. *Trends Genet* **25**: 178-184

Haugen P, Simon DM & Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* **21:** 111-119

Hawkins JS, Kim H, Nason JD, Wing RA & Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium*. *Genome Res* **16**: 1252-1261

Hensgens LA, Grivell LA, Borst P & Bos JL (1979) Nucleotide sequence of the mitochondrial structural gene for subunit 9 of yeast ATPase complex. *Proc. Natl. Acad. Sci. U.S.A* **76:** 1663-1667

Hittinger CT & Carroll SB (2007) Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449:** 677-681

Hoffman CS & Winston F (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* **57**: 267-272

Hughes TR, Roberts CJ, Dai H, Jones AR, Meyer MR, Slade D, Burchard J, Dow S, Ward TR, Kidd MJ, Friend SH & Marton MJ (2000) Widespread aneuploidy revealed by DNA microarray expression profiling. *Nat. Genet* **25**: 333-337

Jacquier A & Dujon B (1983) The intron of the mitochondrial 21S rRNA gene: distribution in different yeast species and sequence comparison between *Kluyveromyces thermotolerans* and *Saccharomyces cerevisiae*. *Mol. Gen. Genet* **192**: 487-499

Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov JP, Zody MC & Lander ES (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res* 13: 91-96

Jaillon O, Aury J, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **449**: 463-467

Jaquet L, Lollier M, Souciet JL & Potier S (1993) Genetic analysis of yeast strains lacking negative feedback control: a one-step method for positive selection and cloning of carbamoylphosphate synthetase-aspartate transcarbamoylase mutants unable to respond to UTP. *Mol. Gen. Genet* **241**: 81-88

Jaquet L, Serre V, Lollier M, Penverne B, Hervé G, Souciet JL & Potier S (1995) Allosteric regulation of carbamoylphosphate synthetase-aspartate transcarbamylase multifunctional protein of *Saccharomyces cerevisiae*: selection, mapping and identification of missense mutations define three regions involved in feedback inhibition by UTP. *J. Mol. Biol* **248**: 639-652

Jeffries TW (2006) Engineering yeasts for xylose metabolism. Curr. Opin. Biotechnol 17: 320-326

Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, Jin Y, Passoth V & Richardson PM (2007) Genome sequence of the lignocellulose-bioconverting and xylose-fermenting yeast *Pichia stipitis*. *Nat. Biotechnol* **25**: 319-326

Jund R & Lacroute F (1970) Genetic and physiological aspects of resistance to 5fluoropyrimidines in *Saccharomyces cerevisiae*. J. Bacteriol **102**: 607-615

Jung PP, Friedrich A, Souciet J, Louis V, Potier S, de Montigny J & Schacherer J (2010) Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species. *Curr Genet* **56**: 507-15.

Jung PP, Schacherer J, Souciet J, Potier S, Wincker P & de Montigny J (2009) The complete mitochondrial genome of the yeast *Pichia sorbitophila*. *FEMS Yeast Res* **9**: 903-910

Kavanaugh LA, Fraser JA & Dietrich FS (2006) Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Mol. Biol. Evol* 23: 1879-1890

Kellis M, Birren BW & Lander ES (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624

Kerscher S, Durstewitz G, Casaregola S, Gaillardin C & Brandt U (2001) The complete mitochondrial genome of *Yarrowia lipolytica*. *Comp. Funct. Genomics* **2:** 80-90

Khutornenko AA, Roudko VV, Chernyak BV, Vartapetian AB, Chumakov PM & Evstafieva AG (2010) Pyrimidine biosynthesis links mitochondrial respiration to the p53 pathway. *Proc. Natl. Acad. Sci. U.S.A* **107:** 12828-12833

Kidwell MG (1993) Lateral transfer in natural populations of eukaryotes. *Annu. Rev. Gen.* 27: 235-56

Knight RD, Landweber LF & Yarus M (2001) How mitochondria redefine the code. *J. Mol. Evol* **53**: 299-313

Koken MH, Reid A, Quignon F, Chelbi-Alix MK, Davies JM, Kabarowski JH, Zhu J, Dong S, Chen S, Chen Z, Tan CC, Licht J, Waxman S, de Thé H & Zelent A (1997) Leukemia-associated retinoic acid receptor alpha fusion partners, *PML* and *PLZF*, heterodimerize and colocalize to nuclear bodies. *Proc. Natl. Acad. Sci. U.S.A* **94:** 10255-10260

Kosa P, Valach M, Tomaska L, Wolfe KH & Nosek J (2006) Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. *Nucleic Acids Res* **34**: 2472-2481

Koszul R, Caburet S, Dujon B & Fischer G (2004) Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J* **23**: 234-243

Kratz E, Dugas JC & Ngai J (2002) Odorant receptor gene regulation: implications from genomic organization. *Trends Genet* **18**: 29-34

Kresse AU, Dinesh SD, Larbig K & Römling U (2003) Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol. Microbiol* **47:** 145-158

Kurtzman CP & Fell JW (1998) The Yeasts, a taxonomic study (fourth edition), Elsevier

Kurtzman CP & Suzuki M (2010) Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Meyerozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. *Mycoscience* **51**: 2-14

Lacefield S, Magendantz M & Solomon F (2006) Consequences of defective tubulin folding on heterodimer levels, mitosis and spindle morphology in *Saccharomyces cerevisiae*. *Genetics* **173**: 635-646

Lacroute F, Piérard A, Grenson M & Wiame JM (1965) The biosynthesis of carbamoyl phosphate in *Saccharomyces cerevisiae*. *J. Gen. Microbiol* **40**: 127-142

Lang BF, Ahne F & Bonen L (1985) The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*. The cytochrome b gene has an intron closely related to the first two introns in the Saccharomyces cerevisiae cox1 gene. *J. Mol. Biol* **184**: 353-366

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M & Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**: 493-497

Lang BF, Laforest M & Burger G (2007) Mitochondrial introns: a critical view. *Trends Genet* 23: 119-125

Laslett D & Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172-175

Lengauer C, Kinzler KW & Vogelstein B (1998) Genetic instabilities in human cancers. *Nature* **396:** 643-649

Li R, Pavelka N, Rancati G, Zhu J, Bradford W, Florens L, Saraf A & Sanderson B (2010) Aneuploidy causes global proteome changes and phenotypic variation in budding yeast. Oral communication in « Experimental Approaches to Evolution and Ecology in Yeast »

Li H, Sethuraman N, Stadheim TA, Zha D, Prinz B, Ballew N, Bobrowicz P, Choi B, Cook WJ, Cukan M, Houston-Cummings NR, Davidson R, Gong B, Hamilton SR, Hoopes JP, Jiang Y, Kim N, Mansfield R, Nett JH, Rios S et al. (2006) Optimization of humanized IgGs in glycoengineered *Pichia pastoris*. *Nat. Biotechnol* **24**: 210-215

Li R, Yu C, Li Y, Lam T, Yiu S, Kristiansen K & Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**: 1966-1967

Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, Dujon B, de Montigny J, Lépingle A, Neuvéglise C, Ozier-Kalogeropoulos O, Potier S, Saurin W, Tekaia F, Toffano-Nioche C, Wésolowski-Louvel M et al. (2000) Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett* **487**: 122-133

Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955-964

Lynch M & Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155

Lynch M & Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**: 459-473

Macino G & Tzagoloff A (1979) Assembly of the mitochondrial membrane system. The DNA sequence of a mitochondrial ATPase gene in *Saccharomyces cerevisiae*. J. Biol. Chem **254**: 4617-4623

Marcotte EM, Pellegrini M, Ng HL, Rice DW, Yeates TO & Eisenberg D (1999) Detecting protein function and protein-protein interactions from genome sequences. *Science* **285**: 751-753

Margolin AA, Greshock J, Naylor TL, Mosse Y, Maris JM, Bignell G, Saeed AI, Quackenbush J & Weber BL (2005) CGHAnalyzer: a stand-alone software package for cancer genome analysis using array-based DNA copy number data. *Bioinformatics* **21**: 3308-3311

Margulis L (1970) Origin of Eukaryotic Cells, Yale Univ. Press, New Haven, CT

Martens JHA & Stunnenberg HG (2010) The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584:** 2662-2669

Marx J (2002) Debate surges over the origins of genomic defects in cancer. Science 297: 544-546

MCCLINTOCK B (1956) Controlling elements and the gene. *Cold Spring Harb. Symp. Quant. Biol* **21**: 197-216

Messenguy F (1979) Concerted repression of the synthesis of the arginine biosynthetic enzymes by aminoacids: a comparison between the regulatory mechanisms controlling aminoacid biosyntheses in bacteria and in yeast. *Mol. Gen. Genet* **169**: 85-95

Messenguy F, Feller A, Crabeel M & Piérard A (1983) Control-mechanisms acting at the transcriptional and post-transcriptional levels are involved in the synthesis of the arginine pathway carbamoylphosphate synthase of yeast. *EMBO J* **2**: 1249-1254

Michel LS, Liberal V, Chatterjee A, Kirchwegger R, Pasche B, Gerald W, Dobles M, Sorger PK, Murty VV & Benezra R (2001) MAD2 haplo-insufficiency causes premature anaphase and chromosome instability in mammalian cells. *Nature* **409**: 355-359

MORTIMER RK (1958) Radiobiological and genetic studies on a polyploid series (haploid to hexaploid) of *Saccharomyces cerevisiae*. *Radiat*. *Res* **9**: 312-326

Nagy M, Laporte J, Penverne B & Hervé G (1982) Nuclear localization of aspartate transcabamoylase in *Saccharomyces cerevisiae*. J. Cell Biol **92**: 790-794

Nakashima K, Yamada L, Satou Y, Azuma J & Satoh N (2004) The evolutionary origin of animal cellulose synthase. *Dev. Genes Evol* **214:** 81-88

NASS MM & NASS S (1963) INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. I. FIXATION AND ELECTRON STAINING REACTIONS. *J. Cell Biol* 19: 593-611

Nikitin D, Tosato V, Zavec AB & Bruschi CV (2008) Cellular and molecular effects of nonreciprocal chromosome translocations in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci.* U.S.A **105**: 9703-9708

Nosek J, Novotna M, Hlavatovicova Z, Ussery DW, Fajkus J & Tomaska L (2004) Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast *Candida parapsilosis*. *Mol. Genet. Genomics* **272**: 173-180

Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K & Nakase T (1993) Non-universal decoding of the leucine codon CUG in several *Candida* species. *Nucleic Acids Res* **21**: 4039-4045

Ohno S (1970) Evolution by gene duplication, Springer Paquin CE & Williamson VM (1984) Temperature Effects on the Rate of Ty Transposition. *Science* 226: 53-55

Peisajovich SG, Rockah L & Tawfik DS (2006) Evolution of new protein topologies through multistep gene rearrangements. *Nat. Genet* **38:** 168-174

Pramateftaki PV, Kouvelis VN, Lanaridis P & Typas MA (2006) The mitochondrial genome of the wine yeast Hanseniaspora uvarum: a unique genome organization among yeast/fungal counterparts. *FEMS Yeast Res* **6**: 77-90

Prescott LM & Jones ME (1969) Modified methods for the determination of carbamyl aspartate. *Anal. Biochem* **32:** 408-419

Ranz JM, Maurin D, Chan YS, von Grotthuss M, Hillier LW, Roote J, Ashburner M & Bergman CM (2007) Principles of genome evolution in the *Drosophila melanogaster* species group. *PLoS Biol* **5**: e152

Rodriques de Miranda L, Appel KR & Seyfarth H (1980) *Pichia sorbitophila* sp nov. *Antonie Van Leeuwenhoek* **46:** 157-159

Roelants F, Potier S, Souciet JL & de Montigny J (1997) Delta sequence of Ty1 transposon can initiate transcription of the distal part of the URA2 gene complex in *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett* **148**: 69-74

Rossi B, Noel P & Bruschi CV (2010) Different Aneuploidies Arise From the Same Bridgeinduced Cchromosomal Translocation Event in *Saccharomyces cerevisiae*. *Genetics* **186**: 775-90.

Sacerdot C, Casaregola S, Lafontaine I, Tekaia F, Dujon B & Ozier-Kalogeropoulos O (2008) Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. *FEMS Yeast Res* 8: 846-857

Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, Sturn A, Snuffin M, Rezantsev A, Popov D, Ryltsov A, Kostukovich E, Borisovsky I, Liu Z, Vinsavich A, Trush V et al. (2003) TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**: 374-378

Saldanha AJ (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20:** 3246-3248

Sánchez-Gracia A, Romero-Pozuelo J & Ferrús A (2010) Two frequenins in *Drosophila*: unveiling the evolutionary history of an unusual neuronal calcium sensor (NCS) duplication. *BMC Evol. Biol* **10**: 54

Sanger F, Nicklen S & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A* 74: 5463-5467

Sanmiguel P & Bennetzen JL (1998) Evidence that a recent increase in Maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot. (Lond.)* **82:** 37-44

Santos MA & Tuite MF (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans. Nucleic Acids Res* 23: 1481-1486

Schacherer J, de Montigny J, Welcker A, Souciet J & Potier S (2005) Duplication processes in *Saccharomyces cerevisiae* haploid strains. *Nucleic Acids Res* **33**: 6319-6326

Schacherer J, Tourrette Y, Potier S, Souciet J & de Montigny J (2007) Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. *DNA Repair (Amst.)* **6:** 1441-1452

Schacherer J, Tourrette Y, Souciet J, Potier S & De Montigny J (2004) Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. *Genome Res* **14**: 1291-1297

Schäfer B (2005) RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe. Gene* **354:** 80-85

Schüller C, Brewster JL, Alexander MR, Gustin MC & Ruis H (1994) The HOG pathway controls osmotic regulation of transcription via the stress response element (STRE) of the *Saccharomyces cerevisiae CTT1* gene. *EMBO J* **13**: 4382-4389

Schvartzman J, Sotillo R & Benezra R (2010) Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat. Rev. Cancer* **10**: 102-115

Sekito T, Okamoto K, Kitano H & Yoshida K (1995) The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. *Curr. Genet* **28**: 39-53

Selmecki A, Forche A & Berman J (2006) Aneuploidy and isochromosome formation in drugresistant *Candida albicans*. *Science* **313**: 367-370

Selmecki AM, Dulmage K, Cowen LE, Anderson JB & Berman J (2009) Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet* **5**: e1000705

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J & Ullrich A (1989) Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**: 707-712

Souciet J, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S, de Montigny J, Dujon B, Durrens P, Gaillardin C, Lépingle A, Llorente B, Malpertuy A, Neuvéglise C, Ozier-Kalogéropoulos O, Potier S, Saurin W, Tekaia F et al. (2000) Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* **487**: 3-12

Souciet JL, Nagy M, Le Gouar M, Lacroute F & Potier S (1989) Organization of the yeast *URA2* gene: identification of a defective dihydroorotase-like domain in the multifunctional carbamoylphosphate synthetase-aspartate transcarbamylase complex. *Gene* **79**: 59-70

Souciet J, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften P, Sherman DJ, Weissenbach J, Westhof E, Wincker P, Jubin C, Poulain J, Barbe V, Ségurens B, Artiguenave F, Anthouard V, Vacherie B, Val M, Fulton RS, Minx P et al. (2009) Comparative genomics of protoploid *Saccharomycetaceae. Genome Res* **19:** 1696-1709

Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol* **98:** 503-517

Sprague GF (1991) Genetic exchange between kingdoms. Curr. Opin. Genet. Dev 1: 530-533

Sussan TE, Yang A, Li F, Ostrowski MC & Reeves RH (2008) Trisomy represses Apc(Min)mediated tumours in mouse models of Down's syndrome. *Nature* **451**: 73-75

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ & Amon A (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**: 916-

924

Torres EM, Williams BR & Amon A (2008) Aneuploidy: cells losing their balance. *Genetics* **179**: 737-746

Tourrette Y, Schacherer J, Fritsch E, Potier S, Souciet J & de Montigny J (2007) Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence of ploidy. *Mol. Microbiol* **64**: 382-395

Unseld M, Marienfeld JR, Brandt P & Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet* **15:** 57-61

Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ & Ried T (2004) Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. *Cancer Res* **64**: 6941-6949

Valach M, Tomaska L & Nosek J (2008) Preparation of yeast mitochondrial DNA for direct sequence analysis. *Curr. Genet* **54**: 105-109

Vogel C & Morea V (2006) Duplication, divergence and formation of novel protein topologies. *Bioessays* 28: 973-978

Vorísek J, Techniková Z, Schwippel J & Benoist P (2002) Enzymatic activities of Ura2 and Ura1 proteins (aspartate carbamoyltransferase and dihydro-orotate dehydrogenase) are present in both isolated membranes and cytoplasm of *Saccharomyces cerevisiae*. *Yeast* **19:** 449-457

Weaver BAA, Silk AD, Montagna C, Verdier-Pinard P & Cleveland DW (2007) Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell* **11:** 25-36
Welcker AJ, de Montigny J, Potier S & Souciet JL (2000) Involvement of very short DNA tandem repeats and the influence of the RAD52 gene on the occurrence of deletions in *Saccharomyces cerevisiae*. *Genetics* **156**: 549-557

Williams BR & Amon A (2009) Aneuploidy: cancer's fatal flaw? Cancer Res 69: 5289-5291

Williams BR, Prabhu VR, Hunter KE, Glazier CM, Whittaker CA, Housman DE & Amon A (2008) Aneuploidy affects proliferation and spontaneous immortalization in mammalian cells. *Science* **322**: 703-709

Williamson D (2002) The curious history of yeast mitochondrial DNA. *Nat. Rev. Genet* **3:** 475-481

Winkler H (1920) Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche. Verlag von Gustav Fischer: Jena.

Wolfe KH & Shields DC (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708-713

Yotsuyanagi Y (1962) Study of yeast mitochondria. I. Variations in mitochondria ultrastructure during the aerobic growth cycle. *J. Ultrastruct. Res.* **7:** 121-140

Zhang J, Rosenberg HF & Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U.S.A* **95:** 3708-3713

Zhou J, Pérès L, Honoré N, Nasr R, Zhu J & de Thé H (2006) Dimerization-induced corepressor binding and relaxed DNA-binding specificity are critical for *PML/RARA*-induced immortalization. *Proc. Natl. Acad. Sci. U.S.A* **103:** 9238-9243

ENSEIGNEMENT ET RECHERCHE

Enseignement et Recherche

En plus de mes travaux de recherche, j'ai également été moniteur durant ces 3 années de thèse. J'ai eu le plaisir d'enseigner les travaux pratiques (TP) et travaux dirigés (TD) de biochimie pour des étudiants en deuxième année de licence. Au niveau des TP, différentes notions ont été abordées lors de 5 séances différentes :

- Les tests d'identification et de dosage des glucides, des protéines et des acides nucléiques.
- La purification du lysozyme par colonne échangeuse d'ions et la réalisation d'un gel d'électrophorèse dénaturant SDS-PAGE.
- Le dosage de l'activité du lysozyme purifié, les calculs de rendement et d'enrichissement ainsi que sa cristallisation.
- L'étude de la fermentation alcoolique par dosage du glucose et de l'éthanol.
- La caractérisation des lipides par détermination des indices d'acide, de saponification et d'iode.

Lors de chaque séance, j'ai eu la chance d'encadrer 4 groupe de 30 étudiants environ. Ces séances de TP sont ponctuées par un contrôle continu qui consiste en une note d'expérimentation et d'exploitation des résultats obtenus, ainsi qu'en une note attribuée sur la rédaction d'un rapport suite à une des séances de TP.

Au niveau des TD, j'ai encadré 6 à 7 groupes de 30 étudiants pour deux séances. La première consistait à revoir les bases sur les calculs de dilutions, l'utilisation et la conversion des unités ou de petits exercices. La seconde séance reposait sur la définition des lipides, leurs constitutions et caractéristiques, mais également sur des exercices reliant ces lipides à divers indices comme l'indice de saponification. Au cours de ces TD, un dernier contrôle continu est réalisé, notamment dans la réalisation de calculs simples.

