

- Thèse présentée pour obtenir le grade de
- Docteur de l'Université Louis Pasteur
- Strasbourg I
-
-
- Discipline : Sciences du Vivant
- Aspects moléculaires et cellulaires de la biologie
- par Emilie Fritsch ▪

Réarrangements chromosomiques chez
Saccharomyces cerevisiae :
Influence du contexte génétique et mécanismes
impliqués dans leur apparition

Soutenue publiquement le 9 juin 2008 ▪

Membres du jury ▪

- Mme Micheline Wesolowski-Louvel (Professeur, Université Claude Bernard Lyon)
Rapporteur externe
- M. Serge Gangloff (Chargé de recherche, CEA Fontenay-aux-Roses)
Rapporteur externe
- M. Ermanno Candolfi (Professeur, Université Louis Pasteur Strasbourg)
Rapporteur interne
- M. Serge Potier (Professeur, Université Louis Pasteur Strasbourg)
Directeur de thèse
- M. Jacky de Montigny (Professeur, Université Louis Pasteur Strasbourg)
Co-directeur de thèse

REMERCIEMENTS

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

Je tiens à remercier Madame Micheline Wesolowski-Louvel (Université Claude Bernard, Lyon) et Messieurs Serge Gangloff (Laboratoire d'Etude de la Réparation de l'ADN-CEA, Fontenay-aux-Roses) et Ermanno Candolfi (Université Louis Pasteur, Strasbourg) d'avoir accepté d'examiner mon travail de thèse.

Merci à Jean-Luc Souciet d'avoir suivi ce projet durant ces quatre années. Merci pour ta confiance, ton soutien et ta disponibilité malgré un emploi du temps de ministre.

Un grand merci à Jacky de Montigny pour son encadrement durant le DEA et la thèse, pour m'avoir encouragée aussi bien dans la partie « recherche » que « enseignement ». Merci pour tes conseils et de m'avoir supportée dans les moments de stress. Nom d'un bachibouzouk, tu es mon tintinophile préféré.

Je tiens également à remercier Joseph Schacherer pour ses conseils avisés, sa disponibilité et son écoute.

Merci à la gente féminine du groupe Levures : Marie-Laure, Catherine, Valérie, Claudine, Laurence et Véronique. Un merci spécial pour Paul, mon padawan devenu jedi. Merci à Yves, ancien levuriste depuis reconverti, pour son aide lors de mon arrivée en DEA. Ce travail a également bénéficié de l'aide de trois supers stagiaires : Céline Graber, Sandrine Vogt et Samantha Haller.

Merci à Audrey, Chris(ounet), Edith, Sandro, Marie, Emilie, Jessica, Mickaël, Tatina, Chris B. et Niamh pour tous les bons moments passés ensemble au laboratoire (et en dehors), les discussions de foot et de génétique et les pauses café. Merci à Christelle, seule compagnie à 7h30 au laboratoire, et à ses prédictions ésotériques. Merci à Sandrine pour sa bonne humeur et son côté tactile ;-).

Je tiens également à remercier tous les membres du laboratoire qui ont contribué de près ou de loin à l'aboutissement de ce travail. Merci également à Jean-Luc Dortet-Bernardet pour son aide avec les statistiques.

Merci à Manu et Natacha pour tous les bons moments passés devant du pain perdu ou des spaghettis carbonara. Partager ces années de thèse avec vous a été un vrai plaisir et a largement contribué à mon enthousiasme.

Enfin, le dernier mais pas le moindre, un énorme merci à Nic, mon binôme et compagnon diurne pendant toutes ces années. Merci pour tous les bons moments et merci aussi pour ton soutien sans faille dans les moments (heureusement plus rares) de doutes.

Une pensée particulière à mes parents et à toute ma famille qui m'ont soutenue pendant ces années de thèse mais également pendant celles passées sur les bancs de la fac. Leur soutien et leur confiance me sont très précieux.

Ce travail a été financé par une allocation du Ministère de l'Enseignement et de la Recherche et par un contrat d'attaché temporaire d'enseignement et de recherche à l'Université Louis Pasteur.

TABLE DES MATIÈRES

INTRODUCTION GENERALE

1. Conséquences des remaniements chromosomiques sur la cellule	22
1.1 Adaptation à l'environnement	22
1.2 Variabilité intra- et inter-espèces	24
1.2.1 Variabilité intra-espèce	24
1.2.2 Variabilité inter-espèces et spéciation	26
1.2.3 Le devenir des événements de duplication	28
1.2.3.1 La conservation de la fonction	29
1.2.3.2 La subfonctionnalisation	31
1.2.3.3 La néofonctionnalisation	31
1.2.3.4 La pseudogénéisation	32
1.3 Les remaniements chromosomiques et les maladies génétiques chez l'Homme	33
2. Apparition des remaniements chromosomiques	34
3. <i>Saccharomyces cerevisiae</i> , un bon modèle d'étude pour les remaniements chromosomiques	35
4. Les mécanismes de recombinaison	36
4.1 Les mécanismes de recombinaison homologue	37
4.1.1 Principe général	37
4.1.2 Les systèmes pour étudier les mécanismes de recombinaison homologue	37
4.1.3 Les différents mécanismes de recombinaison homologue	38
4.1.3.1 Le modèle de réparation des cassures d'ADN double brin (DSBR)	38
4.1.3.2 Le mécanisme de synthesis-dependent strand annealing (SDSA)	39
4.1.3.3 Le mécanisme de break-induced replication (BIR)	39
4.1.3.4 Le mécanisme de single-strand annealing (SSA)	42
4.1.4 Les principales protéines impliquées dans la recombinaison homologue	42
4.2 Le Non-Homologous End Joining (NHEJ)	44
4.2.1 Principe général	44
4.2.2 Les systèmes pour étudier le mécanisme de NHEJ	45

4.2.3 Le mécanisme de NHEJ	45
4.2.4 Les principales protéines intervenant dans le mécanisme de NHEJ	46
4.3 Le Microhomology-Mediated End Joining (MMEJ)	47

MATÉRIEL ET MÉTHODES

1. Organismes et milieux de culture	51
1.1 <i>Saccharomyces cerevisiae</i>	51
1.1.1 Biologie de <i>Saccharomyces cerevisiae</i>	51
1.1.2 Souches utilisées	51
1.2 Autres levures utilisées	52
1.3 Milieux de culture	52
1.4 Conditions de culture	53
1.5 Conservation des souches	53
1.6 Mesure de la croissance cellulaire	53
2. Techniques génétiques de la levure	53
2.1 Obtention de cellules diploïdes	53
2.2 Obtention de cellules haploïdes	54
2.3 Analyse des spores	54
2.3.1 Dissection des tétrades	54
2.3.2 Analyse des spores en vrac par test éther	54
2.4 Analyse phénotypique	54
2.4.1 Détermination des phénotypes	54
2.4.2 Détermination du signe sexuel	55
3. Préparation d'ADN	55
3.1 Préparation rapide d'ADN génomique (Hoffman and Winston, 1987)	55
3.2 Préparation des chromosomes entiers de levure (Vézinhet <i>et al.</i> , 1990)	55
4. Analyse de l'ADN	56
4.1 Précipitation alcoolique de l'ADN	56
4.2 Purification de l'ADN par chromatographie	56
4.3 Détermination de la concentration d'ADN	57
4.4 Analyse par électrophorèse sur gel d'agarose	57
4.4.1 Electrophorèse en gel d'agarose	57
4.4.2 Electrophorèse en champs alternés	57

4.5 Digestion par des endonucléases de restriction	58
5. Hybridation ADN/ADN selon la méthode de Southern (1975)	58
5.1 Transfert	58
5.2 Hybridation	59
5.2.1 Préparation d'une sonde froide	59
5.2.2 Préhybridation et hybridation	60
5.3 Révélation au NBT/BCIP	60
6. Amplification d'un fragment d'ADN par PCR	60
6.1 Paramètres importants lors des différentes étapes	60
6.2 Matériel	61
7. Détermination de la séquence d'un fragment d'ADN	62
8. Système de sélection des révertants Ura ⁺	63
9. Informatique	64

PARTIE I

1 SÉLECTION ET CARACTÉRISATION DES REMANIEMENTS CHROMOSOMIQUES

1. Introduction	65
2. Le système de sélection	68
3. Analyse statistique des événements de mutation	71
3.1 Taux de mutation par cellule par sélection	72
3.2 Intervalle de confiance à 95% associé au taux de mutation	74
3.3 Taux de mutation rapporté à chaque type de réarrangements chromosomiques	74
3.4 Exemple de calcul du taux de mutation ainsi que de son intervalle de confiance	75
3.5 Exemple de calcul du taux de mutation relatif aux différents réarrangements	77
4. Caractérisation moléculaire des remaniements chromosomiques	78
4.1 Analyse du profil de restriction <i>Bam</i> HI par la méthode de Southern blot	78
4.2 Analyse des délétions et des insertions de rétrotransposon Ty1 par PCR	79
4.2.1 Mise en évidence des délétions	79
4.2.2 Mise en évidence des insertions de rétrotransposons Ty1	81

4.3 Analyse des événements de duplication	81
4.3.1 Localisation chromosomique des régions dupliquées	81
4.3.2 Analyse des duplications	83
5. Conclusion	84

2 INFLUENCE DU CONTEXTE GÉNÉTIQUE SUR L'APPARITION DES REMANIEMENTS CHROMOSOMIQUES

1. Introduction	87
2. La variabilité intra-espèce a une influence sur les remaniements sélectionnés	90
2.1 Résultats observés en contexte FL100	90
2.2 Résultats des sélections en contexte S288c	92
2.2.1 Analyse des événements de délétion	94
2.3 Comparaison des résultats obtenus en contextes FL100 et S288c	95
3. L'activité des rétrotransposons est plus faible en contexte S288c	99
3.1 Test de l'activité des rétrotransposons	99
3.2 Résultats des sélections réalisées à 25°C	99
3.3 Comparaison des résultats obtenus à 25 et à 30°C	100
4. Discussion	100

3 MÉCANISMES D'APPARITION DES DÉLÉTIONS ET DES DUPLICATIONS CHROMOSOMIQUES

1. Introduction	105
2. Influence des mécanismes de recombinaison homologue	106
2.1 Résultats obtenus pour la souche aUHT sauvage	107
2.2 Le taux de délétion est accru en contexte $\Delta rad52$	107
2.2.1 Résultats obtenus en contexte $\Delta rad52$	107
2.2.2 Caractérisation des événements de délétions en contexte $\Delta rad52$	108
2.2.3 Comparaison des résultats obtenus en contexte $\Delta rad52$ avec ceux obtenus dans la souche aUHT	109
2.3 Rad59p restreint l'apparition des duplications	110
2.3.1 Résultats des sélections réalisées en contexte $\Delta rad59$	110
2.3.2 Caractérisation des délétions en contexte a $\Delta rad59$	111

2.3.3 Comparaison du contexte a $\Delta rad59$ avec le contexte de référence aUHT	112
2.4 Conclusions	113
3. Implication du mécanisme de NHEJ	114
3.1 L'apparition des délétions et des duplications est indépendante de Yku80p	114
3.1.1 Résultats des sélections réalisées en contexte $\Delta yku80$	114
3.1.2 Caractérisation des événements de délétion	115
3.1.3 Caractérisation des duplications	116
3.1.4 Comparaison des résultats obtenus pour les souches a $\Delta yku80$ et aUHT	117
3.2 Une délétion du gène <i>LIG4</i> conduit à une diminution du taux d'apparition des délétions et des duplications	117
3.2.1 Résultats des sélections réalisées en contexte $\Delta lig4$	117
3.2.2 Caractérisation moléculaire des révertants obtenus en contexte $\Delta lig4$	118
3.2.3 Comparaison des résultats obtenus dans le contexte $\Delta lig4$ et aUHT sauvage	119
3.3 Conclusions	120
4. Un autre mécanisme intervient également dans l'apparition des remaniements chromosomiques	121
4.1 Résultats des sélections réalisées en contexte $\Delta lig4\Delta rad52$	121
4.1.1 Analyse des événements de délétion	121
4.2 Comparaison des résultats obtenus dans la souche $\Delta lig4\Delta rad52$ et la souche aUHT	122
5. Influence de la ploïdie sur l'apparition des remaniements	123
6. Discussion	124
 DISCUSSION ET PERSPECTIVES	 131

PARTIE II

LES FAMILLES MULTIGÉNIQUES *FUR4/FUI1/DAL4* ET *FCY2/FCY21/FCY22* CHEZ LES HÉMIASCOMYCÈTES

1. Introduction	139
2. La famille <i>FUR4/FUI1/DAL4</i>	144

2.1 Les gènes <i>FUR4</i> , <i>FUI1</i> et <i>DAL4</i> chez <i>Saccharomyces cerevisiae</i>	144
2.2 La famille <i>FUR4/FUI1/DAL4</i> chez les hémiascomycètes	147
2.2.1 Recherche d'homologues chez les hémiascomycètes	147
2.2.2 Synténie des différents gènes homologues	150
2.2.3 Etude <i>in vivo</i> des perméases à l'uracile, l'uridine et l'allantoïne chez les hémiascomycètes	153
2.3 Conclusions pour la famille <i>FUR4/FUI1/DAL4</i> chez les hémiascomycètes	155
3. La famille <i>FCY2/FCY21/FCY22</i>	158
3.1 Les gènes <i>FCY2</i> , <i>FCY21</i> et <i>FCY22</i> chez <i>S. cerevisiae</i>	158
3.2 La famille <i>FCY2/FCY21/FCY22</i> chez les hémiascomycètes	159
3.2.1 Recherche d'homologues chez les hémiascomycètes	159
3.2.2 Synténie des différents gènes homologues	163
3.2.3 Etude <i>in vivo</i> de la perméase purine/cytosine chez les hémiascomycètes	163
3.3 Conclusion pour la famille <i>FCY2/FCY21/FCY22</i> chez les hémiascomycètes	165
4. Discussion	167
ENSEIGNEMENT ET RECHERCHE	169
RÉFÉRENCES BIBLIOGRAPHIQUES	173
ANNEXES	
Annexe 1 : Table de Student	191
Annexe 2 : «Influence of the genetic background on the appearance of chromosomal rearrangements in <i>Saccharomces cerevisiae</i> . »	193
Annexe 3: «Spontaneous deletions and reciprocal translocations in <i>Saccharomyces cerevisiae</i> : influence of ploidy. »	217

LISTE DES FIGURES

INTRODUCTION GENERALE

Figure I1	Principe expérimental de l'hybridation génomique comparative	25
Figure I2	Principe expérimental de l'analyse PEM	26
Figure I3	Arbre phylogénétique des hémiascomycètes	27
Figure I4	Devenir des gènes dupliqués	30
Figure I5	Exemple de néofonctionnalisation pour les gènes <i>EDN</i> et <i>ECP</i>	33
Figure I6	Modèle de réparation des cassures d'ADN double-brin (DSBR)	40
Figure I7	Les différents mécanismes de recombinaison homologue	41
Figure I8	Description du mécanisme de non-homologous end joining	46

MATÉRIEL ET MÉTHODES

Figure MM1	Cycle cellulaire de <i>Saccharomyces cerevisiae</i>	51
Figure MM2	Schéma du montage réalisé lors de l'étape de transfert	59
Figure MM3	Représentation du système de sélections	63

PARTIE I

1 SÉLECTION ET CARACTÉRISATION DES REMANIEMENTS CHROMOSOMIQUES

Figure 1.1	Système de sélection <i>URA3 CAN1</i>	66
Figure 1.2	Principe du système de sélection <i>URA3::intron ACT1</i>	67
Figure 1.3	Le gène <i>URA2</i> , la protéine Ura2p et les étapes de la voie de biosynthèse des pyrimidines chez <i>Saccharomyces cerevisiae</i>	69
Figure 1.4	L'allèle <i>ura2_{15,30,72}</i> et les remaniements chromosomiques mis en évidence comme permettant la réactivation de l'activité ATCase	71
Figure 1.5	Table 4 de Lea et Coulson	73
Figure 1.6	Localisation des sites de restriction <i>Bam</i> HI au niveau du locus <i>ura2_{15,30,72}</i>	79

Figure 1.7	Exemple de profil de restriction <i>Bam</i> HI obtenu en Southern blot	79
Figure 1.8	Carte représentative des oligonucléotides utilisés pour mettre en évidence les événements de délétion	80
Figure 1.9	Principe de détection des délétions par amplification PCR	80
Figure 1.10	Exemple de microhomologies présentes aux bornes d'une délétion	81
Figure 1.11	Détection de l'insertion d'un rétrotransposon Ty1 dans le gène <i>URA2</i> par méthode PCR	82
Figure 1.12	Localisation chromosomique des régions dupliquées	82
Figure 1.13	Stratégie d'analyse des duplications géniques et segmentales par hybridation ADN/ADN de type Southern blot	84

2 INFLUENCE DU CONTEXTE GÉNÉTIQUE SUR L'APPARITION DES REMANIEMENTS CHROMOSOMIQUES

Figure 2.1	Généalogie de la souche S288c	89
Figure 2.2	Schéma récapitulatif des différents types de délétions sélectionnées en contexte FL100 haploïde et exemple de microhomologie observée aux bornes des délétions	91
Figure 2.3	Schéma représentatif des délétions obtenues en contexte S288c et des ATG initiateurs putatifs	94
Figure 2.4	Représentation des différentes étapes pour déterminer le nombre et la localisation chromosomique des éléments transposables dans un génome	97
Figure 2.5	Représentation du système de sélection utilisé par Curcio et Garfinkel (1991) pour mesurer la rétrotransposition	98
Figure 2.6	Schéma récapitulatif du mécanisme de rétroposition	103

3 MÉCANISMES D'APPARITION DES DÉLÉTIONS ET DES DUPLICATIONS CHROMOSOMIQUES

Figure 3.1	Représentation des différents ATG initiateurs putatifs utilisés pour les délétions en contexte $\Delta rad52$	109
------------	---	-----

Figure 3.2	Représentation des différents points d'initiation de la traduction prédits <i>in silico</i> en contexte $\Delta rad59$	112
Figure 3.3	Représentation des différents ATG initiateurs prédits <i>in silico</i> en contexte $\Delta yku80$	116
Figure 3.4	Représentation des différents ATG initiateurs prédits <i>in silico</i> en contexte $\Delta lig4$	119
Figure 3.5	Représentation des différents points d'initiation de la traduction prédits <i>in silico</i> en contexte $\Delta lig4\Delta rad52$	122
Figure 3.6	Mécanisme de NHEJ indépendant de la protéine Yku80p proposé comme responsable de l'apparition des délétions	126
Figure 3.7	Mécanisme de NHEJ indépendant de Yku80p proposé comme responsable de l'apparition des duplications segmentales	127
Figure 3.8	Principe général de l'hybridation génomique comparative	128
Figure 3.9	Représentation du mécanisme de MMEJ	129

DISCUSSION ET PERSPECTIVES

Figure C1	Système de sélection utilisant un gène codant pour une protéine fluorescente	135
-----------	--	-----

PARTIE II

LES FAMILLES MULTIGÉNIQUES *FUR4/FUI1/DAL4* ET *FCY2/FCY21/FCY22* CHEZ LES HÉMIASCOMYCÈTES

Figure 4.1	Définition des gènes orthologues et paralogues	140
Figure 4.2	Arbre phylogénétique représentant le groupe des <i>Saccharomyces</i>	141
Figure 4.3	Représentation de l'événement de duplication totale du génome chez les hémiascomycètes	143
Figure 4.4	Voie métabolique des pyrimidines	145
Figure 4.5	Voie métabolique de l'allantoïne	146

Figure 4.6	Comparaison de la séquence de Fui1p avec celle du produit du gène Spar_447_511	150
Figure 4.7	Carte des synténies pour les gènes <i>FUR4</i> , <i>FUI1</i> et <i>DAL4</i> et leurs homologues dans les espèces étudiées	151
Figure 4.8	Arbre phylogénétique des gènes trouvés dans les différentes espèces hémiascomycètes étudiées pour la famille des perméases à l'uracile, l'uridine et l'allantoïne	157
Figure 4.9	Organisation des gènes <i>FCY2</i> , <i>FCY21</i> et <i>FCY22</i>	158
Figure 4.10	Représentation des deux cadres de lecture Spar_c422_6257 et Spar_c421_6259 présentant une similarité de séquences avec Fcy2p (d'après Synteny viewer, SGD)	161
Figure 4.11	Alignements locaux réalisés d'après l'algorithme de Smith et Waterman	162
Figure 4.12	Carte de synténie pour les gènes <i>FCY2</i> , <i>FCY21</i> , <i>FCY22</i> et leurs homologues dans les différentes espèces étudiées	164
Figure 4.13	Arbre phylogénétique des gènes trouvés dans les différentes espèces hémiascomycètes étudiées pour la famille des perméases aux purines et à la cytosine	166

LISTE DES TABLEAUX

MATÉRIEL ET MÉTHODES

Tableau MM1 Souches de <i>Saccharomyces cerevisiae</i> utilisées lors de cette étude	52
Tableau MM2 Souches utilisées dans le cadre de l'étude Génolevures	52
Tableau MM3 Liste des oligonucléotides amorces utilisés lors de cette étude	62

PARTIE I

1 SÉLECTION ET CARACTÉRISATION DES REMANIEMENTS CHROMOSOMIQUES

Tableau 1.1 Nombre de sélections en fonction du nombre de révertants obtenus par sélection	75
Tableau 1.2 Valeurs de s_x obtenues d'après la table 4 de Lea et Coulson en fonction du nombre r de révertants obtenus et du nombre N_x correspondant pour l'estimateur $m_e = 0,30$	75
Tableau 1.3 Valeurs de s_x obtenues d'après la table 4 de Lea et Coulson en fonction du nombre r de révertants obtenus et du nombre N_x correspondant pour l'estimateur $m_e = 0,25$	76

2 INFLUENCE DU CONTEXTE GÉNÉTIQUE SUR L'APPARITION DES REMANIEMENTS CHROMOSOMIQUES

Tableau 2.1 Récapitulatif des taux de mutation globaux et relatifs à chaque événement en contexte FL100 et S288c	93
Tableau 2.2 Résultats des sélections effectuées dans la souche S288c aUHT	92
Tableau 2.3 Récapitulatif des différentes délétions et des microhomologies observées dans le contexte aUHT.	93

3 MÉCANISMES D'APPARITION DES DÉLÉTIONS ET DES DUPLICATIONS CHROMOSOMIQUES

Tableau 3.1 Récapitulatif des résultats obtenus dans la souche aUHT	107
---	-----

Tableau 3.2	Résultats obtenus dans le contexte $\Delta rad52$	108
Tableau 3.3	caractérisation des différentes délétions et de leurs microhomologies	109
Tableau 3.4	Tableau récapitulatif des sélections réalisées dans la souche de référence aUHT et dans la souche a ΔR	109
Tableau 3.5	Récapitulatif des résultats pour les sélections réalisées en contexte a $\Delta rad59$	111
Tableau 3.6	Tableau des différentes délétions obtenues pour les sélections en contexte a $\Delta rad59$	112
Tableau 3.7	Tableau comparatif des résultats obtenus en contextes $\Delta rad59$ et aUHT	113
Tableau 3.8	Tableau récapitulatif des résultats obtenus en contexte $\Delta yku80$	115
Tableau 3.9	Description des différentes délétions caractérisées en contexte $\Delta yku80$	116
Tableau 3.10	Tableau récapitulatif des sélections réalisées dans les souches a $\Delta yku80$ et aUHT	117
Tableau 3.11	Résultats des sélections réalisées en contexte $\Delta lig4$	118
Tableau 3.12	Caractérisation des délétions obtenues en contexte $\Delta lig4$	119
Tableau 3.13	Récapitulatif des sélections réalisées en contexte $\Delta lig4$ et celui de référence	119
Tableau 3.14	Récapitulatif des sélections réalisées en contexte $\Delta lig4 \Delta rad52$	121
Tableau 3.15	Description des délétions caractérisées en contexte $\Delta lig4 \Delta rad52$	122
Tableau 3.16	Tableau récapitulatif des résultats obtenus en contexte aUHT et a $\Delta lig4 \Delta rad52$	122

PARTIE II

LES FAMILLES MULTIGÉNIQUES *FUR4/FUI1/DAL4* ET *FCY2/FCY21/FCY22* CHEZ LES HÉMIASCOMYCÈTES

Tableau 4.1	Pourcentages d'identité obtenus pour les gènes <i>FUR4</i> , <i>FUI1</i> , <i>DAL4</i> et pourcentages de similarité pour leurs produits Fur4p, Fui1p et Dal4p	147
-------------	--	-----

Tableau 4.2	Nombre de gènes présentant une identité significative avec les gènes <i>FUR4</i> , <i>FUII</i> et <i>DAL4</i> de <i>S. cerevisiae</i> dans le génome des autres levures hémiascomycètes séquencées	148
Tableau 4.3	Pourcentages de similarité observés pour les protéines des gènes homologues à <i>FUR4</i> , <i>FUII</i> et <i>DAL4</i>	149
Tableau 4.4	Résistance des souches de différentes espèces du groupe des hémiascomycètes à une gamme de 5-fluorouracile (5FU) (A.) et de 5-fluorouridine (5FUI) (B.).	154
Tableau 4.5	Pourcentages d'identité obtenus pour les gènes <i>FCY2</i> , <i>FCY21</i> , <i>FCY22</i> et de similarité pour leurs produits Fcy2p, Fcy21p et Fcy22p	159
Tableau 4.6	Nombre de gènes présentant une forte identité de séquence avec les gènes <i>FCY2</i> , <i>FCY21</i> et <i>FCY22</i> de <i>S. cerevisiae</i> trouvés dans les autres levures hémiascomycètes séquencées	159
Tableau 4.7	Pourcentages de similarité observés pour les protéines des gènes homologues à <i>FCY2</i> , <i>FCY21</i> et <i>FCY22</i>	161
Tableau 4.8	Résistance des souches de différentes espèces du groupe des hémiascomycètes à une gamme de 5-fluorocytosine	165

LISTE DES GRAPHIQUES

PARTIE I

2 INFLUENCE DU CONTEXTE GÉNÉTIQUE SUR L'APPARITION DES REMANIEMENTS CHROMOSOMIQUES

Graphique 2.1 Représentation des taux de mutation globaux et relatifs aux différents réarrangements pour les contextes FL100 et S288c 95

Graphique 2.2 Représentation des taux de mutation pour la souche S288c aUHT à 30°C et à 25°C 100

3 MÉCANISMES D'APPARITION DES DÉLÉTIONS ET DES DUPLICATIONS CHROMOSOMIQUES

Graphique 3.1 Représentation des taux de mutation globaux et pour chaque type d'événements pour les souches aUHT et a Δ R 110

Graphique 3.2 Représentation des taux de mutation globaux et pour chaque type d'événements obtenus en contexte aUHT et a Δ rad59 113

Graphique 3.3 Comparaison des taux de mutation calculés pour les souches a Δ yku80 et aUHT 117

Graphique 3.4 Comparatif des taux de mutations pour les souches aFL et aUHT 120

Graphique 3.5 Comparaison des taux de mutation déterminés pour les souches a Δ lig4 Δ rad52 et aUHT 123

LISTE DES ABRÉVIATIONS

aa	: acides aminés
ADN	: acide désoxyribonucléique
ARN	: acide ribonucléique
BET	: bromure d'éthidium
Δ	: délétion
dNTP	: 2'-désoxyribonucléoside-5'-triphosphate
DTT	: dithiothréitol
EDTA	: éthylène diamine tétraacétate
min	: minute(s)
pb	: paire(s) de base
qsp	: quantité suffisante pour
rpm	: rotations par minute
sec	: seconde(s)
Tris	: tris-hydroxyméthyl-aminométhane
UV	: ultraviolet

INTRODUCTION

Les premiers travaux de génétique classique réalisés après la découverte des lois de Mendel ont présenté le génome comme une structure plutôt statique. Le fait qu'un certain nombre de fonctions essentielles doivent être conservées dans un organisme donné laissait supposer que seules quelques modifications mineures pouvaient être acceptées. On pensait alors que les modifications intervenaient majoritairement lors du brassage génétique et que les mutations ne jouaient qu'un rôle très faible dans la variation du génome.

Cette idée commença à être remise en cause dans les années 50 avec les travaux de Barbara McClintock (1956). Lors d'études génétiques menées sur des plants de maïs, elle proposa l'existence de gènes mobiles dits « gènes sauteurs » dans leur génome. Ces éléments correspondaient aux premiers éléments transposables mis en évidence dans une espèce. Ce n'est pourtant qu'une trentaine d'années plus tard que le travail de Barbara McClintock fut récompensé, lorsqu'elle obtint le prix Nobel en 1983. Les avancées en génétique moléculaire à partir des années 80 et les développements actuels de la génomique ont permis de confirmer largement que le génome est une structure dynamique, en partie du fait de la présence d'éléments transposables.

Il a été ainsi mis en évidence que les modifications de la structure des chromosomes sont des éléments moteurs essentiels à la plasticité et à l'évolution des génomes. Différents types d'événements ont été décrits comme étant à l'origine de variabilité génomique. Il peut s'agir de délétions, de duplications et d'insertions de transposons. Des translocations réciproques et non-réciproques ainsi que des inversions sont également sources de cette plasticité.

Ces remaniements permettent une flexibilité génomique qui peut avoir des conséquences importantes pour la cellule. Ils peuvent être impliqués dans l'adaptabilité d'une cellule ou d'un organisme à son environnement. Ils peuvent également jouer un rôle au niveau évolutif et peuvent contribuer à la mise en place des phénomènes de spéciation.

Les génomes tels que nous les connaissons à l'heure actuelle grâce aux différents programmes de séquençage résultent de l'accumulation de tels événements. La comparaison des génomes d'espèces proches a montré qu'un nombre très limité de réarrangements permet de les distinguer. Ces événements sont principalement un nombre réduit de translocations, de duplications segmentales ou d'inversions. Par exemple, les génomes de *Saccharomyces cerevisiae* et *Saccharomyces paradoxus* appartenant au groupe des levures hémiascomycètes ne peuvent être différenciés que par quatre inversions et trois duplications segmentales (Kellis

et al., 2003). Le séquençage de différents individus d'une même espèce a également démontré que les remaniements chromosomiques sont également à l'origine de variations intra-espèces. Chez la drosophile, des variations de la taille du génome entre individus de la même espèce ont été décrites (Bosco *et al.*, 2007). Elles seraient dues à des variations au niveau de l'hétéochromatine (Halfer, 1981).

Enfin, les remaniements chromosomiques sont également parfois responsables d'effets plus délétères pour les organismes pluricellulaires, car ils peuvent être impliqués dans certaines maladies génétiques ou dans des cancers.

Au laboratoire de « Génétique Moléculaire, Génomique et Microbiologie », une des thématiques abordées est l'étude des mécanismes conduisant à la formation ainsi qu'à la disparition des gènes, au remodelage des génomes, dans le cadre de l'évolution globale des génomes chez les eucaryotes. Lors de ce travail, nous nous sommes intéressés à la variabilité intra-espèces et à ses conséquences sur l'apparition des remaniements chromosomiques chez l'organisme modèle *Saccharomyces cerevisiae*. Notre étude a également porté sur l'identification des mécanismes intervenant dans l'apparition d'événements de délétions et de duplications.

Dans cette introduction, nous aborderons les différentes conséquences que peuvent avoir les remaniements chromosomiques pour les cellules. Nous traiterons également les conditions pour lesquelles ces réarrangements peuvent apparaître. Enfin, les mécanismes de recombinaison homologue et non-homologue seront détaillés.

1. Conséquences des remaniements chromosomiques sur la cellule

1.1 Adaptation à l'environnement

Les remaniements chromosomiques intervenant dans le génome d'une espèce donnée peuvent lui conférer un avantage sélectif dans certaines conditions environnementales. Ce phénomène a été décrit chez les procaryotes. Il a été montré que des souches de *Pseudomonas aeruginosa* isolés de poumons de patients atteints de mucoviscidose présentent des inversions chromosomiques de grande taille (4 Mb environ) au niveau de leur génome (Kresse *et al.*, 2003) par rapport à la souche standard PAO. Cette maladie conduit à la formation de mucus épais au niveau des poumons. Les inversions chromosomiques décrites lors de cette étude semblent permettre aux souches de s'adapter plus facilement à ce type

d'environnement. A l'heure actuelle, l'analyse moléculaire n'a pas encore permis de comprendre pourquoi ces remaniements conduisent à tel un avantage sélectif.

L'adaptation de certaines espèces à leur habitat par le biais de réarrangements chromosomiques a également été décrite chez les eucaryotes. La levure *S. cerevisiae* est un bon organisme d'étude pour analyser le lien entre les modifications du génome et son environnement. En effet, un nombre important de souches d'origines différentes a été isolé (laboratoire, œnologiques...). De plus, les souches sont en général facilement cultivables pour différentes conditions et les techniques de microbiologie classique peuvent être utilisées. La croissance des souches sur de nombreuses générations peut être menée en chemostat.

En 1992, Adams *et al.* ont développé un système pour caractériser les modifications génomiques survenant dans des souches de laboratoire après croissance en conditions limitantes en phosphate inorganique. Des souches haploïdes et diploïdes ont été cultivées durant environ 1000 générations en chemostat puis ont été isolées et analysées. Il a ainsi été montré quelles étaient porteuses de deux types majeurs de remaniements à savoir des délétions et des duplications isolées ou organisées en tandem. Ces réarrangements étant retrouvés systématiquement dans les expériences, il a été supposé qu'ils confèrent un avantage sélectif direct aux cellules porteuses de ces remaniements.

Par la suite, d'autres analyses ont également été menées sur l'adaptation de souches de laboratoire en conditions de croissance limitantes pour différents types de composés. En 2002, huit souches de *Saccharomyces cerevisiae* ont été étudiées après 500 générations de croissance en milieu limité en glucose (Dunham *et al.*, 2002). Il a été mis en évidence qu'elles étaient toutes porteuses de modifications génomiques complexes telles que des délétions ou duplications de bras de chromosomes conduisant à des aneuploïdies. Une autre étude conduite également en conditions limitées en glucose a révélé que des translocations au sein du génome des souches de laboratoire présentent un effet positif pour la croissance de celles-ci et permet l'adaptation des cellules à leur environnement (Colson *et al.*, 2004).

Le même type d'analyses réalisé pour des souches œnologiques possède en plus de l'intérêt scientifique, un intérêt appliqué. En effet, les caractéristiques phénotypiques que présentent certaines souches suite à l'acquisition de remaniements chromosomiques vont pouvoir être sélectionnées et utilisées à des fins commerciales. Néanmoins, l'étude de ces souches est plus difficile que pour celles de laboratoire car elles sont généralement homothalliques, fréquemment diploïdes, polyploïdes voire aneuploïdes. Ces caractéristiques

rendent les expériences de génétique plus difficiles à mettre en œuvre que pour les souches de laboratoire.

La résistance de souches œnologiques au sulfite a été abordée par Perez-Ortin *et al.* (2002). Le sulfite est un composé utilisé depuis les Egyptiens comme antioxydant et agent antimicrobien lors de la vinification. Certaines souches se sont adaptées à cet environnement et tolèrent des concentrations élevées de ce composé. Elles sont porteuses d'un remaniement chromosomique correspondant à une translocation au niveau d'un gène impliqué dans l'efflux de sulfite. Du fait de cette translocation, le gène se trouve sous contrôle d'un nouveau promoteur et son expression dans ces levures est accrue.

En 2007, Lucena et ses collaborateurs se sont intéressés aux levures *Saccharomyces cerevisiae* lors de la fermentation alcoolique de la canne à sucre. Les conditions de croissance utilisées génèrent de nombreux stress pour les cellules suite à des changements au niveau des concentrations en glucose et en éthanol, à des variations de température ou des pH dans le milieu. Là encore les souches analysées étaient toutes porteuses de remaniements chromosomiques à l'origine d'une meilleure adaptation de ces cellules à ces conditions difficiles.

1.2 Variabilité intra- et inter-espèces

Les remaniements chromosomiques peuvent s'accumuler au sein des génomes et ainsi conduire à une variabilité génomique pour différents individus d'une espèce donnée. Les analyses de génomique comparative permettent, dans le cas de génomes peu éloignés, de retracer l'histoire évolutive de ces espèces. Ces analyses ont montré que ces réarrangements peuvent également être à l'origine d'événements de spéciation.

1.2.1 Variabilité intra-espèce

Avec le séquençage d'un nombre croissant de génomes, une collection importante de données est disponible. Elle peut être utilisée pour des analyses de génomique comparative intra-espèces. Une variabilité intra-espèce a ainsi pu être mise en évidence pour différents organismes procaryotes tels que *Streptococcus pneumoniae* (Hakenbeck *et al.*, 2001), ou eucaryotes comme *Oriza sativa* (Yang *et al.*, 2007), *Drosophila melanogaster* (Bosco *et al.*, 2007) ou encore *S. cerevisiae* (Schacherer *et al.*, 2007).

Plusieurs études similaires ont également été réalisées chez l'Homme. Des remaniements chromosomiques ont été décrits comme responsables de maladies génétiques et de cancers. Il a, par exemple, été décrit que la leucémie promyélocytaire aigüe serait due à la

fusion des gènes PML et RAR α suite à un événement de translocation (Koken *et al.* 1997). Néanmoins, il a également été montré que des variations structurales comme des délétions, des duplications ou des insertions de tailles importantes (supérieures à 50 kb) peuvent également être retrouvées pour des individus sains (Redon *et al.*, 2006).

Une des méthodes les plus utilisées à l'heure actuelle pour détecter des remaniements chromosomiques dans le génome est l'hybridation génomique comparative (CGH). Pour réaliser cette technique, les ADN génomiques d'un individu d'intérêt et de l'individu de référence sont isolés puis marqués avec deux fluorochromes différents. Ils sont ensuite hybridés sur des puces représentant l'ensemble du génome de l'espèce. La comparaison des deux signaux de fluorescence obtenus permet de déterminer les gènes dupliqués et délétés pour chaque individu (Figure I1).

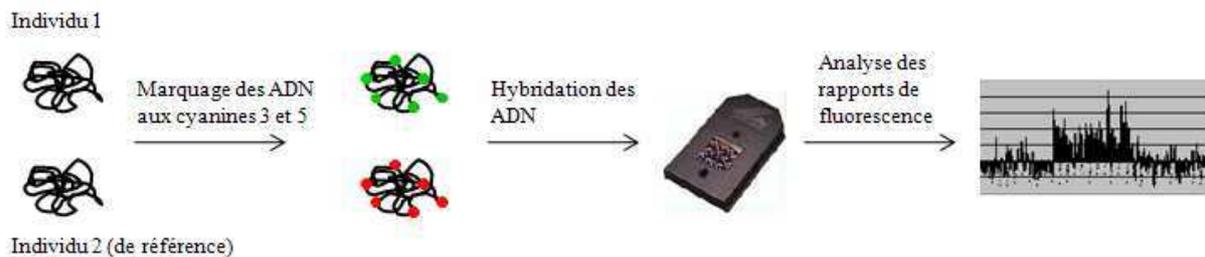


Figure I1 : Principe expérimental de l'hybridation génomique comparative.

Récemment, une nouvelle méthode d'analyse a été développée pour mettre en évidence ces modifications structurales, il s'agit du PEM (Paired-End Mapping) (Korbel *et al.*, 2007). L'expérience est décrite dans la Figure I2. L'ADN de l'individu est isolé puis fragmenté pour obtenir des fragments de 3 kb environ. Des adaptateurs sont ensuite fixés aux bornes des fragments ce qui permet leur circularisation. Ils sont ensuite digérés et les bornes sont séquencées par la technologie de séquençage 454 (Marguiles *et al.*, 2005). Cette technique permet de séquencer rapidement des fragments d'une centaine de paires de bases. Les réarrangements sont ensuite identifiés car les fragments amplifiés présentent des différences de taille avec celle attendue, déterminée par la séquence du génome de référence. Différents types d'événements peuvent être mis en évidence : des délétions, des insertions ou des inversions. Lors de leur étude, les génomes de deux femmes ont été comparés par rapport au génome séquencé. 1297 variations structurales ont ainsi été observées dont 1175 événements d'insertions ou délétions (indels) et 122 inversions d'une taille supérieure à 3 kb.

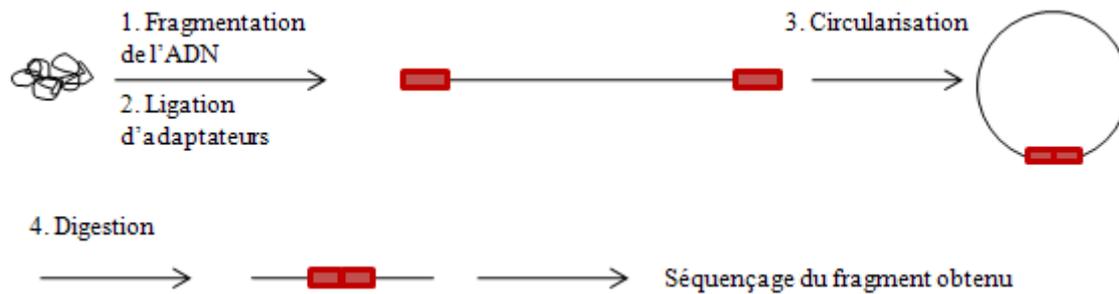


Figure I2 : Principe expérimental de l'analyse PEM (d'après Korbel *et al.*, 2007).

Cette technique développée chez l'Homme peut être adaptée à d'autres organismes pour caractériser les variations intra-espèces. Elle présente l'avantage de pouvoir travailler au niveau nucléotidique et ainsi de pouvoir déterminer directement les séquences aux bornes de réarrangements. Les données obtenues à partir d'analyses CGH permettent de déterminer les gènes impliqués mais pas les séquences au point de jonction. Dans ce type d'expériences, il est donc nécessaire, une fois les données de CGH obtenues, d'amplifier les jonctions des remaniements chromosomiques par PCR puis de séquencer les fragments obtenus pour avoir la séquence précise aux bornes des remaniements chromosomiques.

1.2.2 Variabilité inter-espèces et spéciation

Un des objectifs de la génomique comparative est de pouvoir retracer l'évolution des génomes et déterminer les événements à l'origine de ce phénomène évolutif. En effet, les remaniements chromosomiques se produisant dans les différentes espèces ont ainsi façonné les génomes au cours du temps. Plus les espèces seront distantes évolutivement, plus les remaniements se seront accumulés. Ceci rend l'analyse très difficile car les bases communes des deux génomes sont alors difficilement identifiables. L'étude d'espèces proches sera donc beaucoup plus informative et permettra généralement de mettre en évidence des modifications dues à des remaniements chromosomiques.

Le groupe des *Saccharomyces sensu stricto* comprend la levure *S. cerevisiae* ainsi que des levures dont les ancêtres se sont séparés de celle-ci il y a 5 à 20 millions d'années (Figure I3). L'analyse des séquences de trois membres de ce groupe à savoir *Saccharomyces paradoxus*, *Saccharomyces mikatae* et *Saccharomyces bayanus* a montré que les génomes présentent une bonne conservation de la synténie c'est-à-dire que l'ordre des gènes entre les espèces est fortement maintenu (Kellis *et al.* 2003). Néanmoins, des modifications du génome suite à des remaniements chromosomiques existent (Fischer *et al.*, 2000 ; Kellis *et al.*, 2003). Il a ainsi été remarqué que le génome de *S. paradoxus* diffère de

S. cerevisiae par 4 inversions et 3 duplications segmentales. *S. mikatae* est porteuse de 4 translocations réciproques et 13 inversions et *S. bayanus* comporte 5 translocations réciproques et trois inversions par rapport au génome de *S. cerevisiae*.

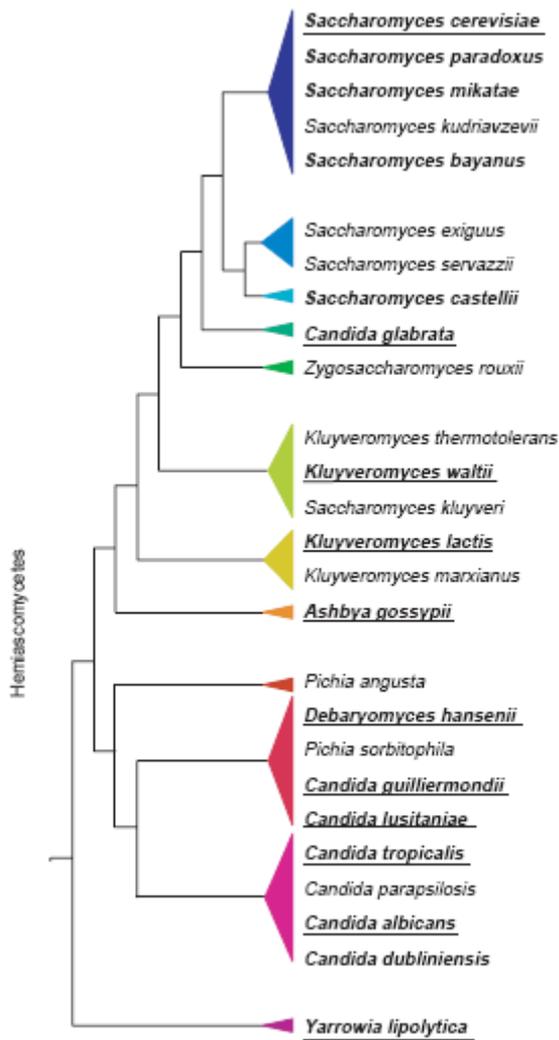


Figure I3 : Arbre phylogénétique des hémiascomycètes (d'après Dujon, 2006).

L'arbre a été réalisé d'après l'étude phylogénétique réalisée par Kurtzman (2003).

Des comparaisons directes inter-espèces ont été menées pour d'autres organismes et notamment pour les génomes de l'Homme et du chimpanzé et ont révélé neuf événements d'inversions. Il a également été mis en évidence que le chromosome 2 chez l'Homme correspond à la fusion de deux chromosomes chez le chimpanzé. Des duplications segmentales ainsi que des délétions ont également été identifiées dans les deux génomes mais à des localisations différentes. Chez l'Homme, 32 Mb d'ADN correspondent à des événements d'insertions d'une taille s'étendant jusqu'à 15 kb et 8 Mb à des insertions de plus de 15 kb (pour revue : Kehrer-Sawatzki et Cooper, 2007).

Si les événements décrits précédemment ne concernent qu'une partie du génome, d'autres, plus complexes, concernant l'ensemble du génome comme notamment la duplication complète des chromosomes, peuvent également se produire.

Ce type de duplication fut caractérisé en 1997 pour le génome de la levure *S. cerevisiae* (Wolfe et Shields, 1997). Les auteurs ont déterminé les régions dupliquées au sein du génome par une approche BLASTP (Altschul *et al.*, 1990). Cette méthode repose sur la recherche de similarité de séquences en acides aminés pour toutes les protéines de levures. Cinquante-cinq régions dupliquées ont ainsi été mises en évidence contenant 376 paires de gènes homologues. Ces blocs de duplication ont été définis selon trois critères : (1) Le score obtenu pour chaque BLASTP devait être supérieur à 200 et avoir une probabilité associée inférieure à 10^{-18} , ceci afin d'éviter que ces blocs soient le fruit du hasard. (2) Au moins trois paires d'homologues avec des distances intergéniques entre ces gènes inférieures à 50 kb devaient se trouver dans chaque bloc. (3) L'ordre des paires de gènes homologues et leur orientation devaient être conservés. Pour 50 régions sur les 55, l'orientation du bloc entier par rapport au centromère est conservée pour les deux copies.

Cette étude a permis de proposer que ces blocs sont des vestiges d'un événement de duplication totale du génome qui se serait produit il y a 100 millions d'années environ après la divergence du groupe des *Kluyveromyces* avec celui des *Saccharomyces* (Figure I3). Des remaniements chromosomiques comme des délétions, des translocations et des duplications se seraient ensuite accumulés chez l'ancêtre des *sensu stricto* conduisant aux génomes de *S. cerevisiae* et des autres membres de ce clade tels que nous les connaissons actuellement.

Ce type d'événement a également été décrit dans d'autres lignées. Dans la lignée des chordées à l'origine des vertébrés, une duplication totale du génome aurait eu lieu (McLysaght *et al.*, 2002). Cet événement se serait également produit pour les xénopes (Hellsten *et al.*, 2007) ainsi que chez les poissons pour les téléostes (Amores *et al.*, 1998).

Ces différents types de réarrangements qu'ils soient circonscrits à quelques gènes ou qu'ils concernent l'ensemble du génome, conduisent à une modification du génome qui peut ensuite conduire à des événements de spéciation et à la variabilité inter-espèces.

1.2.3 Le devenir des événements de duplication

Les comparaisons génomiques inter-espèces permettent de suivre le devenir de remaniements chromosomiques au cours de l'évolution. Le cas des duplications est particulièrement bien documenté. L'apparition de gènes paralogues peut se produire suite à un

événement de duplication pouvant être de natures variées. En effet, plusieurs types de duplications ont été mis en évidence :

- Les duplications géniques
- Les duplications segmentales de plusieurs gènes contigus
- Les duplications de bras de chromosome
- Les duplications totales du génome

Les paralogues ainsi formés peuvent ensuite évoluer de différentes manières.

Plusieurs scénarios sont décrits (Figure I4) :

- La conservation de la fonction : la copie dupliquée conserve la même fonction que la copie originelle.
- La subfonctionnalisation : les deux copies se partagent dans ce cas les fonctions initialement portées par la copie originelle.
- La néofonctionnalisation : l'une des copies conserve la fonction originelle tandis que l'autre va en acquérir une nouvelle.
- La pseudogénéisation : l'accumulation de mutations dans l'une des copies entraîne la perte de fonction.

1.2.3.1 La conservation de la fonction

Une copie dupliquée ayant une valeur sélective neutre par rapport à la copie originelle, elle peut être perdue suite à l'accumulation de mutations ou suite à la délétion du gène. Il a ainsi été décrit que la levure *C. glabrata* aurait subi une perte sélective de gènes du métabolisme (environ 9%) après l'événement de duplication totale du génome ancestral survenu chez les hémiascomycètes. Les gènes intervenant dans la dégradation du galactose ont notamment été perdus (Dujon *et al.*, 2004). Une des hypothèses avancées pouvant expliquer cette observation serait que *C. glabrata* étant une levure pathogène de l'Homme, certaines voies métaboliques ne lui seraient plus nécessaires dans cet environnement.

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 tels que les ARN 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 celle-ci.

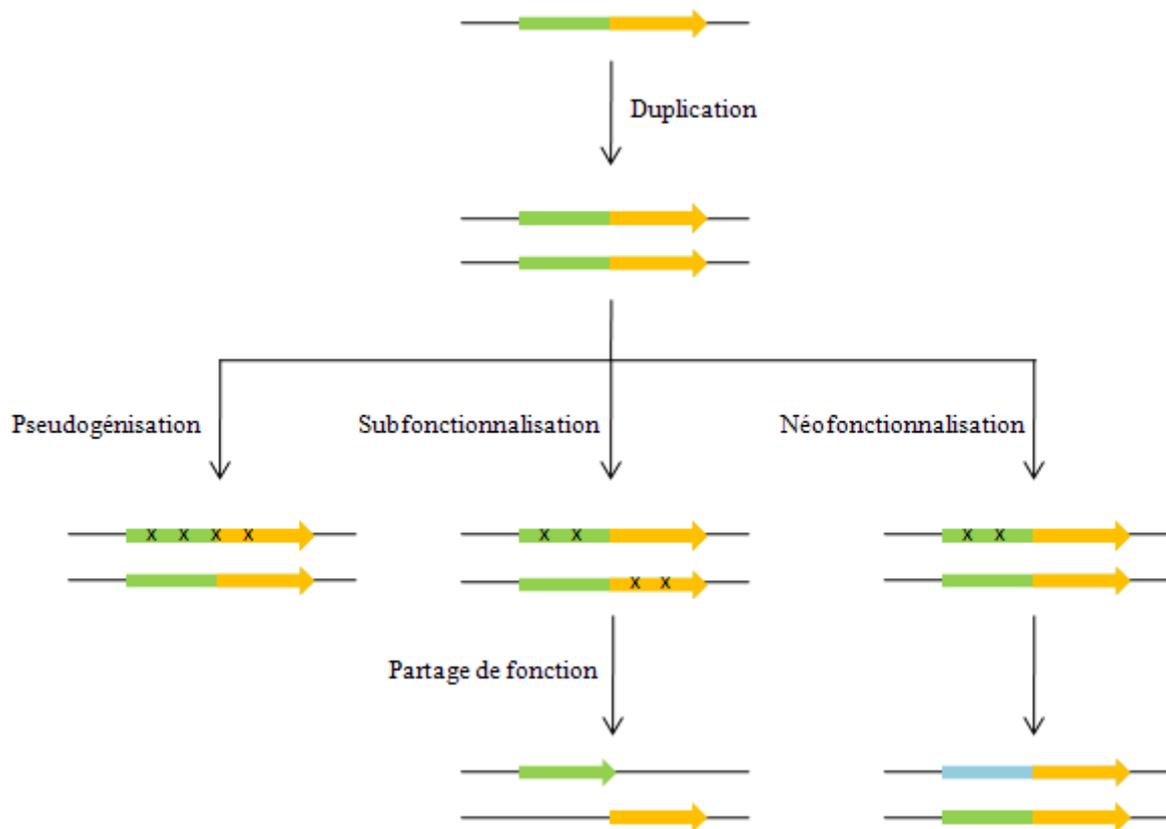


Figure 14 : Devenir des gènes dupliqués.

Le gène code pour une protéine avec deux domaines fonctionnels représentés l'un en vert et l'un en orange.

Chez les murinés, une duplication par rétroposition du gène *Ins2* codant pour les peptides de préproinsuline a généré la présence d'un deuxième gène nommé *Ins1* codant pour une fonction très proche et préservé dans les génomes de ces espèces (Shiao *et al.*, 2008).

Parfois, une régulation post-transcriptionnelle d'une des copies dupliquées permet la conservation du taux d'expression du composé. C'est le cas des gènes *RPS14A* et *RPS14B* chez *S. cerevisiae*. Ces deux gènes dupliqués codent pour la protéine rpS14 qui est un des constituants de la sous-unité ribosomique 40S. Dans une souche sauvage, *RPS14A* est dix fois plus exprimé que *RPS14B*. Ceci résulte de la fixation de rpS14 au niveau de l'ARN pré-messager de *RPS14B* bloquant ainsi son épissage et conduisant à sa dégradation. En revanche, en absence d'activité de *RPS14A* suite à son inactivation ou sa délétion, il y a perte de répression de du gène *RPS14B* qui est alors dix fois plus exprimé et permet de conserver un taux d'expression constant dans la cellule (Fewell et Woolford, 1999).

1.2.3.2 La subfonctionnalisation

La subfonctionnalisation est un processus permettant aux copies dupliquées d'acquérir une différence de fonction ou d'expression. Les deux copies peuvent ainsi être maintenues dans le génome. Un processus de subfonctionnalisation peut avoir lieu pour les gènes codant pour des protéines bifonctionnelles ou multifonctionnelles.

Les gènes *GAL1* et *GAL3* chez *S. cerevisiae* représentent un bon exemple de gènes issus d'une duplication à partir d'un tel gène (Hittinger et Carroll, 2007). Chez la levure, *GAL1* code pour la première enzyme de la voie de dégradation du galactose, la galactokinase tandis que *GAL3* code pour un co-inducteur pour cette voie de catabolisme. Ces deux gènes dupliqués sont issus de la duplication totale du génome. Chez *Kluyveromyces lactis*, espèce pré-duplication, un homologue du gène *GAL1* existe et code pour les deux fonctions, l'activité galactokinase et la co-induction. On peut donc supposer qu'après l'événement de duplication du génome, les deux copies ont évoluées, chacune conservant l'une des deux fonctions.

Les copies dupliquées peuvent également avoir acquis une différence d'expression. C'est par exemple le cas des gènes *URA5* et *URA10* chez *S. cerevisiae*. Il a été montré que ces deux gènes sont issus de la duplication totale du génome (Wolfe et Shields, 1997). Le gène *URA5* code pour une orotate phosphoribosyltransférase (OPRTase) qui catalyse la transformation de l'acide orotique en orotidine monophosphate dans la voie de biosynthèse des pyrimidines. Lorsque ce gène est muté ou délété, un phénotype de bradytrophie est observé pour l'uracile. La souche a alors une croissance lente sur milieu minimum par rapport à la souche sauvage (de Montigny *et al.*, 1989). Le gène *URA10* a ainsi été mis en évidence par complémentation de fonction (de Montigny *et al.*, 1990). L'activité de l'enzyme codée par le gène *URA10* représente 20% tandis que celle du produit du gène *URA5* correspond à 80% de l'activité OPRTase totale de la cellule. Des études de transcription ont montré que le gène *URA5* est exprimé constitutivement tandis que son paralogue *URA10* est régulé par le complexe PPR1-acide dihydroorotique. Ce complexe est également impliqué dans la régulation des gènes *URA4*, *URA1* et *URA3* de la voie de biosynthèse des pyrimidines. Cet exemple représente un exemple de subfonctionnalisation non pas au niveau des fonctions codées par les deux gènes dupliqués mais au niveau de leur régulation.

1.2.3.3 La néofonctionnalisation

Sur le plan évolutif, l'un des événements les plus importants est l'apparition d'une nouvelle fonction à partir de séquences dupliquées. Ce phénomène a été décrit chez les

primates pour deux gènes appartenant à la superfamille des ribonucléases, *EDN* (Eosinophil-derived neurotoxin) et *ECP* (eosinophil cationic protein). Les produits de ces deux gènes sont présents au niveau des granules des leucocytes éosinophiles. La protéine codée par le gène *EDN* joue un rôle d'agent antiviral grâce à son action ribonucléolytique sur l'ARN génomique des rétrovirus. *ECP* code, lui, pour une protéine ayant une faible activité ribonucléasique mais qui agit comme toxine pour différents parasites et bactéries.

Les hominidés et les singes de l'ancien monde comme le macaque ou le babouin possèdent ces deux gènes (Zhang *et al.*, 1998). En revanche, les singes du nouveau monde, comme le tamarin, ne possède qu'un seul gène homologue à *EDN* et *ECN*. Ce gène a été nommé *EDN* car la protéine pour laquelle il code ne présente aucune toxicité contre les bactéries ou les parasites. Néanmoins son activité RNase est beaucoup plus faible que celle observée pour son orthologue chez les singes de l'ancien monde.

Un événement de duplication en tandem du gène *EDN* ancestral serait survenu après la divergence des singes du nouveau monde (platyrhiniens) et de l'ancien monde (catarrhiniens) (Figure I5). Le gène ancestral ne possédait pas l'activité antibactérienne et antiparasitaire, ce qui laisse supposer que cette fonction a été acquise par *ECP* après la duplication.

1.2.3.4 La pseudogénisation

Lorsqu'une copie dupliquée n'apporte pas d'avantage sélectif pour un organisme, aucune pression de sélection ne s'exerce alors pour maintenir ce gène. Dans ce cas, les mutations dans une des copies n'entraîneront pas d'effets délétères pour la cellule et pourront alors être fixées. Ainsi, les mutations de types faux-sens, non-sens, de décalage de cadre de lecture ou au niveau des régions promotrices ou régulatrices vont pouvoir s'accumuler au sein d'une des copies conduisant à son inactivation. Celle-ci peut être due à un blocage de l'expression du gène ou au fait que la protéine ainsi produite n'est plus fonctionnelle. Le gène non-fonctionnel est alors qualifié de pseudogène.

Si l'événement de pseudogénisation est récent, le pseudogène pourra être mis en évidence par recherche de similarité de séquences avec la copie fonctionnelle. Chez *S. cerevisiae*, l'étude de la famille multigénique *DUP240* a déterminé qu'elle était composée de huit gènes fonctionnels mais également de deux pseudogènes et de trois reliques (Wirth *et al.*, 2005). Le terme relique définit un pseudogène dont la séquence est fortement dégénérée suite à l'accumulation de mutations.

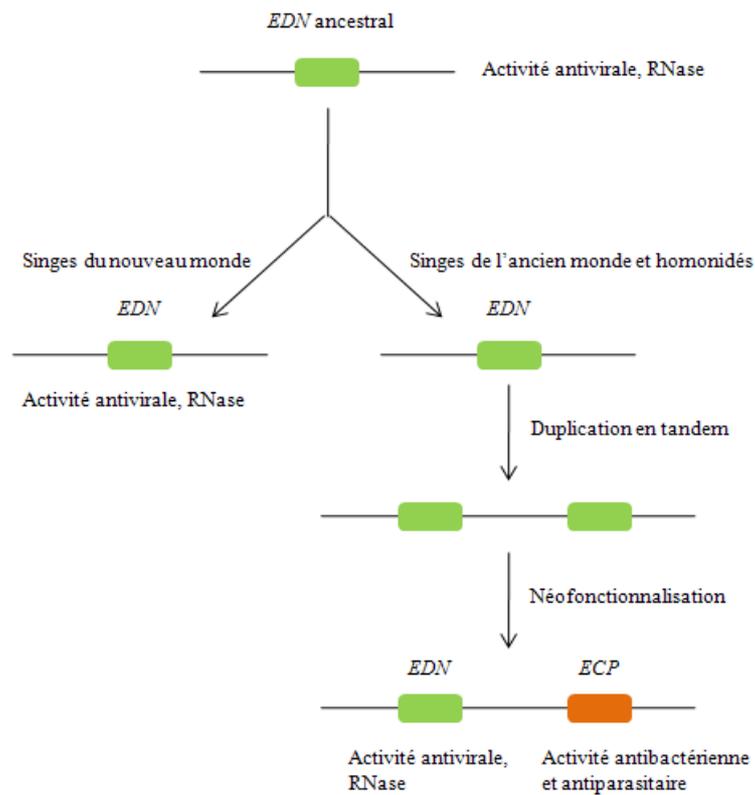


Figure 15 : Exemple de néofonctionnalisation pour les gènes *EDN* et *ECP*.

Chez l'Homme, de nombreux pseudogènes ont également été mis en évidence. Une analyse des chromosomes 21 et 22 chez l'Homme a déterminé que la proportion de pseudogènes pour l'ensemble des gènes de ces chromosomes était d'un pseudogène pour deux gènes fonctionnels (Harrison *et al.*, 2002).

De plus, pour la grande famille des récepteurs olfactifs qui comptent environ 1000 membres par espèce chez les mammifères (906 chez l'Homme), il a été défini qu'au moins 63% ont dégénéré au cours du temps en pseudogènes alors que chez la souris, seuls 20% des gènes ont subi ce processus (Young et Trask, 2002).

1.3 Les remaniements chromosomiques et les maladies génétiques chez l'Homme

Dans certains cas, les remaniements chromosomiques peuvent également présenter des effets délétères chez les organismes pluricellulaires et notamment chez l'Homme. Certaines maladies génétiques peuvent être dues à des duplications, des délétions ou des aneuploïdies. Par exemple, chez les patients atteints du syndrome de Down, une copie supplémentaire du chromosome 21 est présente dans le génome. Le syndrome de Turner

correspond, lui, à une monosomie au niveau de la paire de chromosomes sexuels. La femme atteinte de cette maladie ne possède alors qu'un seul chromosome X.

Des remaniements chromosomiques ont été décrits comme responsables d'oncogenèse (Difilippantonio *et al.*, 2002) ou conduisant à une susceptibilité pour certains cancers chez les individus porteurs de ces modifications (Sharp *et al.*, 2006). Il a par exemple été décrit qu'une délétion des gènes codant pour des glutathion S-transférases est associée avec un risque de cancer de la vessie. Dans le sarcome d'Ewing, une translocation est à l'origine de la fusion des gènes *EWS* et *FLII* ce qui induit une tumeur osseuse maligne chez les patients (Obata *et al.*, 1999). Enfin, un événement de translocation conduisant à la fusion des gènes *PML* et *RAR α* est responsable d'une des formes de leucémie promyélocytaire aigüe (Koken *et al.* 1997).

2. Apparition des remaniements chromosomiques

Les remaniements chromosomiques sont généralement induits suite à une cassure au niveau de l'ADN génomique. Ces cassures d'ADN et par conséquent les modifications génomiques peuvent apparaître de différentes manières au sein d'une cellule. Elles peuvent apparaître de manière spontanée ou être induites par des agents exogènes à la cellule.

Des cassures ou des dommages au niveau de l'ADN conduisant à l'apparition de remaniements chromosomiques peuvent se produire suite au métabolisme cellulaire, par exemple lors de la production de radicaux libres. Dans certains cas, les cassures peuvent apparaître lors de la réplication ou la transcription de l'ADN. Une pause ou le blocage de la fourche de réplication peut être responsable de l'accumulation de mutations ou d'aberrations chromosomiques (Branzei et Foiani, 2005). Cet événement peut se produire dans le cas de structures secondaires au niveau de l'ADN, comme dans le cas des gènes d'ARN de transfert ou d'ARN ribosomiques, au niveau de régions répétées de l'ADN ou encore au niveau de « barrières » de la fourche de réplication RFB (replication fork barriers) (Zhang *et al.*, 2006).

