

- Thèse présentée pour obtenir le grade de
- Docteur de l'Université Louis Pasteur
- Strasbourg I
-
-
-
- Discipline : biologie cellulaire et moléculaire
- par Joseph Schacherer ■

Duplications dans le génome
de *Saccharomyces cerevisiae* :
sélections, caractérisation et mécanismes

Soutenue publiquement le 21 janvier 2005 ■

Membres du jury ■

- M. André Goffeau, rapporteur externe
- M. Alain Bucheton, rapporteur externe
- M. Jean-Marc Reichhart, rapporteur externe
- M. Jacky de Montigny, examinateur
- M. Serge Potier, directeur de thèse

REMERCIEMENTS

Le travail présenté dans ce mémoire a été réalisé au sein du laboratoire de " Dynamique, Évo - lution et Expression de Génomes de Micro-organismes " FRE 2326 / CNRS. Je tiens à remercier Serge Potier et Jean-Luc Souciet de m'avoir accueilli dans leur équipe et d'avoir suivi ce projet tout au long de son développement.

Je tiens à remercier Messieurs Alain Bucheton (Institut de Génétique Humaine, Montpellier), André Goffeau (Université de Louvain-la-Neuve) et Jean-Marc Reichhart (Université Louis - Pasteur, Strasbourg) d'avoir accepté d'examiner mon travail à une période où l'on est en droit d'avoir d'autres distractions.

J'exprime toute ma reconnaissance à Jacky de Montigny qui m'a énormément apporté durant ces années de DEA et thèse, aussi bien sur le plan scientifique que personnel. Merci de m'avoir encadré, conseillé et surtout supporté dans mes moments de doute (très fréquents...).

Merci à tous les membres du groupe "levure" et plus particulièrement à Yvess avec qui j'ai passé de très bons moments à la paillasse mais également lors des congrès. Merci à Marie-Laure Straub qui m'a initié aux techniques de génétique de levure lors de mon tout premier passage au laboratoire et à Catherine Spehner pour ses coups de mains. Ce travail a également bénéficié de la contribution technique de Shiro Horibe, Christelle Sibert et Najwa Kasbaoui et environnementale de Claudine Bleykasten, Valérie Braun, Émilie Fritsch et Marie-Hélène Meyer.

Benôit Kammerer m'a été d'un grand secours tout au long de ma thèse. Sa disponibilité et sa compétence ont largement contribué à la naissance de ce mémoire.

J'en profite pour remercier collectivement tous les membres du laboratoire qui ont contribué de près ou de loin à ce travail.

Une pensée particulière à mes parents et toute ma famille pour leur soutien et leur confiance. Merci Rebeka pour la mise en page du manuscrit et surtout de m'avoir supporté ces derniers mois.

TABLE DES MATIÈRES

INTRODUCTION

1. Préambule historique	19
2. Analyse des régions dupliquées présentes chez les organismes séquencés	21
2.1. Identification de gènes dupliqués	21
2.2. Identifications de segments dupliqués	22
2.3. Proportion de gènes au sein des génomes séquencés	23
3. Comment une séquence peut-elle se dupliquer ?	25
3.1. Duplication génique	25
3.2. Duplication de segment	25
3.3. Duplication de chromosome ou aneuploïdie	27
3.4. Duplication complète du génome ou polyploïdie	28
4. Mécanismes à l'origine des duplications géniques et segmentales	29
4.1. Duplications géniques	30
4.1.1. Duplication génique en tandem	30
4.1.2. Duplications géniques dispersées	30
4.2. Duplication de segments	32
4.2.1. Duplication et recombinaison inégale	32
4.2.2. Duplication et erreur de réplication	32
5. Devenir des gènes dupliqués	33
5.1. La pseudogénéisation	35
5.2. La conservation de la fonction	36
5.3. La subfonctionnalisation	37
5.4. La néofonctionnalisation	38
6. <i>Saccharomyces cerevisiae</i> : un organisme modèle pour l'étude des duplications	40

6.1. Les catégories de duplications présentes chez <i>Saccharomyces cerevisiae</i>	42
6.1.1. Les duplications dispersées	43
6.1.2. Les duplications en tandem	43
6.1.3. Les duplications subtélomériques	45
6.1.4. Les blocs « fossiles »	45
6.2. Les duplications mises en évidence par des approches expérimentales	48
6.2.1. L'amplification génique	48
6.2.2. L'aneuploïdie	48
7. Système de sélection utilisé au laboratoire pour l'étude des duplications	49

MATÉRIELS ET MÉTHODES

1. Organismes et méthodes 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
1.1.3. Milieux de culture	51
1.1.4. Conditions de culture	53
1.2. <i>Escherichia coli</i>	53
1.2.1. Souche utilisée	53
1.2.2. Milieux de culture	53
1.2.3. Conditions de culture	53
1.3. Évaluation de la concentration des cultures	53
1.4. Conservation des souches	54
2. Vecteur de clonage	54
3. Techniques génétiques de la levure	55
3.1. Obtention de cellules diploïdes	55
3.2. Obtention de cellules haploïdes	55
3.3. Détermination du signe sexuel	55
3.4. Analyse des spores	55
3.5. Détermination des phénotypes	55

4. Techniques de transformation par électroporation	56
4.1. Électroporation d' <i>E.coli</i>	56
4.2. Électroporation de <i>Saccharomyces cerevisiae</i>	56
5. Préparation d'ADN	56
5.1. ADN de levure	57
5.1.1. Extraction d'ADN génomique de levure	57
5.1.2. Préparation de chromosomes entiers de levure	58
5.2. ADN plasmidique d' <i>Escherichia coli</i>	58
6. Analyse de l'ADN	59
6.1. Purification de l'ADN par extraction phénol-chloroforme	59
6.2. Précipitation alcoolique	59
6.3. Purification de l'ADN par chromatographie	60
6.4. Détermination de la concentration d'ADN	60
6.5. Analyse par électrophorèse sur gel d'agarose	60
6.5.1. Électrophorèse sur gel d'agarose classique	60
6.5.2. Électrophorèse en champs alternés (ou PFGE)	61
6.6. Digestion par des endonucléases de restriction	61
6.7. Ligation des molécules d'ADN	62
7. Amplification d'un fragment d'ADN par PCR (Polymerase Chain Reaction)	62
7.1. PCR classique	62
7.1.1. Principe	62
7.1.2. Paramètres importants lors des différentes étapes	62
7.1.3. Matériel	63
7.2. Tail - PCR	63
8. Détermination de la séquence d'un fragment d'ADN	66
9. Hybridation ADN/ADN	67
9.1. Transfert	67
9.2. Préparation d'une sonde froide	67
9.3. Hybridation et lavages	68
9.4. Révélation au NBT/BCIP	68
10. Hybridation génomique comparative sur puce à ADN	68
10.1. Puce à ADN utilisée	69
10.2. Extraction de l'ADN génomique	69

10.3. Marquage de l'ADN par amorçage aléatoire	69
10.4. Hybridation et lavages	70
10.5. Lavage et coloration	70
10.6. Scanning et analyse des données	70
11. Informatique	71