Étant d'un naturel introverti, cette expérience de monitorat restera pour moi une expérience professionnelle mais également personnelle. En effet, elle m'aura en outre appris à parler avec aisance devant plusieurs dizaines de personnes. De plus, le fait d'être moniteur permet de rencontrer des personnes scientifiques hors du laboratoire qui peuvent avoir un point de vue externe sur un projet. L'inverse étant également vrai, car on a l'occasion de parler d'autres thématiques de recherche.

L'autre aspect du monitorat a été l'impact de l'enseignement sur la recherche, et celui de la recherche sur l'enseignement. D'un côté, le fait d'enseigner nécessite sans cesse la mise à niveau des techniques ou des connaissances que l'on connaît ou que l'on ne connaît pas. De l'autre côté, il me semble judicieux d'agrémenter et d'illustrer les cours par des exemples précis afin que ceux-ci puissent être plus compréhensibles pour les étudiants.

À côté de ces activités d'enseignement, j'ai également eu l'occasion de participer à 2 projets de vulgarisation scientifique. Le premier, au courant de l'été 2008 et d'actualité à ce moment-là, consistait à créer un dossier sur les tests ADN. Ce dossier comportait notamment des parties sur les bases moléculaires de l'ADN et son historique, ainsi que sur les instruments éthiques et juridiques encadrant ces tests. Le second projet a consisté à réaliser une conférence multidisciplinaire pour des étudiants de niveau licence et master. Lors de cette conférence, en collaboration avec un doctorant en géologie et deux doctorants en microbiologie, nous avons présenté les micro-organismes dans l'environnement ainsi que leurs impacts au niveau de la recherche fondamentale et appliquée.

La participation à ces deux projets m'a appris à simplifier et clarifier mon vocabulaire scientifique couramment utilisé au sein du laboratoire afin d'être compris par le plus grand nombre de personnes. Elle m'a aussi appris à travailler sur un travail rédactionnel en équipe.

Enfin, j'ai aussi participé au Nouveau Chapitre de Thèse (NCT) qui permet de faire un point sur la thèse passée en fonction des compétences acquises et non selon les résultats obtenus. Ce module m'a permis de comprendre que j'avais appris la capacité à exploiter des résultats négatifs ou positifs mais également à gérer de petits projets.





Valorisation des compétences des docteurs « un nouveau chapitre de la thèse »®

Paul Jung

École doctorale des Sciences de la Vie et de la Santé

Université de Strasbourg

Mentor : Mme Pascale Klein

Réarrangements chromosomiques et génomique fonctionnelle chez la levure *Saccharomyces cerevisiae*. Génomique comparative des génomes mitochondriaux des levures hémiascomycètes

Date probable de présentation orale du « NCT » : 15 Juin 2010

Sujet académique de la thèse : Impacts moléculaires et cellulaires des gènes de fusion suite à des réarrangements chromosomiques néoformés chez *Saccharomyces cerevisiae*. Évolution des génomes mitochondriaux chez les levures hémiascomycètes.

Nom du directeur de thèse : Pr. Jacky de Montigny

Cadre général et enjeux de vos recherches

Présentation succinte

Les Mammifères et notamment l'Homme sont composés de tissus et d'organes eux-mêmes constitués de cellules. Chaque cellule possède plusieurs organelles et un noyau qui renferme l'information génétique que l'on connaît sous le nom d'ADN. Chez l'Homme, cet ADN est réparti sur 46 entités : les chromosomes. De nombreuses maladies humaines sont dues à des modifications au niveau de l'ADN ; certaines sont issues de petits changements dans la séquence d'un gène comme dans la Chorée de Huntington où la modification d'un seul gène conduit à la maladie, d'autres pathologies peuvent être dues à des modifications dans la taille, la structure où le nombre de chromosomes. Il y a de nombreux exemples de telles maladies et les plus connues sont sans doute les leucémies où deux chromosomes échangent un morceau de leur ADN.

La levure de boulanger *Saccharomyces cerevisiae* est, comme l'Homme, un organisme eucaryote, c'est-à-dire un organisme contenant son information génétique dans un noyau. Cette levure est étudiée pour de multiples raisons depuis plus d'un siècle car c'est un organisme de petite taille à croissance rapide, facilement conservable. Son génome (c'est-àdire l'ensemble de son ADN) a été séquencé en 1996 et de nombreuses techniques sont utilisées pour divers travaux de génétique, de biochimie, de cytologie, de biologie moléculaire, etc... Mon sujet de thèse s'intéresse aux impacts de divers remaniements chromosomiques sur la levure *S. cerevisiae*, que ce soit d'un point de vue morphologique ou moléculaire. Autrement dit : comment se comporte une cellule eucaryote suite à chamboulement dans le nombre ou la structure des chromosomes.

Votre recherche dans son contexte

L'équipe dans laquelle j'effectue mon travail de thèse étudie les levures sur deux axes distincts : l'étude des réarrangements chromosomiques chez *S. cerevisiae* ainsi que l'étude des génomes des champignons.

Depuis plus de quinze ans, l'un de ces axes est l'étude des remaniements chromosomiques chez *S. cerevisiae*. Plusieurs étudiants avant moi ont tout d'abord sélectionné des cellules de *S. cerevisiae* possédant des réarrangements chromosomiques. Ils ont ensuite caractérisé ces remaniements et ont trouvé des cellules contenant des délétions de taille variable (une partie manquante d'un chromosome) et des duplications, également de taille variable (une partie d'un chromosome est doublée). Ces duplications peuvent être intrachromosomiques ou interchromosomiques et des chromosomes chimères ou surnuméraires peuvent alors être créés, comme c'est le cas dans de nombreux cancers. D'autres étudiants ont également déterminé les mécanismes à la base de la formation de ces divers remaniements chromosomiques. Mon sujet de thèse s'inscrit dans la suite de cette thématique car il permettrait de finaliser un sujet de plus de quinze ans en déterminant les conséquences cellulaires et moléculaires de ces réarrangements chromosomiques.

D'autre part, et ce depuis une dizaine d'année, le second axe de travail de l'équipe consiste à étudier les génomes des champignons et surtout des levures dans le dessein de voir l'évolution de ces espèces au cours du temps. Cette thématique a donné lieu à la création d'un consortium d'une douzaine de laboratoires européens, tout en créant de nombreux contacts internationaux. Un autre aspect de mon sujet de thèse entre pleinement dans ce domaine de recherche car je m'intéresse à l'évolution des génomes mitochondriaux de certaines levures hémiascomycètes.

Vous dans ce contexte

Je suis actuellement doctorant en troisième année de thèse dans le domaine des « Aspects Moléculaires et Cellulaires de la Biologie » à Strasbourg. J'ai également été moniteur pendant ces trois années et j'ai aidé à la réalisation des travaux pratiques et dirigés en biochimie pour des étudiants en deuxième année à la faculté de sciences de la vie.

Je me suis toujours intéressé à tout ce qui est scientifique avec une préférence pour le domaine des sciences du vivant qui est, selon moi, le plus concret. Suite à l'obtention de mon baccalauréat en 2002, je me suis naturellement dirigé vers sur la faculté de Sciences de la vie en m'engageant dans un parcours de biologie générale. Je me suis passionné lors de mes premières années d'université pour la génétique et les mécanismes biologiques. J'ai ensuite intégré le master de Biologie Moléculaire et Structurale avec un parcours « Recherche » car, suite à divers entretiens avec des chercheurs et enseignants-chercheurs, j'ai découvert un métier fascinant que ce soit d'un point de vue professionnel mais également personnel. En effet, ce métier est passionnant car c'est un milieu intellectuellement stimulant où il faut être capable de mener un projet du début à la fin tout en sachant gérer un budget et une équipe de travail. C'est donc tout naturellement que j'ai choisi d'effectuer une thèse en génétique dans l'optique de pouvoir intégrer une équipe de recherche.

Déroulement, gestion et coût de votre projet

Préparation et cadrage du projet

Le sujet de ma thèse s'inscrit dans la continuité des thématiques étudiées dans mon équipe d'accueil, car il s'axe sur une étude de génomique fonctionnelle de la levure *S*. *cerevisiae* ainsi que sur une analyse de génomique comparative des mitochondries des champignons hémiascomycètes. Pour le premier aspect, mon stage de Master a permis de mettre au point les différentes techniques nécessaires au commencement de mon travail de thèse. En plus de ces expériences réalisables au sein du laboratoire une collaboration technique avec la plateforme transcriptomique de l'École Normale Supérieure de Paris a été entreprise. En vue de cette collaboration, un dossier reprenant une grande partie de mes premiers résultats et les perspectives de mes travaux a été déposé puis accepté.

Le second axe de ma thèse repose sur le séquençage et l'analyse de différentes espèces de levures. Ce travail faisant intervenir les laboratoires du consortium Génolevures, seuls les chefs de laboratoires et d'équipes ont mené les directives de ce projet. Le séquençage a été effectué dans deux centres de séquençage : le Génoscope à Evry et la plateforme de séquençage de l'Institut de Génétique, Biologie Moléculaire et Cellulaire (IGBMC) de Strasbourg. Pour mes travaux, des communications avec ces centres de séquençage et des bioinformaticiens ont été nécessaires.

Conduite du projet

Les études de génomiques fonctionnelles de *S. cerevisaie* ont débuté dès le début de ma thèse et ont été le seul sujet abordé durant cette première année de thèse. Les techniques mises au point lors de mon stage de master ont été utilisées sur l'ensemble de ma collection de souches contenant des réarrangements chromosomiques pour obtenir les premiers résultats. Une mise au point de nouvelles expériences comme la RT-PCR quantitative ou la transcription *in vitro* a également été effectuée. De plus la collaboration avec l'ENS de Paris a duré jusqu'à l'obtention des résultats à la fin de cette première année.

Lors de ma deuxième année, mon premier sujet a été un peu remanié. La mise au point et la reproductibilité des RT-PCR quantitatives n'étant pas optimales, ces techniques ont dû être mises de côté. Néanmoins, la perte de ces expériences a pu être compensée par les résultats de l'analyse globale de la transcription de toute ma collection de souches ; une étude plus approfondie de cette analyse transcriptomique a d'ailleurs été entamée durant cette année de thèse. Au niveau de cet axe d'étude, une collaboration avec un autre chercheur de l'équipe a débuté afin de mieux caractériser certains réarrangements chromosomiques. Mon second sujet, portant sur l'analyse comparative des génomes mitochondriaux des levures hémiascomycètes, a consisté en une circularisation et en une annotation manuelle du génome mitochondrial de la levure osmotolérante *Pichia sorbitophila* et en analyse bioinformatique de cet ADN. Ces travaux ont abouti à la publication d'un article scientifique dans une revue internationale.

Dans ma troisième année de thèse, j'ai approfondi l'analyse transcriptomique en apportant d'autres éléments expérimentaux comme le temps de génération, la consommation spécifique de glucose ou encore la production d'éthanol pour l'ensemble de ma collection. Le travail sur la caractérisation des réarrangements chromosomiques a permis la soumission d'un deuxième article, également dans une revue scientifique internationale. Les études sur les génomes mitochondriaux ont continué sur deux autres espèces osmotolérantes : *Zygosaccharomyces rouxii* et *Pichia farinosa*. Malgré beaucoup d'efforts et de temps passé sur la première de ces espèces, seul le génome mitochondrial de *P. farinosa* a pu être entièrement annoté et analysé. En complément de ces génomes, d'autres génomes

mitochondriaux d'organismes proches de *P. farinosa* ont été annotés dans le but de faire une analyse comparative d'espèces très proche phylogénétiquement.

Durant ces trois années de thèse, j'ai beaucoup communiqué avec mes encadrants, ayant des réunions tous les quatre mois avec mon directeur de thèse et mon chef d'équipe afin de faire le point sur l'avancée des travaux et de voir si d'éventuelles réorientations étaient nécessaires. En plus de ces réunions, j'ai dû animer deux à trois réunions d'équipe ou de laboratoire par an.

Estimation et prise en charge du coût du projet

Le coût total de ma thèse est d'environ $150.000 \in$ dont la majeure partie (75%) est requise pour les ressources humaines. Les infrastructures et les produits consommables ne représentent qu'environ 8 et 10%. La partie ressources humaines ainsi qu'une part des consommables est financée par le ministère de la recherche ou par le CNRS. Les autres apports importants sont ceux crédités par l'Université de Strasbourg et une allocation allouée à un jeune chercheur m'encadrant (ANR jeune chercheur).

Nature	Coût (euros)
Ressources humaines	115327
Doctorant	77607
Encadrant	37720
Consommables	19446
Infrastructures	9720
Matériels	3868
Déplacements	500
Formation	2050
Documentation et Communication	1155
Total	152066



Compétences, savoir-faire, qualités professionnelles et personnelles illustrées par des exemples

Mon travail de thèse est ma première expérience professionnelle dans le domaine scientifique et m'a permis d'apprendre et de comprendre les enjeux liés à ce travail. Mon sujet de thèse sur les réarrangements chromosomiques chez *S. cerevisiae* ainsi que sur l'évolution des mitochondries chez les levures hémiascomycètes m'a permis d'acquérir de nombreuses compétences expérimentales mais également intellectuelles. Outre les manipulations génétiques de base, des techniques de biologie moléculaire, de biologie cellulaire, de biochimie ou encore de bioinformatique ont dû être mises au point au cours de ces travaux. La mise en place de ces expériences a nécessité des savoirs qui allaient au-delà de mes connaissances de base et m'a permis d'élargir mes compétences. Ces nouvelles expériences m'ont également permis de m'investir dans la documentation, la recherche ou la réflexion sur de nouveaux outils d'étude.

L'un des principaux apprentissages que je peux retirer de mon travail de thèse est l'aisance orale dans la communication avec autrui. Plutôt réservé de nature, les réunions et séminaires au sein du laboratoire, en corrélation avec le fait d'avoir été moniteur pendant 3 années, m'ont appris à parler en public avec assurance. J'ai par ailleurs acquis la capacité à organiser et optimiser mon temps de travail dans le but de perdre le moins de temps possible. Grâce à l'acquisition de ces compétences, j'ai pu mener à terme des travaux comme les analyses des génomes mitochondriaux, travaux pour lesquels j'ai dû faire des démarches scientifiques que je n'aurais pas crues possibles en début de thèse.

Malgré une grande autonomie, mes travaux n'auraient pu aboutir si j'avais travaillé seul. Le fait d'être entouré d'une équipe compétente, d'avoir des réunions et des discussions régulières avec mon directeur de thèse et mes encadrants, m'ont permis d'acquérir de nombreuses connaissances expérimentales et intellectuelles. Malgré une bonne organisation au sein du laboratoire, certaines tâches restent à la charge des expérimentateurs comme les commandes de produits consommables. Ainsi, j'ai pu comprendre que la commande d'un produit n'est pas si simple et que de nombreuses étapes entrent en compte comme le fait de demander un devis, afin de prévoir au mieux les dépenses futures.

En plus du travail d'expérimentateur, la vie d'un chercheur est rythmée par les congrès nationaux et internationaux et par la rédaction d'articles pour montrer les aboutissements de recherches qui peuvent prendre plusieurs années. L'anglais étant la langue scientifique internationale, sa maîtrise est nécessaire. Ma participation à des cours et à des congrès ainsi que la rédaction de publications ont été l'occasion de perfectionner mon niveau dans cette langue.

L'un des principaux freins à un avancement des travaux est l'aptitude à exploiter les résultats obtenus, qu'ils soient positifs ou négatifs, significatifs ou non. Cette capacité d'analyse et d'exploitation des données a été indispensable pour traiter mes résultats. Pendant ma thèse, j'ai eu la chance de côtoyer de nombreux étudiants en thèse et certains stagiaires dont l'un que j'ai encadré. Cette période d'encadrement m'a permis de proposer, de gérer et de manager un petit projet qui pouvait être mené à son terme dans une période de temps définie.

Malgré un nombre de collaboration assez restreint lors de ma thèse, j'ai eu l'occasion de rencontrer de nombreuses personnes du monde scientifique dont les domaines d'activité sont variés, ce lors de congrès, de formations proposées par l'école doctorale ou

d'autres formations. Grâce aux réunions organisées entre les personnels des laboratoires appartenant au consortium Génolevures, j'ai pu de rencontrer de nombreux scientifiques reconnus sur le plan international. De plus, je reste en contact avec d'anciens doctorants du laboratoire qui me donnent d'autres points de vue sur ma thèse et dont les parcours professionnels me permettent de voir des perspectives de carrières.

Toutes les connaissances expérimentales que j'ai pu acquérir durant mes trois années de thèse ne sont pas directement transférables à tous les domaines scientifiques même si certaines techniques de biologie moléculaire sont assez ubiquitaires, comme les amplifications de type PCR par exemple. Par contre, le fait de pouvoir standardiser de nouveaux protocoles et de nouvelles expériences, d'analyser les résultats avec rigueur et sérieux, de mettre en place et de gérer un projet, ainsi que l'aisance à communiquer à l'oral et à l'écrit sont des compétences que j'ai acquises et que je pense pouvoir adapter dans d'autres domaines.

Résultats, impacts des recherches

Le premier axe de recherche de ma thèse, reposant sur les impacts des réarrangements chromosomiques, est la suite et la finalité d'une étude qui a commencé il y a plus de quinze ans. Mon apport dans cette analyse a eu comme résultat la soumission d'un article scientifique portant sur les mécanismes d'apparition de certains remaniements chromosomiques. Un second article sur les impacts de ces réarrangements chromosomiques est en préparation. L'étude portant sur l'analyse des génomes mitochondriaux des levures hémiascomycètes s'inscrit dans la thématique de mon équipe d'accueil et du consortium de laboratoires Génolevures. Ces études ont permis la publication d'un article scientifique et la soumission d'un autre dans des revues internationales. Mon apport dans les études du consortium Génolevures devrait également aboutir à la préparation d'un article scientifique.

Les résultats obtenus à partir de mon sujet de thèse pourraient mener à la publication de quatre articles dans des revues scientifiques d'ici à la fin de l'année. Mon implication dans le consortium Génolevures pourrait même apporter une cinquième publication. Même si ces études n'ont pas d'impacts d'un point de vue économique ou industriel, elles peuvent en avoir sur la recherche fondamentale car elles permettent de voir les effets de divers réarrangements chromosomiques naturellement sélectionnés sur la cellule et de comprendre d'autres aspects de certaines maladies de type cancer. De plus, la compréhension de l'Évolution des espèces reste un challenge scientifique très intéressant car elle permet de voir comment certains organismes s'adaptent à un environnement et comment ils évoluent. Les études sur l'évolution des génomes mitochondriaux pourraient ouvrir la voie à de nouvelles investigations dans des espèces de levures proches de *Pichia stipitis*, champignon connu pour former du bioéthanol à partir du xylose, un sucre présent dans la lignocellulose des plantes.

La publication de quatre ou cinq articles scientifiques dans des revues internationales dans une courte période de temps atteste de mes qualités dans les domaines scientifiques centraux de ma thèse. Le fait d'avoir deux axes d'études différents, avec des outils d'analyse et des organismes différents montrent mon implication et ma capacité à devenir rapidement performant dans divers domaines.

Identifications de pistes professionnelles

Le métier de chercheur ou d'enseignant-chercheur est un métier passionnant qui nécessite de nombreuses connaissances et de rester au courant des avancées scientifiques. Le fait de travailler dans un environnement intellectuellement stimulant, d'être en contact permanent avec des personnes curieuses d'apprendre est également un aspect intéressant de ce métier. Néanmoins, la recherche plus appliquée, pharmaceutique ou clinique, m'intéresse particulièrement. Même si mes connaissances expérimentales ne sont pas exactement celles qui sont recherchées par des recruteurs, le fait de pouvoir transférer mes compétences dans ces domaines et ma capacité d'adaptation semblent être des atouts.