Des agents exogènes physiques ou chimiques peuvent également être impliqués dans l'apparition des remaniements (pour revue : Kupiec, 2000). Par exemple, les radiations ionisantes telles que les rayons X induisent des lésions au niveau de l'ADN notamment des cassures simple et double-brins. Dans le cas des radiations aux UV, les lésions formées sont généralement des dimères de pyrimidine. Ce type de lésions est réparé par le système de

réparation NER (Nucleotide Excision Repair) (Singh et Krishna, 2005). Ce mécanisme intervient pour réparer les bases ayant subi des dommages dans l'ADN. Néanmoins, il arrive que cette réparation entraîne par la suite une cassure d'ADN. Si cette cassure n'est pas réparée correctement, il peut alors y avoir apparition de différents types de réarrangements (di Caprio et Cox, 1981).

Certains agents chimiques comme le MMS (methyl methanesulfonate) sont également recombino-gènes. Les lésions causées par le MMS, notamment la méthylation de l'adénine, peuvent entraîner une cassure de l'ADN et ensuite conduire à l'arrêt de la réplication.

Enfin, certains agents comme le cisplatine vont créer des liens inter-brins en interagissant avec deux guanines au niveau de l'ADN. Ce pont va empêcher la séparation des brins d'ADN et ainsi bloquer la réplication et la transcription (Zdraveski *et al.*, 2000). Il est à noter que le cisplatine est utilisée en chimiothérapie. En effet, cet agent affecte préférentiellement la réplication et la transcription des cellules à croissance rapide comme les cellules cancéreuses (Skov et MacPhail, 1991).

3. *Saccharomyces cerevisiae*, un bon modèle d'étude pour les remaniements chromosomiques

La levure *Saccharomyces cerevisiae* est un outil particulièrement adapté à l'étude des remaniements chromosomiques. Cet eucaryote possède un cycle haplodiplobiontique et les souches utilisées en laboratoire sont hétérothalliques. Ainsi les analyses pourront être menées dans ces contextes stables haploïdes ou diploïdes.

Son génome est compact et de petite taille (12 Mb). La séquence complète de la souche S288c a été obtenue en 1996 (Goffeau *et al.*, 1996). Il s'agissait alors du premier génome complet eucaryote séquencé. L'identification des ORFs est exhaustive chez *Saccharomyces cerevisiae* car plusieurs phases de modifications et de corrections ont succédé à la phase d'annotation initiale du génome.

D'un point de vue expérimental, la levure est facilement cultivable et de nombreux outils génétiques ont été développés pour cet organisme. Il est ainsi possible de construire des souches mutées rapidement et aisément par des techniques de microbiologie classique et de biologie moléculaire. Les techniques de transformation notamment ont été développées dès la fin des années 70 pour cet organisme (Hinnen *et al.*, 1978). La levure a également été utilisée pour la cartographie génétique de nombreux gènes dans les années 60-70 (Hawthorne et Mortimer, 1960).

Ainsi, une collection de délétions systématiques de tous les gènes de *Saccharomyces 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 *KANMX4* entraînant une résistance à la drogue G418. Dans le cas des gènes essentiels, des souches diploïdes hétérozygotes présentant un allèle délété ont été construites. Il s'agit de la collection EUROSCARF (Brachmann *et al.*, 1998).

L'utilisation de cette collection nous a permis d'étudier l'apparition des remaniements chromosomiques dans différents contextes mutés afin de déterminer les gènes ayant un impact sur la formation des événements. En effet, des souches dont les gènes codant pour des protéines impliquées dans les différentes voies de recombinaison ont été construites par croisement de la souche porteuse du marqueur de sélection utilisé au laboratoire avec les souches porteuses des gènes d'intérêt délétés.

De nombreuses souches d'origines différentes ont été caractérisées. Il peut s'agir de souches de laboratoire, naturelles, isolées à partir de fromages, de pâtes à pain, ou œnologiques, trouvées au niveau de grappes de raisin ou de chais.

A l'heure actuelle, trois souches ont été séquencées. Il s'agit de la souche de laboratoire de référence S288c (Goffeau *et al.*, 1996) et d'une souche œnologique RM11 (Broad Institute). La troisième souche séquencée, YJM789, est une souche opportuniste prélevée chez un patient immunodéprimé atteint du SIDA (Wei *et al.*, 2007). De plus, les séquences partielles pour d'autres souches de *S. cerevisiae* sont disponibles (SGRP Sanger Institute). L'ensemble de ces données a permis de réaliser des comparaisons intra-espèces.

Enfin, d'autres espèces appartenant comme *S. cerevisiae* au groupe des hémiascomycètes ont été séquencées. Les données obtenues servent de base pour les analyses de génomique comparative inter-espèces et l'étude de l'évolution du groupe des hémiascomycètes. Ces analyses sont majoritairement réalisées *in silico* car les expériences de génétique n'ont pas encore été développées pour toutes ces espèces.

4. Les mécanismes de recombinaison

Tous les organismes présentent des mécanismes de recombinaison qui leur permettent d'assurer une diversité génétique mais également de maintenir l'intégrité de leur génome lors d'une cassure d'ADN. Ils sont également à l'origine du redémarrage de la fourche de réplication (Rothstein *et al.*, 2000 ; Branzei et Foiani, 2005). Ces mécanismes peuvent être

séparés en deux catégories. La première comporte les mécanismes associés à la recombinaison homologue. La deuxième correspond aux mécanismes de recombinaison non-homologue dont le principal est celui de Non-Homologous End Joining (NHEJ).

4.1 Les mécanismes de recombinaison homologue

4.1.1 Principe général

Chez les eucaryotes, la recombinaison homologue est un processus universel trouvé dans tous les organismes. Elle joue un rôle important en générant une diversité génétique et en assurant une ségrégation correcte des chromosomes durant la méiose. Dans les cellules en mitose, la fonction principale de la recombinaison homologue est de réparer les cassures d'ADN double-brin qui peuvent se former de manière spontanée ou suite à l'action d'agents endogènes (pour revues : Pâques et Haber, 1999 ; Symington, 2002 ; Aylon et Kupiec, 2004). La recombinaison homologue intervient également dans le redémarrage des fourches de réplication en pause. Ce phénomène a notamment été mis en évidence chez *Escherichia coli* (Cox *et al.*, 2000). Elle est également nécessaire pour initier les réarrangements programmés comme le processus de changement de signe sexuel chez les souches de *S. cerevisiae* homothalliques.

Le mécanisme repose sur l'échange d'ADN entre des régions homologues d'au moins 60 pb (Sugawara et Haber, 1992). Lors d'une cassure d'ADN, il a été montré que le mécanisme de Non-Homologous End Joining (NHEJ) précéderait le mécanisme de recombinaison homologue et serait très efficace pour réparer l'ADN. Néanmoins, le NHEJ agirait préférentiellement lors de la phase G1 du cycle cellulaire puis en phases S et G2, la recombinaison homologue prendrait le dessus (Frank-Vaillant et Marcand, 2002).

4.1.2 Les systèmes pour étudier les mécanismes de recombinaison homologue

La recombinaison homologue peut être induite lors d'une cassure d'ADN double-brin. Plusieurs systèmes pour sélectionner des événements de réparation par recombinaison homologue ont utilisés cette observation. Dans les cellules en mitose, la recombinaison peut être initiée par induction d'une endonucléase site-spécifique. Deux systèmes développés chez la levure seront décrits ici.

Le premier est basé sur l'activité de l'endonucléase HO. Cette enzyme reconnaît une séquence de 22 pb et présente uniquement un site de coupure dans une cellule sauvage localisé au locus *MAT* impliqué dans le signe sexuel. Des souches ont été construites pour lesquelles le gène *HO* est sous le contrôle du promoteur inductible *GAL1* (Inbar et Kupiec,

1999). Lorsque les cellules sont placées sur un milieu contenant du galactose comme source de carbone, l'endonucléase HO est produite, coupe au niveau du site spécifique de coupure et engendre ainsi une cassure d'ADN double-brin. Pour que la souche survive, le dommage doit être réparé.

Le second système utilisé repose sur l'enzyme I-SceI. Cette endonucléase est normalement codée et exprimée dans les mitochondries chez la levure où elle permet la mobilité des introns au niveau de l'ADN mitochondrial. Une version synthétique de ce gène a été construite et a ensuite été placée sous contrôle du promoteur inductible *GALI*. Sur milieu contenant du galactose, le gène est activé et l'endonucléase est produite. Ceci conduit à une cassure d'ADN double-brin qui doit être réparée pour que les cellules survivent (Storici *et al.*, 2003).

Les résultats obtenus dans ces deux systèmes démontrent que la recombinaison homologue est impliquée dans la réparation de l'ADN suite à une cassure d'ADN double-brin.

4.1.3 Les différents mécanismes de recombinaison homologue

4.1.3.1 Le modèle de réparation des cassures d'ADN double brin (DSBR)

La conversion génique est définie comme un transfert d'information génétique non-réciproque entre deux molécules présentant un pourcentage élevé d'identité. Généralement ce phénomène se produit entre deux chromatides sœurs (Kadyk et Hartwell, 1992). La conversion génique a tout d'abord été mise en évidence en méiose par l'observation de la ségrégation non-mendélienne de certains allèles dans les asques de plusieurs espèces d'ascomycètes et en particulier *S. cerevisiae*. Elle a par la suite également été observée pour les cellules en mitose. Des travaux ont ensuite montré que la conversion génique pouvait être associée à des crossing-over (Fogel et Hurst, 1967). Afin d'expliquer ce phénomène, plusieurs modèles ont été développés dont notamment le modèle de Holliday (1964). Finalement, Szostak et ses collaborateurs développèrent le modèle de réparation des cassures d'ADN double-brin (DSBR) (Szostak *et al.*, 1983). Ce modèle a ensuite été confirmé par les travaux de Nicolas *et al.* (1989). Il est présenté dans la Figure I6.

Suite à une cassure d'ADN double-brin, les extrémités 5' de l'ADN sont digérées sur une distance pouvant s'étendre sur plus d'un kb. Il y a alors invasion des extrémités 3' de la cassure d'ADN sur les brins intacts. Ces extrémités serviront alors d'amorces pour l'initiation de la synthèse de nouveaux brins d'ADN en utilisant comme matrice la région intacte d'ADN. Ce processus va conduire à la formation de deux jonctions

de Holliday. Si les deux jonctions de Holliday sont clivées de la même manière, une conversion génique simple est observée. En revanche, si une des jonctions est clivée au niveau des brins croisés et l'autre au niveau des brins non-croisés, il y a alors crossing-over. Ces deux jonctions peuvent être coupées par des résolvases. Ces endonucléases ont été mises en évidence chez les bactéries (RuvC et RusA) (Mahdi *et al.*, 1996), chez les archaebactéries (Hjc et Hje) (Middleton *et al.*, 2004) ainsi que chez les bactériophages (T4 endonucléase VII) (Mueller *et al.*, 1988). Chez les eucaryotes, aucune résolvasse n'a été trouvée jusqu'à présent. Néanmoins, il a été proposé que le complexe Emel-Mus81 chez *Schizosaccharomyces pombe* (Mms4p-Mus81p chez *S. cerevisiae*) pourrait jouer un rôle dans le clivage des jonctions de Holliday (Boddy *et al.*, 2001). Des études *in vitro* ont cependant montré que le complexe Mus81p-Mms4, seul, ne peut pas réaliser ce clivage ce qui suggérerait la présence d'un coactivateur (Ehmsen et Heyer, 2008).

4.1.3.2 Le mécanisme de synthesis-dependent strand annealing (SDSA)

Ce modèle a été établi pour répondre à l'observation qui indiquait qu'en mitose peu de conversions géniques étaient associées à des crossing-over (Kostriken *et al.*, 1983 ; Nassif *et al.*, 1994). La première partie de ce mécanisme est la même que pour le modèle de DSBR (Figure I7A). L'ADN subit une cassure double-brin et les extrémités 5' sont alors dégradées. Seule une des extrémités 3' utilise comme matrice le brin intact de la région homologue. L'autre extrémité 3' au niveau de la cassure va utiliser l'ADN néosynthétisé comme matrice plutôt que le brin intact complémentaire. Ce type de processus résulte généralement en une conversion génique sans crossing-over.

4.1.3.3 Le mécanisme de break-induced replication (BIR)

Dans certains cas, la réparation se produit par invasion du chromosome matrice par un seul des brins issus de la cassure d'ADN double-brin (Figure I7B) (Smith *et al.*, 2007 ; Llorente *et al.*, 2008). Le brin peut alors être copié jusqu'au télomère. Le chromosome ayant subi la cassure peut perdre la région allant de la cassure jusqu'à l'extrémité du chromosome et possède en échange le segment d'ADN qu'il a recopié à partir du brin matrice. Ce mécanisme a été décrit lors du maintien des télomères dans les cellules pour lesquelles les gènes codant pour l'activité télomérase sont délétés (Lundblad et Blackburn, 1993).

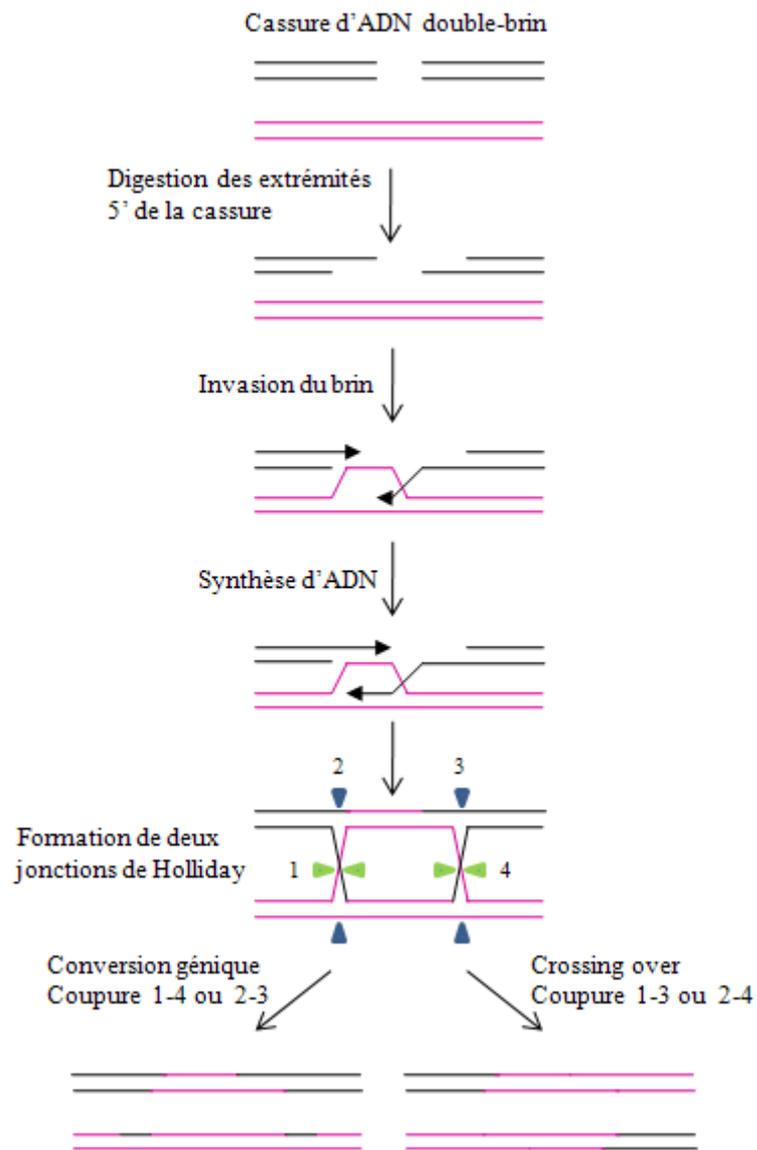


Figure I6 : Modèle de réparation des cassures d'ADN double-brin (DSBR) par Szostak (d'après Symington, 2002).

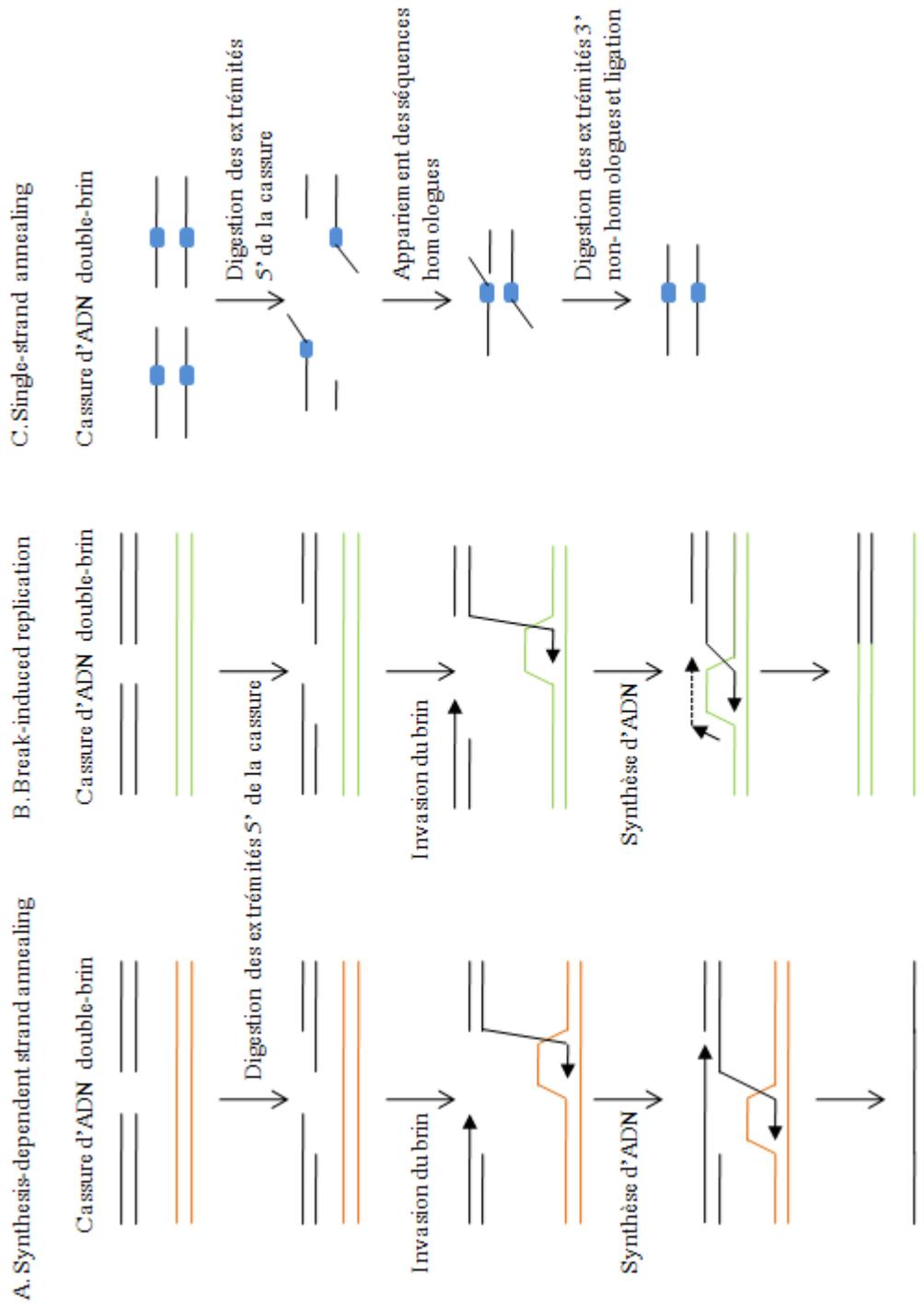


Figure 17 : Les différents mécanismes de recombinaison homologue (d'après Symington, 2002).

4.1.3.4 Le mécanisme de single-strand annealing (SSA)

Lors d'une cassure d'ADN double-brin, les extrémités 5' des deux brins d'ADN sont dégradées par une exonucléase ce qui conduit à la production de longues queues d'ADN simple-brin (Figure I7C). S'il y a présence de séquences complémentaires homologues alors le réappariement des deux brins peut avoir lieu à ce niveau. Il y a ensuite dégradation des régions 3' non-appariées et ligation. Ce mécanisme conduit à une délétion de la région contenue entre les deux séquences homologues (Maryon et Carrol, 1991 ; Fishman-Lobell *et al.*, 1992 ; Daley et Wilson, 2005).

4.1.4 Les principales protéines impliquées dans la recombinaison homologue

Plusieurs gènes impliqués dans la recombinaison homologue ont été identifiés pour des mutants de *Saccharomyces cerevisiae* sensibles aux radiations. La plupart de ces gènes ont d'ailleurs été nommés d'après cette propriété (rad pour radiation sensitive). Il s'agit des gènes *RAD50*, *RAD51*, *RAD52*, *RAD54*, *RAD55*, *RAD57*, *XRS2* et *MRE11* dont les mutants sont sensibles aux rayons X mais pas aux UV (Game et Mortimer, 1974). L'ensemble de ces gènes correspond au groupe d'épistasie *RAD52*. Tous les membres de ce groupe sont extrêmement bien conservés chez les eucaryotes. Par exemple la protéine codée par le gène *RAD51* chez *S. cerevisiae* présente 59% d'identité avec ses homologues chez l'Homme ou la souris.

- la protéine Rad51p

La protéine Rad51p est une des protéines les plus importantes pour la recombinaison homologue. Il s'agit de l'homologue de RecA chez les bactéries. Elle recouvre l'ADN simple-brin formé par dégradation des extrémités 5'. Elle forme alors un nucléofilament qui va permettre la réaction d'invasion du brin. Les mécanismes de SDSA et BIR nécessite la présence de cette protéine. Davis et Symington (2004) ont montré que Rad51p est nécessaire dans 95% des événements de BIR. D'autres études ont montré que dans certains cas, le mécanisme de BIR pourrait également se produire en absence du filament présynaptique (Malkova *et al.*, 2001). Le mécanisme de SSA, quant à lui, ne nécessite pas la présence de cette protéine. Dans le cas du SSA, il n'y pas d'invasion du brin au niveau du chromosome matrice, ce qui pourrait expliquer que ce mécanisme puisse se réaliser en absence de Rad51p.

- la protéine Rad52p

Une mutation dans le gène *RAD52* conduit à un défaut de fonctionnement de l'ensemble des mécanismes de recombinaison homologue. Son rôle est donc essentiel pour que les mécanismes puissent agir efficacement. La protéine Rad52p peut se fixer à l'ADN simple-brin sous forme de multimères (Shinohara *et al.*, 1998 ; Stasiak *et al.*, 2000). Il a été montré *in vitro* qu'elle accélère l'appariement de l'ADN simple brin sur lequel elle est fixée avec le brin d'ADN complémentaire. Elle peut également interagir avec la protéine Rad51p, elle agit alors comme médiateur pour permettre la fixation de Rad51p au niveau de l'ADN simple-brin en détachant les protéines RPA (replication factor A) qui l'entourent (Sung *et al.*, 2003). La fixation du complexe Rad52p-Rad51p servirait alors de sites de nucléation pour l'assemblage du filament présynaptique (Song et Sung, 2000). Rad52p intervient également dans les mécanismes indépendant de Rad51p (Symington, 2002).

- la protéine Rad59p

La protéine codée par le gène *RAD59* possède une similarité de séquence avec la région N-terminale de la protéine Rad52p. Elle participe à l'appariement des régions d'ADN simple-brin (Bai et Symington, 1996). Elle est notamment impliquée dans le mécanisme de SSA (Storici *et al.*, 2006).

- Les protéines Rad55p et Rad57p

Les produits des gènes *RAD55* et *RAD57* forment un hétérodimère qui interagit avec les séquences à proximité de la cassure d'ADN. *In vivo*, leur fixation nécessite la présence des protéines Rad51p et Rad52p. Cet hétérodimère intervient deux fois au cours de la recombinaison homologue. La première fois, il interviendrait lors de l'invasion du brin d'ADN au niveau du brin intact. La deuxième fois, ce complexe serait nécessaire pour permettre à la conversion génique d'avoir lieu (Aylon et Kupiec, 2004). Ce complexe pourrait intervenir dans la stabilisation de l'interaction de Rad51p avec l'ADN simple-brin. Il pourrait également jouer un rôle dans la stabilisation du complexe présynaptique (Johnson et Symington, 1995).

- La protéine Rad54p

La protéine Rad54p est un membre de la famille Swi2/Snf2 dont les protéines sont impliquées dans le remodelage de la chromatine. Il s'agit d'une ATPase qui interagit directement avec la protéine Rad51p. Elle est nécessaire pour la réaction d'invasion du brin

intact par le brin issu de la cassure d'ADN en créant un superenroulement négatif de l'ADN conduisant à l'ouverture transitoire des deux brins d'ADN. Rad54p peut ainsi promouvoir la formation de jonctions paranémiques au niveau de l'ADN (van Komen *et al.*, 2000). Rad54p permettrait aussi au filament présynaptique de scanner activement le duplex d'ADN à la recherche d'une région homologue (Sung *et al.*, 2003).

- le complexe MRX

Les protéines Mre11p, Rad50p et Xrs2p s'assemblent pour former le complexe MRX. Celui-ci n'est pas spécifique de la recombinaison homologue puisqu'il intervient aussi dans le mécanisme de NHEJ et dans le maintien des télomères. Le complexe possède une activité endonucléase de l'ADN simple-brin et une activité exonucléase 3'-5' (Paull et Gellert, 1998). Ce complexe interviendrait dans le traitement des extrémités 5' après cassure pour faciliter l'étape d'invasion du brin pour les mécanismes de SDSA et BIR et l'appariement des séquences homologues dans le mécanisme de SSA (Krishna *et al.*, 2007). La génération d'ADN simple-brin aux extrémités des cassures d'ADN double-brin nécessite une activité 5'-3' exonucléase. Le complexe MRX serait responsable d'une partie de cette activité puisqu'une délétion de ces protéines conduit à une diminution de la résection des extrémités *in vivo*. Il agirait alors avec l'exonucléase Exo1p (Nakada *et al.*, 2004).

Ce complexe intervient également dans le mécanisme de « checkpoint ». Il se fixerait au niveau de la cassure d'ADN permettant ainsi le recrutement de Tel1p, une kinase intervenant dans la phosphorylation des histones H2A (Nakada *et al.*, 2003, pour revue Harrison et Haber, 2006).

4.2 Le Non-Homologous End Joining (NHEJ)

4.2.1 Principe général

Le mécanisme repose sur la réparation de cassures d'ADN double-brin en absence de séquence homologue. Il agit en l'absence d'identité de séquences ou en présence de courtes séquences homologues de taille inférieure à 10 pb (pour revue : Lewis and Resnick, 2000 ; Dudášová *et al.*, 2004 ; Daley *et al.*, 2005). La réparation de la cassure peut être précise ou conduire à des délétions (jusqu'à 13 kb) (Yu et Gabriel, 2003).

L'importance du NHEJ dans la réparation de l'ADN n'a été mise en évidence que tardivement chez la levure par rapport aux mécanismes de recombinaison homologue. Ceci est dû à la prévalence de la recombinaison homologue chez *S. cerevisiae*. Le terme de NHEJ ne fut déterminé qu'en 1996 quand Moore et Haber l'utilisèrent pour décrire la

réparation de cassure d'ADN double-brin s'étant produite chez la levure suite à une coupure de l'ADN par l'endonucléase HO en absence de séquence homologue (Moore et Haber, 1996). Il intervient généralement dans les souches pour lesquelles la recombinaison homologue est non-fonctionnelle suite à l'inactivation de *RAD52* et dans les situations où il y a absence de régions homologues.

4.2.2 Les systèmes pour étudier le mécanisme de NHEJ

Le mécanisme de NHEJ ne nécessite que peu ou pas de séquences homologues pour pouvoir réparer la cassure d'ADN double-brin. Les différentes protéines impliquées dans le NHEJ ont été déterminées en utilisant des systèmes expérimentaux basés sur des plasmides linéarisés au niveau de régions ne présentant pas de séquence homologue dans le génome. D'autres systèmes utilisés reposent sur la réparation de l'ADN après cassure suite à la coupure d'une enzyme de restriction.

Plusieurs systèmes basés sur la recircularisation de plasmides ont été développés. L'un d'eux repose sur l'utilisation d'un plasmide navette entre *S. cerevisiae* et *Escherichia coli* porteur d'un gène marqueur de sélection (Boulton et Jackson, 1996). Le plasmide est linéarisé par coupure par des enzymes de restriction au niveau d'une région ne présentant pas de séquence homologue dans le génome de la levure. Il est ensuite introduit dans une levure présentant une auxotrophie pour le gène marqueur porté par le plasmide. Les cellules ayant poussé sur milieu sélectif sont celles pour lesquels le plasmide a été recircularisé et qui exprime donc le gène marqueur. Les études ont montré que le mécanisme de NHEJ est à l'origine de la recircularisation.

Ce système a ensuite été adapté par Daley et Wilson (2005). Des oligonucléotides de tailles variables ont été ajoutés aux bornes de plasmides linéarisés créant ainsi des microhomologies. Ceci a permis d'étudier l'importance de la taille de ces séquences pour la mise en place du mécanisme de NHEJ. Il a ainsi été montré que des microhomologies d'une taille inférieure à 10 pb étaient suffisantes pour la réparation de la cassure. De plus, il a été observé que la réparation pouvait aussi se produire pour des extrémités franches.

4.2.3 Le mécanisme de NHEJ

Le mécanisme est présenté dans la Figure I8. Bien que les protéines intervenant dans le mécanisme aient été identifiées pour ce mécanisme, l'ordre d'action de ces différents intervenants n'est pas encore clairement identifié. Le modèle présenté est celui défini par la revue d'Aylon et Kupiec (2004).

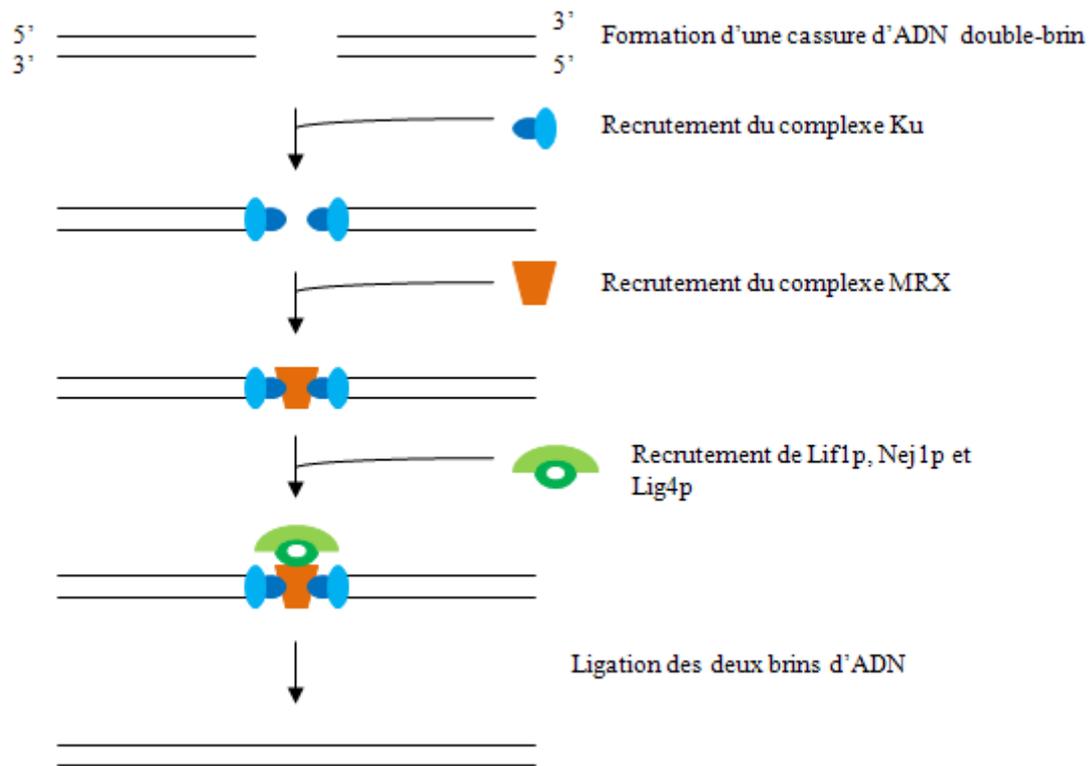


Figure 18 : Description du mécanisme de non-homologous end joining (d'après Aylon et Kupiec, 2004).

Suite à une cassure d'ADN double-brin, le complexe Ku est recruté au niveau des régions d'ADN double-brin aux bornes de la cassure. Le complexe MRX vient alors se fixer. Les deux complexes vont créer un pont entre les deux extrémités d'ADN. Le complexe Lig4p, Nej1p, Lif1p est ensuite recruté à son tour ce qui conduit à la ligation des brins d'ADN et à la réparation de la cassure.

4.2.4 Les principales protéines intervenant dans le mécanisme de NHEJ

Comme nous venons de le voir, différentes protéines interviennent dans le mécanisme de NHEJ. Nous ne présenterons ici que celles qui sont strictement nécessaires pour la réparation de la cassure d'ADN.

Le complexe Ku est composé d'un hétérodimère de deux protéines Yku80p et Yku70p chez la levure. Ces deux gènes proviennent d'une duplication d'un gène ancestral indépendante de l'événement de duplication totale du génome (Doherty *et al.*, 2001). Il est normalement localisé au niveau des télomères où il assure la protection de leurs extrémités mais se déplace lors d'une cassure d'ADN pour venir se placer aux bornes de la cassure (Martin *et al.*, 1999 ; Bertuch et Lundblad, 2003). Des hypothèses basées sur des études menées chez la levure et chez les eucaryotes supérieurs ont été émises quand à la fonction

exacte de ce complexe. Il pourrait protéger les deux extrémités d'ADN, issues de la cassure, des nucléases cellulaires. Il a également été supposé que le complexe Ku créerait un pont entre les deux extrémités facilitant ainsi l'étape de ligation. Enfin, il permettrait le recrutement d'endonucléases et/ou d'exonucléases au niveau des sites de cassures. Ceci entraînerait le traitement des extrémités afin de faciliter par la suite l'étape de ligation. Un mécanisme de NHEJ en absence du complexe Ku a été décrit (Moore et Haber, 1996 ; Yu et Gabriel, 2003). Dans ce cas, la réparation n'est pas précise et est souvent accompagnée de délétions.

Le complexe MRX correspond à l'association des trois protéines Mre11p, Rad50p et Xrs2p. Comme nous l'avons vu plus haut, ce complexe intervient également dans le mécanisme de recombinaison homologue. Il interviendrait dans le traitement des extrémités pour faciliter l'étape de ligation (Nakada *et al.*, 2004). Avec le complexe Ku, il pourrait aussi intervenir dans la création d'un pont entre les deux extrémités d'ADN issues de la cassure.

L'étape essentielle de ce mécanisme est la ligation des brins d'ADN aux extrémités de la cassure. Celle-ci est réalisée par la protéine codée par le gène *LIG4* (Teo *et al.*, 1997). Cette ligase n'intervient pas dans la réplication mais uniquement dans la réparation de l'ADN (Wilson *et al.*, 1997). Elle forme un complexe avec ses deux cofacteurs Lif1p et Nej1p. La protéine codée par le gène *LIF1* fonctionne comme un adaptateur entre le complexe Ku et la ligase (Herrmann *et al.*, 1998). Le rôle de la protéine Nej1p est moins bien défini. Elle pourrait intervenir dans l'import de la ligase au niveau du noyau (Valencia *et al.*, 2001). Cependant, des études structurales ont montré que la protéine Lig4p possède un signal de localisation nucléaire (Daley *et al.*, 2005). D'autres études ont nommé cette protéine Lif2p et ont suggéré que Lif2p pourrait être impliquée dans le recrutement de Lig4p-Lif1p au niveau de l'ADN ou pourrait altérer la structure de la chromatine pour permettre l'accès à l'ADN (Frank-Vaillant et Marcand, 2001).

4.3 Le Microhomology-Mediated End Joining (MMEJ)

Il y a une dizaine d'années, un troisième mécanisme a été ajouté aux mécanismes de NHEJ et de recombinaison homologue pour réparer les cassures d'ADN double-brin. Ce mécanisme a été défini pour de cellules dans lesquelles le gène codant pour *YKU80* a été délété (Boulton et Jackson, 1996). Il a ensuite été montré qu'il était également indépendant du gène *RAD52* impliqué dans la recombinaison homologue (Ma *et al.*, 2003). Ce mécanisme est mutagène et semblerait contribuer à des instabilités génomiques trouvées dans certains cancers (Zhong *et al.*, 2002). Il implique un appariement entre microhomologies conduisant généralement à la délétion de la séquence contenue entre celles-ci (Ma *et al.*,

2003). Malgré cet impact sur la stabilité des génomes, les constituants de ce mécanisme sont encore peu connus. Chez *S. cerevisiae*, il a été montré que le MMEJ est indépendant de Yku80p et de Rad52p. En revanche, il dépend de Rad1p, endonucléase impliquée dans le traitement des extrémités simple-brin 3' et du complexe MRX (Ma *et al.*, 2003). Lee et Lee (2007) ont également trouvé que Nej1p impliqué dans le NHEJ serait également nécessaire au MMEJ ainsi que différentes polymérases, Pol4p, Rad30p, Rad3p et Pol32p. Enfin, ils ont aussi démontré que le MMEJ est favorisé en présence de Sae2p et Tel1p alors que ces protéines ont un effet de contrôle négatif sur le NHEJ. Chez *Schizosaccharomyces pombe*, Decottignies (2007) a montré que le mécanisme de MMEJ est différent du NHEJ car il est indépendant de la ligase Lig4p. Contrairement à ce qui est observé chez *S. cerevisiae*, ce mécanisme serait dépendant de Rad22p (homologue de Rad52p chez *S. pombe*). Il nécessite également Exo1, exonucléase impliquée dans la dégradation 5'-3' des extrémités d'ADN simple-brin ainsi que la polymérase Pol4p. Chez l'Homme, le MMEJ nécessiterait la présence des ligases I et III mais pas celle de la ligase IV impliquée dans le NHEJ (Liang *et al.*, 2008). Ainsi le MMEJ est mécaniquement similaire au NHEJ et au SSA. De plus il dépendrait de produits de gènes impliqués dans ces deux voies de recombinaison. Des études supplémentaires restent donc à être réalisées pour déterminer précisément si le MMEJ est une voie de sauvetage lorsque la cellule est déficiente pour la recombinaison homologe et le NHEJ ou s'il s'agit d'une variante d'un de ces deux mécanismes.

Les remaniements chromosomiques peuvent jouer un rôle important au niveau de la cellule. Ils permettent l'adaptation des organismes à leur environnement, interviennent dans les mécanismes évolutifs et peuvent dans certains cas, avoir des conséquences plus délétères chez les organismes pluricellulaires. Plusieurs travaux sur les mécanismes d'apparition des remaniements chromosomiques ont été réalisés. Ils présentent souvent l'inconvénient d'utiliser des agents mutagènes pour induire les réarrangements ou des plasmides porteurs de gènes marqueurs qui ne reflètent pas nécessairement ce qui se passe au niveau du génome. Ainsi, les mécanismes d'apparition et les gènes impliqués dans leur formation spontanée sont encore mal connus à l'heure actuelle.

Afin de les étudier, il est donc nécessaire d'élaborer des cribles génétiques permettant la sélection de remaniements chromosomiques spontanés *in vivo*. Au laboratoire de Génétique Moléculaire, Génomique et Microbiologie, un crible de sélection a été développé (Exinger et Lacroute, 1979 ; Roelants *et al.*, 1995). Il repose sur un allèle particulier du gène *URA2* et permet la sélection de remaniements chromosomiques spontanés chez la levure

Saccharomyces cerevisiae. Il sera décrit plus en détail dans le chapitre 1. Ce système permet la sélection de révertants porteurs de délétions, de duplications ou d'insertions de rétrotransposons Ty1.

Au cours de ce travail, nous avons utilisé le crible de sélection *ura2*_{15,30,72} pour étudier l'impact de la variabilité génomique intra-espèce sur l'obtention des remaniements chromosomiques et pour déterminer les mécanismes impliqués dans leur formation. La stratégie utilisée a reposé sur la sélection de remaniements chromosomiques dans différents contextes génétiques.

Pour tester l'influence des variations génomiques intra-espèces sur l'apparition des remaniements chromosomiques, nous avons comparé les réarrangements sélectionnés dans deux souches différentes à savoir FL100 utilisée au laboratoire et S288c, souche de référence séquencée. Les résultats seront détaillés dans le chapitre 2 de la partie I.

Dans un deuxième temps, nous nous sommes intéressés aux différents mécanismes permettant l'apparition des remaniements chromosomiques. Pour ce faire, des souches *ura2*_{15,30,72} délétées ont été construites pour des gènes dont les produits sont impliqués dans la recombinaison homologue (*RAD52* et *RAD59*) et dans le NHEJ (*LIG4* et *YKU80*). Il a ainsi été possible de comparer ces différents contextes quant aux fréquences d'apparition et aux types de remaniements obtenus. Nous avons pu déterminer quels gènes étaient importants pour la formation des remaniements et suggérer le mécanisme impliqué. Cette partie sera présentée dans le chapitre 3 de la partie I.

Enfin, dans le cadre du projet Génolevures 3, nous avons participé à l'annotation de quatre génomes d'espèces appartenant au groupe des *Kluyveromyces*. A partir des données obtenues par le consortium, nous nous sommes focalisés sur l'étude de la variabilité inter-espèces pour deux familles de gènes dupliqués codant pour des perméases en effectuant une analyse de génomique comparative inter-espèces. Ce travail sera détaillé dans la partie II.

MATÉRIEL ET MÉTHODES

1. Organismes et milieux de culture

1.1 *Saccharomyces cerevisiae*

1.1.1 Biologie de *Saccharomyces cerevisiae*

Saccharomyces cerevisiae est un champignon hémiascomycète hétérothallique. Son cycle cellulaire est haplodiplobiontique c'est-à-dire qu'il peut se maintenir sous forme haploïde et diploïde par multiplication végétative (bourgeoisement) (Figure MM1). En conditions défavorables (carence en carbone ou en azote), les cellules diploïdes vont sporuler et donner naissance à un asque contenant quatre spores non ordonnées. Dans des conditions plus favorables, les spores vont germer pour donner deux cellules haploïdes de signe a et deux cellules haploïdes de signe α . Ces dernières, si elles sont de signes opposés, vont pouvoir fusionner pour générer une cellule diploïde. Ces étapes peuvent être reproduites au laboratoire.

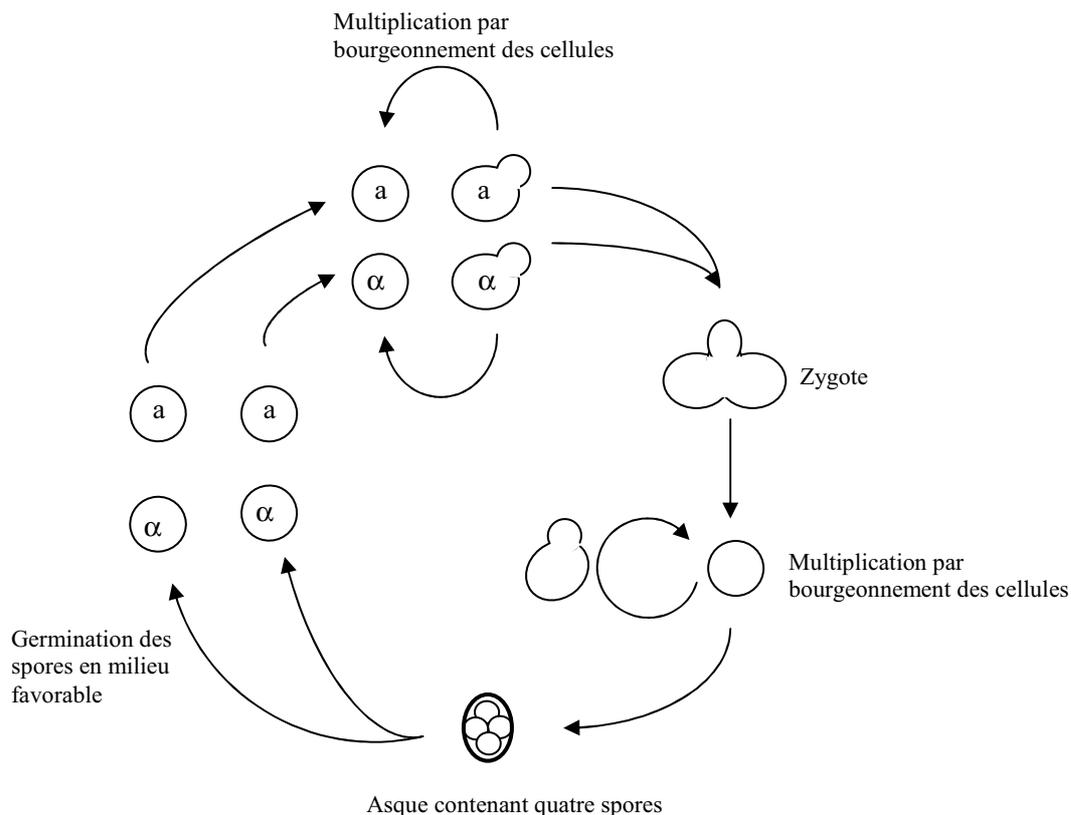


Figure MM1 : Cycle cellulaire de *Saccharomyces cerevisiae*

1.1.2 Souches utilisées

Les caractéristiques des différentes souches utilisées ou construites lors de ce travail sont répertoriées dans le tableau MM1.

Souches	Génotype	Phénotype
FL100 AW1	<i>MATa ura2_{15,30,72} trp1Δ63</i>	Ura ⁻ Trp ⁻
aUHT S288c	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63</i>	Ura ⁻ His ⁻ Trp ⁻
aFL	<i>MATa ura2_{15,30,72} his3Δ200 lys2Δ0 lig4::KANMX4</i>	Ura ⁻ His ⁻ Lys ⁻ G418 ^R
aFY	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63 yku80::KANMX4</i>	Ura ⁻ His ⁻ Trp ⁻ G418 ^R
aFR	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63 rad59::KANMX4</i>	Ura ⁻ His ⁻ Trp ⁻ G418 ^R
aΔR	<i>MATa ura2_{15,30,72} his3Δ200 lys2Δ0 rad52::KANMX4</i>	Ura ⁻ His ⁻ Lys ⁻ G418 ^R
aLR	<i>MATa ura2_{15,30,72} his3Δ200 leu2Δ0 lys2Δ0 lig4::KANMX4 rad52::KANMX4</i>	Ura ⁻ His ⁻ Leu ⁻ Lys ⁻ G418 ^R

Tableau MM1 : Souches de *Saccharomyces cerevisiae* utilisées lors de cette étude.

Les différentes souches construites ont été obtenues par croisement entre la souche aUHT et les souches de la collection EUROSCARF.

1.2 Autres levures utilisées

Les caractéristiques des souches utilisées lors de l'analyse Génolevures sont décrites ci-après (Tableau MM2).

Espèce	N° Accession	Phénotype	Génotype	Origine
<i>Saccharomyces paradoxus</i>	CBS 432	sauvage		Kellis <i>et al.</i> , 2003
<i>Candida glabrata</i>	ATCC200989	His ⁻ Trp ⁻	<i>ura3Δ his3::URA3 trp1::URA3</i>	ATCC
<i>Zygosaccharomyces rouxii</i>	CLIB491	sauvage		CIRM
<i>Kluyveromyces lactis</i>	CLIB210	Met ⁻	<i>MATa met-</i>	CIRM
<i>Kluyveromyces thermotolerans</i>	CLIB292	sauvage		CIRM
<i>Debaryomyces hansenii</i>	CLIB197	sauvage		CIRM

Tableau MM2 : Souches utilisées dans le cadre de l'étude Génolevures

1.3 Milieux de culture

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

- milieu complet YPD : bactopectone 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 pour tester l'assimilation des différentes sources d'azote YCB : yeast carbon base 117 g/l
- milieu de présporulation GNA : glucose 50 g/l ; bacto beef extract 11,25 g/l ; bactopectone 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 ajout de 2% d'agar. Les milieux YNB et YCB peuvent être supplémentés par ajout de bases ou d'acides aminés (50 µg/ml) en fonction des auxotrophies des souches utilisées.

1.4 Conditions de culture

Les cellules sont cultivées soit en milieu liquide sous agitation soit en milieu solide. La température optimale d'incubation est de 30°C pour *S. cerevisiae*. En conditions optimales de croissance, lors de la phase exponentielle, le temps de génération est de 1h30 environ sur milieu complet et de 2h30 environ sur milieu minimum.

1.5 Conservation des souches

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

1.6 Mesure de la croissance cellulaire

La concentration cellulaire d'une culture de *Saccharomyces cerevisiae* peut être déterminée de deux manières :

- en milieu liquide, par la mesure de la turbidité de la culture à 600 nm par spectrophotomètre (ici Spectronic® 20 Genesys™)

Une unité d'absorbance équivaut à :

- 7×10^7 cellules haploïdes/ml
- 4×10^7 cellules diploïdes/ml

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

2. Techniques génétiques de la levure

2.1 Obtention de cellules diploïdes

Des diploïdes peuvent être obtenus par croisement en milieu solide d'une souche haploïde de signe a et d'une souche haploïde de signe 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.

2.2 Obtention de cellules haploïdes

Les souches diploïdes en phase exponentielle de croissance sont étalées sur milieu GNA et incubées une nuit à 30°C. Une anse de milieu est ensuite transférée sur milieu Ack et le milieu ensemencé est incubé à 30°C. Les premiers asques apparaissent au bout de deux à quatre jours d'incubation, en fonction des souches.

2.3 Analyse des spores

2.3.1 Dissection des tétrades

Elle permet d'analyser les phénotypes des spores issues de la méiose. Une anse de milieu de sporulation est mélangée à 200 µl d'eau stérile à laquelle seront ajoutés 150 µl de zymolyase 20T à 10 mg/ml. 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 peuvent être séparées puis alignées sur boîtes 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 individuellement.

2.3.2 Analyse des spores en vrac par test éther

Ce test permet d'évaluer la proportion des différents types de spores issues de la méiose. L'éther tue les diploïdes, détruit la paroi des asques et libère les spores (Dawes et Hardie, 1974).

Un ml d'éther est ajouté à 0,5 ml d'une suspension de milieu de sporulation. Ce mélange est vortexé pendant 30 s, incubé 20 min à température ambiante puis dilué en cascade au 1/10 et au 1/100. Ces dilutions sont étalées sur milieu complet puis analysées.

2.4 Analyse phénotypique

2.4.1 Détermination des phénotypes

Elle se fait en comparant la capacité de croissance des souches variées sur différents milieux sélectifs selon trois méthodes :

- les répliques velours : cette technique présente l'avantage de transférer toutes les colonies d'une boîte-mère sur un velours stérile puis sur une ou plusieurs boîtes-filles en une seule étape.

- 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 d'apprécier des différences de croissance pour les souches étudiées sur des gammes de concentration de substrats.

2.4.2 Détermination du signe sexuel

Le signe sexuel est déterminé par croisement sur milieu complet de la souche de signe inconnu avec une souche de signe a et une souche de signe α . On observe après 3 ou 4 h la présence ou non de zygotes suivant le croisement et on en déduit le signe.

3. Préparation d'ADN

3.1 Préparation rapide d'ADN génomique (Hoffman and Winston, 1987)

Dix ml de cellules cultivées en milieu complet liquide jusqu'en phase stationnaire sont centrifugées 5 min à 2500 rpm. Le culot est repris dans 1 ml d'eau et transféré en tube Eppendorf. Après centrifugation 2 min à 12400 rpm, les cellules sont reprises dans 250 μ l de tampon de broyage (TrisHCl pH8 10 mM ; EDTA 1 mM ; NaCl 100 mM ; Triton 2% ; SDS 1%) puis des billes de verre sont ajoutées ainsi que 250 μ l de phénol-chloroforme. Le mélange est vortexé 3 min puis centrifugé 5 min à 12400 rpm. La phase supérieure est transférée dans un autre tube Eppendorf et un volume de phénol-chloroforme est ajouté. Le mélange est vortexé 30 s environ puis centrifugé 2 min à 12400 rpm. La phase supérieure est à nouveau transférée dans un tube Eppendorf. L'ADN est ensuite précipité par ajout de 20 μ l d'acétate d'ammonium 4 M et de 600 μ l d'éthanol absolu. Le mélange est laissé 30 min dans la glace puis centrifugé 5 min à 12400 rpm. Le culot est séché et repris dans 400 μ l de TE (Tris 10 mM, EDTA 1 mM). 30 μ g de RNase A sont ajoutés et le mélange est incubé 15 min à 37°C. L'ADN est ensuite précipité par ajout de 10 μ l d'acétate de sodium 4 M et de 1 ml d'éthanol absolu et incubé 30 min dans la glace. Après centrifugation 10 min à 12400 rpm, le culot est rincé par ajout d'un ml d'éthanol 80% et centrifugé 5 min à 12400 rpm. Le culot est ensuite mis à sécher et repris dans 50 μ l de TE ou d'eau distillée.

3.2 Préparation des chromosomes entiers de levure (Vézinhel *et al.*, 1990)

L'analyse des caryotypes de levures par électrophorèse en champs alternés nécessite d'obtenir des chromosomes intacts. Les procédures habituelles de préparation

d'ADN ne permettent pas d'obtenir des molécules de haut poids moléculaire. Pour éviter les coupures dues aux forces mécaniques lors des différents mélanges et agitations, les cellules entières incorporées dans des blocs d'agarose sont lysées et déprotéinisées *in situ*. La matrice d'agarose protège l'ADN des forces de scission et fournit des échantillons faciles à manipuler pour l'analyse par électrophorèse en champs alternés.

50 ml de culture sont centrifugés 10 min à 2500 rpm. Les cellules sont ensuite lavées dans 25 ml d'EDTA 50 mM froid puis dans 10 ml de CPES froid formé d'un mélange de CPE×1 (acide citrique 40 mM, Na₂HPO₄ 120 mM, EDTA 20 mM), sorbitol 1,2 M et DTT 5 mM. Les cellules sont alors placées dans la glace puis reprises dans du CPES+zymolyase (200 µg/souche) tel que la concentration finale soit de $3,9 \times 10^9$ cellules/ml. Un ml de cellules est alors placé dans un bain-marie à 37°C pendant 3 min pour activer la zymolyase. 1 ml d'agarose BioRad 1,2% dans du CPE×1 est ajouté. Le mélange est vortexé et prélevé avec une pipette Pasteur puis coulé dans les puits du moule. Une fois pris en masse, les blocs sont démoulés, recouverts par 10 ml de CPE+zymolyase (100 µg/ml) puis incubés 2 h à 37°C. Le tampon est ensuite remplacé par 10 ml de tampon de lyse (Tris 10 mM pH8 ; EDTA 0,45 M ; SDS 1% ; Protéinase K 0,8 mg/ml) et les blocs sont incubés une nuit à 50°C. Le lendemain, deux lavages de 30 min à l'eau à 50°C sont réalisés suivis de deux lavages de 30 min à l'eau à température ambiante. Les blocs sont ensuite repris et conservés dans 10 ml d'EDTA 0,25 M à 4°C.

4. Analyse de l'ADN

4.1 Précipitation alcoolique de l'ADN

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

On ajoute 2,5 volumes d'éthanol 100% à -20°C et 1/10 de volume d'acétate de sodium 3 M pH7. Le mélange est incubé à -20°C pendant au moins 30 min puis est centrifugé 15 min à 12400 rpm. Le culot d'ADN est ensuite lavé avec de l'éthanol 70% à -20°C puis l'ADN est dissout dans le volume désiré d'eau distillée ou de tampon TE. Cette solution peut ensuite être conservée à -20°C.

4.2 Purification de l'ADN par chromatographie

Cette méthode est basée sur la technique de chromatographie par gel-filtration (ou exclusion). Elle repose sur une séparation différentielle des molécules en fonction de leur

taille sur une résine de Séphacryl. Cette technique a été utilisée pour séparer des fragments d'ADN double-brin amplifiés par PCR des autres molécules présentes dans le milieu réactionnel (dNTPs, sels, amorces en excès...). Ces impuretés de tailles plus petites vont être retenues sur la résine tandis que l'ADN amplifié sera élué rapidement par centrifugation.

La colonne est remise en suspension, placée dans un microtube puis centrifugée 1 min à 2400 rpm pour éliminer l'excédent de tampon. Après avoir placé la colonne dans un nouveau microtube, 25 à 50 µl des échantillons à purifier sont déposés délicatement à la surface de la résine puis centrifugés 2 min à 2400 rpm. On récupère ainsi la solution d'ADN pure dans un nouveau microtube.

4.3 Détermination de la concentration d'ADN

La concentration d'une solution d'ADN peut être calculée en mesurant l'absorbance de la solution à une longueur d'onde de 260 nm. En effet, à cette longueur d'onde, une unité de DO correspond à une concentration d'ADN bicaténaire de 50 µg/ml et à une concentration d'ADN monocaténaire de 40 µg/ml. Dans cette étude, les mesures ont été réalisées à l'aide d'un Eppendorf BioPhotometer.

4.4 Analyse par électrophorèse sur gel d'agarose

4.4.1 Electrophorèse en gel d'agarose

Les molécules d'ADN peuvent être séparées suivant leur taille par électrophorèse sur gel d'agarose 1% dans le tampon d'électrophorèse TAE×1 (Tris 4,84 g ; acétate de sodium 0,68 g ; acide acétique 1,35 ml ; EDTA 0,336 g ; H₂O qsp 1l) sous une tension de 30 V à 120 V. L'ajout de bromure d'éthidium (BET) dans le gel ou dans un bain après migration permet d'observer l'ADN sous UV.

4.4.2 Electrophorèse en champs alternés

Elle permet de séparer des fragments d'ADN de quelques kilobases (kb) à plus de 10000 kb ce qui va ainsi permettre d'observer les chromosomes de *Saccharomyces cerevisiae* dont la taille varie de 250 à 2500 kb.

Le système utilisé est le CHEF (Contour-clamped Homogeneous Electric Fields) où les électrodes sont organisées en hexagone permettant de définir deux champs électriques séparés de 120° (Chu, 1989). Les fragments d'ADN sont soumis à ces deux champs électriques alternativement. Le principe de séparation est basé sur le fait qu'il existe

une vitesse différentielle de réorientation des molécules lors du changement d'orientation du champ électrique. Ainsi les grandes molécules se réorientent moins rapidement que les petites et l'accumulation de ces retards va permettre la séparation des molécules en fin de migration. La migration s'effectue sur gel d'agarose 1% dans le tampon TBE×0,5 (Tris 5,4 g ; acide borique 2,75 g ; EDTA 0,5M 2 ml ; H₂O qsp 1 l) à une température de 14°C. Les conditions utilisées sont 22 h de migration à 6 V/cm avec un temps de commutation initial de 40 s et final de 120 s, l'angle entre les deux champs étant de 120°C.

4.5 Digestion par des endonucléases de restriction

Les enzymes de restriction sont des endonucléases qui coupent l'ADN bicaténaire au niveau d'une séquence spécifique le plus souvent palindromique. La coupure produit des extrémités cohésives 5' ou 3' sortantes ou franches. Les différentes enzymes de restriction sont utilisées dans des conditions précises de température, de force ionique et de pH décrites par le fournisseur (ROCHE) dans les tampons qui leur sont associés.

La réaction est réalisée en microtube Eppendorf contenant dans un volume final de 50 µl, un dixième de volume de tampon dix fois concentré, 20 U d'enzyme de restriction à 10 U/µl et de 0,5 à 1 µg d'ADN. Le temps de digestion peut varier de 2 h à une digestion réalisée sur la nuit.

5. Hybridation ADN/ADN selon la méthode de Southern (1975)

Cette technique permet de visualiser sur support solide des fragments d'ADN par l'intermédiaire d'une sonde spécifique de l'ADN à révéler. Elle a été utilisée dans le cadre de deux approches différentes :

- l'hybridation ADN/ADN classique : il s'agit d'une hybridation sur de l'ADN génomique digéré séparé par électrophorèse classique.
- le chromo-blot : il s'agit d'une hybridation de chromosomes entiers de levure séparés par électrophorèse en champs alternés.

5.1 Transfert

L'ADN localisé sur le gel est transféré sur une membrane de nylon Hybond N⁺ (MP Biomedicals). Un dispositif permet de réaliser ce transfert par capillarité. Le gel est tout d'abord soumis à trois bains successifs : un bain de dépurination (HCl fumant 20 ml ;

H₂O qsp 1 l) pendant 30 min, un bain de dénaturation (NaCl 88g/l; NaOH 20g/l) pendant 45 min et enfin un bain de neutralisation (HCl fumant 69ml ; Tris-Base 121g/l; NaCl 88g/l).

Le transfert est ensuite réalisé selon le montage suivant (Figure MM2) dans du tampon 20×SSC (NaCl 174 g/l, Na₃Citrate 88 g/l).

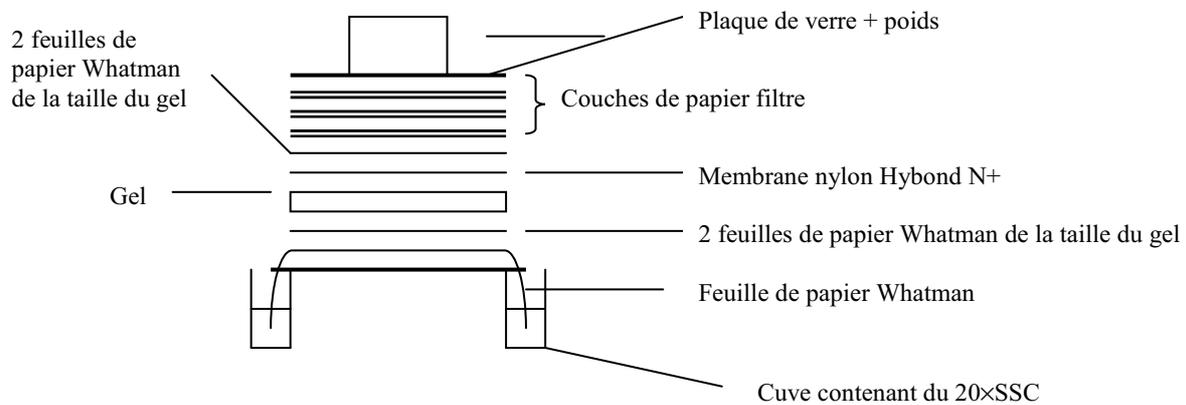


Figure MM2 : Schéma du montage réalisé lors de l'étape de transfert

Pour les molécules de haut poids moléculaires telles que les chromosomes de levure, le gel est uniquement soumis à un bain de dénaturation (NaCl 88g/l; NaOH 20g/l) et c'est dans ce même tampon qu'est réalisé le transfert sur la nuit.

5.2 Hybridation

5.2.1 Préparation d'une sonde froide

L'ADN spécifique est synthétisé par PCR puis purifié par chromatographie.

De 500 ng à 3 µg d'ADN linéaire dans un volume de 15 µl sont dénaturés par chauffage à 95°C pendant 10 min puis placés dans la glace. 2 µl d'hexanucléotides (×10), 2 µl de dNTPs marqués à la digoxygénine×10 et 1 µl du fragment de Klenow de l'ADN polymérase I (2 U/µl) sont ajoutés. Le mélange est incubé à 37°C pendant la nuit. L'ADN est ensuite précipité par ajout de 2,5 µl de chlorure de lithium 4M et de 75 µl d'éthanol absolu à -20°C. Le mélange est incubé à -80°C pendant 30 min (ou à -20°C pendant 2 h) puis centrifugé à 13400 g pendant 20 min. Le culot est lavé dans 50 µl d'éthanol 70% puis centrifugé 10 min à 13400 g. L'ADN est séché puis le culot est repris dans 25 µl de TE ou d'eau distillée et conservé à -20°C.

5.2.2 Préhybridation et hybridation

L'ADN est fixé à la membrane soit par chauffage 2 h à 80°C, soit par exposition aux UV 5 min. La membrane est ensuite préhybridée à 68°C pendant au moins 2 h dans un sac scellé avec 20 ml de tampon d'hybridation (5×SSC ; N-lauroylsarcosine 1 mg/ml ; SDS 0,02% ; solution bloquante×10 10 mg/ml). 10 µl de sonde sont dénaturés par chauffage 10 min à 95°C puis placés dans la glace. Le tampon est remplacé par le tampon d'hybridation préincubé à 68°C contenant les sondes. La membrane est alors incubée toute la nuit à 68°C.

Dans le cas d'un chromoblot, la membrane est tout d'abord neutralisée dans un bain de TrisHCl 0,5 M pH 7,5, NaCl 1 M avant d'être préhybridée.