CHAPITRE 1 : SYSTÈME DE SÉLECTION ET CARACTÉRISATION MOLÉCULAIRE DES DUPLICATIONS

Résumé	73
Introduction	75
1. Le système de sélection	75
2. Caractérisation moléculaire des duplications	79
2.1. Analyse du profil de restriction <i>Bam</i> HI au locus <i>URA2</i>	79
2.2. Analyse génétique	80
3. Localisation chromosomique et estimation de la taille de l'élément dupliqué	
3.1. Localisation chromosomique des régions dupliquées	82
3.2. Estimation de la taille des régions dupliquées	84
3.2.1. Événements de duplication interchromosomique	85
3.2.2. Événements de duplication intrachromosomique	85
3.2.2.1. Marche sur le fragment de restriction <i>Bam</i> HI du locus <i>URA2</i>	
3.2.2.2. Marche sur les fragments de restriction <i>Bam</i> HI suivants	
Conclusion	90

CHAPITRE 2 : DUPLICATIONS GÉNIQUES PAR RÉTROPOSITION EN PHASE HAPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

Résumé	91
Introduction	92
Résultats	97
1. Caractérisation des sites d'insertions de la région dupliquée	97

1.1. Détermination du site d'insertion en amont de la duplication	97
1.1.1. Stratégie d'insertion-excision de plasmide	97
1.1.2. Analyse des séquences à la jonction	101
1.2. Détermination du site d'insertion en aval de la duplication	104
2. La surexpression d'un rétrotransposon Ty1 augmente la fréquence des événements de duplication de la région codant l'ATCase	104
2.1. Sélection des révertants	105
2.2. Analyse des révertants porteurs d'un événement de duplication	105
2.2.1. Analyse du profil de restriction <i>Bam</i> HI au locus <i>URA2</i>	106
2.2.2. Localisation chromosomique de la région dupliquée	106
2.2.3. Détermination de la taille de la région dupliquée	108
2.2.4. Détermination des sites d'insertion de la région dupliquée	108
2.2.4.1. Site d'insertion en amont de la duplication	109
2.2.4.2. Site d'insertion en aval de la duplication	111
Discussion	112

CHAPITRE 3 : DUPLICATIONS SEGMENTALES EN PHASE HAPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

Résumé	117
Introduction	118
Résultats	120
1. Détermination de la taille des régions dupliquées	120
2. Analyse de la séquence des sites d'insertions de la région dupliquée	125
2.1. Détermination du site d'insertion en amont de la duplication	125
2.2. Détermination du site d'insertion en aval du segment dupliqué	128
3. Influence de la recombinaison homologue sur les événements de duplication segmentale et de duplication génique	131
3.1. Sélection	131
3.1. Caractérisation moléculaire	132
3.2.1. Localisation chromosomique de la région dupliquée	132
3.2.2. Détermination de la taille de la région dupliquée	132
Discussion	135

CHAPITRE 4 : ANEUPLOÏDIES ET TRANSLOCATIONS NON-RÉCIPROQUES À L'ÉTAT DIPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

Résumé	141
Introduction	142
Résultats	144
1. Caractérisation moléculaire et génétique des événements de duplication	144
1.1. Analyse du profil de restriction <i>Bam</i> HI au locus <i>URA2</i>	144
1.2. Analyse génétique	144
2. Localisation chromosomique de la région dupliquée	146
3. Détermination de la taille de la région dupliquée	146
3.1. Événements de duplication interchromosomique	148
3.2. Événements de duplication intrachromosomique	148
4. Analyse de la séquence des sites d'insertion en amont de la région dupliquée	150
5. Analyse globale du nombre de copies de gènes par hybridation génomique comparative par microarray	154
6. Détermination du site d'insertion en aval du segment dupliqué pour le révertant B286	160
Discussion	161

CHAPITRE 5 : LES PARALOGUES *URA5* ET *URA10* : ÉVOLUTION DE DEUX GÈNES ISSUS DE LA DUPLICATION COMPLÈTE DU GÉNOME DE *SACCHAROMYCES CEREVISIAE*

Résumé	167
Introduction	168
1. Les gènes <i>URA5</i> (YML106w) et <i>URA10</i> (YMR271c) de <i>Saccharomyces cerevisiae</i>	172
1.1. Fonction et régulation	172
1.2. <i>URA5</i> et <i>URA10</i> sont éléments du bloc de duplication 36	174
2. Les gènes <i>URA5</i> et <i>URA10</i> chez les autres levures hémiascomycètes séquencées	175

2.1. Recherche des orthologues des gènes <i>URA5</i> et <i>URA10</i>	175
2.2. Synténie au niveau des régions contenant <i>URA5</i> et <i>URA10</i>	176
2.3. Évolution des deux copies de gène	183
Conclusion	185
CONCLUSIONS GÉNÉRALES ET PERSPECTIVES	187
RÉFÉRENCES BIBLIOGRAPHIQUES	195

LISTES DES FIGURES

INTRODUCTION

<i>Figure I.1</i>	Conservation de synténie entre le génome de la souris et de l'homme	18
<i>Figure I.2</i>	Relations de parentés entre gènes dupliqués chez deux espèces	22
<i>Figure I.3</i>	Identification des segments dupliqués	23
<i>Figure I.4</i>	Événements moléculaires susceptibles de produire des duplications	31
<i>Figure I.5</i>	Devenir des séquences dupliquées	34
<i>Figure I.6</i>	Gènes de l'hémoglobines présents chez l'homme	40
<i>Figure I.7</i>	Levures hémiascomycètes partiellement ou entièrement séquencées	41
<i>Figure I.8</i>	Type de duplications observées chez <i>Saccharomyces cerevisiae</i>	44
<i>Figure I.9</i>	Unité répétée d'ADN ribosomique chez <i>Saccharomyces cerevisiae</i>	44
<i>Figure I.10</i>	Modèle de la duplication totale du génome	47

MATÉRIEL ET MÉTHODES

<i>Figure M.1</i>	Cycle cellulaire de <i>Saccharomyces cerevisiae</i>	52
<i>Figure M.2</i>	Diagramme schématique de la technique de Tail-PCR	64

CHAPITRE 1

<i>Figure 1.1</i>	Le gène <i>URA2</i> et les étapes de la voie de biosynthèse de novo des nucléotides pyrimidiques	76
-------------------	--	----

<i>Figure 1.2</i>	Allèle mutant <i>ura2 15-30-72</i> et types d'événements identifiés pouvant mener à la réactivation fonctionnelle du domaine ATCase	78
<i>Figure 1.3</i>	Exemple de profil de restriction <i>Bam</i> HI dans le cas d'un événement de duplication	80
<i>Figure 1.4</i>	Ségrégation observée dans le croisement d'un révertant par la souche <i>ura2 15-30-72</i>	81
<i>Figure 1.5</i>	Analyse du caryotype des révertants haploïdes porteurs d'un événement de duplication	83
<i>Figure 1.6</i>	Marche par hybridation de type Southern directement sur les chromosomes séparés par électrophorèse en champs alternés	85
<i>Figure 1.7</i>	Stratégie de marche sur le fragment de restriction <i>Bam</i> HI du locus <i>URA2</i>	87
<i>Figure 1.8</i>	Stratégie de marche sur les fragments de restriction <i>Bam</i> HI contigus au locus <i>URA2</i>	89