Les postes de chargé de recherche dans différentes sociétés comme ROCHE nécessitent souvent quelques années d'expériences dans le domaine de la recherche ou des études précliniques ou cliniques. Étant très intéressé par ce genre de postes, une adaptation à ce style de recherche dans des entreprises de type CRO, dans la recherche préclinique ou dans leur département de Recherche et Développement, me semble être un prérequis.

Enfin, les profils de postes offerts par des groupes pharmaceutiques tel que NOVARTIS sont très attractifs car ils offrent de nombreuses possibilités de recherche et développement au sein de leur entreprise. La recherche de diagnostic médical pour déterminer divers types de maladies retient tout spécialement mon attention.

<u>Annexes</u>

Annexe 1

The complete mitochondrial genome of the yeast *Pichia sorbitophila*

Annexe 2

Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species

Annexe 3

The Ty1 LTR-retrotransposon population in *Saccharomyces cerevisiae* genome: dynamics and sequence variations during mobility

RESEARCH ARTICLE



The complete mitochondrial genome of the yeast Pichia sorbitophila

Paul P. Jung¹, Joseph Schacherer¹, Jean-Luc Souciet¹, Serge Potier¹, Patrick Wincker² & Jacky de Montigny¹

¹Unité de Génétique Moléculaire, Génomique et Microbiologie, Institut de Botanique, UMR7156, Université de Strasbourg, Strasbourg, France; and ²Génoscope, Centre National de Séquençage, Evry, France

Correspondence: Jacky de Montigny, Unité de Génétique Moléculaire, Génomique et Microbiologie, Institut de Botanique, UMR7156, Université de Strasbourg, 28 rue Goethe, 67083 Strasbourg Cedex, France. Tel.: +33 3 90 24 20 23; fax: +33 3 90 24 20 28; e-mail: demontigny@gem.u-strasbg.fr

Received 6 May 2009; revised 8 June 2009; accepted 9 June 2009.

DOI:10.1111/j.1567-1364.2009.00540.x

Editor: André Goffeau

Keywords

mitochondria; *Pichia sorbitophila*; tRNA; intron; organelle.

Introduction

Mitochondria are organelles that are ubiquitously present in eukaryotic organisms and always perform the same fundamental role: they allow the production of energy through oxidative phosphorylation or the citric acid cycle (Gray *et al.*, 1999). mtDNA is a small molecule consisting mostly of coding sequences that lack the complexity of the nuclear genome and is used to explore the evolution of higher eukaryotes. The organelle genome database GOBASE (O'Brien *et al.*, 2009) contains $> 900\,000$ mitochondrionencoded sequences from a wide range of eukaryotic taxa, including ascomycetous yeasts that have a remarkable variation in size and gene content.

The length of mtDNA in ascomycetous fungi varies from 20 kb in *Schizosaccharomyces pombe* (Lang *et al.*, 1985) to > 100 kb in *Podospora anserina* (Cummings *et al.*, 1990). In spite of this variation in size, ascomycetous yeasts contain the same core of mitochondrial genes: three subunits of cytochrome oxidase (*COX1*, *COX2* and *COX3*), three subunits of ATP synthase (*ATP6*, *ATP8* and *ATP9* – except *P. anserina*, which lacks the *ATP9*), the apocytochrome *b* (*COB*), the small and the large rRNAs and a set of tRNAs

Abstract

Here, we report the complete nucleotide sequence of the 39 107-bp mitochondrial genome of the yeast *Pichia sorbitophila*. This genome is closely related to those of *Candida parapsilosis* and *Debaryomyces hansenii*, as judged from sequence similarities and synteny conservation. It encodes three subunits of cytochrome oxidase (*COX1*, *COX2* and *COX3*), three subunits of ATP synthase (*ATP6*, *ATP8* and *ATP9*), the seven subunits of NADH dehydrogenase (*NAD1-6* and *NAD4L*), the apocytochrome b (*COB*), the large and small rRNAs and a complete set of tRNAs. Although the mitochondrial genome of *P. sorbitophila* contains the same core of mitochondrial genes observed in the ascomycetous yeasts, those coding for the RNAse P and the ribosomal protein VAR1p are missing. Moreover, the mtDNA of *P. sorbitophila* contains several introns in its genes and has the particularity of possessing an intron, which is not linked to any upstream exon.

able to recognize all codons situated within the proteincoding genes. By contrast, some genes are absent in certain organisms such as those encoding the seven hydrophobic subunits of NADH dehydrogenase complex I in *Saccharomyces cerevisiae* (Foury *et al.*, 1998). Moreover, the genes coding for the ribosomal protein VAR1p and the RNA subunit of the mitochondrial RNAse P (*RPM1*) are missing in most of the completely sequenced hemiascomycetous yeast mitochondrial genomes (Table 1).

Pichia sorbitophila is a hemiascomycetous yeast that has been isolated as a contaminant of 70% sorbitol solution and has the special feature of being highly resistant to osmotic stress, in particular to salt stress (4 M NaCl). Here, we have determined the complete nucleotide (nt) sequence of the mtDNA of *P. sorbitophila* (CBS7064 strain). It is a circular molecule of 39 107 bp, containing 14 protein-coding genes: three subunits of cytochrome oxidase (*COX1*, *COX2* and *COX3*), the apocytochrome b (*COB*), three subunits of ATP synthase (*ATP6*, *ATP8* and *ATP9*) and the seven subunits of NADH dehydrogenase (*NAD1-6* and *NAD4-L*). It also contains the small and the large rRNAs as well as a complete set of tRNAs, whereas the *VAR1* and *RPM1* genes are missing. The mtDNA of *P. sorbitophila* contains several

Table 1. Summary of the mitochondrial features of previously sequenced mitochondrial DNA of hemiascomycetous yeasts

		Genome		СОВ	ATP 6,	NADH	VAR	Intronic			RNAse
Species	Topology	size (nt)	% GC	COX 1, 2, 3	7,8	dehydrogenase	1	ORFs	tRNAs	rRNAs	Р
Saccharomyces cerevisiae	С	85779	17	4	3		1	11	24	2	1
Saccharomyces servazzii	С	30782	22	4	3		1	4	23	2	1
Saccharomyces castelli	С	25753	20	4	3		1	1	23	2	1
Candida glabrata (torpulosis)	С	20063	17	4	3		1	3	23	2	1
Vanderwaltozyma polyspora	С	21684	15	4	3		1		23	NA	NA
Kluyveromyces thermotolerans	С	23584	24	4	3		1	3	24	2	1
Kluyveromyces lactis	С	40291	26	4	3		1	1	22	2	1
Ashbya gossypii	С	23564	18	4	3		1		23	2	
Hanseniaspora uvarum	L	18844	29	4	3			1	23	2	
Pichia canadensis	С	27694	18	4	3	7	1	2	25	2	
Debaryomyces hansenii	С	29462	27	4	3	7		4	25	2	
Candida metapsilosis	L	24152	25	4	3	7		1	24	2	
Candida neerlandica	С	32141	24	4	3	7		4	24	2	
Candida orthopsilosis	С	22528	25	4	3	7		1	24	2	
Candida parapsilosis	L	32745	23	4	3	7		6	24	2	
Candida albicans	С	40420	32	2	3	7		NA	30	2	
Yarrowia lipolytica	С	47916	22	4	3	7		10	27	2	
Candida zemplinina	С	23114	21	4	3		1	7	25	2	
Pichia sorbitophila	С	39107	23	4	3	7		11	25	2	

Mitochondrial DNA can be circular (C) or linear (L) and its size varies from 85 779 bp *in Saccharomyces cerevisiae* to 18 844 bp in *Hanseniaspora uvarum* always with a low GC content. All these species contain the same core of genes except for the genes coding the NADH dehydrogenase, the VAR1 protein and the RNAse P.

NA, not applicable.

introns in its genes and has the particularity of possessing an intron, which is not linked to any upstream exon.

Materials and methods

Strain, genome sequencing and sequence assembly

The *P. sorbitophila* strain used is CBS7064. Complete genome sequences (determined at Genoscope) were automatically assembled from shotgun Sanger sequencing reads using ARACHNE (Jaffe *et al.*, 2003). One 31-kb contig showing identity with known yeast mtDNA was obtained. Further PCR analysis and sequencing reactions were performed in order to complete the mtDNA sequence.

Gene annotation

The BLASTX program (Altschul *et al.*, 1997) was used to compare the *P. sorbitophila* mtDNA with proteins from annotated yeast mitochondrial genomes of *S. cerevisiae* (Foury *et al.*, 1998), *Hanseniaspora uvarum* (Pramateftaki *et al.*, 2006), *Candida orthopsilosis* (Kosa *et al.*, 2006), *Candida metapsilosis* (Kosa *et al.*, 2006), *Pichia canadensis* (Sekito *et al.*, 1995), *Candida parapsilosis* (Nosek *et al.*, 2004), *Candida albicans* (Anderson *et al.*, 2001), *Yarrowia lipolytica* (Kerscher *et al.*, 2001), *Kluyveromyces lactis* (Zivanovic *et al.*, 2005), *Kluyveromyces thermotolerans* (Talla *et al.*, 2005), *Kluyveromyces certer et al.*, 2005), *Kluyveromyces cer*

2005), Candida glabrata (Koszul et al., 2003), Candida neerlandica (Valach et al., 2008), Candida zemplinina (Pramateftaki et al., 2008), Debaryomyces hansenii (unpublished data), Saccharomyces servazii (Langkjaer et al., 2003) and Saccharomyces castellii (Langkjaer et al., 2003). Large and small ribosomal genes were identified by BLASTN comparison with *D. hansenii*, *C. parapsilosis* and *S. cerevisiae* whereas tRNA genes were assigned using two programs: TRNASCAN-SE and ARWEN (Lowe & Eddy, 1997; Laslett & Canbäck, 2008).

Results and discussion

General features and protein-coding genes

The complete mitochondrial sequence from *P. sorbitophila* is a circular molecule of 39 107 bp, characterized by a low GC content of 23% (Fig. 1). As for the other yeast mitochondrial genomes available, the size of this DNA is smaller than the *S. cerevisiae* one, but relatively close to those of *K. lactis* and *C. albicans* (Table 1). This mtDNA molecule is compact because intergenic sequences represent only 7.8% of it. The rest of this genome is composed of 41 genes encoded and transcribed from the same DNA strand except for a dubious ORF (Fig. 2, Table 2). Among these 41 genes, there are 14 protein-coding genes, which encode three subunits of cytochrome oxidase (*COX1*, *COX2* and *COX3*), three subunits of ATP synthase (*ATP6*, *ATP8* and *ATP9*), the apocytochrome



Fig. 1. Circular DNA map of the mitochondrial genome of *Pichia sorbitophila*. The outer circle indicates the localization of protein-coding (blue), tRNA (red) and rRNA (green) genes; some of these contain several introns (white). All genes are transcribed clockwise, except a dubious ORF (yellow). The inner circle represents the transcript units determined thanks to the position of the initiation and cleavage sites as well as the position of the RNA genes, used for RNA punctuation processing mode.

b (*COB*) and the seven subunits of NADH dehydrogenase (*NAD1-6* and *NAD4-L*). The other 27 genes represent the small and the large subunits of the rRNA as well as a complete set of tRNAs.

Noncoding RNAs

rRNAs

Completely sequenced mitochondrial yeast genomes are characterized by the small and the large rRNAs containing none, one or two introns (Jacquier & Dujon, 1983). The *P. sorbitophila* mtDNA possesses these two rRNAs. The large rRNA has two introns and the first one has exactly the same intron–exon boundaries as in *S. cerevisiae* and *K. thermo-tolerans* (Jacquier & Dujon, 1983). The presence of these introns allows us to classify this species as omega-plus (Jacquier & Dujon, 1985). Nevertheless, none of it codes for endonuclease as described for *S. cerevisiae* (Jacquier & Dujon, 1983).

tRNAs

A complete set of 25 tRNA genes has been identified in the mitochondrial genome of *P. sorbitophila*, which is able to

ATP9_PISO	MQLVLAAKYMGAAMATLGLGGAAMGIALVFVALMNGTTRNPSIRATIFPQAMLGFATAEA	60
ATP9_DEHA	MQLALAAKYIGASMATLGLGGAAIGIALVFVALINGTSRNPSLRATLFPQAILGFALAEA	60
ATP9_CAPA	MQIALAAKYIGASIATLGLGGAAIGIALVFVALINGTSRNPSLRSTLFPQAILGFALSEA	60
ATP9_KLTH	MQLVLAAKYIGAGISTIGLLGAGIGIAIVFAALINGVSRNPSLKDTLFPFAILGFALSEA (60
ATP9_SACE	MQLVLAAKYIGAGISTIGLLGAGIGIAIVFAALINGVSRNPSIKDTVFPMAILGFALSEA	60
ATP9_YALI	MQLVLAGKYIGAGLASIGLVGAGIGIAIVFAALINGVSRNPALKGQLFTYSILGFALSEA	60
	:..**:**.:::*** **.:***.**:**.:**::**.:**::**:	
ATP9_PISO	CGLFCLMMSFLLIYAV 76	
ATP9_DEHA	CGLFCLMMSFLLLYAV 76	
ATP9_CAPA	CGLFCLMISFLLLYAV 76	
ATP9_KLTH	TGLFCLMISFLLLYGV 76	
ATP9_SACE	TGLFCLMVSFLLLFGV 76	
ATP9 YALI	TGLFALMIAFLLLYAV 76	
	.**::::.*	

decode all codons present in the ORFs (Table 2). The tRNA^{Met} gene is present in three copies and all of them recognize the same AUG codon. On the other hand, tRNA^{Arg}, tRNA^{Leu} and tRNA^{Ser} genes are represented twice, but are characterized in each case by different anticodons.

As in all sequenced yeast mitochondrial genomes, the *P. sorbitophila* one codes for a $tRNA^{Trp}$ that recognizes UGA codons and, like in the majority of these mtDNA, $tRNA^{Arg}$ decodes CGN codons. All tRNAs contain 7 nt in their anticodon loop as in the universal genetic code, except for $tRNA^{Thr}$ and $tRNA^{Phe}$, which possess 9 nt. Because of the size of these loops, tRNAs that recognize CUN codons are considered to accept leucine, contrary to *S. cerevisiae*, in which these codons are decoded as threonine (Miranda *et al.*, 2006).

Introns

Introns of protein-coding genes

Mitochondrial introns can be classified into two groups (I and II) depending on their structure and splicing mechanism (Lang *et al.*, 2007). They can either be linked in frame with the upstream exon or be free-standing (Haugen *et al.*, 2005). Group I introns code for proteins (RNA maturases or endonucleases) that are classified into four distinct families in accordance to their secondary structure and their protein motifs: LAGLIDADG, GIY-YIG, HNH and His-Cys box. On the other hand, group II introns code for reverse transcriptases that are necessary for their own homing mechanism.

The *P. sorbitophila* mitochondrial genome contains 14 introns in protein-coding ORFs distributed between three genes: *COX1*, *COB* and *NAD5*. Among these 14 introns, 11 code for proteins that have been classified in accordance to their secondary structure by the ANABENCH program (Badidi *et al.*, 2003): 10 of these introns belong to group I, all of which are classified into the LAGLIDADG family, and the last one belongs to group II. In the *COX1* gene, there are five group I and one group II introns, whereas in *COB* and *NAD5*, only group I is present (three and two introns, respectively). All these intron-coding proteins are linked in

Fig. 2. Multiple alignments of the mitochondrial protein ATP9 of *Pichia sorbitophila* (PISO), *Debaryomyces hansenii* (DEHA), *Candida parapsilosis* (CAPA), *Kluyveromyces thermotolerans* (KLTH), *Saccharomyces cerevisiae* (SACE) and *Yarrowia lipolytica* (YALI), using the mitochondrial genetic code of *S. cerevisiae*. Gray and white blocks correspond to the AUA and CUN codons, and allow the assignment of these codons to isoleucine and threonine instead of methionine and leucine.

frame with the upstream exon. They start with an ATG codon and finish with a TAA in the introns, except for the first *COX1* intron that ends with TAG.

Sign of intron mobility

Homing and transposition are two mechanisms that confer groups I and II introns with mobility within the genome. The first one allows for lateral transfer between two alleles, while the second is used to create free-standing introns (Belfort & Perlman, 1995; Chevalier & Stoddard, 2001). One intron found in the mitochondrial genome of *P. sorbitophila* can be defined as a relic of intron mobility by transposition as it does not take place on the transcribed DNA strand and is not located in any gene.

Genetic code and codon usage

Mitochondria often use the genetic code differently from the standard genetic code (Knight et al., 2001). As described previously, the mtDNA of *P. sorbitophila* codes for a tRNA^{Trp} able to decode the UGA codon like in all completely sequenced hemiascomycetous yeast mitochondrial genomes. Furthermore, multiple alignments of proteins from P. sorbitophila, D. hansenii, C. parapsilosis, K. thermotolerans, S. cerevisiae and Y. lipolytica allow us to analyze this genetic code more in detail. They highlight the fact that the P. sorbitophila code differs from that of S. cerevisiae (Fig. 2). Indeed, as proven for C. parapsilosis and P. canadensis, codons AUA and CUN are decoded as isoleucine and leucine instead of methionine and threonine in S. cerevisiae (Sekito et al., 1995; Nosek et al., 2004). This is in direct relation to the size of the anticodon loop of the corresponding tRNAs: they contain 9 and 7 nt in the case of S. cerevisiae and P. sorbitophila, respectively.

The codon usage of exonic and intronic ORFs shows a strong bias toward codons that are rich in A or T, preferentially for those ending with one of these two nucleotides (Table 3). All codons are represented at least once, even if six of these codons are specific to intronic ORFs.

Genes	Coordinates (nt)	Function
tRNA ^{Ser}	37–120	Anticodon: UGA
tRNA ^{Leu}	124–196	Anticodon: UAG
ND4	295–1602	NADH dehydrogenase, subunit 4
tRNA ^{Met}	1602–1673	Anticodon: CAU
tRNA ^{Met}	1679–1749	Anticodon: CAU
small rRNA	1768–3066	15S rRNA gene
tRNA ^{lle}	3097–3167	Anticodon: GAU
tRNA ^{GIn}	3186–3258	Anticodon: UUG
tRNA ^{Asp}	3260–3331	Anticodon:GUC
tRNA ^{ser}	3349–3428	Anticodon: GCU
tRNA ^{Arg}	3432–3503	Anticodon: ACG
COB	Join (3565–3957, 5215–5250, 6388–6438,	Apocytochrome b
	6801–6827, 8045–8293, 9402–9797)	
tRNA ^{Arg}	9851–9921	Anticodon: UCU
tRNA ^{var}	9969–10040	Anticodon: UAC
tRNA ^{Irp}	10043–10114	Anticodon: UCA
tRNA ^{Phe}	10123–10195	Anticodon: GAA
tRNA ^{Lys}	10226–10296	Anticodon: UUU
COX3	10359–11168	Cytochrome c oxidase, subunit 3
ND4L	11233–11487	NADH dehydrogenase, subunit 4L
ND5	Join (11489–12070, 13209–13415, 14636–15502)	NADH dehydrogenase, subunit 5
COX1	Join (15602–15856, 17185–17328, 18514–18834,	Cytochrome c oxidase, subunit 1
	20202–20357, 21555–21653, 24084–24173,	
	25478–25528, 27931–28392)	
large rRNA	Join (28451–30372, 30634–30688, 30975–31281)	21S rRNA gene
tRNA~"	31344–31415	Anticodon: UGC
COX2	31453-32178	Cytochrome c oxidase, subunit 2
tRNA	32180-32250	Anticodon: GUU
ND6	32313-32753	NADH dehydrogenase, subunit 6
ND1	32753-33694	NADH dehydrogenase, subunit 1
ATP9	33785-34015	ATP synthase, subunit 9
	35485-35564	Anticodon: UAA
tRNA ^{-y-}	355/1-35654	Anticodon: GUA
tRNA	35057-35727	Anticodon: GUG
TRIVA ^{Thr}	35/33-35803	Anticodon: CAU
TRIVA"" +DNIAGlu	35813-35883	Anticodon: UGU
+DALAGIV	25902-55972 25090-26051	Anticodon. UCC
think for the second se	260E2 2612E	Anticodon: UCC
τηνα · +DNIAPro	26202 26275	Anticodon: UGG
	2622A 26A70	
AIFO	20224-304/U 2007	ATP synthase, subunit 8
	27221 20E20	ATP Synthase, Subunit 6
	3/32I-38038	NADH denydrogenase, subunit 2
ND3	38702-39085	NADH denydrogenase, subunit 3

Coordinates of protein-coding genes correspond to the start and stop codons whereas tRNA coordinates (light gray) have been deduced thanks to their secondary structure predictions.