5.3 Révélation au NBT/BCIP

La membrane est lavée deux fois 15 min à température ambiante dans du tampon SSC×2, SDS 0,1% puis deux fois 15 min à 68°C dans du tampon SSC×0,1, SDS 0,1%. Toutes les étapes se déroulent ensuite à température ambiante. La membrane est équilibrée pendant 1 min dans du tampon de lavage (TrisHCl 100 mM pH 7,5 ; NaCl 150 mM) puis est incubée durant 15 min dans un tampon de saturation (TrisHCl 100 mM pH 7,5 ; NaCl 150 mM ; lait en poudre demi-écrémé 5%). Ce traitement permet de saturer les sites spécifiques aux anticorps anti-DIG-AP. La membrane est ensuite placée dans un sac scellé dans 20 ml de tampon de saturation (pour une membrane de 100 cm²) additionné d'anticorps anti-digoxygénine (150 mU/ml final) pendant 45 min. Elle est ensuite lavée deux fois pendant 15 min dans le tampon de lavage puis équilibrée 2 min dans la solution de révélation (TrisHCl 100 mM pH 9,5 ; NaCl 100 mM ; MgCl₂ 50 mM). Elle est finalement placée à l'obscurité dans 5 ml de cette même solution additionnée de 80 µl de NBT/BCIP jusqu'à apparition de la coloration (de 2 à 16 heures).

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

6.1 Paramètres importants lors des différentes étapes

La PCR permet l'amplification exponentielle et spécifique d'un fragment d'ADN d'intérêt en utilisant des oligonucléotides amorces spécifiques du fragment d'ADN, les quatre dNTPs et une ADN polymérase thermostable. Elle comprend trois étapes qui seront répétées pendant x cycles. Le nombre de cycles dépend de la quantité de matrice introduite au départ. Les trois étapes sont les suivantes :

- la dénaturation : permet la séparation de l'ADN double brin en ADN simple brin. La température utilisée est généralement de 94°C et sa durée est d'environ 30 s.
- l'hybridation : est l'étape qui va déterminer la spécificité et le rendement de la réaction. C'est la séquence des amorces qui va déterminer la température d'hybridation (T_m). La durée de cette étape varie de 30 s à 1 min.
- l'élongation : permet la synthèse du brin complémentaire au brin matrice à partir des amorces. Elle est réalisée à une température un peu inférieure à la température optimale de l'enzyme utilisée (72°C pour la *Taq* DNA polymérase). Si l'on utilise la *Taq* DNA polymérase, la durée d'élongation est d'environ 1 min/kb.

Le nombre de cycles dépend de la quantité de matrice au départ. Il faut en général 25 à 35 cycles pour obtenir 1 mg d'ADN pour 100 µl de réaction à partir de quelques picogrammes d'ADN. Une étape de post-élongation est en général effectuée après la fin des cycles pour terminer la synthèse de certains brins éventuellement inachevés.

6.2 Matériel

Les réactions de PCR sont réalisées dans des tubes de 500 µl de volume. Le volume de milieu réactionnel est dans la plupart des cas de 50 µl. La réaction se fait dans un thermocycleur de type MastercycleurTM Gradient d'Eppendorf. L'enzyme utilisée est la *Taq* DNA Polymerase commercialisée par MP Biomedicals. Elle est fournie avec un tampon×10 dont la composition est la suivante : 10 mM Tris-HCl pH 9.0, 50 mM KCl, 1.5 mM MgCl₂, 0.1% Triton X-100, 0.2 mg/ml BSA ainsi qu'avec une solution contenant les 4 dNTPs à une concentration de 10 mM.

La composition du milieu réactionnel est de : 5 µl de tampon×10 ; 0,5 µl de la solution contenant les 4 dNTPs, 1 µl de chaque oligonucléotide à 20 µM ; environ 100 ng ; 0,2 unités d'enzyme ; eau bidistillée qsp 50 µl.

La liste des oligonucléotides utilisés lors de cette étude est présentée dans le tableau MM3.

Oligonucléotide	Séquence (5'-3')	Taille (nt)	Position	T _m (°C)
38	ATATTGGTTGTCTACTACG	19	-2608/-2590	52
38,2	GGGACAAGATAAAGGACG	18	-1647/-1630	54
38,3	TGCCTACGTACATTTCTG	18	-1117/-1100	52
38,6	GGACAAATGGTATGGAC	17	-799/-783	50
38,8	GGAAACATTATAATCCGTTA	20	-419/-400	52
FR4	ATCACGTGAAACTTCGA	17	3500/3484	48
AW1	AGCAGCTGATGTTGCTA	18	4190/4174	54
δB41	ATTCCATTGTTGATAAAGG	19		50
δB41R	TATCAACAATGGAATCC	17		46

Tableau MM3 : Liste des oligonucléotides amorces utilisés lors de cette étude.

Les positions sont indiquées par rapport à la position +1 du gène *URA2*.

7. Détermination de la séquence d'un fragment d'ADN

Cette technique décrite par Sanger *et al.* (1977) fait intervenir la synthèse interrompue d'ADN simple brin. A partir de l'extrémité 3'-OH de l'amorce appariée, l'ADN polymérase peut synthétiser le brin complémentaire en présence d'un mélange des quatre 2'-desoxyribonucléoside-5'-triphosphates (dATP, dCTP, dGTP et dTTP) et des quatre 2',3'-dideoxyribonucléoside-5'-triphosphates marqués chacun par un fluorochrome différent (ddATP*, ddCTP*, ddGTP*, ddTTP*). L'ensemble est soumis à une succession de cycles de polymérisation au cours desquels l'ADN polymérase peut, au niveau de chaque nucléotide de l'ADN matrice, incorporer un desoxyribonucléotide ou un didésoxyribonucléotide. Quand un didésoxyribonucléotide est incorporé, la réaction s'arrête. Cette incorporation se produit de manière aléatoire ainsi le milieu réactionnel contient l'ensemble des molécules néosynthétisées possibles. Ces molécules sont ensuite dénaturées puis séparées par électrophorèse capillaire selon leur taille. On peut ainsi reconstituer la séquence en analysant la nature du fluorochrome terminant chacun de ces fragments néosynthétisés, du plus petit (premier nucléotide de la matrice) au plus grand (dernier nucléotide de la matrice).

Le séquençage d'ADN a été effectué par la plate-forme technique de l'IBMP (Institut de Biologie Moléculaire des Plantes) à l'aide d'un séquenceur à électrophorèse capillaire ABI3130 xl d'Applied Biosystems. Les réactions de séquençage ont été réalisées en utilisant le Big Dye Terminator cycle sequencing kit V3.1 d'Applied Biosystems.

8. Système de sélection des révertants Ura^+

Le système utilisé a été mis au point par Roelants *et al.* (1995) et amélioré par Welcker *et al.* (2000). Ce système de sélection nous permet de sélectionner des révertants spontanés provenant d'événements indépendants.

Le protocole pour une sélection est représenté dans la Figure MM3 ci-dessous.

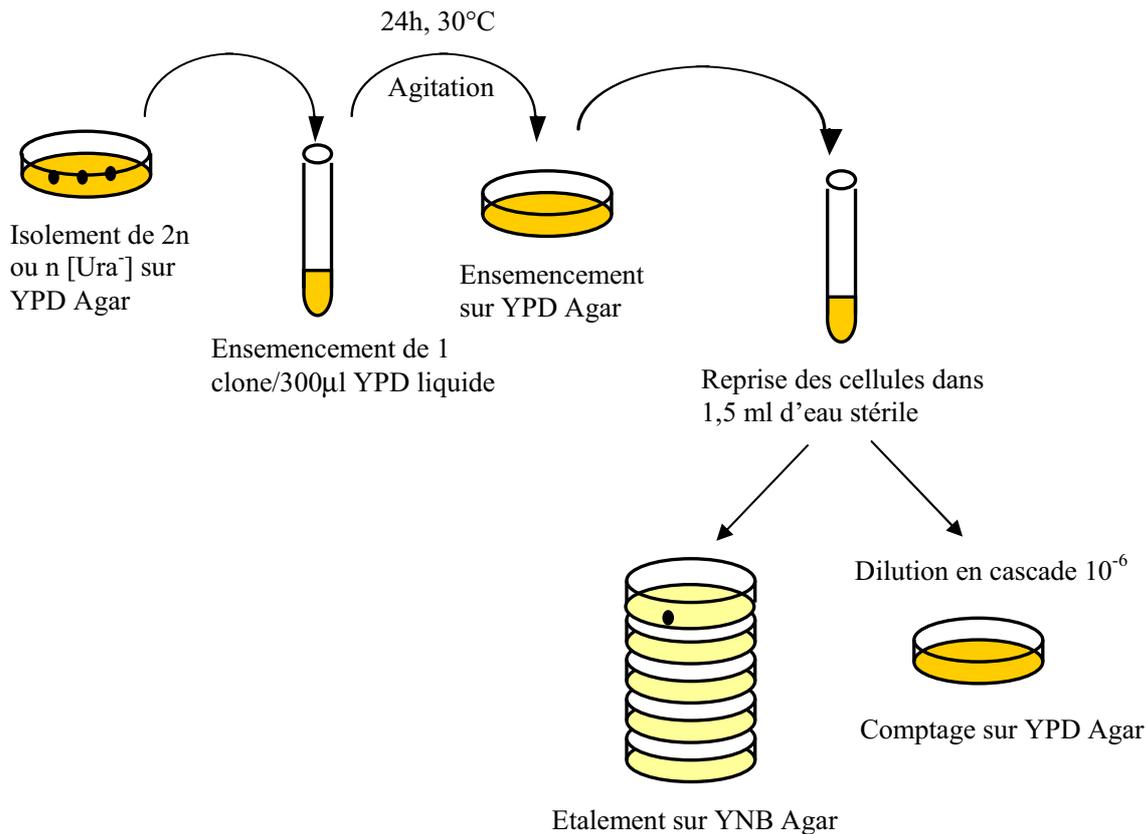


Figure MM3 : Représentation du système de sélections

Une colonie isolée d'une souche haploïde ou diploïde *ura2_{15,30,72}* est prélevée et mise en suspension dans 300 µl de milieu complet YPD liquide. Le milieu est incubé 24 h à 30 °C sous agitation puis étalé sur milieu complet YPD solide. On incube ensuite ce milieu ensemencé pendant 4 à 5 jours à 30 °C afin d'augmenter la masse cellulaire. La tapis de cellules ainsi obtenu est récolté et mis en suspension dans 1,5 ml d'eau distillée stérile. 100 µl de cette suspension sont prélevés, dilués à un facteur 10⁻⁶ puis étalés sur YPD solide pour estimer le nombre total de cellules étalées par sélection. Le milieu restant est réparti puis étalé sur 5 boîtes de milieu minimum YNB supplémenté en fonction des auxotrophies de chaque

souche. Les premiers révertants sont généralement obtenus après une semaine d'incubation à 30 °C. Ils sont alors prélevés et analysés.

Deux révertants sont dits indépendants s'ils proviennent de deux sélections différentes. S'ils sont issus d'une même sélection, ils ont pu être obtenus à la suite de deux événements distincts, dans ce cas ils présenteront des réarrangements différents et seront alors considérés comme indépendants. Ils peuvent également provenir de la division d'une cellule dans laquelle l'événement de réactivation a eu lieu de manière précoce. Dans ce cas, le réarrangement observé sera le même dans tous les révertants de la sélection. L'analyse moléculaire permettra de distinguer ce deux cas de Figure.

9. Informatique

La recherche de séquences d'acides nucléiques ou de protéines par similitude dans les banques de séquences Internet (Genbank, EMBL ou Swiss-Prot, par exemple) a été réalisée à l'aide du programme BLAST (Altschul *et al.*, 1990). Ce programme utilise une matrice de substitution d'acides aminés ou nucléiques (BLOSUM ou PAM par exemple). L'analyse et les comparaisons de séquences nucléotidiques ont été réalisées à l'aide du logiciel DNA StriderTM 1.3 ainsi qu'à l'aide du package EMBOSS (<http://bips.u-strasbg.fr/EMBOSS/>).

Les banques de données utilisées sont les suivantes :

- *Saccharomyces* Genome Database (SGD)
<http://www.yeastgenome.org/>
- Génolevures
<http://cbi.labri.fr/Genolevures/>
- *Ashbya* Genome database (AGD)
<http://agd.unibas.ch/index.html>
- Alignement multiple par T-Coffee
http://www.igs.cnrs-mrs.fr/Tcoffee/tcoffee_cgi/index.cgi
- Genbank
<http://www.ncbi.nlm.nih.gov/>

Partie I

Chapitre 1

Sélection et caractérisation des remaniements chromosomiques

Chapitre 2

Influence du contexte génétique sur l'apparition des remaniements chromosomiques

Chapitre 3

Mécanismes d'apparition des délétions et des duplications chromosomiques

Discussion et perspectives

1 SÉLECTION ET CARACTÉRISATION DES REMANIEMENTS CHROMOSOMIQUES

1. Introduction

Les modifications chromosomiques telles que les délétions, les duplications ou les translocations sont des remaniements rares apparaissant à une fréquence très faible de l'ordre de 10^{-10} chez *Saccharomyces cerevisiae*. Ces réarrangements sont donc difficiles à mettre en évidence et des systèmes de sélection ont été mis au point pour permettre leur étude *in vivo*. Parmi les nombreux cribles de sélection existants, quatre exemples permettant l'isolement de différents réarrangements chromosomiques chez *S. cerevisiae* seront détaillés :

- Le système *CANI URA3* (Figure 1.1)

Ce système de sélection a été développé par Chen et Kolodner (1999). Dans ce crible de sélection, le gène *HXT13*, localisé à environ 7,5 kb du gène *CANI*, a été remplacé par le gène *URA3*. Le gène *CANI* code pour une perméase à l'arginine permettant l'entrée de celle-ci dans la cellule mais également celle de la canavanine, son analogue toxique pouvant conduire en sa présence à la mort de la cellule. Le gène *URA3*, quant à lui, code pour l'orotidine-5'-phosphate décarboxylase qui convertit l'OMP en UMP mais également l'acide 5-fluoroorotique (5FOA) en son dérivé toxique le 5-fluorouridine monophosphate (5FUMP) (Figure 1.3).

Ainsi une souche sauvage est sensible à la canavanine et au 5FOA et ce crible positif repose sur la sélection de révertants résistants à ces deux composés suite à la perte de fonction des produits des gènes *CANI* et *URA3*. Les modifications observées à l'origine de cette perte sont des remaniements chromosomiques comme délétions, des translocations ainsi que des événements plus complexes comme des fusions de chromosomes. Ces événements peuvent apparaître de manière spontanée (Chen and Kolodner, 1999 ; Motegi and Myung, 2007) ou être induits par l'action d'agents mutagènes comme le MMS (méthyl méthane sulfonate) (Myung and Kolodner, 2002).

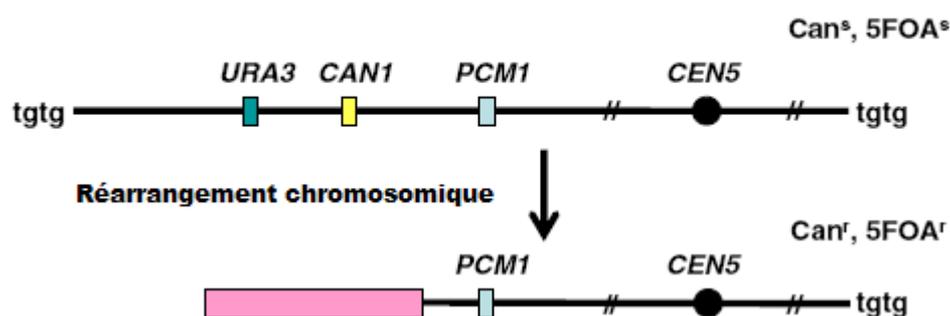


Figure 1.1 : Système de sélection *URA3 CAN1* (d'après Motegi and Myung, 2007).

Le rectangle rose représente l'un des réarrangements chromosomiques responsable de la perte de la cassette *URA3 CAN1*.

- Le système *URA3::intron ACT1* (Figure 1.2)

Ce crible de sélection, mis au point par Yu et Gabriel (1999), repose sur un allèle *URA3* modifié par l'addition d'une copie de l'intron du gène *ACT1*, codant pour l'actine. Cet intron a été modifié par l'ajout au niveau de sa séquence d'un site de coupure pour l'endonucléase HO. Cette enzyme est impliquée chez *S. cerevisiae* dans la conversion génique au locus *MAT* entraînant le changement de signe sexuel pour les souches homothalliques. Ceci est réalisé suite à la génération d'une cassure d'ADN double-brin par l'enzyme (Kostriken *et al.*, 1983).

Dans la souche utilisée, le gène codant pour l'endonucléase HO est sous contrôle d'un promoteur inductible en présence de galactose. Ainsi sur milieu contenant du galactose, le gène *HO* est exprimé et l'endonucléase synthétisée va créer une cassure d'ADN double-brin unique dans le gène *URA3* modifié. Celle-ci doit être réparée pour que la cellule survive. Si une religation simple du site de coupure se produit, l'intron du gène *ACT1* sera épissé et un retour à la prototrophie vis-à-vis de l'uracile sera observé. En revanche, si la réparation de la cassure implique un remaniement chromosomique plus complexe, l'épissage n'aura pas lieu suite à une modification des sites essentiels d'épissage et le révertant sera auxotrophe pour l'uracile mais résistant au 5FOA.

Plusieurs types d'événements peuvent conduire à la réparation de la cassure double-brin et ainsi conférer la résistance au 5FOA : l'insertion de rétrotransposons Ty1 (Yu and Gabriel, 1999) ou la mise en place de délétions (Yu and Gabriel, 2003) et de translocations (Yu and Gabriel, 2004).

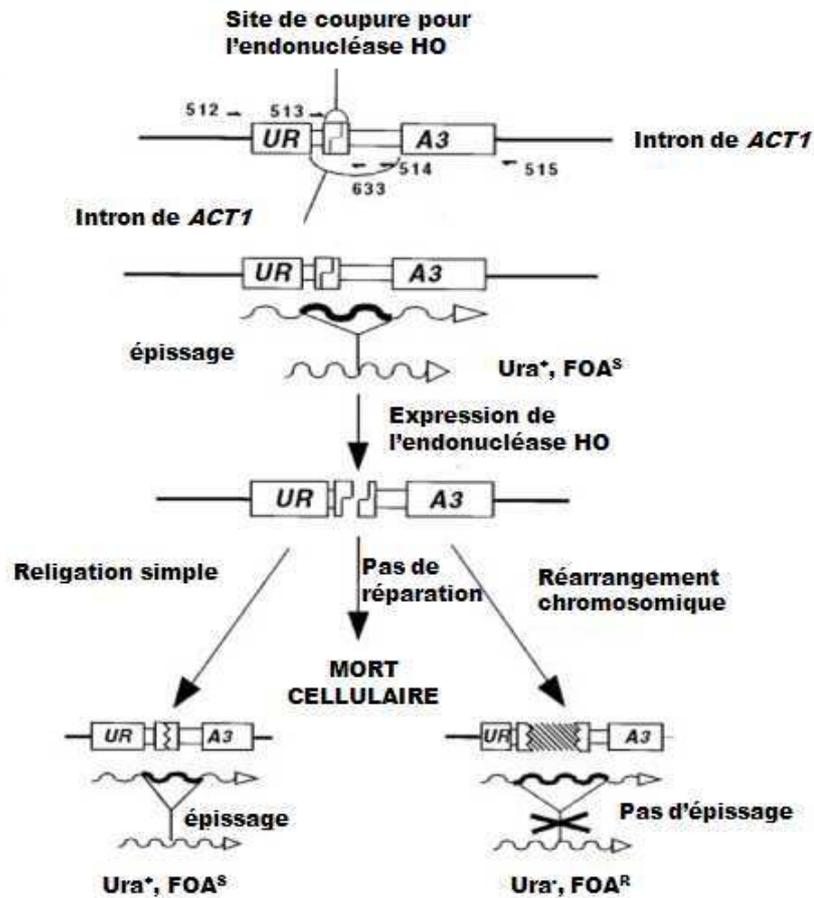


Figure 1.2 : Principe du système de sélection *URA3::intron ACT1* (d'après Yu and Gabriel, 1999)

- Le système RPL20A

Ce crible de sélection (Koszul *et al.*, 2004) est basé sur une différence de croissance existant entre une souche sauvage et une souche dans laquelle le gène *RPL20A* a été délété, conduisant à une vitesse de croissance réduite. Ce gène possède 99% d'identité avec son paralogue *RPL20B*. Dans une souche $\Delta RPL20A$, il a été montré qu'une copie supplémentaire de *RPL20B* permet la restauration d'une vitesse de croissance normale. Ainsi, à partir de la souche de départ $\Delta RPL20A$, des révertants ayant retrouvé une vitesse de croissance identique à celle du sauvage peuvent être sélectionnés. La caractérisation des souches obtenues montre que les événements à l'origine de la restauration de la vitesse de croissance sont des duplications du gène *RPL20B* qui peuvent être intra ou interchromosomiques ainsi que des aneuploïdies dans lesquelles le chromosome surnuméraire porte une copie supplémentaire du gène *RPL20B*.

- Réponse à une pression de sélection

Ce dernier exemple ne repose pas sur un crible de sélection associé à un ou plusieurs gènes marqueurs mais sur l'isolement de souches porteuses de remaniements chromosomiques suite à une pression de sélection (Dunham *et al.*, 2002). Une souche de *S. cerevisiae* a été cultivée en chemostat en conditions limitantes en glucose. Après plusieurs générations dans ces conditions, des clones issus de cette culture ont été analysés et il a été montré qu'ils pouvaient être porteurs de modifications chromosomiques telles des délétions, des duplications et des translocations.

Au laboratoire, un système de sélection a également été mis au point (Exinger et Lacroute, 1979 ; Roelants *et al.*, 1995). 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 nombreux types de remaniements spontanés différents comme des délétions, des duplications ou des insertions de rétrotransposons Ty1 dans un contexte haploïde comme diploïde.

2. Le système de sélection

Ce système de sélection est basé sur un allèle particulier du gène *URA2* qui diffère de l'allèle sauvage par trois mutations dans la partie proximale du gène. Situé sur le chromosome X, le gène *URA2* code pour une protéine multifonctionnelle impliquée dans les deux premières étapes de la voie de biosynthèse des pyrimidines (Figure 1.3). Cette protéine est composée à partir de l'extrémité N-terminale d'un domaine actif GATase-CPSase (glutamine amidotransférase-carbamylphosphate synthétase), d'un domaine cryptique DHOase-like (dihydroorotase-like) et d'un domaine actif ATCase (aspartyltranscarbamylase). L'allèle utilisé *ura2_{15,30,72}* possède deux mutations non-sens dans la région codant pour le domaine GATase et une mutation décalage du cadre de lecture dans la région codant la CPSase (Figure 1.4). Du fait de ces trois mutations, il y a arrêt prématuré de la traduction entraînant la perte des activités GATase, CPSase et ATCase. De plus, la présence des trois mutations conduit également à une diminution de la stabilité de l'ARN messager de l'allèle *ura2_{15,30,72}* ayant pour conséquence sa dégradation plus rapide que celui du gène sauvage (Losson et Lacroute, 1979 ; Souciet *et al.*, 1982). La souche présente donc un phénotype d'auxotrophie à l'uracile. Elle est ATCase⁻ mais reste cependant CPSase⁺ car cette activité est compensée par deux isoenzymes de la voie de biosynthèse de l'arginine (Figure 1.3).

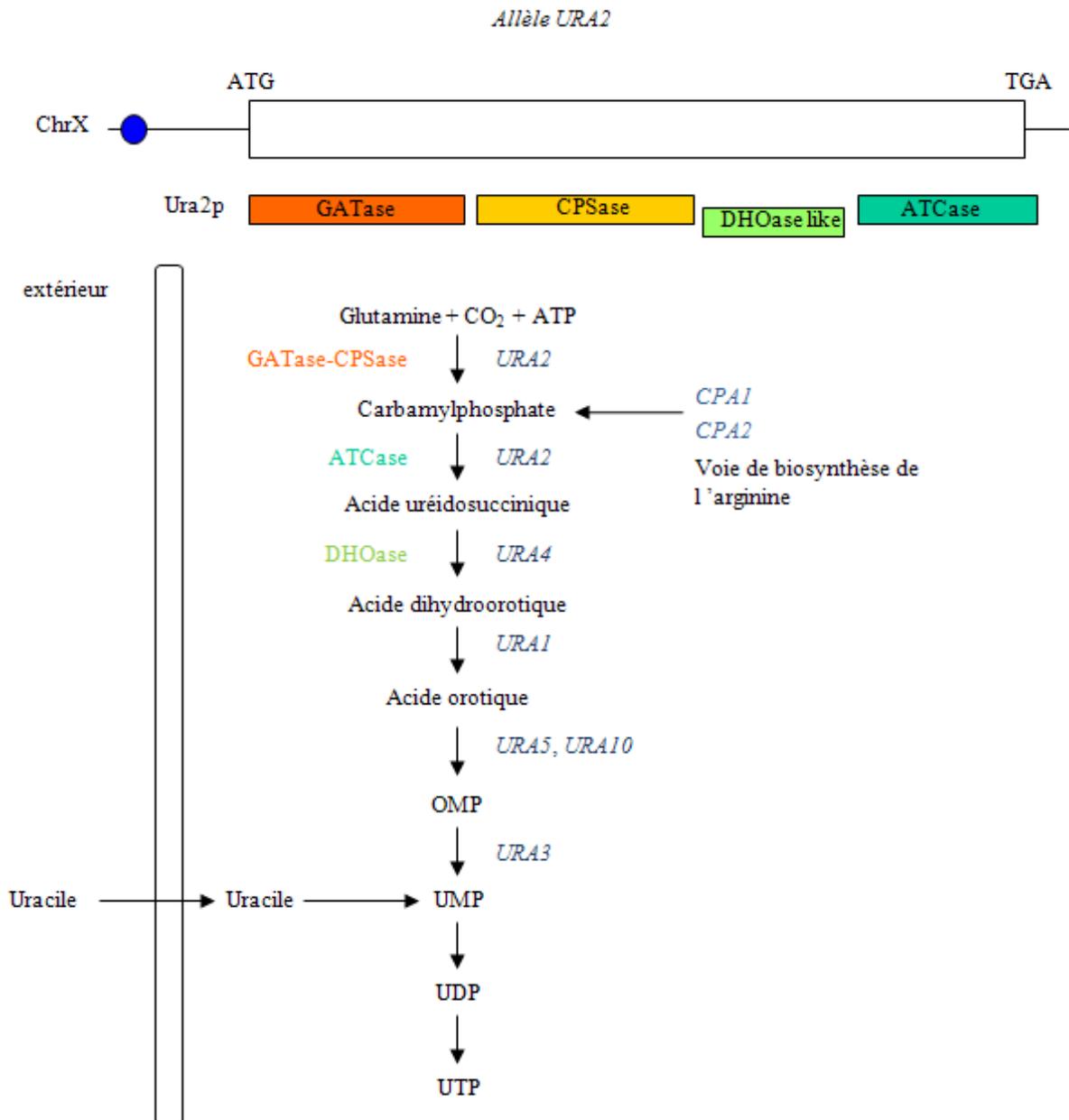


Figure 1.3 : Le gène *URA2*, la protéine Ura2p et les étapes de la voie de biosynthèse des pyrimidines chez *Saccharomyces cerevisiae*.

GATase-CPSase : Glutamine amidotransférase-Carbamylphosphate synthétase ; DHOase : Dihydroorotase ; ATCase : Aspartyltranscarbamylase

CPA1 et *CPA2* sont les deux gènes dont les produits assurent la synthèse du carbamylphosphate dans la voie de biosynthèse de l'arginine.

Une réactivation du domaine ATCase sera suffisante pour assurer un retour à la prototrophie. La réversion spontanée des trois mutations peut conduire à une prototrophie mais se produira à une fréquence de l'ordre de 10^{-18} . Ce type de réactivation est donc improbable avec les effectifs analysés (10^{12} cellules environ). Ainsi d'autres événements complexes seront à l'origine de la réactivation du domaine ATCase.

La sélection de tels révertants réalisée lors de travaux précédemment menés au laboratoire ont montré que trois types d'événements peuvent conduire à la réactivation de l'activité ATCase (Figure 1.4) :

- **L'insertion d'un rétrotransposon Ty1** en aval des trois mutations et en amont de la région codant pour l'ATCase (Roelants *et al.*, 1995).

Dans ce cas, la transcription du nouvel allèle est initiée à partir de la séquence δ du rétrotransposon qui est orientée en sens inverse par rapport à la région du gène *ura2*_{15,30,72} codant pour le domaine ATCase (Roelants *et al.*, 1997).

- **Une délétion de la région contenant les 3 mutations** (Welcker *et al.*, 2000).

La région codant pour l'ATCase sera alors transcrite à partir d'une région promotrice, ce qui permet la réactivation de cette activité.

- **Une duplication de la région codant pour le domaine ATCase** suivie de sa fusion avec une nouvelle région promotrice localisée à une autre position dans le génome.

Deux types de duplications ont été mis en évidence (Schacherer *et al.*, 2004) : les **duplications géniques** pour lesquelles seule la région codant pour l'ATCase est dupliquée et les **duplications segmentales** pour lesquelles la région codant pour l'ATCase ainsi qu'un nombre variable de gènes contigus ont été codupliqués.

Dans ce travail, nous avons utilisé ce système de sélection pour étudier l'impact de la variabilité génomique intra-espèces sur la sélection de remaniements chromosomiques et les mécanismes impliqués dans leur formation. La stratégie utilisée a consisté en l'isolement de révertants porteurs de remaniements chromosomiques dans différents contextes génétiques.

Afin de pouvoir comparer les résultats entre ces différents contextes, une analyse statistique consistant à déterminer les taux de mutation pour les différents événements a été mise au point et menée systématiquement pour l'ensemble des contextes. En parallèle, nous sommes aussi attachés à déterminer les réarrangements chromosomiques à l'origine des

événements de réversion et pour ce faire, une analyse moléculaire des différents révertants a été systématiquement réalisée.

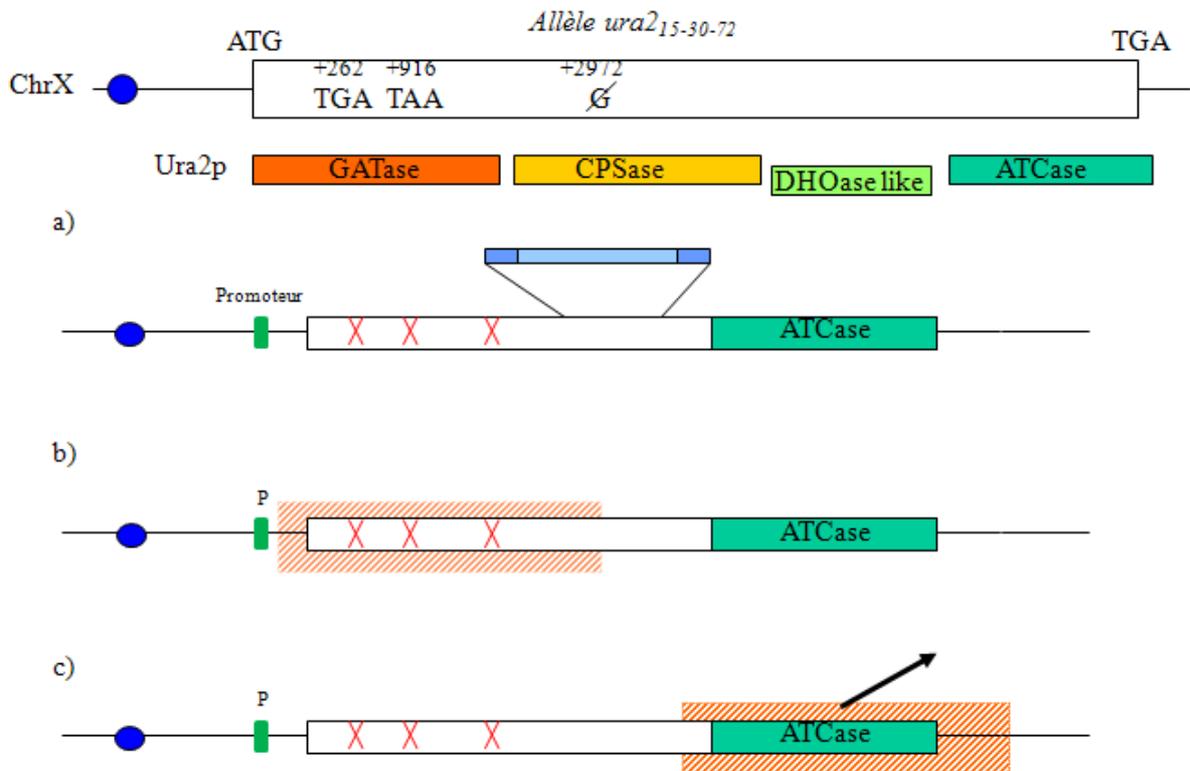


Figure 1.4 : L'allèle *ura2*_{15,30,72} et les remaniements chromosomiques mis en évidence comme permettant la réactivation de l'activité ATCase.

- Insertion d'un rétrotransposon Tyl
- Délétion de la région contenant les trois mutations
- Duplication de la région codant pour l'ATCase

3. Analyse statistique des événements de mutation

L'analyse statistique des événements de mutation entraînant la réactivation de l'ATCase a été réalisée d'après la méthode de Lea et Coulson (1949). Cette approche a été développée pour étudier la distribution du nombre de mutants au sein de populations bactériennes. Elle a ensuite été utilisée pour analyser les taux de mutations chez la levure *Saccharomyces cerevisiae*, par exemple dans le cas de la détermination des fréquences de rétrotransposition (Curcio and Garfinfel, 1991). Elle a aussi été développée chez l'Homme, notamment pour étudier les fréquences de mutation au locus *hprt* codant pour une hypoxanthine phosphoribosyltransferase (Green *et al.*, 1995).

3.1 Taux de mutation par cellule par sélection

La méthode de Lea et Coulson est adaptée pour le calcul de fréquences de mutation dans le cas d'événements rares. Cette méthode est donc particulièrement préconisée dans notre cas, puisque la fréquence d'apparition des événements de réversion est très faible (de l'ordre de 10^{-10} à 10^{-12}).

Cette analyse permet de calculer le taux de mutation global (pour l'ensemble des remaniements chromosomiques) ainsi que le taux pour chaque type d'événement mutationnel (délétions, duplications, insertions de rétrotransposons...). Une sélection correspond à une expérience réalisée pour un clone indépendant.

Afin de déterminer le taux de mutation global, il est nécessaire de connaître au préalable plusieurs données expérimentales :

- n_c le nombre de cellules étalées par sélection
- N le nombre total de sélections réalisées
- N_0 le nombre de sélections n'ayant donné aucun événement de réactivation
- N_x les nombres de sélections pour lesquelles x mutants ont été obtenus

Exemple : N_1 correspond au nombre de sélections pour lesquelles un seul révertant a été mis en évidence.

A partir de ces valeurs, m_e , une estimation du nombre moyen de mutations par sélection (noté m), peut être calculée. Cet estimateur est déterminé à partir de la proportion p_0 de sélections n'ayant abouti à aucun mutant par rapport au nombre total de sélections réalisées :

$$m_e = -\ln(p_0) = -\ln(N_0/N)$$

P_0 est la proportion de sélections n'ayant abouti à aucun révertant par rapport au nombre total de sélections. N_0 correspond au nombre de sélections sans révertant et N au nombre total de sélections réalisées.

Cette valeur de l'estimateur m_e est ensuite utilisée pour calculer les termes s_x qui permettront de définir S , le terme de maximum de vraisemblance défini par Lea et Coulson (1949). La table 4 de Lea et Coulson (Figure 1.5) contient les valeurs r_x , définies comme le nombre de révertants obtenus par sélection en fonction des différentes valeurs possibles de m , nombre moyen de mutations par sélection. Dans la table, à l'intersection de ces deux paramètres, on trouve la valeur notée s_x .

A partir de ces valeurs s_x déterminées pour chaque sélection réalisée, la valeur de S , terme de maximum de vraisemblance est déterminée en utilisant la formule suivante : $S = \Sigma(N_x s_x)$ pour m_e . M_e a été déterminé expérimentalement et a donc une valeur approximative ne correspondant pas nécessairement aux valeurs théoriques définies dans la table. On procède donc au calcul de S pour les valeurs de m définies dans la table encadrant m_e déterminé par le calcul. Le meilleur estimateur de m est ensuite déterminé comme le nombre moyen de mutations par sélection pour lequel S est la plus proche de 0.

Table 4. $\frac{t_r - p_r}{p_r}$ and $\frac{\sigma_m}{m} \sqrt{N}$

This table is used in estimating mutation rate by the maximal likelihood method.

$\frac{t_0 - p_0}{p_0} = -1$ for all values of m .

$m \setminus r$	1	2	3-4	5-8	9-16	17-32	33-64	>64	$\sigma_m \sqrt{N}/m$
0-05	19-000	19-723	20-019	20-179	20-211	20-176	20-124	20-057	4-527
0-10	9-000	9-698	9-998	10-166	10-206	10-174	10-124	10-057	3-239
0-15	5-667	6-341	6-644	6-821	6-868	6-839	6-790	6-723	2-674
0-20	4-000	4-652	4-957	5-142	5-196	5-171	5-123	5-057	2-341
0-25	3-000	3-632	3-939	4-130	4-191	4-170	4-123	4-057	2-116
0-30	2-333	2-946	3-254	3-452	3-519	3-501	3-456	3-390	1-951
0-35	1-857	2-451	2-760	2-964	3-038	3-023	2-979	2-914	1-824
0-40	1-500	2-077	2-386	2-596	2-676	2-665	2-622	2-557	1-722
0-45	1-222	1-783	2-092	2-307	2-393	2-386	2-344	2-279	1-638
0-50	1-000	1-545	1-855	2-075	2-166	2-162	2-121	2-057	1-568
0-55	0-818	1-349	1-658	1-882	1-979	1-978	1-939	1-875	1-507
0-60	0-667	1-184	1-492	1-720	1-822	1-825	1-787	1-723	1-455
0-65	0-538	1-043	1-350	1-582	1-689	1-695	1-658	1-595	1-409
0-70	0-429	0-920	1-227	1-462	1-574	1-584	1-548	1-485	1-368
0-75	0-333	0-813	1-119	1-357	1-474	1-487	1-452	1-390	1-332
0-80	0-250	0-719	1-023	1-264	1-386	1-402	1-369	1-307	1-299
0-85	0-176	0-634	0-937	1-181	1-308	1-327	1-295	1-233	1-269
0-90	0-111	0-559	0-860	1-107	1-237	1-259	1-229	1-168	1-242
0-95	0-053	0-491	0-790	1-039	1-174	1-199	1-170	1-109	1-217
1-0	0-000	0-429	0-727	0-978	1-117	1-145	1-117	1-057	1-194
1-2	-0-167	0-228	0-520	0-777	0-931	0-971	0-948	0-890	1-118
1-4	-0-286	0-080	0-364	0-627	0-793	0-845	0-828	0-771	1-059
1-6	-0-375	-0-034	0-243	0-508	0-686	0-749	0-736	0-682	1-012
1-8	-0-444	-0-125	0-144	0-411	0-599	0-672	0-665	0-612	0-973
2-0	-0-500	-0-200	0-063	0-330	0-526	0-609	0-608	0-557	0-940
2-2	-0-545	-0-262	-0-007	0-260	0-464	0-556	0-560	0-511	0-913
2-4	-0-583	-0-315	-0-066	0-200	0-409	0-510	0-520	0-473	0-888
2-6	-0-615	-0-361	-0-119	0-147	0-362	0-471	0-486	0-441	0-867
2-8	-0-643	-0-401	-0-164	0-099	0-319	0-436	0-456	0-413	0-848
3-0	-0-667	-0-436	-0-205	0-057	0-280	0-404	0-430	0-390	0-831
3-2	-0-688	-0-467	-0-242	0-018	0-245	0-376	0-407	0-369	0-816
3-4	-0-706	-0-495	-0-275	-0-017	0-213	0-350	0-386	0-350	0-802
3-6	-0-722	-0-520	-0-305	-0-050	0-183	0-326	0-367	0-334	0-789
3-8	-0-737	-0-542	-0-333	-0-080	0-155	0-304	0-350	0-319	0-777
4-0	-0-750	-0-563	-0-358	-0-107	0-129	0-284	0-334	0-305	0-766
4-2	-0-762	-0-581	-0-381	-0-133	0-105	0-264	0-320	0-293	0-756
4-4	-0-773	-0-598	-0-402	-0-157	0-082	0-246	0-306	0-282	0-747
4-6	-0-783	-0-614	-0-422	-0-179	0-061	0-229	0-293	0-272	0-738
4-8	-0-792	-0-629	-0-441	-0-200	0-041	0-213	0-282	0-263	0-730
5-0	-0-800	-0-642	-0-458	-0-220	0-021	0-197	0-271	0-254	0-722
6	-0-833	-0-697	-0-530	-0-304	-0-062	0-130	0-223	0-219	0-690
7	-0-857	-0-737	-0-584	-0-369	-0-128	0-073	0-184	0-194	0-665
8	-0-875	-0-768	-0-626	-0-422	-0-184	0-025	0-151	0-174	0-644
9	-0-889	-0-792	-0-660	-0-465	-0-231	-0-016	0-121	0-157	0-627
10	-0-900	-0-812	-0-689	-0-502	-0-272	-0-053	0-095	0-143	0-613
11	-0-909	-0-828	-0-712	-0-534	-0-307	-0-086	0-071	0-131	0-600
12	-0-917	-0-842	-0-733	-0-561	-0-339	-0-116	0-048	0-120	0-589
13	-0-923	-0-853	-0-750	-0-586	-0-367	-0-143	0-028	0-110	0-580
14	-0-929	-0-863	-0-765	-0-607	-0-393	-0-167	0-008	0-101	0-571
15	-0-933	-0-872	-0-779	-0-626	-0-416	-0-190	-0-010	0-092	0-564

Figure 1.5 : Table 4 de Lea et Coulson (1949)

Cette table permet l'estimation du taux de mutation par la méthode du maximum de vraisemblance.

A partir de cette valeur m , on peut calculer μ , le taux de mutation exprimé en nombre de mutations par cellule par sélection en utilisant la formule $\mu = m/n_c$ avec m , le nombre moyen de mutation par sélection et n_c , le nombre de cellules étalées par sélection.

3.2 Intervalle de confiance à 95% associé au taux de mutation

Il est essentiel de pouvoir comparer les résultats obtenus dans les différents contextes. Or, la valeur de μ , définie comme le taux de mutation par cellule par sélection, est à elle seule peu informative puisqu'il nous est impossible de définir si les taux de mutation globaux dans les différents contextes sont statistiquement différents ou pas. Il a donc été nécessaire de déterminer des intervalles de confiance pour les différents taux de mutation calculés. Deux taux de mutation sont estimés significativement différents si les intervalles de confiance ne se chevauchent pas.

Le calcul de l'intervalle de confiance des taux de mutation est réalisé à partir d'un test t de Student bilatéral. La table 4 de Lea et Coulson (Figure 1.5) donne une valeur correspondant à $\sigma_m\sqrt{N/m}$ pour chaque valeur de m (estimateur du nombre moyen de mutations), permettant ainsi de calculer la valeur de l'écart-type σ_m lié à m . On peut ensuite déterminer σ_μ , écart-type pour le taux de mutation en utilisant la formule $\sigma_\mu = \sigma_m/n_c$ avec σ_m la valeur de l'écart-type associée à m et à n_c le nombre total de cellules par sélection.

Compte tenu de l'écart-type σ_μ associé au taux de mutation μ et de la valeur critique de la loi de Student t_α pour le risque donné fixé à 5%, l'intervalle de confiance à 95% peut être calculé de la façon suivante : $[\mu - t_{0,05}\sigma_\mu ; \mu + t_{0,05}\sigma_\mu]$. La valeur $t_{0,05}$ pour $N-1$ degrés de liberté est déterminée d'après la table de Student (Annexe 1) ou à partir du logiciel MINITAB.

3.3 Taux de mutation rapporté à chaque type de réarrangements chromosomiques

L'influence d'un contexte génétique peut être définie comme l'impact du contexte sur la fréquence globale d'apparition des réarrangements chromosomiques mais aussi sur les différents types de réarrangements sélectionnés (délétions, duplications, insertions de rétrotransposons...). Pour déterminer ce paramètre, nous avons défini le taux de mutation relatif à chaque type de remaniements chromosomiques.

Pour déterminer ces taux, il est nécessaire de calculer la proportion de chaque type de réarrangements par rapport au nombre total de remaniements sélectionnés. Le taux de

mutation pour chaque type de réarrangement est obtenu en multipliant ce pourcentage par le taux de mutation global correspondant à celui de tous les remaniements confondus.

3.4 Exemple de calcul du taux de mutation ainsi que de son intervalle de confiance

Le calcul est détaillé ici pour les sélections effectuées dans la souche de référence aUHT.

Dans le contexte aUHT, 155 sélections ont été réalisées ($N=155$). Le nombre total de cellules par sélection n_c est de $2,439 \times 10^9$. Le tableau suivant (Tableau 1.1) indique le nombre de sélections N_x ayant abouti à x révertants.

r_x	N_x
0	117
1	23
2	9
3	2
4	1
5	2
10	1

Tableau 1.1 : Nombre de sélections en fonction du nombre de révertants obtenus par sélection

Avec ces valeurs, on peut estimer le nombre moyen de mutations par sélection noté m_e par la relation suivante :

$$m_e = -\ln(N_0/N) = -\ln(117/155) = 0,281$$

Cette valeur de m_e ne se trouvant pas directement dans la table 4 de Lea et Coulson, on détermine le terme S de maximum de vraisemblance pour les valeurs de la table encadrant m_e à savoir pour m égal à 0,25 et m égal à 0,30. S est défini d'après la formule $S = \sum(N_x s_x)$. N_x est le nombre de sélections ayant abouti à x révertants et s_x , la valeur obtenue dans la table à l'intersection entre m (le nombre moyen de mutations par sélection) et r (le nombre de révertants pour une sélection donnée) (Tableau 1.2 et Tableau 1.3).

r	N_x	$s_x (m_e=0,30)$
0	117	-1
1	23	2,333
2	9	2,946
3	2	3,254
4	1	3,254
5	2	3,452
10	1	3,519

Tableau 1.2 : Valeurs de s_x obtenues d'après la table 4 de Lea et Coulson en fonction du nombre r de révertants obtenus et du nombre N_x correspondant pour l'estimateur $m_e = 0,30$.

On calcule S pour $m_e = 0,30$:

$$S = (N_0 \times s_0) + (N_1 \times s_1) + (N_2 \times s_2) + (N_3 \times s_3) + (N_4 \times s_4) + (N_5 \times s_5) + (N_{10} \times s_{10})$$

$$S = (117 \times -1) + (23 \times 2,333) + (9 \times 2,946) + (2 \times 3,254) + (1 \times 3,254) + (2 \times 3,452) + (1 \times 3,51)$$

$$S = -16,6$$

r	N_x	$s_x (m_e=0,25)$
0	117	-1
1	23	3
2	9	3,632
3	2	3,939
4	1	3,939
5	2	4,13
10	1	4,191

Tableau 1.3 : Valeurs de s_x obtenues d'après la table 4 de Lea et Coulson en fonction du nombre r de révertants obtenus et du nombre N_x correspondant pour l'estimateur $m_e = 0,25$.

On calcule S pour $m_e = 0,25$:

$$S = (N_0 \times s_0) + (N_1 \times s_1) + (N_2 \times s_2) + (N_3 \times s_3) + (N_4 \times s_4) + (N_5 \times s_5) + (N_{10} \times s_{10})$$

$$S = (117 \times -1) + (23 \times 3) + (9 \times 3,632) + (2 \times 3,939) + (1 \times 3,939) + (2 \times 4,13) + (1 \times 4,191)$$

$$S = 8,956$$

En comparant les deux valeurs de S obtenues, on peut définir le meilleur estimateur de m pour lequel S est la plus proche de 0. Il s'agit de $m=0,25$. A partir de cette valeur, μ , le taux de mutation correspondant au nombre de mutations par cellule par sélection, est défini par la relation suivante :

$$\mu = m/n_c = 0,25/2,439 \times 10^9 = 1,025 \times 10^{-10} \text{ mutation/cellule/sélection}$$

L'écart-type σ_m est obtenu d'après la table de Lea et Coulson.

Pour $m=0,25$, $\sigma_m \sqrt{N}/m = 2,116$ d'où :

$$\sigma_m = 2,116 \times m / \sqrt{N} = 2,116 \times 0,25 / \sqrt{155} = 0,043.$$

σ_μ , l'écart-type associé au taux de mutation, peut ensuite être défini :

$$\sigma_\mu = \sigma_m / n_c = 0,043 / 2,439 \times 10^9 = 1,763 \times 10^{-11}$$

A partir de ces données, il est possible de déterminer l'intervalle de confiance dans le contexte aUHT. Pour cela, on définit la valeur critique de la loi de Student pour le risque fixé à 5% et pour N-1 degrés de liberté. D'après le logiciel MINITAB, $t_{0,05}(N-1=154) = 1,9755$.

On en déduit les bornes de l'intervalle de confiance (IC) :

$$\mu - t_{0,05}\sigma_{\mu} = 1,025 \times 10^{-10} - 1,9755 \times 1,763 \times 10^{-11} = \mathbf{0,676 \times 10^{-10} \text{ mutation/cellule/sélection}}$$

$$\mu + t_{0,05}\sigma_{\mu} = 1,025 \times 10^{-10} + 1,9755 \times 1,763 \times 10^{-11} = \mathbf{1,373 \times 10^{-10} \text{ mutation/cellule/sélection}}$$

Dans le contexte aUHT, le taux de mutation est donc de :

$1,025 \times 10^{-10}$ mutation/cellule/sélection [IC : $0,676 \times 10^{-10}$; $1,373 \times 10^{-10}$] avec IC, l'intervalle de confiance.

3.5 Exemple de calcul du taux de mutation relatif aux différents réarrangements

Dans la souche aUHT, 20 délétions et 9 duplications ont été mises en évidence soit 69% de délétions et 31% de duplications.

On présentera ici le calcul du taux de mutation et des écarts-type uniquement dans le cas des délétions.

Pour les délétions, le taux de mutation est obtenu en multipliant la proportion d'événements de délétion par le taux de mutation global trouvé précédemment :

$$\mu = 0,69 \times 1,025 \times 10^{-10} = \mathbf{0,707 \times 10^{-10} \text{ mutation/cellule/sélection}}$$

$$\mu - t_{0,05}\sigma_{\mu} = 0,69 \times 0,676 \times 10^{-10} = \mathbf{0,466 \times 10^{-10} \text{ mutation/cellule/sélection}}$$

$$\mu + t_{0,05}\sigma_{\mu} = 0,69 \times 1,373 \times 10^{-10} = \mathbf{0,947 \times 10^{-10} \text{ mutation/cellule/sélection}}$$

Le taux de délétion dans le contexte aUHT est de : **$0,707 \times 10^{-10}$ délétion/cellule/sélection [IC : $0,466 \times 10^{-10}$; $0,947 \times 10^{-10}$]** avec IC, l'intervalle de confiance.

Dans cette étude, les remaniements chromosomiques obtenus ont été comparés entre eux en utilisant les fréquences d'apparition. Afin de valider ces résultats, nous avons également comparé les effectifs obtenus par la méthode de Khi Deux. Une bonne corrélation est observée entre les deux méthodes. L'approche du Khi Deux ne sera pas détaillée ici.

Lors de notre étude, nous avons testé plusieurs souches mutées pour des gènes dont les produits sont impliqués dans les différentes voies de recombinaison. Afin de déterminer si le produit du gène délété a un impact sur la fréquence d'apparition globale des remaniements chromosomiques ainsi que sur le type de remaniements observés, des sélections de révertants porteurs de réarrangements chromosomiques ont ainsi été réalisées dans toutes ces souches. La méthode d'analyse statistique présentée ci-dessus a été répétée pour tous les

contextes génétiques. Ceci nous a permis de comparer les résultats entre eux et de déterminer l'existence ou non de différences statistiques. Dans la suite du texte, les calculs ne seront plus détaillés.

4. Caractérisation moléculaire des remaniements chromosomiques

4.1 Analyse du profil de restriction *Bam*HI par la méthode de Southern blot

Pour tous les révertants, une analyse du profil de restriction *Bam*HI au locus *URA2* a été réalisée. Le choix de l'enzyme de restriction repose sur le fait qu'elle ne devait pas couper à l'intérieur de la région codant pour l'ATCase dans le gène *URA2* et que le fragment de restriction obtenu devait avoir une taille inférieure à 15 kb pour faciliter sa détection lors de l'hybridation ADN/ADN de type Southern sur gel d'agarose 1% en électrophorèse classique. L'enzyme de restriction *Bam*HI répond à tous ces critères puisque la région codant pour l'ATCase est préservée, les sites de coupure se trouvant de part et d'autre de la région codant pour cette activité. Le fragment ainsi obtenu a une taille de 6,8 kb dans la souche mutée servant à la sélection (Figure 1.6).

Pour réaliser le profil de restriction, l'ADN génomique des différents révertants a été digéré par l'enzyme *Bam*HI puis une hybridation a été réalisée en utilisant comme sonde un fragment obtenu par PCR correspondant à la région codant pour l'ATCase. Les différents profils obtenus sont présentés dans la Figure 1.7.

Pour la souche *ura2_{15,30,72}*, une seule bande de 6,8 kb correspondant au locus *URA2* est détectée (piste a). En revanche, pour tous les révertants, on observe un profil modifié. Dans le cas d'une duplication (pistes d et e), le profil de restriction présente deux bandes : l'une de 6,8 kb correspondant au locus muté et une seconde de taille variable correspondant à la copie dupliquée de la région codant pour l'ATCase insérée dans une région différente du génome. Dans le cas d'une délétion ou d'une insertion de rétrotransposon Ty1, une seule bande de taille supérieure à 6,8 kb est observable (pistes b et c).

L'analyse du profil de restriction *Bam*HI permet donc dans un premier temps de discriminer les événements de duplication des autres types d'événements. Les délétions et les insertions de rétrotransposons Ty1 sont ensuite différenciées par analyse PCR.

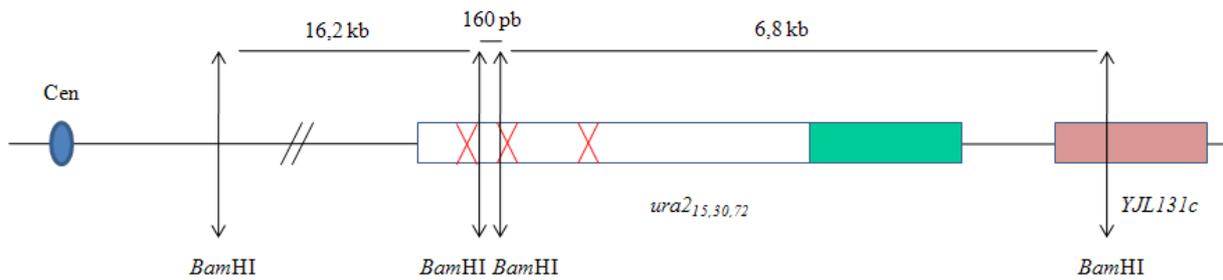


Figure 1.6 : Localisation des sites de restriction *Bam*HI au niveau du locus *ura2*_{15,30,72}.

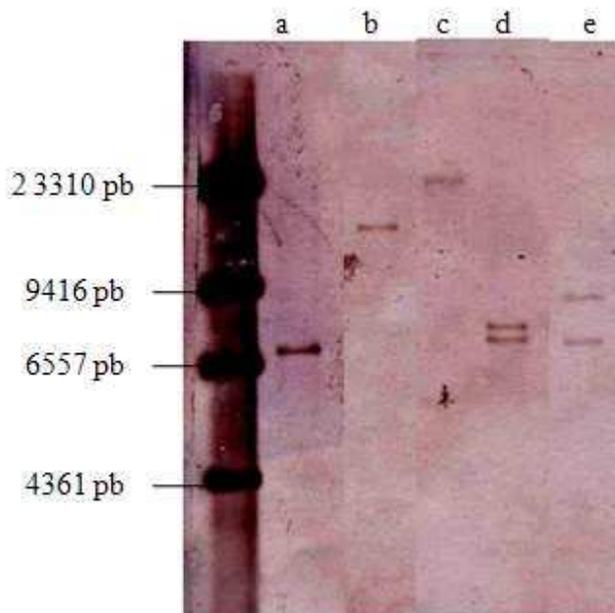


Figure 1.7 : Exemple de profil de restriction *Bam*HI obtenu en Southern blot.

Piste a : souche *ura2*_{15,30,72}

Pistes b et c : révertants aUHT01-1 et aUHT24-1

Pistes d et e : révertants aUHT119-1 et aUHT120-1

4.2 Analyse des délétions et des insertions de rétrotransposon Ty1 par PCR

4.2.1 Mise en évidence des délétions

Lors des études précédentes (Roelants *et al.*, 1995 ; Welcker *et al.*, 2000), il a été montré que les délétions obtenues couvrent généralement la séquence promotrice ainsi que la région contenant les trois mutations ponctuelles au niveau du gène *URA2*. Ces événements peuvent être mis en évidence par amplification PCR en utilisant des oligonucléotides spécifiques situés de part et d'autre de l'allèle *ura2*_{15,30,72} (Figure 1.8). L'amplification d'un fragment PCR permettra de confirmer qu'une délétion est à l'origine de la réactivation de l'activité ATCase. De plus la taille des amplifications mises en évidence permettra de déduire l'étendue de la délétion.

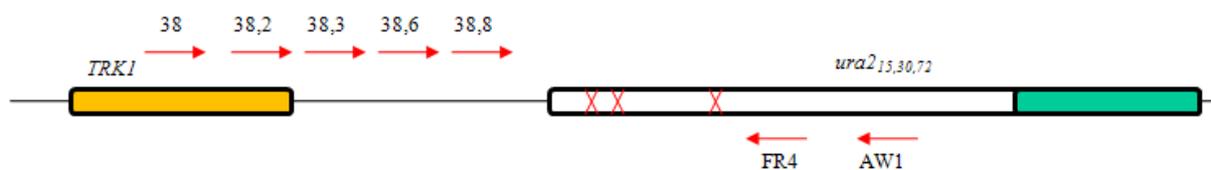


Figure 1.8 : Carte représentative des oligonucléotides utilisés pour mettre en évidence les événements de délétion.

Dans la souche *ura2*_{15,30,72}, les deux oligonucléotides les plus proches utilisés à savoir 38,8 et FR4 sont séparés de 3919 pb. C'est donc la taille minimale à amplifier entre les oligonucléotides pour la souche *ura2*_{15,30,72}. Dans les conditions expérimentales que nous avons utilisées, notamment un temps d'élongation de 1 min 30 sec, l'amplification d'un tel fragment est impossible. En revanche dans le cas d'une délétion dans laquelle la région mutée est éliminée, la longueur du fragment à amplifier est diminuée et une amplification d'un fragment PCR est alors possible dans ces mêmes conditions (Figure 1.9). En cas de délétions plus longues, les autres oligonucléotides représentés dans la Figure 1.8 sont utilisés selon le même principe.

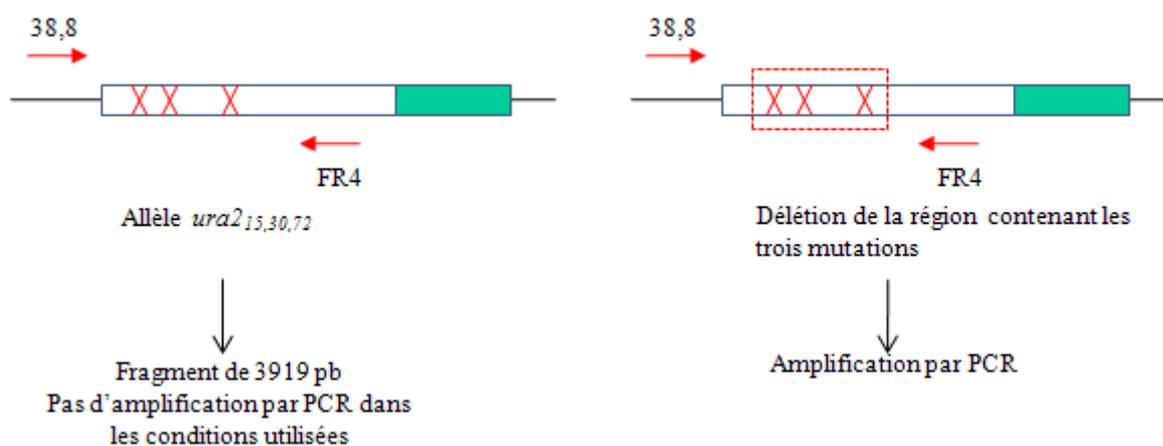


Figure 1.9 : Principe de détection des délétions par amplification PCR.

38,8 est un oligonucléotide en amont de l'ATG initiateur et FR4 est localisé dans la région codant pour la CPSase en aval des trois mutations.

La détermination de la séquence des différents produits PCR obtenus a permis de caractériser les bornes des délétions. Dans la plupart des cas, l'analyse montre que des répétitions directes de courtes tailles appelées microhomologies sont présentes aux bornes des délétions (Figure 1.10). Ces différentes microhomologies seront détaillées dans les chapitres suivants.

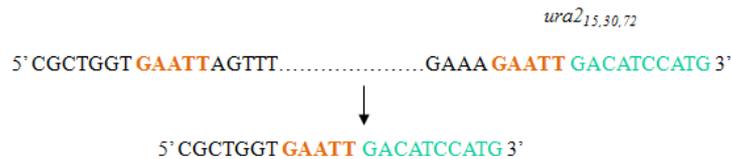


Figure 1.10 : Exemple de microhomologies présentes aux bornes d'une délétion.

La séquence répétée est représentée en orange. La région codant pour l'activité ATCase est représentée en vert.

4.2.2 Mise en évidence des insertions de rétrotransposons Ty1

Un rétrotransposon Ty1 peut s'insérer en amont de la région codant pour l'ATCase dans le gène *ura2_{15,30,72}*. Dans la plupart des cas de réactivation, l'insertion du rétrotransposon a lieu en sens inverse du sens transcriptionnel d'*ura2_{15,30,72}*. Dans ce cas, une séquence contenue dans la région LTR (Long Terminal Repeat) du rétrotransposon est alors utilisée comme séquence promotrice (Roelants *et al*, 1997). Dans un faible nombre de révertants, l'insertion se fait dans le même sens transcriptionnel. Dans ce cas de Figure, la séquence précise, utilisée comme région promotrice n'a pas encore été déterminée.

Si un rétrotransposon Ty1 est inséré en amont de la région codant pour l'ATCase, un fragment PCR pourra être amplifié en utilisant un oligonucléotide spécifique de la séquence LTR du rétrotransposon et un oligonucléotide spécifique de la séquence codante du domaine ATCase (Figure 1.11).

L'utilisation des deux jeux d'amorces (AW1,δB41 et AW1,δB41R) va permettre de vérifier l'événement d'insertion d'un rétrotransposon, de déterminer son sens d'insertion et de localiser approximativement son site d'insertion.

4.3 Analyse des événements de duplication

4.3.1 Localisation chromosomique des régions dupliquées

Lors d'un événement de duplication, la copie dupliquée peut s'insérer sur le même chromosome que l'allèle originel, on parle alors d'une duplication intrachromosomique. Lorsque la copie est située sur un autre chromosome, on définit la duplication comme interchromosomique.

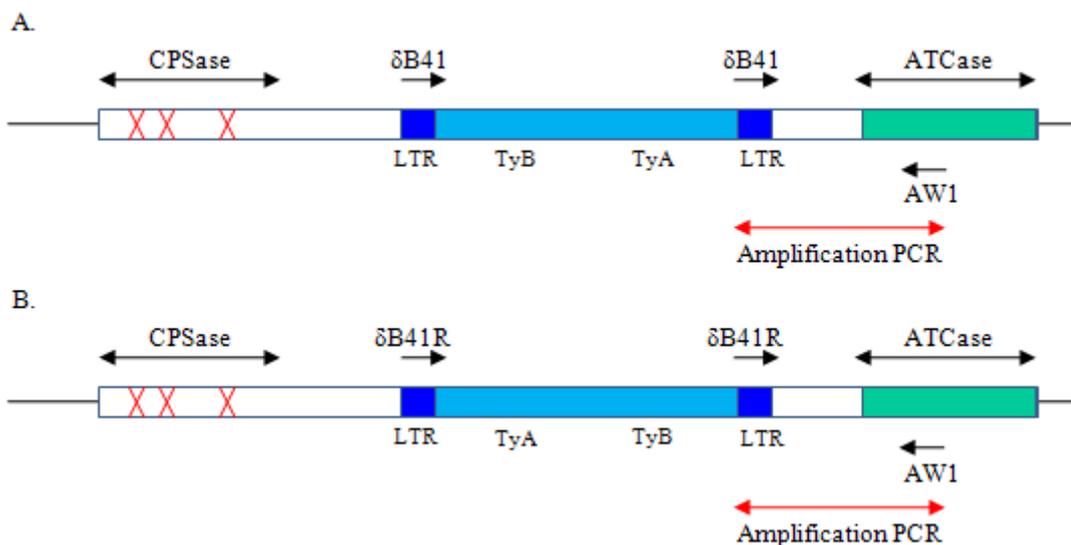


Figure 1.11 : Détection de l'insertion d'un rétrotransposon Ty1 dans le gène *URA2* par méthode PCR.