CHAPITRE 2

<i>Figure 2.1</i>	Cycle réplicatif des éléments Ty	95
<i>Figure 2.2</i>	Détermination de la région de départ de la partie du gène <i>URA2</i> dupliquée	98
<i>Figure 2.3</i>	Stratégie de clonage de l'extrémité 5' de la région dupliquée	100
<i>Figure 2.4</i>	Site d'insertion chromosomique en 5' et 3' des duplications géniques présentes chez 4 révertants spontanés	102
<i>Figure 2.5</i>	Présence de microhomologie à la jonction 5' des duplications géniques.	101
<i>Figure 2.6</i>	Région située en amont des séquences Ty1 chez les révertants Rev 25 et Rev 56	103
<i>Figure 2.7</i>	Exemples de caryotypes après électrophorèse en champs alternés de quelques révertants sélectionnés à partir de la souche FRJ1	107
<i>Figure 2.8</i>	Site d'insertion chromosomique en 5' des duplications géniques présentes chez les 23 révertants isolés à partir de la souche FRJ1.	109

<i>Figure 2.9</i>	Représentation schématique des sites d'insertion 5'et 3' des duplications géniques présentes chez les 23 révertants isolés à partir de la souche FRJ1	110
<i>Figure 2.10</i>	Duplication génique via un mécanisme de rétroposition.	112

CHAPITRE 3

<i>Figure 3.1</i>	Identification des segments dupliqués chez les 4 révertants spontanés	122
<i>Figure 3.2</i>	Stratégie de marche sur le chromosome par analyse des profils de restriction	123
<i>Figure 3.3</i>	Corrélation entre la migration du chromosome X et la taille des segments dupliqués	124
<i>Figure 3.4</i>	Site d'insertion chromosomique en 5'des segments dupliqués	127
<i>Figure 3.5</i>	Site d'insertion chromosomique en 3'des segments dupliqués	128
<i>Figure 3.6</i>	Duplications segmentales en tandem direct	130
<i>Figure 3.7</i>	Caryotypes après électrophorèse en champs alternés des révertants sélectionnés à partir de la souche <i>ura2 15-30-72 rad52 :: kanMX4</i>	132
<i>Figure 3.8</i>	Identification des segments dupliqués chez les révertants <i>rad52</i>	134
<i>Figure 3.9</i>	Modèle de formation du segment dupliqué en tandem direct par crossing-over inégal	136
<i>Figure 3.10</i>	Modèle de formation du segment dupliqué en tandem direct	138

CHAPITRE 4

<i>Figure 4.1</i>	Analyse génétique des révertants diploïdes porteurs d'une duplication	145
<i>Figure 4.2</i>	Exemples de caryotypes après électrophorèse en champs alternés de quelques révertants sélectionnés à partir de la souche diploïde <i>ura2 15-30-72</i>	147
<i>Figure 4.3</i>	Stratégie de marche sur le chromosome par hybridation de type Southern	149

<i>Figure 4.4</i>	Site d'insertion chromosomique en 5'des segments dupliqués	153
<i>Figure 4.5</i>	Caractérisation des gènes dupliqués du chromosome X chez les révertants diploïdes	155
<i>Figure 4.6</i>	Duplication d'une région localisée sur un autre chromosome que le chromosome X	157
<i>Figure 4.7</i>	Délétions présentes chez les révertants B025 et B132	158
<i>Figure 4.8</i>	Remaniements présents chez les révertants diploïdes	159
<i>Figure 4.9</i>	Duplication segmentale en tandem direct chez le révertant B286	160
<i>Figure 4.10</i>	Modèle de réparation d'une cassure double brin par un mécanismes de type BIR	165

CHAPITRE 5

<i>Figure 5.1</i>	Arbre phylogénétique des levures hémiascomycètes	170
<i>Figure 5.2</i>	Voie de biosynthèse de novo des pyrimidines	173
<i>Figure 5.3</i>	Localisation chromosomique des gènes paralogues <i>URA5</i> et <i>URA10</i>	175
<i>Figure 5.4</i>	Synténie au niveau de la région dupliquée contenant le gène <i>URA5</i> chez les <i>Saccharomyces</i>	178
<i>Figure 5.5</i>	Synténie au niveau de la région dupliquée contenant le gène <i>URA10</i> chez les <i>Saccharomyces</i>	179
<i>Figure 5.6</i>	Synténie au niveau de la région dupliquée contenant le gène <i>URA5</i> à partir des espèces du genre <i>Kluyveromyces</i>	181
<i>Figure 5.7</i>	Régions orthologues chez <i>S. cerevisiae</i> et <i>K. waltii</i>	182

CONCLUSION

<i>Figure C.1</i>	Stratégie d'étude du mécanisme de rétroposition	189
-------------------	---	-----

LISTE DES TABLEAUX

INTRODUCTION

<i>Tableau 1.1</i>	Proportion de gène dupliqués au sein de certains génomes séquencés	24
--------------------	--	----

MATÉRIEL ET MÉTHODES

<i>Tableau M.1</i>	Paramètres d'amplification utilisés pour la technique de Tail-PCR	65
--------------------	---	----

CHAPITRE 1

<i>Tableau 1.1</i>	Localisation chromosomique des régions dupliquées présentes chez les 8 révertants spontanés	83
--------------------	---	----

CHAPITRE 2

<i>Tableau 2.1</i>	Distribution des éléments Ty dans le génome de <i>Saccharomyces cerevisiae</i>	94
<i>Tableau 2.2</i>	Révertants sélectionnés à partir de la souche <i>ura2 15-30-72</i> dans laquelle la transposition a été augmentée	106
<i>Tableau 2.3</i>	Localisation chromosomique des régions dupliquées présentes chez les 23 révertants isolés à partir de la souche FRJ1	107

CHAPITRE 3

<i>Tableau 3.1</i>	Taille des segments dupliqués présents chez les révertants spontanés	122
<i>Tableau 3.2</i>	Taille des segments dupliqués présents chez les révertants <i>rad52</i>	134

CHAPITRE 4

<i>Tableau 4.1</i>	Localisation chromosomique des régions dupliquées présentes chez les 8 révertants diploïdes	147
<i>Tableau 4.2</i>	Caractéristiques des régions dupliquées présentes chez les 14 révertants diploïdes	151
<i>Tableau 4.3</i>	Estimation de la taille des chromosomes porteurs de la séquence ATCase dupliquée	156

CHAPITRE 5

<i>Tableau 5.1</i>	Nombre d'homologues aux gènes <i>URA5</i> et <i>URA10</i> de <i>S. cerevisiae</i> dans le génome des autres levures hémiascomycètes séquencées	175
<i>Tableau 5.2</i>	Pourcentages d'identité de séquence protéique des homologues par rapport aux séquences protéiques Ura5p et Ura10p de <i>S. cerevisiae</i>	184

INTRODUCTION

Dans les années 1950, Barbara Mc Clintock observa un phénomène inhabituel lors d'études génétiques sur des plants de maïs (*Zea mays*) : l'existence de « gènes sauteurs », appelés éléments génétiques mobiles ou transposables (Mc Clintock, 1956). Néanmoins, à l'époque, ses travaux ne suscitèrent qu'un intérêt mineur. L'importance de ses découvertes fut reconnue seulement trente ans plus tard lors de la remise de son prix Nobel en 1983. De nos jours, nous savons que le génome n'est pas un ensemble rigide et statique, mais est plutôt dynamique et flexible. En effet, une stabilité totale signifierait un état statique qui s'opposerait à l'apparition de nouvelles formes de vie en réponse aux modifications extérieures. En conséquence, le génome est sujet à des variations, la vie nécessitant un équilibre entre ancienneté et nouveauté.