Coordinates of rRNA genes (dark gray) have been determined by similarity to *Debaryomyces hansenii*. The anticodon of each tRNA has been ascertained by the TRNASCAN-SE and ARWEN programs.

Synteny

Gene order in the mitochondrial genome of *P. sorbitophila* is particularly close to that of *D. hansenii* and *C. parapsilosis* (Fig. 3). Indeed, large blocks of genes are conserved between these species. Among these syntenic blocks, some genomic rearrangement can be highlighted as inversions. For example, genes located between the small rRNA and the tRNA^{Ser} are found in different positions in the mitochondrial genomes of *D. hansenii* and *C. parapsilosis* (Fig. 3).

By contrast, comparison of the synteny between the mtDNA of *P. sorbitophila* and other sequenced yeast

5

Codon	a.a.	Exonic	Intronic												
TTT	F	151	294	ТСТ	S	108	144	TAT	Y	198	375	TGT	С	36	64
TTC	F	104	94	TCC	S	3	13	TAC	Υ	30	37	TGC	С	2	7
TTA	L	518	550	TCA	S	92	134	TAA	*	14	10	TGA	W	52	97
TTG	L	1	14	TCG	S	1	5	TAG	*	-	1	TGG	W	2	8
CTT	Т	38	50	CCT	Р	89	100	CAT	Н	63	124	CGT	R	20	32
CTC	Т	1	11	CCC	Р	1	5	CAC	Н	12	22	CGC	R	-	2
CTA	Т	20	36	CCA	Р	51	58	CAA	Q	60	112	CGA	R	-	2
CTG	Т	1	6	CCG	Р	-	1	CAG	Q	-	12	CGG	R	1	1
ATT	1	147	273	ACT	Т	116	155	AAT	Ν	146	536	AGT	S	89	139
ATC	1	52	50	ACC	Т	1	8	AAC	Ν	29	46	AGC	S	3	9
ATA	Μ	298	459	ACA	Т	86	118	AAA	Κ	87	451	AGA	R	65	135
ATG	Μ	113	162	ACG	Т	1	4	AAG	Κ	1	35	AGG	R	1	10
GTT	V	107	103	GCT	А	131	127	GAT	D	82	193	GGT	G	147	225
GTC	V	7	13	GCC	А	6	13	GAC	D	8	3	GGC	G	-	3
GTA	V	130	155	GCA	А	78	99	GAA	Е	89	169	GGA	G	90	129
GTG	V	2	12	GCG	А	6	4	GAG	Е	14	23	GGG	G	13	7

Table 3. Codon usage of exonic and intronic protein-coding genes in the mitochondrial genome of Pichia sorbitophila

Amino acids (a.a.) are represented with one-letter codes.

*Stop codon.



Fig. 3. Synteny of different hemiascomycetous yeast mitochondrial genomes sequenced. Random synteny blocks have been designed with *Pichia sorbitophila* as the origin. The order of genes of *P. sorbitophila*, *Debaryomyces hansenii* and *Candida parapsilosis* is well conserved, showing a close relationship of these species. Among their mtDNA, some genomic rearrangements such as translocations are found.

mitochondrial genomes show large differences, for example with *Y. lipolytica* (Fig. 3). Nevertheless, several groups of genes are conserved among hemiascomycetous yeast species and are always located one beside the other: *ATP6-ATP8*, *NAD1-NAD6*, *NAD2-NAD3* and *NAD5-NAD4-L*. These pairs of genes are also situated in the same polycistronic transcripts.

Putative transcription initiation and endonucleolytic cleavage sites

The nonanucleotide motif sequence TATAAGTAA is identified as the transcription initiation site in *S. cerevisiae*, and similar sequences are found in other yeast mitochondrial genomes (Schäfer, 2005). The TATAAGAA motif is found twice in the mtDNA of *P. sorbitophila*, located upstream from the genes coding for the *COX3* protein and tRNA^{Leu}, leading to two putative preprocessing transcripts.

On the other hand, the dodecanucleotide motif AATAA-TATTCTT has been defined as the sequence that is required for RNA processing in *S. cerevisiae*; these sequences are present downstream from most of the protein-coding genes (Schäfer, 2005). Similar motifs are found among yeast mitochondria. In *P. sorbitophila*, the motif HWTA-TATTCWT is found six times, downstream from the protein-coding genes *COB*, *NAD5*, *NAD1* and *ATP9*, and the small rRNA and tRNA^{Phe}. Moreover, the localization of the tRNA genes in this genome leads us to think that the RNA punctuation mode is also used for RNA processing (Schäfer, 2005). Indeed, they are present in clusters except for few tRNAs that flank the *NAD4* and *COX2* genes, leading to the release of the monocistronic mRNA (Fig. 1).

Considering this RNA punctuation processing mode as well as the initiation and cleavage motif sites, five monocistronic (*NAD4*, *COB*, *COX1*, *COX2* and *ATP9*), one bicistronic (*NAD6-NAD1*) and two polycistronic transcripts (*COX3-NAD4L-NAD5* and *ATP8-ATP6-NAD2-NAD3*) have been defined and are cotranslated (Fig. 1).

Conclusion

The mitochondrial genome of *P. sorbitophila* is similar to many fungal mtDNAs in that it contains the usual set of genes. Furthermore, it also possesses the seven hydrophobic subunits of the NADH dehydrogenase, but *VAR1* and *RPM1* genes are missing. In addition, the high proteic similarity rate and the syntenic block conservation show a very close phylogenic relationship with *D. hansenii* and *C. parapsilosis*. The high proportion of introns in the mtDNA of *P. sorbitophila* and the dubious ORF (corresponding to the free-standing intron) can explain the difference of length between these species (about 39kb against 29 and 32 kb, respectively). The mtDNA of *P. sorbitophila* also allows confirmation of the fact that the *S. cerevisiae* mitochondrial genetic code is not conserved across the hemiascomycetes fungi. Indeed, codons AUA and CUN are decoded as isoleucine and leucine like in the universal code instead of methionine and threonine as in the case of *S. cerevisiae* mitochondria.

Many more questions remain unanswered. For example, no origins of replication were found in the *P. sorbitophila* mitochondrial genome from sequence comparison with the consensus pattern from *S. cerevisiae*. Moreover, the DNA sequence of *P. sorbitophila*, completely sequenced and under annotation, will probably highlight some nuclear mtDNA such as *VAR1* and *RPM1* genes. It will also provide useful information for further studies on its function in the future.

Acknowledgements

This work was a part of the Génolevures Consortium (CNRS-GDR2345) and was supported by the ANR (GEN-ARISE). We thank all the members of the Génoscope, who have contributed to the whole genome sequencing of *P. sorbitophila*, out of which the present mitochondrial sequence was assembled. We also thank Joern Pütz for his help in the tRNA structure prediction and Daniel Medina Stacey for proofreading.

Statement

The sequence has been deposited in EMBL under the accession number FN356025.

References

- Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.
- Anderson JB, Wicken C, Khan M, Cowen LE, Federspiel N, Jones T & Kohn LM (2001) Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans. J Bacteriol* **183**: 865–872.
- Badidi E, De Sousa C, Lang BF & Burger G (2003) AnaBench: a Web/CORBA-based workbench for biomolecular sequence analysis. *BMC Bioinformatics* **4**: 63–72.
- Belfort M & Perlman PS (1995) Mechanisms of intron mobility. *J Biol Chem* **270**: 30237–30240.
- Chevalier BS & Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/ intein mobility. *Nucleic Acids Res* **29**: 3757–3774.
- Cummings JD, McNally KL, Domenico JM & Matsuura ET (1990) The complete DNA sequence of the mitochondrial genome of *Podospora anserina*. *Curr Genet* **17**: 375–402.

Foury F, Roganti T, Lecrenier N & Purnelle B (1998) The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. *FEBS Lett* **440**: 325–331.

Gray MW, Burger G & Lang BF (1999) Mitochondrial evolution. *Science* **283**: 1476–1481.

Haugen P, Simon DM & Bhattacharya D (2005) The natural history of group I introns. *Trends Genet* **21**: 111–119.

Jacquier A & Dujon B (1983) The intron of the mitochondrial 21S rRNA gene: distribution in different yeast species and sequence comparison between *Kluyveromyces thermotolerans* and *Saccharomyces cerevisiae*. *Mol Gen Genet* **192**: 487–499.

Jacquier A & Dujon B (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell* **41**: 383–394.

Jaffe DB, Butler J, Gnerre S, Mauceli E, Lindblad-Toh K, Mesirov JP, Zody MC & Landler ES (2003) Whole-genome sequence assembly for mammalian genomes: Arachne 2. *Genome Res* 13: 91–96.

Kerscher S, Durstewitz G, Casaregola S, Gaillardin C & Brandt U (2001) The complete genome of *Yarrowia lipolytica. Comp Funct Genom* **2**: 80–90.

Knight RD, Landweber LF & Yarus M (2001) How mitochondria redefine the code. *J Mol Evol* **53**: 299–313.

Kosa P, Valach M, Tomaska L, Wolfe KH & Nosek J (2006) Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. *Nucleic Acids Res* 34: 2472–2481.

Koszul R, Malpertuy A, Frangeul L, Bouchier C, Wincker P, Thierry A, Duthoy S, Ferris S, Hennequin C & Dujon B (2003) The complete mitochondrial genome sequence of the pathogenic yeast *Candida (Torulopsis) glabrata. FEBS Lett* **534**: 39–48.

Lang BF, Ahne F & Bonen L (1985) The mitochondrial genome of the fission yeast *Schizosaccharomyces pombe*. The cytochrome b gene has an intron closely related to the first two introns in the *Saccharomyces cerevisiae* cox1 gene. *J Mol Biol* **184**: 353–366.

Lang BF, Laforest MJ & Burger G (2007) Mitochondrial introns: a critical view. *Trends Genet* **23**: 119–125.

Langkjaer RB, Casaregola S, Ussery DW, Gaillardin C & Piskur J (2003) Sequence analysis of three mitochondrial DNA molecules reveals interesting differences among *Saccharomyces* yeasts. *Nucleic Acids Res* **31**: 3081–3091.

Laslett D & Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* 24: 172–175.

Lowe TM & Eddy SR (1997) tRNAscan-SE: a program for improved detection of RNA genes in genomic sequence. *Nucleic Acids Res* 25: 955–964.

Miranda I, Silva R & Santos MAS (2006) Evolution of the genetic code in yeasts. *Yeast* 23: 203–213.

Nosek J, Novotna M, Hlavatovicova Z, Ussery DW, Fajikus J & Tomaska L (2004) Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast *Candida parapsilosis*. *Mol Genet Genomics* **272**: 173–180.

O'Brien EA, Zhang Y, Wang E, Marie V, Badejoko W, Lang BF & Burger G (2009) GOBASE: an organelle genome database. *Nucleic Acids Res* **37**: 946–950.

Pramateftaki PV, Kouvelis VN, Lanaridis P & Typas MA (2006) The mitochondrial genome of the wine yeast *Hanseniaspora uvarum*: a unique genome organization among yeast/fungal counterparts. *FEMS Yeast Res* **6**: 77–90.

Pramateftaki PV, Kouvelis VN, Lanaridis P & Typas MA (2008) Complete mitochondrial sequence of the wine yeast *Candida zemplinina*: intraspecies distribution of a novel group-IIB1 intron with eubacterial affiliations. *FEMS Yeast Res* 8: 311–327.

Schäfer B (2005) RNA maturation in mitochondria of *S. cerevisiae* and *S. pombe. Gene* **354**: 80–85.

Sekito T, Okamoto K & Kitano H (1995) The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. *Curr Genet* 28: 39–53.

Talla E, Anthouard V, Bouchier C, Frangeul L & Dujon B (2005) The complete mitochondrial genome of the yeast *Kluyveromyces thermotolerans. FEBS Lett* **579**: 30–40.

Valach M, Tomaska L & Nosek J (2008) Preparation of yeast mitochondrial DNA for direct sequence analysis. *Curr Genet* 54: 105–109.

Zivanovic Y, Wincker P, Vacherie B, Bolotin-Fukuhara M & Fukuhara H (2005) Complete nucleotide sequence of the mitochondrial DNA from *Kluyveromyces lactis. FEMS Yeast Res* **5**: 315–322. **RESEARCH ARTICLE**

Complete mitochondrial genome sequence of the yeast *Pichia farinosa* and comparative analysis of closely related species

Paul P. Jung · Anne Friedrich · Jean-Luc Souciet · Véronique Louis · Serge Potier · Jacky de Montigny · Joseph Schacherer

Received: 26 July 2010 / Revised: 20 August 2010 / Accepted: 23 August 2010 © Springer-Verlag 2010

Abstract Yeasts of the *Pichia* genus have been isolated from different natural environments. Phylogenies based on multigene sequence analysis have shown that the genus is polyphyletic. Some species of this genus are member of the CTG group. In order to have a better insight into the relationship among species assigned to the yeast genera Pichia into the CTG group, we first sequenced the mitochondrial genome of the osmotolerant yeast Pichia farinosa. We then compared this genome with mitochondrial genomes of yeasts of the CTG group. The P. farinosa mitochondrial DNA is a circular-mapping genome of 32,065 bp, which contains 43 genes transcribed from both strands. It contains a complete set of tRNAs, the small and the large rRNAs, as well as 14 protein-coding genes. Yeasts of the CTG group contain the same core of mitochondrial genes. Phylogenetic analysis based on mitochondrial sequences clearly shows that the CTG group is divided into two distinct clades: the first one contains diploid Candida species, whereas the second mainly contains haploid Pichia species. Moreover, this analysis provides clear evidence that Pichia farinosa and

Communicated by L. Tomaska.

J. de Montigny and J. Schacherer contributed equally to this work.

Accession number: the sequence has been deposited in EMBL under accession number FN870377.

Electronic supplementary material The online version of this article (doi:10.1007/s00294-010-0318-y) contains supplementary material, which is available to authorized users.

P. P. Jung \cdot A. Friedrich \cdot J.-L. Souciet \cdot V. Louis \cdot S. Potier \cdot J. de Montigny \cdot J. Schacherer (\boxtimes)

Department of Genetics, Genomics and Microbiology,

University of Strasbourg, CNRS, UMR7156, Strasbourg, France e-mail: schacherer@unistra.fr

Pichia sorbitophila, which were known to be unique species, are two distinct species.

Keywords Mitochondrial genome · *Pichia farinosa* · CTG group · Phylogeny

Abbreviation

Mt Mitochondrial

Introduction

Hemiascomyceteous yeasts are fungi that spread from the well-studied *Saccharomyces cerevisiae* to *Yarrowia lipolytica*. This subphylum of Ascomycota is divided into three major groups: the *Saccharomycetaceae* family, the CTG group and the *Dipodasceae* family (Dujon 2010). They are composed by more than a thousand species that cover large phylogenetic distances, and hence, they represent a unique opportunity for evolutionary genomics studies.

The Saccharomycetaceae family, which has been phylogenetically well studied and divided into 14 clades, contains notably the genera Saccharomyces, Kluyveromyces and Zygosaccharomyces (Kurtzman and Robnett 2003). The Dipodasceae and related families comprise species such as Arxula adenivorans, Pichia pastoris and Yarrowia lipolytica (Dujon 2010). The CTG group is composed predominantly of Candida and Pichia species that have the specificity to decode CUG codons as serine rather than leucine as described in the universal genetic code (Ohama et al. 1993; Santos and Tuite 1995). Most of the human pathogens are known to belong to this group (e.g. Candida albicans and Candida parapsilosis). In addition to Candida species, this group also contains the marine yeast Debaryomyces hansenii, the xylose-fermenting yeast Pichia stipitis and other *Pichia* species such as *Pichia anomala* and *Pichia farinosa*, which were collected from immunocompromised patients or people affected by a cancer (Adler et al. 2007; Bakir et al. 2004)

The CTG group is monophyletic. Nevertheless, species belonging to this group are taxonomically classified in different genus. Some genera contain species, which do not necessarily share the same properties such as the codon reassignment previously mentioned. This is particularly the case within the Pichia genus, which is polyphyletic. This genus is therefore considered as artificial, and placement of Pichia species into phylogenetically genera has been partly resolved from the sequence analysis (Kurtzman and Fell 1998). Recently, Pichia species that form coenzyme Q-9 were phylogenetically analyzed from D1/D2 domains of the large subunit (LSU) and small subunit (SSU) rRNA genes (Kurtzman and Suzuki 2010). Based on their phylogenetic analysis, the authors proposed five new genera such as Meyerozyma and Millerozyma. The latest includes the species P. farinosa. Both genera (Meyerozyma and Millerozyma) contain yeasts that are classified into the CTG group and have been isolated from different environments. P. farinosa, which is also osmotolerant, has been found in a Jopen beer and in some patient with cancer, whereas Pichia sorbitophila has been isolated from a contaminant of a 70% sorbitol solution (Rodriques de Miranda et al. 1980; Adler et al. 2007). These two yeasts have physiological differences, but previous taxonomic analysis described them as the same species (Kurtzman and Fell 1998; Maresová and Sychrová 2003).

To assess genetic relationships between P. farinosa and *P. sorbitophila* as well as among species of the CTG group, we compared the mitochondrial DNA sequences, considering the high degree of sequence conservation of these organelles. To better characterize mitochondrial genome evolution in the CTG group, we first completely sequenced and annotated the mtDNA of Pichia farinosa (CBS 185). Moreover, we have also annotated the mt genomes of closely related species, which have already been sequenced: Pichia stipitis, Pichia guilliermondii, Candida tropicalis and Lodderomyces elonisporus (Jeffries et al. 2007; Butler et al. 2009). All these mt genomes were compared to mtDNA of related yeasts of the CTG group: Debaryomyces hansenii, Pichia sorbitophila, Candida albicans, Candida parapsilosis, Candida metapsilosis, Candida orthopsilosis and Candida neerlandica (Sacerdot et al. 2008; Jung et al. 2009; Anderson et al. 2001; Nosek et al. 2004, Kosa et al. 2006; Valach et al. 2008).

These mt genomes are variable in length and GC content. Nevertheless, they all contain the same core of mitochondrial genes. The results of our phylogenetic analysis show that the CTG group is divided into two distinct clades. Finally, this analysis also provides clear evidence that *P. sorbitophila* and *P. farinosa* are two distinct species.

Materials and methods

Strains, genome sequencing and genome assembly

The *Pichia farinosa* strain used is CBS 185. Genomic DNA was extracted, and whole genome shotgun sequences were obtained using the Illumina genome analyzer. A total of 70,695,778 paired-end reads of 36 bp length were generated. Genome sequences were automatically assembled from sequencing reads using SOAP2 software (Li et al. 2009). One 32-kb scaffold showing identity with known yeast mtDNA was obtained. Further PCR analysis and sequencing reactions were performed in order to complete the mtDNA sequence.