$\delta B41$ et $\delta B41R$ sont des oligonucléotides spécifiques du LTR de type δ du rétrotransposon Ty1.

AW1 est un oligonucléotide spécifique de la région codant pour l'ATCase.

A. Principe de détection dans le cas d'une insertion de transposon Ty1 en sens par rapport à la transcription du gène *ura2*_{15,30,72}.

B. Principe de détection dans le cas d'une insertion de transposon Ty1 en antisens par rapport à la transcription du gène *ura2*_{15,30,72}.

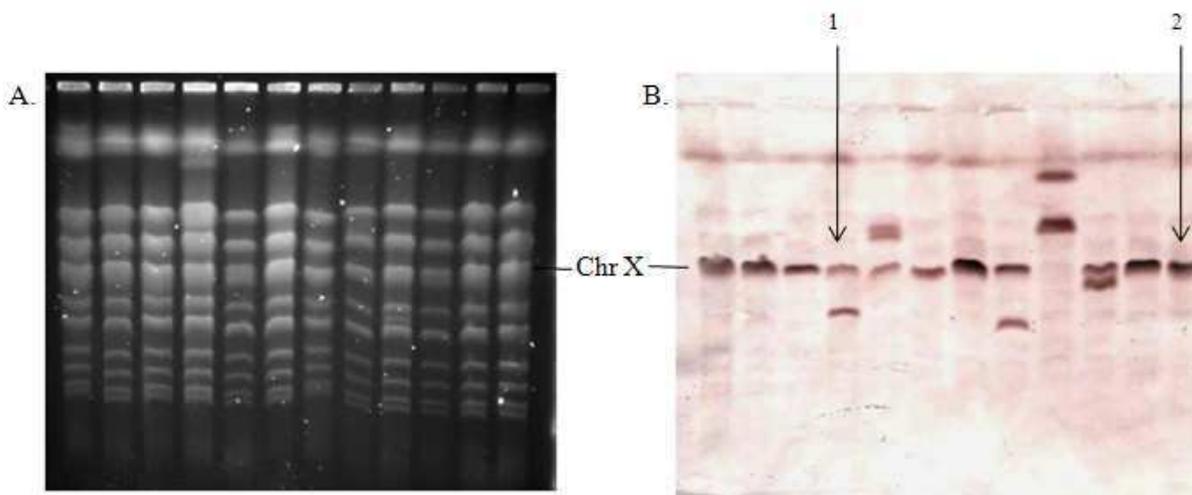


Figure 1.12 : Localisation chromosomique des régions dupliquées.

A. Séparation des chromosomes par électrophorèse en champs alternés (200V, 22 H, temps de pulsation : 40 sec jusqu'à 120 sec).

B. Hybridation ADN/ADN de type Southern avec une sonde spécifique de la région codant pour l'ATCase.

(1) Deux chromosomes sont mis en évidence dont le chromosome X, la copie dupliquée est située sur un autre chromosome, il s'agit d'une duplication interchromosomique.

(2) Seul le chromosome X est détecté, la copie s'est insérée sur le même chromosome que l'allèle original, on peut en conclure qu'il s'agit d'une duplication intrachromosomique.

Afin de caractériser plus précisément les duplications, le chromosome porteur de la copie dupliquée est déterminé en utilisant un caryotype électrophorétique obtenu par séparation des chromosomes entiers de levure par électrophorèse en champs alternés. On réalise une hybridation de type Southern sur les chromosomes ainsi séparés en utilisant la région ATCase comme sonde (Figure 1.12). Si seul le chromosome X est détecté, la copie dupliquée se situe alors sur ce chromosome et on peut conclure qu'il s'agit d'une duplication intrachromosomique. Si le chromosome X et un autre chromosome sont mis en évidence, la copie dupliquée s'est insérée sur un autre chromosome et il s'agit alors d'une duplication interchromosomique.

4.3.2 Analyse des duplications

Deux types de duplications ont également été caractérisés dans les travaux précédents (Schacherer *et al.*, 2004 ; Schacherer *et al.*, 2005), les duplications géniques et les duplications segmentales. Si elles sont géniques, seule la région codant pour l'ATCase est dupliquée. Dans le cas des duplications segmentales, la région codant pour l'ATCase ainsi que plusieurs gènes contigus sont codupliqués, dont notamment le gène *YJL131c* directement adjacent à *URA2* (Figure 1.13).

Pour déterminer la nature de ces duplications, nous avons effectué une analyse du profil de restriction *Bam*HI en utilisant une sonde correspondant à la région intergénique entre *URA2* et *YJL131c*. Le fragment de restriction *Bam*HI correspondant à la région codant pour l'ATCase comprend en plus de cette partie d'*URA2* la partie 5' du gène *YJL131c*.

Si la duplication est génique, seule la région codant pour l'ATCase est dupliquée. Ainsi l'analyse du profil *Bam*HI révèle une seule bande correspondant à la région intergénique entre l'allèle *ura2_{15,30,72}* originel et le gène *YJL131c*. Par contre, si la duplication est segmentale, le gène *YJL131c* ainsi que la région intergénique entre *YJL131c* et *URA2* sont codupliqués. Deux bandes sont alors mises en évidence, l'une correspondant à la région intergénique entre *YJL131c* et *URA2* et l'autre à la copie dupliquée de cette région à une autre position dans le génome.

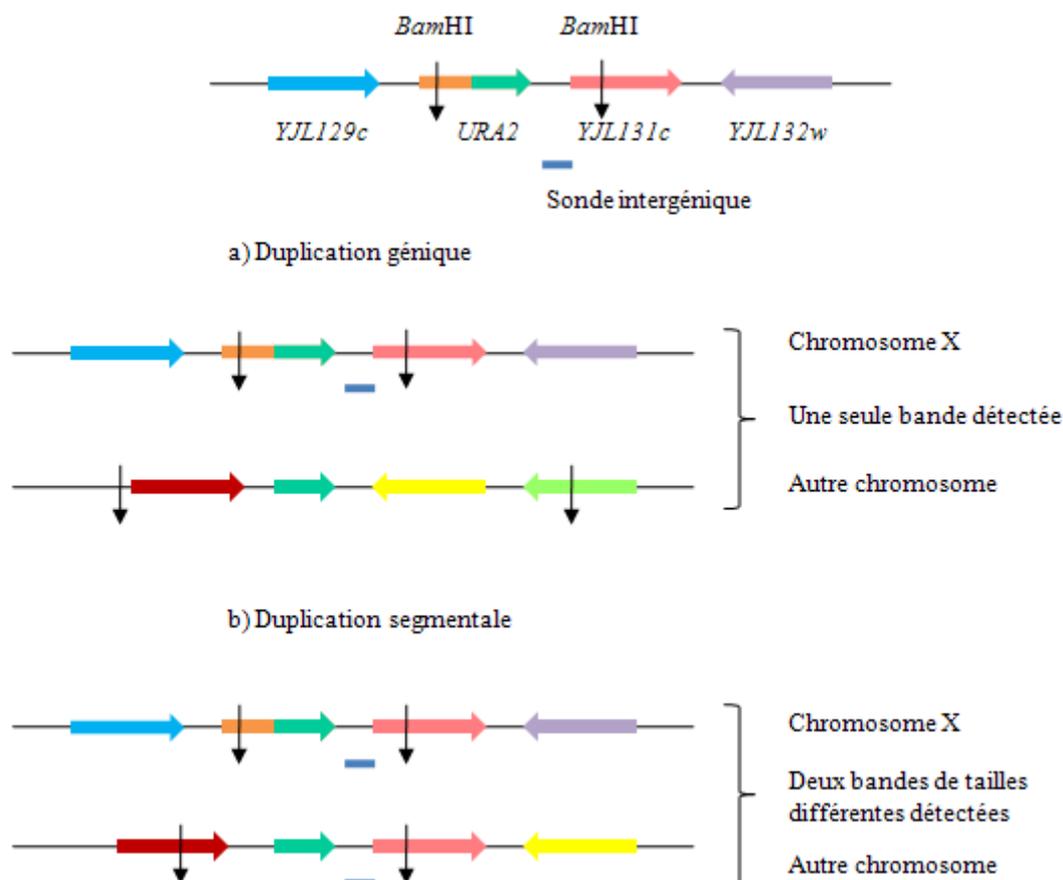


Figure 1.13 : Stratégie d'analyse des duplications géniques et segmentales par hybridation ADN/ADN de type Southern blot.

a) Dans le cas d'une duplication génique, une seule bande correspondant à l'allèle originel est obtenue en utilisant une sonde correspondant à la région intergénique entre *ura2*_{15,30,72}.

b) Dans le cas d'une duplication segmentale, deux bandes seront mises en évidence. Le profil *Bam*HI obtenu avec la sonde intergénique est alors identique à celui obtenu avec la sonde correspondant à la région codant pour l'ATCase.

5. Conclusion

Ce système de sélection performant ainsi que les outils d'analyse à notre disposition nous ont permis d'analyser l'impact de la variabilité intra-espèce sur l'apparition de remaniements chromosomiques. En effet, plusieurs souches de *Saccharomyces cerevisiae* ont été séquencées et il a été décrit que la variabilité génomique observée entre ces souches peut avoir de nombreuses conséquences phénotypiques. Pour étudier l'impact des variations intra-espèce quant à la fréquence de mutation et aux différents types de remaniements chromosomiques sélectionnés, deux souches différentes ont été comparées, la souche de référence, séquencée

S288c et la souche utilisée au laboratoire FL100. Les résultats obtenus seront décrits dans le chapitre 2.

Ce type d'approche a également été utilisé pour étudier les mécanismes d'apparition des remaniements chromosomiques en utilisant des contextes différents pour lesquels des gènes dont les produits sont impliqués dans les voies de recombinaison ont été délétés. L'influence de l'inactivation de ces gènes sur l'apparition des délétions et des duplications a ainsi pu être étudiée et sera présentée dans le chapitre 3.

2 INFLUENCE DU CONTEXTE GÉNÉTIQUE SUR L'APPARITION DES REMANIEMENTS CHROMOSOMIQUES

1. Introduction

La détermination de la séquence complète d'un seul individu de référence pour une espèce donnée permet d'étudier la variabilité inter-espèce mais ne permet pas d'analyser des différences au niveau génomique entre différents individus. Néanmoins, avec le séquençage d'un nombre toujours croissant de génomes, des expériences de génomique comparative sont à présent réalisées pour différents membres d'une même espèce. Ces variations intra-espèces ont ainsi été décrites dans plusieurs organismes comme *Streptococcus pneumoniae* (Hakenbeck *et al.*, 2001), *Oriza sativa* (Yang *et al.*, 2007), *Drosophila melanogaster* (Bosco *et al.*, 2007) ou encore chez l'Homme (Korbel *et al.*, 2007) par exemple. La levure *Saccharomyces cerevisiae* est un très bon modèle eucaryote pour l'étude de la variabilité génomique intra-espèces. En effet, de nombreuses souches d'origines différentes ont été caractérisées. Il peut s'agir de souches de laboratoire, de souches opportunistes isolées à partir de patients immunodéprimés ou encore de souches œnologiques. Plusieurs d'entre elles sont partiellement ou complètement séquencées. Des études utilisant des techniques d'hybridation comparative entre différentes souches de laboratoire ont été réalisées (Winzler *et al.*, 2002 ; Schacherer *et al.*, 2007) et ont montré que celles-ci présentent un certain nombre de variations nucléotidiques comme les SNPs (Single Nucleotide Polymorphism) ainsi que des délétions ou des inversions dont le nombre est variable en fonction des souches. Par exemple, la souche RM11 isolée à partir de plants de vigne, récemment séquencée présente 29508 SNPs et 85 délétions de plus de 200 pb par rapport à la souche de référence S288c (Broad Institute). Le séquençage complet d'une souche isolée à partir d'un patient immunodéprimé, nommée YJM789 (Wei *et al.*, 2007) a confirmé la présence d'un polymorphisme par rapport à la souche de référence S288c.

Cette variabilité n'est pas uniquement localisée dans les régions intergéniques mais peut également être trouvée dans les régions codantes. Ceci peut avoir des conséquences phénotypiques importantes. Il a, en effet, été montré que le polymorphisme observé pour les gènes *AMNI* et *GPAI* dans des souches de levure de laboratoire et naturelles affecte respectivement l'expression de gènes impliqués dans la séparation des cellules filles et dans la

réponse aux phéromones (Yvert *et al.*, 2003). Une autre étude a également démontré que des variations nucléotidiques intra-espèces au niveau des gènes *TAO3* ou *MTK1* pour différentes souches ont des conséquences sur leur efficacité de sporulation (Deutschbauer et Davis, 2005).

Au travers de ces études, il a été démontré que différentes souches de *S. cerevisiae* présentent des variations nucléotidiques ainsi que des modifications génomiques plus importantes suite à des remaniements chromosomiques conduisant à de possibles conséquences phénotypiques. Nous avons donc cherché à comprendre si la variabilité génomique intra-espèces pouvait avoir une influence sur l'apparition des remaniements chromosomiques. Pour aborder cet aspect, nous avons comparé deux souches de *Saccharomyces cerevisiae* utilisées en laboratoire quant à l'apparition de délétions, de duplications et d'insertions de rétrotransposon Ty1. Le choix de travailler avec deux souches de laboratoire repose sur des commodités expérimentales notamment sur la facilité à les manipuler (culture, croisement...). En utilisant le crible génétique *ura2_{15,30,72}* présenté dans le chapitre 1, des remaniements chromosomiques ont été sélectionnés dans les deux contextes. Il a ainsi été possible de comparer les types de réarrangements ainsi que les taux de mutation obtenus pour déterminer si des différences sont observées entre ces deux souches. Les deux contextes utilisés sont les suivants :

- Le contexte FL100

Cette souche est celle utilisée dans tous les travaux précédemment menés au laboratoire. Elle a été obtenue par François Lacroute (Lacroute, 1968).

- Le contexte S288c

La souche S288c a été isolée à partir de nombreux croisements par Robert Mortimer dans les années 50 (Mortimer et Johnston, 1986) dans le but d'obtenir une souche parentale permettant ensuite de construire des mutants pour des études biochimiques et génétiques (Figure 2.1). Cette souche a été construite pour répondre à plusieurs conditions : elle devait être non-flocculante et ne nécessiter qu'un nombre limité de nutriments pour sa croissance à savoir la biotine, une source d'azote, du glucose et différents sels minéraux.

Cette souche a été utilisée dans les années 60 pour la cartographie génétique de nombreux gènes puis pour le clonage et la caractérisation de plusieurs d'entre eux dans les années 80 (Hawthorne et Mortimer, 1960). De plus, cette souche a été choisie pour réaliser le

séquençage complet du génome de la levure (Goffeau *et al.*, 1996). Elle a également été utilisée pour la construction de la banque de délétions systématiques des gènes de *Saccharomyces cerevisiae* EUROSCARF (Brachmann *et al.*, 1998). Cette souche présente néanmoins l'inconvénient de difficilement sporuler.

Le choix de cette souche pour notre analyse comparative s'est donc facilement imposé puisqu'elle représente actuellement la souche de référence pour toutes les analyses de variabilité intra-espèce. De plus, expérimentalement, elle présente de nombreux avantages puisqu'elle est facilement cultivable et que des expériences de génétiques peuvent aisément être menées.

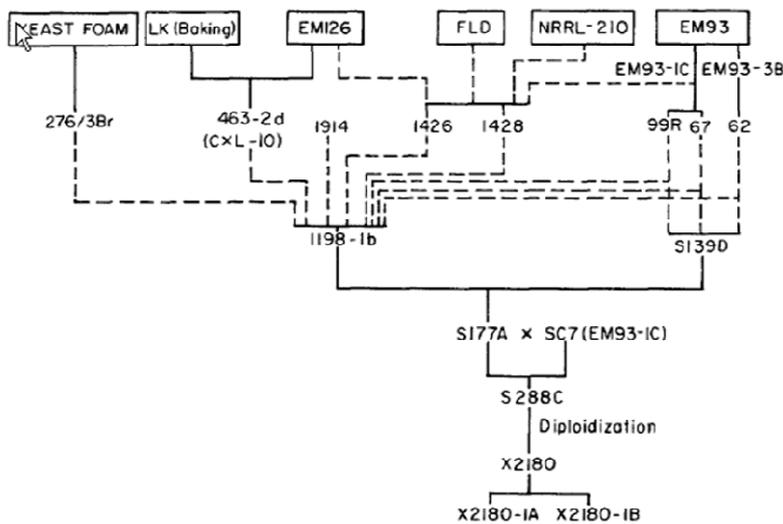


Figure 2.1 : Généalogie de la souche S288c (d'après Mortimer et Johnston, 1986).

Les traits pleins représentent les croisements directs entre souches tandis que les traits en pointillés indiquent que des croisements intermédiaires ont été réalisés mais ne sont pas décrits.

Les expériences d'hybridation comparative de ces deux souches (J. Schacherer, communications personnelles) ont permis de montrer que FL100 diffère de S288c par 22446 SNPs, ce qui représente une variation de 0,21% du génome de FL100 par rapport à S288c. De plus, 53 délétions de plus de 200 pb ont également pu être mises en évidence dans le génome de FL100 par rapport à celui de S288c. Les deux souches présentent donc des différences nucléotidiques qui pourraient avoir une influence sur la nature des remaniements et/ou sur les taux de mutation observés.

2. La variabilité intra-espèce a une influence sur les remaniements sélectionnés

2.1 Résultats observés en contexte FL100

Les études antérieures menées au laboratoire ont été réalisées en contexte FL100. Lors des sélections dans les souches haploïdes, rappelons que trois types d'événements ont été mis en évidence :

- L'insertion d'un rétrotransposon Ty1 (Roelants *et al.*, 1997)

Un élément transposable Ty1 s'insère en aval de la troisième mutation et en amont de la région codant l'ATCase en sens ou en antisens par rapport à la transcription du gène *URA2*. La transcription de la région codant pour l'ATCase est initiée à partir d'une séquence contenue dans la région LTR du rétrotransposon Ty1 la plus proche de la région codant pour l'ATCase.

- Une délétion de la région contenant les 3 mutations (Welcker *et al.*, 2000)

3 types de délétions peuvent être définies (Figure 2.2) : le premier type est une délétion interne au locus *ura2_{15,30,72}*. Dans ce cas, l'ATG initiateur reste en phase avec la région codant pour l'ATCase et est donc utilisé pour l'initiation de la traduction. C'est le type d'événements le plus fréquemment observé dans notre système de sélection.

Le deuxième type de délétions correspond à une perte de la région englobant les 3 mutations, l'ATG initiateur et la région promotrice. Cette dernière est composée du point d'initiation de la transcription en position -63 par rapport à l'ATG initiateur pour les transcrits majeurs et -44 pour les transcrits mineurs ainsi que d'une boîte TATAT en position -123 (Souciet *et al.*, 1989). Le codon d'initiation de la traduction ayant été perdu, on suppose qu'un ATG interne au locus *ura2_{15,30,72}* est alors utilisé. En revanche, la séquence utilisée comme nouvelle région promotrice pour l'initiation de la transcription est encore inconnue.

Enfin le dernier type de délétions regroupe les événements de fusion-délétion pour lesquelles l'allèle *ura2_{15,30,72}* délété est fusionné en phase avec un autre gène du chromosome X.

La présence du gène essentiel *GCD14* à 20 kb environ en amont du gène *URA2* limite la taille des délétions en contexte haploïde. Les plus longues délétions trouvées dans ce contexte ont une taille de 10 kb environ et correspondent à des fusions avec le gène *PBS2*. En contexte diploïde, en revanche, de plus longues délétions peuvent être obtenues, la plus longue sélectionnée s'étendant sur 128 kb.

Il a également été montré que de courtes séquences répétées dont la taille varie de 1 à 13 pb sont trouvées aux bornes de la plupart des délétions.

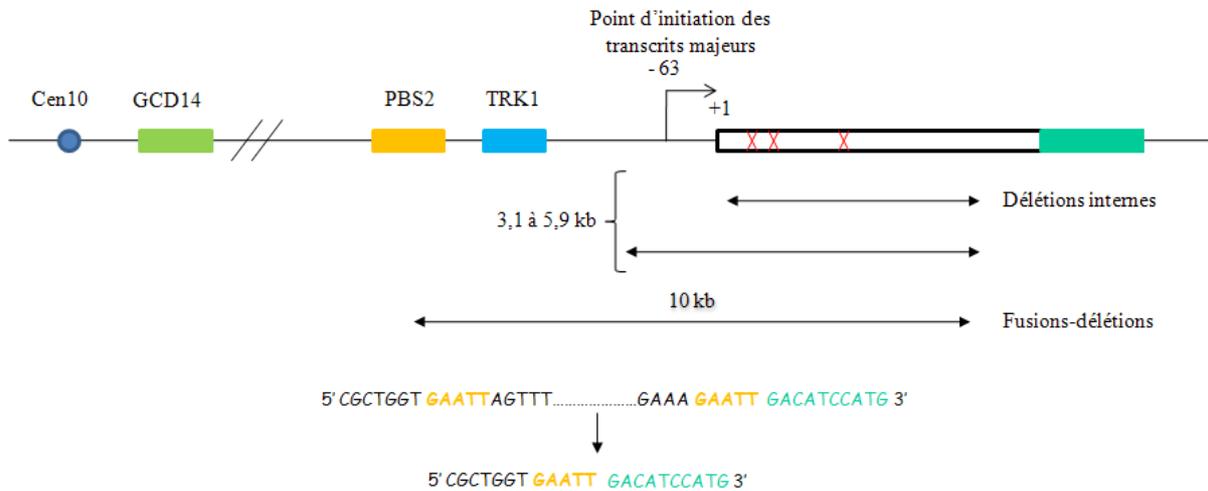


Figure 2.2 : Schéma récapitulatif des différents types de délétions sélectionnées en contexte FL100 haploïde et exemple de microhomologie observée aux bornes des délétions.

- Une duplication de la région codant pour l'ATCase suivie de la fusion de la copie avec une nouvelle région promotrice (Schacherer *et al.*, 2004)

Deux types de duplications ont été caractérisés : les duplications géniques pour lesquelles seule la région codant pour l'ATCase est dupliquée et les duplications segmentales de tailles variables pour lesquelles la région codant pour l'ATCase ainsi qu'un certain nombre de gènes contigus sont dupliqués en fonction de la taille de la duplication. Dans les études précédentes, la taille de région dupliquée pour la plus courte des duplications segmentales est de 6 kb correspondant à une duplication de la région codant pour l'ATCase ainsi que 3 gènes en aval. La plus longue région dupliquée s'étend sur 90 kb ce qui correspond à une duplication de la région s'étendant de la région codant pour l'ATCase au gène *YJL189w*.

Dans le contexte génétique FL100, la répartition des différents événements à l'origine de la réactivation de l'ATCase est de 1/6 de délétions, 1/6 de duplications et 2/3 d'insertions de rétrotransposons Ty1 (Roelants *et al.*, 1995).

Les taux de mutation global et relatifs à chaque type d'événements mutationnels ont été déterminés par la méthode de Lea et Coulson (1949) décrite dans le chapitre 1. Les différentes valeurs sont présentées dans le tableau 2.1.

2.2 Résultats des sélections en contexte S288c

Afin de comparer le taux de mutation global des remaniements chromosomiques ainsi que les types de remaniements observés dans les souches FL100 et S288c, des sélections ont été réalisées dans la souche aUHT (S288c) possédant l'allèle *ura2*_{15,30,72}. Les différents remaniements obtenus suite aux 155 sélections réalisées dans le contexte S288c sont présentées dans le tableau 2.2.

Souche	N	Nb de révertants indépendants	Type de remaniements obtenus		
			délétions	duplications	insertions Ty1
aUHT	155	29	20	9	0

Tableau 2.2 : Résultats des sélections effectuées dans la souche S288c aUHT.

N représente le nombre total de sélections effectuées.

29 révertants indépendants ont été obtenus à partir de $3,77 \times 10^{11}$ cellules étalées. La caractérisation moléculaire des révertants a été réalisée d'après les méthodes décrites dans le chapitre 1. L'analyse du profil *Bam*HI par hybridation ADN/ADN de type Southern a révélé deux bandes pour 9 révertants, une de 6,8 kb correspondant à l'allèle *ura2*_{15,30,72} et une bande de taille variable. Une seule bande de taille différente de celle de l'allèle muté a été observée pour les 20 autres révertants. A partir de ces profils, nous avons ainsi déterminé que parmi les 29 révertants sélectionnés, 9 sont porteurs d'une duplication et 20 d'une délétion ou d'une insertion d'un rétrotransposon Ty1. L'analyse par PCR selon le principe décrit dans le chapitre 1 nous a permis de discriminer ces deux types d'événements. Aucune amplification n'a été obtenue lors de la recherche d'insertions de rétrotransposon Ty1. Nous avons donc pu conclure que pour les 20 révertants, il s'agit de délétions.

Le taux de mutation global ainsi que pour celui pour chaque type d'événements a été déterminé pour cette souche et les résultats sont présentés dans le tableau 2.1.

Souches	N	$n_c (\times 10^9)$	Taux de mutation		Proportion de chaque type de réarrangements			Origine
			global ($\times 10^{-10}$)	$\times 10^{-10}$	Délétions	Duplications	Insertions de Tyl	
FL100 AW1	333	2,65	0,671 (0,173;1,169)	17%	17%	66%	Tourrette Y.	
S288C aUHT	155	2,43	1,025 (0,677;1,373)	69%	31%	0	ce travail	
			0,112 (0,029;0,195)	0,112 (0,029;0,195)	0,317 (0,209;0,426)	0,447 (0,115;0,779)	-	

Tableau 2.1 : Récapitulatif des taux de mutation globaux et relatifs à chaque événement en contexte FL100 et S288c.

Borne 5'	Borne 3'	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
-626	+3346	3972	G	1	+3352	1097	aUHT198-1
+56	+4005	3948	GGTGA	5	+1	878	aUHT01-1
+50	+3999	3948	TCTAC	5	+1	898	aUHT121-1
+56	+4005	3948	GGTGA	5	+1	878	aUHT123-1
+138	+4006	3267	CTGGTGAA	8	+1	925	aUHT153-1
+138	+4006	3267	CTGGTGAA	8	+1	925	aUHT158-1
+138	+4006	3267	CTGGTGAA	8	+1	925	aUHT24-1
+196	+3416	3219	CCATCTTATG	11	+1	1141	aUHT149-1
-512	3347	3858	AAAGAATTGA	11	+3352	1097	aUHT156-1
-512	3347	3858	AAAGAATTGA	11	+3352	1097	aUHT163-1
-512	3347	3858	AAAGAATTGA	11	+3352	1097	aUHT190-1

Tableau 2.3 : Récapitulatif des différentes délétions et des microhomologies observées dans le contexte aUHT.

En bleu sont présentées les délétions internes au locus *ura2*_{15,30,72} et en orange les délétions plus longues incluant l'ATG initiateur et/ou la région promotrice du gène.

2.2.1 Analyse des événements de délétion

Pour 11 délétions parmi les 20 obtenues, les séquences aux bornes de la délétion ont pu être déterminées par séquençage du produit d'amplification par PCR. Les résultats sont présentés dans le tableau 2.3.

De courtes séquences répétées dont la taille varie de 1 à 11 pb ont été mises en évidence pour les 11 délétions. Pour 7 révertants, les délétions observées sont internes au locus *ura2*_{15,30,72} et encadrent la région contenant les trois mutations. Dans toutes les souches, la région codant pour le fragment ATCase se trouve en phase avec l'ATG initiateur qui peut donc vraisemblablement être utilisé pour l'initiation de la traduction (Figure 2.3).

Pour les 4 autres révertants, il s'agit de délétions plus longues englobant la région contenant les trois mutations ainsi que la région promotrice et le codon initiateur. Le codon d'initiation de la traduction étant perdu, un codon ATG interne est alors probablement utilisé. La position de l'ATG initiateur a été déterminée *in silico* et reste donc putative. Le premier ATG trouvé en phase a été choisi. Il pourrait donc s'agir de l'ATG en position +3352 dans les 4 cas sans cependant exclure d'autres positions possibles. Une analyse *in vivo* du point d'initiation de la transcription ainsi que le séquençage des premiers acides aminés de la protéine ainsi produite permettraient de connaître précisément le site d'initiation de la transcription et l'ATG utilisé pour la traduction.

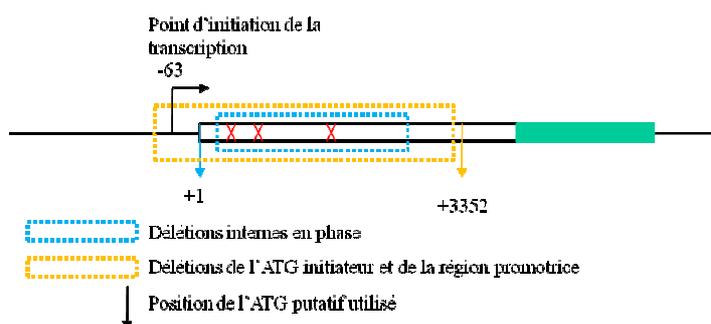


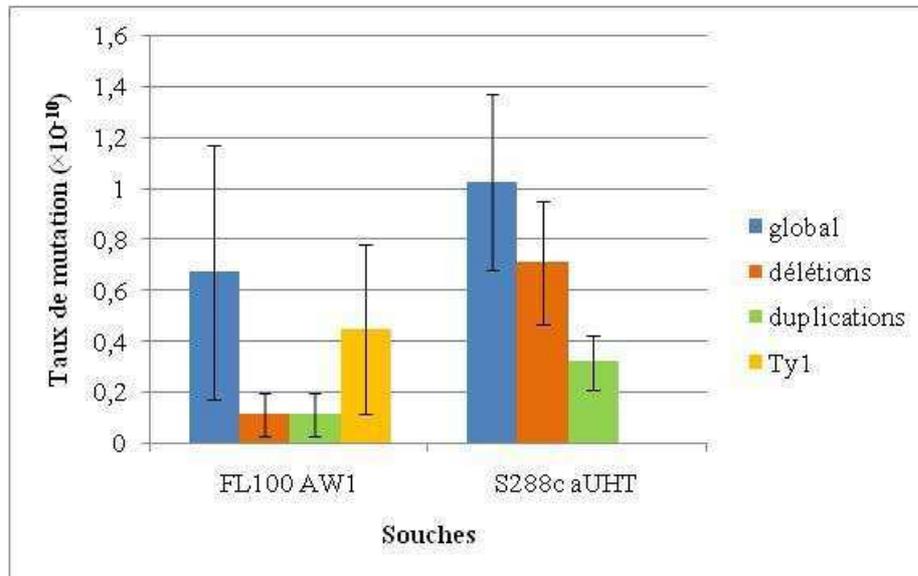
Figure 2.3 : Schéma représentatif des délétions obtenues en contexte S288c et des ATG initiateurs putatifs.

Pour les 9 autres délétions, aucun fragment PCR n'a pu être obtenu. Une explication possible serait que la délétion soit plus longue en 5' que celles pouvant être amplifiées avec les oligonucléotides amorces utilisés, des expériences en utilisant des amorces en amont seraient donc nécessaires pour caractériser les 9 délétions restantes sachant que la

borne de la délétion en 5' est limitée à 20 kb en amont du gène *ura2_{15,30,72}* du fait de la présence du gène essentiel *GCD14*.

2.3 Comparaison des résultats obtenus en contextes FL100 et S288c

Un récapitulatif des différents événements ainsi que des taux de mutation calculés est présenté dans le tableau 2.1 et dans le graphique 2.1.



Graphique 2.1 : Représentation des taux de mutation globaux et relatifs aux différents réarrangements pour les contextes FL100 et S288c. Les intervalles de confiance à 95% sont représentés par les barres d'erreur.

Aucune différence significative du taux global de mutation ne peut être mise en évidence entre les deux contextes. En effet, pour que des taux de mutation soient considérés comme statistiquement différents, il est nécessaire que les intervalles de confiance entre les deux taux soient non-chevauchants. Dans notre cas les intervalles de confiance se recouvrent, on peut donc conclure que les taux d'apparition globaux sont similaires.

En revanche, une augmentation du taux de délétions et de duplications est observée en contexte S288c. Ceci peut être expliqué par un déséquilibre au niveau du nombre relatif des événements observés. En effet, la proportion de délétions et de duplications est plus élevée du fait de l'absence de réactivation par insertion de rétrotransposons Ty1.

Cette absence est particulièrement remarquable car la réactivation de l'activité ATCase par insertion de rétrotransposon Ty1 est un événement majeur en contexte FL100 puisqu'il représente 2/3 des remaniements sélectionnés.

Au moins deux hypothèses peuvent être formulées pour expliquer cette différence entre les deux souches.

La première hypothèse serait une différence au niveau du nombre de Ty1 entre la souche S288c et la souche FL100. Des variations au niveau du nombre et de la localisation des éléments transposables ont été décrites dans plusieurs souches de *S. cerevisiae*. Il a été montré par exemple que la souche RM11 ne possède aucune copie de rétrotransposon Ty1 (Gabriel *et al.*, 2006). De plus, des études comparatives réalisées pour le groupe des *Saccharomyces sensu stricto* montrent aussi une variabilité intra et inter-espèces du contenu en Ty1 (Liti *et al.*, 2005). Une souche de *Saccharomyces paradoxus* présentant une absence totale de rétrotransposons Ty1, Ty2, Ty4 et Ty5 a été caractérisée (Moore *et al.*, 2004). La souche S288c séquencée en 1996 (Goffeau *et al.*, 1996), possède 31 rétrotransposons Ty1 (Kim *et al.*, 1998). Pour FL100, le nombre et la localisation exacts des éléments Ty1 ne sont pas connus, néanmoins la proportion de Ty1 entre les souches S288c et FL100 semble être la même (Schacherer, communication personnelle). Suite à cette observation, on peut donc conclure que la variation observée entre les deux souches quant à l'absence de réactivation par insertion de Ty1 peut difficilement être expliquée par une variation du contenu en rétrotransposons Ty1.

Le nombre d'éléments Ty1 et leur localisation chromosomique dans la souche FL100 peuvent être déterminés en utilisant la technique d'hybridation comparative mise au point par Gabriel *et al.* (2006) (Figure 2.4). Cette technique repose sur le fait que même si les membres d'une même famille de transposons sont très conservés, leurs régions flanquantes sont, quant à elles, uniques. Ainsi la position des transposons est définie par identification des séquences contiguës. Les fragments d'ADN contenant les séquences flanquant les différents rétrotransposons sont isolés, marqués avec deux fluorochromes différents et hybridés sur des puces à ADN représentant le génome complet de la levure à l'exception des séquences de rétrotransposons et des séquences répétées. Par analyse de la fluorescence obtenue, il est alors possible de préciser le nombre et la localisation des différents éléments transposables. De plus, cette technique permet de comparer deux souches en cohybridant les deux ADN génomiques. Une variante de ce système, appelée TIP-chip (transposon insertion site profiling chip) a été développée par Wheelan *et al.* (2006) permettant également de déterminer le contenu et la position dans le génome des éléments Ty1. Cependant, seul le séquençage de la souche FL100 nous permettra de connaître ces informations avec exactitude.

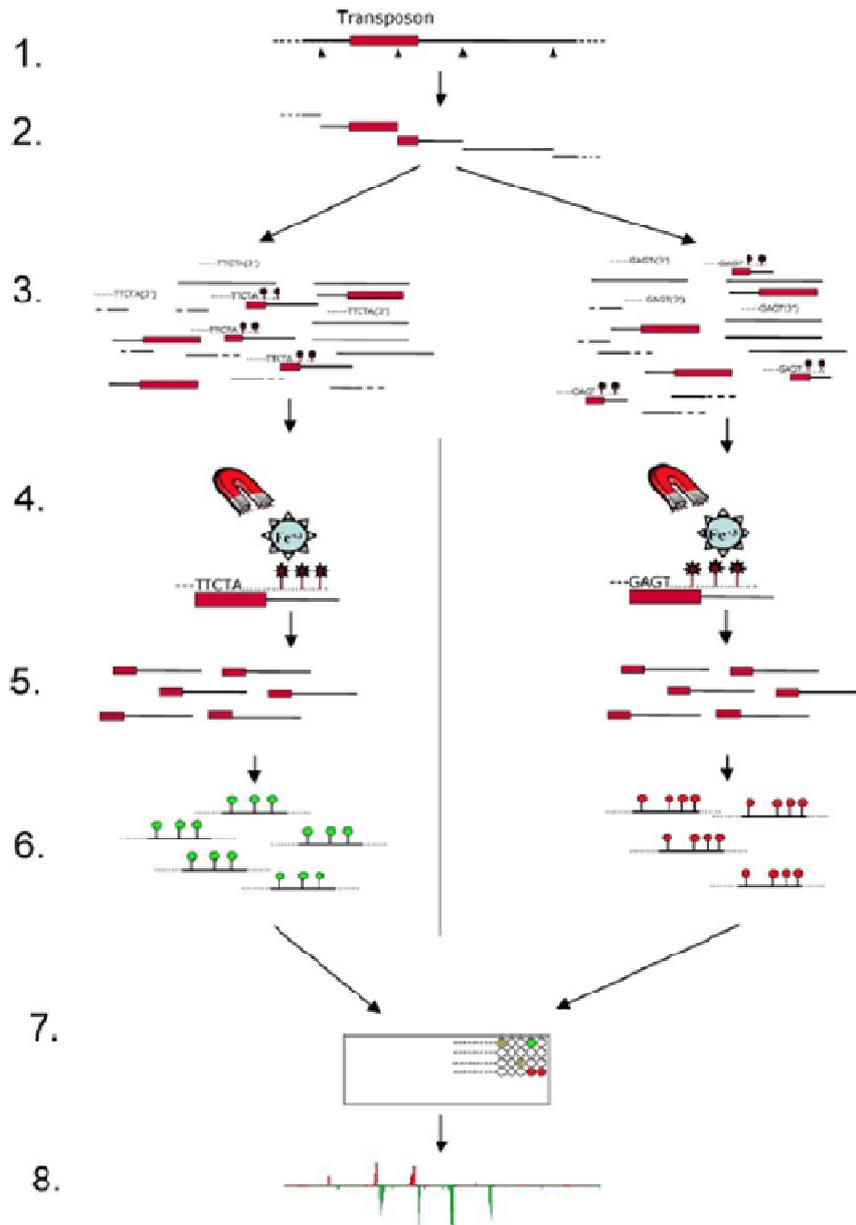


Figure 2.4 : Représentation des différentes étapes pour déterminer le nombre et la localisation chromosomique des éléments transposables dans un génome (d'après Gabriel *et al.*, 2006).

1. L'ADN génomique de la souche d'intérêt est digéré par des enzymes de restriction (triangles noirs)
2. L'ADN digéré est ensuite mélangé aux sondes s'hybridant spécifiquement à certaines séquences de rétrotransposons.
3. Le mélange est incubé en présence d'ADN polymérase et de dNTPS dont un est biotinyllé. Cette étape permet l'amplification spécifique des régions flanquantes situées en 3' de la séquence du transposon.
4. Les fragments biotinyllés sont purifiés en utilisant des billes magnétiques couplées à la streptavidine.
- 5, 6. Les fragments sont ensuite libérés des billes et marqués à la cyanine 5 ou 3 par la technique de random priming.
- 7, 8. Les ADN marqués sont ensuite hybridés sur une puce à ADN représentant le génome complet de la levure. L'analyse de la fluorescence permet ensuite de déterminer la position et le contenu en rétrotransposons dans la souche d'intérêt.

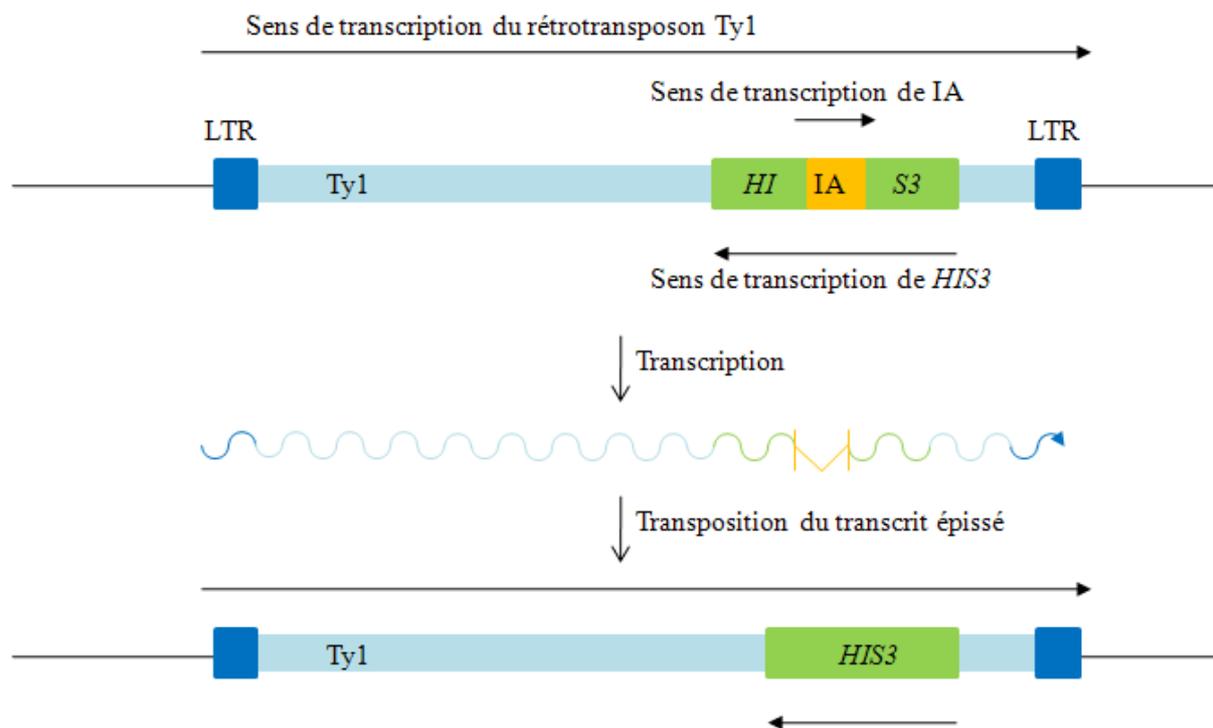


Figure 2.5 : Représentation du système de sélection utilisé par Curcio et Garfinkel (1991) pour mesurer la rétrotransposition.

La deuxième hypothèse formulée serait une variation de l'activité des rétrotransposons entre les deux souches. Le taux de rétrotransposition pourrait être calculé dans les deux contextes en utilisant un rétrotransposon contenant un gène marqueur (Curcio et Garfinkel, 1991) (Figure 2.5). Le gène *HIS3* interrompu par un intron artificiel (IA) en orientation antisens a servi de marqueur. Le segment *HIS3AI* a été introduit dans un élément Ty1 de telle manière que l'intron soit en orientation sens par rapport à l'élément transposable. Lors de la rétrotransposition, l'ARNm est mûré et l'intron est éliminé. La copie de l'élément transposable insérée dans le génome contiendra le gène *HIS3* sans intron. On observera alors l'apparition de révertants His^+ à partir d'une souche auxotrophe pour l'histidine. L'activité de rétrotransposition est donc directement corrélée au nombre de révertants His^+ et le taux de réversion est donc lié au taux de rétrotransposition. La comparaison des taux de rétrotransposition en contexte FL100 et S288c permettrait de déterminer ainsi si une différence d'activité des rétrotransposons existe entre les deux souches.

Nous avons choisi une méthode plus indirecte mais plus rapide car ne nécessitant pas de construction complexe pour déterminer si les contextes FL100 et S288c présentent une différence d'activité des rétrotransposons. Cette méthode repose sur le fait que la transposition

est favorisée en travaillant à une température inférieure à la température optimale de croissance de *Saccharomyces cerevisiae*. Cette expérience sera décrite plus en détails dans le paragraphe suivant.

3. L'activité des rétrotransposons est plus faible en contexte S288c

3.1 Test de l'activité des rétrotransposons

Paquin et Williamson (1984) ont mis au point une méthode leur permettant de détecter la transposition des éléments Ty1 basée sur la capacité de ces éléments à changer l'expression des gènes adjacents à leur site d'insertion. Le principe du système de sélection utilisé dans cette étude repose sur la sélection de révertants pour lesquels il y a eu perte de la régulation et donc expression constitutive du gène *ADH2* codant pour une alcool-déshydrogénase chez *S. cerevisiae*.

En analysant les événements à l'origine de cette activité constitutive, ils ont constaté que dans 80% des cas, un élément Ty1 était inséré en amont du gène. A partir de ce crible génétique, le taux d'activité des transposons Ty1 a été déterminé à différentes températures de croissance et il a été montré que la température joue un rôle sur l'activité des transposons. En effet, à une température inférieure à 30°C, température optimale de croissance pour *Saccharomyces cerevisiae*, la transposition est accrue.

En partant de cette observation, nous avons testé l'activité des rétrotransposons Ty1 en contexte S288c et pour ce faire, des sélections de remaniements chromosomiques ont été réalisées à la température de 25°C. En effet, si l'absence d'insertion de rétrotransposons Ty1 observée à 30°C est due à une diminution d'activité pour la souche S288c par rapport à la souche FL100, le fait d'augmenter cette activité bien qu'indirectement devrait nous permettre, malgré tout, de sélectionner ce type d'événement.

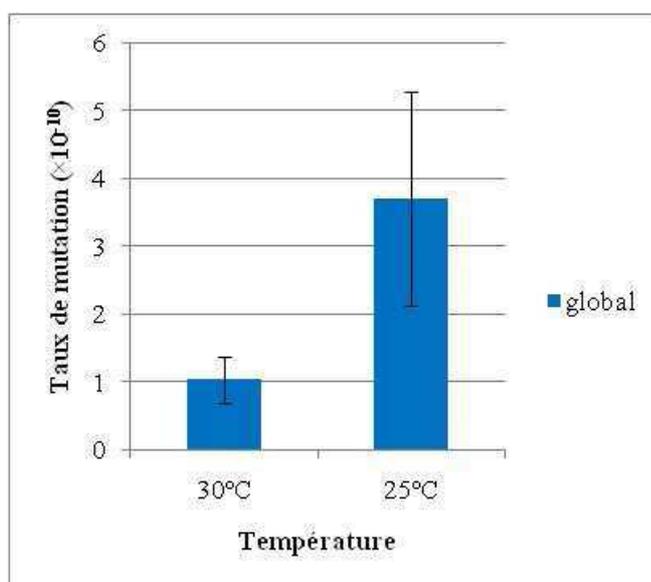
3.2 Résultats des sélections réalisées à 25°C

A cette température, 24 sélections, correspondant à $1,04 \times 10^{11}$ cellules étalées, ont été réalisées et 16 révertants indépendants ont ainsi été obtenus. Tous ces révertants ont été testés par PCR pour l'insertion d'un rétrotransposon Ty1 au niveau du locus *ura2*_{15,30,72}. Quatre d'entre eux présentent une insertion soit 25% des événements totaux. Les révertants porteurs d'autres types d'événements n'ont pas été analysés davantage.

Dans ces conditions, le taux de mutation global est de $3,70 \times 10^{-10}$ mutation/cellule/sélection (IC : $2,13 \times 10^{-10}$; $5,27 \times 10^{-10}$) à cette température. Le taux d'insertion des rétrotransposons Ty1 est de $0,925 \times 10^{-10}$ mutation/cellule/sélection (IC : $0,532 \times 10^{-10}$; $1,318 \times 10^{-10}$).

3.3 Comparaison des résultats obtenus à 25 et à 30°C

La comparaison des taux de mutation globaux a permis de montrer qu'il existe une différence statistique entre le taux de mutation à 25°C et celui à 30°C, les intervalles de confiance étant non-chevauchants. En effet, il est 3,6 fois plus important à 25°C qu'à 30°C (Graphique 2.2).



Graphique 2.2 : Représentation des taux de mutation pour la souche S288c aUHT à 30°C et à 25°C. Les barres d'erreur représentent l'intervalle de confiance à 95%.

Le fait d'obtenir 4 réactivations par insertion de rétrotransposon Ty1 à 25°C alors qu'aucune n'était observée à 30°C nous permet de conclure que les rétrotransposons sont actifs en contexte S288c mais que leur activité est plus faible que celle observée en contexte FL100.

4. Discussion

Dans ce travail, nous nous sommes intéressés à l'influence que peut avoir la variabilité intra-espèce sur l'apparition des remaniements chromosomiques. Pour ce faire, nous avons

sélectionné des révertants dans deux contextes génétiques différents, FL100 et S288c. L'analyse de ces révertants nous a permis de déterminer les différents types de remaniements chromosomiques ainsi que les taux de mutation obtenus.

La comparaison des taux de mutation globaux obtenus dans les différents contextes a montré qu'il n'existe aucune différence statistique pour ces taux à 30°C. Ce résultat nous permet de conclure que la variabilité génomique observée entre les deux souches n'a pas d'impact sur le taux global de mutation.

Une absence de réactivation de l'ATCase par insertion de rétrotransposon Ty1 a été observée à 30°C dans le contexte S288c. Ce résultat est remarquable puisque ce type de remaniement représente 2/3 des événements sélectionnés en contexte FL100. Lors de ce travail, nous avons testé l'activité des rétrotransposons par une méthode indirecte. Cette dernière repose sur le fait que la transposition est favorisée à une température inférieure à la température optimale de croissance pour *S. cerevisiae* (Paquin et Williamson, 1984). Des sélections ont été réalisées à 25°C et des révertants porteurs d'insertions de Ty1 ont été obtenus à cette température. Ce résultat nous permet d'établir que les rétrotransposons sont actifs dans la souche S288c mais que leur activité est plus faible que celle observée en contexte FL100.

Cette différence d'activité pourrait être expliquée par une différence de régulation des rétrotransposons dans ces deux souches. En effet, chez *S. cerevisiae*, deux types de régulation qui limitent la rétrotransposition ont été décrits : une régulation transcriptionnelle (Jiang, 2002) et une régulation post-transcriptionnelle (Garfinkel *et al.*, 2003), toutes deux dépendantes du nombre de copies d'éléments Ty1 dans la cellule. La régulation transcriptionnelle ou TGS (transcriptional gene silencing) sert de mécanisme de contrôle négatif régulant la transcription des rétrotransposons Ty1. Chez la levure, ces mécanismes sont encore mal connus. En revanche, chez les plantes, ce mécanisme passe généralement par une méthylation de l'ADN conduisant à l'arrêt de la transcription de la région. La régulation post-transcriptionnelle ou PTGS (post-transcriptional gene silencing), encore appelée RNAi (RNA interference), quant à elle, n'influe pas sur la transcription et n'a donc pas d'impact sur l'accumulation des ARNm. Chez les eucaryotes, le RNAi passe par une étape de clivage de l'ARNm par une RNase nommée Dicer (Almeida et Allshire, 2005). Cependant, la levure *Saccharomyces cerevisiae* ne possède pas cette voie de régulation, le gène codant pour Dicer notamment étant absent. Le mécanisme de régulation post-transcriptionnelle chez la levure reste actuellement encore méconnu. Une analyse plus approfondie de la régulation de

l'activité des rétrotransposons dans la souche FL100 est donc nécessaire pour compléter cette étude.

Dans le cas des délétions, on observe une augmentation du taux de délétions en contexte S288c par rapport à FL100. Celle-ci peut être expliquée par un déséquilibre des proportions des différents événements suite à l'absence des insertions de rétrotransposons Ty1. On retrouve les mêmes catégories de délétions qu'en contexte FL100 ainsi que la présence de microhomologies aux bornes ce qui nous permet de conclure que le polymorphisme existant entre les deux contextes n'a pas d'impact sur le taux ni sur les mécanismes d'apparition des délétions.

Pour les duplications, on observe comme dans le cas des délétions, une augmentation du taux de duplications pour le contexte S288c. La même explication que précédemment pour expliquer le cas des délétions peut être émise. Néanmoins, la diminution d'activité des rétrotransposons Ty1 dans la souche S288c par rapport à la souche FL100 pourrait avoir des conséquences sur le type de duplications sélectionnées. Il a, en effet, été proposé que le mécanisme de rétroposition est à l'origine de l'apparition des duplications géniques (Schacherer *et al.*, 2004) et ce mécanisme est lié à la rétrotransposition (Figure 2.6). Selon ce mécanisme, l'ARNm de l'allèle *ura2*_{15,30,72}, une fois dans le cytoplasme, se trouverait encapsidé dans les VLPs (virus-like particules) avec les ARNm d'éléments Ty1. La reverse transcription serait initiée alors en aval de la région codant pour l'ATCase au niveau de la queue polyA. Puis il y aurait un changement de matrice entraînant la synthèse d'un ADNc de fusion entre la région codant pour l'ATCase et une partie du rétrotransposon Ty1. Cet ADNc de fusion serait ensuite intégré dans le génome ce qui permettrait la réactivation de l'activité ATCase.

On peut donc émettre l'hypothèse que la diminution d'activité des rétrotransposons Ty1 entraînera également une diminution du mécanisme de rétroposition et que les duplications ainsi obtenues en contexte S288c devraient être majoritairement segmentales puisque leur mécanisme d'apparition est indépendant de la rétrotransposition. Pour confirmer cette hypothèse, la caractérisation moléculaire des duplications sélectionnées dans la souche S288C aUHT pourra être menée.

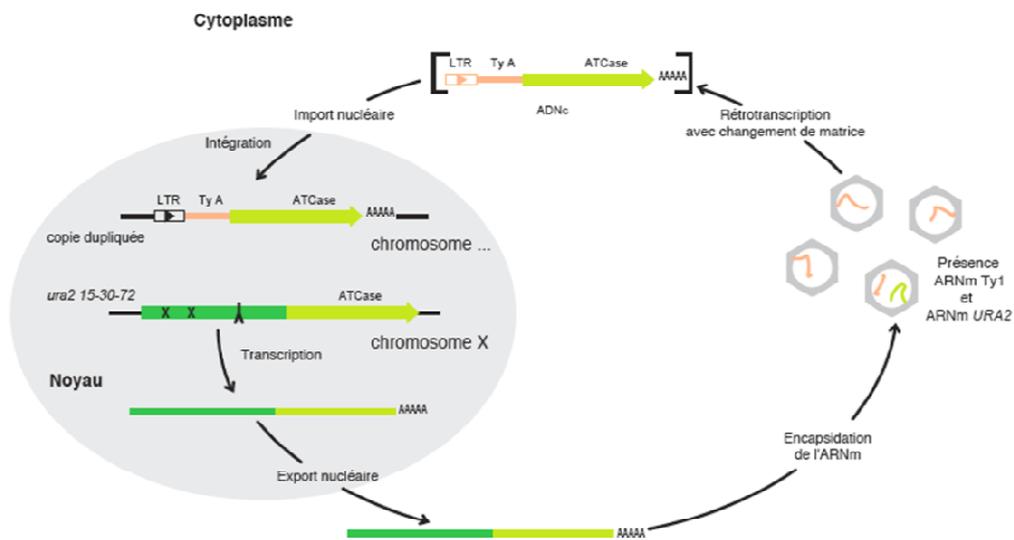


Figure 2.6 : Schéma récapitulatif du mécanisme de rétroposition (d'après Schacherer, 2005)

Lors de cette étude, nous avons montré qu'une variabilité intra-espèce au niveau de l'apparition des réarrangements existe entre la souche S288c et la souche FL100 et qu'elle est due à une différence d'activité des rétrotransposons. Afin d'approfondir cette étude, il serait intéressant de connaître le nombre et la localisation exacte des rétrotransposons Ty1 dans la souche FL100 ainsi que d'étudier les facteurs à l'origine de la différence d'activité des éléments mobiles entre les deux souches. De plus, d'autres souches de *S. cerevisiae*, comme la souche RM11 ne contenant pas de Ty1, pourraient être testées quant à l'apparition des remaniements chromosomiques afin de déterminer si la variabilité observée concerne uniquement les rétrotransposons ou si des différences peuvent également être détectées pour les autres types de remaniements comme les délétions ou les duplications.

3 MÉCANISMES D'APPARITION DES DÉLÉTIONS ET DES DUPLICATIONS CHROMOSOMIQUES

1. Introduction

Chaque organisme dispose de mécanismes de recombinaison qui lui permettent de maintenir l'intégrité du génome lorsqu'ils sont impliqués dans la réparation des cassures d'ADN mais également d'assurer une diversité génétique. Ces mécanismes peuvent être séparés en deux catégories. La première regroupe l'ensemble des mécanismes associés à la recombinaison homologue. Quatre mécanismes peuvent être distingués : le double-strand break repair model (DSBR), le synthesis-dependent strand annealing (SDSA), le mécanisme de break-induced replication (BIR) et enfin le single-strand annealing (SSA). Ces mécanismes ont été décrits dans l'introduction générale. Ils font tous intervenir des protéines codées par les gènes du groupe d'épistasie *RAD52* et nécessitent la présence de régions homologues pour pouvoir réparer les cassures d'ADN double-brin (Introduction Figure I7).

La deuxième catégorie englobe les mécanismes de recombinaison non-homologue dont le principal est le non-homologous end joining (NHEJ). Celui-ci peut fonctionner en présence de courtes homologies de séquence (moins de 10 pb) ou en leur absence.

Dans ce travail, nous nous sommes attachés à déterminer les mécanismes conduisant à la formation des délétions et des duplications. En utilisant le crible de sélection *ura2_{15,30,72}*, il a été montré que les délétions peuvent être géniques c'est-à-dire que seule la région contenant les trois mutations est délétée, ou segmentales pour lesquelles la séquence contenant les trois mutations ainsi qu'une région variable en amont du gène *URA2* est perdue. Pour la plupart des délétions, de courtes microhomologies dont la taille varie de 1 à 11 pb, ont été mises en évidence. Des études antérieures ont suggéré que le mécanisme d'apparition des délétions est dépendant de la taille des microhomologies (Daley et Wilson, 2005). Il a ainsi été proposé que le NHEJ est impliqué dans la genèse des délétions sans microhomologies aux bornes ou en présence de très courtes microhomologies de 1 à 4 pb. Un mécanisme de type SSA-like impliquant Rad59p et Rad52p serait à l'origine de l'apparition des délétions dès que la taille des séquences homologues est supérieure à 10 pb. Enfin un mécanisme indépendant du SSA et du NHEJ nécessitant la ligase Cdc9p permettrait l'apparition de délétions présentant des microhomologies variant de 5 à 9 pb. D'autres études ont également proposé que le

mécanisme de NHEJ serait à l'origine de l'apparition des délétions en utilisant de courtes séquences répétées (7 à 22 pb) (Moore et Haber, 1996 ; Yu et Gabriel, 2003).

Deux types de duplications ont également été caractérisés, en utilisant le système de sélection *ura2_{15,30,72}*. Les premières sont les duplications géniques pour lesquelles seule la région codant pour l'ATCase est dupliquée puis intégrée dans le génome. Le mécanisme d'apparition proposé est la rétroposition (Schacherer *et al.*, 2004). Dans ce cas, un ADN de fusion formé de la région codant pour l'ATCase et d'une partie d'un rétrotransposon Ty1, est synthétisé par rétrotranscription à partir des deux ARNs messagers contenus dans une Virus-Like Particule (VLP). Cet ADN est ensuite intégré dans le génome de la levure.

Le deuxième type de duplications correspond aux duplications segmentales pour lesquelles la région codant pour l'ATCase est codupliquée avec un certain nombre de gènes localisés en aval. Ces segments dupliqués ont toujours une localisation intrachromosomique et sont en tandem direct avec la région originelle. Les duplications segmentales décrites dans notre système présentent toutes des microhomologies dont la taille varie de 1 à 11 pb au niveau de la borne 5' du segment dupliqué. Il a été proposé que le mécanisme de BIR serait à l'origine de l'apparition des telles duplications (Kozsul *et al.*, 2004). D'autres études suggèrent que le NHEJ pourrait être à l'origine de l'apparition de ces duplications en contexte diploïde (Schacherer *et al.*, 2007).

Afin de mettre en évidence les différents mécanismes nécessaires à l'apparition des remaniements chromosomiques, nous avons construit différentes souches *ura2_{15,30,72}* portant des mutations dans différents gènes codant pour des protéines impliquées dans la recombinaison homologue (HR) (Rad52p ou Rad59p) ou dans le NHEJ (Lig4p ou Yku80p). Dans ces contextes modifiés, des révertants ont été sélectionnés et analysés. Les résultats obtenus, concernant les taux de mutation et les types de remaniements, ont été comparés à ceux obtenus dans la souche de référence aUHT dans le but de déterminer l'implication des mécanismes de HR ou de NHEJ dans la formation des délétions et des duplications.

2. Influence des mécanismes de recombinaison homologue

L'ensemble des sélections a été réalisé en contexte S288c et à 30°C. La construction des différentes souches délétées pour les gènes, dont le produit est impliqué dans l'HR ou le NHEJ, est présentée dans le chapitre Matériel et Méthodes.

2.1 Résultats obtenus pour la souche aUHT sauvage

Les résultats obtenus pour cette souche ont été détaillés dans le chapitre 2. Ils sont récapitulés dans le tableau 3.1. Le taux de mutation global et spécifique à chaque type d'événements ont été déterminés par la méthode de Lea et Coulson présentée dans le chapitre 1.

Souche	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677 ; 1,373)	29	20 (69%) 0,707 (0,466 ; 0,948)	9 (31%) 0,317 (0,209 ; 0,426)

Tableau 3.1 : Récapitulatif des résultats obtenus dans la souche aUHT.

N représente le nombre de sélections réalisées et n_c , le nombre cellules étalées par sélection.

Dans ce contexte, 155 sélections ont été réalisées permettant la sélection de 29 révertants indépendants. Après caractérisation moléculaire, il a été déterminé que 20 étaient porteurs d'une délétion et 9 d'une duplication. Il a été montré dans le chapitre précédent que l'activité des rétrotransposons est fortement diminuée en contexte S288c et à température optimale de croissance. Ainsi aucun événement d'insertion de rétrotransposons Ty1 n'a été sélectionné dans ce contexte génétique.

La caractérisation moléculaire de 11 événements de délétion a montré que des microhomologies dont la taille varie de 1 à 11 pb sont présentes aux bornes des délétions (Chapitre 2 Tableau 2.3).

2.2 Le taux de délétion est accru en contexte $\Delta rad52$

2.2.1 Résultats obtenus en contexte $\Delta rad52$

La protéine codée par le gène *RAD52* est un élément majeur de l'HR. Elle se lie aux régions d'ADN simple brin issues de la digestion des extrémités 5' au niveau des cassures d'ADN double-brin. Elle participe également à l'hybridation d'un ADN simple-brin complémentaire (Mortensen *et al.*, 1996). Une délétion du gène *RAD52* a pour conséquence une mauvaise réparation des cassures double-brin et un fonctionnement défectueux de tous les mécanismes impliquant la recombinaison homologe (Symington, 2002).

Dans le contexte *ura2_{15,30,72}Δrad52*, 83 sélections, correspondant à $2,4 \times 10^{11}$ cellules étalées, ont été réalisées et 21 révertants indépendants ont pu être obtenus. Le taux de

mutation global et de chaque type d'événements, calculés pour le contexte $\Delta rad52$, sont récapitulés dans le tableau 3.2.

La caractérisation moléculaire par analyse du profil de restriction *Bam*HI par hybridation ADN/ADN de type Southern et par PCR a révélé que parmi les 21 révertants analysés, 18 sont porteurs de délétions et 3 de duplications.

Souche	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aΔR	83	2,89	1,69 (1,133 ; 2,247)	21	18 (86%) 1,45 (0,974 ; 1,932)	3 (14%) 0,237 (0,158 ; 0,314)

Tableau 3.2 : Résultats obtenus dans le contexte $\Delta rad52$.

N représente le nombre de sélections réalisées et n_c , le nombre de cellules étalées par sélection.

2.2.2 Caractérisation des événements de délétions en contexte $\Delta rad52$

Les séquences aux bornes des jonctions pour 7 délétions ont été obtenues par séquençage des fragments d'amplification PCR. Dans 6 cas, on observe la présence de microhomologies dont la taille varie de 1 à 11 pb. Pour une délétion, aucune séquence homologue n'a été trouvée à la jonction (tableau 3.3)

On peut diviser les délétions obtenues en deux catégories. Deux délétions sont internes au locus *ura2*_{15,30,72}. La région codant pour le fragment ATCase se trouve alors en phase avec l'ATG initiateur qui peut alors vraisemblablement être utilisé pour l'initiation de la traduction. Les 5 autres cas correspondent à des délétions plus longues englobant les trois mutations, l'ATG initiateur ainsi que la région promotrice. Le codon d'initiation étant perdu lors de la délétion, un ATG interne est probablement utilisé. La position de ce codon a été déterminée *in silico* en prenant le premier ATG en phase avec la région codant pour l'ATCase. Cette position reste prédictive et d'autres positions existent. Le tableau 3.3 et la figure 3.1 présentent les différents types de délétions ainsi que les positions putatives des ATG initiateurs.