Les gènes et les génomes actuels sont le résultat de l'accumulation d'événements qui se sont produits dans le passé. Les génomes sont notamment le siège de grands remaniements chromosomiques telles les délétions, les inversions, les translocations qui peuvent être conservés au cours de l'évolution. Ils peuvent aussi être à l'origine de la mort cellulaire mais également la cause de maladies génétiques notamment dans les populations humaines.

Les événements de délétion, translocation, inversion sont importants quantitativement lorsque l'on compare différentes espèces. Cependant le rôle de ce type de réarrangements dans le processus évolutif n'est pas bien défini. De nombreuses espèces apparentées possèdent des génomes qui diffèrent par un certain nombre de réarrangements chromosomiques. La comparaison des génomes de l'homme et de la souris permet d'illustrer cette dynamique chromosomique. Le caryotype humain comporte 23 paires de chromosomes et celui de la souris seulement 20. La cartographie assez avancée de la souris et de l'homme a montré que de nombreux blocs de gènes étroitement liés chez l'homme se retrouvent également associés chez la souris. De tels blocs de gènes sont

qualifiés de synténiques. Une carte de l'ensemble de ces blocs a été établie (Fig. I.1). Il semble que 50 à 100 événements distincts d'inversions et de translocations ont été nécessaires pour mélanger et échanger ces blocs afin d'aboutir aux génomes actuels de l'homme et de la souris (Gibbs *et al.*, 2004). Il est également intéressant à noter que la conservation du chromosome X est quasi-parfaite (Grutzner *et al.*, 2002), contrairement au chromosome Y qui a été le siège de nombreux remaniements (Skaletsky *et al.*, 2003).

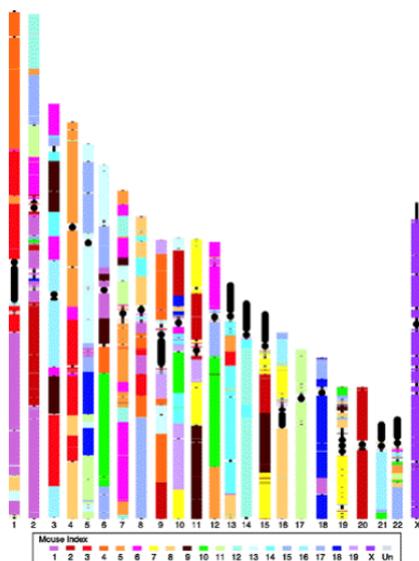


FIGURE I.1 – Conservation de synténie entre le génome de la souris et de l'homme (d'après Eichler & Sankoff, 2003).

Comme les réarrangements précédents, les événements de duplication de séquences d'ADN contribuent très largement à la plasticité des génomes. Cependant, il est plus facile de comprendre leur rôle dans le processus évolutif et de les considérer ainsi comme un des moteurs essentiels de l'évolution. En effet, ce type de remaniement engendre une copie surnuméraire d'un gène. L'une des régions dédoublées par une duplication est libre d'accumuler des mutations ponctuelles tandis que l'autre copie

prend en charge la fonction essentielle de la région considérée. Dès lors, les gènes en question peuvent se mettre à diverger fonctionnellement, ce qui peut présenter un avantage potentiel pour l'évolution des génomes. En effet, dans bien des cas où l'on a pu comparer différents produits de gènes assumant des fonctions voisines mais spécifiques comme par exemple les globines, il existe de sérieux indices montrant que ces gènes sont issus d'événements de duplication.

Une des thématiques abordées au Laboratoire de Microbiologie et Génétique est la dynamique des génomes et plus particulièrement l'étude des mécanismes impliqués dans la formation de remaniements chromosomiques de grande taille. Les mécanismes impliqués dans les événements de duplication étant encore peu connus à ce jour, nous avons donc entrepris l'étude de ces réarrangements génomiques dans un organisme modèle, la levure *Saccharomyces cerevisiae*.

1. PRÉAMBULE HISTORIQUE

L'étude des événements de duplication est une préoccupation ancienne. En 1918, C. Bridges est le premier à mentionner, lors d'un congrès, la possibilité de l'existence de gènes dupliqués. Il évoque déjà à cette époque l'hypothèse d'une implication de ces gènes dans le processus évolutif. D'après lui, ces gènes identiques peuvent évoluer de manière distincte et diversifier ainsi leur effet. Par la suite, il apportera une des premières observations de duplication, celle du gène *Bar* chez un mutant de *Drosophila melanogaster*, présentant une réduction importante de la taille de l'œil (Bridges, 1936). L'étude cytologique des chromosomes polytènes a mis en évidence un dédoublement de bande sur le chromosome X correspondant à la duplication en tandem du gène *Bar*. Le rôle potentiel des duplications dans le processus évolutif a alors été mis à jour et de nombreuses hypothèses sur le devenir des éléments dupliqués ont été proposées (Stephens, 1951).

En 1959, Markert a montré l'existence de formes multiples d'une même enzyme, les isozymes, dont l'étude a révélé, entre autres, l'existence de familles de gènes. L'analyse des protéines a précédé, ici, l'identification des gènes qui les codent du fait des techniques disponibles à l'époque. À partir d'exemples de séquences hautement répétées comme les gènes codant les ARN ribosomiques, les histones ou encore l'hémoglobine de

mammifères, Susumu Ohno a réellement mis en avant l'importance du rôle des duplications à travers l'évolution dans sa monographie "Evolution by gene duplication" (1970). Il reformule l'hypothèse selon laquelle la duplication de gènes permet l'émergence de nouvelles fonctions. Il a ainsi proposé que la duplication génique était le principal mécanisme d'évolution : une des deux copies d'un gène, en absence de pression de sélection, va pouvoir accumuler des mutations aléatoirement et acquérir ainsi une nouvelle fonction.

Bien que la duplication soit alors considérée comme le moyen le plus simple de produire de la diversité génétique, l'observation de l'apparition d'une duplication est très rare. Chez *Saccharomyces cerevisiae*, il faut attendre 1977 pour voir la première étude portant sur de tels remaniements. Hansche et Beres (1977) décrivent la duplication du gène *ACPI* codant l'acide phosphatase. La duplication du gène est assortie d'un événement de translocation. Par la suite, d'autres duplications de gènes ont pu être sélectionnées chez *S. cerevisiae* : *HIS4* (Greer & Fink, 1979), *ADH2* (Paquin *et al.*, 1992) et *ADH4* (Dorsey *et al.*, 1992). Ces événements se produisent à des fréquences très faibles de l'ordre de 10^{-10} à 10^{-12} duplication / cellule / génération et il est nécessaire pour les sélectionner d'utiliser un crible de sélection positif.

Cependant, ce n'est qu'à la fin des années 90, après la détermination et l'analyse de la séquence complète de certains génomes dont celui de la levure *Saccharomyces cerevisiae* (Goffeau *et al.*, 1996), que l'importance des événements de duplication est clairement démontrée et reconnue. L'analyse des séquences complètes permet d'identifier l'ensemble des gènes dupliqués qui n'auraient pu l'être avec les différentes approches expérimentales jusqu' alors utilisées. De plus, l'importance quantitative de ces séquences s'avère être un dénominateur commun des eucaryotes, archaebactéries et procaryotes.