In addition, scaffolds corresponding to mitochondrial DNA of *Pichia stipitis* strain CBS 6054, *Pichia guillier-mondii* strain ATCC 6260, *Candida tropicalis* strain MYA-3404 and *Lodderomyces elongisporus* strain NRLL YB-4239 were also annotated (Jeffries et al. 2007; Butler et al. 2009).

Gene annotation

The BLASTX program (Altschul et al. 1997) was used to compare mtDNA with proteins from annotated yeast mitochondrial genomes of *Candida parapsilosis* (Nosek et al. 2004), *Debaryomyces hansenii* (Sacerdot et al. 2008), *Kluyveromyces thermotolerans* (Talla et al. 2005), *Pichia canadensis* (Sekito et al. 1995) and *Pichia sorbitophila* (Jung et al. 2009). Large and small ribosomal RNA genes were identified by BLASTN comparison with *Saccharomyces cerevisaie* (Foury et al. 1998), *Candida glabrata* (Koszul et al. 2003), *D. hansenii* and *C. parapsilosis*, whereas tRNA genes were assigned combining three programs: BLASTN, TRNAscan-SE and ARWEN (Lowe and Eddy 1997; Laslett and Canbäck 2008).

Phylogenetic studies

Phylogenetic studies were made using the mtDNA of the yeasts analyzed in this study. Amino acid and nucleic sequences were aligned using MUSCLE (Edgar 2004) and TRANalign programs, respectively (available on the EMBOSS package, http://bips.u-strasbg.fr/EMBOSS/). Phylogenetic relationships among yeast species were analyzed by both the maximum-likelihood and bayesian methods. The phylogenetic trees were constructed from an unambiguously alignment of the concatenated protein sequences of Cox2, Cox3, Atp6, Atp8, Atp9, Nad1, Nad2,

Fig. 1 Circular map of the mitochondrial genome of the yeast *P. farinosa*. The outer *circle* indicates genes that are encoded by both strands



Nad3, Nad4, Nad4 l, Nad5 and Nad6 using a total of 3,242 amino acid positions. Maximum-likelihood and bayesian analyses were performed by PHYML and MRBAYES, respectively (Guindon and Gascuel 2003; Dereeper et al. 2008).

Results and discussion

General organization: genome size, gene content and gene order

In a previous study, we reported the complete nucleotide sequence of the mitochondrial genome of the yeast *P. sorbitophila* CBS 7064 (Jung et al. 2009). In order to provide comparative data from another *Pichia* species to better define the mitochondrial genomes in the CTG group, we have completely sequenced and annotated the mtDNA of *P. farinosa* CBS 185. This mt DNA is a circular-mapping genome of 32,065 bp, characterized by a low GC content of 25% (Fig. 1). The mtDNA molecule is compact, and the size of it is smaller than the *S. cerevisiae* one, but relatively close to those of *P. sorbitophila* (Table 1). It is composed of 43 genes that represent 80% of it and are transcribed by

both strands. Among the 43 genes, there are 14 proteincoding genes: three subunits of cytochrome c oxidase (COX1, COX2 and COX3), the apocytochrome b (COB), three subunits of ATP synthase (ATP6, ATP8 and ATP9) and seven subunits of NADH dehydrogenase complex 1 (NAD1-6 and NAD4L). LSU and SSU genes code the large and the small rRNA genes, whereas the other genes represent 27 tRNA genes able to carry all amino acids.

To have a global view of the general organization of mitochondrial genomes, we have annotated four mt genomes, which have already been sequenced: Pichia stipitis, Pichia guilliermondii, Candida tropicalis and Lodderomyces elonisporus (Jeffries et al. 2007; Butler et al. 2009) (Fig. 2; Table S1). All these genomes were compared to mtDNA of yeasts of the CTG group, which have already been analyzed (Sacerdot et al. 2008; Jung et al. 2009; Anderson et al. 2001; Nosek et al. 2004). The general features of all these mt genomes are shown in Table 1. A low GC content characterizes all these sequences, and their sizes are variable. Their length spreads from almost 24 kb in P. guillermondii to more than 50 kb in C. tropicalis. Most of the size variation (from 23 to 39 kb) is due to the presence or absence of introns. However, the large size variation (from 39 to more than 50 kb) can be explained partly

Species	Strain	Genome size (nt)	Ploidy	GC(%)	COB COX1, 2, 3	ATP 6, 7, 8	NADH dehydrogenase complex	Intergenic region (%)	i-ORFs	tRNAs	rRNA
Pichia species											
Pichia farinosa	CBS 185	32,065	n	27	4	3	7	20	4	27	2
Pichia sorbitophila	CBS 7064	39,107	2 <i>n</i>	23	4	3	7	8	11	25	2
Pichia stipitis	CBS 6054	30,649	n	25	4	3	7	16	4	24	2
Pichia guilliermondii	ATCC 6260	23,890	n	25	4	3	7	16	1	25	2
Other yeasts of the CTG g	roup										
Candida albicans	SC 5314	40,420	2 <i>n</i>	32	4	3	7	44	NA	30	2
Candida metapsilosis	MCO 448	24,152	2 <i>n</i>	25	4	3	7	12	1	24	2
Candida neerlandica	NRRL Y 27057T	32,141	2 <i>n</i>	24	4	3	7	26	4	24	2
Candida orthopsilosis	MCO 456	22,528	2 <i>n</i>	25	4	3	7	13	1	24	2
Candida parapsilosis	CBS 7157	32,745	2 <i>n</i>	23	4	3	7	8	6	24	2
Candida tropicalis	MYA 3404	50,304	2 <i>n</i>	37	4	3	7	49	3	30	2
Debaryomyces hansenii	CBS 767	29,462	n	27	4	3	7	25	4	25	2
Lodderomyces elongisporus	NRLL YB-4239	35,601	2 <i>n</i>	29	4	3	7	42	0	25	2

 Table 1
 General features of mitochondrial genomes of yeasts of the CTG group

by the length of the intergenic regions. Non-coding DNA accounts for only 8% of the mtDNA of *Pichia sorbitophila*, whereas it accounts for 49% in *C. tropicalis*. The function of these non-coding regions, if any, is unknown. All these genomes contain the same protein-coding genes as those of *P. farinosa*, the small and large genes (*LSU* and *SSU*) and a set of tRNA genes. In contrast to yeasts of the *Saccharomy-cetacea* family, it is interesting to note that the genes coding for the ribosomal protein Var1p and the RNA subunit of the mitochondrial RNAseP (*RPM1*) are missing in all the mtDNA of yeast of the CTG group.

Gene order in the mtDNA of P. farinosa is particularly close to P. stipitis, P. guillermondii and P. sorbitophila. In contrast, it is dissimilar to that observed in the mtDNA of Candida species (data not shown). Information derived from this study indicates that Pichia species are more closely related to each other than Candida species. The close relationship of *P. farinosa* and *P. stipitis* is further reflected in sequence similarities of mitochondrial genes and in phylogenetic analyses grouping them as sister species of exclusion of *P. sorbitophila* (see Fig. 5). When comparing gene organization across Pichia species and D. hansenii, large blocks of synteny are conserved (Fig. 3). Some of them are conserved in all species such as the block from LSU to NAD1 (Fig. 3). Genomic rearrangements as inversions can be highlighted among these syntenic blocks. Finally, gene pairs ATP8-ATP6, NAD1-NAD6, NAD2-NAD3 and NAD4L-NAD5 are also juxtaposed and co-transcribed in all mt genomes of species analyzed here. Nevertheless, this situation was previously described in other species of yeast such as Yarrowia lipolytica and is not specific to yeasts of the CTG group, but specific to hemiascomyceteous yeasts that possess these genes in their mtDNA (Kerscher et al. 2001).

Introns of protein-coding genes

Two groups (I and II) of mitochondrial introns were defined depending on their structure and splicing mechanism (Lang et al. 2007). Introns can either be linked in frame with the upstream exon or be free-standing (Haugen et al. 2005). Group I introns code for proteins (RNA maturases or endonucleases), which are classified into four distinct families in accordance to their secondary structure and their proteic motifs: LAGLIDADG, GIY-YIG, HNH and His-Cys box. In contrast, group II introns code for reverse transcriptases that are necessary for their own homing mechanism.

As mentioned previously, mitochondrial genomes presented here contain the same 14 protein-coding genes. The number of intron is variable. As described in our previous study, *P. sorbitophila* contains 14 introns distributed in four genes: *COX1*, *COB*, *LSU* and *NAD5* (Jung et al. 2009). In contrast, no introns are found in the protein-coding genes of the mt genome of *L. elongisporus*.

Nevertheless, several coding introns are found in *COX1* and *COB* genes in the mt genomes analyzed here. All these introns are linked in frame with previous exon. They belong to class I intron and are divided into two superfamilies based on their LAGLIDADG or GIY-YIG patterns (Chevalier and Stoddard 2001). In *P. farinosa* and *P. stipitis*, four introns that encode endonucleases are distributed between the *COX1* and *COB* genes. Two introns present in *P. farinosa* contain the GIY-YIG motif, whereas the two others contain the LAGLIDADG motif. In *P. stipitis*, all the

Fig. 2 Linear maps of the mitochondrial DNA of P. stipitis, P. guilliermondii, C. tropicalis and L. elongisporus. All genes are transcribed by the minus strand in P. stipitis and P. guilliermondii but by both strands in C. tropicalis and L. elongisporus. Mt genomes of P. stipitis and P. guilliermondii are supposed to be circular because genes LSU and COXI are found respectively on each side of the corresponding scaffold. By contrast, topologies of mtDNA of C. tropicalis and L. elongisporus are not defined



introns are members of the LAGLIDADG superfamily. In the mtDNA of *C. tropicalis*, only the *COX1* gene contains three LAGLIDADG coding introns. In *P. guilliermondii*, we only found one coding GIY-YIG intron in the *COB* gene, but this genome could have more endonucleases because the *COX1* gene is incomplete, and a part of it is found at each side of the mt scaffold (Fig. 2).

Finally, some hemiascomyceteous fungi such as *S. cerevisiae* or *K. thermotolerans* have the particularity to encode an endonuclease, which is translated from a coding intron residing in the mitochondrial *LSU* gene (Jacquier and Dujon 1983). Here, only *P. guilliermondii* contains one intron in the *LSU* gene, but it does not encode any nuclease.

Transcription and RNA maturation

In the mitochondrial genome of *S. cerevisiae*, transcription starts are defined by the presence of the nonanucleotide sequence TATAAGTAA (Schäfer 2005). By the same way,

processing can occur at the dodecanucleotide motif AATAATATTCTT or with the cleavage of tRNA in transcript units. In mtDNAs of species annotated here, tRNA genes are not only distributed in clusters but are also dispersed and surround several protein-coding genes such as COX3 in P. farinosa (Figs. 1, 2). This suggests that the tRNA punctuation mode is used for RNA processing (Schäfer 2005). Moreover, the core motif ATAATATT, similar to the dodecanucleotide pattern AATAATATTCTT described in S. cerevisiae, is found in the mtDNA of P. farinosa downstream COX2, COX3, NAD1 and LSU genes and is a potential cleavage sequence (Fig. 1). The same motif is found in mtDNA of P. stipitis, P. guilliermondii and L. elongisporus downstream protein-coding genes by in silico approach. Nevertheless, this motif is absent in mt genome of C. tropicalis.

Considering these two kinds of processing, monocistronic or polycistronic transcripts can be generated. As described for other yeasts such as *C. metapsilosis* and



Fig. 3 Conservation of mitochondrial synteny among *Pichia* species. Large blocks of synteny are conserved between *D. hansenii* and *Pichia* species analyzed here showing a close relationship between these species

C. orthopsilosis (Kosa et al. 2006), polycistronic transcripts always contain gene pairs *ATP8–ATP6*, *NAD1–NAD6*, *NAD2–NAD3* and *NAD4L–NAD5* that seem to be also co-translated because of the small intergenic sequence between these genes (Table S1).

Genetic code

Mitochondria often use different genetic code (Knight et al. 2001). As in all completely sequenced hemiascomyceteous yeasts, mtDNA of *Pichia* species code for a tRNA^{Trp} able to decode the UGA codon. Moreover, this mitochondrial genetic code also differs in *P. sorbitophila* by the fact that codons AUA and CUN are decoded as isoleucine and leucine instead of methionine and threonine (Jung et al. 2009). When comparing multiple alignments of mitochondrial proteins from different yeasts such as *S. cerevisaie*, *C. glabrata*, *K. thermotolerans* or *Y. lipoytica*, conserved isoleucine

are found at the level of AUA codons in other *Pichia* species and *D. hansenii* showing that AUA codons are decoded as isoleucine (Fig. 4). By the same way, CUN codon can be assigned to leucine.

Because few species of the *Pichia* genus have been analyzed, it is impossible to conclude that this genetic code is a particularity of these species. Nevertheless, it seems to be specific of yeasts of the CTG group because the same genetic code has been demonstrated in *C. parapsilosis* and supposed in *C. tropicalis* and *L. elongisporus* (Nosek et al. 2004).

Phylogenetic analysis

Mt genomes of *Pichia* species have the same genetic code than yeasts of the CTG group, and syntenic approach shows strong relationships between them. In order to better assess genetic relationships between *P. farinosa* and *P. sorbitophila*



Fig. 4 Multiple alignment of protein. Here, there is an example of alignment of the Atp9 protein of different species. Translation of the *ATP9* gene of *P. farinosa* was performed using the mitochondrial genetic code described in *S. cerevisiae*. We used available *ATP9* sequences: *S. cerevisiae* (NP_009319.1), *C. glabrata* (NP_818784.1), *K. thermotolerans* (YP_184726.1) and *Y. lipolytica* (NP_075437.1). The conservation of some isoleucine in *S. cerevisiae, C. glabrata*,

as well as among species of the CTG group, we inferred a phylogenetic tree using amino acid of 12 concatenated mitochondrial proteins (Fig. 5). The phylogenetic tree was constructed from an unambiguously alignment of the concatenated protein sequences of Cox2, Cox3, Atp6, Atp8, Atp9, Nad1, Nad2, Nad3, Nad4, Nad4 l, Nad5 and Nad6 using a total of 3,242 amino acid positions. Maximumlikelihood and bayesian phylogenies were constructed and compared. Overall, there is a high congruence between the two trees (Fig. 5). There is a minor conflict, however, between our two trees regarding the positioning of C. neerlandica. The maximum-likelihood tree places C. neerlandica beside C. albicans with a boostrap support value of 53%. Conversely, the bayesian method infers that C. neerlandica is more closely related to C. tropicalis with a boostrap support value of 100%. Nevertheless, the two methods inferred a congruent topology, which clearly shows that the CTG group is divided into two subgroups with a bootstrap support value of almost 100%. The first subgroup contains the Pichia species as well as D. hansenii. The second group contains the diploid Candida species analyzed here (Fig. 5).

Interestingly, the *Pichia* subgroup is also divided. Our topology places *P. sorbitophila* at the base of the *Pichia* subgroup with a high boostrap support. This study shows that *P. farinosa*, *P. stipitis*, *D. hansenii* and *P. guilliermondii* are more closely related to each other than either is to *P. sorbitophila* (Fig. 5). Interestingly, all these species but *P. sorbitophila* are haploids. In fact, *P. sorbitophila* is a natural hybrid, which may explain the observed topology (Dujon 2010).

Consequently, *P. sorbitophila* and *P. farinosa* are obviously two distinct species. This analysis clearly shows that *P. farinosa* and *P. stipitis* are sister species as they are grouped together with a bootstrap support value of 100%. *P. stipitis* is used to ferment xylose from plant

K. thermotolerans and *Y. lipolytica* corresponds to methionine in *P. farinosa*. The codon corresponding to this amino acid in *P. farinosa* is AUA, suggesting the reassignment of this codon to isoleucine. By the same way, the same reassignment is found in mt genomes of *P. stipitis*, *P. guilliermondii*, *C. tropicalis* and *L. elongisporus*. Moreover, in all these five species, codons CUN can be reassigned to leucine rather than threonine as described in *S. cerevisiae*

lignocellulose and to produce bioethanol with the highest efficiency comparing to other microorganisms (Jeffries et al. 2007). Nevertheless, species such as D. hansenii and Candida intermedia are also known to grow on xylose (Nobre et al. 1999; Leandro et al. 2006) as well as novel yeast species closely related to P. stipitis (i.e. Pichia segobiensis, Candida shehatae, Pachysolen tannophilus and Enteroramus dimorbhus) (Jeffries 2006; Tsui et al. 2008). Knowledge of genomes of these organisms is important to know if these organisms consume this pentose by the same way or by different manner as in P. stipitis. Recently, yeast genomes analyzed by Butler et al. (2009) show that some nuclear genes have wide similarities with protein-coding genes occurring in xylose metabolism and transport in *P. stipitis.* To know if the use of xylose as carbon source is a particularity of all species of the CTG group, many more species have to be studied. Notably, the knowledge of the nuclear genome of P. farinosa, which is very close to P. stipitis, could bring new clues.

Conclusion

The comparative analysis presented here allows a more comprehensive understanding of the mitochondrial genomes found in the yeast of the CTG group. Even if their length is variable, mt genomes of these yeasts contain the same 14 protein-coding genes. The variable intron content and the length of the intergenic regions can explain this size variation. In contrast to yeasts of the *Saccharomycetacea* family, the genes coding for the ribosomal protein Var1p and the RNAseP (*RPM1*) are missing in all the mtDNA of yeast of the CTG group.

Our phylogenetic analysis based on mitochondrial sequences clearly shows that the CTG group is divided into two distinct clades: the first one mainly contains haploid



◄ Fig. 5 Phylogeny of yeasts of the CTG group obtained by maximumlikelihood (a) and bayesian (b) analyses, based on the concatenation of the sequence of 12 protein-coding genes. *Numbers* represent bootstrap values from 100 replicates for ML method and the probability obtained with the Bayesian approach. MrBayes was configured for 10,000 iterations, a sample frequency of 10 and a burn-in of 250

Pichia species, whereas the second contains diploid *Candida* species. This observation supports recent phylogenic studies based on whole genome analysis (Fitzpatrick et al. 2006; Butler et al. 2009). Their phylogenetic analyses based on 2,146 putative orthologous gene families provide strong evidence for two subgroups within the CTG group (Fitzpatrick et al. 2006). One contains the fully sexual species *C. lusitaniae*, *P. guillermondii* and *D. hansenii*, whereas the second contains species with at best a cryptic sexual cycle such as *C. albicans*, *C. dubliniensis*, *C. tropicalis*, *C. parapsilosis* and *L. elongisporus*. Interestingly, there is a high congruence of phylogenies based on nuclear and mitochondrial sequences.

Finally, our phylogenetic analysis shows that *P. farinosa* and *P. sorbitophila*, which were known to be unique species, are very distant and represent two distinct species even if large syntenic blocks are conserved. In fact, *P. farinosa* and *P. sorbitophila* should be classified as separate species.

Acknowledgments We thank Thomas Jeffries for providing us the mitochondrial sequence of *Pichia stipitis*. We are very grateful to Irwin Davidson, Bernard Jost, and Stéphanie Legras from the IGBMC Microarray and Sequencing Platform. We also thank Jean-Marc Aury for assistance in computational analysis. This work was supported in part by the Génolevures Consortium GDR 2354. JPP is supported by a grant from the French "Ministère de l'Enseignement Supérieur et de la Recherche". JS was supported by a CNRS PEPS 2009 grant.