Borne 5' <i>URA2</i> (inclus)	Borne 3' <i>URA2</i> (inclus)	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
-717	+3275	3992	-	0	+3301	1114	aΔR14-1
-645	+3231	3876	A	1	+3262	1127	aΔR33-1
-710	+3267	3977	T	1	+3301	1114	aΔR39-1
-645	+3231	3876	A	1	+3262	1127	aΔR49-1
-504	+5394	5898	ACC	3	+5407	412	aΔR47-1
+177	+3370	3192	TGAATCT	7	+1	1150	aΔR26-1
+196	+3416	3219	TCCATCTTATG	11	+1	1141	aΔR32-1

Tableau 3.3 : Caractérisation des différentes délétions et de leurs microhomologies.

Les délétions internes au locus *ura2*_{15,20,72} sont représentées en bleu tandis que les délétions de la région promotrice et de l'ATG initiateur sont présentées en orange.

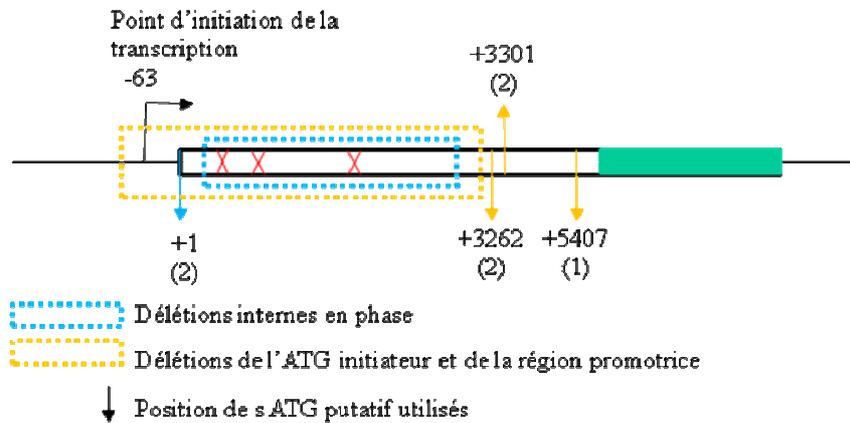


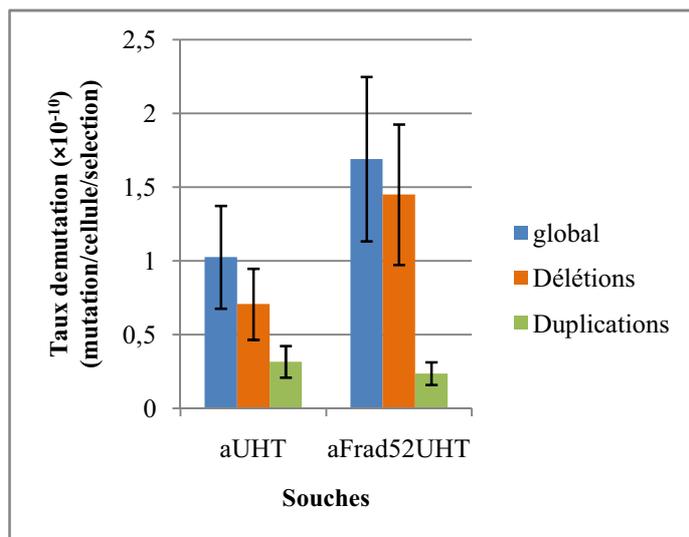
Figure 3.1 : Représentation des différents ATG initiateurs putatifs utilisés pour les délétions en contexte $\Delta rad52$. Le chiffre entre parenthèses indique le nombre de révertants pour lesquels cet ATG est prédit.

2.2.3 Comparaison des résultats obtenus en contexte $\Delta rad52$ avec ceux obtenus dans la souche aUHT

Le récapitulatif des résultats obtenus dans les contextes aUHT et *ura2*_{15,30,72} $\Delta rad52$ est présenté dans le tableau 3.4 et dans le graphique 3.1.

Souches	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677;1,373)	29	20 (69%)	9 (31%)
					0,707 (0,466;0,948)	0,317 (0,209;0,426)
aΔR	83	2,89	1,69 (1,133 ; 2,247)	21	18 (86%)	3 (14%)
					1,45 (0,974 ; 1,932)	0,237 (0,158 ; 0,314)

Tableau 3.4 : Tableau récapitulatif des sélections réalisées dans la souche de référence aUHT et dans la souche aΔR. N indique le nombre de sélections et n_c , le nombre de cellules par sélection.



Graphique 3.1 : Représentation des taux de mutation globaux ainsi que pour chaque type d'événements pour les souches aUHT et a Δ R. Les barres d'erreur correspondent aux intervalles de confiance.

Les intervalles de confiance étant chevauchants pour les taux d'apparition des duplications, aucune différence statistiquement significative n'est observée entre les deux contextes.

En revanche, il existe une différence significative pour le taux de délétions puisqu'il est environ deux fois plus élevé en absence du gène *RAD52* par rapport à la souche de référence aUHT.

On constate que le taux global de mutation dans la souche *ura2_{15,30,72} Δ rad52* est 1,6 fois plus important que celui observé en contexte aUHT. Néanmoins, ce résultat ne peut pas être validé statistiquement car les intervalles de confiance sont chevauchants. Ceci est dû à un biais dans la méthode statistique utilisée qui ne permet pas de mettre en évidence de différence si la variation entre les taux de mutation est inférieure à un facteur 2.

Ces résultats nous permettent de conclure que le mécanisme de recombinaison homologue n'est pas impliqué dans l'apparition des délétions et des duplications. Cependant, on observe une augmentation de l'apparition des délétions ce qui suggère que le mécanisme impliqué est favorisé lorsque l'HR est inactivée.

2.3 Rad59p restreint l'apparition des duplications

2.3.1 Résultats des sélections réalisées en contexte Δ rad59

La protéine codée par le gène *RAD59* est impliquée dans le mécanisme de SSA (Storici *et al.*, 2006). Elle présente des fonctions communes avec Rad52p et participe à la réparation des cassures double-brin en appariant deux séquences d'ADN simple-brin (Bai et Symington, 1996).

Dans le contexte *ura2_{15,30,72}Δrad59*, 130 sélections représentant $2,88 \times 10^{11}$ cellules étalées ont été réalisées et 37 révertants indépendants ont ainsi été isolés. Les taux de mutation global et pour les différents événements mutationnels ont été calculés d'après la méthode de Lea et Coulson et sont présentés dans le tableau 3.5.

Souche	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements	
					taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aFR	130	2,22	1,81 (1,27 ; 2,35)	37	20 (54%) 0,977 (0,685 ; 1,269)	17 (46%) 0,832 (0,584 ; 1,081)

Tableau 3.5 : Récapitulatif des résultats pour les sélections réalisées en contexte aΔrad59.

L'analyse moléculaire des réarrangements à l'origine de la réactivation de l'ATCase par hybridation ADN/ADN de type Southern a révélé deux bandes pour 17 révertants, une correspondant à l'allèle de départ et une de taille variable. Une seule bande de taille différente de celle de l'allèle muté a été mise en évidence pour les 20 autres révertants. Ces profils nous permettent de conclure que 17 révertants sont porteurs d'une duplication et les 20 autres d'une délétion ou d'une insertion de rétrotransposon Ty1. L'analyse PCR décrite dans le chapitre 1 nous a permis de discriminer ces deux types d'événements. Aucune amplification n'a été obtenue lors de la recherche d'insertions de rétrotransposons Ty1. Nous avons donc pu conclure que les 20 révertants sont porteurs d'une délétion.

2.3.2 Caractérisation des délétions en contexte aΔrad59

Pour 14 révertants porteurs de délétions, la jonction de la délétion a été amplifiée par PCR. La détermination de la séquence par séquençage des fragments d'amplification PCR a permis d'obtenir, comme dans les contextes précédents, les séquences aux bornes des délétions (tableau 3.6).

Douze révertants présentent des microhomologies de 1 à 11 pb tandis que deux ont des bornes franches sans microhomologie. Sept délétions sont internes au locus *ura2_{15,30,72}*. La région ATCase se trouve alors en phase avec l'ATG initiateur du gène *URA2* qui peut ainsi être utilisé lors de la traduction. Les 7 autres délétions englobent la région codant pour les trois mutations, l'ATG initiateur et la région promotrice. Comme précédemment, le premier ATG initiateur interne en phase avec la région codant pour

l'ATCase a été retenu comme codon d'initiation. Les résultats obtenus sont détaillés dans le tableau 3.6 et la figure 3.2.

Borne 5' <i>URA2</i> (inclus)	Borne 3' <i>URA2</i> (inclus)	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
-696	+4023	4719	-	0	+4156	829	aFR17-1
-573	+3341	3914	-	0	+3352	1097	aFR96-1
115	+3179	3063	G	1	+1	1193	aFR28-1
222	+3595	3372	C	1	+1	1090	aFR45-1
-643	+3100	3743	T	1	+3184	1153	aFR46-1
-66	+4054	4120	G	1	+4156	829	aFR93-1
228	+4096	3867	ATTG	4	+1	925	aFR08-1
+56	+4005	3948	GGTGA	5	+1	878	aFR59-1
+56	+4005	3948	GGTGA	5	+1	878	aFR85-1
+56	+4005	3948	GGTGA	5	+1	878	aFR86-1
-352	+3891	4243	GCCTTA	6	+3988	885	aFR27-1
+177	+3370	3192	CTGAATCT	8	+1	1150	aFR42-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFR10-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFR12-2

Tableau 3.6 : Tableau des différentes délétions obtenues pour les sélections en contexte a Δ rad59.

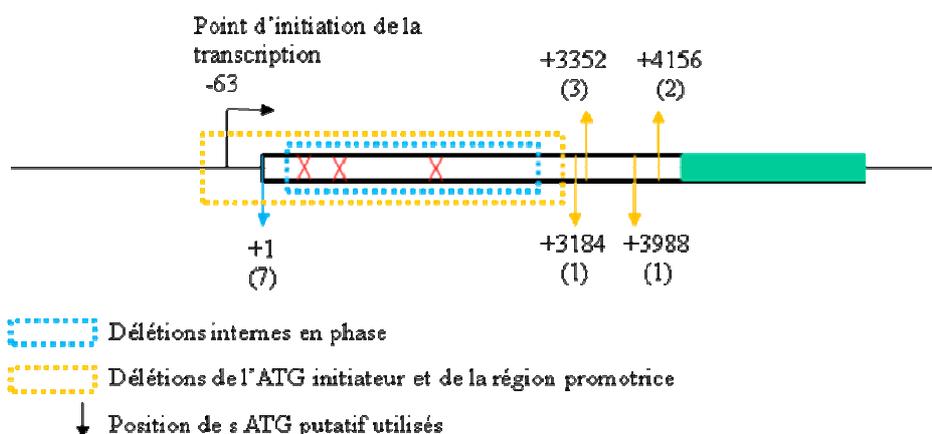


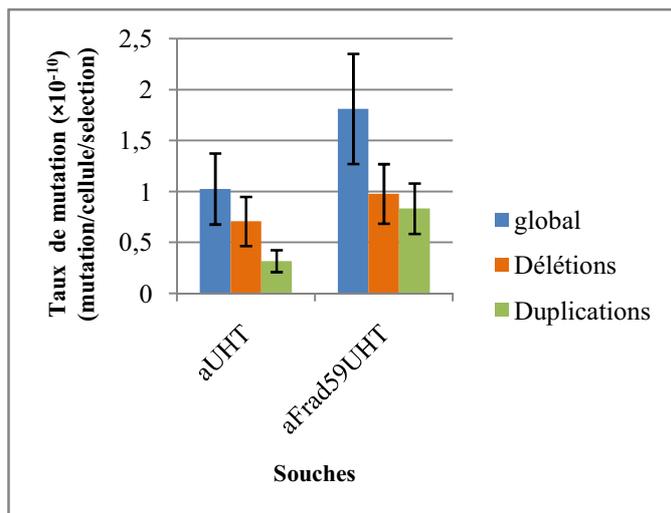
Figure 3.2 : Représentation des différents points d'initiation de la traduction prédits *in silico* en contexte a Δ rad59.

Le chiffre entre parenthèses représente le nombre de révertants pour lesquels cet ATG est prédit.

2.3.3 Comparaison du contexte a Δ rad59 avec le contexte de référence aUHT

Le récapitulatif des résultats obtenus dans ces deux contextes est présenté dans le tableau 3.7 et dans le graphique 3.2.

Souches	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677 ; 1,373)	29	20 (69%) 0,707 (0,466 ; 0,948)	9 (31%) 0,317 (0,209 ; 0,426)
aFR	130	2,22	1,81 (1,27 ; 2,35)	37	20 (54%) 0,977 (0,685 ; 1,269)	17 (46%) 0,832 (0,584 ; 1,081)

Tableau 3.7 : Tableau comparatif des résultats obtenus en contextes $\Delta rad59$ et aUHTGraphique 3.2 : Représentation des taux de mutation globaux et pour chaque type d'événements obtenus en contexte aUHT et $a\Delta rad59$. Les barres d'erreur correspondent aux intervalles de confiance observés pour les différents taux de mutation.

Aucune différence statistique n'est observée pour les taux de délétions, les intervalles de confiance étant chevauchants. En revanche, les intervalles de confiance sont non-chevauchants dans le cas des taux de duplications, indiquant que ceux-ci ne sont pas statistiquement équivalents dans les deux contextes. Le taux de duplications est 2,5 fois plus important en absence de *RAD59*, ce qui nous permet de conclure que Rad59p restreint l'apparition des duplications et qu'en l'absence du mécanisme de SSA, ce type d'événement est favorisé.

On constate que le taux global de mutation en contexte *ura2_{15,30,72} Δrad59* est 1,8 fois plus important que celui observé en contexte aUHT. Comme pour le contexte *ura2_{15,30,72} Δrad52*, ce résultat ne peut pas être validé statistiquement car les intervalles de confiance sont chevauchants.

2.4 Conclusions

Dans cette partie, nous avons testé l'influence des mécanismes de recombinaison homologue sur l'apparition de délétions et de duplications. Pour la souche *ura2_{15,30,72} Δrad52*, nous avons ainsi pu montrer que le taux de duplications n'est pas modifié, en revanche, le

taux de délétions est augmenté. Ce résultat suggère donc qu'en absence du mécanisme de recombinaison homologue, l'apparition des délétions est favorisée.

L'influence du mécanisme de SSA a également été analysée en sélectionnant des remaniements dans un contexte muté pour *RAD59*. Le taux de duplications étant deux fois et demie plus important en contexte $\Delta rad59$, nous pouvons donc conclure que l'inactivation du mécanisme de SSA entraîne une augmentation du nombre de duplications sélectionnées.

3. Implication du mécanisme de NHEJ

Le mécanisme de NHEJ est impliqué dans la réparation de cassure d'ADN double-brin. On retrouve souvent à l'origine de ce mécanisme de courtes séquences identiques (moins de 10 pb). Il peut cependant fonctionner en absence de ce type de séquences.

L'implication du mécanisme de NHEJ dans l'apparition des délétions et des duplications a été testée. Pour ce faire, deux mutants de délétion dont les gènes sont impliqués dans cette voie à savoir $\Delta yku80$ et $\Delta lig4$ ont été construits en contexte *ura2_{15,30,72}*. La protéine codée par le gène *YKU80* est une sous-unité du complexe Ku impliqué dans la fixation de l'ADN aux sites de cassure d'ADN double-brin (Bertuch et Lundblad, 2003). Ce complexe intervient également dans le maintien des télomères. Le complexe Ku est formé d'un hétérodimère composé d'une sous-unité Yku80p et d'une sous-unité Yku70p. Il se fixe au niveau de la cassure d'ADN double-brin et a une action de pont entre les deux brins issus de la cassure. Il permet ensuite le recrutement et la fixation de la ligase codée par le gène *LIG4* au niveau de la cassure (Dudasova *et al.*, 2004).

La protéine Lig4p codée par le gène *LIG4* correspond à la ligase nécessaire au NHEJ, qui forme un complexe avec ses deux cofacteurs Lif1p et Nej1p (Daley *et al.*, 2005). Elle n'est pas impliquée dans la réplication de l'ADN et à l'heure actuelle, aucun autre rôle ne lui est connu à part la ligation des brins d'ADN suite à une cassure d'ADN (Daley *et al.*, 2005).

3.1 L'apparition des délétions et des duplications est indépendante de Yku80p

3.1.1 Résultats des sélections réalisées en contexte $\Delta yku80$

Dans la souche *ura2_{15,30,72} $\Delta yku80$* , 135 sélections ont été réalisées correspondant à $2,7 \times 10^{11}$ cellules étalées. Nous avons ainsi pu isoler 22 révertants. Le taux global de mutation ainsi que les taux pour chaque type d'événements mutationnels ont été calculés d'après la méthode de Lea et Coulson, comme décrit précédemment. Le tableau 3.8 présente les résultats obtenus dans ce contexte.

Souches	N	n _c (×10 ⁹)	Taux de mutation global (×10 ⁻¹⁰)	Nb de révertants indépendants	Proportion de chaque type de réarrangements	
					Délétions	Duplications
aFY	135	2	0,997 (0,604 ; 1,39)	22	16 (73%) 0,728 (0,441 ; 1,015)	6 (27%) 0,269 (0,163 ; 0,375)

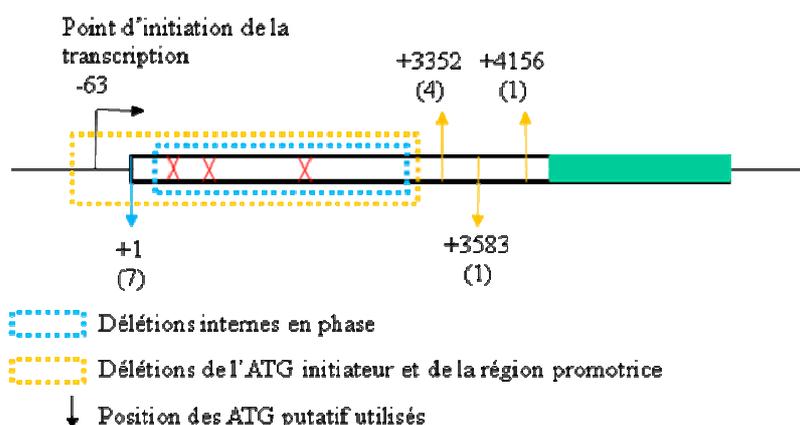
Tableau 3.8 : Tableau récapitulatif des résultats obtenus en contexte $\Delta yku80$.

L'analyse des différents révertants par hybridation ADN/ADN de type Southern a montré que parmi les 22 révertants obtenus, 6 présentaient deux bandes en profil *Bam*HI et les 16 autres, une seule bande nous permettant de conclure que 6 révertants sont porteurs d'une duplication et que les 16 autres contiennent une délétion ou une insertion de Ty1. La caractérisation par PCR de ces révertants a déterminé qu'un événement de délétion est responsable de la réactivation de l'ATCase pour les 16 souches isolées.

3.1.2 Caractérisation des événements de délétion

Pour 13 délétions, une amplification par PCR suivie d'un séquençage a permis d'obtenir les bornes à la jonction des délétions. Des microhomologies dont la taille varie de 2 à 11 pb ont été observées aux bornes de toutes les délétions (tableau 3.9). Sept délétions sont internes au locus *ura2*_{15,30,72} tandis que les 6 autres englobent la région contenant les trois mutations ainsi que la région promotrice. Pour les délétions internes, l'ATG initiateur du gène *URA2* se trouve en phase avec la région codant pour l'ATCase, ce qui permet de supposer qu'il est utilisé lors de l'initiation de la traduction. Dans les 6 cas de délétions plus longues, l'ATG initiateur de la traduction a été déterminé *in silico*, comme décrit précédemment. La figure 3.3 présente les ATG putatifs utilisés pour l'initiation de la traduction.

Borne 5' <i>URA2</i> (inclus)	Borne 3' <i>URA2</i> (inclus)	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
-435	+4090	4525	TT	2	+4156	829	aFy82-1
+43	+3992	3948	AATGG	5	+1	898	aFy27-1
+43	+3992	3948	AATGG	5	+1	898	aFy29-1
+56	+4005	3948	GGTGA	5	+1	878	aFy56-1
+56	+4005	3948	GGTGA	5	+1	878	aFy59-1
+138	+4006	3267	CTGGTGAA	8	+1	925	aFy58-2
+139	+3579	3439	TGGTGAATT	9	+3583	1020	aFy58-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFy12-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFy16-3
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFy21-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFy38-1
+196	+3416	3219	TCCATCTTATG	11	+1	1141	aFy55-1
+196	+3416	3219	TCCATCTTATG	11	+1	1141	aFy71-1

Tableau 3.9 : Description des différentes délétions caractérisées en contexte $\Delta yku80$.Figure 3.3 : Représentation des différents ATG initiateurs prédits *in silico* en contexte $\Delta yku80$.

3.1.3 Caractérisation des duplications

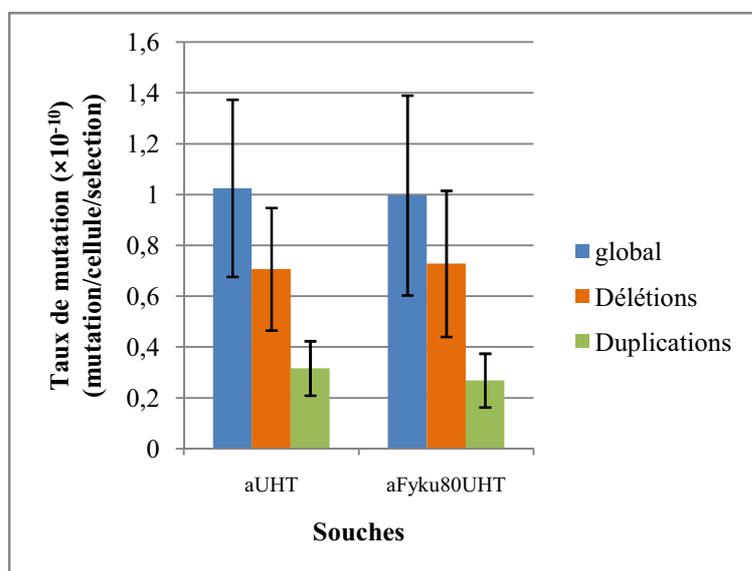
Pour les 6 révertants porteurs d'une duplication, nous avons réalisé un caryotype électrophorétique dans le but de localiser le chromosome porteur de la duplication. Pour cela, les chromosomes ont été séparés par électrophorèse en champs alternés puis une hybridation de type Southern a été réalisée en utilisant une sonde correspondant à la région ATCase. Dans toutes les souches, seul le chromosome X a été révélé, nous permettant de conclure qu'elles sont toutes intrachromosomiques. De plus, le profil *Bam*HI par hybridation ADN/ADN en utilisant une sonde correspondant à la région intergénique entre *URA2* et *YJL131c* présente deux bandes pour les différentes duplications. Ce profil à deux bandes indique que la région intergénique entre les gènes *URA2* et *YJL131c* est coduplicquée avec la région codant pour l'ATCase, nous permettant de conclure qu'il s'agit donc de duplications segmentales.

3.1.4 Comparaison des résultats obtenus pour les souches $a\Delta yku80$ et aUHT

Le récapitulatif des résultats obtenus est présenté dans le tableau 3.10 et dans le graphique 3.3.

Souches	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677 ; 1,373)	29	20 (69%) 0,707 (0,466 ; 0,948)	9 (31%) 0,317 (0,209 ; 0,426)
aFY	135	2	0,997 (0,604 ; 1,39)	22	16 (73%) 0,728 (0,441 ; 1,015)	6 (27%) 0,269 (0,163 ; 0,375)

Tableau 3.10 : Tableau récapitulatif des sélections réalisées dans les souches $a\Delta yku80$ et aUHT.



Graphique 3.3 : Comparaison des taux de mutation calculés pour les souches $a\Delta yku80$ et aUHT.

Aucune différence significative n'a été mise en évidence pour les taux de mutation qu'ils soient globaux ou relatifs aux délétions et aux duplications. Ce résultat nous permet de conclure que les délétions et les duplications apparaissent de manière indépendante de la protéine Yku80p.

3.2 Une délétion du gène *LIG4* conduit à une diminution du taux d'apparition des délétions et des duplications

3.2.1 Résultats des sélections réalisées en contexte $\Delta lig4$

Dans la souche $ura2_{15,30,72}\Delta lig4$, 151 sélections ont été réalisées correspondant à $3,7 \times 10^{11}$ cellules étalées. A partir de ces sélections, 9 révertants indépendants ont été obtenus.

Les taux de mutation global et relatifs à chaque type d'événements sont présentés dans le tableau 3.11.

Les neuf révertants ont été caractérisés par les approches habituelles décrites plus haut. Aucune duplication ni insertion de rétrotransposon Ty1 n'a été mise en évidence. Ainsi les neuf révertants sont porteurs d'une délétion. Le taux de délétion dans ce contexte est donc identique au taux global de mutation.

Souche	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements	
					taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aFL	151	2,45	0,204 (0,055;0,353)	9	9 (100%) 0,204 (0,055;0,353)	0

Tableau 3.11 : Résultats des sélections réalisées en contexte $\Delta lig4$.

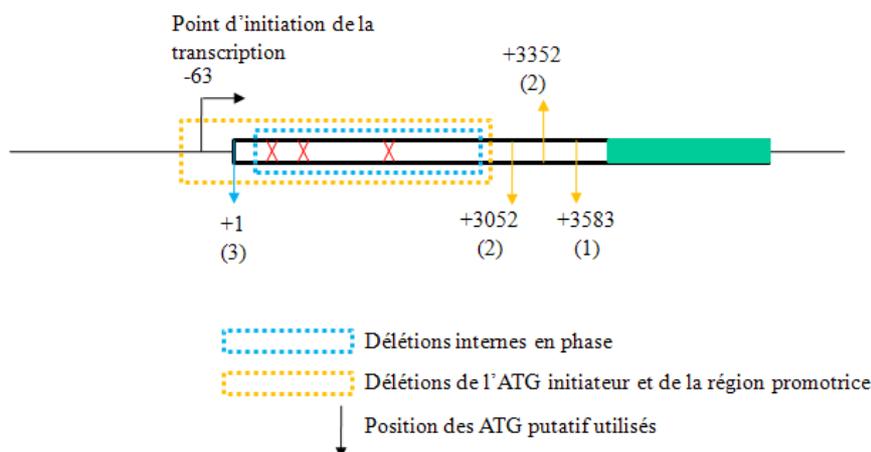
3.2.2 Caractérisation moléculaire des révertants obtenus en contexte $\Delta lig4$

Huit délétions ont été caractérisées par PCR. Toutes présentent des microhomologies dont la taille varie de 1 à 11 pb.

Quatre des délétions sont internes. Pour trois d'entre elles, l'ATG initiateur du gène *URA2* est en phase avec la région codant pour la région ATCase et peut donc vraisemblablement être utilisé lors de la traduction. Pour la quatrième délétion, en revanche, l'ATG initiateur du gène *URA2* ne se trouve plus en phase avec la région codant pour l'ATCase après la délétion. Dans ce cas, celui-ci ne peut pas être utilisé pour l'initiation de la traduction. Un codon initiateur interne au locus *ura2*_{15,30,72} a donc été prédit *in silico* comme dans le cas des délétions englobant la région promotrice décrit précédemment.

Les quatre autres délétions sont plus longues et englobent les trois mutations, le codon initiateur et la région promotrice. La position du codon initiateur interne a été déterminée par la même procédure que celle décrite plus haut. Le tableau 3.12 et la figure 3.4 présentent l'ensemble de ces résultats.

Borne 5' <i>URA2</i> (inclus)	Borne 3' <i>URA2</i> (inclus)	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
159	+3241	3081	G	1	+1	1187	aFL80-1
401	+2853	2451	TGA	3	+3052	1197	aFL12-1
144	+3583	3438	GTT	3	+1	1068	aFL94-1
-741	+2925	3666	CGATTT	6	+3052	1197	aFL06-1
+177	+3370	3192	CTGAATCT	8	+1	1150	aFL16-1
+139	+3579	3439	TGGTGAATT	9	+3583	1020	aFL22-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFL19-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aFL38-1

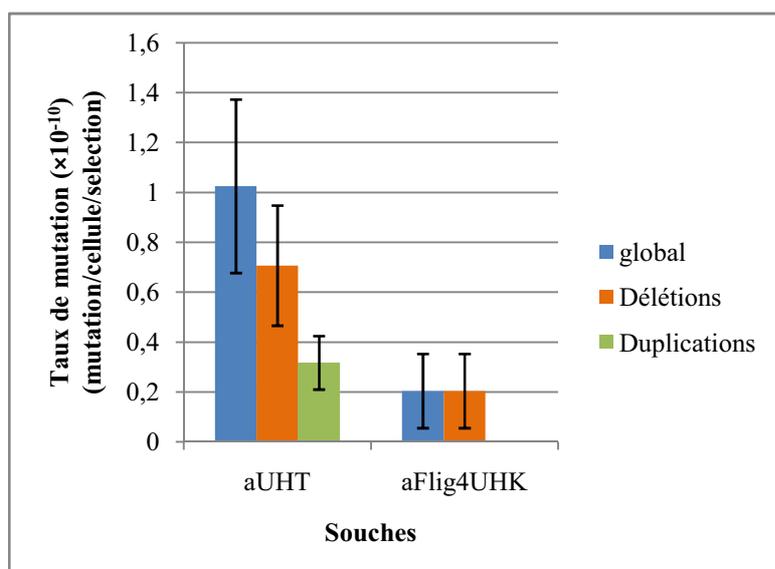
Tableau 3.12 : Caractérisation des délétions obtenues en contexte $\Delta lig4$.Figure 3.4 : Représentation des différents ATG initiateurs prédits *in silico* en contexte $\Delta lig4$.

3.2.3 Comparaison des résultats obtenus dans le contexte $\Delta lig4$ et aUHT sauvage

Les résultats obtenus dans les deux contextes sont présentés dans le tableau 3.13 et sur le graphique 3.4.

Souches	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677 ; 1,373)	29	20 (69%) 0,707 (0,466 ; 0,948)	9 (31%) 0,317 (0,209 ; 0,426)
aFL	151	2,45	0,204 (0,055 ; 0,353)	9	9 (100%) 0,204 (0,055 ; 0,353)	0

Tableau 3.13 : Récapitulatif des sélections réalisées en contexte $\Delta lig4$ et celui de référence.



Graphique 3.4 : Comparatif des taux de mutations pour les souches aFL et aUHT.

Le taux global de mutation est 5 fois plus faible lorsque l'on mute *LIG4* par rapport à celui observé dans la souche aUHT. Ceci se traduit par une diminution du nombre de révertants obtenus. Ce résultat peut expliquer l'absence totale de duplications sélectionnées, la population de révertants étant vraisemblablement trop faible pour que l'on puisse observer ce type de remaniements. Des délétions sont, quant à elles, observées mais là encore le taux de délétions est environ 3,5 fois plus faible que dans la souche sauvage.

Ces observations permettent de conclure que la protéine codée par le gène *LIG4* joue un rôle dans l'apparition des délétions comme dans celle des duplications.

3.3 Conclusions

Après l'influence de la recombinaison homologue, nous nous sommes intéressés à l'impact du mécanisme de NHEJ sur l'apparition des délétions et des duplications. Lorsque que l'on mute le gène *YKU80*, aucune différence n'est observée par rapport à la souche de référence que ce soit au niveau des taux de mutation que des types de remaniements obtenus. Ce résultat montre que l'apparition des délétions et des duplications est indépendante de la protéine Yku80p.

En revanche, lorsque le gène *LIG4* est délété, on observe une diminution du taux global de mutation ainsi que des taux de délétions et duplications, nous permettant de conclure que la protéine Lig4p joue un rôle dans l'apparition de ces deux types d'événements.

4. Un autre mécanisme intervient également dans l'apparition des remaniements chromosomiques

Il a été démontré que deux grands mécanismes permettent la réparation de cassures d'ADN double-brin : la recombinaison homologue et le NHEJ. Dans cette partie, nous nous sommes interrogés sur l'apparition des délétions et des duplications en absence de ces deux mécanismes majeurs. Pour ce faire, nous avons construit un double mutant $\Delta lig4\Delta rad52$ puis sélectionné et analysé des révertants dans ce contexte.

4.1 Résultats des sélections réalisées en contexte $\Delta lig4\Delta rad52$

Dans la souche $ura2_{15,30,72}\Delta lig4\Delta rad52$, 155 sélections ont été réalisées correspondant à $2,7 \times 10^{11}$ cellules étalées et 11 révertants indépendants ont ainsi été sélectionnés. Le taux de mutation global ainsi que ceux déterminés pour les différents types de réarrangements sont présentés dans le tableau 3.14.

La caractérisation moléculaire des révertants par analyse du profil *Bam*HI par hybridation ADN/ADN et par PCR a montré qu'il s'agit de 8 délétions et 3 duplications.

Souche	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aLR	155	1,74	0,574 (0,279 ; 0,869)	11	8 (73%) 0,419 (0,204 ; 0,634)	3 (27%) 0,155 (0,075 ; 0,235)

Tableau 3.14 : Récapitulatif des sélections réalisées en contexte $\Delta lig4\Delta rad52$.

4.1.1 Analyse des événements de délétion

Quatre délétions ont été caractérisées par analyse PCR et séquençage du fragment d'amplification. Trois d'entre elles présentent une microhomologie de 11 pb tandis que la quatrième a une taille de 5 pb. Cette dernière délétion est interne au locus $ura2_{15,30,72}$. L'ATG initiateur du gène *URA2* se trouve en phase avec la région codant pour l'ATCase et est vraisemblablement utilisé pour la traduction. Pour les trois autres révertants, la délétion comprend la région contenant les trois mutations, le codon initiateur et la région promotrice. Le codon d'initiation de la traduction étant perdu, un codon interne au locus $ura2_{15,30,72}$ a été prédit *in silico* comme décrit précédemment (tableau 3.16 et figure 3.5).

Borne 5' <i>URA2</i> (inclus)	Borne 3' <i>URA2</i> (inclus)	Taille de la délétion (nt)	Séquence à la jonction	Taille de la microhomologie	Position de l'ATG initiateur putatif	Taille de la protéine putative (aa)	Révertants
+43	+3992	3948	AATGG	5	+1	898	aLR20-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aLR18-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aLR31-1
-512	3347	3858	GAAAGAATTGA	11	+3352	1097	aLR78-1

Tableau 3.15 : Description des délétions caractérisées en contexte $\Delta lig4\Delta rad52$.

En bleu sont présentées les délétions internes et en orange les délétions englobant la région contenant les trois mutations ainsi que la région promotrice et l'ATG initiateur du gène *URA2*.

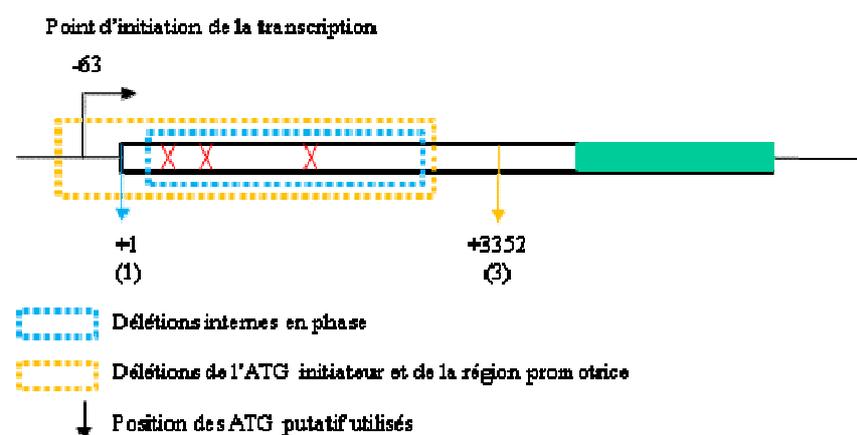


Figure 3.5 : Représentation des différents points d'initiation de la traduction prédits *in silico* en contexte $\Delta lig4\Delta rad52$.

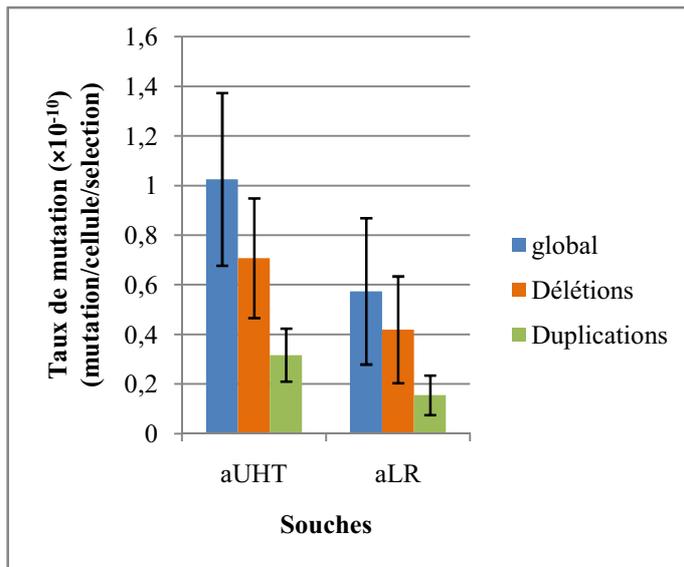
Le chiffre entre parenthèses représente le nombre de révertants pour lesquels cet ATG est prédit.

4.2 Comparaison des résultats obtenus dans la souche $\Delta lig4\Delta rad52$ et la souche aUHT

Le récapitulatif des résultats obtenus dans le contexte aUHT et dans le contexte *ura2_{15,30,72}* $\Delta lig4\Delta rad52$ est présenté dans le tableau 3.16 et dans le graphique 3.5.

Souches	N	$n_c (\times 10^9)$	Taux de mutation global ($\times 10^{-10}$)	Nb de révertants indépendants	Proportion de chaque type de réarrangements taux de mutation ($\times 10^{-10}$) (intervalle de confiance)	
					Délétions	Duplications
aUHT	155	2,43	1,025 (0,677 ; 1,373)	29	20 (69%) 0,707 (0,466 ; 0,948)	9 (31%) 0,317 (0,209 ; 0,426)
aLR	155	1,74	0,574 (0,279 ; 0,869)	11	8 (73%) 0,419 (0,204 ; 0,634)	3 (27%) 0,155 (0,075 ; 0,235)

Tableau 3.16 : Tableau récapitulatif des résultats obtenus en contexte aUHT et a $\Delta lig4\Delta rad52$.



Graphique 3.5 : Comparaison des taux de mutation déterminés pour les souches $a\Delta lig4\Delta rad52$ et aUHT.

Aucune différence statistique n'est observée entre le double mutant et la souche de référence tant au niveau du taux de mutation global qu'au niveau des taux relatifs aux différents types de réarrangements. Ce résultat suggère qu'un mécanisme supplémentaire différent de la recombinaison homologe et du NHEJ, intervient dans l'apparition des délétions et des duplications. Ce mécanisme est d'ailleurs favorisé en absence des deux mécanismes majeurs.

5. Influence de la ploïdie sur l'apparition des remaniements

Lors de cette étude, nous nous sommes également intéressés à l'impact que peut avoir la ploïdie sur l'apparition des remaniements chromosomiques. Pour ce faire, nous avons voulu comparer les fréquences d'apparition et les types de remaniements obtenus en contextes haploïde a et/ou α et diploïdes a/α et a/a .

Dans le diploïde a/α , un répresseur $a1/\alpha2$ est formé. Il inhibe l'expression des gènes spécifiques de la phase haploïde et favorise indirectement la méiose. Un des gènes connus pour être réprimé en phase diploïde a/α est le gène *NEJ1* qui code pour un cofacteur de la protéine Lig4p impliquée dans le mécanisme de NHEJ (Frank-Vaillant et Marcand, 2001 ; Valencia *et al.*, 2001). On peut donc supposer qu'en phase diploïde, ce mécanisme sera réprimé et n'interviendrait donc pas dans l'apparition des remaniements. En contexte diploïde a/a , le répresseur n'est pas formé et la souche se comporte donc comme un haploïde. Il a été montré dans des études précédentes (Tourrette *et al.*, 2007) que le taux de mutation est diminué en contexte diploïde FL100. Un diploïde a/α a été construit en contexte S288c mais aucune sélection n'a actuellement été réalisée dans ce contexte.

Nous avons construit une souche *MATa/MATa ura2_{15,30,72}/ura2_{15,30,72}* par la technique de fusion de sphéroplastes. 101 sélections correspondant à $1,37 \times 10^{11}$ cellules étalées ont été réalisées dans ce contexte et un seul révertant a été obtenu pour cette souche. Le calcul du taux de mutation global ne peut pas être effectué dans ce cas car la population de révertants est trop faible pour que la valeur obtenue soit statistiquement significative. Ces observations suggèrent que la sélection de révertants est difficile dans ce contexte par rapport aux contextes haploïdes ou diploïdes *a/α*. Il semble peu vraisemblable que ceci soit lié à la présence de deux allèles *ura2_{15,30,72}* puisque des sélections ont déjà été réalisées en contexte diploïde *a/α* et ont permis l'obtention de révertants (Tourrette *et al.*, 2007). De plus il est difficile de penser qu'il s'agisse d'un problème lié à au l'état diploïde *a/a*. En effet, des études concernant les mécanismes d'apparition des remaniements chromosomiques ont déjà été menées dans ce contexte (Clikeman *et al.*, 2001). La raison pour laquelle l'obtention de révertants porteurs de remaniements chromosomiques est restreinte dans le contexte *a/a* reste donc encore inconnue.

6. Discussion

Dans cette étude, nous nous sommes intéressés aux mécanismes impliqués dans l'apparition des délétions et des duplications. Nous discuterons dans un premier temps les résultats obtenus dans le cas des délétions, puis nous aborderons les duplications.

Dans la souche *ura2_{15,30,72}Δrad52*, le taux de délétions est deux fois plus élevé dans ce contexte que dans la souche sauvage. Ceci nous permet de conclure que la recombinaison homologue n'est pas impliquée dans l'apparition des délétions et suggère de plus que leur mécanisme d'apparition est favorisé en absence de la recombinaison homologue. Ce résultat complète les résultats préliminaires observés dans la souche FL100 (Welcker *et al.*, 2000). Pour qu'une cassure double-brin soit réparée par recombinaison homologue, une très forte identité portant sur au moins 60 pb est nécessaire. Dans notre cas, aucune séquence aussi longue, homologue à la région codant pour l'ATCase n'est présente en amont des trois mutations. Ainsi la réactivation de l'activité ATCase est impossible en utilisant cette voie de recombinaison homologue. On trouve néanmoins des séquences répétées aux bornes des délétions mais de tailles beaucoup plus courtes (1 à 11 pb) ce qui indique qu'un mécanisme de recombinaison non-homologue ne nécessitant que des régions homologues de tailles réduites est probablement à l'origine de l'apparition des délétions.

Nous avons donc testé l'impact du NHEJ sur l'apparition des délétions. Dans la souche *ura2_{15,30,72}Δyku80*, aucune différence significative n'a été observée pour le taux de mutation

global ou le taux relatif aux événements de délétion. Ceci démontre que les délétions apparaissent de façon indépendante de la protéine Yku80p. En revanche lorsque le gène *LIG4*, codant pour la ligase essentielle au mécanisme de NHEJ, est délété, une diminution d'environ 5 fois du taux global de mutation ainsi que du taux de délétions a été mise en évidence, nous permettant de conclure que la protéine Lig4p joue un rôle dans l'apparition des délétions. Deux mécanismes différents de NHEJ ont été décrits chez la levure (Boulton et Jackson, 1996) ainsi que chez les eucaryotes supérieurs (Wang *et al.*, 2003). Le premier, dépendant de la protéine Yku80p, permet une réparation exacte de la cassure tandis que le deuxième, indépendant du produit du gène *YKU80*, utilise de courtes séquences répétées. Ce dernier est à l'origine de délétions ou d'insertions lors de la réparation de la cassure d'ADN. Yu et Gabriel (2003) ont montré dans leur système que les deux mécanismes peuvent conduire à l'apparition de remaniements chromosomiques différents chez *S. cerevisiae*.

Dans notre étude, le mécanisme proposé pour l'apparition des délétions dans le système de sélection *ura2_{15,30,72}* est un mécanisme de NHEJ indépendant de Yku80p (Figure 3.6). Dans ce modèle, les délétions résultent d'une cassure d'ADN double-brin suivie de la réparation par NHEJ en utilisant de courtes séquences homologues. En absence de la recombinaison homologue, ce mécanisme est alors privilégié, augmentant ainsi le nombre d'événements de délétions. On peut donc supposer que dans une cellule sauvage, il existe un équilibre entre les différents mécanismes de recombinaison homologue et non-homologue et que si l'on inactive l'un d'eux, un déséquilibre s'installe en faveur des autres.

En ce qui concerne les duplications, on constate qu'en contexte *ura2_{15,30,72}Δrad52*, aucune variation n'est observée avec la souche de référence. En revanche, lorsque l'on délète le gène *RAD59*, on observe une augmentation de 2,5 fois du taux d'apparition des duplications suggérant que le mécanisme de SSA restreint l'apparition des duplications. Pour pouvoir réparer efficacement les cassures d'ADN par le mécanisme de SSA, la présence de séquences homologues d'une taille minimum de 29 pb est nécessaire. Or dans les études précédentes réalisées en contexte FL100 (Schacherer *et al.*, 2007), les microhomologies trouvées aux bornes des duplications segmentales avaient une taille de 1 à 13 pb.

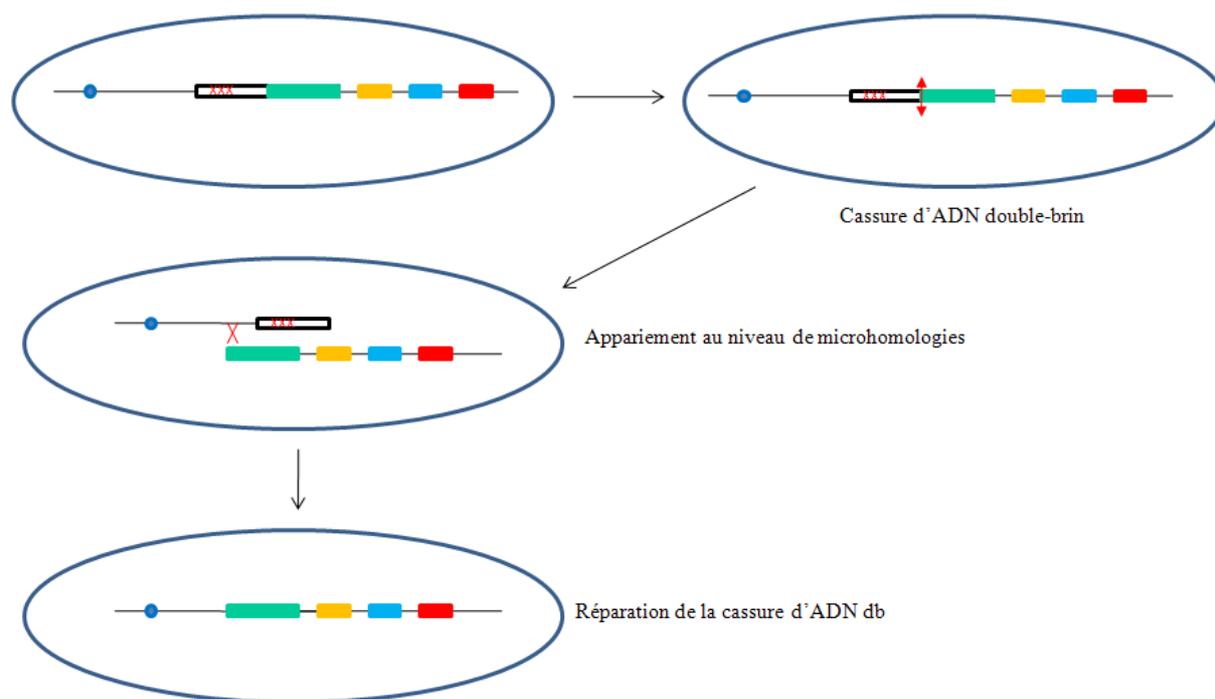


Figure 3.6 : Mécanisme de NHEJ indépendant de la protéine Yku80p proposé comme responsable de l'apparition des délétions.

On peut déduire de ces différentes observations qu'en présence des mécanismes de recombinaison homologue et de single-strand annealing l'apparition des duplications est limitée.

Aucune différence n'a été observée quant aux types de remaniements et aux taux de mutation observés entre le contexte $\Delta yku80$ et la souche de référence. Le mécanisme d'apparition des duplications est donc indépendant de la protéine Yku80p. Dans le cas d'une délétion du gène *LIG4*, le taux global de mutation est cinq fois plus faible que dans la souche de référence et le taux de duplications est également affecté puisqu'aucun remaniement de ce type n'a pu être mis en évidence lors des sélections dans ce contexte. Dans ce cas, on peut donc également émettre l'hypothèse qu'un mécanisme de NHEJ indépendant de la protéine Yku80p est à l'origine de l'apparition des duplications (Figure 3.7). Selon ce modèle, les duplications segmentales seraient formées à partir de deux cassures d'ADN double-brin se produisant en mitose, suivies de leur réparation au niveau de microhomologies par un mécanisme de NHEJ. Ce modèle a également été proposé pour l'apparition des duplications segmentales pour les souches diploïdes (Schacherer *et al.*, 2007). Comme dans le cas des duplications, on peut constater que ce mécanisme est privilégié en absence des mécanismes de recombinaison homologue et plus particulièrement du SSA.

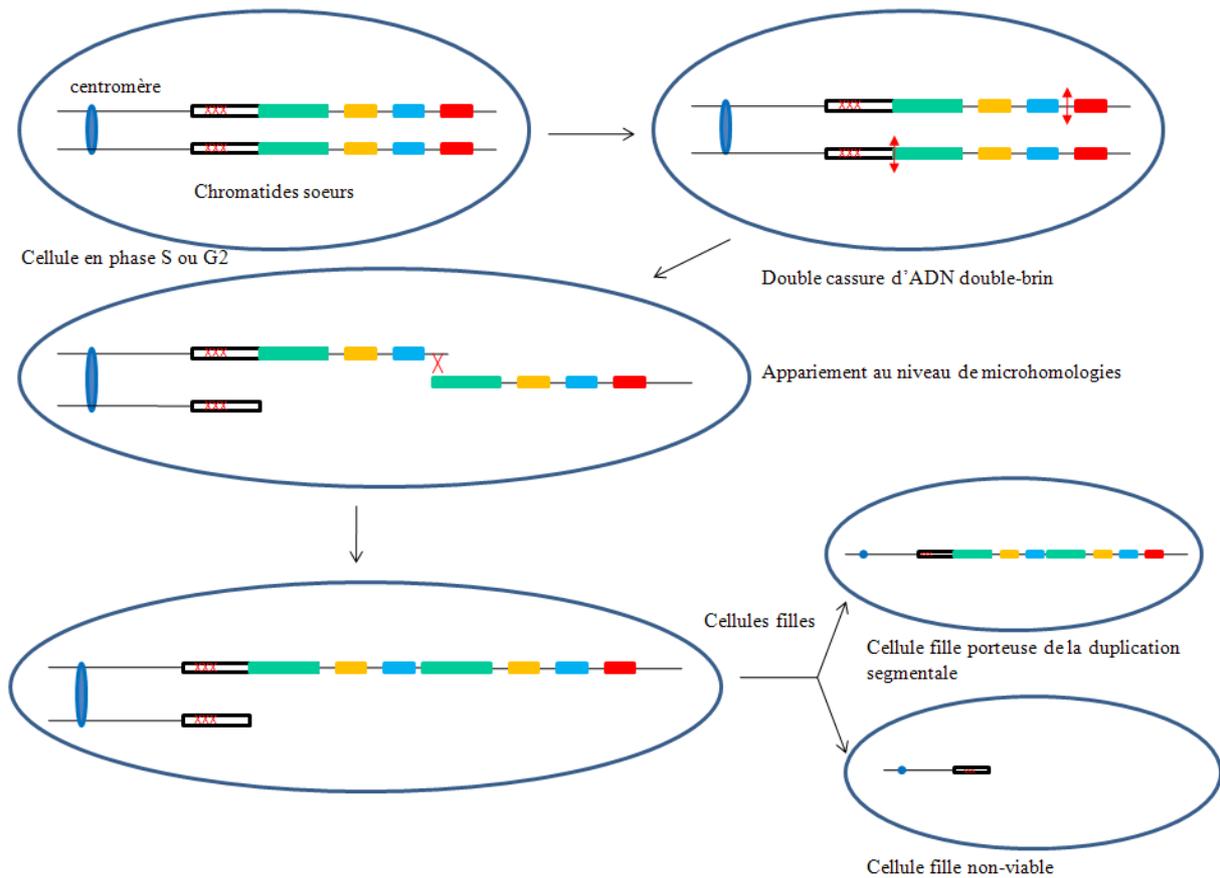


Figure 3.7 : Mécanisme de NHEJ indépendant de Yku80p proposé comme responsable de l'apparition des duplications segmentales en phase S ou G2.

L'analyse des duplications dans le contexte *ura2_{15,30,72}Δyku80* a montré que seules des duplications segmentales sont sélectionnées. Ce résultat confirme l'hypothèse émise dans le chapitre 2 selon laquelle la diminution d'activité des éléments Ty1 dans le contexte S288c entraînerait une diminution de la réactivation de l'ATCase par duplication génique. Une caractérisation plus poussée par microarray CGH (Comparative Genomic Hybridization) pourrait être menée pour préciser la longueur du fragment dupliqué. Pour réaliser cette approche, les ADN génomiques d'un révertant présentant une duplication segmentale et de la souche de départ *ura2_{15,30,72}Δyku80* sont marqués aux cyanines 5 ou 3 respectivement puis hybridés sur la même puce représentant l'ensemble des gènes de *Saccharomyces cerevisiae*. La comparaison des deux signaux de fluorescence obtenus permet de déterminer les gènes dupliqués et délétés pour chaque souche (Figure 3.8). En complément, l'analyse par les approches « classiques » d'hybridation ADN/ADN de type Southern permettra, dans un premier temps, de discriminer les duplications segmentales des duplications géniques dans les différents autres contextes. Ces résultats permettront de valider définitivement l'hypothèse émise.

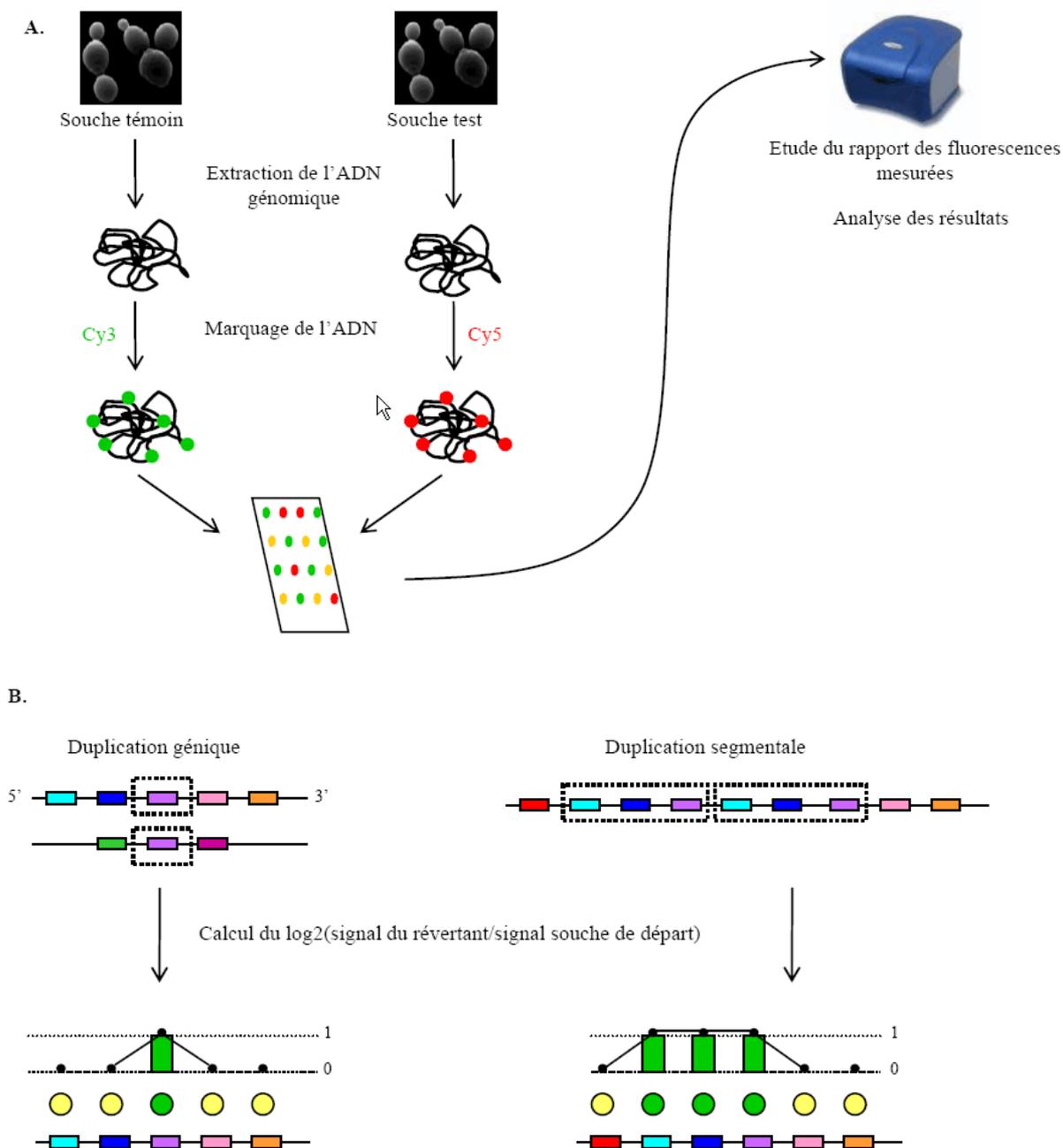


Figure 3.8 : A. Principe général de l'hybridation génomique comparative. Les ADN du témoin et de la souche porteuse d'une duplication sont préparés puis marqués à la cyanine 3 (Cy3) et 5 (Cy5) respectivement. Ils sont ensuite hybridés sur une puce représentant l'ensemble du génome de *Saccharomyces cerevisiae*. La fluorescence obtenue est ensuite analysée en scannant la lame.

B. Analyse des rapports de fluorescence dans le cas de duplications segmentales ou géniques. Pour chaque gène le rapport de fluorescence est mesuré pour les Cy5 et Cy3 et l'intensité de la fluorescence est proportionnelle au nombre de copies. On calcule le logarithme du rapport du signal obtenu pour le révertant sur celui obtenu pour la souche de référence. Si la fluorescence mesurée est la même dans les deux cas, le rapport de fluorescence est égal à 1 et le logarithme à 0. En revanche, dans le cas d'une duplication, une copie supplémentaire est présente, le rapport de fluorescence est de 2 et le logarithme est donc égal à 1.

Enfin, nous nous sommes intéressés à l'apparition des remaniements chromosomiques lorsque l'on inactive la recombinaison homologe et le NHEJ. Dans la souche *ura2_{15,30,72}Δrad52Δlig4*, on a constaté que le taux de mutation global, de duplications et de délétions sont inchangés par rapport à ceux obtenus dans la souche de référence. Ce résultat suggère qu'un troisième mécanisme utilisant de courtes séquences homologues, indépendant des deux mécanismes majeurs permet également l'apparition des délétions et des duplications en utilisant de courtes séquences homologues. De plus, lorsque l'HR et le NHEJ sont inactivés, le déséquilibre semble se faire en faveur de ce troisième mécanisme. Un mécanisme potentiel pourrait être le microhomology-mediated end joining (MMEJ) qui nécessite Mre11p, Rad1p et Rad50p (Ma *et al.*, 2003). Celui-ci est indépendant de la recombinaison homologe et du non-homologous end joining (Figure 3.9).

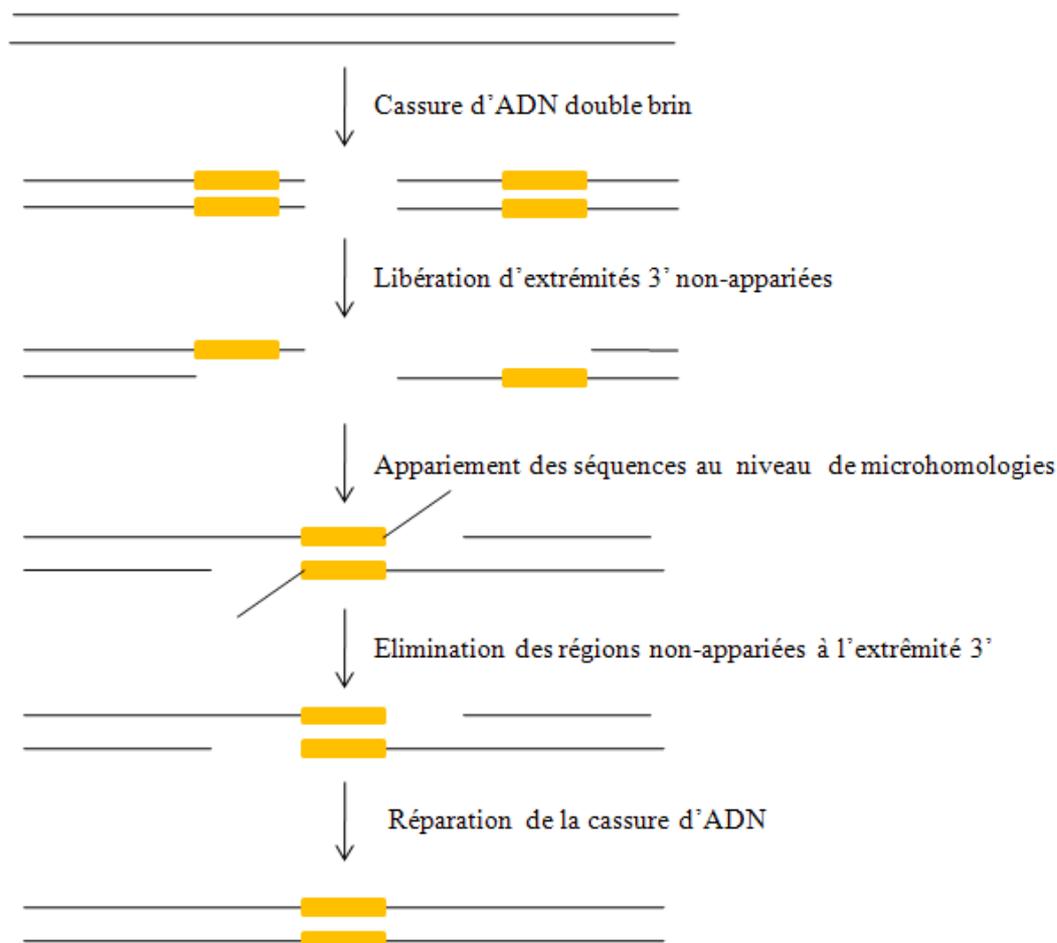


Figure 3.9 : Représentation du mécanisme de MMEJ.

Lors d'une cassure d'ADN double-brin, l'activité exonucléotidique 5'-3' du complexe Mre11p/Rad50p/Xrs2p libère des régions simple-brin en 3' de la cassure (Nakada *et al.*, 2004). Il se produit ensuite un appariement des deux extrémités au niveau de microhomologies. Les régions non-appariées sont ensuite éliminées par le complexe endonucléase Rad1p/Rad10p. Il y a enfin synthèse d'ADN à partir des microhomologies puis ligation par la ligase Cdc9p. Pour tester l'implication du MMEJ dans l'apparition des délétions et des duplications, la même stratégie que celle utilisée lors de cette étude pourrait être mise en œuvre. Des sélections de remaniements chromosomiques pourraient être réalisées dans des souches délétées pour des gènes dont les protéines sont impliquées dans le mécanisme de MMEJ telles que Rad1p ou Rad10p. Cette approche présente pourtant l'inconvénient majeur que les produits de ces deux gènes sont pléiotropiques et qu'ils sont donc retrouvés dans de nombreux mécanismes différents tels que le NHEJ ou le SSA ce qui pourrait compliquer les interprétations.

Pour conclure, un mécanisme de NHEJ indépendant du complexe Ku est responsable de l'apparition des délétions et des duplications segmentales. En absence de recombinaison homologue, ce mécanisme est favorisé, ce qui suggère que dans une cellule sauvage, un équilibre existe entre les différents mécanismes de recombinaison afin de préserver l'intégrité du génome. Ces résultats observés en contexte S288c sont complémentaires de ceux proposés en contexte FL100 (Welcker *et al.*, 2000 ; Schacherer *et al.*, 2007).

DISCUSSION ET PERSPECTIVES

Dans ce chapitre, nous allons résumer l'ensemble des résultats obtenus dans la partie I concernant l'influence du contexte génétique et les mécanismes d'apparition des remaniements chromosomiques. Nous présenterons une discussion de ces résultats ainsi que des perspectives plus générales.