2. ANALYSE DES RÉGIONS DUPLIQUÉES PRÉSENTES CHEZ LES ORGANISMES SÉQUENCÉS

Durant ces 10 dernières années, les séquences complètes d'un certain nombre d'organismes ont permis d'établir l'architecture de leur génome. Dans le cadre de l'étude des duplications de gène, l'un des nombreux intérêts de ces programmes de séquençage a été de pouvoir comparer l'ensemble des ORFs présentes au sein d'un organisme (comparaison intraspécifique) afin de caractériser leur redondance. Cette analyse a permis de

donner une vision globale sur la nature des séquences d'ADN dupliquées c'est-à-dire sur leur proportion, leur taille, leur localisation... Il a ainsi été possible de quantifier la proportion de gènes dupliqués présents chez ces organismes. Par ailleurs, une analyse de comparaison des ORFs, notamment dupliquées, entre les différents génomes séquencés (comparaison interspécifique) a été possible et a permis d'apporter certains éléments de réponses sur le plan évolutif.

Ainsi, l'analyse des génomes séquencés a permis la mise en évidence de duplications de gènes, mais également de duplications de segments chromosomiques portant une série de gènes contigus. Dans un premier temps, nous allons présenter les stratégies d'identification des duplications de gènes et segments au sein des génomes séquencés.

2.1. Identification de gènes dupliqués

Le séquençage d'un organisme est suivi par une phase d'annotation, c'est-à-dire d'identification des ORFs présentes dans son génome. Cette étape est indispensable afin de pouvoir comparer les ORFs les unes aux autres. La recherche de similarité entre séquences peut être réalisée par des méthodes optimales d'alignement global (Needleman and Wunsch, 1970), local (Smith *et al.*, 1981), ou encore à l'aide de méthodes heuristiques comme le font les programmes FASTA (Pearson and Lipman, 1988) et BLAST (Altschul *et al.*, 1990). Ces algorithmes permettent de rechercher des éléments similaires à une séquence requête dans une banque de données. La ou les séquences de la banque qui présentent des éléments similaires avec les séquences requêtes sont alors conservées. Ces algorithmes attribuent un score à la comparaison effectuée. Si les séquences de deux gènes présentent une similarité élevée, on pourra en déduire que ces deux séquences descendent probablement d'une même séquence ancestrale, et qu'elles sont donc homologues. La fixation du seuil au-dessus duquel la similarité observée entre deux séquences est suffisamment forte pour que l'homologie puisse être proposée de manière convaincante est une des étapes essentielles de la comparaison de séquence. Une validation manuelle est souvent nécessaire.

Par ailleurs, si deux séquences homologues appartiennent aux génomes de deux espèces différentes et sont donc issues d'une séquence présente chez un ancêtre commun, ces séquences seront qualifiées d'orthologues (Fig. I.2). Par contre, si les deux séquences homologues appartiennent au même génome et sont donc issues d'un événement de duplication, elles seront qualifiées de paralogues (Fig. I.2).

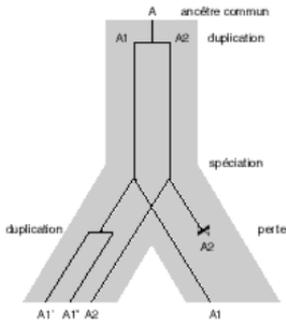


FIGURE I.2 – Relations de parentés entre gènes dupliqués chez deux espèces 1 et 2. Dans l'espèce 1, les gènes A1', A1'' et A2 sont paralogues entre eux. De plus, A1' et A1'' sont des orthologues de A1 de l'espèce 2.

2.2. Identification de segments dupliqués

La cartographie de tous les gènes homologues au sein d'un ou de plusieurs génomes permet d'identifier des séries de gènes qui ont conservé le même ordre et la même orientation (Fig. I.3). Ces deux régions sont alors appelées blocs dupliqués si elles appartiennent au même génome ou de synténie si elles sont dans deux génomes différents. Le principe de parcimonie nous indique que, très vraisemblablement, de tels blocs contenant des séries de gènes homologues sont eux-mêmes homologues. Les matrices de similarité (ou dot plot) permettent de comparer chaque chromosome avec les autres et de faire éventuellement apparaître de tels blocs homologues. Ces matrices peuvent être construites soit à partir des séquences brutes, soit à partir des relations d'homologie déjà caractérisées.

Comme illustré figure I.3, pour chaque position de la séquence du Chr. A une fenêtre coulissante est déplacée sur le Chr. B, et à chaque déplacement un score est calculé en fonction de la similarité entre les deux séquences. Si le score pour cette position sur B dépasse un certain seuil, alors un point est indiqué à l'intersection avec la position de A. Une suite de points sur une diagonale indique une région de similarité entre les deux séquences. Sur la figure I.3, on voit que les ORFs 1, 3 et 4 du chromosome A sont homologues et synténiques aux ORFs 1, 2 et 3 du chromosome B. Ces trois ORFs définissent donc un bloc de synténie entre les chromosomes A et B.

Si les chromosomes A et B appartiennent au même génome, ces deux blocs résultent probablement d'un événement de duplication ancestral et sont donc paralogues. Par contre, si les chromosomes se trouvent dans deux génomes différents, les deux blocs peuvent être orthologues et descendre de la même séquence présente dans l'ancêtre commun.

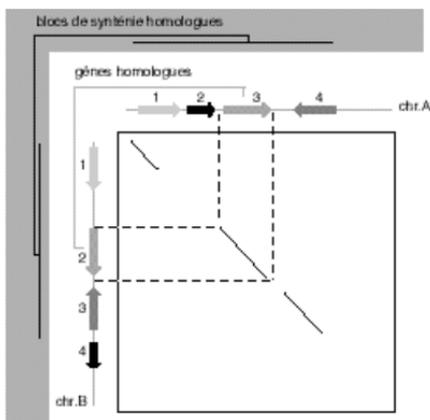


FIGURE 1.3 – Identification des segments dupliés. Les flèches grises représentent les ORFs homologues et synténiques entre les chromosomes X et Y. Par contre, les flèches en noir représentent les ORFs sans homologues détectés dans la région observée.

2.3. Proportion de gènes au sein des génomes séquencés

L'analyse des séquences montre que tous les organismes étudiés à ce jour présentent une proportion importante de duplications. Le nombre de gènes dupliqués a été estimé dans des organismes représentatifs des bactéries, des archaeobactéries et des eucaryotes (Tab. I.1). Les résultats montrent que dans les trois règnes du vivant, une proportion importante de gènes est issue d'un événement de duplication. Schématiquement, on peut dire que la fraction de gènes dupliqués présents chez un organisme correspond en moyenne à 1/3 de son génome. Cependant, il y a des exceptions comme par exemple chez *Arabidopsis thaliana* où cette fraction dépasse les 2/3 de son génome. De plus, il est très vraisemblable que ces proportions soient sous-estimées car la divergence entre deux séquences dupliquées peut être telle qu'aucune similarité ne soit trouvée.

Lynch et Conery (2000) ont estimé que la duplication d'un gène se produit et se fixe dans un génome à une fréquence d'une duplication par gène tous les 100 millions d'années chez les eucaryotes tels *H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, *A. thaliana* et *S. cerevisiae*. Cette fréquence est comparable à celle observée pour la substitution de nucléotides qui est de 0,1-0,5 substitution par site tous les 100 millions d'années au niveau des génomes nucléaires de vertébrés (Li, 1997).