References

- Adler A, Hidalgo-Grass C, Boekhout T, Theelen B, Sionov E, Polacheck I (2007) *Pichia farinosa* bloodstream infection in a lymphoma patient. J Clin Microbiol 45:3456–3458
- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402
- Anderson JB, Wickens C, Khan M, Cowen LE, Federspiel N, Jones T, Kohn LM (2001) Infrequent genetic exchange and recombination in the mitochondrial genome of *Candida albicans*. J Bacteriol 183:865–872
- Bakir M, Cerikcioğlu N, Tirtir A, Berrak S, Ozek E, Canpolat C (2004) Pichia anomala fungaemia in immunocompromised children. Mycoses 47:231–235
- Butler G, Rasmussen MD, Lin MF, Santos MAS, Sakthikumar S, Munro CA, Rheinbay E, Grabherr M, Forche A, Reedy J-L, Agrafioti I, Arnaud M-B, Bates S, Brown AJP, Brunke S,
Costanzo MC, Fitzpatrick DA, de Groot PWJ, Harris D, Hoyer LL et al (2009) Evolution of pathogenicity and sexual reproduction in eight *Candida* genomes. Nature 459:657–662

- Chevalier BS, Stoddard BL (2001) Homing endonucleases: structural and functional insight into the catalysts of intron/intein mobility. Nucleic Acids Res 29:3757–3774
- Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J, Guindon S, Lefort V, Lescot M et al (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:465–469
- Dujon B (2010) Yeast evolutionary genomics. Nat Rev Genet 11:512– 524
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113
- Fitzpatrick DA, Logue ME, Stajich JE, Butler G (2006) A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis. BMC Evol Biol 6:99
- Foury F, Roganti T, Lecrenier N, Purnelle B (1998) The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. FEBS Lett 440:325–331
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704
- Haugen P, Simon DM, Bhattacharya D (2005) The natural history of group I introns. Trends Genet 21:111–119
- Jacquier A, Dujon B (1983) The intron of the mitochondrial 21S rRNA gene: distribution in different yeast species and sequence comparison between *Kluyveromyces thermotolerans* and *Saccharomyces cerevisiae*. Mol Gen Genet 192:487–499
- Jeffries TW (2006) Engineering yeasts for xylose metabolism. Curr Opin Biotechnol 17:320–326
- Jeffries TW, Grigoriev IV, Grimwood J, Laplaza JM, Aerts A, Salamov A, Schmutz J, Lindquist E, Dehal P, Shapiro H, Jin Y, Passoth V, Richardson PM (2007) Genome sequence of the lignocellulosebioconverting and xylose-fermenting yeast *Pichia stipitis*. Nat Biotechnol 25:319–326
- Jung PP, Schacherer J, Souciet JL, Potier S, Wincker P, de Montigny J (2009) The complete mitochondrial genome of the yeast *Pichia sorbitophila*. FEMS Yeast Res 9:903–910
- Kerscher S, Durstewitz G, Casaregola S, Gaillardin C, Brandt U (2001) The complete mitochondrial genome of *Yarrowia lipolytica*. Comp Funct Genomics 2:80–90
- Knight RD, Landweber LF, Yarus M (2001) How mitochondria redefine the code. J Mol Evol 53:299–313
- Kosa P, Valach M, Tomaska L, Wolfe KH, Nosek J (2006) Complete DNA sequences of the mitochondrial genomes of the pathogenic yeasts *Candida orthopsilosis* and *Candida metapsilosis*: insight into the evolution of linear DNA genomes from mitochondrial telomere mutants. Nucleic Acids Res 34:2472–2481
- Koszul R, Malpertuy A, Frangeul L, Bouchier C, Wincker P, Thierry A, Duthoy S, Ferris S, Hennequin C, Dujon B (2003) The complete mitochondrial genome sequence of the pathogenic yeast *Candida (Torulopsis) glabrata.* FEBS Lett 534:39–48
- Kurtzman CP, Fell JW (1998) The yeasts, a taxonomic study, 4th edn. Elsevier, Amsterdam

- Kurtzman CP, Robnett CJ (2003) Phylogenetic relationships among yeasts of the 'Saccharomyces complex' determined from multigene sequence analyses. FEMS Yeast Res 3:417–432
- Kurtzman CP, Suzuki M (2010) Phylogenetic analysis of ascomycete yeasts that form coenzyme Q-9 and the proposal of the new genera *Babjeviella*, *Meyerozyma*, *Millerozyma*, *Priceomyces*, and *Scheffersomyces*. Mycoscience 51:2–14
- Lang BF, Laforest M, Burger G (2007) Mitochondrial introns: a critical view. Trends Genet 23:119–125
- Laslett D, Canbäck B (2008) ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24:172–175
- Leandro MJ, Gonçalves P, Spencer-Martins I (2006) Two glucose/ xylose transporter genes from the yeast *Candida* intermedia: first molecular characterization of a yeast xylose-H+ symporter. Biochem J 395:543–549
- Li R, Yu C, Li Y, Lam T, Yiu S, Kristiansen K, Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967
- Lowe TM, Eddy SR (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955–964
- Maresová L, Sychrová H (2003) Physiological characterization of osmotolerant yeast *Pichia sorbitophila* and comparison with a putative synonym *Pichia farinosa*. Folia Microbiol 48:211–217
- Nobre A, Lucas C, Leão C (1999) Transport and utilization of hexoses and pentoses in the halotolerant yeast *Debaryomyces hansenii*. Appl Environ Microbiol 65:3594–3598
- Nosek J, Novotna M, Hlavatovicova Z, Ussery DW, Fajkus J, Tomaska L (2004) Complete DNA sequence of the linear mitochondrial genome of the pathogenic yeast *Candida parapsilosis*. Mol Genet Genomics 272:173–180
- Ohama T, Suzuki T, Mori M, Osawa S, Ueda T, Watanabe K, Nakase T (1993) Non-universal decoding of the leucine codon CUG in several *Candida* species. Nucleic Acids Res 21:4039–4045
- Rodriques de Miranda L, Appel KR, Seyfarth H (1980) *Pichia sorbitophila* sp. nov. Antonie Van Leeuwenhoek 46:157–159
- Sacerdot C, Casaregola S, Lafontaine I, Tekaia F, Dujon B, Ozier-Kalogeropoulos O (2008) Promiscuous DNA in the nuclear genomes of hemiascomycetous yeasts. FEMS Yeast Res 8:846–857
- Santos MA, Tuite MF (1995) The CUG codon is decoded in vivo as serine and not leucine in *Candida albicans*. Nucleic Acids Res 23:1481–1486
- Schäfer B (2005) RNA maturation in mitochondria of *S. pombe*. Gene 354:80–85
- Sekito T, Okamoto K, Kitano H, Yoshida K (1995) The complete mitochondrial DNA sequence of *Hansenula wingei* reveals new characteristics of yeast mitochondria. Curr Genet 28:39–53
- Talla E, Anthouard V, Bouchier C, Frangeul L, Dujon B (2005) The complete mitochondrial genome of the yeast *Kluyveromyces thermotolerans*. FEBS Lett 579:30–40
- Tsui CKM, Daniel H, Robert V, Meyer W (2008) Re-examining the phylogeny of clinically relevant *Candida* species and allied genera based on multigene analyses. FEMS Yeast Res 8:651–659
- Valach M, Tomaska L, Nosek J (2008) Preparation of yeast mitochondrial DNA for direct sequence analysis. Curr Genet 54:105–109





http://mc.manuscriptcentral.com/fems

The Ty1 LTR-retrotransposon population in Saccharomyces cerevisiae genome: dynamics and sequence variations during mobility.

Journal:	FEMS Yeast Research		
Manuscript ID:	FEMSYR-10-10-0156.R1		
Manuscript Type:	Research Paper		
Date Submitted by the Author:	03-Jan-2011		
Complete List of Authors:	Bleykasten-Grosshans, Claudine; Université de Strasbourg/CNRS Jung, Paul; Université de Strasbourg/CNRS Fritsch, Emilie; European Molecular Biology Laboratory (EMBL) Potier, Serge; Université de Strasbourg/CNRS de Montigny, Jacky; Université de Strasbourg/CNRS Souciet, Jean-Luc; Université de Strasbourg/CNRS		
Keywords:	transposable element, reverse transcription, recombination, non- autonomous transposable element		



3
4
5
6
7
8
à
3
10
11
12
13
14
15
16
17
18
19
20
20
∠ I 22
22
23
24
25
26
27
28
29
30
31
22
3Z
33
34
35
36
37
38
39
40
41
12
12
-T-J //
44 15
40
46
47
48
49
50
51
52
53
54
55
55
00
5/
58
59
60

1 The Ty1 LTR-retrotransposon population 2 in Saccharomyces cerevisiae genome: 3 dynamics and sequence variations during mobility. 4 Claudine Bleykasten-Grosshans, Paul P. Jung, Emilie S. Fritsch¹, Serge Potier, Jacky de 5 6 **Montigny and Jean-Luc Souciet** 7 8 Laboratoire de Génétique Moléculaire, Génomique et Microbiologie UMR7156 (Université 9 de Strasbourg and CNRS), 67083 Strasbourg, France 10 ¹Present address: European Molecular Biology Laboratory (EMBL), 69117 Heidelberg, 11 12 Germany. 13 14 Abstract 15 Transposable element (TE) evolution in genomes has mostly been deduced from 16 comparative genome analyses. TEs often account for a large proportion of the eukaryotic 17 nuclear genome (up to 50%, depending on the species). Among the many existing genomic 18 copies, only a small fraction may contribute to the mobility of a TE family. We have 19 identified here, using a genetic screening procedure to trap Tv1 LTR-retrotransposon

insertions in *Saccharomyces cerevisiae*, which among the population of resident Ty1 copies are responsible for Ty1 mobility. Although the newly inserted Ty1 copies resulting from a single round of transposition were found to originate from a limited subset of Ty1 resident copies, they showed a high degree of diversity at the nucleotide level, mainly due to the reverse transcription-mediated recombination. In this process, highly expressed and strikingly non-autonomous mutant Ty1 were found to be the most frequently used resident copies, 27 family.

28 Introduction

Transposable elements (TEs) are hosted by genomes, where their presence and functions contributes to genome variations at both inter and intra-specific levels. Because of their repetitive nature and their mobility (often replicative), TEs induce fast genome reshaping processes, and their regulatory signals also modulate adjacent gene expression (Kidwell & Lisch, 2000). Although TEs are practically ubiquitous in the genomes of Archaea, Bacteria and Eukaryota, there exist some extremely marked differences in their genomic contents, in terms of variety, number of copies, genomic fraction, distribution and the respective proportions of full-length, mutated and fossil copies (Le Rouzic & Capy, 2006; Pritham, 2009). These differences partly result from distinct host defense reactions and TE persistence mechanisms. Recent genome sequencing projects have shown the occurrence of various past TE bursts in eukaryote genomes (for examples, see Cordaux & Batzer, 2009 and Ungerer et al., 2009) and identified the currently active families (Abrusan & Krambeck, 2006 and the references therein) and the subfamily organizations (see for example Du et al., 2010). TE traits may partly govern their expansion, persistence and extinction in genomes as predicted by mathematical models in which the TE families in a genome are taken to be different species interacting in an ecosystem (Pritham, 2009; Venner et al., 2009). In particular, the production of mutated non-autonomous TE copies which are able to transpose by using the machinery provided by autonomous elements may threaten the persistence of the family to which they belong (Le Rouzic & Capy, 2006; Le Rouzic et al., 2007). Therefore, to investigate TE persistence, direct methods of measuring the contribution to mobility of individual TE copies in their natural chromosomic context are required, but to our knowledge, no such approaches have been developed so far. This fundamental issue could be addressed by identifying, among the resident TE copies, the parental copies from which new TE insertions originate, using a model organism that meets the following criteria: (i) means of

> detecting new TE insertions arising from the non-marked resident TE copies and inserting in their natural chromosomic target must be achievable and (ii) a high-quality genomic sequence and annotations must be available, especially for dealing with repetitive TEs. Due to the existence of sequence similarities between resident TE copies, the newly inserted copies will have to be characterized at the nucleotide level, and an additional degree of complexity is expected to arise when studying the inheritance of externally-primed retrotransposons (Beauregard et al., 2008), which are recombined during the reverse-transcription (RT) steps on which transposition depends.

Saccharomyces cerevisiae reference strain S288c is a highly suitable organism for developing such an approach. Long Terminal Repeat (LTR) externally-primed retrotransposons are the only TEs present in the S. cerevisiae genome. They belong to five families called Ty1 to Ty5, showing great variability in the number of copies present in various strains (Wei *et al.*, 2007; Borneman et al., 2008; Liti et al., 2009). In S288c, the most fully sequenced and best studied S. cerevisiae strain, the Ty1 subset has the largest number of copies (32 copies), as well as being the most active (Lesage & Todeschini, 2005) and the best documented subset in terms of individual copy traits (Morillon et al., 2002; Servant et al., 2008). Due to the existence of efficient control pathways, Ty1 transposition shows a very low rate of occurrence and is restricted to the haploid state. The 'URA2' genetic screening method (Roelants et al., 1995) was developed to detect large spontaneous chromosomal rearrangements, including Ty1 insertions. This screening procedure depends on the URA2 gene involved in pyrimidine biosynthesis, which encodes a multifunctional enzyme with carbamoylphosphate synthetase (CPSase) and aspartyltranscarbamylase (ATCase) activities. In S. cerevisiae, URA2, CPA1 and CPA2 may supply the CPSase activities, whereas the terminal part of URA2 is responsible for the single ATCase activity. The ATCase-defective strains used here showed point mutations located upstream of the ATCase-coding domain, which cause premature translation

FEMS Yeast Research

arrest (Fig. 1). Chromosomal rearrangements can restore ATCase expression, resulting in
Ura⁺ revertants. In a FL100 background, this positive screening has been found to act as a real
transposon-trap, since up to 70% of the rearrangements selected consisted of de novo
integrase-mediated Ty1 insertions (Roelant *et al.*, 1995; Tourrette *et al.*, 2007).

The *URA2* screening system was adapted here to the S288c strain, in order to enable the identification of the parental elements resulting in new Ty1 insertions. Newly inserted Ty1 copies will be referred to below as NETs, which stands for <u>NEw Ty1</u>, and the pre-existing Ty1 copies will be referred to as RETs, which stands for <u>REsident-Ty1</u>. This approach was used to address two fundamental aspects of the Ty1 dynamics: the contribution of individual resident copies to Ty1 mobility and the diversity generated in their progeny.

89 Materials and methods

90 Yeast strains, growth and media

Yeast cells were grown at 30°C in liquid or solid (2% agar) yeast peptone dextrose (YPD) or supplemented yeast nitrogen base (YNB). The strain FY67 (MATa trp1/163) (Winston et al., 1995) was used to obtain an ATCase defective strain with the S288c background by a one-step replacement using a PCR fragment amplified from the strain FL100 ura2 15-30-72 (Schacherer *et al.*, 2004). The resulting T21 strain carries the *ura2-s* allele with two non-sense point mutations at positions 262 and 916, with respect to the URA2 coding sequence (Fig. 1). The reference strain $ura2::KANMX4 \ leu2\Delta0$ used for cross-analyses was obtained from the Euroscarf strain BY4741.

100 Isolation and analysis of Ura⁺ revertants

 In order to isolate spontaneous independent Ura⁺ revertants, a previously described standardized procedure was used (Fritsch *et al.*, 2009). The mutation rate

(mutations/cell/selection) was determined using a maximum-likelihood method (Lea & Coulson, 1949). The 95% confidence limits were calculated using Student's t-test. To identify the types of rearrangement resulting in ATCase reactivation, molecular analyses were performed as previously described (Schacherer et al., 2004; Fritsch et al., 2009). Ty1 insertions detected in Southern blot hybridizations were confirmed by PCR amplification using Taq DNA polymerase from MP Biomedicals as described by the manufacturer. The primers used for PCR amplifications and sequencing were chosen on the basis of the genomic sequence of S288c (SGD: Saccharomyces Genome Database, http://www.yeastgenome.org): a set of URA2-specific primers was used as well as a set of Ty1-specific primers designed to hybridize with any of the complete Ty1 elements in S288c (listed in Table S1). Extremities of Ty1 elements were amplified using one specific primer in the corresponding genomic flanking region and the second in the coding region of the Ty1 element; to amplify whole Ty1 elements, specific primers were chosen in their flanking regions and amplifications were performed with iProof High Fidelity DNA Polymerase from Biorad, according to the manufacturer's protocol. For sequencing purposes, PCR products were purified using MicroSpin S400 (GE Healthcare). DNA sequencing was performed with AmpliTag FS DNA polymerase and BIGDYE TM terminators and sequence reactions were analyzed with an Applied Biosystems 373XL sequencer.

6 121

122 Tyl and Ty2 data

Sequences of the 31 resident Ty1 of S288c (RETs) were downloaded at
ftp://ftp.yeastgenome.org/yeast/data_download/sequence/genomic_sequence/other_features,
using the 06/06/2008 version. An additional RET (YLR035C-A) was assembled from
chromosome XII coordinates 215081 to 221006 (Lesage & Todeschini, 2005). To simplify

127 the presentation, a number ranging from 1 to 32 was allotted to each parental RET, depending

FEMS Yeast Research

on its genomic localization. The correspondence with the SGD names is indicated in Table 1,
as well as the correspondence with MIPS CYGD names (http://mips.helmholtzmuenchen.de/genre/proj/yeast/) that were used in Morillon *et al.* (2002). As some RETs are
Ty1/2 hybrids (Jordan & McDonald, 1998), Ty2 data were also downloaded. Nucleotide
positions in the Ty1 elements are given relative to the reference TyH3 sequence (Boeke *et al.*,
1988; accession number M18706).

125 5

135 Sequence alignments

Sequence alignments were performed using the Blast (Altschul *et al.*, 1990) and ClustalW2
(Larkin *et al.*, 2007) programs with default parameters.

139 RNA extraction and RT-PCR assays

Yeast strains were grown in liquid YNB media supplemented with tryptophan at 30°C. Cells in the mid-log phase were harvested and 10 ml of each culture were frozen using liquid nitrogen. RNAs were extracted using a standard hot acid phenol method combined with TRIZOL reagent (Invitrogen), followed by a Qiagen RNeasy Mini Kit step according to the manufacturers protocols. Samples were treated with DNAseI from Invitrogen and the absence of DNA was monitored by performing PCR reactions. SuperScriptIII Reverse Transcriptase from Invitrogen was used to carry out RT-PCR reactions with 100 ng RNA in a final reaction volume of 25 µl.

- 3 149 **Results**

150 Characteristics of spontaneous Ura⁺ revertants

151 Using the standardized screening procedure, 79 independent Ura^+ revertants were isolated 60 152 from the T21 strain carrying the *ura2-s* allele. The mutation rate determined using a

maximum-likelihood method was 4.07×10^{-10} mutations/cell/selection (confidence interval CI: 3.40x10⁻¹⁰ – 4.73x10⁻¹⁰). Molecular analyses were performed on all 79 revertants to identify the rearrangements resulting in ATCase reactivation, as described in Schacherer *et al.*, 2004 and Fritsch *et al.*, 2009. In 62 of the revertants, the ATCase-reactivating rearrangement corresponds to: deletion events (27 revertants), translocation events (2 revertants) and duplication events (33 revertants). All these categories have been previously described in the FL100 background (Welcker *et al.*, 2000; Schacherer *et al.*, 2004; Tourrette *et al.*, 2007).