Le génome, longtemps considéré comme statique, est maintenant largement reconnu comme une structure dynamique grâce notamment à l'importance des éléments transposables et à la mise en évidence tant *in silico* qu'*in vivo* des remaniements chromosomiques. L'évolution des génomes nécessite la modification de l'information génétique. Ceci passe, entre autres, par l'acquisition de réarrangements chromosomiques et leur fixation dans le génome s'ils présentent un avantage sélectif.

Ces remaniements sont des événements rares apparaissant à une fréquence faible dans les cellules (de l'ordre de 10^{-10} à 10^{-12}). Afin de pouvoir les étudier, il est donc nécessaire de pouvoir isoler ces événements en utilisant un système de sélection efficace. Au laboratoire, un crible de sélection basé sur l'allèle *ura2*_{15,30,72} a été développé chez la levure *Saccharomyces cerevisiae*. Il diffère du gène *URA2* par uniquement trois mutations ponctuelles. Il ne fait appel à aucune construction complexe et permet de s'approcher au mieux des conditions naturelles pour lesquelles les remaniements peuvent apparaître dans les génomes. A partir de ce crible génétique, différents remaniements chromosomiques ont été sélectionnés chez *S. cerevisiae* : insertions de rétrotransposons Ty1, délétions et duplications de régions chromosomiques de taille variable.

Dans cette étude, nous nous sommes intéressés aux conséquences de la variabilité génomique intra-espèce sur la formation de remaniements chromosomiques.

Deux contextes génétiques ont été comparés quant à la fréquence d'apparition et aux types de remaniements observés en utilisant le système de sélection basé sur le gène *URA2*. Le contexte FL100 correspond à la souche qui a été utilisée au laboratoire pour toutes les études antérieures tandis que le contexte S288c est le contexte de référence séquencé dans lequel la majorité des analyses fonctionnelles post-séquençage a été réalisée. Nous avons sélectionné des révertants porteurs de remaniements chromosomiques pour ces deux souches et calculé les taux de mutation globaux et spécifiques à chaque type d'événements. Aucune différence statistique pour les taux globaux de mutation n'a pu être observée entre les deux contextes.

Ceci nous permet de conclure que la variabilité génomique entre les deux souches n'a pas d'impact sur la fréquence globale d'apparition des remaniements chromosomiques.

En revanche, aucune réactivation par insertions de rétrotransposon Ty1 n'a pu être sélectionnée en contexte S288c à la température optimale de croissance (30°C). Ce résultat est remarquable car il s'agit d'un événement majeur en contexte FL100. Le nombre de révertants étant voisin entre les deux souches, nous avons donc proposé qu'une baisse d'activité des rétrotransposons dans le contexte S288c serait à l'origine de ce résultat.

Cette hypothèse a été testée en réalisant des sélections en contexte S288c à 25°C. En effet, il a été montré que la rétrotransposition est accrue à une température inférieure à la température optimale de croissance de la levure. A 25°C, quatre révertants porteurs d'insertions de rétrotransposons ont été sélectionnés. Ce résultat permet de conclure que les rétrotransposons sont actifs en contexte S288c mais que leur activité est moindre que celle observée dans les souches FL100.

Dans la suite du travail, il sera intéressant de déterminer les facteurs à l'origine de cette différence d'activité entre les deux contextes. La régulation de la rétrotransposition pourrait être différente entre les contextes FL100 et S288c. Deux types de régulation ont été décrites chez *S. cerevisiae*, l'une transcriptionnelle et l'autre post-transcriptionnelle. Dans un premier temps, il serait informatif de déterminer le taux de transcription des rétrotransposons Ty1 dans les deux contextes afin de déterminer si des différences peuvent être observées. En absence de différence, on pourrait supposer que la régulation pourrait être plutôt post-transcriptionnelle puisque celle-ci n'a pas d'impact sur la quantité de transcrits.

A plus long terme, des sélections pourraient être réalisées dans un nombre plus important de souches de *S. cerevisiae* afin d'affiner l'analyse. En effet, lors de notre étude, la seule différence notable entre les souches FL100 et S288c concerne l'activité des rétrotransposons Ty1. Augmenter la population testée permettra de déterminer si la variabilité peut avoir des conséquences pour les autres types de remaniements observés tels que les délétions ou les duplications ainsi que sur le taux global de mutation. De plus, la variabilité intra-espèces est plus importante pour les souches naturelles. Il serait donc intéressant de tester des souches provenant d'habitats différents comme des souches œnologiques, issues de pâte à pain ou de fromage ainsi que des souches opportunistes isolées à partir de patients immunodéprimés. Ce travail nécessiterait de construire des mutants *ura2*_{15,30,72} dans les diverses souches puis de réaliser des sélections de remaniements chromosomiques. L'inconvénient majeur est que les souches naturelles sont fréquemment diploïdes ou aneuploïdes et homothalliques pouvant compliquer la stratégie notamment pour les constructions des mutants.

La souche RM11 présente un intérêt particulier car elle ne contient pas de rétrotransposons Ty1. Tous les événements qui seraient sélectionnés seraient donc indépendants de la rétrotransposition des éléments Ty1. Il n'y aurait alors ni insertion de rétrotransposons Ty1 ni rétroposition. La comparaison des remaniements chromosomiques obtenus dans RM11 par rapport aux souches pour lesquelles la rétrotransposition des éléments Ty1 est fonctionnelle devrait permettre de comprendre le rôle des éléments transposables dans la flexibilité du génome.

Dans un deuxième temps, nous avons tenté de déterminer les mécanismes conduisant à la mise en place des différents réarrangements de type délétions et duplications. Pour cela, différentes souches ont été construites présentant des délétions pour des gènes dont les protéines sont impliquées dans la recombinaison homologue (*RAD52* et *RAD59*) ou dans le non-homologous end-joining (*YKU80* et *LIG4*). De nombreuses sélections de révertants porteurs de réarrangements ont ensuite été effectuées dans ces différents contextes modifiés. Les résultats que nous avons obtenus nous ont permis de déterminer l'impact des voies de recombinaison sur les taux de mutation observés ainsi que sur les types de remaniements sélectionnés. Nous avons ainsi pu montrer que l'inactivation de la recombinaison homologue par délétion des gènes *RAD52* ou *RAD59* n'a pas d'influence sur l'apparition des délétions et des duplications. Leur mécanisme de formation serait néanmoins favorisé en absence de recombinaison homologue. Par ailleurs pour les délétions analysées, des microhomologies de courtes tailles (1 à 11 pb) ont été mises en évidence à la jonction. Ce résultat suggère fortement qu'un mécanisme de recombinaison non-homologue serait à l'origine de l'apparition des délétions.

Des sélections ont ensuite été menées dans un contexte délété pour le gène *YKU80* dont le produit est impliqué dans le mécanisme de NHEJ. Il a ainsi été montré que l'apparition des délétions et des duplications peut se réaliser de manière indépendante de la protéine Yku80p. En revanche, lorsque le gène *LIG4* codant pour la ligase essentielle au mécanisme de NHEJ est muté, une forte diminution des taux d'apparition des délétions et des duplications est observée. Ce résultat nous a permis de conclure que la protéine Lig4p joue un rôle important dans l'apparition des délétions et des duplications. Il a ainsi été proposé qu'un mécanisme de NHEJ indépendant du complexe Ku formé par l'hétérodimère Yku80p/Yku70p serait à l'origine de la formation des délétions et des duplications en utilisant de courtes séquences répétées. En absence de recombinaison homologue, ce mécanisme est privilégié. L'ensemble des résultats nous permet de supposer qu'un équilibre entre les différents mécanismes de

recombinaison existe dans une cellule sauvage. Lorsque l'une des voies est déficiente, le déséquilibre produit serait alors en faveur des autres mécanismes.

Enfin, nous nous sommes interrogés sur l'existence d'un mécanisme permettant la formation des délétions et des duplications en absence de la recombinaison homologe et du NHEJ. Pour cela, un double mutant *ura2_{15,30,72}Δrad52Δlig4* a été construit et des sélections de révertants porteurs de remaniements chromosomiques ont été effectués dans ce contexte muté. Aucune différence quant au type de remaniements sélectionnés et au taux global de mutation n'a été observée dans ce contexte par rapport à la souche de référence. Ce résultat suggère qu'un troisième mécanisme permet l'apparition des délétions et des duplications. Le mécanisme impliqué pourrait être le microhomology-mediated end-joining (MMEJ). Ce mécanisme repose sur la réparation de cassures d'ADN double-brin en utilisant de courtes séquences répétées. Ces séquences interagissent entre elles et recombinent. Il y a alors perte de la région entre les deux microhomologies. Ce mécanisme nécessite notamment les protéines Rad1p et Rad10p qui forment un complexe à activité endonucléase.

Dans la suite du travail, l'implication de ce mécanisme pourrait être testée en réalisant des sélections de révertants dans des contextes mutés pour ces deux gènes. Néanmoins, ces deux gènes codent pour des protéines pléiotropiques, intervenant dans d'autres voies de recombinaison. L'interprétation des résultats sera plus difficile que dans le cas de protéines impliquées dans une unique voie comme ceci est le cas pour Lig4p, la ligase spécifique du NHEJ.

Par ailleurs, pour obtenir une analyse plus fine des mécanismes impliqués dans la formation des remaniements chromosomiques, il serait utile de travailler non pas avec des souches portant des délétions dans les gènes d'intérêt mais plutôt qui possèdent des mutations faux-sens. Ceci présente notamment un intérêt dans le cas des complexes protéiques. Une mutation faux-sens ciblée au niveau d'un site actif permettrait de maintenir la structure de la protéine tout en perdant son activité. L'environnement protéique pour les autres constituants du complexe ne serait ainsi pas perturbé. L'étude de l'influence de la protéine sur l'apparition des remaniements chromosomiques serait alors réalisée dans des conditions se rapprochant au mieux des conditions naturelles. Cette analyse est plus difficile à mettre en place car elle requiert de construire les souches mutées dans les différents gènes impliqués dans les mécanismes de recombinaison et de connaître les éventuelles positions à muter pour que la structure soit conservée tout en bloquant l'activité de la protéine.

Le système de sélection *ura2*_{15,30,72} est un crible génétique performant pour la sélection de révertants porteurs de remaniements chromosomiques. Lors de cette étude, plus de 800 sélections ayant permis l'obtention d'une centaine de révertants ont été menées. Il présente l'avantage de sélectionner des révertants spontanés. De plus, plusieurs types de remaniements différents peuvent être sélectionnés : délétions, duplications ou insertions de rétrotransposons Ty1. Il constitue un outil puissant permettant de tester de nombreux contextes génétiques différents et utilisable aussi bien en contexte haploïde que diploïde. D'un point de vue pratique, il est peu couteux et relativement rapide à mettre en œuvre.

Ce système présente néanmoins ses limites. Tout d'abord, d'un point de vue expérimental, du fait de la fréquence d'apparition très faible des événements, il est nécessaire d'étaler de nombreuses boîtes de Pétri pour obtenir des révertants indépendants. Par exemple le travail présenté ici a nécessité l'étalement d'environ 5000 boîtes pour l'obtention de 129 révertants. En effet, 5 boîtes doivent être étalées par sélection.

Un système de sélection permettant d'obtenir des remaniements en milieu liquide faciliterait le travail du manipulateur. Dans ce cas, il serait nécessaire d'avoir un marqueur permettant de séparer les cellules porteuses de remaniements des autres cellules. Une possibilité serait d'utiliser un système de sélection reposant sur un gène codant pour une protéine fluorescente comme la GFP (green fluorescent protein) (Figure C1). Le gène codant pour une protéine fluorescente sous contrôle d'un promoteur non-fonctionnel, porteur de trois mutations est inséré dans le génome. Trois mutations au niveau de la région promotrice sont nécessaires pour éviter les événements de réversion spontanée. Dans ce cas, seuls des remaniements chromosomiques devraient être sélectionnés. En effet, on peut supposer que comme dans le cas du crible génétique *ura2*_{15,30,72}, plusieurs types de remaniements chromosomiques pourront être obtenus, permettant ainsi la réactivation du gène marqueur.

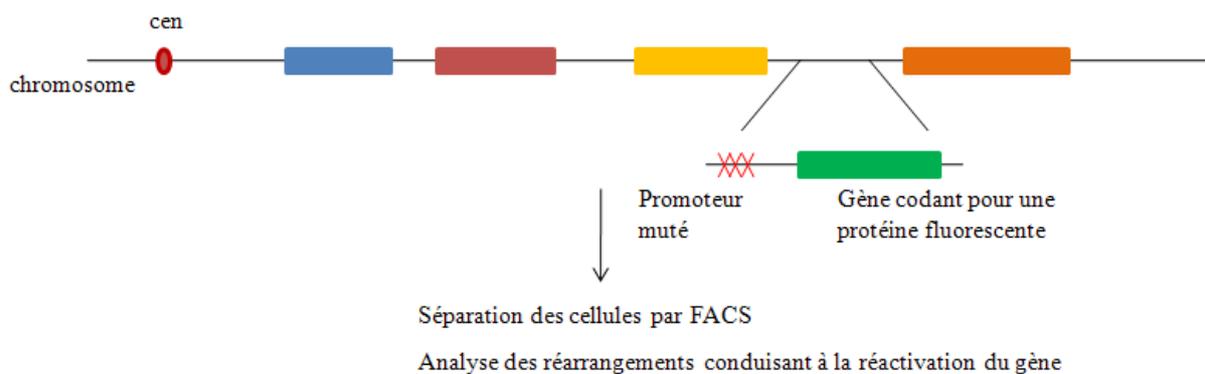


Figure C1 : Système de sélection utilisant un gène codant pour une protéine fluorescente.

Dans la souche de référence, la protéine fluorescente n'est pas produite et aucune fluorescence n'est émise. En revanche si un événement permet la réactivation du gène, le gène sera exprimé et la protéine, ainsi produite, conduira à l'apparition de fluorescence. Ces cellules pourront alors être séparées par la technique de FACS (fluorescence-activated cell sorting). Cette technique de cytométrie de flux permet de séparer des cellules à partir d'un mélange en se basant sur l'intensité de fluorescence. Ce crible de sélection permettrait une sélection moins laborieuse des révertants porteurs de réarrangements chromosomiques.

Un autre point limitant de notre système repose sur l'utilisation d'un gène marqueur. La nécessité de réactivation du gène pour que la cellule puisse survivre induit nécessairement un biais dans le type de remaniements sélectionnés. Certains remaniements ne sont peut-être jamais observés car ils ne permettent pas la réactivation de l'activité ATCase et donc leur sélection.

Une analyse à plus grande échelle sur l'ensemble du génome permettrait de définir des «hotspots» de recombinaison c'est-à-dire des régions plus recombino-gènes. Le système à la GFP permettrait cette étude en insérant notre marqueur avec le promoteur muté à différentes régions du génome puis en comparant les fréquences d'apparition et les types de remaniements obtenus en fonction du lieu d'insertion. Cependant cette étude nécessiterait, comme pour le système de sélection *ura2*_{15,30,72}, de construire les souches porteuses du gène marqueur. Ceci n'est pas toujours aisé notamment pour les souches « non-laboratoires » qui sont généralement plus difficilement manipulables.

Idéalement, il serait intéressant de pouvoir sélectionner des souches ayant acquis des remaniements chromosomiques au sein de leurs génomes au cours du temps en absence de pressions de sélection. Les souches seraient cultivées sur de nombreuses générations. Après un certain temps de culture, les souches seraient analysées et les différents réarrangements obtenus seraient caractérisés. Cette sélection présenterait l'avantage d'obtenir des remaniements spontanés sans contraintes et de se rapprocher ainsi des conditions naturelles dans lesquelles les réarrangements se forment. Néanmoins à l'heure actuelle, aucun système expérimental n'est au point pour pouvoir différencier facilement les souches remaniées, porteuses de remaniements des souches de départ en absence d'un gène marqueur.

Pour conclure, lors de ce travail, nous avons abordé la question de l'influence la variabilité des génomes sur la formation des remaniements génomiques. En effet, si la variabilité représente un moteur pour l'évolution, elle entraîne également des effets plus délétères dans le cas des maladies génétiques ou des cancers chez l'Homme. Nous avons ainsi pu montrer que

le contexte génétique a une influence sur la sélection des remaniements chromosomiques. Par ailleurs, nous avons également déterminé en utilisant le système de sélection *ura2_{15,30,72}* qu'un mécanisme de NHEJ indépendant du complexe Ku est à l'origine de l'apparition des délétions et des duplications.

Partie II

Les familles multigéniques *FUR4/FUI1/DAL4* et
FCY2/FCY21/FCY22 chez les hémiascomycètes

LES FAMILLES MULTIGÉNIQUES *FUR4/FUI1/DAL4* ET *FCY2/FCY21/FCY22* CHEZ LES HÉMIASCOMYCÈTES

1. Introduction

Le nombre croissant de génomes eucaryotes séquencés permet la mise à disposition d'une importante collection de données génomiques. Ces données peuvent être utilisées pour réaliser des comparaisons de génomes et étudier les mécanismes impliqués dans leur évolution. Après le séquençage en 1996 de la levure de boulanger *Saccharomyces cerevisiae*, les programmes de séquençage ont principalement été réalisés pour d'autres organismes tels que *Caenorhabditis elegans* (The *C. elegans* Sequencing Consortium, 1998), *Drosophila melanogaster* (Adams *et al.*, 2000), *Arabidopsis thaliana* (The *Arabidopsis* Genome Initiative, 2000) ou encore *Homo sapiens* (International Human Genome Sequencing Consortium, 2001). Cependant ces espèces appartiennent toutes à des lignées éloignées évolutivement. Des comparaisons inter-espèces pour comprendre l'architecture des génomes ainsi que la détection des remaniements ayant pu conduire à l'apparition des différentes espèces sont rendues ardues voire impossibles, les différences entre les génomes étant trop nombreuses pour être véritablement informatives.

En effet, des remaniements chromosomiques tels que des délétions, des duplications, des inversions ou des translocations vont s'accumuler au cours du temps dans les génomes. Plus le temps de divergence entre deux espèces est grand, plus il sera difficile de les comparer car l'accumulation des réarrangements va conduire à la perte de la base commune. Il est donc plus informatif de travailler avec des génomes d'espèces proches. En effet, deux espèces avec un ancêtre commun proche vont non seulement présenter un pourcentage d'identité de séquences important mais également une architecture des génomes plus ou moins conservée.

Deux gènes seront définis comme homologues quand ils proviennent d'un ancêtre commun. Ces gènes peuvent être des orthologues c'est-à-dire qu'ils sont issus d'un ancêtre commun et sont présents dans des espèces différentes. Ils seront dits paralogues s'ils dérivent d'un ancêtre commun et sont présents dans la même espèce (Figure 4.1). Par extension, ces termes sont utilisés dans de nombreuses études pour décrire des gènes ou des protéines présentant un fort pourcentage d'identité ou de similarité respectivement. Dans ce cas, la définition repose uniquement sur des données génomiques. Elle est basée sur le postulat que si des séquences ont un pourcentage d'identité élevé, elles dérivent probablement d'un même gène ancestral.

Cette définition élargie est à utiliser avec précaution. En effet, certains gènes peuvent provenir d'un ancêtre commun et ne présenter qu'un faible pourcentage d'identité, notamment dans le cas d'espèces éloignées. A l'inverse, certains gènes peuvent présenter un pourcentage d'identité significatif supérieur à 30% et n'avoir aucun ancêtre commun. C'est le cas par exemple des deux gènes codant pour des dihydroorotate déhydrogénases (DHODase) chez *Saccharomyces kluyveri*. Les produits de ces deux gènes ont des localisations cellulaires différentes, l'une étant mitochondriale tandis que l'autre est située dans le cytoplasme. Le gène codant pour la DHODase cytoplasmique proviendrait d'un transfert horizontal à partir du gène de *Lactococcus lactis* (Gojković *et al.*, 2004). Les produits de ces deux gènes présentent un pourcentage de similarité de 32,7%, pourtant on peut conclure qu'ils ne sont pas apparentés et ne peuvent donc pas être considérés comme paralogues. Dans la suite du texte, les termes homologues, paralogues et orthologues seront utilisés d'après la définition élargie.

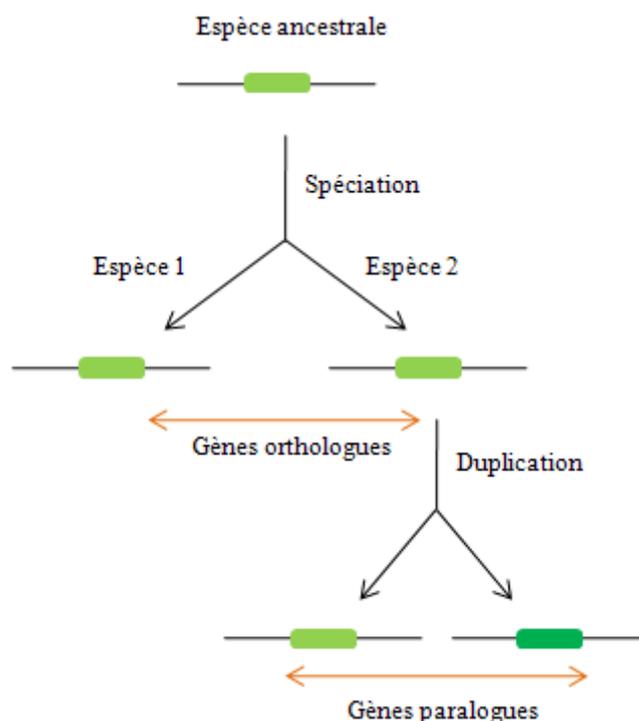


Figure 4.1 : Définition des gènes orthologues et paralogues.

Le séquençage de nombreux génomes dans le groupe des hémiascomycètes représente un exemple de projet permettant la comparaison d'espèces proches. Cette étude a débuté par le programme Génolevures I (Souciet *et al.*, 2000) qui a consisté en un séquençage avec une couverture génomique faible sans assemblage (0,4×) des génomes de 13 levures représentant l'ensemble des clades des hémiascomycètes (Figure 4.2). Des espèces appartenant au groupe des *Saccharomyces sensu stricto* (clade 1), auquel appartient *Saccharomyces cereviae*, ont

ensuite été séquencées avec une couverture génomique plus importante de 3 à 7 × (Kellis *et al.*, 2003 ; Cliften *et al.*, 2003). Par la suite, d'autres séquences de génomes de levures du groupe des hémiascomycètes ont été obtenues. Le séquençage d'*Ashbya* (*Eremothecium*) *gossypii* appartenant au clade 12 a été réalisé en 2004 (Dietrich *et al.*, 2004), celui de *Kluyveromyces polysporus* (clade 6) en 2007 (Scannell *et al.*, 2007).

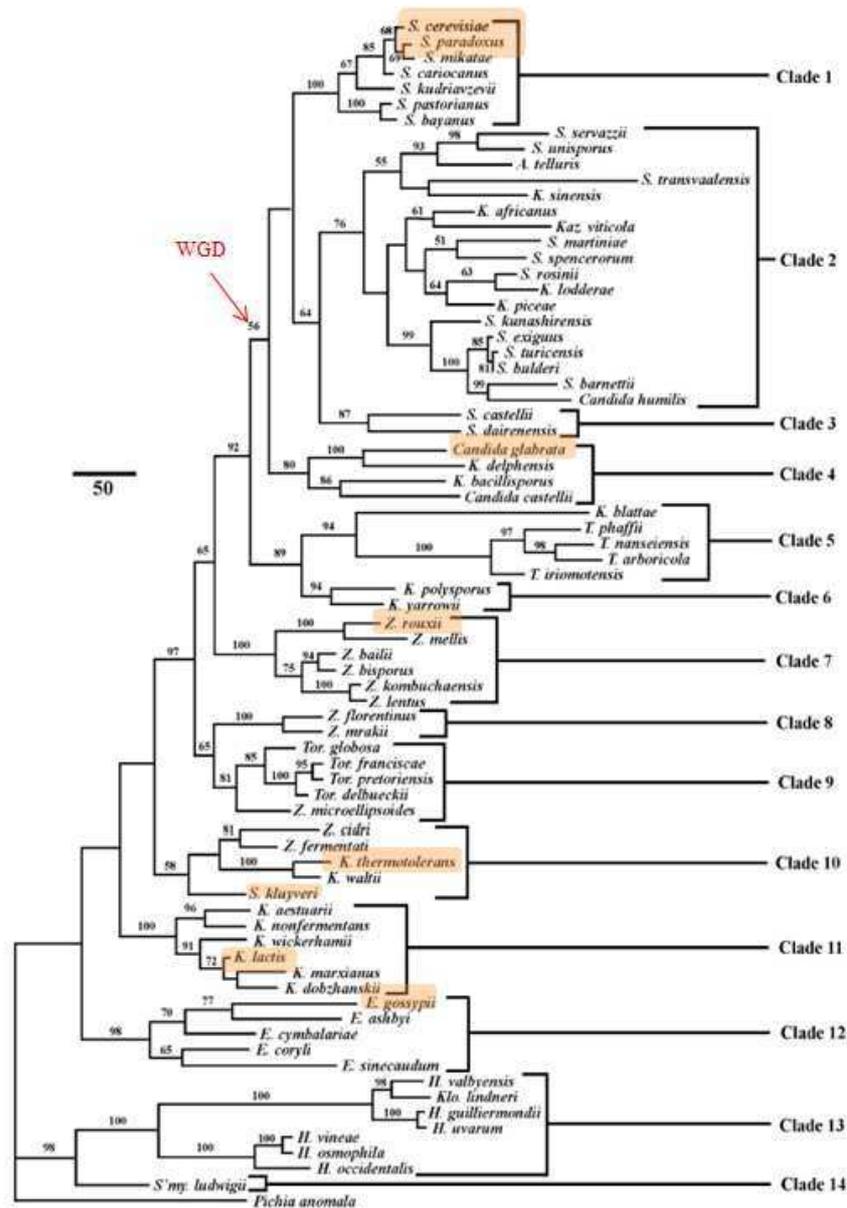


Figure 4.2 : Arbre phylogénétique représentant le groupe des *Saccharomyces* (d'après Kurtzman et Robnett, 2003). Les espèces plus distantes comme *Y. lipolytica* et *D. hansenii* ne sont pas représentées sur l'arbre.

L'arbre a été construit en utilisant les séquences nucléotidiques des ARNs ribosomiques 18S, 5,8S et 26S (3 régions), EF1- α , COXII et l'ARN ribosomique mitochondrial.

WGD (whole genome duplication) représente l'événement de duplication totale du génome.

Le projet Génolevures 2 a permis la détermination de la séquence complète des génomes de *Candida glabrata* (clade 4), *Kluyveromyces lactis* (clade 11), *Debaryomyces hansenii* et *Yarrowia lipolytica* (Dujon *et al.*, 2004). Ces projets ont permis d'obtenir une collection de données et de réaliser des comparaisons entre les génomes.

Il a ainsi été mis en évidence que les génomes de ces levures présentent un certain nombre de points communs au niveau de la structure. Tout d'abord elles possèdent toutes des génomes compacts de petite taille, *Ashbya gossypii* également nommée *Eremothecium gossypii* possédant le génome le plus petit (9,2 Mb) et *Yarrowia lipolytica* le plus grand avec 20,5 Mb. Leurs génomes présentent un faible nombre d'introns, un nombre réduit de transposons. Enfin, ils ne comportent que très peu d'éléments répétés non-codants. En effet, La densité des gènes chez ces différentes espèces est très grande puisque l'on trouve en moyenne un gène tous les 2 kb (contre un gène tous les 40-45 kb chez l'Homme par exemple).

Au niveau de l'architecture des génomes, la synténie, c'est-à-dire l'organisation des gènes les uns par rapport aux autres, est très bien conservée dans le groupe des *Saccharomyces sensu stricto*. La majorité des gènes de *S. cerevisiae* possède un gène présentant une forte identité de séquences dans les espèces *S. paradoxus*, *S. mikatae* et *S. bayanus*. De plus, les blocs de synténie couvrent 90% du génome de *S. cerevisiae* dans ces trois espèces. La comparaison de l'architecture de ces génomes montre que seul un faible nombre d'événements dont des translocations réciproques, des duplications segmentales et des inversions les différencient. Par exemple, le génome de *S. paradoxus* diffère de celui de *S. cerevisiae* uniquement par quatre inversions et trois duplications segmentales (Kellis *et al.*, 2003). Pour les autres espèces plus éloignées, les blocs de synténie sont moins bien conservés. Ainsi la comparaison des cartes génomiques entre *S. cerevisiae* et la levure pathogène *Candida glabrata* appartenant au groupe des *Saccharomyces sensu lato* (clade 4) indique qu'un nombre de réarrangements chromosomiques plus important que celui observé chez les *sensu stricto* s'est produit dans les deux espèces depuis le moment où elles ont divergé. Pour *Yarrowia lipolytica*, l'espèce la plus éloignée, la synténie est totalement perdue par rapport à *S. cerevisiae*.

L'analyse des génomes d'espèces hémiascomycètes plus éloignées appartenant au groupe des *Kluyveromyces* (clades 10 et 11) a permis de mettre en évidence un événement majeur dans l'évolution des génomes de ce groupe. En effet, il a été montré que le génome de *S. cerevisiae* tel qu'on le connaît à l'heure actuelle résulte d'une duplication ancestrale complète de l'ensemble des chromosomes (Wolfe et Shields, 1997) (Figure 4.2). Les analyses génomiques et notamment les comparaisons avec d'autres espèces du clade des *Kluyveromyces*

(Génolevures 3) ont révélé que cette duplication complète du génome se serait produite après la divergence entre les *Saccharomyces* et *Kluyveromyces* (Wolfe et Shields, 1997). On va donc définir des espèces pré-duplication telles que le groupe des *Kluyveromyces*, *Ashbya gossypii*, *Debaryomyces hansenii* ou *Y. lipolytica*. Par contre, le groupe des *Saccharomyces sensu stricto* et des *Saccharomyces sensu lato* font partie des espèces post-duplications. On retrouve dans ces génomes des blocs de duplication définis comme fossiles, c'est-à-dire correspondant aux traces de l'événement de duplication ancestrale (Figure 4.3). Cinquante-cinq blocs de duplication contenant 376 paires de gènes appelés ohnologues ont été mis en évidence chez *S. cerevisiae* (Wolfe et Shields, 1997).

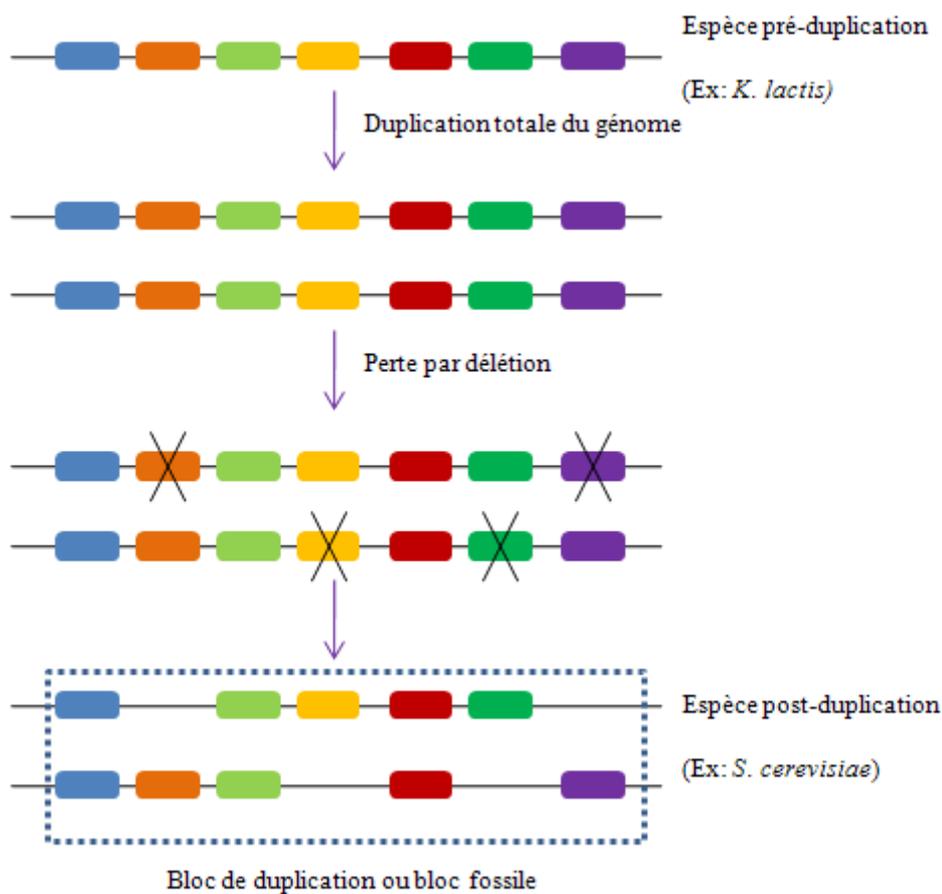


Figure 4.3 : Représentation de l'événement de duplication totale du génome chez les hémiascomycètes.

Le projet Génolevures 3 a consisté à la détermination de la séquence complète et l'annotation des génomes de quatre espèces pré-duplications appartenant au groupe des *Kluyveromyces* à savoir *Zygosaccharomyces rouxii* (clade 7), *K. thermotolerans* (clade 10), *Saccharomyces kluyveri* (clade 10) et *K. lactis* (clade 11). La phase d'annotation a été réalisée de façon manuelle en utilisant le génome de *S. cerevisiae* comme référence. Chaque protéine putative

codée par les gènes des quatre espèces a été comparée avec celles de *S. cerevisiae*. Des pourcentages de similarité ont ainsi pu être déterminés. Les gènes ont été regroupés par famille définies par les taux de similarité et également par les fonctions putatives des protéines codées par ces gènes. L'annotation est également passée par la recherche d'introns éventuels dans les différents gènes ainsi que celle de pseudogènes (gènes ayant accumulé les mutations et perdu leurs fonctions). Dans le cadre de ce projet, j'ai pris part, durant une période de trois mois, à l'annotation de 1200 CDS (coding sequence) sur les 24000 présentes au total.

A partir des données obtenues par l'ensemble du consortium, nous nous sommes intéressés au devenir des duplications de gènes dans des espèces pré- et post-duplication. Pour ce faire, notre analyse a porté sur deux familles multigéniques codant pour des perméases : la famille *FUR4/FUI1/DAL4* (famille des perméases à l'uracile, l'uridine et l'allantoïne) et la famille *FCY2/FCY21/FCY22* (famille de perméases aux purines et à la cytosine) dont les fonctions exactes chez *S. cerevisiae* seront décrites par la suite.

Les espèces considérées couvrent l'ensemble de l'arbre phylogénétique des hémiascomycètes. *S. cerevisiae* sera l'espèce utilisée comme référence pour les différentes comparaisons. *S. paradoxus* représente le groupe des *Saccharomyces sensu stricto* (clade 1). Au sein des *Saccharomyces sensu lato*, le choix a porté sur *Candida glabrata* (clade 4) dont le génome a été séquencé lors du programme Génolevures 2. Elle est le second agent pathogène responsable de candidoses chez l'Homme après *C. albicans*. Les quatre nouvelles espèces séquencées et annotées dans le programme Génolevures 3 ont été ajoutées à l'étude et représentent les clades 7, 10 et 11 des *Kluyveromyces*. Enfin, nous avons choisi trois espèces plus distantes : *Ashbya gossypii*, un pathogène de plantes (clade 12) (Dietrich *et al.*, 2004), *Debaryomyces hansenii*, levure halotolérante et *Yarrowia lipolytica*, levure utilisant les alcanes comme source de carbone (Dujon *et al.*, 2004).

2. La famille *FUR4/FUI1/DAL4*

2.1 Les gènes *FUR4*, *FUI1* et *DAL4* chez *Saccharomyces cerevisiae*

Le gène *FUR4* (*YBR021w*) code pour la perméase spécifique de l'uracile. Le gène a été caractérisé chez *Saccharomyces cerevisiae* au début des années 70 à partir de mutants résistants au 5-fluorouracile (Jund et Lacroute, 1970). Dans les années 80, il a été cloné et séquencé (Chevallier, 1982, Jund *et al.*, 1988). Il code pour la protéine Fur4p de 633 aa qui transporte l'uracile en symport avec des protons. Une levure prototrophe est capable de

synthétiser *de novo* l'uridine triphosphate par la voie de biosynthèse des pyrimidines, néanmoins elle peut aussi importer l'uracile exogène présent dans le milieu de culture par l'intermédiaire de la perméase Fur4p (Figure 4.4). Cette perméase est également responsable de l'import d'un analogue de l'uracile, le 5-fluorouracile (5FU). Une fois dans la cellule, celui-ci est transformé en 5-fluorouridine monophosphate par l'uracile phosphoribosyltransférase et être intégré dans les ARN conduisant à une perte de leur stabilité. Ceci conduit notamment à une synthèse protéique erronée dans le cas des ARN messagers, pouvant ainsi résulter en la mort de la cellule.

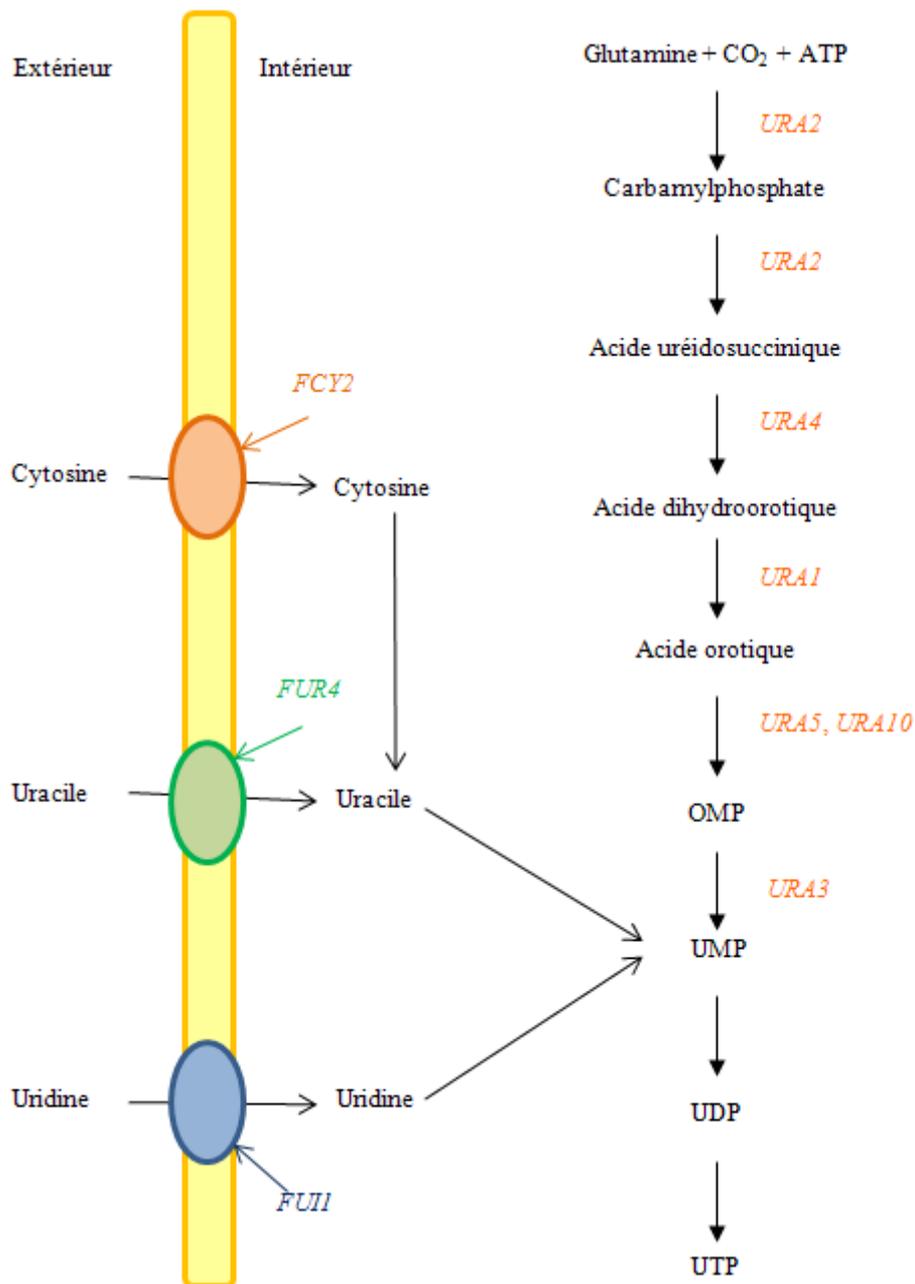


Figure 4.4 : Voie métabolique des pyrimidines.

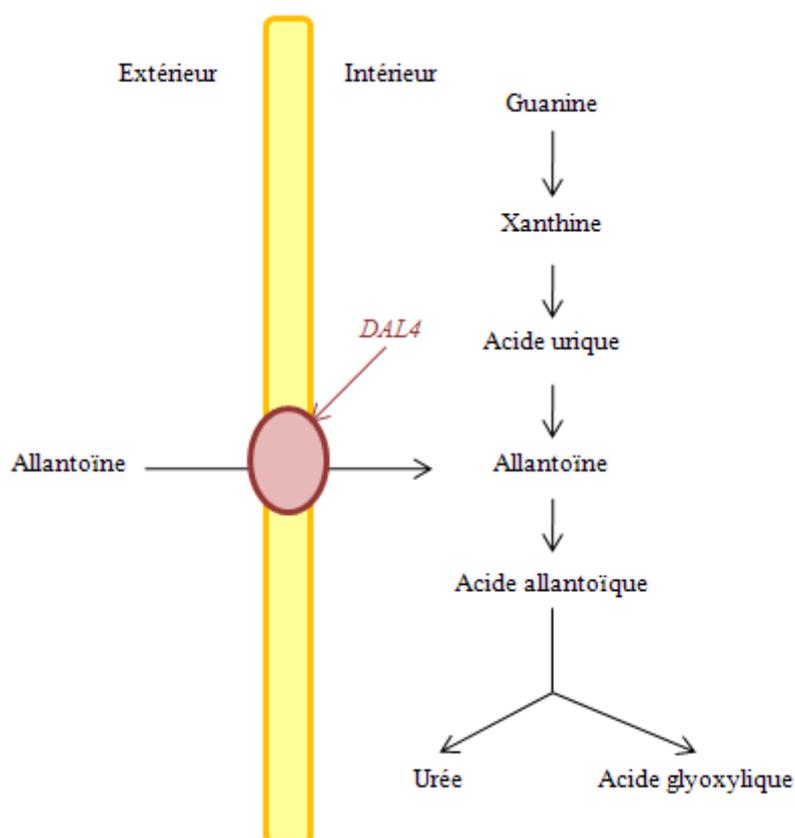


Figure 4.5 : Voie métabolique de l'allantoïne.

Le gène *FUR4* fait partie du bloc de duplication 3 défini par Wolfe et Shields (1997). Ce bloc est composé de deux régions dupliquées, l'une située sur le chromosome II et l'autre sur le chromosome IV, comportant 14 paires de paralogues. Néanmoins, l'orthologue de *FUR4* est absent chez *Saccharomyces cerevisiae* qui ne présente donc plus qu'un seul gène.

Le gène *FUI1* (*YBL042c*) code pour la perméase spécifique de l'uridine (Figure 4.4). Le gène et la fonction de la protéine correspondante ont été caractérisés par Wagner *et al.* (1998). Elle permet également l'entrée dans la cellule de l'analogue de l'uridine, la 5-fluorouridine. Cette molécule sera convertie en 5-fluorouridine monophosphate conduisant aux mêmes conséquences délétères décrites précédemment.

Le gène *DAL4* (*YIR028w*) code pour une protéine de 635 aa. Il s'agit de la perméase spécifique de l'allantoïne (Sumrada *et al.*, 1977), substrat utilisé par *S. cerevisiae* comme source d'azote et issu de la dégradation des purines (Figure 4.5). Il a été montré que le gène *DAL4* est réprimé sur milieu complet ainsi l'allantoïne n'est que très faiblement

importée et ne représente donc pas la source majeure d'azote pour la cellule (Cooper *et al.*, 1987). En revanche, si l'allantoïne est la seule source d'azote présente dans le milieu, la faible quantité importée induit l'expression du gène *DAL4* ce qui permet ensuite un transport efficace du substrat.

Les pourcentages d'identité pour les séquences nucléiques et de similarité en acides aminés sont présentés dans le tableau 4.1. Du fait de la forte ressemblance des gènes, il a été proposé que la famille multigénique *FUR4/FUI1/DAL4* proviendrait de duplications d'un gène ancêtre possédant un spectre large permettant l'import de l'uracile, l'uridine et l'allantoïne. Il a été montré que dans certaines circonstances, la perméase Fur4p est capable d'importer les trois substrats et non uniquement l'uracile. Cette observation permet de proposer que le gène *FUR4* serait le plus proche du gène ancêtre et que *FUI1* et *DAL4* correspondraient aux copies ayant divergé et acquis une spécificité de substrat (Wagner, 1999).

Protéines	Fur4p	Fui1p	Dal4p
Fur4p	100.0%	64,0% (46,2%)	81,7% (67,0%)
Fui1p	64,0% (46,2%)	100.0%	62,9% (45,7%)
Dal4p	81,7% (67,0%)	62,9% (45,7%)	100.0%
Acides nucléiques	<i>FUR4</i>	<i>FUI1</i>	<i>DAL4</i>
<i>FUR4</i>	100.0%	56,10%	66,90%
<i>FUI1</i>	56,10%	100.0%	53,90%
<i>DAL4</i>	66,90%	53,90%	100.0%

Tableau 4.1 : Pourcentages d'identité obtenus pour les gènes *FUR4*, *FUI1*, *DAL4* et pourcentages de similarité pour leurs produits Fur4p, Fui1p et Dal4p. Dans le cas des protéines, le nombre indiqué entre parenthèses indique le pourcentage d'identité. Les pourcentages ont été déterminés par alignement global d'après l'algorithme de Needleman et Wunsch.

2.2 La famille *FUR4/FUI1/DAL4* chez les hémiascomycètes

2.2.1 Recherche de gènes homologues chez les hémiascomycètes

Nous nous sommes intéressés au nombre de gènes possédant une identité de séquences avec les gènes codant pour les perméases chez *S. cerevisiae* dans les différentes levures afin de déterminer si elles possèdent ou non trois gènes codant pour les différentes perméases. Les gènes présentant une identité de séquence importante pour *FUR4*, *FUI1* et

DAL4 ont été recherchés à partir de leurs produits par similarité de séquences protéiques avec celles des protéines trouvées chez *S. cerevisiae* en utilisant la méthode heuristique BLASTP (Altschul *et al.*, 1990). Cette recherche a été effectuée dans les banques de données issues des séquençages partiels (SGD) ou complets (Génolevures) de levures hémiascomycètes (Figure 4.2). *Saccharomyces paradoxus* possède trois gènes présentant une identité significative avec *FUR4*, *FUI1* et *DAL4* tout comme *S. cerevisiae* (clade 1) (Tableau 4.2). Chez *Candida glabrata* (clade 4) ainsi que chez *Kluyveromyces lactis*, *Kluyveromyces thermotolerans*, *Zygosaccharomyces rouxii* et *Ashbya gossypii* (clades 7, 10 et 11), seuls deux gènes homologues sont retrouvés. Enfin, dans les espèces plus éloignées à savoir *Debaryomyces hansenii* et *Yarrowia lipolytica*, un unique gène présentant une identité de séquence importante a été mis en évidence. *Saccharomyces kluyveri* constitue un cas particulier puisque cette espèce possède cinq gènes soit deux de plus que *S. cerevisiae*.

Espèces		Nombres de gènes
<i>Saccharomyces sensu stricto</i>	<i>S. cerevisiae</i>	3
	<i>S. paradoxus</i>	3
<i>Saccharomyces sensu lato</i>	<i>C. glabrata</i>	2
Clade <i>Kluyveromyces</i>	<i>K. lactis</i>	2
	<i>K. thermotolerans</i>	2
	<i>S. kluyveri</i>	5
	<i>Z. rouxii</i>	2
Espèces plus distantes	<i>A. gossypii</i>	2
	<i>D. hansenii</i>	1
	<i>Y. lipolytica</i>	1

Tableau 4.2 : Nombre de gènes présentant une identité significative avec les gènes *FUR4*, *FUI1* et *DAL4* de *S. cerevisiae* dans le génome des autres levures hémiascomycètes séquencées.

L'absence d'un ou deux gènes chez les espèces plus éloignées de *S. cerevisiae* peut s'expliquer par le fait que certaines des perméases sont absentes et que l'import de certains substrats exogènes n'est pas effectué dans ces espèces. Une autre explication serait que les perméases présentes ont un spectre de substrats plus large que ce qui est observé chez *S. cerevisiae* et *S. paradoxus*. L'étude *in vivo* des substrats importés dans la cellule pour ces différentes espèces permettra de donner des éléments de réponse. Il serait nécessaire de

réaliser des mutants knock-out des différents gènes afin de déterminer leur impact sur le transport de l'uracile, l'uridine et l'allantoïne.

Les pourcentages de similarité entre les protéines des différents gènes considérés sont présentés dans le tableau 4.3.

Protéines	SAPA			CAGL	
	Spar_c197_1075	Spar_c447_511	Spar_c29_11377	CAGL0B04433g	CAGL0F04499g
Fur4p	99,7% (97,8%)	31,7% (23,8%)	81,9% (67,1%)	82,4% (72,8%)	66,9% (47,5%)
Fui1p	64,7% (46,6%)	42,9% (42,1%)	62,6% (45,2%)	64,5% (45,4%)	77,0% (65,7%)
Dal4p	81,4% (67,2%)	31,6% (23,1%)	99,4% (98,3%)	78,3% (63,0%)	63,1% (45,8%)
	KLLA		ZYRO		DEHA
	KLLA0D03454g	KLLA0E05479g	ZYRO0A04092g	ZYRO0G04092g	DEHA2G08426g
Fur4p	72,6% (56,4%)	64,3% (44,5%)	75,9% (59,6%)	63,2% (45,6%)	59,4% (44,7%)
Fui1p	63,9% (45,1%)	67,1% (52,9%)	59,9% (40,3%)	71,6% (56,2%)	61,9% (43,2%)
Dal4p	74,3% (57,3%)	63,1% (45,1%)	73,1% (57,6%)	62,0% (42,6%)	61,4% (46,7%)
	SAKL				
	SAKL0C11594g	SAKL0F00286g	SAKL0G04532g	SAKL0H03476g	SAKL0H03498g
Fur4p	80,3% (66,1%)	64,5% (47,6%)	69,5% (54,6%)	66,1% (48,6%)	66,4% (48,8%)
Fui1p	67,2% (49,3%)	70,0% (57,4%)	65,4% (47,4%)	71,9% (58,9%)	74,3% (62,8%)
Dal4p	78,9% (66,5%)	63,5% (46,7%)	69,6% (54,5%)	66,2% (49,0%)	63,6% (46,8%)
	KLTH		ASGO		YALI
	KLTH0D13926g	KLTH0G03432g	ABR124c	AFR598c	YALI0D05621g
Fur4p	65,0% (47,5%)	70,4% (53,8%)	68,5% (50,2%)	61,3% (42,4%)	59,4% (42,1%)
Fui1p	67,2% (53,3%)	63,7% (44,7%)	55,4% (38,7%)	72,4% (57,9%)	56,3% (39,6%)
Dal4p	63,8% (48,0%)	69,5% (55,8%)	65,4% (49,5%)	61,5% (42,6%)	57,6% (41,8%)

Tableau 4.3 : Pourcentages de similarité observés pour les protéines des gènes homologues à *FUR4*, *FUI1* et *DAL4*.

SAPA : *Saccharomyces paradoxus*, CAGL : *Candida glabrata*, SAKL : *Saccharomyces kluyveri*, KLLA : *Kluyveromyces lactis*, ZYRO : *Zygosaccharomyces rouxii*, KLTH : *kluyveromyces thermotolerans*, ASGO : *Ashbya gossypii*, DEHA : *Debaryomyces hanseni* et YALI : *Yarrowia lipolytica*.

Le nombre indiqué entre parenthèses correspond au pourcentage d'identité. Les pourcentages ont été déterminés d'après les alignements globaux réalisés avec l'algorithme de Needleman et Wunsch.

Dans le cas de *S. paradoxus*, deux des trois protéines trouvées par BLASTP présentent un pourcentage de similarité supérieur à 90%, l'une pour Fur4p et l'autre pour Dal4p. En revanche pour le troisième candidat, la similarité la plus importante est observée avec la protéine Fui1p. Cependant elle reste faible puisqu'elle est inférieure à 50%. Si l'on observe l'alignement plus en détail (Figure 4.6), on constate que la protéine de *S. paradoxus* ne s'aligne que sur la région centrale de la protéine Fui1p. De plus, la protéine de *S. paradoxus* (*Spar_447_511*) a une taille de 276 aa tandis que Fui1p est composée de 639 aa. Une étude *in vivo* de l'import de l'uridine chez *S. paradoxus* permettrait de déterminer si cette protéine est active ou non. Dans le cas où l'import ne serait pas effectué, on pourrait alors supposer que le gène *Spar_447_511* est un pseudogène ou qu'il code pour une autre fonction que celle de perméase à l'uridine.

Fui1p	201	IYFSIWIVINRVVMACVWNSTLAYIGSQCVQLMLKAIFGTNLNTRIKDTI	250
Spar_447_511	1	MACVWNSTLAYIGSQCVQLMLKAIFGVDLNTRIKDTI	37
Fui1p	251	KNPNLTNFEFMCVMVFWVACLPPFLWFPPDKLRHIFALKSAITPFAAFGFL	300
Spar_447_511	38	KNPNLTNFEFMCVMVFWVACLPPFLWFPPDKLRHVFALKSAITPFAAFGFL	87
Fui1p	301	IWTLCRAKAGHLALGSLNDNGGAIKTVLAWSVIRAIMSALDNFSTLILNA	350
Spar_447_511	88	IWTLCRAKAGHLALGSLNDNGGAIKTVLAWSVIRAIMSALDNFSTLILNA	137
Fui1p	351	PDFTRFGKTYKSSVYSQLIALPVCYAIISLIGILSVSAAYTLYGVNYWSP	400
Spar_447_511	138	PDFTRFGKTYKSSVYSQLVALPVCYGIISLIGILSVSAAYTLYGINYWSP	187
Fui1p	401	LDILNRYLDNYTSGNRAGVFLISFIFAFDQLGANLSCNSIPAGTDLTALL	450
Spar_447_511	188	LDILNRYLDNYTSGNRAGVFLISFIFAFDQVGANLSCNSIPAGTDLTALL	237
Fui1p	451	PKFINIRRGSYICALISLAICPWDLLSSSSKFTTALAAYAVFLSAIAGVI	500
Spar_447_511	238	PKFINIRRGSYICALISLAICPWDLLSSSSKFTTALAAY	276
Fui1p	501	SADYFIVRKGVMNIFHCYTDKPGSYMYNKYGTNWRRAVVAYIFGIAPNFA	550
Spar_447_511	277		276

Figure 4.6 : Comparaison de la séquence de Fui1p avec celle du produit du gène Spar_447_511. L'alignement global a été réalisé d'après l'algorithme de Needleman et Wunsch.

2.2.2 Synténie des différents gènes homologues

La synténie pour un ensemble de gènes est dite conservée entre deux espèces si l'ordre des gènes est le même entre les deux. On parlera de rupture de synténie si cet ordre n'est pas respecté.

Nous avons déterminé les régions contenant les différents gènes homologues trouvés par BLASTP en utilisant les bases de données Génolevures et SGD. Nous avons ensuite déterminé les conservations de synténie entre les différentes espèces. Si dans une espèce donnée, la synténie est conservée pour un des gènes, on pourra considérer d'après la définition élargie qu'il s'agit vraisemblablement de l'orthologue du gène présent chez *S. cerevisiae*. Les synténies obtenues sont présentées dans la Figure 4.7.

Chez les *Saccharomyces sensu stricto*, on constate que les régions environnantes des gènes codant les perméases sont très similaires entre *S. cerevisiae* et *S. paradoxus*. En revanche, pour les autres espèces comme celle du groupe des *Kluyveromyces*, les synténies sont beaucoup moins bien conservées. Ceci peut s'expliquer par le fait que plus les espèces sont distantes, plus la synténie sera perdue du fait de l'accumulation des remaniements chromosomiques.

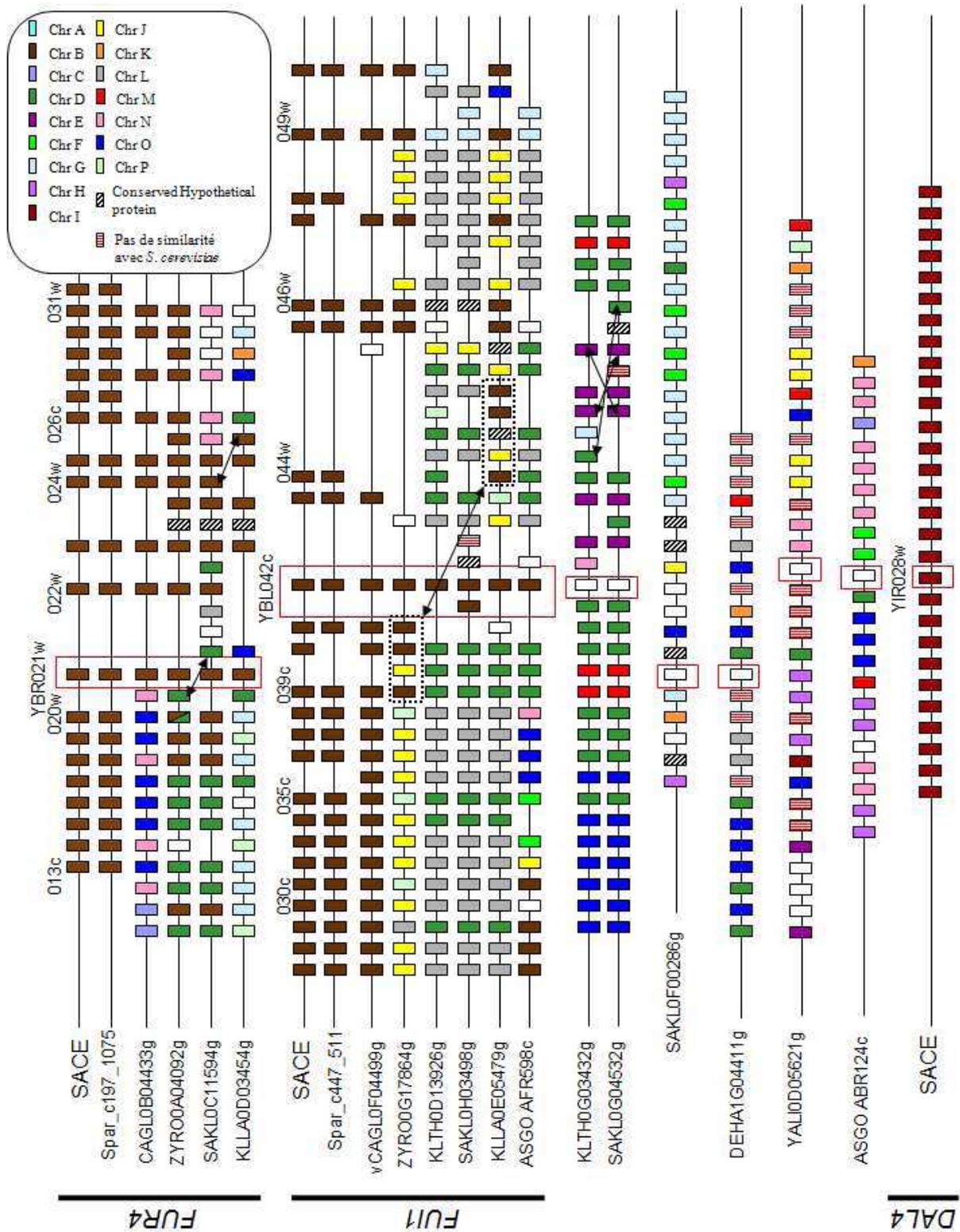


Figure 4.7 : Carte des synténies pour les gènes *FUR4*, *FUI1* et *DAL4* et leurs homologues dans les espèces étudiées. Chaque rectangle représente un gène. Les couleurs des gènes ont été attribuées par correspondance avec les chromosomes de *S. cerevisiae*. Les flèches représentent les événements d'inversions.

Pour le gène *FUR4*, quatre espèces présentent un orthologue à *S. cerevisiae* : *C. glabrata*, *Z. rouxii*, *S. kluyveri* et *K. lactis*. D'un point de vue fonctionnel, on peut donc supposer que ces quatre espèces possèdent une perméase à l'uracile. Chez *C. glabrata*, la région en aval de l'orthologue à *FUR4* est bien conservée puisque huit gènes consécutifs sur dix présents chez *S. cerevisiae* sont retrouvés. En revanche, il y a rupture de synténie pour toute la région en amont du gène. Pour *Z. rouxii* et *S. kluyveri*, les synténies sont bien conservées de part et d'autres du gène orthologue. *FUR4* fait partie du bloc de duplication 3 chez *S. cerevisiae* qui est localisé sur les chromosomes II et IV. Il est à noter que pour *Z. rouxii* et *S. kluyveri* qui sont des espèces pré-duplication, la synténie est constituée de l'imbrication de gènes orthologues à ceux trouvés sur ces deux chromosomes chez les espèces post-duplications. Dans le cas de *K. lactis*, quatre gènes consécutifs sont conservés en aval de l'orthologue à *FUR4*. En revanche, il y a rupture de synténie pour la région en amont du gène.

Pour le gène *FUII*, trois espèces présentent une conservation de synténie. Il s'agit de *C. glabrata*, pour laquelle l'ordre et l'orientation des gènes sont très bien conservés, *Z. rouxii* et *K. lactis*, qui présentent une conservation de synténie partielle. Dans le cas de *K. lactis*, l'orientation de l'ordre des gènes conservés est inversée par rapport à celle trouvée chez *S. cerevisiae*. On peut supposer que ces trois espèces possèdent donc une perméase à l'uridine. Pour *K. thermotolerans* et *S. kluyveri*, il n'y a pas de conservation de synténie avec *S. cerevisiae*. En revanche, la région en amont des gènes homologues trouvés dans ces deux espèces est conservée par rapport à celle présente chez *K. lactis*. Il est à noter que chez *S. kluyveri*, les deux gènes homologues à *FUII* sont présents en tandem direct. Enfin pour *A. gossypii*, il n'y a pas de conservation de synténie avec *S. cerevisiae* mais la région en aval de l'orthologue est similaire à celle trouvée chez *K. thermotolerans* et *S. kluyveri*.

Aucune espèce, à l'exception de *S. paradoxus*, ne présente de conservation de synténie avec le gène *DAL4* de *S. cerevisiae*. Ce résultat laisse supposer que soit les différentes espèces ne possèdent pas de perméase à l'allantoïne soit la synténie n'a pas été conservée au cours de l'évolution. Aucune espèce, à l'exception de *S. paradoxus* et *S. kluyveri*, ne possède trois gènes ayant un pourcentage de similarité significatif par rapport à *FUR4*, *FUII* et *DAL4*. Cette observation suggère que ces espèces ne possèderaient pas la perméase à l'allantoïne.

Pour les espèces comme *D. hansenii* ou *Y. lipolytica*, la conservation de la synténie est inexistante. L'accumulation de réarrangements au cours de l'histoire évolutive rend impossible l'étude de la synténie entre ces souches et *S. cerevisiae* car les blocs de synténie ont disparu.

2.2.3 Etude *in vivo* des perméases à l'uracile, l'uridine et l'allantoïne chez les hémiascomycètes

Le gène *FUR4* code pour la perméase spécifique de l'uracile. Le 5-fluorouracile, analogue de l'uracile, peut également être importé dans la cellule par cette perméase. Une fois dans la cellule, le 5FU est converti en 5-fluorouridine monophosphate. Ce composé est alors toxique pour la cellule et entraîne la mort de la cellule à une concentration supérieure à 10^{-5} M chez *S. cerevisiae*.

Afin de déterminer si les différentes espèces utilisées pour l'analyse *in silico* possèdent une perméase à l'uracile, nous avons testé la résistance de ces souches à une gamme de concentrations de 5FU. Si la perméase est présente, le 5FU sera importé dans la cellule et l'on pourra déterminer la concentration entraînant un effet toxique. Pour les *Saccharomyces sensu stricto*, l'espèce testée est *S. paradoxus*. *C. glabrata* est l'espèce utilisée pour le groupe des *sensu lato*. Trois espèces de *Kluyveromyces* ont été analysées, à savoir *K. lactis*, *K. thermotolerans* et *Zygosaccharomyces rouxii* ainsi qu'une espèce distante *Debaryomyces hansenii*. Ces espèces ont été choisies pour des raisons expérimentales. Il était impératif que ces souches soient toutes prototrophes pour l'uracile. En effet, l'ajout d'uracile dans les milieux aurait faussé les résultats car les concentrations de 5FU seraient diluées.