	Nombre total de gène	% de gènes dupliqués	Références
Bactérie			
<i>Mycoplasma pneumoniae</i>	677	44	Himmreich et al., 1996
<i>Helicobacter pylori</i>	1590	17	Tomb et al., 1997
<i>Haemophilus influenzae</i>	1709	17	Rubin et al., 2000
Archéobactérie			
<i>Archaeoglobus fulgidus</i>	2436	30	Klenk et al., 1997
Eucaryotes			
<i>Saccharomyces cerevisiae</i>	6241	30	Rubin et al., 2000
<i>Caenorhabditis elegans</i>	18 424	49	Rubin et al., 2000
<i>Drosophila melanogaster</i>	13 601	41	Rubin et al., 2000
<i>Arabidopsis thaliana</i>	25 498	65	The Arabidopsis Genome Initiative, 2000
<i>Homo sapiens</i>	environ 30 000	38	Li et al., 2001

TABLEAU 1 – Proportion de gène dupliqués au sein de certains génomes séquencés.

3. COMMENT UNE SÉQUENCE PEUT-ELLE SE DUPLIQUER ?

Les possibilités par lesquelles des séquences d'ADN au sein d'un génome peuvent se dupliquer sont multiples. L'apparition de gènes paralogues peut se faire par la duplication d'un seul gène, d'un segment composé de plusieurs gènes contigus, d'un chromosome entier ou encore d'un génome entier.

3.1. Duplication génique

Les événements de duplication concernant un seul gène sont qualifiés de duplication génique. Ces éléments dupliqués peuvent être situés l'un à côté de l'autre, ce que l'on appelle une duplication en tandem ou bien ils peuvent être dispersés au sein

du génome. Ces paralogues forment des familles de gènes dont la taille a été déterminée pour certains génomes entièrement séquencés (Gu *et al.*, 2002). La distribution des tailles varie en fonction des espèces mais également en fonction des familles de gènes (Lespinet *et al.*, 2002). Par exemple, la plus grande famille chez *D. melanogaster* est représentée par celle des trypsines avec 111 membres alors que chez les mammifères, elle est représentée par la famille des récepteurs olfactifs avec environ 1000 membres (Mombaerts, 2001 ; Zhang and Firestein, 2002). Chez *Saccharomyces cerevisiae*, il existe 722 familles dont 457 comprennent deux gènes. Les 265 familles restantes peuvent être composées de 3 gènes (124 familles) à plus de 20 gènes (11 familles) (Blandin *et al.*, 2000).

Pour les familles de gènes dont les membres sont dispersés dans le génome, le nombre de membres est très variable. Par exemple, l'ovalbumine est codée par une famille de 3 gènes alors que les histones sont codés par une famille contenant de 100 à 1000 gènes. Généralement, ces séquences d'ADN peuvent évoluer et donner naissance à des pseudogènes ou à des gènes codant des protéines ayant des fonctions différentes.

D'autres gènes ont évolué en familles répétées en tandem. L'organisateur nucléolaire en est un très bon exemple. Il abrite les répétitions en tandem des gènes d'ARN ribosomique. Chez la drosophile, les organisateurs nucléolaires présents sur les chromosomes X et Y contiennent respectivement 250 et 150 copies en tandem de gènes d'ARNr. Celui de l'homme en compte 250 copies. Les gènes codant les histones peuvent également être organisés en répétition en tandem chez certaines espèces. Pour ces répétitions en tandem, les copies multiples sont identiques car leur présence semble plutôt liée à la nécessité pour la cellule d'avoir une grande quantité du produit de gène correspondant et non pas à l'apport d'une fonction supplémentaire.

3.2. Duplication de segment

Avant l'ère du séquençage systématique de génomes, il était très difficile de prendre en compte l'environnement chromosomique d'un gène dupliqué. Comme nous l'avons vu précédemment, un gène peut être dupliqué en même temps que les gènes qui lui sont adjacents conduisant à une duplication de segment. Ces événements sont importants car ils constituent comme toute duplication un médiateur de l'évolution des gènes, mais également de l'évolution des génomes.

Pour l'évolution des gènes, les segments dupliqués constituent un réservoir plus important de matériel permettant la spécialisation ou l'apparition de nouvelles fonctions qu'une duplication génique. Ces nouveaux gènes codent des protéines qui ne sont pas essentielles pour la viabilité des organismes mais peuvent être, par exemple, importants pour l'adaptation des espèces à certaines niches écologiques (Duda and Palumbi, 1999). Chez les mammifères, un certain nombre de gènes dupliqués sont impliqués dans l'évolution récente du système immunitaire, de la détoxification et de la reproduction (Copley *et al.*, 2003).

Pour l'évolution des génomes, les duplications segmentales sont des régions très similaires et par conséquent sont des sites préférentiels pour les réarrangements chromosomiques. De récentes études comparatives entre des espèces eucaryotiques très proche (comme l'homme et la souris ou les levures hémiascomycètes) ont montré que des séquences répétées très similaires sont présentes aux jonctions de grands remaniements chromosomiques (Dehal *et al.*, 2001; Kellis *et al.*, 2003).

À l'heure actuelle, il est très clair qu'une proportion importante des gènes dupliqués est issue d'un événement de duplication segmentale chez les organismes séquencés. Une des surprises lors de l'analyse du génome humain (International Human Genome Sequencing Consortium, 2001) a été l'abondance de grands blocs d'ADN dupliqués avec une similarité de séquence très élevée (>90 %). Plusieurs méthodes informatiques et expérimentales (Bailey *et al.* 2001, 2002) ont permis d'estimer à 5-6% la part du génome humain qui est dupliquée. En comparaison avec d'autres organismes comme la drosophile ou le ver, le génome humain est enrichi en duplications segmentales récentes (Bailey *et al.*, 2002). La proportion de duplications segmentales récentes varie considérablement entre les organismes eucaryotes séquencés. Chez le rat et la souris, la proportion correspond respectivement à 2,92 % et 1-2% du génome (Tuzun *et al.*, 2004; Bailey *et al.*, 2004).

Les blocs dupliqués présents dans le génome humain partagent une similarité de séquence importante (> 90 %) et ont une taille allant de quelques kb à plus de 100kb. Contrairement aux duplications en tandem, ils sont dispersés à travers le génome humain. Deux catégories de duplications segmentales peuvent être distinguées : les duplications intrachromosomiques et interchromosomiques. Les duplications intrachromosomiques sont constituées de blocs dispersés au niveau des régions euchromatiques sur un même chromosome. Par contre, les duplications interchromosomiques font

référence aux duplications réparties sur deux chromosomes non homologues. Chez l'homme, ces dernières sont principalement en position centromérique ou subtélomérique. Une telle architecture génomique n'avait jamais été observée chez les invertébrés séquencés.

Les duplications segmentales ont un impact très important sur la dynamique et l'évolution du génome humain. Les duplications intrachromosomiques et interchromosomiques augmentent les probabilités de réarrangements secondaires pouvant entraîner des inversions, des délétions ou encore d'autres duplications. Cette possibilité de réarrangements est due à la présence des grands blocs avec une similarité de séquence importante qui peuvent servir de substrat pour les événements de recombinaison homologue.