The 17 remaining revertants showed a Ty1 copy inserted upstream of the ATCase-coding domain (NETs named A to Q). The insertion sites and orientation of NETs are indicated in Fig. 1 and in Table S2: in four revertants, the NET was inserted in the same transcriptional orientation as the URA2 gene (the 'sense' orientation), whereas the opposite orientation was observed in the other 13 revertants (the 'antisense' orientation). Sense and antisense NETs all corresponded to canonical integrase-mediated insertions, as confirmed by the presence of the typical five-bp target-site-duplication signature (TSD) (Table S2). After crossing each of the 17 revertants with a *ura2::KANMX4 leu2\Delta 0* strain, the subsequent tetrad analyses showed that the Ura⁺ phenotype cosegregates with the Ty1 insertion, which means that the functional reactivation of ATCase requires no additional events, whatever the orientation of the inserted NET. Reactivation of ATCase as the result of antisense Tyl insertion was previously described, but no sense NET insertions were obtained in a larger collection of Ura⁺ revertants with a different FL100 background (60 independent Ty1 insertions: Roelants et al., 1995 and A. J. Welcker, unpublished data). In order to further investigate the effects of sense NETs on adjacent downstream sequences, RT-PCR assays were performed, using a unique reverse primer anchored to the ATCase-coding domain and various Ty1-specific forward primers that hybridize at the sense NETs 3' extremity (Table S1). Using forward primers designed to span the Ty1-ura2 junction, a specific fragment was amplified (Fig. 2, a): in each revertant, the

fragment size was correlated with the distance between the NET insertion point and the ATCase-coding domain. Similar successful amplifications were obtained with the primer LTRs2 complementary to 3' LTR positions 298 to 318 (Fig. 2, b), but no amplification was obtained when a forward primer complementary to the *TYB* end was used (Fig. 2, c). These results indicate that *URA2*-related transcripts are initiated in the 3' extremity of sense NETs which can act as a promoter area on adjacent sequences, like previously established in the case of the 5' extremity of Ty1 elements (Roelants *et al.*, 1997; Servant *et al.*, 2008).

186 Origin of NETs: identifying their parents among RETs

The nucleotide sequences of the NET extremity located near the ATCase-coding domain are similar but not identical: there are 15 variant sequences among the 17 NETs (see the sequence alignments available in Files S1 and S2). This suggested that ATCase reactivation does not require a specific RET, and made it relevant to identify which of the S288c RETs were the parents of the NETs selected. Two areas in each of the 17 NETs were systematically sequenced: (i) the 5' LTR and the first 350 bp of the TYA sequence and (ii) the 600 bp long terminal part of TYB plus the complete 3' LTR. The mobility of Ty elements results from a discontinuous RT process requiring two obligatory strand-transfers to produce the complete (U3-R-U5) LTR-flanked cDNA from the Ty1 RNA with unique subterminal 5'(U5) and 3'(U3) segments and repetitive terminal (R) regions. This process occurs in virus-like particles (VLP) containing two Ty1 RNA molecules, which are possibly transcribed from distinct parental copies and may thus result in mosaic cDNA molecules (Gabriel & Mules, 1999). Therefore, we have performed Blast alignments between the corresponding NET and RET segments: the inheritance of (U3) and (R-U5) segments is particularly elaborate, (U3) segments from NETs being inherited from the 3' RET LTR and (R-U5) segments from the 5' RET LTR. Due to local identities between RET sequences, alignments mostly identified

several RETs for each NET segment. We therefore performed a 'reconstitution' of the NET RT process, in order to discard any identified RETs that were not suitable for obtaining combinations of two parents at most. An example of the detailed procedure for parental RET identification is described in Fig. S1. We were able to unambiguously deduce the parental origins of six NETs (Fig. 3 and Table S2: NETs A to F). In the case of nine other NETs (NETs G to O), one of the two parents was clearly identified. Their second parent could not be identified among several possible RETs, due to either (i) the existence of local sequence identities between RETs (NET-J, for example) or (ii) because one RET matched all the sequenced NET segments, which suggested the minimum but non-exclusive hypothesis that they were generated within homozygous VLPs (NET-L, for example). (iii) The case of NETs G and H will be discussed below. The two remaining NETs P and Q may have originated from several possible combinations of RETs. Sequencing of additional NET segments (not shown) did not improve the resolution of RET identification and finally 21 out of the whole set of the 17 NET parents (34 parental RETs) were unambiguously identified.

It should be pointed out here that Ty1 and Ty2 elements correspond to two different but closely related families and some of the 32 S288c existing RETs have been characterized as Ty1/2 hybrids (Jordan & McDonald, 1998). However, the results obtained using the reconstitution strategy pointed to the conclusion that none of the S288c Ty2 elements were involved in the formation of the NETs selected here.

8 222

223 RET usage in NETs

The 21 NET parents identified were generated by only nine RETs out of all those in the
S288c 32 RET reservoir. Two subgroups of RETs were previously defined in expression
studies: a highly expressed RET subgroup responsible for 75% of the total RET transcription,
and a weakly expressed set (Morillon *et al.*, 2002 and Table 1). The present data indicate that

four of the 21 NET parents belong to the weakly expressed subgroup (RET-1, 23, 26 and 32), with each of them generating only one NET. The other 17 parents were found corresponding to five RETs from the highly expressed subgroup (RET-6, 8, 11, 20 and 30) with almost each of them generating several NETs (Table 1 and Fig. 3). Interestingly, this pattern of distribution between highly and weakly expressed RETs correlates well with the overall rates of expression of both RET subpopulations (Table S3), suggesting that, in our assay, RET mobility may correlate with RET expression. Consistently, the genomic RNA all four weakly expressed RETs generating NETs was "co-packaged" with one from the highly expressed RETs.

Among the present NET parents, two of the RETs were particularly noteworthy, namely RET-20 and RET-6, which have a progeny of eight and four NETs, respectively. When calculated on the basis of the individual expression rates established by Morillon *et al.* (2002), the rate of occurrence of RET-20 was significantly higher than expected, whereas those of all the other RETs were consistent with the individual rates of expression (Table S4). This suggests that either the expression rate of RET-20 is even higher in our assay or that mobility of some RETs may not only depend on its transcription level. For example, RET-20 may have a higher intrinsic level of mobility and/or it may escape repressive processes reducing Ty1 mobility. Remarkably, the nucleotide sequence of RET-20 shows that it belongs to a subclass of mutated RETs (Fig. 3) which also includes RET-1 and RET-6: these are non-autonomous Ty1 elements encoding intact TyA protein but non-functional TyA-B polyprotein. These non-autonomous TEs have been described as hyperparasites in cis that use proteins encoded by other genomic TEs belonging to the same family to propagate (Sabot & Schulman, 2006). Additional sequencing was performed on the progeny of these non-autonomous RETs (12 NETs out of 17), and the results obtained showed that in the case of five NETs, the RT

252 process yielded the original parental mutation: three NETs carry the 428 bp deletion typical of

RET-6, and two other NETs contain the two-nucleotide addition in *TYB* arising from RET-20 (Fig. 3). By contrast, recombination during the RT process restored the corresponding autonomous Ty1 sequence in the seven other NETs, including the particular case of potentially autonomous NET-E, which has two mutated parents: RET-6 and RET-1. We have however confirmed by sequencing that the specific mutation present in RET-1 pre-existed at the corresponding genomic locus.

260 NET variability resulting from recombination during the RT process

As mentioned before, 15 polymorphic sequences were detected in the set of the 17 new Ty1 copies generated in this study, which provided an opportunity of investigating the variability arising during a single Ty1 replication cycle. The variability of NET segments was found to range from 0.6 to 13.6% (from pair-wise global alignments) which is similar to the variation of the corresponding segments in all the RET population but most importantly, 12 NETs turned out to be new variant Ty1 elements that differ from all the preexisting RETs. All these 12 NETs are chimeras involving two distinct parental RETs. These data show that the recombination into heterozygous particles occurring during the RT process is the main source of variability. In eight of the chimeric NETs, the 3'(R-U5) and 3'(U3) were inherited from distinct RETs, indicating that the mode of minus-strong-stop-DNA transfer was inter-molecular (Table S2 and Fig. 3). In the case of NETs G and H, where 3'(R-U5) and 3'(U3) were inherited from the same RET, intra-molecular transfer obviously occurred, but NETs A and E both had parents with an identical 5' (R-U5), and either inter- or intra-molecular transfer may therefore have been involved. During cDNA elongation, template switch(es) have occurred between the two encapsidated RNA molecules, resulting either in 10 NETs in which the TYA segment, central segment or TYB segment have been inherited from different

277 RETs or in NETs in which the 5' and 3'(R-U5) segments are inherited from different parents
278 (two NETs: F and G) (Table S2 and Fig. 3).

To investigate the *in vivo* recombination processes occurring along the entire Ty1 genome at the nucleotide level, and map the regions where switches occurred, full-length sequencing was performed on NETs A, B and C: whole sequences are available in File S1. These NETs were chosen because they have clearly identified parents with divergent TYA-B segments. Template switches were detected using multiple alignments and mapped in Fig. 3. The frequency of the template switches ranged from one to six events per NET. It is worth noting that these differences are not attributable to local identities between parent sequences, which would prevent the detection of switches: NET-A (six switches detected) has less divergent parents (96.8 % identities) than NET-C (two switches detected), whose parents showed 93% identity along the whole TYA-B segment. In the case of NET-B (one switch detected), the parental RETs showed 97% identity along the whole TYA-B segment.

The number of switches detected in NETs (19 switches for 12 chimerical NETs corresponding to roughly 4 x 10^4 sequenced bp) from our assay suggests that Ty1 may have evolved a recombination rate in the same range as human immunodeficiency virus (HIV, recombination rate estimated to 3 x 10^{-4} crossovers per bp per replication round in Jetzt *et al.*, 2000), which was described as one of the most recombinogenic of all the retroviruses (Wain-Hobson *et al.*, 2003).

⁸ 296

297 NET variability resulting from mutations during the RT process

An additional source of variability is RT infidelity, which gives rise to mutations. Toward the SGD reference sequence, we have detected three NETs that may have undergone an RTinduced mutation (Table S2 and Fig. 3). One substitution may have occurred in NET-D at *TYA* position 567, where an ACC threonine codon was replaced by a non-synonymous CCC

proline codon: as RET-30 has provided the corresponding segment only once in NETs, we have confirmed by sequencing that the corresponding chromosomal locus is nevertheless identical to the published S288c sequence. Two other substitutions may have occurred in NETs C and K, in the 5' U3 positions 210 and 333, respectively, because these positions are not symmetrical in the 3' U3. These mutations were detected in a total number of 40154 bp sequenced in the 17 NETs, which are equivalent to 115566 bp replicated by the RT process (Gabriel *et al.*, 1996). The resulting mutation rate of 2.59×10^{-5} per bp per replication cycle is in the same range as previously reported when using a Ty1 element induced from a plasmid $(2.54 \times 10^{-5} \text{ per bp per replication cycle, in Gabriel$ *et al.*, 1996). Significantly our own estimateof RT infidelity was measured in a natural Ty1 population.

Interestingly, some divergences were detected that may not result from mutations during RT but from preexisting divergences in RETs. Actually, NETs G and H both show 5'(R-U5) segments with identical sequences, but which do not match with any of the S288c Ty elements and differ from all the known RET sequences by 2 to 9 sites. It is therefore possible that one RET which is missing from the reference sequence, might be involved in the generation of NETs G and H. This RET was referred to as 'uk', which stands for 'unknown' (Fig. 3). Consistently, several additional RETs were detected in the S288c genome (Gabriel et al., 2006; Wheelan et al., 2006; Shibata et al., 2009). In addition, we identified the 'uk' sequence at a single locus on S288c chromosome III in a non-annotated solo-LTR (coord. 169203 to 169540) next to the annotated solo-LTR YCRWdelta11 (Fig. 4a). In this region that has been described as a hot-spot of Ty insertions in chromosome III (Far Right-Arm transposition Hot-spot, Wicksteed et al., 1994), we have detected three unannotated Ty1 and Ty3 solo-LTRs in the SGD sequence. Additional PCR results confirmed that, like in Gabriel et al. (2006), two full-length Ty1 elements are in fact present at this locus (Fig. 4b). The present findings therefore support the suggestion put forward by Shibata et al. (2009) that a

FEMS Yeast Research

few revisions of the S288c genome annotations and sequence may be required to describe the Ty-related sequence landscape more accurately.

Discussion

In order to study the dynamics of Ty1 mobility at the nucleotide level, two non-sense mutations were introduced into the S288c background at the URA2 locus in order to directly select in an almost native chromosomic locus, newly inserted Ty1 copies arising from the natural Ty1 population. The coding sequence between the second non-sense mutation and the beginning of the ATCase-coding domain traps integrase-mediated Ty1 insertions. We characterized 17 newly inserted Ty1 elements showing polymorphic nucleotide sequences, which are able to reactivate ATCase expression by acting as a promoter. Four NETs inserted in the same transcriptional direction as URA2 (sense NETs) and 13 NETs with the opposite orientation (antisense NETs) were selected. The recovery of gene transcription by sense Ty1 is rather unusual, whereas antisense Ty1 were previously reported to mediate gene transcription on URA2 in a different genetic background (FL100), as well as on various reporter genes (see for examples Boeke et al., 1986 and Servant et al., 2008). It has been suggested that the TyA-coding sequence, along with the 5' LTR, may contribute to the transcription process controlled by antisense Ty1 (Servant *et al.*, 2008), but the finding that sense NETs may also play a similar role, based on the present RT-PCR data, suggests the existence of alternative promoting pathway(s) involving regulatory motifs located in the Ty1 3' extremity.

The sequence variability of the 17 NETs produced in this study suggested no specific RET was selected to reactivate ATCase and the relevance to assess the relative contribution of each individual genomic element in an LTR retrotransposon family. The findings obtained here show that to generate the NETs isolated in our assay, only a few of the S288c RETs available

contribute importantly to Ty1 mobility. In particular, two 'Hot Ty1s' (RET-6 and RET-20) were found to be involved in the formation of 70% of the new copies, by analogy with the 'Hot L1' non-LTR retrotransposons described in the human genome (Brouha et al., 2003). The NETs do, however, gain considerable nucleotidic variability during the RT process. The occurrence of in vivo RT infidelity and RT-mediated inter-element recombination was previously observed (Xu & Boeke, 1987; Boeke et al., 1988), and further investigations were carried out using a single Ty1 sequence in a multicopy plasmid as the main source of Ty1 RNA and proteins (Gabriel et al., 1996; Wilhelm et al., 1999). Here we assessed the occurrence of these processes in a native Ty1 population, where cDNA may be generated in mosaic VLPs assembled from variant TyA/TyA-B isoforms expressed by distinct RETs. Our results suggest that Ty1 may have evolved efficient recombination, thanks to (i) inter-molecular minus-DNA-strong-stop transfer, which seem to be more prevalent than previously reported (Wilhelm et al., 1999) (in eight NETs out of 10 chimerical NETs for which the mode of transfer was identified), and (ii) internal template switches with similar rates of occurrence to those observed in highly recombinogenic RT-dependant viruses. Even insertions consisting in non-functional/mutated NETs (NETs A, F, L, M, N) were detectable in our experiments. In this respect, the system used here was found to provide a particularly useful means of analyzing natural RT mediated variability, by comparison to equivalent approaches on RT-dependent viral isolates from which only recombinant genomes that remain infective can be isolated.

We also established here that mutated non-autonomous RETs are not only able to transpose thanks to inter-element complementation (Curcio & Garfinkel, 1994), but they also generate NETs efficiently: both 'Hot Ty1' RET-6 and 20 are non-autonomous. Whether and how nonautonomous RETs are privileged for mobility in *S. cerevisiae* and if an additional regulation based on non-autonomous copies takes place in order to minimize the impact of Ty1 on its

host are exciting questions that still remain to be answered. It is particularly interesting that RNAs originating from RET-20 or RET-6 have translation defects: they show a +2 frameshift or a deletion, which may result in premature termination of the translation process (in addition to the Ty1 programmed +1 frameshift between the TYA and TYB genes). Such defects are likely to affect their RNA dynamics through interactions with components of the non-sense-mediated-decay pathway involved in RNA storage/degradation (Stalder & Muhlemann, 2008) and contributing to Ty1 transposition (Berretta et al., 2008; Checkley et al., 2010; Dutko et al., 2010) and to Ty3 assembling (Beliakova-Bethell et al., 2006). The fact that non-autonomous Ty1 are highly mobile also addresses the issues of how the selection of the genomic RNA packaged into VLP operates and how sufficient functional TyA-B can be incorporated in such VLP. In the case of HIV-1 and HIV-2, proteins translated from their own RNA are preferentially packaged (Anderson & Lever, 2006 and references therein) but the presence of NETs arising from VLP carrying two defective RNAs (NET-E and potentially homozygous NET-L, M and N) suggests that different rules may govern these processes in the case of Ty1 retrotransposon. The fact that the Ty1 genomic RNA contains no intron could be one of the bases of such differences.

The high success of non-autonomous TE may affect the persistence of their family if competition with autonomous members reaches a critical level (Le Rouzic et al., 2007). It is thus of particular interest that the intense recombination we described above results in the progeny of the prolific non-autonomous RETs recovering potentially autonomous Ty1 elements (7 out of the 12 NETs that were generated by non-autonomous RETs). These dynamics may result in a particularly delicate balance between the birth and death of Ty1 elements, as previously suggested (Le Rouzic, et al., 2007). Thus, the data obtained here again address the issue of a potential importance of non-autonomous Ty1 elements in the persistence of the Ty1 family. This question is even more interesting when considering that

the S. cerevisiae genome seems to be highly restrictive to TE persistence: contrary to what occurs in basidiomycetes (see Martin et al., 2010, for an example), in distant ascomycetes and in multicellular eukaryotes, TE communities may have undergone massive extinction in Saccharomycetaceae species (Génolevures-consortium et al., 2009; Dujon, 2010). Moreover, its lifestyle prevents the occurrence of further TE reinvasions through limited out-crossing (Knop, 2006). Therefore, the Ty1 family is particularly noteworthy and the present study provides an important starting point to further investigate its dynamics and the reasons of its persistence across S. cerevisiae strains.

Acknowledgements

We thank L. Despons for bioinformatic tools to automate Blast analyses and J. Gagneur for his help with the statistical analyses. We would like to thank M. Zeniou-Meyer and N. Jauniaux for technical assistance.

This research was partly supported by the ANR-05-BLAN-0331-03 (GENARISE) and the

CNRS GDR 2354 Génolevures-consortium. PPJ and ESF were supported by grants from the

French Ministry for Higher Education and Research.

Higher Eur.

417 **References**

1 2 3

- 418 [1] Abrusan G & Krambeck HJ (2006) Competition may determine the diversity of transposable elements. *Theor Popul Biol* **70**: 364-375.
- 420 [2] Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ (1990) Basic local alignment
 421 search tool. *J Mol Biol* 215: 403-410.
- 422 [3] Anderson EC & Lever AML (2006) Human immunodeficiency virus type 1 Gag polyprotein modulates its own translation. *J Virol* 80: 10478–10486.
- 424 [4] Beauregard A, Curcio MJ & Belfort M (2008) The take and give between
 425 retrotransposable elements and their hosts. *Annu Rev Genet* 42: 587-617.
- 426 [5] Beliakova-Bethell N, Beckham C, Giddings TH, Jr., Winey M, Parker R & Sandmeyer S
 427 (2006) Virus-like particles of the Ty3 retrotransposon assemble in association with P-body
 428 components. *RNA* 12: 94-101.
- 429 [6] Berretta J, Pinskaya M & Morillon A (2008) A cryptic unstable transcript mediates
 430 transcriptional trans-silencing of the Ty1 retrotransposon in *S. cerevisiae. Genes Dev* 22: 615431 626.
- 432 [7] Boeke JD, Styles CA & Fink GR (1986) Saccharomyces cerevisiae SPT3 gene is required
- 433
 433
 434
 434
 435
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
 434
- 435 [8] Boeke JD, Eichinger D, Castrillon D & Fink GR (1988) The Saccharomyces cerevisiae
 436 genome contains functional and nonfunctional copies of transposon Ty1. Mol Cell Biol 8:
 437 1432-1442.
- 438 [9] Borneman AR, Forgan AH, Pretorius IS & Chambers PJ (2008) Comparative genome analysis of a *Saccharomyces cerevisiae* wine strain. *FEMS Yeast Res* 8: 1185-1195.
- 440 [10] Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV & Kazazian HH,
- 33 441 Jr. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc*34 442 *Natl Acad Sci U S A* 100: 5280-5285.
- 443 [11] Checkley MA, Nagashima K, Lockett SJ, Nyswaner KM & Garfinkel DJ (2010) P-body
 444 components are required for Ty1 retrotransposition during assembly of retrotransposition445 competent virus-like particles. *Mol Cell Biol* **30**: 382-398.
- 446 [12] Cordaux R & Batzer MA (2009) The impact of retrotransposons on human genome
 40 447 evolution. *Nat Rev Genet* 10: 691-703.
- 45 45 45 450 [14] Du J, Tian Z, Bowen NJ, Schmutz J, Shoemaker RC & Ma J (2010) Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR
 46 452 Swapping in soybean. *Plant Cell* 22: 48-61.
- 47 453 [15] Dujon B (2010) Yeast evolutionary genomics. *Nat Rev Genet* **11**: 512-24.
- 48 454 [16] Dutko JA, Kenny AE, Gamache ER & Curcio MJ (2010) 5' to 3' mRNA decay factors colocalize with Ty1 Gag and human APOBEC3G and promote Ty1 retrotransposition. *J Virol* 84.
- 457 [17] Fritsch ES, Schacherer J, Bleykasten-Grosshans C, Souciet JL, Potier S & de Montigny J
 458 (2009) Influence of genetic background on the occurrence of chromosomal rearrangements in
 54 459 Saccharomyces cerevisiae. BMC Genomics 10: 99.
- 460 [18] Gabriel A & Mules EH (1999) Fidelity of retrotransposon replication. Ann N Y Acad Sci
 461 870: 108-118.
- 462 [19] Gabriel A, Willems M, Mules EH & Boeke JD (1996) Replication infidelity during a
 59 463 single cycle of Ty1 retrotransposition. *Proc Natl Acad Sci U S A* 93: 7767-7771.
- 464 [20] Gabriel A, Dapprich J, Kunkel M, Gresham D, Pratt SC & Dunham MJ (2006) Global
 465 mapping of transposon location. *PLoS Genet* 2: e212.