Les résultats des différents tests sont présentés dans le tableau 4.4. La souche de *S. cerevisiae* est sensible au 5FU à partir d'une concentration de 10^{-5} M. Une souche de *S. cerevisiae* résistante au 5FU, car mutée pour le gène codant pour la perméase à l'uracile a été utilisée comme témoin de résistance. Elle présente une sensibilité à partir d'une concentration en 5FU supérieure à 10^{-3} M. En effet, à cette concentration, le composé peut pénétrer dans la cellule par diffusion passive.

Toutes les souches testées présentent une résistance inférieure à la souche témoin de résistance avec un maximum de résistance pour *D. hansenii* pour une concentration de 10^{-4} M. La souche la moins résistante est *S. paradoxus* qui devient sensible au 5FU pour des concentrations supérieures à 10^{-6} M.

Ces résultats nous permettent de conclure que pour toutes les souches testées, le 5FU peut être activement importé dans la cellule. Toutes les espèces testées présentent donc vraisemblablement une perméase à l'uracile.

Concentration 5FU	5×10 ⁻⁷	10 ⁻⁶	5×10 ⁻⁶	10 ⁻⁵	5×10 ⁻⁵	10 ⁻⁴	5×10 ⁻⁴	10 ⁻³	5×10 ⁻³	10 ⁻²	Ø
SACE	+	+	+	-	-	-	-	-	-	-	+
SACE Résistante	+	+	+	+	+	+	+	+/-	-	-	+
CAGL	+	+	+/-	-	-	-	-	-	-	-	+
KLTH	+	+	+	+	+/-	-	-	-	-	-	+
KLLA	+	+	+/-	-	-	-	-	-	-	-	+
ZYRO	+	+	+	+/-	-	-	-	-	-	-	+
DEHA	+	+	+	+	+	+	-	-	-	-	+
SAPA	+	+	-	-	-	-	-	-	-	-	+

A.

Concentration 5FU	5×10 ⁻⁷	10 ⁻⁶	5×10 ⁻⁶	10 ⁻⁵	5×10 ⁻⁵	10 ⁻⁴	5×10 ⁻⁴	10 ⁻³	5×10 ⁻³	Ø
SACE FL100	+	+	+	+	-	-	-	-	-	+
SACE Résistante	+	+	+	+	+	+	+	+/-	-	+
CAGL	+	+	+	+	-	-	-	-	-	+
KLTH	+	+	+	+	+	+/-	-	-	-	+
KLLA	+	+	+	+	+	+	+	+	+/-	+
ZYRO	+	+	+	-	-	-	-	-	-	+
DEHA	+	+/-	-	-	-	-	-	-	-	+
SAPA	+	+	+	+/-	-	-	-	-	-	+

B.

Tableau 4.4 : Résistance des souches de différentes espèces du groupe des hémiاسcomycètes à une gamme de 5-fluorouracile (5FU) (A.) et de 5-fluorouridine (5FUI) (B.).

Les espèces sont nommées SACE pour *S. cerevisiae*, CAGL pour *C. glabrata*, KLTH pour *K. thermotolerans*, KLLA pour *K. lactis*, ZYRO pour *Z. rouxii*, DEHA pour *D. hansenii* et SAPA pour *S. paradoxus*.

L'étude de l'import de l'uridine a également été réalisée lors de ce travail. La perméase à l'uridine codée par le gène *FUI1* chez *S. cerevisiae* permet également l'entrée dans la cellule du 5FUI (5-fluorouridine). Ce composé est transformé dans la cellule en 5FUMP, analogue toxique pour la cellule. Les mêmes espèces que précédemment ont été testées quant à leur résistance au 5FUI pour des concentrations croissantes allant de 5×10^{-7} M à 5×10^{-3} M. Les résultats obtenus pour les différentes espèces sont présentés dans le tableau 4.4. La souche de *S. cerevisiae* possède une sensibilité au 5FUI à partir d'une concentration de 10^{-5} M tandis que la souche résistante mutée pour le gène codant pour la perméase présente une sensibilité à partir de 5×10^{-3} M de 5FUI dans le milieu. *K. lactis* est résistante au 5FUI jusqu'à la concentration de 5×10^{-3} M. Les autres espèces, en revanche, sont sensibles au 5FUI à partir d'une concentration variant de 5×10^{-6} M pour la plus sensible (*D. hansenii*) à 5×10^{-4} M pour la plus résistante (*K. thermotolerans*).

Ces résultats montrent que *K. lactis* ne peut pas importer le 5FUI dans la cellule ce qui suggère que cette espèce ne possède pas de perméase à l'uridine. Pour toutes les autres espèces, l'entrée du 5-fluorouridine peut avoir lieu. Elles possèdent donc une perméase capable de faire rentrer de l'uridine de manière active dans la cellule.

L'analyse des perméases à l'allantoïne n'a pas pu être effectuée. Pour tester l'activité de la perméase, un milieu contenant l'allantoïne comme unique source d'azote est utilisé. Les souches présentant une croissance sur ce milieu sont celles possédant la perméase. Dans notre étude, plusieurs souches présentaient des auxotrophies, ce qui nécessitait d'ajouter les composés correspondants dans le milieu pour assurer leur croissance. Les acides aminés ainsi que les bases pouvant être utilisés comme source d'azote, l'allantoïne n'est donc plus l'unique source d'azote du milieu. Nous n'avons donc pas pu définir un milieu utilisable pour tester la présence de cette perméase.

2.3 Conclusions pour la famille *FUR4/FUI1/DAL4* chez les hémiascomycètes

Dans ce travail, nous nous sommes intéressés à la famille mutigénique composée des gènes *FUR4*, *FUI1* et *DAL4* mis en évidence chez *S. cerevisiae*. Ces trois gènes codent pour des perméases spécifiques à l'uracile, l'uridine et l'allantoïne respectivement.

Dans un premier temps, notre étude a consisté en l'analyse *in silico* de cette famille chez différentes espèces séquencées et annotées appartenant au groupe des hémiascomycètes. Cette analyse a consisté en la recherche des gènes homologues dans ces différentes espèces par une analyse BLASTP. La recherche de synténie pour les différents homologues a ensuite été effectuée. Dans un deuxième temps, nous avons testé l'activité *in vivo* de ces perméases

dans les différentes espèces par l'étude de la résistance à des composés toxiques à l'uracile ou à l'uridine.

Chez les *Saccharomyces sensu stricto*, *S. paradoxus* présente un homologue pour chacun des trois gènes de la famille et la synténie est très bien conservée entre cette espèce et *S. cerevisiae*. L'analyse de l'homologue à la protéine codée par *FUII* a montré que la protéine était tronquée chez *S. paradoxus*, laissant suggérer que cette protéine était peut-être non-fonctionnelle. Néanmoins, l'analyse *in vivo* a révélé que *S. paradoxus* est capable d'importer l'uracile et l'uridine dans la cellule ce qui suggère que la protéine, bien que tronquée, pourrait être fonctionnelle. Une deuxième hypothèse serait que l'une des protéines trouvées chez *S. paradoxus* présente deux activités et permette ainsi l'import des deux substrats. Enfin, une troisième hypothèse serait qu'une autre protéine ne possédant pas une similarité significative avec *Fui1p* est capable d'importer l'uridine dans la cellule. Des souches de *S. paradoxus* délétées pour les différents gènes trouvés lors de cette étude devrait être construites afin d'analyser si elles sont toujours capables de transporter l'uracile et l'uridine à l'intérieur de la cellule. Cette expérience permettra de discriminer les différentes hypothèses.

Pour *C. glabrata*, appartenant au groupe des *Saccharomyces sensu lato*, deux gènes homologues à *FUII* et à *FUR4* respectivement, ont été trouvés. La synténie est conservée pour ces deux gènes. De plus l'analyse *in vivo* de l'activité des perméases confirme ces résultats puisque la souche est capable d'importer aussi bien l'uridine que l'uracile. L'analyse de l'arbre phylogénétique réalisé à partir des séquences nucléotidiques des gènes dans les différentes espèces, montre aussi que les deux candidats sont plus proches de *FUR4* et *FUII* (Figure 4.8). Ces résultats laisseraient supposer que *C. glabrata* ne possède pas de perméase à l'allantoïne. Des expériences *in vivo* de l'import de ce substrat devraient être réalisées pour *C. glabrata* en utilisant l'allantoïne comme seule source d'azote dans le milieu.

A partir du clade des *Kluyveromyces*, la synténie est beaucoup moins bien conservée et l'analyse s'avère beaucoup plus difficile. Il nous a été possible de définir *in vivo* que *K. thermotolerans* est capable de transporter l'uracile et l'uridine. En revanche, la synténie n'est conservée pour aucun de ces deux gènes avec *S. cerevisiae*. *Z. rouxii* possède également la capacité d'importer l'uracile et l'uridine. Dans ce cas, il est possible de faire le lien avec la synténie qui est conservée pour les deux homologues avec celle de *S. cerevisiae*. Enfin pour *K. lactis*, deux gènes sont synténiques avec *FUII* et *FUR4* respectivement. La souche est incapable *in vivo* de faire entrer l'uridine dans la cellule, suggérant qu'elle ne possède pas la perméase pour ce composé. Une analyse plus poussée des perméases et

notamment le test de croissance de la souche en présence d'allantoïne serait nécessaire pour déterminer la fonction de cet homologue.

Enfin pour *D. hansenii*, un seul gène a pu être mis en évidence. Cette espèce étant très éloignée, la synténie entre cette espèce et *S. cerevisiae* n'est pas du tout conservée suite à l'accumulation de réarrangement. *In vivo*, il a été montré que la souche est capable d'importer l'uracile et l'uridine à l'intérieur de la cellule. Ce résultat nous permet de supposer que soit la seule protéine trouvée est capable d'importer ces deux substrats et possède une spécificité plus large que dans le cas de *S. cerevisiae*, soit une autre perméase codée par un gène non-homologue à la famille *FUR4/FUI1/DAL4* est utilisé. Pour préciser ces résultats, un mutant de *D. hansenii* pour ce gène devrait être testé quant à sa capacité à faire entrer l'uracile, l'uridine et l'allantoïne.

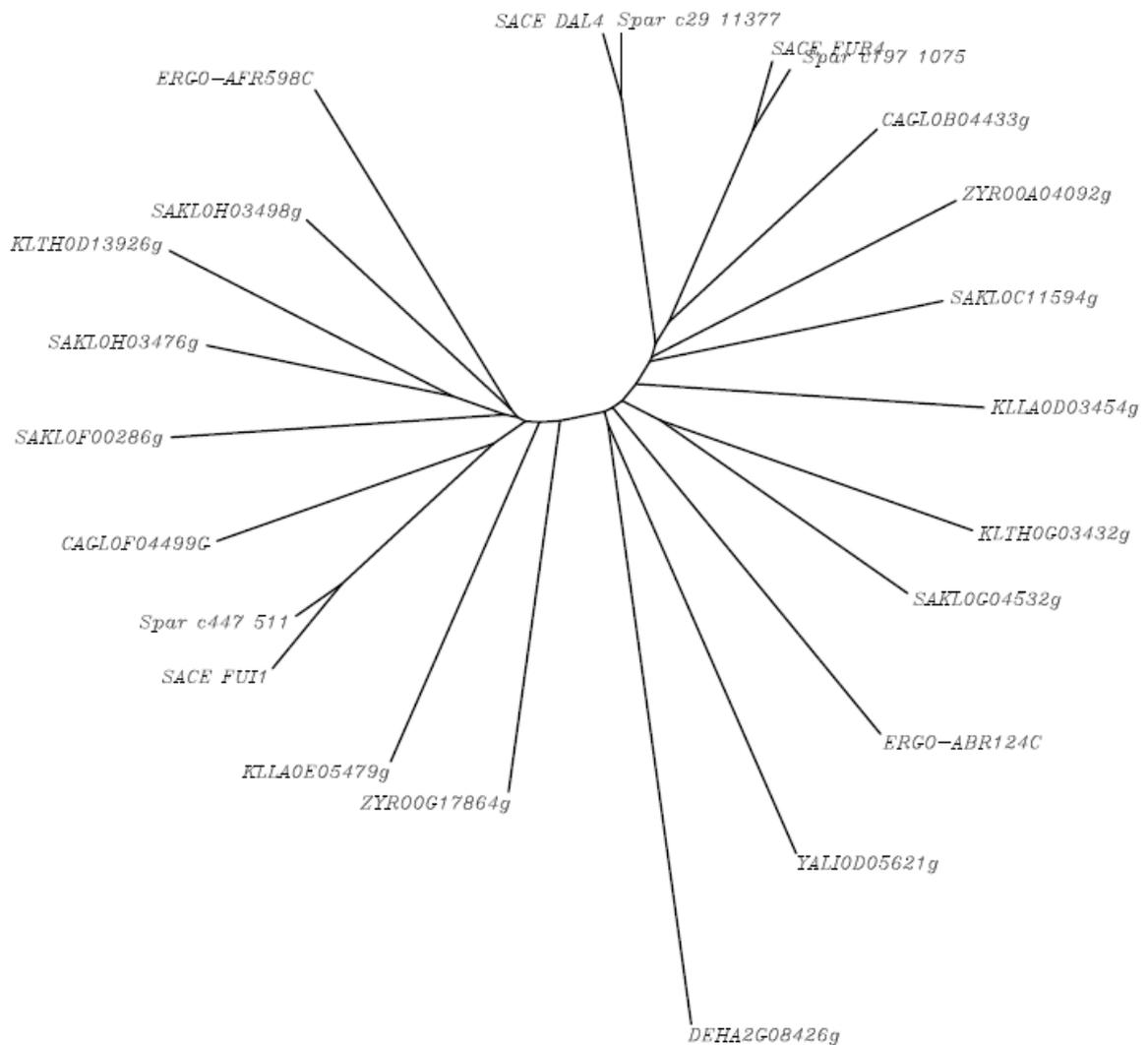


Figure 4.8 : Arbre phylogénétique des gènes trouvés dans les différentes espèces hémiascomycètes étudiées pour la famille des perméases à l'uracile, l'uridine et l'allantoïne. L'arbre a été réalisé par la méthode de Neighbor Joining à partir des alignements de séquences nucléotidiques avec le programme TreeView.

3. La famille *FCY2/FCY21/FCY22*

3.1 Les gènes *FCY2*, *FCY21* et *FCY22* chez *S. cerevisiae*

La protéine codée par le gène *FCY2* (*YER056c*) est une perméase aux purines et à la cytosine qui permet le transport actif de l'adénine, la guanine, l'hypoxanthine et de la cytosine à l'intérieur de la cellule (Figure 4.4). Le gène *FCY2* a été caractérisé au début des années 70 suite à l'obtention de mutants résistants à la 5-Fluorocytosine (5FC) (Jund et Lacroute, 1970). Cet analogue de la cytosine, une fois dans la cellule, est converti par la suite en 5FUMP analogue toxique de l'UMP. Il a ensuite été montré que la perméase codée par *FCY2* permettait l'import spécifique de l'adénine et l'hypoxanthine (Polak et Grenson, 1973) et de la guanine (Chevallier *et al.*, 1975).

Deux gènes présentant une identité significative à *FCY2* ont été mis en évidence (Wagner *et al.*, 2001), il s'agit des gènes *FCY21* (*YER060w*) et *FCY22* (*YER060w-A*). Les protéines Fcy21p et Fcy22p ne peuvent pas compléter une délétion de *FCY2*. La fonction de ces deux gènes reste encore inconnue. Chez *Saccharomyces cerevisiae*, les trois gènes sont situés sur le chromosome V. Les gènes *FCY21* et *FCY22* sont organisés en tandem direct tandis que *FCY2* est situé en amont de ces deux gènes et séparés de *FCY21* par quatre gènes (Figure 4.9). Ils font partie du bloc de duplication 27 défini par Wolfe et Shields. Néanmoins, les ohnologues de ces gènes sont absents chez *S. cerevisiae*. L'apparition de ces gènes est donc indépendante de la duplication totale du génome.



Figure 4.9 : Organisation des gènes *FCY2*, *FCY21* et *FCY22*.

Ces gènes présentent une très forte identité de séquences. Les pourcentages d'identité entre les gènes *FCY2*, *FCY21* et *FCY22* ainsi que les pourcentages de similarité de leurs produits sont présentés dans le tableau 4.5.

Protéines	SACE		
	Fcy2p	Fcy21p	Fcy22p
Fcy2p	100%	83,2% (72,3%)	87,2% (80,2%)
Fcy21p	83,2% (72,3%)	100%	84,5% (72,6%)
Fcy22p	87,2% (80,2%)	84,5% (72,6%)	100%

Acides nucléiques	SACE		
	FCY2	FCY21	FCY22
FCY2	100%	71,10%	81,70%
FCY21	71,10%	100%	73,40%
FCY22	81,70%	73,40%	100%

Tableau 4.5 : Pourcentages d'identité obtenus pour les gènes *FCY2*, *FCY21*, *FCY22* et de similarité pour leurs produits Fcy2p, Fcy21p et Fcy22p. Dans le cas des protéines, le nombre indiqué entre parenthèses indique le pourcentage d'identité. Les pourcentages ont été déterminés par alignement global d'après l'algorithme de Needleman et Wunsch.

3.2 La famille *FCY2/FCY21/FCY22* chez les hémiascomycètes

3.2.1 Recherche d'homologues chez les hémiascomycètes

Comme précédemment, les gènes présentant une forte identité de séquence avec les trois gènes trouvés chez *S. cerevisiae* ont été recherchés en utilisant le programme BLASTP à partir des séquences protéiques de leurs produits dans les banques de données Génolevures et SGD. Les résultats obtenus sont présentés dans le tableau 4.6.

Espèces	Nombres de gènes trouvés	
<i>Saccharomyces sensu stricto</i>	<i>S. cerevisiae</i>	3
	<i>S. paradoxus</i>	3
<i>Saccharomyces sensu lato</i>	<i>C. glabrata</i>	3
Clade <i>Kluyveromyces</i>	<i>K. lactis</i>	1
	<i>K. thermotolerans</i>	2
	<i>S. kluyveri</i>	1
	<i>Z. rouxii</i>	2
Espèces plus distantes	<i>A. gossypii</i>	1
	<i>D. hansenii</i>	1
	<i>Y. lipolytica</i>	1

Tableau 4.6 : Nombre de gènes présentant une forte identité de séquence avec les gènes *FCY2*, *FCY21* et *FCY22* de *S. cerevisiae* trouvés dans les autres levures hémiascomycètes séquencées.

Parmi les *Saccharomyces sensu stricto*, trois gènes présentant une identité de séquence importante avec les gènes trouvés chez *S. cerevisiae* ont été mis en évidence chez *S. paradoxus*.

Pour les *Saccharomyces sensu lato*, trois gènes ont été mis en évidence chez *C. glabrata*. Dans le clade des *Kluyveromyces*, *K. thermotolerans* et *Z. rouxii* possèdent deux candidats tandis que *K. lactis* et *S. kluyveri* en ont un seul. Les espèces plus distantes présentent toutes un seul représentant de la famille FCY2/FCY21/FCY22.

Les pourcentages de similarité entre les protéines homologues des différentes espèces sont présentés dans le tableau 4.7. Pour *S. paradoxus*, parmi les trois protéines homologues trouvées, deux présentent un fort pourcentage de similarité avec Fcy21p et Fcy22p respectivement. Le troisième candidat chez *S. paradoxus* (*Spar_c421_6259*) présente un pourcentage de similarité plus élevé pour Fcy2p, néanmoins celui-ci reste faible par rapport aux deux autres puisqu'il n'est que de l'ordre de 50%. La protéine prédite à partir de la séquence de *Spar_c421_6259* a une taille de 277 aa alors que Fcy2p est constituée de 533 aa. En réalisant un alignement local d'après l'algorithme de Smith et Waterman, on constate que cette protéine chez *S. paradoxus* est parfaitement alignée avec le début de la protéine de *S. cerevisiae*. Le pourcentage de similarité sur cette région est alors de 99,3%. Par recherche d'homologue en utilisant le programme Fungal Alignment Viewer sur le site SGD, l'ORF *Spar_c422_6257* a été mise en évidence (Figure 4.10). Elle est présente sur le contig 422 juste en amont du contig 421 sur lequel se trouve *Spar_c421_6259* et code pour une protéine de 193 aa présentant une faible similarité avec Fcy2p. Par alignement local, nous avons montré que la protéine putative s'aligne parfaitement avec la région C-terminale de Fcy2p et ces deux régions présentent un pourcentage de similarité de 95,8% (Figure 4.11). On peut donc supposer que *S. paradoxus* possédait un homologue à Fcy2p mais que suite à l'accumulation de mutation, il y a eu pseudogénéisation et perte de l'homologue. Pour confirmer cette hypothèse, il serait nécessaire de séquencer cette région chez *S. paradoxus* pour déterminer s'il s'agit bien d'un pseudogène et non d'une erreur de séquençage éventuelle.

	SACE			SAKL	YALI	ASGO
	FCY2	FCY21	FCY22	SAKL0F08624g	YALI0F30569g	ADL074W
Fcy2p	100%	83,2% (72,3%)	87,2% (80,2%)	79,0% (65,9%)	50,4% (38,7%)	69,4% (55,8%)
Fcy21p	83,2% (72,3%)	100%	84,5% (72,6%)	74,4% (61,3%)	50,6% (38,4%)	68,6% (54,3%)
Fcy22p	87,2% (80,2%)	84,5% (72,6%)	100%	73,1% (59,8%)	50,4% 38,7%	68,9% (54,4%)
	ZYRO		KLTH		DEHA	KLLA
	ZYRO0C08140g	ZYRO0D16874g	KLTH0A04290g	KLTH0D00264g	DEHA2E24200g	KLLA0C08283g
Fcy2p	76,1% (59,9%)	76,2% (65,0%)	76,4% (63,9%)	72,5% (57,1%)	71,0% (53,1%)	76,2% (63,2%)
Fcy21p	77,9% (64,2%)	74,5% (61,7%)	74,5% (60,0%)	72,9% (57,8%)	70,0% (51,1%)	75,7% (62,1%)
Fcy22p	76,3% (61,2%)	72,9% (60,2%)	72,7% (57,9%)	71,9% (57,1%)	70,1% (51,5%)	72,2% (58,7%)
	SAPA			CAGL		
	Spar_c421_6259	Spar_c421_6271	Spar_c421_6277	CAGL0C00231g	CAGL0J02948g	CAGL0L00671g
Fcy2p	51,4% (50,7%)	84,9% (74,0%)	85,7% (78,0%)	69,7% (54,2%)	75,6% (63,5%)	81,3% (68,4%)
Fcy21p	39,7% (33,5%)	97,2% (96,2%)	82,9% (72,4%)	71,6% (55,1%)	81,4% (68,3%)	80,7% (66,1%)
Fcy22p	42,1% (36,4%)	84,0% (72,1%)	94,8% (92,1%)	71,2% (55,1%)	77,8% (64,6%)	77,6% (63,4%)

Tableau 4.7 : Pourcentages de similarité observés pour les protéines des gènes homologues à *FCY2*, *FCY21* et *FCY22*.

SAPA : *Saccharomyces paradoxus*, CAGL : *Candida glabrata*, SAKL : *Saccharomyces kluyveri*, KLLA : *Kluyveromyces lactis*, ZYRO : *Zygosaccharomyces rouxii*, KLTH : *kluyveromyces thermotolerans*, ASGO : *Ashbya gossypii*, DEHA : *Debaryomyces hanseni* et YALI : *Yarrowia lipolytica*.

Le nombre indiqué entre parenthèses correspond au pourcentage d'identité. Les pourcentages ont été déterminés d'après les alignements globaux réalisés avec l'algorithme de Needleman et Wunsch.

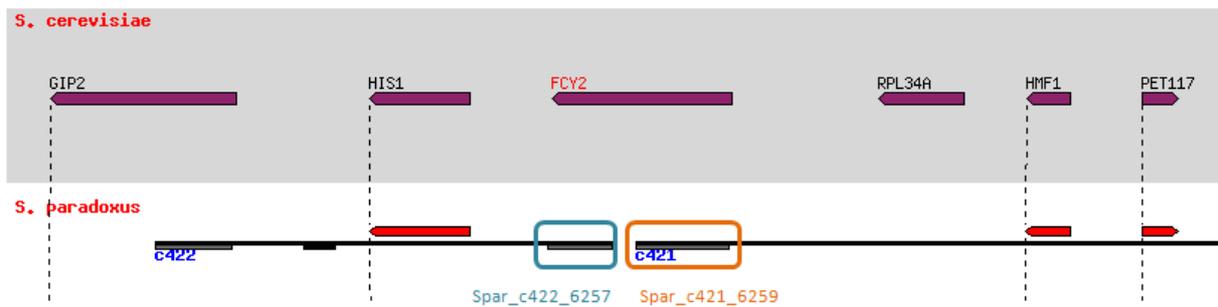


Figure 4.10 : Représentation des deux cadres de lecture Spar_c422_6257 et Spar_c421_6259 présentant une similarité de séquences avec Fcy2p (d'après Synteny viewer, SGD).

A.

Fcy2p	1	MLEEGNNVYEIQDLEKRSPVIGSSLENEKKVAASETFTATSEDDQQYIVE	50
Sparc421_6259	1	MLEEGNNVYEIQDLEKRSPVIGSSLENEKKVAASETFTATSEDDQQYIVE	50
Fcy2p	51	SSEATKLSWFHKFFASLNAETKGVPEPVTEDKTDSDSILNAASMWFSANMV	100
		:	
Sparc421_6259	51	SSEATKLSWFHKFFASLNAETKGVPEPVTEDKSDSDSILNAASMWFSANMV	100
Fcy2p	101	IASYALGALGPMVFGVFNFGQSVLVIIFFNIMGLIFVAFFSVFGAELGLRQ	150
Sparc421_6259	101	IASYALGALGPMVFGVFNFGQSVLVIIFFNIMGLIFVAFFSVFGAELGLRQ	150
Fcy2p	151	MILSRYLGVNVTARIFSLINVIACVGGIVNTSVSAQLLNMMVNEGSGHVC	200
Sparc421_6259	151	MILSRYLGVNVTARIFSLINVIACVGGIVNTSVSAQLLNMMVNEGSGHVC	200
Fcy2p	201	PIWAGCLIIIGGTVLVTFVFGYSVIHAYEKWSWVPMFAVFLVIIAQLSRSG	250
		: .	
Sparc421_6259	201	PLWAGCLIIIGGTVLVTFVFGYSVIHAYEKWSWVPMFAAFLVIIAQLSRSG	250
Fcy2p	251	KFKGGEWVGGATTAGSVLSFGSSIFGFAAGWTTYAADYTVYMPKSTNKYK	300
		. :	
Sparc421_6259	251	KFKGGEWVGGATTAGGVLSFGSAVFGS	277
Fcy2p	301	IFFSLVAGLAFPLFFFTMILGAASAMAALNDPTWKAYYDKNAMGGVIYAIL	350
Sparc421_6259	278		277

B.

Fcy2p	301	IFFSLVAGLAFPLFFFTMILGAASAMAALNDPTWKAYYDKNAMGGVIYAIL	350
Sparc422_6257	1	MGGVIYAIL	9
Fcy2p	351	VPNSLNGFGQFCCVLLALSTIANIPNMYTVALSAQALWAPLAKIPRVVW	400
Sparc422_6257	10	VPNSLNGFGQFCCVLLALSTIANIPNMYTVALSAQALWAPLAKIPRVVW	59
Fcy2p	401	TMAGNAATLGLISIPATYYFDGFMENFMDSIGYYLAIYIAISCSEHFFYRR	450
		.	
Sparc422_6257	60	TMAGNAATLGLISIPATYYFDSFMENFMDSIGYYLAIYIAISCSEHFFYRR	109
Fcy2p	451	SFSAYNIDDWDNWEHLPIGIAGTAALIVGAFGVGALGMCQTYWVGEIGRLI	500
		: . :	
Sparc422_6257	110	SFSAYNVDDWDNWGRPLPIGIAGTSAALIVGAFGVGALGMCQTYWVGEIGRLI	159
Fcy2p	501	GKYGGDIGFELGASWAFIIFNLRPLELKYFGR	533
		: : :	
Sparc422_6257	160	GEYGGDIGFELGASWAFIIFNLRPLEIKYFGR	192

Figure 4.11 : Alignements locaux réalisés d'après l'algorithme de Smith et Waterman.

A. Comparaison de Fcy2p avec Spar_C421_6259.

B. Comparaison de Fcy2p avec Spar_c422_6257.

3.2.2 Synténie des différents gènes

Comme dans le cas de la famille multigénique *FUR4/FUI1/DAL4*, nous avons déterminé la synténie pour les différents gènes puis nous les avons comparées pour les différentes espèces étudiées (Figure 4.12).

Chez *S. paradoxus*, la synténie est très bien conservée, les trois gènes étant présents sur le même chromosome et les régions environnantes étant très similaires à celles trouvées chez *S. cerevisiae*.

Pour *C. glabrata*, la synténie est moins bien conservée. Deux gènes CAGL0L00671g et CAGL0J02948g mis en évidence sur des chromosomes différents semblent synténiques de *FCY21*. Dans le cas du gène situé sur le chromosome XII (CAGL0L00671g), la région en aval du gène n'est pas conservée. Pour *C. glabrata*, le troisième gène, CAGL0C00231g, n'a pas de synténie conservée avec *S. cerevisiae* et est situé proche du télomère du chromosome III.

Dans le clade des *Kluyveromyces*, *Z. rouxii* possède deux gènes homologues qui présentent une conservation de synténie avec *FCY21*. Pour ZYRO0C08140g, trois gènes en amont du gène ainsi que toute la région en aval est conservée tandis que pour ZYRO0D16874g, seule la région en amont du gène est similaire à la synténie trouvée chez *S. cerevisiae*. *K. thermotolerans* présente également deux gènes dont un seul a une synténie conservée par rapport à *S. cerevisiae*. On peut supposer que le deuxième gène provient d'une duplication génique dans cette espèce. Enfin pour *K. lactis* et *S. kluyveri*, le seul gène homologue trouvé dans cette espèce semble, au vu de l'organisation des gènes, être l'orthologue de *FCY21*. Pour *S. kluyveri*, on observe une très bonne conservation de la synténie par rapport à *FCY21*. En revanche pour *K. lactis*, on observe une rupture de synténie avant le cinquième gène en amont de l'orthologue.

Pour les espèces plus distantes, *A. gossypii* présente une synténie conservée avec *FCY21* pour le gène *ADL074w*. En revanche, *D. hansenii* et *Y. lipolytica* sont trop distantes évolutivement pour que les blocs de synténie soient conservés.

3.2.3 Etude *in vivo* de la perméase purine/cytosine chez les hémiascomycètes

Le gène *FCY2* code pour une perméase à la cytosine et aux purines. Cette perméase permet l'entrée dans la cellule de la 5-fluorocytosine (5FCY). Cet analogue de la cytosine, une fois dans la cellule, est alors transformé en 5FUMP toxique à une concentration supérieure à 10^{-6} M chez *S. cerevisiae*. La même approche que celle réalisée pour la famille des perméases à l'uracile, l'uridine et l'allantoïne a été effectuée.

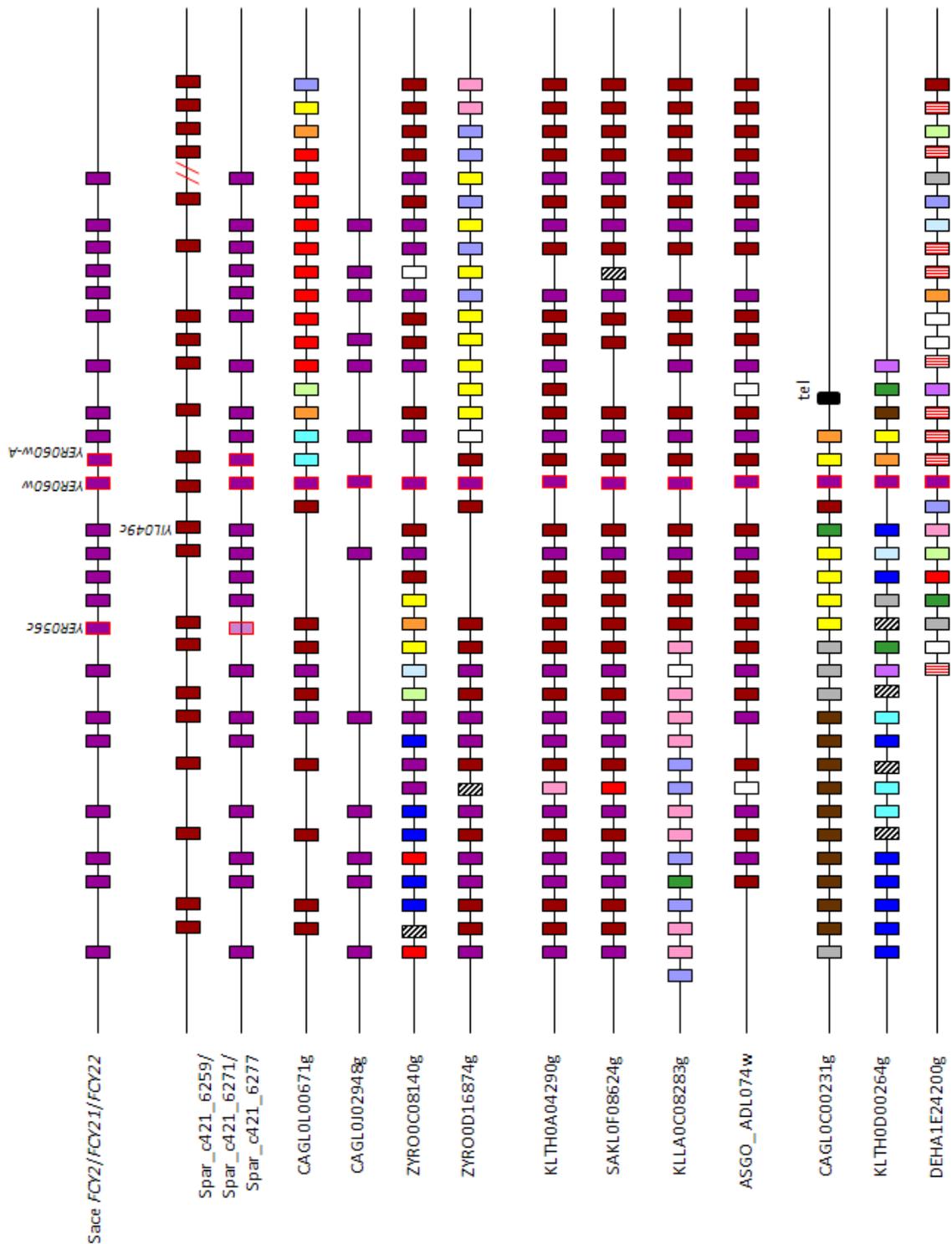


Figure 4.12 : Carte de synténie pour les gènes *FCY2*, *FCY21*, *FCY22* et leurs homologues dans les différentes espèces étudiées. Chaque rectangle représente un gène. Les couleurs des gènes ont été attribuées par correspondance avec les chromosomes de *S. cerevisiae*. (voir Figure 4.7 pour la légende des couleurs). Les flèches représentent les événements d'inversions.

Concentration 5FCY	5×10^{-7}	10^{-6}	5×10^{-6}	10^{-5}	5×10^{-5}	10^{-4}	5×10^{-4}	10^{-3}	5×10^{-3}	10^{-2}	Ø
SACE FL100	+	+/-	-	-	-	-	-	-	-	-	+
CAGL	+	+/-	-	-	-	-	-	-	-	-	+
KLTH	+	+	+	-	-	-	-	-	-	-	+
KLLA	+	+	+/-	-	-	-	-	-	-	-	+
ZYRO	+/-	-	-	-	-	-	-	-	-	-	+
DEHA	+	+	-	-	-	-	-	-	-	-	+
SAPA	+/-	-	-	-	-	-	-	-	-	-	+

Tableau 4.8 : Résistance des souches de différentes espèces du groupe des hémascomycètes à une gamme de 5-fluorocytosine.

Les espèces sont nommées SACE pour *S. cerevisiae*, CAGL pour *C. glabrata*, KLTH pour *K. thermotolerans*, KLLA pour *K. lactis*, ZYRO pour *Z. rouxii*, DEHA pour *D. hansenii* et SAPA pour *S. paradoxus*.

Nous avons déterminé les concentrations pour lesquelles les souches des différentes espèces deviennent sensibles à l'analogue de la cytosine. Pour cela, une gamme de concentration de 5FCY variant de 5×10^{-7} M à 10^{-2} M a été réalisée. Les résultats obtenus pour les souches des différentes espèces sont présentés dans le tableau 4.8. Toutes les souches présentent une forte sensibilité au 5FCY. En effet, une concentration supérieure à 5×10^{-6} M de 5FCY empêche leur croissance. Ce résultat suggère que toutes espèces possèdent une perméase permettant l'entrée de la cytosine dans la cellule.

L'import de l'adénine, de la guanine et de l'hypoxanthine dans ces espèces pourrait également être testé. Ces composés ne possédant pas d'analogues toxiques utilisables, il serait donc nécessaire de construire des mutants pour des gènes impliqués dans la voie de biosynthèse de ces composés, dans les différentes espèces. Si ces souches mutées peuvent se développer en milieu possédant ces substrats en quantités limitantes, nous pourrions conclure qu'elles ont la capacité d'importer l'adénine, la guanine et l'hypoxanthine à l'intérieur de la cellule.

3.3 Conclusion pour la famille FCY2/FCY21/FCY22 chez les hémiacomycètes

L'analyse *in silico* nous a permis de déterminer que seules les espèces proches appartenant aux groupes de *S. sensu stricto* et *lato*, présentent trois gènes possédant une identité de séquence significative à ceux trouvés chez *S. cerevisiae*. De plus, *S. paradoxus* est la seule à posséder les trois gènes sur le même chromosome avec une organisation identique que *S. cerevisiae*. Toutes les autres espèces analysées ne présentent qu'un seul homologue à l'exception de *Z. rouxii* et *K. thermotolerans* qui en ont deux. L'analyse de la synténie a montré que toutes les espèces à l'exception de *D. hansenii* ont au moins un gène homologue présentant une conservation de synténie avec le gène *FCY21*. De plus, l'analyse *in vivo* a mis

en évidence que toutes les espèces sont capables d'importer la cytosine dans la cellule et possède donc une perméase pour ce substrat.

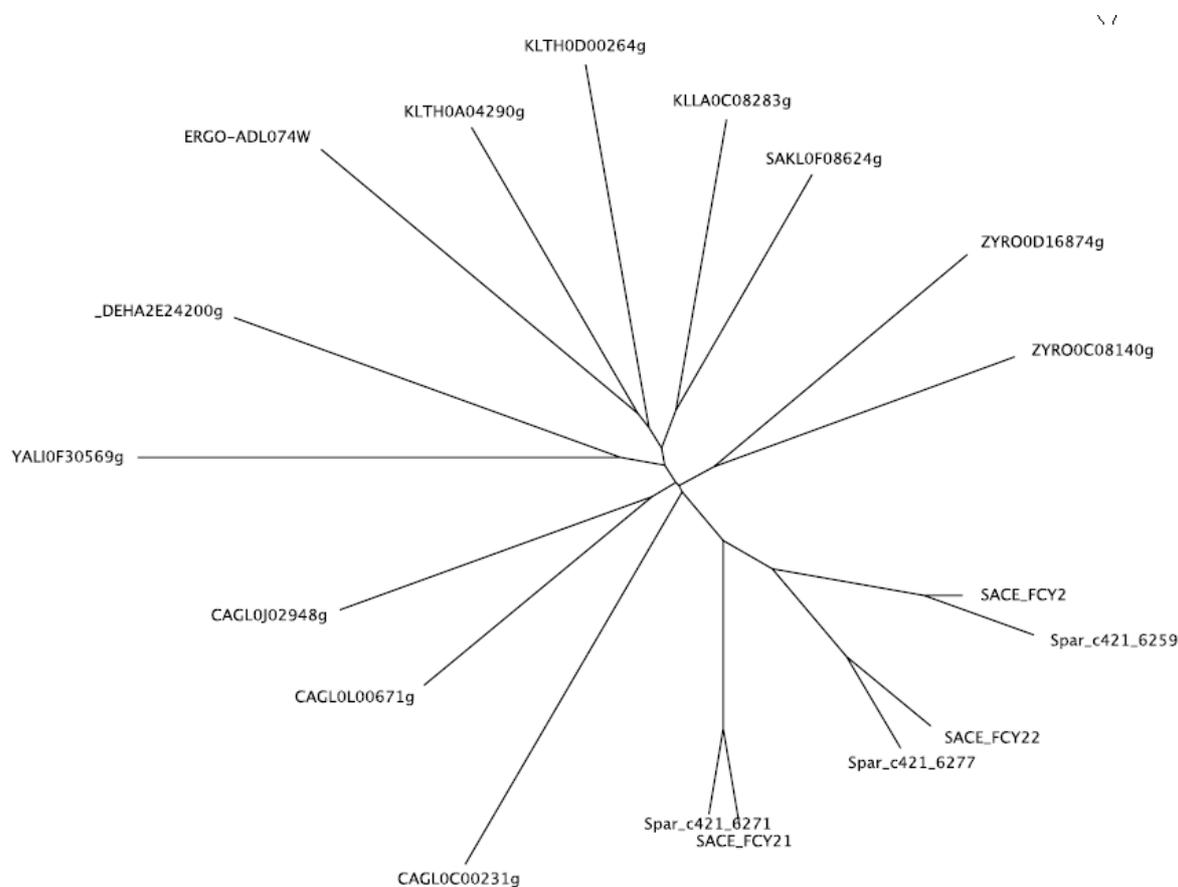


Figure 4.13 : Arbre phylogénétique des gènes trouvés dans les différentes espèces hémiascomycètes étudiées pour la famille des perméases aux purines et à la cytosine. L'arbre a été réalisé par la méthode de Neighbor Joining à partir des alignements de séquences nucléotidiques avec le programme TreeView.

Ces résultats permettent d'émettre l'hypothèse que chez l'ancêtre commun de *S. paradoxus* et *S. cerevisiae*, le gène *FCY2* serait issu d'une duplication à partir du gène ancestral *FCY21*. En effet, toutes les espèces à l'exception de *D. hansenii* présente une synténie conservée avec ce gène. *FCY2* aurait alors maintenu la fonction de perméase à la cytosine tandis que *FCY21* aurait perdu cette fonction. *C. glabrata* ne possède pas cette organisation, on peut donc supposer que cet événement de duplication se serait produit après la divergence de *C. glabrata* et des *Saccharomyces*. De plus l'analyse de l'arbre phylogénétique créé à partir des séquences nucléotidiques des gènes trouvés dans les différentes espèces montre que *FCY21* aurait divergé de *FCY2* (Figure 4.13). Puis *FCY22* aurait lui-même divergé de *FCY2*. Ce résultat permet de supposer qu'après l'événement de

duplication de *FCY21* ancestral, une duplication de *FCY2* se serait ensuite produite conduisant à l'apparition de *FCY22*.

Pour *C. glabrata*, le troisième gène, CAGL0C00231g, n'a pas de synténie conservée avec *S. cerevisiae*. Il est situé proche du télomère du chromosome III. L'analyse de l'arbre phylogénétique montre que le gène situé sur le chromosome III est plus proche de CAGL0L00671g que de CAGL0J02948g (Figure 4.13). On peut supposer que CAGL0C00231g serait issu d'une duplication génique de CAGL0L00671g suivie de son insertion à au niveau du chromosome III.

4. Discussion

Lors de cette étude, les données obtenues par l'ensemble du consortium Génolevures ont été utilisées pour aborder la question du devenir des familles multigéniques au sein d'espèces pré- et post-duplication. Pour cela, notre analyse a porté sur deux familles multigéniques chez *S. cerevisiae*. La première est la famille *FUR4/FUII/DAL4* dont la fonction est connue et différente pour les trois gènes paralogues chez *S. cerevisiae*. Il s'agit de perméases à l'uracile, l'uridine et l'allantoïne respectivement. La deuxième famille étudiée est *FCY2/FCY21/FCY22* dont la fonction n'a été déterminée que pour le gène *FCY2* qui code pour une perméase à la cytosine et aux purines. L'étude *in silico* de ces familles ainsi que l'analyse *in vivo* pour différentes espèces séquencées ont permis de montrer que l'organisation et/ou le nombre de ces gènes sont très variables d'une espèce à l'autre. Cependant, il est très difficile de retracer l'histoire évolutive de ces familles de gènes tant les génomes sont modifiés d'une espèce à l'autre au cours du temps. Il est donc nécessaire de travailler avec des espèces très proches pour pouvoir déterminer les réarrangements à l'origine des modifications du génome. Avec des espèces plus éloignées, trop de modifications se sont accumulées, rendant l'analyse très ardue.

Par ailleurs, bien que l'analyse *in silico* apporte de nombreuses informations et permette d'émettre des hypothèses, l'analyse *in vivo* reste incontournable. En effet, le cas de l'homologue à *FUII* chez *K. lactis* représente un bon exemple de complémentarité entre les informations obtenues *in silico* et *in vivo*. En effet, l'analyse *in silico* montre que ce gène présente un fort pourcentage d'identité avec le gène *FUII* codant pour la perméase spécifique à l'uridine chez *S. cerevisiae*. De plus, la synténie est conservée en partie avec ce gène. L'analyse *in silico*, à elle seule, laisserait supposer que la souche de *K. lactis* possède une perméase à l'uridine. Néanmoins, lors de l'étude *in vivo*, il a été montré que cette souche ne

permet pas l'entrée de l'uridine dans la cellule et ne possède donc vraisemblablement pas de perméase pour ce composé.

Pour confirmer l'ensemble des fonctions putatives décrites lors de cette étude, une analyse fonctionnelle des produits des différents gènes considérés doit être menée. Pour cela, des mutants de ces gènes dans les différentes espèces devraient être construits. Si ces gènes codent effectivement pour des perméases, leur inactivation devrait bloquer l'import des différents composés.

L'analyse *in silico* et *in vivo* doivent donc être menées en parallèle pour obtenir des informations les plus précises possibles.

Cette analyse a, également, confirmé l'importance des remaniements chromosomiques dans l'apparition de nouvelles fonctions et dans les mécanismes de spéciation.

Enseignement et Recherche

ENSEIGNEMENT ET RECHERCHE

J'ai tenu à ajouter cette partie plus personnelle sur mes activités d'enseignements car, bien que moins scientifiques, elles font partie intégrante de mon travail de thèse. Dans ce chapitre, je présenterai les différentes expériences que j'ai eues en tant que monitrice puis en tant qu'attachée temporaire d'enseignement et de recherche. Je développerai ensuite ce qu'a représenté pour moi l'enseignement au cours de ces quatre années.

J'ai eu l'opportunité d'obtenir après mon DEA un poste de monitrice à de l'Université Louis Pasteur. Mon expérience a débuté au sein de l'équipe pédagogique de biochimie en travaux pratiques (TP)/travaux dirigés (TD) de biochimie pour les licences 2^{ème} année.

Cette unité d'enseignement a pour but de présenter les principales propriétés structurales et physicochimiques des macromolécules biologiques au travers d'expériences réalisées lors de travaux pratiques et d'exercices en travaux dirigés. Les notions pratiques abordées en TP sont les suivantes :

- Les tests d'identification et de dosage des glucides, des protéines et des acides nucléiques.
- La purification du lysozyme par colonne échangeuse d'ions puis sa cristallisation et le dosage de son activité.
- L'étude de la fermentation alcoolique par dosage des glucides et de l'éthanol.
- La caractérisation des lipides par détermination des indices d'acide, de saponification et d'iode.

Lors des TP, mon travail d'encadrement a consisté à superviser les étudiants de deux groupes pendant six séances, à les évaluer sur la rédaction d'un rapport et lors d'un test pratique.

Mon enseignement en TD a été réalisé pour les séances « calculs » ayant pour objectif de revoir les bases des calculs de dilutions et l'utilisation des unités, et « lipides » dont le but est de donner aux étudiants une vue globale des propriétés physicochimiques et structurales des lipides.

J'ai ensuite poursuivi en tant qu'attachée temporaire d'enseignement et de recherche. Durant cette période, j'ai réalisé des TDs de génomique comparative pour des étudiants de master 1^{ère} et 2^{ème} année.

L'utilisation de l'informatique lors de ces TD m'a permis d'avoir une très bonne interactivité avec les étudiants et d'approfondir, d'un point de vue pratique, les notions théoriques abordées lors des cours magistraux.

Pour les étudiants en master 2^{ème} année, les TD ont permis d'aborder l'utilisation des bases de données, les alignements de séquences et l'annotation des génomes. Pour la partie « annotations des génomes », les étudiants devaient utiliser les différents programmes vus lors des séances précédentes (Genmark, ORF Finder, BLASTN, BLASTP...) afin d'identifier les ORF au niveau d'une région d'ADN et les annoter le plus précisément possible. Ceci réalisé, une discussion a été menée avec les étudiants sur l'intérêt et les limites des informations obtenues après annotation.

Les travaux dirigés pour les étudiants en master 1^{ère} année se sont déroulés en deux temps. La première partie consistait à préparer les travaux pratiques en génétique dont le but était d'élaborer une expérience de remplacement génique en une étape par un fragment PCR. Chaque étudiant devait déterminer pour le gène qui lui était assigné (auxotrophie, catabolisme, essentiel ou sans fonction connue) les séquences des oligonucléotides à utiliser pour synthétiser le fragment de remplacement et pour vérifier son insertion au locus d'intérêt, les souches et les milieux à utiliser pour sélectionner les transformants.

La deuxième partie des travaux dirigés portait sur des analyses de génomique comparative par l'étude des gènes *URA5* et *URA10* chez *Saccharomyces cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, *Candida glabrata* et enfin *Kluyveromyces waltii*. Cette partie permettait d'appréhender les notions d'homologie, de paralogie et d'orthologie et de réfléchir aux mécanismes permettant l'apparition de paralogues dans une espèce ainsi qu'au devenir des gènes dupliqués au cours de l'évolution.

Ces quatre années passées à faire de la recherche et de l'enseignement ont été particulièrement enrichissantes tant au niveau personnel que professionnel. Il est vrai que le temps passé à enseigner est du temps que l'on ne peut pas accorder au travail de recherche à proprement parler. Je souhaite, tout d'abord, remercier mes directeurs de thèse ainsi que les autres enseignants-chercheurs du laboratoire qui m'ont toujours soutenue dans mon envie d'enseigner et de m'investir dans cette activité tout en conciliant le travail de thèse. Ce temps passé en dehors du laboratoire est l'occasion de diversifier ses activités, d'avoir des relations avec d'autres moniteurs ou enseignants et également avec les étudiants, permettant ainsi d'avoir des contacts humains extérieurs et différents de ceux du laboratoire.

Les expériences d'enseignement en biochimie puis en génomique comparative m'ont permis de mettre à profit les connaissances acquises durant mon cursus universitaire et au cours de ma thèse. Cela a été particulièrement le cas pour l'enseignement de génomique comparative puisque j'ai pu aborder avec les étudiants des notions que j'utilise au laboratoire. Cela a aussi été l'occasion, à l'inverse, d'approfondir des connaissances qui par la suite m'ont été utiles lors de mon travail de thèse. Lors de la préparation des cours de génomique pour les masters 2^{ème} année, j'ai été amenée à utiliser des programmes notamment pour les alignements de séquences qui m'ont également servi lors de l'analyse des familles multigéniques.

Par ailleurs, devoir faire passer un message aux étudiants nécessite un travail d'explication et de vulgarisation qui est également nécessaire lors de communications scientifiques.

D'un point de vue plus personnel, j'ai pris beaucoup de plaisir à enseigner. Le fait que certains étudiants sortent de cours en ayant appris quelque chose est extrêmement gratifiant. De plus, je trouve que l'opportunité d'enseigner a eu des répercussions très positives sur la partie recherche de mon travail de thèse. C'est une autre façon de passer « de l'autre côté de la barrière », apprendre à devenir autonome, à prendre des responsabilités et à proposer des idées.

L'enseignement et la recherche me paraissent complémentaires car ils permettent une dynamique entre l'acquisition des connaissances et sa transmission. Cette expérience m'a confortée dans l'objectif d'orienter ma vie professionnelle vers une carrière universitaire m'offrant l'opportunité d'allier ces deux facettes.

Références bibliographiques

BIBLIOGRAPHIE

Adams J., Puskas-Rozsa S., Simlar J. and Wilke C.M., 1992 Adaptation and major chromosomal changes in populations of *Saccharomyces cerevisiae*. *Curr Genet* 22(1):13-9.

Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L., Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H., Shue B.C., Sidén-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M. and Venter J.C., 2000 The genome sequence of *Drosophila melanogaster*. *Science* 287(5461):2185-95.

- Almeida R. and Allshire R.C., 2005 RNA silencing and genome regulation. *Trends Cell Biol* 15(5):251-8.
- Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J., 1990 Basic local alignment search tool. *J Mol Biol* 215(3):403-10.
- Amores A., Force A., Yan Y.L., Joly L., Amemiya C., Fritz A., Ho R.K., Langeland J., Prince V., Wang Y.L., Westerfield M., Ekker M. and Postlethwait J.H., 1998 Zebrafish hox clusters and vertebrate genome evolution. *Science* 282(5394):1711-4.
- Arabidopsis* Genome Initiative, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408(6814):796-815.
- Aström S.U., Okamura S.M. and Rine J., 1999 Yeast cell-type regulation of DNA repair. *Nature* 397(6717):310.
- Aylon Y. and Kupiec M., 2004 DSB repair: the yeast paradigm. *DNA Repair (Amst)* 3(8-9):797-815.
- Bai Y. and Symington L.S., 1996 A Rad52 homolog is required for RAD51-independent mitotic recombination in *Saccharomyces cerevisiae*. *Genes Dev.* 10(16):2025-37.
- Bertuch A.A. and Lundblad V., 2003 The Ku heterodimer performs separable activities at double-strand breaks and chromosome termini. *Mol Cell Biol* 23(22):8202-15.
- Blanc V. M., and J. Adams, 2003 Evolution in *Saccharomyces cerevisiae*: identification of mutations increasing fitness in laboratory populations. *Genetics.* 165(3):975-83
- Boddy M.N., Gaillard P.H., McDonald W.H., Shanahan P., Yates J.R. 3rd and Russell P., 2001 Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* 107(4):537-48.
- Bosco G., Campbell P., Leiva-Neto J.T. and Markow T.A., 2007 Analysis of *Drosophila* species genome size and satellite DNA content reveals significant differences among strains as well as between species. *Genetics* 177(3):1277-90.

- Boulton S.J. and Jackson S.P., 1996 *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J* 15(18):5093-103.
- Brachmann C.B., Davies A., Cost G.J., Caputo E., Li J., Hieter P. and Boeke J.D., 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(2):115-32.
- Branzei D. and Foiani M., 2005 The DNA damage response during DNA replication. *Curr Opin Cell Biol* 17(6):568-75.
- C. elegans* Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282(5396):2012-8.
- Chen C. and Kolodner R.D., 1999 Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat Genet* 23(1):81-5.
- Chevallier M.R., 1982 Cloning and transcriptional control of a eucaryotic permease gene. *Mol Cell Biol* 2(8):977-84.
- Chevallier M.R., Jund R. and Lacroute F., 1975 Characterization of cytosine permeation in *Saccharomyces cerevisiae*. *J Bacteriol* 122(2):629-41.
- Chu G., 1989 Pulsed field electrophoresis in contour-clamped homogeneous electric fields for the resolution of DNA by size or topology. *Electrophoresis* 10(5-6):290-5.
- Cliften P., Sudarsanam P., Desikan A., Fulton L., Fulton B., Majors J., Waterston R., Cohen B.A. and Johnston M., 2003 Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* 301(5629):71-6.
- Clikeman J.A., Khalsa G.J., Barton S.L. and Nickoloff J.A., 2001 Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. *Genetics* 157(2):579-89.
- Colson I., Delneri D. and Oliver S.G., 2004 Effects of reciprocal chromosomal translocations on the fitness of *Saccharomyces cerevisiae*. *EMBO Rep* 5(4):392-8.

- Cooper T.G., Chisholm V.T., Cho H.J. and Yoo H.S., 1987 Allantoin transport in *Saccharomyces cerevisiae* is regulated by two induction systems. *J Bacteriol* 169(10):4660-7.
- Cox M.M., Goodman M.F., Kreuzer K.N., Sherratt D.J., Sandler S.J. and Marians K.J., 2000 The importance of repairing stalled replication forks. *Nature* 404(6773):37-41.
- Curcio M.J. and Garfinfel D.J., 1991 Single-step selection for Ty1 element retrotransposition. *Proc Natl Acad Sci U S A.* 88(3):936-40.
- Daley J.M. and Wilson T.E., 2005 Rejoining of DNA double-strand breaks as a function of overhang length. *Mol Cell Biol* 25(3):896-906.
- Daley J.M., Palmboos P.L., Wu D. and Wilson T.E., 2005 Nonhomologous end joining in yeast. *Annu Rev Genet.* 39:431-51.
- Davis A.P. and Symington L.S., 2004 RAD51-dependent break-induced replication in yeast. *Mol Cell Biol* 24(6):2344-51.
- Dawes I.W. and Hardie I.D., 1974 Selective killing of vegetative cells in sporulated yeast cultures by exposure to diethyl ether.
- de Montigny J., Belarbi A., Hubert J.C. and Lacroute F., 1989 Structure and expression of the URA5 gene of *Saccharomyces cerevisiae*. *Mol Gen Genet* 215(3):455-62.
- de Montigny J., Kern L., Hubert J.C. and Lacroute F., 1990 Cloning and sequencing of *URA10*, a second gene encoding orotate phosphoribosyl transferase in *Saccharomyces cerevisiae*. *Curr Genet* 17(2):105-11.
- Decottignies A., 2007 Microhomology-mediated end joining in fission yeast is repressed by pku70 and relies on genes involved in homologous recombination. *Genetics* 176(3):1403-15.
- Deutschbauer A.M. and Davis RW., 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet.* 37(12):1333-40.
- Di Caprio L. and Cox B.S., 1981 DNA synthesis in UV-irradiated yeast. *Mutat Res* 82(1):69-85.

- Dietrich F.S., Voegeli S., Brachat S., Lerch A., Gates K., Steiner S., Mohr C., Pöhlmann R., Luedi P., Choi S., Wing R.A., Flavier A., Gaffney T.D. and Philippsen P., 2004 The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* 304(5668):304-7.
- Difilippantonio M.J., Petersen S., Chen H.T., Johnson R., Jasin M., Kanaar R., Ried T. and Nussenzweig A., 2002 Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J Exp Med* 196(4):469-80.
- Doherty A.J., Jackson S.P. and Weller G.R., 2001 Identification of bacterial homologues of the Ku DNA repair proteins. *FEBS Lett* 500(3):186-8.
- Dudášová Z., Dudás A. and Chovanec M., 2004 Non-homologous end-joining factors of *Saccharomyces cerevisiae*. *FEMS Microbiol Rev* 28(5):581-601.
- Dujon B., 2006 Yeasts illustrate the molecular mechanisms of eukaryotic genome evolution. *Trends Genet* 22(7):375-87.
- 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.M., Beyne E., Bleykasten C., Boisramé A., Boyer J., Cattolico L., Confanioleri F., De Daruvar A., Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A., Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R., Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H., Nicaud J.M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O., Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A., Swennen D., Tekaiia F., Wésolowski-Louvel M., Westhof E., Wirth B., Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A., Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J., Wincker P. and Souciet J.L., 2004 Genome evolution in yeasts. *Nature* 430(6995):35-44.
- Dunham M.J., Badrane H., Ferea T., Adams J., Brown P.O., Rosenzweig F. and Botstein D., 2002 Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99(25):16144-9.
- Ehmsen K.T. and Heyer W.D., 2008 *Saccharomyces cerevisiae* Mus81-Mms4 is a catalytic, DNA structure-selective endonuclease. *Nucleic Acids Res* 36(7):2182-95.

- Exinger F. and Lacroute F., 1979 Genetic evidence for the creation of a reinitiation site by mutation inside the yeast *ura2* gene. *Mol Gen Genet* 173(1):109-13.
- Fewell S.W. and Woolford J.L. Jr, 1999 Ribosomal protein S14 of *Saccharomyces cerevisiae* regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Mol Cell Biol* 19(1):826-34.
- Fischer G., James S.A., Roberts I.N., Oliver S.G. and Louis E.J., 2000 Chromosomal evolution in *Saccharomyces*. *Nature* 405(6785):451-4.
- Fishman-Lobell J., Rudin N. and Haber J.E., 1992 Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. *Mol Cell Biol* 12(3):1292-303.
- Fogel S. and Hurst D.D., 1967 Meiotic gene conversion in yeast tetrads and the theory of recombination. *Genetics* 57(2):455-81.
- Frank-Vaillant M. and Marcand S., 2001 NHEJ regulation by mating type is exercised through a novel protein, Lif2p, essential to the ligase IV pathway. *Genes Dev* 15(22):3005-12.
- Frank-Vaillant M. and Marcand S., 2002 Transient stability of DNA ends allows nonhomologous end joining to precede homologous recombination. *Mol Cell* 10(5):1189-99.
- Gabriel A., Daprich J., Kunkel M., Gresham D., Pratt S. C. and Dunham M.J., 2006 Global mapping of transposon location. *PLoS Genet.* 2(12):e212.
- Game J.C. and Mortimer R.K., 1974 A genetic study of x-ray sensitive mutants in yeast. *Mutat Res* 24(3):281-92.
- Garfinkel D. J., K. Nyswaner, J. Wang, and J. Y. Cho, 2003 Post-transcriptional cosuppression of Ty1 retrotransposition. *Genetics* 165:83-99.
- Goffeau A., Barrell B.G., Bussey H., Davis R.W., Dujon B., Feldmann H., Galibert F., Hoheisel J.D., Jacq C., Johnston M., Louis E.J., Mewes H.W., Murakami Y., Philippsen P., Tettelin H. and Oliver S.G., 1996 Life with 6000 genes. *Science* 274(5287):546, 563-7.

- Green M.H., O'Neill J.P. and Cole J., 1995 Suggestions concerning the relationship between mutant frequency and mutation rate at the hprt locus in human peripheral T-lymphocytes. *Mutat Res* 334(3):323-39.
- Hakenbeck R., Balmelle N., Weber B., Gardès C., Keck W. and de Saizieu A., 2001 Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect Immun* 69(4):2477-86.
- Halfer C., 1981 Interstrain heterochromatin polymorphisms in *Drosophila melanogaster*. *Chromosoma* 84(2):195-206.
- Harrison J.C. and Haber J.E., 2006 Surviving the breakup: the DNA damage checkpoint. *Annu Rev Genet* 40:209-35.
- Harrison P.M., Hegyi H., Balasubramanian S., Luscombe N.M., Bertone P., Echols N., Johnson T. and Gerstein M., 2002 Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Res* 12(2):272-80.
- Hawthorne D.C. and Mortimer R.K., 1960 Chromosome Mapping in *Saccharomyces*: Centromere-Linked Genes. *Genetics* 45(8):1085-110.
- Hellsten U., Khokha M.K., Grammer T.C., Harland R.M., Richardson P. and Rokhsar D.S., 2007 Accelerated gene evolution and subfunctionalization in the pseudotetraploid frog *Xenopus laevis*. *BMC Biol* 5:31.
- Herrmann G., Lindahl T. and Schär P., 1998 *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *EMBO J* 17(14):4188-98.
- Hinnen A., Hicks J.B. and Fink G.R., 1978 Transformation of yeast. *Proc Natl Acad Sci USA* 75(4):1929-33.
- Hittinger C.T. and Carroll S.B., 2007 Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449(7163):677-81.
- Hoffman C.S and Winston F., 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57:267-272.
- Holliday R., 1964 A mechanism for gene conversion in fungi. *Genet Research* 5:282-304.

- Inbar O. and Kupiec M., 1999 Homology search and choice of homologous partner during mitotic recombination. *Mol Cell Biol* 19(6):4134-42.
- International Human Genome Sequencing Consortium, 2001 Initial sequencing and analysis of the human genome. *Nature*. 2001 Feb 15;409(6822):860-921.
- Jiang Y. W., 2002 Transcriptional cosuppression of yeast Ty1 retrotransposons. *Genes Dev*. 16(4):467-78.
- Johnson R.D. and Symington L.S., 1995 Functional differences and interactions among the putative RecA homologs Rad51, Rad55, and Rad57. *Mol Cell Biol* 15(9):4843-50.
- Jund R. and Lacroute F., 1970 Genetic and physiological aspects of resistance to 5-fluoropyrimidines in *Saccharomyces cerevisiae*. *J Bacteriol* 102(3):607-15
- Jund R., Weber E. and Chevallier M.R., 1988 Primary structure of the uracil transport protein of *Saccharomyces cerevisiae*. *Eur J Biochem* 171(1-2):417-24.
- Kadyk L.C. and Hartwell L.H., 1992 Sister chromatids are preferred over homologs as substrates for recombinational repair in *Saccharomyces cerevisiae*. *Genetics* 132(2):387-402.
- Kehrer-Sawatzki H. and Cooper D.N., 2007 Understanding the recent evolution of the human genome: insights from human-chimpanzee genome comparisons. *Hum Mutat* 28(2):99-130.
- Kellis M., Patterson N., Endrizzi M., Birren B. and Lander E.S., 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* 423(6937):241-54.
- Kim J.M., Vanguri S., Boeke J.D., Gabriel A. and Voytas DF., 1998 Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res* 8(5):464-78.
- Koken M.H., Reid A., Quignon F., Chelbi-Alix M.K., Davies J.M., Kabarowski J.H., Zhu J., Dong S., Chen S., Chen Z., Tan C.C., Licht J., Waxman S., de Thé H. and Zelent A., 1997 Leukemia-associated retinoic acid receptor alpha fusion partners, PML and PLZF, heterodimerize and colocalize to nuclear bodies. *Proc Natl Acad Sci USA* 94(19):10255-60.