En 1998, Lupski propose le concept d'«anomalie génomique» humaine (genomic disorder) qui résulte d'un réarrangement chromosomique. Ce concept est à mettre en parallèle avec le concept d'«anomalie monogénique» résultant d'une mutation ponctuelle dans un gène suite à une erreur de réplication et/ou de réparation de l'ADN. Un grand nombre de ces anomalies génomiques sont associées à des duplications segmentales. En effet, ces anomalies résultent, la plupart du temps, de recombinaisons homologues ectopiques entre des segments dupliqués (Stankiewicz and Lupski, 2002). Plusieurs maladies humaines sont attribuées à ces remaniements chromosomiques dont le syndrome de Williams Beuren et de DiGeorge par exemple (Mazzarella and Schlessinger, 1998 ; Lupski, 1998; Emanuel and Shaikh, 2001).

Sur le plan évolutif, les réarrangements liés aux duplications segmentales pourraient également être impliqués dans la formation de barrières de spéciation. Ainsi, la comparaison de la séquence du chromosome 19 de l'homme et son homologue chez la souris révèle que des segments dupliqués sont présents aux jonctions des remaniements observés entre ces 2 espèces (Dehal *et al.*, 2001). De la même manière, la caractérisation d'une inversion péricentrique observée entre l'homme et le chimpanzé a mis en évidence la présence de duplications segmentales aux jonctions de ce remaniement (Nickerson and Nelson, 1998). Ces données suggèrent une corrélation entre les segments dupliqués et les sites de remaniements chromosomiques entre espèces. De plus, bien que la signification biologique de ces événements ne soit pas connue, ces réarrangements semblent participer à l'isolement des espèces d'un même genre comme cela a été suggéré pour les espèces du genre *Saccharomyces* (Samonte and Eichler, 2001; Fischer *et al.*, 2000).

3.3. Duplication de chromosome ou aneuploïdie

L'aneuploïdie correspond à une aberration chromosomique due à un nombre anormal de chromosomes. Dans la majorité des cas, l'aneuploïdie diffère du type sauvage par un chromosome en moins ou en plus. On parle alors respectivement de monosomie ($2n-1$) et de trisomie ($2n+1$). Pour une trisomie, le matériel génétique situé sur l'ensemble d'un chromosome est dupliqué.

Les végétaux sont en général plus tolérants à l'aneuploïdie que le règne animal. Les états aneuploïdes sont très étudiés chez l'homme comme par exemple le syndrome de Down (trisomie 21), le syndrome de Klinefelter (XXY) et le syndrome de Turner (X0). Le taux d'apparition spontanée d'aneuploïdie est élevé (1/650 naissances pour la trisomie 21) et produit une forte proportion des maladies d'origine génétique chez l'homme. Par ailleurs, la perte ou le gain de chromosomes sont également fréquents au niveau des cellules cancéreuses chez les mammifères (Lengauer *et al.*, 1998 ; Pihan and Doxsey, 2003). Chez *Drosophila melanogaster*, les seules aneuploïdes autosomiques qui survivent jusqu'à l'âge adulte correspondent aux trisomies et monosomies du chromosome 4, le plus petit des chromosomes. Par contre, l'aneuploïdie semble être très répandue chez *Saccharomyces cerevisiae* (Hughes *et al.*, 2000).

La cause de la plupart des états aneuploïdes est la non-disjonction des chromatides sœurs au cours de la méiose ou de la mitose. Cette non-disjonction correspond à un défaut du processus de ségrégation des chromosomes conduisant à l'aneuploïdie. Elle se produit spontanément.

3.4. Duplication complète du génome ou polyplôïdie

La manière la plus rapide d'augmenter le nombre de gènes est constituée par une duplication complète du génome. Cette duplication est qualifiée de polyplôïdisation. Chez les polyplôïdes, il faut distinguer les autopolyploïdes, composés de multiples jeux provenant d'une même espèce, des allopolyploïdes, comportant des jeux de chromosomes issus d'au moins deux espèces.

L'autopolyploïdie résulte d'un dysfonctionnement au cours des divisions cellulaires. Cet événement peut se produire par erreur durant la méiose conduisant à la formation de gamètes diploïdes plutôt qu'haploïdes. Si deux gamètes diploïdes fusionnent, une cellule tétraploïde contenant 4 copies de chaque chromosome va être formée.

La polyplôïdie est un processus très commun dans le monde végétal. Les autopolyploïdes sont souvent viables car chaque chromosome a forcément un homologue pouvant former des bivalents durant la méiose et conduire à l'obtention de gamètes permettant la reproduction. Mais généralement, le croisement avec l'organisme original qui a permis sa formation n'est pas fertile. En effet, le croisement, par exemple, d'un tétraploïde avec un diploïde va engendrer un triploïde qui sera stérile car un jeu de chromosomes n'aura pas de partenaire homologue au cours de la division méiotique.

La polyplôïdie est un mécanisme qui peut être à la source d'une spéciation. La génération de nouvelles espèces chez les plantes par autopolyploïdie a été observée, notamment par Hugo de Vries. Durant son travail sur les primevères, *Oenothera lamarckiana*, de Vries a isolé une forme tétraploïde de cette plante normalement diploïde qu'il a nommé *Oenothera gigas*.

Par contre, la polyplôïdie est un processus très peu commun chez les animaux. Il existe, cependant, des cas apparus naturellement comme par exemple chez les vers plats, les sangsues et les crevettes. Récemment, l'analyse comparative de génomes de levures a mis à jour l'existence d'une duplication totale de génome au cours du processus évolutif des levures hémiascomycètes comme nous allons le voir plus en détail par la suite.

4. MÉCANISMES À L'ORIGINE DES DUPLICATIONS GÉNIQUES ET SEGMENTALES

La plupart des duplications de matériel génétique par modification du nombre de chromosomes résulte clairement d'un dysfonctionnement au cours des divisions cellulaires, mitotiques ou méiotiques. Par contre, les mécanismes impliqués dans les duplications génique et segmentale sont plus controversés et aucun mécanisme n'a été clairement établi. La nature de l'événement qui est à l'origine de ce type de duplications n'est pas facile à établir, notamment lorsque la duplication est ancienne, car les indices de cette origine sont souvent effacés par l'accumulation de mutations ou d'autres réarrangements. Plusieurs mécanismes dont certains hypothétiques ont été proposés par Maeda et Smithies (1986) pour expliquer la formation des gènes dupliqués. Quatre types d'événements moléculaires sont susceptibles de conduire à la duplication de gène (Fig.I.4) :

- la recombinaison ectopique entre séquences répétées localisées de part et d'autre du gène dupliqué ;
- la rétrotranscription d'un ARN messager et l'insertion de l'ADNc dans le génome ;
- la sur-réplication d'une région par erreur durant le processus de réplication
- les cassures aléatoires non identiques sur deux chromosomes homologues suivies de leur réunion.

4.1. Duplications géniques

4.1.1. Duplication génique en tandem

Comme nous l'avons vu précédemment, certaines familles de gènes sont constituées par des régions répétées en tandem. La présence de telles régions dans les génomes est principalement expliquée par des événements de crossing-over inégaux, mitotiques ou méiotiques (Ohta, 1976 ; Smith, 1976). Ces événements consistent en une recombinaison ectopique entre des séquences répétées localisées de part et d'autre de la région dupliquée sur deux chromosomes ou deux chromatides sœurs (Fig. I.4). L'échange inégal d'information génétique est un mécanisme démontré comme étant à la base des expansions et contractions des séquences répétées en sens direct, comme c'est le cas des loci des gènes codant les ARNs ribosomiques (Petes, 1980 ; Szostak and Wu, 1980). En effet, chez *Saccharomyces cerevisiae*, on compte 100 à 200 copies répétées d'ADN ribosomique localisées sur le chromosome XII. Szostak et Wu (1980) ont mesuré la perte et le gain d'un marqueur inséré dans cette région chez une souche haploïde. Ceci a permis d'estimer la fréquence des événements de crossing-over inégal dans cette région durant la mitose qui est d'environ 10^{-2} par cellule et par génération. Le nombre de copies résulte donc d'un équilibre entre gain et perte de copies de gènes. Petes (1980) est arrivé au même résultat en méiose montrant que les échanges entre chromatides sœurs par crossing-over inégal étaient importants dans cette région.