1 2		
3	466	[21] Génolevures-consortium, Souciet JL, Dujon B. et al. (2009) Comparative genomics of
4	467	protoploid Saccharomycetaceae. <i>Genome Res</i> 19 : 1696-1709.
5	468	[22] Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD & Dougherty JP (2000) High rate of
7	469	recombination throughout the human immunodeficiency virus type 1 genome. J Virol
8	470	74 :1234-1240.
9	471	[23] Jordan IK & McDonald JF (1998) Evidence for the role of recombination in the
10	472	regulatory evolution of Saccharomyces cerevisiae Ty elements. J Mol Evol 47: 14-20.
11 12	473	[24] Kidwell MG & Lisch DR (2000) Transposable elements and host genome evolution.
13	474	Trends Ecol Evol 15: 95-99.
14	475	[25] Knop M (2006) Evolution of the hemiascomycete yeasts: on life styles and the
15	476	importance of inbreeding. <i>Bioessays</i> 28: 696-708.
16	477	[26] Larkin MA, Blackshields G, Brown NP, et al. (2007) Clustal W and Clustal X version
17 10	478	2.0. Bioinformatics 23: 2947-2948.
19	479	[27] Le Rouzic A & Capy P (2006) Population genetics models of competition between
20	480	transposable element subfamilies. Genetics 174: 785-793.
21	481	[28] Le Rouzic A, Boutin TS & Capy P (2007) Long-term evolution of transposable elements.
22	482	<i>Proc Natl Acad Sci U S A</i> 104 : 19375-19380.
23	483	[29] Lea DE & Coulson CA (1949) The distribution of numbers of mutants in bacterial
24	484	populations. <i>J Genet</i> 49 : 264-285.
26	485	[30] Lesage P & Todeschini AL (2005) Happy together: the life and times of Ty
27	486	retrotransposons and their hosts. Cytogenet Genome Res 110: 70-90.
28	487	[31] Liti G, Carter DM, Moses AM et al. (2009) Population genomics of domestic and wild
29	488	yeasts. Nature 458 : 337-341.
31	489	[32] Martin F, Kohler A, Murat C, et al. (2010) Perigord black truffle genome uncovers
32	490	evolutionary origins and mechanisms of symbiosis. <i>Nature</i> 464 : 1033-1038.
33	491	[33] Morillon A, Benard L, Springer M & Lesage P (2002) Differential effects of chromatin
34	492	and Gcn4 on the 50-fold range of expression among individual yeast Ty1 retrotransposons.
35	493	<i>Mol Cell Biol</i> 22 : 2078-2088.
30	494	[34] Pritham EJ (2009) Transposable elements and factors influencing their success in
38	495	eukaryotes. J Hered 100: 648-655.
39	496	[35] Roelants F, Potier S, Souciet JL & de Montigny J (1995) Reactivation of the ATCase
40	497	domain of the URA2 gene complex: a positive selection method for Ty insertions and
41 42	498	chromosomal rearrangements in Saccharomyces cerevisiae. Mol Gen Genet 246: 767-773.
4∠ ∡2	499	[36] Roelants F, Potier S, Souciet JL & de Montigny J (1997) Delta sequence of Ty1

- 43
 43
 44
 500
 457
 46
 479
 479
 479
 479
 479
 479
 479
 470
 470
 470
 470
 470
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 471
 <
- 46 502 [37] Sabot F & Schulman AH (2006) Parasitism and the retrotransposon life cycle in plants: a
 47 503 hitchhiker's guide to the genome. *Heredity* 97: 381-388.
- 504 [38] Schacherer J, Tourrette Y, Souciet JL, Potier S & De Montigny J (2004) Recovery of a
 505 function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. *Genome*506 *Res* 14: 1291-1297.
- 51 507 [39] Servant G, Pennetier C & Lesage P (2008) Remodeling yeast gene transcription by
 53 508 activating the Ty1 long terminal repeat retrotransposon under severe adenine deficiency. *Mol* 54 509 *Cell Biol* 28: 5543-5554.
- ⁵⁵ 510 [40] Shibata Y, Malhotra A, Bekiranov S & Dutta A (2009) Yeast genome analysis identifies
- 56 511 chromosomal translocation, gene conversion events and several sites of Ty element insertion.
 58 512 Nucleic Acids Res 37: 6454-6465.
- 59 513 [41] Stalder L & Muhlemann O (2008) The meaning of nonsense. *Trends Cell Biol* **18**: 315-
- 60 514 321.

 ${}^{3}_{F}$ 515 [42] Tourrette Y, Schacherer J, Fritsch E, Potier S, Souciet JL & de Montigny J (2007) 516 Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence

5 517 of ploidy. *Mol Microbiol* **64**: 382-395.

12

- 518 [43] Ungerer MC, Strakosh SC & Stimpson KM (2009) Proliferation of Ty3/gypsy-like
 519 retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biol* 7: 40.
- 520 [44] Venner S, Feschotte C & Biemont C (2009) Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* 25: 317-323.
- 523 of human and simian immunodeficiency virus sequence sets reveals massive recombination
 14 524 resulting in shorter pathways. *J Gen Virol* 84: 885-895.
- 15 525 [46] Wei W, McCusker JH, Hyman RW, *et al.* (2007) Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc Natl Acad Sci U S A* 104: 12825-12830.
- 528 [47] Welcker AJ, de Montigny J, Potier S & Souciet JL (2000) Involvement of very short 529 DNA tandem repeats and the influence of the *RAD52* gene on the occurrence of deletions in
- 21 530 Saccharomyces cerevisiae. Genetics **156**: 549-557.
- 531 [48] Wheelan SJ, Scheifele LZ, Martinez-Murillo F, Irizarry RA & Boeke JD (2006)
 532 Transposon insertion site profiling chip (TIP-chip). *Proc Natl Acad Sci U S A* 103: 17632533 17637.
- 25 535 17057.
 26 534 [49] Wicksteed BL, Collins I, Dershowitz A, et al. (1994) A physical comparison of chromosome III in six strains of *Saccharomyces cerevisiae*. *Yeast* 10: 39-57.
- 536 [50] Wilhelm M, Boutabout M, Heyman T & Wilhelm FX (1999) Reverse transcription of the
 537 yeast Ty1 retrotransposon: the mode of first strand transfer is either intermolecular or
 538 intramolecular. *J Mol Biol* 288: 505-510.
- 539 [51] Winston F, Dollard C & Ricupero-Hovasse SL (1995) Construction of a set of 33 540 convenient *Saccharomyces cerevisiae* strains that are isogenic to S288C. *Yeast* 11: 53-55.
- 34 541 [52] Xu H & Boeke JD (1987) High-frequency deletion between homologous sequences
- ³⁵ 542 during retrotransposition of Ty elements in Saccharomyces cerevisiae. Proc Natl Acad Sci U
 ³⁶ 543 SA 84: 8553-8557.



Table 1. Correspondence between the MIPS CYGD, the SGD and the RET nomenclatures for

S288c Ty1 elements and the usage of RETs as NET parents

MIPS name	SGD name	n°	Occurrence ^a
A(*)	YARCTv1-1	RET-1	1
BL	YBLWTv1-1	RET-2	Ĩ
BR	YBRWTy1-2	RET-3	
DR1	YDRCTy1-1	RET-4	
DR3	YDRCTy1-2	RET-5	
DR4	YDRCTv1-3	RET-6 ^{h}	4
DR5	YDRWTv1-4	RET-7	
DR6	YDRWTv1-5	RET-8 ^h	2
ER1	YERCTv1-1	RET-9 ^{<i>h</i>}	
ER2	YERCTv1-2	RET-10	
GR1	YGRWTv1-1	RET-11 ^h	2
GR2	YGRCTv1-2	RET-12	_
GR3	YGRCTv1-3	RET-13	
H(*)	YHRCTv1-1	RET-14	
JR1	YJRWTv1-1	RET-15	
JR2	YJRWTv1-2	RET-16 ^{<i>h</i>}	
LR1	YLR035-CA	RET-17	
LR2	YLRCTv1-1	RET-18	
LR3	YLRWTv1-2	RET-19 ^{<i>h</i>}	
LR4	YLRWTv1-3	RET-20 ^{<i>h</i>}	8
ML1	YMLWTv1-1	RET-21	
ML2	YMLWTv1-2	RET-22 ^{h}	
MR1	YMRCTy1-3	RET-23	
MR2	YMRCTv1-4	RET-24	
NL1	YNLCTy1-1	RET-25	
NL2	YNLWTy1-2	RET-26	1
OL	YOLWTy1-1	RET-27	
OR	YORWTy1-2	RET-28	
PL	YPLWTy1-1	RET-29 ^{<i>h</i>}	
PR1	YPRCTy1-2	RET-30 ^{<i>h</i>}	1
PR2	YPRWTy1-3	RET-31	
PR3	YPRCTv1-4	RET-32	1

^{*a*} for the RETs found to act as NET parents

^{*h*}RET belonging to the highly expressed subgroup

 (\ast) has been added to the MIPS CYGD names of RETs when they are identical to

the names of NETs isolated in the present study.

FEMS Yeast Research

546 Figure legend547

Fig. 1. Ty1 insertions reactivating ATCase in *ura2-s*. The *URA2*-coding region is represented by a box including the terminal ATCase-coding domain. Black asterisks mark the two nonsense point mutations in the ATCase-defective allele *ura2-s* with the nucleotide positions above (the initiating ATG is the reference position). Size scale is indicated. The positions of NETs A to Q inserted into *ura2-s* are given by flags directed in the corresponding NET orientation.

) I 554

> Fig. 2. RT-PCR assays used to detect URA2-related transcripts initiated at the sense NET 3' extremity. Coding region of the sense NET inserted into *ura2-s* is indicated by a striped box and its LTR, by black arrows. Lines below the map mark the DNA fragments amplified by RT-PCR: their extremities correspond to the positions of the primers specific to URA2 or Ty1 sequences. Lines (a) and (b) correspond to fragments amplified using a forward primer encompassing the NET-ura2 junction or hybridizing inside the LTR (positions 278-298), respectively. The dotted line (c) means that the amplification failed with the corresponding forward primer in the TYB region.

Fig. 3. Reconstitution of NET genealogy. The general map of a Ty1 element is presented at the top of the figure (figure is not to scale): the light grey box indicates the Ty1 coding sequence (TYA and TYB genes), LTRs are colored in dark grey and the (U3) and (RU5) segments are delimited by vertical bars. Specific features of non-autonomous RET-1, 6 and 20 and their coordinates are indicated in red with frameshift (fs) as asterisks. In NET maps (NET names are listed on the right, with 'S' footnote pointing to sense NETs), sequenced segments are given by bold lines and non-sequenced segments, by dotted lines. LTRs are surrounded by thin lines. A single map is given when several NETs showed the same

structure of segment inheritance from RET parents. The parental origin of each NET segment is indicated: figures above each segment correspond to the individual parental RET using the numbering scheme shown in Table 2 (names of non autonomous RETs are in red). Segments from highly expressed RETs are shaded in dark grey and those from weakly expressed RETs are in light grey tones, while segments that have not been unambiguously attributed a parental RET are in white. Segments from the 'uk' RET are presented in dark blue. Mutations inherited from non-autonomous RETs are shown in red. NETs are given in decreasing order of complete characterization, in terms of their parents identification: NETs with both parents identified are listed first (NETs A to F), then NETs with one identified parent and several possible second parents (G to O) and NETs with several possible parental combinations (P and Q). When several RETs are suitable for one segment, the following conventions were used: '11/16' indicates that either RET-11 or RET-16 are suitable; 'y' means that the parent belongs to the following subgroup of RETs: (6, 8, 9, 18, 25, 26, 29, 30); numbers in parenthesis are the name of presumed parents according to the minimum hypothesis that the NET involved is the progeny of a single RET, although the possibility cannot be ruled out that the corresponding segment may have been inherited from another RET showing local sequence identity. Additional data about NETs are given as follows: non-symmetrical 5' and 3' (R-U5) segments are underlined; RT-derived mutations and their coordinates are marked with green tags. NETs having at least one non-autonomous parent but that are nevertheless potentially autonomous have their name highlighted in yellow. For chimerical NETs, the minus-strong-stop DNA transfer mode is indicated on the right ('inter'- or 'intra'-molecular mode; 'nd': not determined). Internal template shifts and their coordinates, when available, are indicated by light blue boxes.

FEMS Yeast Research

Fig. 4. Structure of the Far Right-Arm transposition Hot-spot region of S288c chromosome III. (a) Chromosomal coordinates are indicated and sequence features specified according to the SGD are represented: this region is occupied by two Ty1 solo-LTRs/delta sequences and one Ty3 LTR carrying an additional nested Ty1 LTR (arrow heads give the corresponding U3-R-U5 orientation). With the exception of YCRWdelta11 (in black), the LTR sequences in grey and in white are not annotated in the SGD but were detected in Blast alignments. One of them contains the same 'uk' (R-U5) sequence (bold dashed line) as in NETs G and H. (b) Proposed structure of the corresponding locus deduced from NET analyses and from additional PCR amplifications performed with primers specific to *YCR028c* and *TYB* (*) or primers specific to *YCR027c* and *TYB* (**): two complete Ty1 elements are present in a head to head orientation, the second element is the potential RET-uk that carries sequence features identical to the *TYA-B* sequence of NET-H.

PP-P

FIGURE 1



FIGURE 2



..... (C)

FIGURE 3

U3 I	RU5	ΤΥΑ	1 20	ТҮВ	6	U3	RU5		
			* ** 2124 2545 fs+1 fs+2		5085-5 deleti	512 on			
6	425-510 590-1	6 26 610 675-830 980-	6 2470-2600 1350	26	6 4200-4350	6	6/26	nd	NET-A
	11	11		3700-3980	32	32		inter	NET-B
<u>20</u>	23 20 420-42	23 5			20 4600-4625	20	23	inter	NET-C
20	30	30	30		20	20	30	inter	NET-D
6	6/1	1	6		1	<u>6</u> 512-5545	6/1	nd	NET-E
6	6	11			6	6	<u>11</u>	inter	NET-F
8	uk	8			8	8	8	intra	NET-G ^s
20	<u>uk</u>	uk	uk		20	20	<u>20</u>	intra	NET-H ^s
20		20/y	y		20	20	У	inter	NET-i/J/K
20	20	(20)	20 fs+2		20	20	20		NET-L ^s /M
6	(6)	(6)			6	6	(6)		NET-N
(8)	(8)	(8)				(8)	<u>(8)</u>		NET-O ^S
(19)	(19)	(19)) (19)	(19)		NET-P
11/16	У	11/16			11/1	6 11/16	ò y	inter	NET-Q



FIGURE 4


Réarrangements chromosomiques et génomique fonctionnelle chez la levure *Saccharomyces cerevisiae*. *G*énomique comparative des génomes mitochondriaux des levures hémiascomycètes

Les mutations ponctuelles et les remaniements chromosomiques comme les insertions, duplications ou translocations sont les moteurs essentiels de la plasticité et de l'évolution des génomes. Deux approches ont été menées dans ce travail.

La première approche repose sur l'étude des impacts des réarrangements chromosomiques et des gènes de fusion néoformés chez *Saccharomyces cerevisiae* et de leurs conséquences sur le fonctionnement cellulaire. Au laboratoire, grâce à un crible de sélection positive fondé sur un allèle mutant du gène *URA2* qui rend une souche auxotrophe pour l'uracile, nous avons retenu une collection de souches révertantes possédant différents types de réarrangements chromosomiques. Ces remaniements sont accompagnés de la fusion de la région codant l'ATCase du gène *URA2* avec différents gènes receveurs. Lors de ce travail, nous avons déterminé de manière précise l'influence des réarrangements chromosomiques et celle des gènes de fusion néoformés sur le fonctionnement de la cellulaire Les résultats obtenus montrent que ces réarrangements perturbent le fonctionnement de la cellule non pas selon la nature du remaniement mais en fonction de la ploïdie. Les gènes de fusion peuvent également engendrer des dysfonctionnements car les fonctions associées aux deux gènes impliqués dans la fusion peuvent être altérées comme le montre la variabilité de l'activité enzymatique de l'ATCase.

La seconde orientation s'appuie sur une analyse des génomes mitochondriaux des levures osmotolérantes *Pichia sorbitophila* et *Pichia farinosa* ainsi que leur comparaison avec d'autres séquences mitochondriales de levures hémiascomycètes. Cette étude indique que *P. sorbitophila* et *P. farinosa*, sont des espèces distantes d'un point de vue phylogénétique.

Chromosomal rearrangement and functional genomic in the yeast *Saccharomyces cerevisiae* and comparative genomic of mitochondrial genomes from hemiascomyceteous yeasts

Point mutations and gross chromosomal rearrangements (GCRs) such as insertions, duplications or translocations are key parameters of genome evolution. Two approaches have been used in this study.

The first one focused on the impacts of GCRs and fusion genes in *Saccharomyces cerevisiae*. In our laboratory, a positive selection screen based on a mutated allele of the *URA2* gene was used to obtain a set of mutants possessing different GCRs. In all cases, GCRs are linked to the fusion of the ATCase part of *URA2* with other genes. During this work, we have first precisely determined the impacts of GCRs and gene fusions on cell function. Main results show that GCRs impair cell function according to ploidy rather than the kind of rearrangement. Fusion genes can also lead to dysfunctions because functions associated to genes implicated in this fusion can be impaired as it is the case for the ATCase activity.

The second approach of this study focused on the analysis of mitochondrial genomes of the osmotolerant yeasts *Pichia sorbitophila* and *Pichia farnosa* and their comparison with other mitochondrial sequences of hemiascomyceteous yeasts. This study indicates that *P. sorbitophila* and *P. farinosa* are two phylogenetically distant species.

Discipline : Aspects Moléculaires et Cellulaires de la Biologie Mots clés : *Saccharomyces cerevisiae*, remaniements chromosomiques, génomique comparative, groupe CTG Laboratoire de Génétique Moléculaire, Génétique et Microbiologie UMR7156 UDS/CNRS 28 rue Goethe 67083 Strasbourg Cedex