- Korbel J.O., Urban A.E., Affourtit J.P., Godwin B., Grubert F., Simons J.F., Kim P.M., Palejev D., Carriero N.J., Du L., Taillon B.E., Chen Z., Tanzer A., Saunders A.C., Chi J., Yang F., Carter N.P., Hurles M.E., Weissman S.M., Harkins T.T., Gerstein M.B., Egholm M. and Snyder M., 2007 Paired-end mapping reveals extensive structural variation in the human genome. *Science* 318(5849):420-6.
- Kostriken R., Strathern J.N., Klar A.J., Hicks J.B. and Heffron F., 1983 A site-specific endonuclease essential for mating-type switching in *Saccharomyces cerevisiae*. *Cell* 35(1):167-74.
- Kozul R, Caburet S, Dujon B and Fischer G., 2004 Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J* 23(1):234-43.
- Kresse A.U., Dinesh S.D., Larbig K. and Römling U., 2003 Impact of large chromosomal inversions on the adaptation and evolution of *Pseudomonas aeruginosa* chronically colonizing cystic fibrosis lungs. *Mol Microbiol* 47(1):145-58.
- Krishna S., Wagener B.M., Liu H.P., Lo Y.C., Sterk R., Petrini J.H. and Nickoloff J.A., 2007 Mre11 and Ku regulation of double-strand break repair by gene conversion and break-induced replication. *DNA Repair (Amst)* 6(6):797-808.
- Kupiec M., 2000 Damage-induced recombination in the yeast *Saccharomyces cerevisiae*. *Mutat Res* 451(1-2):91-105.
- Kurtzman C.P. and Robnett C.J., 2003 Phylogenetic relationships among yeasts of the '*Saccharomyces* complex' determined from multigene sequence analyses. *FEMS Yeast Res* 3(4):417-32.
- Lacroute F., 1968 Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol.* 95(3):824-32.
- Lea D.E. and Coulson C.A., 1949 The distribution of the number of mutants in bacterial populations. *Journal of Genetics* 49:264-285.
- Lee K. and Lee S.E., 2007 *Saccharomyces cerevisiae* Sae2- and Tel1-dependent single-strand DNA formation at DNA break promotes microhomology-mediated end joining. *Genetics* 176(4):2003-14.

Lewis L.K. and Resnick M.A., 2000 Tying up loose ends: nonhomologous end-joining in *Saccharomyces cerevisiae*. *Mutat Res* 451(1-2):71-89.

Liang L., Deng L., Nguyen S.C., Zhao X., Maulion C.D., Shao C. and Tischfield J.A., 2008 Human DNA ligases I and III, but not ligase IV, are required for microhomology-mediated end joining of DNA double-strand breaks. *Nucleic Acids Res.*

Liti G., A. Peruffo, S. A. James, I. N. Roberts, and E. J. Louis, 2005 Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu strict* complex. *Yeast* 22:177-192.

Llorente B., Smith C.E. and Symington L.S., 2008 Break-induced replication: what is it and what is it for? *Cell Cycle* 7(7):859-64.

Losson R., Fuchs R.P. and Lacroute F., 1979 In vivo transcription of a eukaryotic regulatory gene. *EMBO J* 2(12):2179-84.

Lucena B.T., Silva-Filho E.A., Coimbra M.R., Morais J.O., Simões D.A. and Morais M.A. Jr, 2007 Chromosome instability in industrial strains of *Saccharomyces cerevisiae* batch cultivated under laboratory conditions. *Genet Mol Res* 6(4):1072-84.

Lundblad V. and Blackburn E.H., 1993 An alternative pathway for yeast telomere maintenance rescues est1- senescence. *Cell* 73(2):347-60.

Ma J.L., Kim E.M., Haber J.E. and Lee S.E., 2003 Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* 23(23):8820-8.

Ma J.L., Kim E.M., Haber J.E. and Lee S.E., 2003 Yeast Mre11 and Rad1 proteins define a Ku-independent mechanism to repair double-strand breaks lacking overlapping end sequences. *Mol Cell Biol* 23(23):8820-8.

Mahdi A.A., Sharples G.J., Mandal T.N. and Lloyd R.G., 1996 Holliday junction resolvases encoded by homologous *rusA* genes in *Escherichia coli* K-12 and phage 82. *J Mol Biol* 257(3):561-73.

- Malkova A., Signon L., Schaefer C.B., Naylor M.L., Theis J.F., Newlon C.S. and Haber J.E., 2001 RAD51-independent break-induced replication to repair a broken chromosome depends on a distant enhancer site. *Genes Dev* 15(9):1055-60.
- Margulies M., Egholm M., Altman W.E., Attiya S., Bader J.S., Bemben L.A., Berka J., Braverman M.S., Chen Y.J., Chen Z., Dewell S.B., Du L., Fierro J.M., Gomes X.V., Godwin B.C., He W., Helgesen S., Ho C.H., Irzyk G.P., Jando S.C., Alenquer M.L., Jarvie T.P., Jirage K.B., Kim J.B., Knight J.R., Lanza J.R., Leamon J.H., Lefkowitz S.M., Lei M., Li J., Lohman K.L., Lu H., Makhijani V.B., McDade K.E., McKenna M.P., Myers E.W., Nickerson E., Nobile J.R., Plant R., Puc B.P., Ronan M.T., Roth G.T., Sarkis G.J., Simons J.F., Simpson J.W., Srinivasan M., Tartaro K.R., Tomasz A., Vogt K.A., Volkmer G.A., Wang S.H., Wang Y., Weiner M.P., Yu P., Begley R.F. and Rothberg J.M., 2005 Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376-80.
- Martin S.G., Laroche T., Suka N., Grunstein M. and Gasser S.M., 1999 Relocalization of telomeric Ku and SIR proteins in response to DNA strand breaks in yeast. *Cell* 97(5):621-33.
- Maryon E. and Carroll D., 1991 Characterization of recombination intermediates from DNA injected into *Xenopus laevis* oocytes: evidence for a nonconservative mechanism of homologous recombination. *Mol Cell Biol* 11(6):3278-87.
- McClintock B., 1956 Controlling elements and the gene. *Cold Spring Harb Symp Quant Biol* 21:197-216.
- McLysaght A., Hokamp K. and Wolfe K.H., 2002 Extensive genomic duplication during early chordate evolution. *Nat Genet* 31(2):200-4.
- Middleton C.L., Parker J.L., Richard D.J., White M.F. and Bond C.S., 2004 Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic Acids Res* 32(18):5442-51.
- Moore J.K and Haber J.E., 1996 Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in *Saccharomyces cerevisiae*. *Mol Cell Biol* 16(5):2164-73.
- Moore S. P., G. Liti, K. M. Stefanisko, K. M. Nyswaner, C. Chang, Louis E.J. and Garfinkel D.J., 2004 Analysis of a Ty1-les variant of *Saccharomyces paradoxus*: the gain and loss of Ty1 elements. *Yeast* 21:649-660.

- Mortensen U.H., Bendixen C., Sunjevaric I. and Rothstein R., 1996 DNA strand annealing is promoted by the yeast Rad52 protein. *Proc Natl Acad Sci U S A.* 93(20):10729-34.
- Mortimer R.K. and Johnston J.R., 1986 Genealogy of principal strains of the yeast genetic stock center. *Genetics* 113(1):35-43.
- Motegi A. and Myung K., 2007 Measuring the rate of gross chromosomal rearrangements in *Saccharomyces cerevisiae*: A practical approach to study genomic rearrangements observed in cancer. *Methods* 41(2):168-76.
- Mueller J.E., Kemper B., Cunningham R.P., Kallenbach N.R. and Seeman N.C., 1988 T4 endonuclease VII cleaves the crossover strands of Holliday junction analogs. *Proc Natl Acad Sci U S A* 85(24):9441-5.
- Myung K. and Kolodner R.D., 2002 Suppression of genome instability by redundant S-phase checkpoint pathways in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99(7):4500-7.
- Nakada D., Hirano Y. and Sugimoto K., 2004 Requirement of the Mre11 complex and exonuclease 1 for activation of the Mec1 signaling pathway. *Mol Cell Biol* 24(22):10016-25.
- Nakada D., Shimomura T., Matsumoto K. and Sugimoto K., 2003 The ATM-related Tell protein of *Saccharomyces cerevisiae* controls a checkpoint response following phleomycin treatment. *Nucleic Acids Res* 31(6):1715-24.
- Nassif N., Penney J., Pal S., Engels W.R. and Gloor G.B., 1994 Efficient copying of nonhomologous sequences from ectopic sites via P-element-induced gap repair. *Mol Cell Biol* 14(3):1613-25.
- Nicolas A, Treco D, Schultes NP, Szostak JW. 1989 An initiation site for meiotic gene conversion in the yeast *Saccharomyces cerevisiae*. *Nature* 338(6210):35-9.
- Obata K., Hiraga H., Nojima T., Yoshida M.C. and Abe S., 1999 Molecular characterization of the genomic breakpoint junction in a t(11;22) translocation in Ewing sarcoma. *Genes Chromosomes Cancer* 25(1):6-15
- Pâques F. and Haber J.E., 1999 Multiple pathways of recombination induced by double-strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* 63(2):349-404.

- Paquin C. E., and V. M. Williamson, 1984 Temperature Effects on the Rate of Ty Transposition. *Science*. 226(4670):53-55.
- Paull T.T. and Gellert M., 1998 The 3' to 5' exonuclease activity of Mre 11 facilitates repair of DNA double-strand breaks. *Mol Cell* 1(7):969-79.
- Pérez-Ortín J.E., Querol A., Puig S. and Barrio E., 2002 Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Res* 12(10):1533-9.
- Polak A. and Grenson M., 1973 Evidence for a common transport system for cytosine, adenine and hypoxanthine in *Saccharomyces cerevisiae* and *Candida albicans*. *Eur J Biochem* 32(2):276-82.
- Redon R., Ishikawa S., Fitch K.R., Feuk L., Perry G.H., Andrews T.D., Fiegler H., Shapero M.H., Carson A.R., Chen W., Cho E.K., Dallaire S., Freeman J.L., González J.R, Gratacòs M., Huang J., Kalaitzopoulos D., Komura D., MacDonald J.R., Marshall C.R., Mei R., Montgomery L., Nishimura K., Okamura K., Shen F., Somerville M.J, Tchinda J., Valsesia A., Woodwark C., Yang F., Zhang J., Zerjal T., Zhang J., Armengol L., Conrad D.F., Estivill X., Tyler-Smith C., Carter N.P., Aburatani H., Lee C., Jones K.W., Scherer S.W. and Hurles M.E., 2006 Global variation in copy number in the human genome. *Nature* 444(7118):444-54.
- Roelants F., Potier S., Souciet J. L., and 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.
- Roelants F., Potier S., Souciet J.L. and de Montigny J., 1995 Reactivation of the ATCase domain of the *URA2* gene complex: a positive selection method for Ty insertions and chromosomal rearrangements in *Saccharomyces cerevisiae*. *Mol Gen Genet* 246(6):767-73.
- Rothstein R., Michel B. and Gangloff S., 2000 Replication fork pausing and recombination or "gimme a break". *Genes Dev* 14(1):1-10.
- Sanger F., Nicklen S. and Coulson A.R., 1977 DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467.

Scannell D.R., Frank A.C., Conant G.C., Byrne K.P., Woolfit M. and Wolfe K.H., 2007 Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc Natl Acad Sci USA* 104(20):8397-402.

Schacherer J., de Montigny J., Welcker A., Souciet J.L. and Potier S., 2005 Duplication processes in *Saccharomyces cerevisiae* haploid strains. *Nucleic Acids Res* 33(19):6319-26.

Schacherer J., Ruderfer D.M., Gresham D., Dolinski K., Botstein D. and Kruglyak L., 2007 Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. *PLoS ONE*. 2(3):e322.

Schacherer J., Tourrette Y., Potier S., Souciet J.L. and de Montigny J., 2007 Spontaneous duplications in diploid *Saccharomyces cerevisiae* cells. *DNA Repair (Amst)* 6(10):1441-52.

Schacherer J., Tourrette Y., Souciet J. L., Potier S. and de Montigny J., 2004 Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. *Genome Res*. 14(7):1291-7.

Sharp A.J., Cheng Z. and Eichler E.E., 2006 Structural variation of the human genome. *Annu Rev Genomics Hum Genet* 7:407-442.

Shiao M.S., Liao B.Y., Long M. and Yu H.T., 2008 Adaptive evolution of the insulin two-gene system in mouse. *Genetics* 178(3):1683-91.

Shinohara A., Shinohara M., Ohta T., Matsuda S. and Ogawa T., 1998 Rad52 forms ring structures and co-operates with RPA in single-strand DNA annealing. *Genes Cells* 3(3):145-56.

Singh R.K. and Krishna M., 2006 DNA damage induced nucleotide excision repair in *Saccharomyces cerevisiae*. *Mol Cell Biochem* 290(1-2):103-12.

Skov K. and MacPhail S., 1991 Interaction of platinum drugs with clinically relevant x-ray doses in mammalian cells: a comparison of cisplatin, carboplatin, iproplatin, and tetraplatin. *Int J Radiat Oncol Biol Phys* 20(2):221-5.

Smith C.E., Llorente B. and Symington L.S., 2007 Template switching during break-induced replication. *Nature* 447(7140):102-5.

- Song B. and Sung P., 2000 Functional interactions among yeast Rad51 recombinase, Rad52 mediator, and replication protein A in DNA strand exchange. *J Biol Chem* 275(21):15895-904.
- 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., Lépling A., Llorente B., Malpertuy A., Neuvéglise C., Ozier-Kalogéropoulos O., Potier S., Saurin W., Tekaiia F., Toffano-Nioche C., Wésolowski-Louvel M., Wincker P. and Weissenbach J., 2000 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* 487(1):3-12.
- Souciet J.L., Hubert J.C. and Lacroute F., 1982 Cloning and restriction mapping of the yeast URA2 gene coding for the carbamyl phosphate synthetase aspartate-transcarbamylase complex. *Mol Gen Genet* 186(3):385-90.
- Souciet J.L., Nagy M., Le Gouar M., Lacroute F. and 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(1):59-70.
- Southern E.M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* 98:503-517.
- Stasiak A.Z., Larquet E., Stasiak A., Müller S., Engel A., Van Dyck E., West S.C. and Egelman E.H., 2000 The human Rad52 protein exists as a heptameric ring. *Curr Biol* 10(6):337-40.
- Storici F., Durham C.L., Gordenin D.A. and Resnick M.A., 2003 Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc Natl Acad Sci USA* 100(25):14994-9.
- Storici F., Snipe J.R., Chan G.K., Gordenin D.A. and Resnick M.A., 2006 Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. *Mol Cell Biol* 26(20):7645-57.
- Sugawara N. and Haber J.E., 1992 Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. *Mol Cell Biol* 12(2):563-75.

- Sumrada R. and Cooper T.G., 1977 Allantoin transport in *Saccharomyces cerevisiae*. J Bacteriol 131(3):839-47.
- Sung P., Krejci L., Van Komen S. and Sehorn M.G., 2003 Rad51 recombinase and recombination mediators. J Biol Chem 278(44):42729-32.
- Symington L.S., 2002 Role of *RAD52* epistasis group genes in homologous recombination and double-strand break repair. Microbiol Mol Biol Rev 66(4):630-70.
- Szostak J.W., Orr-Weaver T.L., Rothstein R.J. and Stahl F.W., 1983 The double-strand-break repair model for recombination. Cell 33(1):25-35.
- Teo S.H. and Jackson S.P., 1997 Identification of *Saccharomyces cerevisiae* DNA ligase IV: involvement in DNA double-strand break repair. EMBO J 16(15):4788-95.
- The ENCODE Project Consortium, 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447(7146):799-816.
- Tourrette Y., 2004 Sélection et analyse de remaniements chromosomiques chez *Saccharomyces cerevisiae* en contexte diploïde : origine des délétions et des translocations réciproques. Thèse. Université Louis-Pasteur, Strasbourg I.
- Tourrette Y., Schacherer J., Fritsch E., Potier S., Souciet J.L. and de Montigny J., 2007 Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence of ploidy. Mol Microbiol 64(2):382-95.
- Valencia M., Bentele M., Vaze M.B., Herrmann G., Kraus E., Lee S.E., Schär P. and Haber J.E., 2001 NEJ1 controls non-homologous end joining in *Saccharomyces cerevisiae*. Nature 414(6864):666-9.
- Van Komen S., Petukhova G., Sigurdsson S., Stratton S. and Sung P., 2000 Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. Mol Cell 6(3):563-72.
- Vézinhet F., Blondin B. and Hallet J.N., 1990 Chromosomal DNA patterns and mitochondrial DNA polymorphism as tool for identification of œnological strains of *Saccharomyces cerevisiae*. Appl Microbiol Biotechnicol 32:568-571.

- Wagner R., 1999 Variation allélique dans les familles multigéniques chez *Saccharomyces cerevisiae* : sélection *in vivo* et caractérisation moléculaire de nouveaux allèles de gènes codant des transporteurs membranaires. Thèse. Université Louis-Pasteur, Strasbourg I.
- Wagner R., de Montigny J., de Wergifosse P., Souciet J.L. and Potier S., 1998 The ORF YBL042 of *Saccharomyces cerevisiae* encodes a uridine permease. FEMS Microbiol Lett 159(1):69-75
- Wagner R., Straub M.L., Souciet J.L., Potier S. and de Montigny J., 2001 New plasmid system to select for *Saccharomyces cerevisiae* purine-cytosine permease affinity mutants. J Bacteriol 183(14):4386-8
- Wang H., Perrault A.R., Takeda Y., Qin W., Wang H. and Iliakis G., 2003 Biochemical evidence for Ku-independent backup pathways of NHEJ. Nucleic Acids Res 31(18):5377-88.
- Wei W., McCusker J.H., Hyman R.W., Jones T., Ning Y., Cao Z., Gu Z., Bruno D., Miranda M., Nguyen M., Wilhelmy J., Komp C., Tamse R., Wang X., Jia P., Luedi P., Oefner P.J., David L., Dietrich F.S., Li Y., Davis R.W. and Steinmetz L.M., 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. Proc Natl Acad Sci USA 104(31):12825-30.
- Welcker A., de Montigny J., Potier S. and Souciet J.L., 2000 Involvement of very short tandem repeats and the influence of the *RAD52* gene on the occurrence of deletions in *Saccharomyces cerevisiae*. Genetics 156:549-557.
- Wheelan S.J., Scheifele L.Z., Martínez-Murillo F., Irizarry R.A. and Boeke J.D., 2006 Transposon insertion site profiling chip (TIP-chip). Proc Natl Acad Sci USA 103(47):17632-7.
- Wilson T.E., Grawunder U. and Lieber M.R., Yeast DNA ligase IV mediates non-homologous DNA end joining. Nature 388(6641):495-8.
- Winzeler E. A., C.I. Castillo-Davis, G. Oshiro, D. Liang, D. R. Richards, Zhou Y. and Hartl D.L., 2003 Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. Genetics. 163(1):79-89

- Wirth B., Louis V.L., Potier S., Souciet J.L. and Despons L., 2005 Paleogenomics or the search for remnant duplicated copies of the yeast *DUP240* gene family in intergenic areas. *Mol Biol Evol* 22(9):1764-71.
- Wolfe K.H. and Shields D.C., 1997 Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* 387(6634):708-13.
- Yang S., Jiang K., Araki H., Ding J., Yang Y.H. and Tian D., 2007 A molecular isolation mechanism associated with high intra-specific diversity in rice. *Gene* 394(1-2):87-95.
- Young J.M. and Trask B.J., 2002 The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet* 11(10):1153-60.
- Yu X. and Gabriel A., 1999 Patching broken chromosomes with extranuclear cellular DNA. *Mol Cell* 4(5):873-81.
- Yu X. and Gabriel A., 2003 Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics* 163(3):843-56.
- Yu X. and Gabriel A., 2004 Reciprocal translocations in *Saccharomyces cerevisiae* formed by nonhomologous end joining. *Genetics* 166(2):741-51.
- Yvert G., Brem R.B., Whittle J., Akey J.M., Foss E., Smith E.N., Mackelprang R. and Kruglyak L., 2003 Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* 35(1):57-64.
- Zdraveski Z.Z., Mello J.A., Marinus M.G. and Essigmann J.M., 2000 Multiple pathways of recombination define cellular responses to cisplatin. *Chem Biol* 7(1):39-50.
- Zhang C., Roberts T.M., Yang J., Desai R. and Brown G.W., 2006 Suppression of genomic instability by SLX5 and SLX8 in *Saccharomyces cerevisiae*. *DNA Repair (Amst)* 5(3):336-46.
- Zhang J., Rosenberg H.F. and Nei M., 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* 95(7):3708-13.
- Zhong Q., Chen C.F., Chen P.L. and Lee W.H., 2002 BRCA1 facilitates microhomology-mediated end joining of DNA double strand breaks. *J Biol Chem* 277(32):28641-7.

Annexe

Annexe 1

Table de Student

Annexe 2

Influence of the genetic background on the appearance of chromosomal rearrangements in *Saccharomces cerevisiae*

Annexe 3

Tourrette *et al.*, 2007

Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence of ploidy.

Table de Student

Table des valeurs critiques positives **bilatérales** de la loi de Student en fonction du seuil α et du degré de liberté dl . Cette table a été construite avec le logiciel SAS.

$n \backslash \alpha$	0,350	0,300	0,250	0,200	0,150	0,100	0,050	0,020	0,010	0,001
1	1,63185	1,96261	2,41421	3,07768	4,16530	6,31375	12,7062	31,8205	63,6567	636,619
2	1,20963	1,38621	1,60357	1,88562	2,28193	2,91999	4,3027	6,9646	9,9248	31,599
3	1,10452	1,24978	1,42263	1,63774	1,92432	2,35336	3,1824	4,5407	5,8409	12,924
4	1,05730	1,18957	1,34440	1,53321	1,77819	2,13185	2,7764	3,7469	4,6041	8,610
5	1,03055	1,15577	1,30095	1,47588	1,69936	2,01505	2,5706	3,3649	4,0321	6,869
6	1,01335	1,13416	1,27335	1,43976	1,65017	1,94318	2,4469	3,1427	3,7074	5,959
7	1,00137	1,11916	1,25428	1,41492	1,61659	1,89458	2,3646	2,9980	3,4995	5,408
8	0,99254	1,10815	1,24032	1,39682	1,59222	1,85955	2,3060	2,8965	3,3554	5,041
9	0,98578	1,09972	1,22966	1,38303	1,57374	1,83311	2,2622	2,8214	3,2498	4,781
10	0,98043	1,09306	1,22126	1,37218	1,55924	1,81246	2,2281	2,7638	3,1693	4,587
11	0,97608	1,08767	1,21446	1,36343	1,54756	1,79588	2,2010	2,7181	3,1058	4,437
12	0,97249	1,08321	1,20885	1,35622	1,53796	1,78229	2,1788	2,6810	3,0545	4,318
13	0,96948	1,07947	1,20415	1,35017	1,52992	1,77093	2,1604	2,6503	3,0123	4,221
14	0,96690	1,07628	1,20014	1,34503	1,52310	1,76131	2,1448	2,6245	2,9768	4,140
15	0,96468	1,07353	1,19669	1,34061	1,51723	1,75305	2,1314	2,6025	2,9467	4,073
16	0,96275	1,07114	1,19369	1,33676	1,51213	1,74588	2,1199	2,5835	2,9208	4,015
17	0,96105	1,06903	1,19105	1,33338	1,50766	1,73961	2,1098	2,5669	2,8982	3,965
18	0,95954	1,06717	1,18871	1,33039	1,50371	1,73406	2,1009	2,5524	2,8784	3,922
19	0,95819	1,06551	1,18663	1,32773	1,50019	1,72913	2,0930	2,5395	2,8609	3,883
20	0,95699	1,06402	1,18476	1,32534	1,49704	1,72472	2,0860	2,5280	2,8453	3,850
21	0,95590	1,06267	1,18308	1,32319	1,49419	1,72074	2,0796	2,5176	2,8314	3,819
22	0,95491	1,06145	1,18155	1,32124	1,49162	1,71714	2,0739	2,5083	2,8188	3,792
23	0,95401	1,06034	1,18016	1,31946	1,48928	1,71387	2,0687	2,4999	2,8073	3,768
24	0,95318	1,05932	1,17888	1,31784	1,48714	1,71088	2,0639	2,4922	2,7969	3,745
25	0,95242	1,05838	1,17772	1,31635	1,48517	1,70814	2,0595	2,4851	2,7874	3,725
26	0,95173	1,05752	1,17664	1,31497	1,48336	1,70562	2,0555	2,4786	2,7787	3,707
27	0,95108	1,05673	1,17564	1,31370	1,48169	1,70329	2,0518	2,4727	2,7707	3,690
28	0,95048	1,05599	1,17472	1,31253	1,48014	1,70113	2,0484	2,4671	2,7633	3,674
29	0,94993	1,05530	1,17386	1,31143	1,47870	1,69913	2,0452	2,4620	2,7564	3,659
30	0,94941	1,05466	1,17306	1,31042	1,47736	1,69726	2,0423	2,4573	2,7500	3,646
31	0,94892	1,05406	1,17232	1,30946	1,47611	1,69552	2,0395	2,4528	2,7440	3,633
32	0,94847	1,05350	1,17162	1,30857	1,47494	1,69389	2,0369	2,4487	2,7385	3,622
33	0,94804	1,05298	1,17096	1,30774	1,47384	1,69236	2,0345	2,4448	2,7333	3,611
34	0,94764	1,05248	1,17035	1,30695	1,47281	1,69092	2,0322	2,4411	2,7284	3,601
35	0,94726	1,05202	1,16976	1,30621	1,47184	1,68957	2,0301	2,4377	2,7238	3,591
36	0,94691	1,05158	1,16922	1,30551	1,47092	1,68830	2,0281	2,4345	2,7195	3,582
37	0,94657	1,05117	1,16870	1,30485	1,47005	1,68709	2,0262	2,4314	2,7154	3,574
38	0,94625	1,05077	1,16821	1,30423	1,46923	1,68595	2,0244	2,4286	2,7116	3,566
39	0,94595	1,05040	1,16774	1,30364	1,46846	1,68488	2,0227	2,4258	2,7079	3,558
40	0,94566	1,05005	1,16730	1,30308	1,46772	1,68385	2,0211	2,4233	2,7045	3,551
50	0,94343	1,04729	1,16387	1,29871	1,46199	1,67591	2,0086	2,4033	2,6778	3,496
60	0,94194	1,04547	1,16160	1,29582	1,45820	1,67065	2,0003	2,3901	2,6603	3,460
70	0,94088	1,04417	1,15998	1,29376	1,45550	1,66691	1,9944	2,3808	2,6479	3,435
80	0,94009	1,04320	1,15876	1,29222	1,45349	1,66412	1,9901	2,3739	2,6387	3,416
90	0,93948	1,04244	1,15782	1,29103	1,45192	1,66196	1,9867	2,3685	2,6316	3,402
100	0,93899	1,04184	1,15707	1,29007	1,45067	1,66023	1,9840	2,3642	2,6259	3,390

Influence of genetic background on the occurrence of chromosomal rearrangements in *Saccharomyces cerevisiae*

Emilie S. Fritsch, Joseph Schacherer, Claudine Bleykasten-Grosshans, Jean-Luc Souciet,
Serge Potier and Jacky de Montigny*

Laboratory of Molecular Genetics, Genomics and Microbiology, UMR7156, Louis-Pasteur
University and CNRS, Strasbourg, France

Corresponding author: Jacky de Montigny E-mail: demontigny@gem.u-strasbg.fr

Abstract

Chromosomal rearrangements such as duplications and deletions are key factors in evolutionary processes because they promote genomic plasticity. Although the genetic variations in the *Saccharomyces cerevisiae* species have been fairly well documented, little is known so far about the impact of the genetic background on the appearance of rearrangements. Here we compared two different *S. cerevisiae* strains, FL100 and S288c, in terms of the type of rearrangements and the mutation rates observed using the same genetic screening. No transposon rearrangements were detected in S288c, whereas these arrangements constitute a major event in FL100. We also studied the mechanisms responsible for the occurrence of deletions and duplications in mutating genes, the product of which is involved in either the homologous recombination (HR) or the non-homologous end joining (NHEJ). The results obtained here indicate that a Yku80p-independent NHEJ pathway is involved in the occurrence of these rearrangements, which is in agreement with what was observed in the FL100 background. By comparing these two backgrounds, it was thus established that intraspecies genomic variations lead to a variability for the appearance of chromosomal rearrangements due to differences in Ty1 retrotransposon activity.

Introduction

DNA double strand breaks can either occur spontaneously or result from DNA damaging agents such as ionizing radiations or chemical reagents. If this damage is not properly repaired, it can lead to the appearance of chromosomal rearrangements such as duplications, deletions and translocations, which can affect cell growth and survival. These rearrangements are key events in genome reshaping and evolution processes. Many of the genomes sequenced to date show traces of these rearrangements. In a recent study, the *Saccharomyces cerevisiae* clinical strain YJM789 was compared with the reference strain S288c, and many differences were observed between the two genomes, such as large inversions and indels (Wei *et al.*, 2007). In another study, *Saccharomyces cerevisiae* strains isolated after growth in glucose-limiting conditions were found to carry chromosomal rearrangements (Dunham *et al.*, 2002). The resulting duplications and deletions can thus lead to subfunctionalization (Hittinger and Carroll, 2007), neofunctionalization (Young and Trask, 2002) or speciation (Zhang *et al.*, 1998). In multicellular organisms, however, chromosomal rearrangements are often responsible for oncogenesis (Difilippantonio *et al.*, 2002 ; Sharpless *et al.*, 2001) and for many human genetic diseases (for a review, see McKinnon and Caldecott, 2007).

DNA repair mechanisms are essential to each organism, since they preserve the integrity of the genome and prevent the deleterious effects of chromosomal rearrangements. These mechanisms can be classified in two distinct pathways: the homologous recombination (HR) pathway, where long homologous sequences are required for a DNA double strand break to be repaired, and the non-homologous end-joining (NHEJ) pathway, where little or no homology is required.

In order to select spontaneous chromosomal rearrangements, a genetic screening based on a particular allele of the *URA2* gene was developed (Roelants *et al.*, 1995) (Figure 1). The *URA2* gene is located on chromosome X and encodes a multifunctional protein involved in the two first steps of the pyrimidine biosynthesis pathway. The *ura2*_{15,30,72} allele used in this study has three point mutations located in the proximal region of the gene coding for the GATase and CPSase domains, resulting in the loss of all the activities. The loss of GATase and CPSase activities is compensated for by two isoenzymes of the arginine biosynthesis pathway, but not the loss of ATCase activity. This strain is therefore auxotrophic towards uracil, and it becomes prototrophic after the reactivation of the ATCase resulting from complex chromosomal rearrangements. This powerful tool can be used to perform *in vivo*

experiments without the side-effects observed when using mutagenic agents or reporter genes on a plasmid.

In previous studies, rearrangements of several kinds were observed using this screening (Roelants *et al.*, 1995; Welcker *et al.*, 2000). In the haploid context, three types of rearrangements leading to ATCase reactivation were detected: Ty1 insertions upstream from the last point mutation, deletions of the region containing the 3 mutations, and duplications of the region encoding the ATCase followed by the fusion of the duplicated copy with a new promoting sequence. Roelants *et al.* (1997) studied the ATCase reactivation resulting from Ty1 insertions and established that the transcription process is initiated at a sequence present in the LTR region of the Ty1. Deletions have been described by Welcker *et al.* (2000) and duplications by Schacherer *et al.* (2004).

An analysis of the variations at the nucleotide level in some commonly used *Saccharomyces cerevisiae* strains was carried out by Schacherer *et al.* (2007), who detected SNPs (Single Nucleotide Polymorphism) and deletions in various strains. These sequence differences may have important effects on several biological pathways and phenotypes. To assess whether the genetic background affects the mutation rates and the type of chromosomal rearrangements, our genetic screening was used to select spontaneous rearrangements in the S288c context, and the results obtained were compared with those obtained in previous studies on a FL100 background. A total number of 22,446 SNPs and 53 deletions were identified in comparison between the FL100 and S288c strains (Schacherer *et al.*, 2007). The divergence observed between the two strains amounted to 0.21%.

Interestingly, in the S288c background, duplications and deletions events were obtained but no Ty1 insertions were detected as responsible for the ATCase reactivation: it was therefore concluded that the occurrence of chromosomal rearrangements is background-dependent. In addition, the possible involvement of various repair mechanisms in the frequency and the types of rearrangements occurring in the haploid context was studied, starting with the impact of homologous recombination selecting revertants in the $\Delta rad52$ strain. Since Rad59p plays an important role in Single-Strand Annealing processes between direct repeats, a $\Delta rad59$ mutant was constructed and selections were performed in this background. In order to establish whether these genes might play a role, the mutation rates and the types of rearrangements obtained were compared between these contexts and a reference *ura2_{15,30,72}* strain. Inhibiting the RAD52-dependent homologous recombination increased the deletion rates, whereas inactivation of the SSA pathway increased the duplication rate. Secondly, the

effects of non-homologous end joining were studied by mutating *LIG4* and *YKU80* in our *ura2_{15,30,72}* background. *LIG4* mutation was found to affect the mutation rates in the case of deletions and duplications, which was not the case with *YKU80* deletion: this finding points to the conclusion that a Yku80p-independent NHEJ mechanism is responsible for the occurrence of chromosomal deletions and duplications.

Materials and Methods

Yeast strains and media Yeast cells were grown at 30°C in liquid or solid (2% agar) yeast peptone dextrose (YPD) and yeast nitrogen base (YNB) supplemented. The *ura2*_{15,30,72} strains are closely related to BY4741 and BY4742 backgrounds (Table 1). Single deletion mutants were obtained by crossing our laboratory strain *ura2*_{15,30,72} with yeast deletion strains from the EUROSCARF collection.

Selection of Ura⁺ mutants One isolated colony was grown in 300 µl YPD medium overnight, plated on YPD medium and incubated at 30°C for 4-5 days. The cells were then harvested, resuspended in 1.5 ml water and spread on YNB appropriately supplemented without uracil. The cells were then incubated at 30°C thus allowing the growth of spontaneous Ura⁺ revertants. In order to determine the total number of cells plated, a 100 µl aliquot of the culture was diluted 10⁶ times, plated on YPD medium and incubated at 30°C. This procedure corresponds to one independent selection. The same procedure was used to perform selections at 25°C.

Mutation rate determination The mutation rates (mutations/cell/selection) were determined using the maximum-likelihood method described by Lea and Coulson (1949). The 95% confidence limits were calculated using Student's *t*-test. To compare the proportion of rearrangements obtained with the control values, a χ^2 test was performed.

Southern blot analysis Genomic DNA from *S. cerevisiae* was prepared as described by Hoffman and Winston (1997). Digestion of this DNA was then performed with the restriction endonuclease *Bam*HI (ROCHE) as described by the manufacturer. DNA digestions were migrated in a 1% agarose gel by electrophoresis and transferred onto a Hybond N⁺ membrane. Digoxigenin-labeled DNA probes were prepared using the DNA labeling kit (ROCHE). Detection was then carried out using a NBT/BCIP colorimetric method (ROCHE).

PCR amplification, DNA sequencing and sequence analysis The primers used for PCR amplification and sequencing were chosen on the basis of the published genomic sequence of S288c. DNA fragments were obtained by performing PCR amplification using *Taq* DNA

polymerase from MP, and the PCR conditions used were those described by the manufacturers. The Ty1 insertions in the *URA2* coding sequence were characterized by performing PCR using a specific primer for the LTR of the retrotransposon Ty1 and a second specific primer for the coding sequence of the ATCase. The deletions were detected by performing primer-walking PCR and determining whether a PCR amplification product was present or absent between close primer pairs. This made it possible to define the boundaries of the deletions.

The PCR products were purified using MicroSpin S400 (GE Healthcare). DNA sequencing was performed on the purified fragments as described by Sanger *et al.* (1977). The sequencing was carried out using AmpliTaq FS DNA polymerase and BIGDYE™ terminators. Sequence reactions were analyzed with an Applied Biosystems 373XL sequencer.

Results

The genetic background affects the occurrence of chromosomal rearrangements.

In previous studies (Roelants *et al.*, 1995; Welcker *et al.*, 2000), all the selections were carried out in the FL100 *ura2_{15,30,70}* context (Lacroute, 1968). Among the chromosomal rearrangements leading to ATCase reactivation in FL100, Ty1 insertions usually account for 66%, deletions for 17% and duplications for 17%.

In this study, a S288c background was used to assess whether a genetic background may affect the mutation rate and the type of rearrangements detected with our genetic screening. Revertants were therefore selected from a *MATa ura2_{15,30,72}* strain by performing 155 selections, amounting to 3.78×10^{11} spread cells. From these selections, 29 independent revertants were obtained. The mutation rate determined using a maximum-likelihood method was 1.025×10^{-10} mutations/cell/selection (0.676×10^{-10} ; 1.373×10^{-10}) (Figure 2 and Table 2).

The type of chromosomal rearrangements present in the various revertants selected was determined by performing a *Bam*HI restriction pattern analysis based on Southern blot hybridization. Using a DNA probe specific for the ATCase domain, a single 6.8-kb band was observed with the *ura2_{15,30,72}* strain. A modified restriction profile was detected in all the selected revertants. In the case of a deletion or a Ty1 retrotransposon insertion, a single band differing in size from 6.8 kb was observed, whereas in the case of a duplication, two bands were detected: one corresponding to the *ura2_{15,30,72}* allele and one corresponding to the duplicated ATCase coding region. We discriminated between deletions and Ty1 insertions events using a PCR approach, as described in Material and Methods section. It was thus established that among the 29 events leading to an ATCase reactivation, 20 (69%) were deletions and 9 (31%) were duplications. The boundaries of deletions were determined and microhomologies ranging from 1 to 11 bp in size were observed.

Contrary to what was observed by Roelants *et al.* (1995), a larger number of deletions and duplications were obtained here, and no reactivation by Ty1 insertions was found to occur in the S288c background. These findings suggest that Ty1 retrotransposons may be less active in S288c than in FL100, and are thus less likely to insert upstream of ATCase and lead to a reactivation of the activity. The activity of Ty1 retrotransposons was therefore tested.

The Ty1 activity is lower in the S288c than the FL100 background.

Paquin and Williamson (1984) established that the temperature affects the rate of Ty transposition. They showed that the rate of transposition increases at temperatures below

30°C, the optimal growth temperature for *Saccharomyces cerevisiae*. To determine whether the absence of Ty1 insertions in our context is due to a decrease in Ty1 transposition activity, selections were performed at 25°C. Twenty-four selections amounting to 1.04×10^{11} spread cells were performed at that temperature and 16 independent revertants were obtained. The mutation rate was determined: 3.70×10^{-10} mutations/cell/selection (2.12×10^{-10} ; 5.27×10^{-10}). This mutation rate was 3.6 times higher than at 30°C. In addition, at 25°C, four Ty1 insertions (25%) were detected, whereas none were selected at the optimal growth temperature. These results led to the conclusion that the Ty1 retrotransposons are active in the S288c background but that they show lower rates of activity than those observed in the FL100 background.

Using this S288c background, we studied the mechanisms leading to the occurrence of chromosomal deletions and duplications. We therefore worked at 30°C, since at that temperature, we were able to neglect the Ty1 insertions and focus on the deletions and duplications events. Daley and Wilson (2005) showed that deletions can occur due to HR, SSA and NHEJ, depending on the DNA overhang length. Schacherer *et al.* (2004) proposed in addition that the mechanism involved in the occurrence of genic duplications was a retroposition mechanism requiring HR. Lastly Koszul *et al.* (2004) suggested that Break Induced Repair (BIR) might lead to the appearance of segmental duplications. Here we decided to focus on the impact of homologous recombination and non-homologous end joining processes in the occurrence of deletions and duplications.

Involvement of a homologous recombination mechanism in the occurrence of chromosomal deletions and duplications

The deletion rate increases in the $\Delta rad52$ context.

The first target used to investigate the impact of homologous recombination on the selection of chromosomal rearrangements was *RAD52*. Rad52p is a major component of HR: it binds to ssDNA and is involved in searching for and annealing a complementary ssDNA. A deletion in *RAD52* results in defective double-strand break repair and single-strand annealing processes. In this context, 83 selections corresponding to 2.4×10^{11} cells were performed and 21 independent revertants were obtained. The mutation rate in the $\Delta rad52$ context was 1.69×10^{-10} mutations/cell/selection (1.13×10^{-10} ; 2.25×10^{-10}), as shown in Table 2.

Among the 21 revertants characterized, 3 (14%) duplications and 18 (86%) deletions with microhomologies ranging from 1 to 11 bp were obtained (Table 2 and Figure 2). There were no significant differences between the duplication rates, but the deletion rate was found to be twice as high as that recorded in the reference strain, which indicated that a defective homologous recombination leads to an increased number of repair leading to a deletion.

The occurrence of duplications is repressed in RAD59

Using the approach described above, we used a strain with a mutated gene, the product of which is involved in the SSA pathway (Storici *et al.*, 2006). Among the various candidates available, *RAD59* was chosen. Rad59p is a protein which contributes to repairing DNA double-strand breaks by annealing complementary single-strand DNA. In this *ura2_{15,30,72} Δrad59*, 130 selections corresponding to 2.88×10^{11} cells were carried out and 37 revertants were isolated. The mutation rate in this context was found to be 1.81×10^{-10} mutations/cell/selection (1.27×10^{-10} ; 2.35×10^{-10}), and no statistical differences were observed in comparison with the reference strain.

The 37 revertants could be identified as 20 deletions (54%) and 17 duplications (46%). The microhomologies observed at the deletion boundaries ranged from 1 bp to 11 bp in size. The duplication rate was 2.5 times higher in this context, but no significant differences were observed in the deletion rates. This result allows us to conclude that Rad59p restrains the appearance of duplications.

Involvement of NHEJ in the occurrence of deletions and duplications

In order to test the impact of NHEJ on the occurrence of deletions and duplications, two deletion mutants in this pathway, *Δyku80* and *Δlig4* were constructed in the *ura2_{15,30,72}* background. Yku80p is a subunit of the Ku complex involved in telomere maintenance and in DNA binding at DSB sites, and Lig4p is the DNA ligase required by NHEJ along with the cofactors Lif1p and Nej1p.

The occurrence of deletions and duplications is Yku80p-independent.

In the *ura2_{15,30,72} Δyku80* context, 135 selections corresponding to 2.7×10^{11} cells were performed and 22 independent revertants were thus obtained. The mutation rate in the *Δyku80* context was 0.997×10^{-10} mutations/cell/selection (0.604×10^{-10} ; 1.39×10^{-10}) (Figure 2 and

Table 2). No significant differences were therefore observed in this respect between *ura2_{15,30,72}Δyku80* and the reference strain.

Among the 22 revertants characterized, 6 duplications (27%) and 16 deletions (73%) with microhomologies ranging from 2 to 11 bp in size were obtained. These results are in agreement with the data obtained for the reference strain. It was therefore concluded that in our system, deletions and duplications can occur in the Yku80p-independent context.

A Δlig4 context leads to a decrease of DNA repair

In the *ura2_{15,30,72}Δlig4* background, 151 selections corresponding to 3.7×10^{11} cells were performed and 9 independent revertants were isolated. The mutation rate was found to be 0.204×10^{-10} mutations/cell/selection (0.055×10^{-10} ; 0.353×10^{-10}), which corresponds to a statistically significant 5-fold decrease. Molecular characterization of the 9 revertants showed that all the reversion events were due to deletions with microhomologies ranging from 1 to 11 bp in size. When the NHEJ pathway is inactivated, a decrease of efficient repair is observed. Moreover since the deletion rate was found to be 3.4 times lower in this context, it was concluded that NHEJ has an impact on the occurrence of deletions. No duplications could be selected, which suggests that the inactivation of NHEJ also inhibits the appearance of duplications in the *ura2* system.

Discussion

It was recently established that DNA sequence variations frequently occur among *S. cerevisiae* strains (Winzeler *et al.*, 2003; Schacherer *et al.*, 2007). These sequence differences have important effects on the phenotypes (Steinmetz *et al.*, 2002; Yvert *et al.*, 2003; Deutschbauer and Davis, 2005; Wei *et al.*, 2007). However, little is known so far about the impact of genetic variations on the occurrence of chromosomal rearrangements.

Using a *URA2*-based genetic screening, we investigated whether the genetic background affects the types of chromosomal rearrangements selected and the mutation rates. We also studied the impact of the repair pathways on the occurrence of various chromosomal rearrangements in a S288c background to determine whether these recombination mechanisms were involved or not. Furthermore, to assess whether the repair mechanisms are conserved from one strain to another, we compared these results with what was observed in FL100.

Different genetic backgrounds lead to a variation of the type of chromosomal rearrangements selected.

To test the impact of the genetic background, two strains, the S288c and the FL100 strains were compared in terms of the rates of occurrence of various rearrangements such as deletions, duplications and Ty1 retrotransposon insertions. Interestingly, in the S288c context, no Ty1 retrotransposition insertions were selected, whereas these rearrangements account for 2/3 of the selected events in FL100. This finding points to the conclusion that either there may be fewer Ty1 retrotransposons in S288c than in FL100 or that they show lower levels of activity.

The location of the transposons and their number were compared by Gabriel *et al.* (2006) between several *S. cerevisiae* laboratory strains. They demonstrated that no Ty1 is present in the genome of the RM11 strain. The retrotransposon content therefore varies from one species to another. Variations have been reported concerning the retrotransposons in the *Saccharomyces sensu stricto* complex (Liti *et al.*, 2005) and a strain which is a Ty1-less variant of *S. paradoxus* has been described (Moore *et al.*, 2004). In addition, Ty elements are known to be key factors contributing to chromosomal rearrangements, and these elements can be selectively advantageous in experimental evolution studies (Dunham *et al.*, 2002; Blanc and Adams, 2003). However, since the proportion of Ty1 elements seems to be similar in FL100 and S288c backgrounds (Schacherer *et al.*, 2007), the differences between the Ty1

insertions observed between the two backgrounds can probably be accounted for by a difference of Ty1 activity.

Paquin and Williamson (1984) reported that Ty transposition is temperature-sensitive and that this process is enhanced at temperatures below the optimal growth temperature. The activity of the retrotransposons in the S288c background was therefore tested by selecting revertants at 25°C. At that temperature, 4 Ty1 insertions leading to ATCase reactivation were obtained. This finding shows that Ty1 retrotransposons are active in the S288c background but their activity is probably decreased in that background contrary to what was observed in the case of the FL100 background.

In this study, intra-species variations were found to occur as regards the appearance of chromosomal rearrangements between the S288c and the FL100 strains, and this variability was attributed to differences in the activity of the Ty1 elements. In addition, the regulation of the Ty1 retrotransposition might differ between the two strains. For instance, it has been established that *S. cerevisiae* undergoes transcriptional (Jiang, 2002) and post-transcriptional (Garfinkel *et al.*, 2003) control which limit transposition in a copy-number dependent manner. Differences in these control mechanisms might also explain the transpositional differences observed between the two strains. Further studies on the FL100 strain are now required to determine how the activity of the Ty1 retrotransposons is regulated and establish their exact localization.

Non-Homologous End Joining contributes to the occurrence of deletions.

Contrary to what occurs in higher eukaryotes, where non-homologous recombination processes predominate, the main repair pathway in yeast is the homologous recombination pathway (Aylon and Kupiec, 2004). It was therefore proposed to investigate the contribution of HR on the appearance of chromosomal deletions.

In the $\Delta rad52$ context, the deletion rate was found to be twice as high as in the reference strain, which suggests that the occurrence of deletions is favoured when *RAD52*-dependent homologous recombination is inactivated. This finding is in agreement with what was previously observed in the FL100 background (Welcker *et al.*, 2000).

For deletions to be repaired by homologous recombination, a homologous sequence at least 60 bp in length is required (Sugawara and Haber, 1992). However, since no sequence homologous to the ATCase-coding domain is present upstream of the *URA2* gene, no ATCase

reactivation is possible via the HR repair pathway. In addition, since microhomologies (1 to 11 bp) were detected at the junctions between the deletions in every context studied, it was concluded that a mechanism of non-homologous recombination was probably responsible for the occurrence of the deletions.

When mutating *LIG4*, the ligase required for NHEJ to occur, we observed a decrease in the deletion rate. However, when *YKU80*, which is also involved in NHEJ, was mutated, no differences were observed in comparison with the reference strain.

Boulton and Jackson (1996) established that two different NHEJ pathways can be distinguished in *S. cerevisiae*. One of these pathways is accurate and Ku-dependent, whereas the other one is error-prone, Ku-independent and involves short homologies. In addition, it has been suggested that the two NHEJ pathways might lead to different chromosomal rearrangements in *S. cerevisiae* (Yu and Gabriel, 2003).

It therefore seems likely that the occurrence of deletions in the *ura2*_{15,30,72} context may be attributable to a Yku80p-independent NHEJ mechanism. These deletions would result from a DSB followed by a NHEJ repair involving short sequence homologies. In addition, when HR is inactivated, the rates of HR and NHEJ seem to be unbalanced and repair by NHEJ is favoured, thus leading to a higher rate of deletions.

The occurrence of duplications depends on NHEJ.

Rad59p is involved in SSA, which is an error-free mechanism involving short homologous sequences (Sugawara and Haber, 2000). In the $\Delta rad59$ context, a 2.5-fold increase in the duplication rate was observed, which suggests that inactivation of the SSA pathway may favour the occurrence of duplications.

On the other hand, in our studies on the $\Delta lig4$ context, the fact that no duplications were selected suggested that the duplication rate was greatly impaired. However, when *YKU80* was mutated, no significant differences were observed in the duplication rates. It was therefore concluded that a Ku-independent NHEJ mechanism is involved in the occurrence of duplications in our system. Duplications may possibly result from two DSBs events followed by a NHEJ repair process occurring between two microhomologies at the DNA ends.

The results obtained here suggest that a Non-Homologous End Joining mechanism independent of the Ku complex may be responsible for the occurrence of deletions and duplications. The fact that this repair mechanism is highlighted in absence of HR and SSA

suggests that in a wild-type strain, a balance between the various mechanisms exists to maintain the integrity of the genome. The results obtained on the occurrence of deletions in the S288c context were similar to those previously obtained on a FL100 background (Welcker *et al.*, 2000). Although phenotypic differences and sequence variations can be observed from one strain to another, the use of major repair pathways such as HR and NHEJ therefore remain unchanged.

By contrast, no Ty1 insertions were detected in S288c whereas these are major events contributing to ATCase reactivation in the FL100 background. This difference may be attributable to a difference in Ty1 activity rather than a difference between the mechanisms involved, since the repair mechanisms are the same between FL100 and S288c. It would be worth further analyzing the activity of retrotransposons and their regulation in the FL100 background and comparing the results with those obtained here on S288c in order to confirm this point.

ACKNOWLEDGEMENTS

We thank Malek Alioua for their help with DNA sequencing in the Strasbourg CNRS/IBMP department facilities. We thank Nicolas Jauniaux for his helpful comments on the manuscript. This work was partly supported by the ANR-05-BLAN-0331-03 (GENARISE) and the Genolevure-3 sequencing consortium GDR CNRS 2354. E. F. was supported by a grant from the French Ministère de l'Enseignement Supérieur et de la Recherche.

LITERATURE CITED

Aylon Y., and M. Kupiec, 2004 DSB repair: the yeast paradigm. *DNA Repair (Amst)*. 3(8-9):797-815

Blanc V. M., and J. Adams, 2003 Evolution in *Saccharomyces cerevisiae*: identification of mutations increasing fitness in laboratory populations. *Genetics*. 165(3):975-83

Boulton S. J., and S. P. Jackson, 1996 *Saccharomyces cerevisiae* Ku70 potentiates illegitimate DNA double-strand break repair and serves as a barrier to error-prone DNA repair pathways. *EMBO J*. 15(18):5093-103.

Daley J. M., and T. E. Wilson, 2005 Rejoining of DNA double-strand breaks as a function of overhang length. *Mol Cell Biol*. 25(3):896-906.

Deutschbauer A. M., and R. W. Davis, 2005 Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet*. 37(12):1333-40

Difilippantonio M. J., S. Petersen, H. T. Chen, R. Johnson, M. Jasin *et al.*, 2002 Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J Exp Med*. 196(4):469-80.

Dunham M. J., H. Badrane, T. Ferea, J. Adams, P. O. Brown *et al.*, 2002 Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* 99(25):16144-49.

Gabriel A., J. Dapprich, M. Kunkel, D. Gresham, S. C. Pratt *et al.*, 2006 Global mapping of transposon location. *PLoS Genet*. 2(12):e212.

Garfinkel D. J., K. Nyswaner, J. Wang, and J. Y. Cho, 2003 Post-transcriptional cosuppression of Ty1 retrotransposition. *Genetics* 165:83-99.

Hittinger C. T., and S. B. Carroll, 2007 Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* 449:677-682.

Hoffman C. S., and F. Winston, 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57:267-272.

Jiang Y. W., 2002 Transcriptional cosuppression of yeast Ty1 retrotransposons. *Genes Dev.* 16(4):467-78.

Koszul R., S. Caburet, B. Dujon, and G. Fischer, 2004 Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J.* 23(1):234-43.

Lacroute F., 1968 Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J Bacteriol.* 95(3):824-32.

Lea D. E., and C. A. Coulson, 1949 The distribution of numbers of mutants in bacterial populations. *J. Genet.* 49:264-285.

Liti G., A. Peruffo, S. A. James, I. N. Roberts, and E. J. Louis, 2005 Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric-associated sequences in the *Saccharomyces sensu strict* complex. *Yeast* 22:177-192.

McKinnon P. J., and K.W. Caldecott, 2007 DNA strand break repair and human genetic disease. *Annu. Rev. Genom. Human Genet.* 8:37-55.

Moore S. P., G. Liti, K. M. Stefanisko, K. M. Nyswaner, C. Chang *et al.*, 2004 Analysis of a Ty1-les variant of *Saccharomyces paradoxus*: the gain and loss of Ty1 elements. *Yeast* 21:649-660.

Paquin C. E., and V. M. Williamson, 1984 Temperature Effects on the Rate of Ty Transposition. *Science.* 226(4670):53-55.

Roelants F., S. Potier, J. L. Souciet, and J. de Montigny, 1995 Reactivation of the ATCase domain of the URA2 gene complex: a positive selection method for Ty insertions and chromosomal rearrangements in *Saccharomyces cerevisiae*. *Mol Gen Genet.* 246(6):767-73.

Roelants F., S. Potier, J. L. Souciet, and J. de Montigny, 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.

Sanger F., S. Nicklen, and A. R. Coulson, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74:5463-5467.

Schacherer J., Y. Tourrette, J. L. Souciet, S. Potier, and J. de Montigny, 2004 Recovery of a function involving gene duplication by retroposition in *Saccharomyces cerevisiae*. Genome Res. 14(7):1291-7.

Schacherer J., D. M. Ruderfer, D. Gresham, K. Dolinski, D. Botstein *et al.*, 2007 Genome-wide analysis of nucleotide-level variation in commonly used *Saccharomyces cerevisiae* strains. PLoS ONE. 2(3):e322.

Sharpless N. E., D. O. Fergusson, R. C. O'Hagan, D. H. Castrillon, C. Lee *et al.*, 2001 Impaired nonhomologous end-joining provokes soft tissue sarcomas harboring chromosomal translocations, amplifications and deletions. Mol Cell. 8(6):1187-96.

Steinmetz L. M., H. Sinha, D. R. Richards, J. I. Spiegelman, P. J. Oefner *et al.*, 2002 Dissecting the architecture of a quantitative trait locus in yeast. Nature. 416(6878):326-30

Storici F., J. R. Snipe, G. K. Chan, D. A. Gordenin, and M. A. Resnick, 2006 Conservative repair of a chromosomal double-strand break by single-strand DNA through two steps of annealing. Mol Cell Biol. 26(20):7645-57.

Sugawara N., and J.E. Haber, 1992 Characterization of double-strand break-induced recombination : homology-requirements and single-stranded DNA formation. Mol. Cell. Biol. 12(2):563-575.

Sugarawa N., G. Ira, and J. E. Haber, 2000 DNA length dependence of the single-strand annealing pathway and the role of *Saccharomyces cerevisiae* *RAD59* in double-strand break repair. *Mol Cell Biol.* 20(14):5300-9.

Wei, W., J. H. McCusker, R. W. Hyman, T; Jones, Y. Ning *et al.*, 2007 Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strain YJM789. *Proc Natl Acad Sci USA* 104(31):12825-30.

Welcker A., J. de Montigny, S. Potier, and J. L. Souciet, 2000 Involvement of very short tandem repeats and the influence of the *RAD52* gene on the occurrence of deletions in *Saccharomyces cerevisiae*. *Genetics* 156:549-557.

Winzeler E. A., C.I. Castillo-Davis, G. Oshiro, D. Liang, D. R. Richards *et al.*, 2003 Genetic diversity in yeast assessed with whole-genome oligonucleotide arrays. *Genetics.* 163(1):79-89

Young J. M., and B. J. Trask, 2002 The sense of smell: genomics of vertebrate odorant receptors. *Hum Mol Genet.* 11(10):1153-60.

Yu X., and A. Gabriel, 2003 Ku-dependent and Ku-independent end-joining pathways lead to chromosomal rearrangements during double-strand break repair in *Saccharomyces cerevisiae*. *Genetics.* 163(3):843-56.

Yvert G., R. B. Brem, J. Whittle, J. M. Akey, E. Foss *et al.*, 2003 Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet.* 35(1):57-64.

Zhang J., H. F. Rosenberg, and M. Nei, 1998 Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A* 95(7):3708-13.

Table 1: Yeast strains used in this study.

Strains	Genotype
aUHT	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63</i>
aΔ <i>lig4</i>	<i>MATa ura2_{15,30,72} his3Δ200 lys2Δ0 lig4::KANMX4</i>
aΔ <i>yku</i>	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63 yku80::KANMX4</i>
aΔ <i>rad52</i>	<i>MATa ura2_{15,30,72} his3Δ200 lys2Δ0 rad52::KANMX4</i>
aΔ <i>rad59</i>	<i>MATa ura2_{15,30,72} his3Δ200 trp1Δ63 rad59::KANMX4</i>

Table 2: Number of revertants obtained in the contexts studied and the corresponding mutation rates.

The mutation rates (mutations/cell/selection) were determined using the maximum-likelihood method described by Lea and Coulson (1949). The 95% confidence limits were calculated using Student's *t*-test.

Strains	No. of revertants or ATCase reactivation events (mutation rate $\times 10^{-10}$) (95% confidence interval)		
	Global	Deletion	Duplication
aUHT	29 1,025 (0,676 ; 1,373)	20 (69%) 0,707 (0,466 ; 0,947)	9 (31%) 0,317 (0,210 ; 0,426)
aΔ <i>rad52</i>	21 1,69 (1,133 ; 2,247)	18 (86%) 1,45 (0,974 ; 1,932)	3 (14%) 0,237 (0,158 ; 0,314)
aΔ <i>rad59</i>	37 1,81 (1,27 ; 2,35)	20 (54%) 0,977 (0,685 ; 1,269)	17 (46%) 0,832 (0,584 ; 1,081)
aΔ <i>yku80</i>	22 0,997 (0,604 ; 1,39)	16 (73%) 0,728 (0,441 ; 1,015)	6 (27%) 0,269 (0,163 ; 0,375)
aΔ <i>lig4</i>	9 0,204 (0,055;0,353)	9 (100%) 0,204 (0,055;0,353)	0

FIGURE LEGENDS

Figure 1: The *ura2*_{15,30,72} allele and its corresponding multifunctional protein Ura2p. ns15 and ns 30 correspond to the positions of the non-sense mutations and fs30 to the position of the frameshift mutation. GATase stands for Glutamine AmidoTransferase, CPSase for carbamylphosphate synthetase, DHOase-like for dihydroorotase-like and ATCase for aspartyltranscarbamylase.

Figure 2: Mutation rates in the genetic contexts studied. Error bars give the standard deviations.

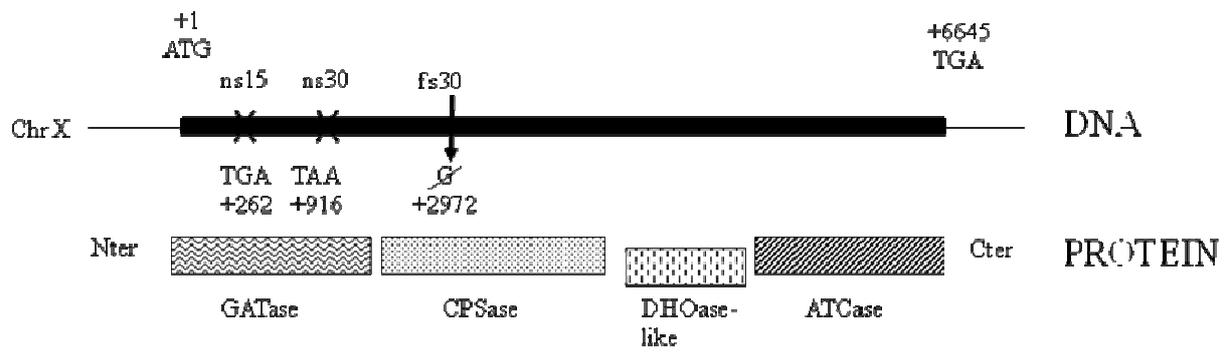


Figure 1

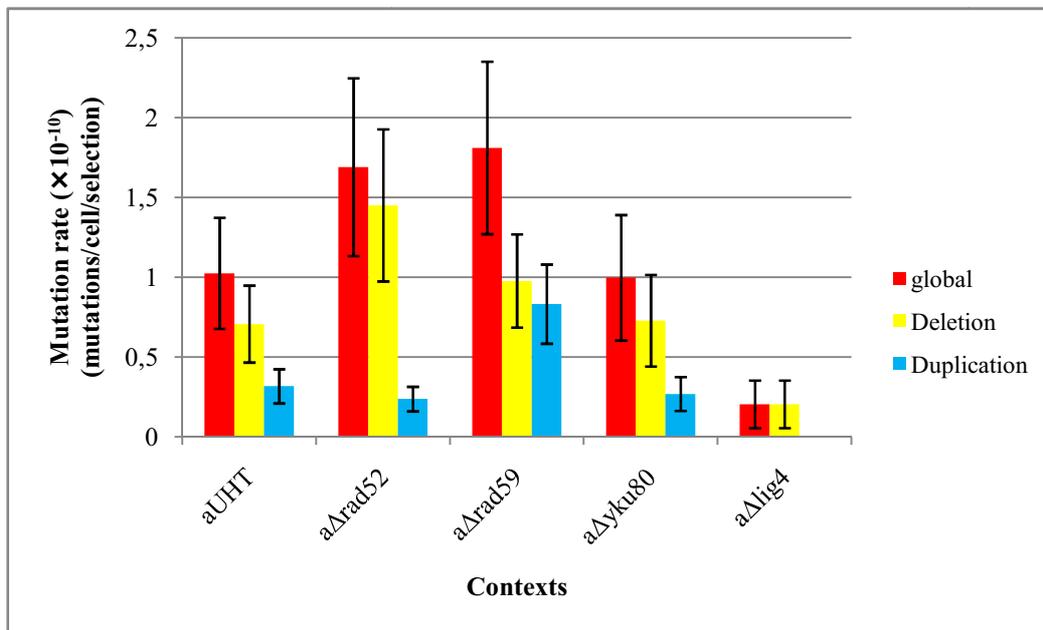


Figure 2

Spontaneous deletions and reciprocal translocations in *Saccharomyces cerevisiae*: influence of ploidy

Yves Tourrette, Joseph Schacherer, Emilie Fritsch, Serge Potier, Jean-Luc Souciet, Jacky de Montigny

Molecular Microbiology, 2007, Vol. 64, N° 2, Pages 382-395

Annexe 3 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Les utilisateurs de l'ULP peuvent consulter cette publication sur le site de l'éditeur :

<http://dx.doi.org/10.1111/j.1365-2958.2007.05660.x>

La version imprimée de cette thèse peut être consultée à la bibliothèque ou dans un autre établissement via une demande de prêt entre bibliothèques (PEB) auprès de nos services :

<http://www-sicd.u-strasbg.fr/services/peb/>