4.1.2. Duplications géniques dispersées

Dans la plupart des génomes eucaryotes, on retrouve des séquences appelées pseudogènes processés qui correspondent à des copies ADNc de gènes cellulaires fonctionnels mais qui sont dépourvus d'introns et de promoteurs. Ils possèdent en

revanche une queue poly(A) et ils sont souvent tronqués en 5' (Mighell *et al.*, 2000 ; Weiner *et al.*, 1986). Ces caractéristiques suggèrent que les pseudogènes trouvent leur origine dans la transcription inverse d'ARNm cellulaires (Weiner *et al.*, 1986).

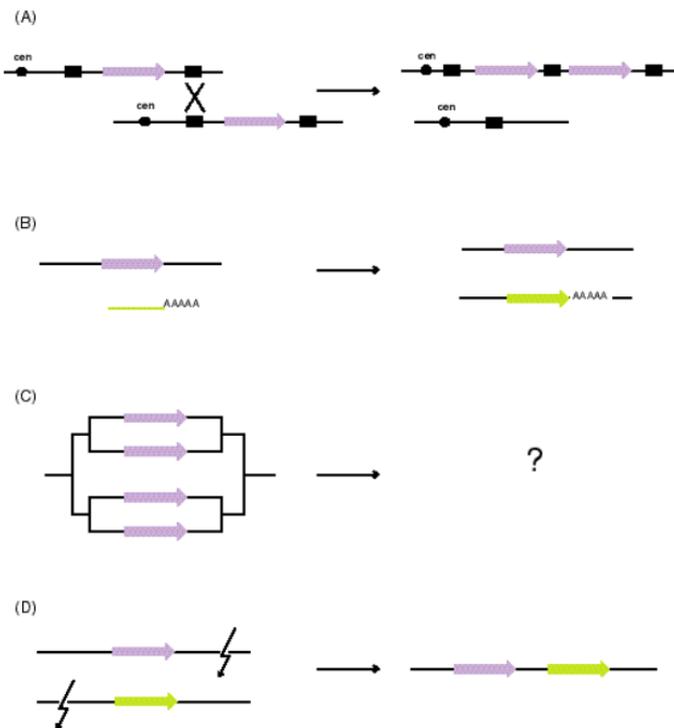


FIGURE 1.4 – Événements moléculaires susceptibles de produire des duplications. (d'après Maeda et Smithies, 1986).

A. Recombinaison ectopique entre séquences répétées localisées de part et d'autre d'un gène sur deux chromosomes homologues.

B. Rétrotranscription d'un ARNm et insertion de l'ADNc correspondant dans le génome.

C. Sur-réplication d'une région par initiations multiples de la réplication au niveau d'une origine de réplication.

D. Cassures aléatoires non identiques sur deux chromosomes homologues suivies de leur réunion.

D'où vient l'activité transcriptase inverse qui permet la formation de ces ADNc? Une approche a consisté à rechercher des traitements induisant la formation de pseudogènes processés puis de corréler cette induction avec l'expression de certains rétroéléments. Un traitement à la 5-azacytidine, un agent déméthylant, permet d'augmenter plus de 10 fois la formation de pseudogènes à partir d'un gène rapporteur, sans que la transcription de ce rapporteur ne soit affectée. En revanche, la transcription des rétrotransposons sans LTR de type LINE est fortement induite suggérant qu'ils puissent être la source de l'activité transcriptase inverse recherchée (Tchenio *et al.*, 1993). L'implication des LINEs a pu être vérifiée expérimentalement en coexprimant un gène rapporteur de la rétroposition ou un gène contenant des introns et un élément LINE actif sous le contrôle de promoteurs forts, ce qui aboutit à la formation d'ADNc à une fréquence élevée (Dhellen *et al.*, 1997). Ces ADNc sont intégrés dans le génome avec toutes les caractéristiques d'un pseudogène processé (Esnault *et al.*, 2000). Ces résultats indiquent clairement que l'activité transcriptase inverse des rétrotransposons sans LTR de type LINE est capable de rétrotranscrire des ARNm cellulaires dont les produits forment des pseudogènes.

4.2. Duplication de segments

4.2.1. Duplication et recombinaison inégale (Bailey *et al.*, 2003)

Comme pour les gènes en répétition en tandem, il a été proposé que les événements de crossing-over inégaux entre régions homologues pouvaient être à l'origine de la formation de duplications de segments en tandem. Chez l'homme, l'analyse des séquences localisées de part et d'autre des segments dupliqués, décrits précédemment, a permis d'apporter des éléments de réponse quant aux mécanismes impliqués (Bailey *et al.*, 2003). En effet, 27 % des segments dupliqués sont encadrés par des séquences répétées *Alu* suggérant fortement que ces éléments répétés ont pu servir de substrat à des événements de recombinaison allélique et ont conduit à ce type de duplication. Cependant, toutes les duplications détectées ne résultent pas d'un événement de recombinaison non allélique. Les jonctions caractérisées aux extrémités de certaines duplications observées dans le génome ne peuvent résulter de tels événements. Ainsi, un autre mécanisme doit permettre la formation de duplications segmentales.

4.2.2. Duplication et erreur de réplication (Koszul *et al.*, 2004)

En utilisant un crible de dosage génique basé sur les gènes *RPL20A* et *RPL20B* chez *Saccharomyces cerevisiae*, Koszul *et al.* (2004) ont isolé des révertants dans lesquels un événement de duplication segmentale s'est produit. Les événements correspondent majoritairement à des duplications segmentales dupliquées en tandem direct. Ces segments d'ADN ont une taille qui varie de 41 à 288 kb. L'analyse des jonctions des blocs dupliqués révèle trois types de séquences impliquées dans les réarrangements : des séquences LTR d'environ 300 pb, des séquences d'ADN de faible complexité (trinuécléotides et polyA/T) et des séquences microhomologues de 2 à 9 nucléotides. Des événements de recombinaison inégale entre éléments répétés ne sont donc pas envisageables pour au moins ces deux derniers types d'éléments. Certaines observations comme notamment le type de séquence à la jonction (les trinuécléotides) ainsi que la superposition entre ces jonctions et les sites de terminaison de réplication laisse penser que ces réarrangements pourraient résulter d'accidents de réplication. Cependant, seule l'étude de mutants appropriés peut confirmer ces hypothèses.

5. DEVENIR DES GÈNES DUPLIQUÉS

L'apparition de gènes paralogues peut se faire par la duplication d'un seul gène, d'un segment composé de plusieurs gènes contigus, d'un chromosome entier ou encore d'un génome entier. La duplication d'un gène se produit à une fréquence d'une duplication par gène tous les 100 millions d'années chez les eucaryotes tels l'*H. sapiens*, *M. musculus*, *D. melanogaster*, *C. elegans*, *A. thaliana* et *S. cerevisiae*. Cette fréquence est comparable à celle observée pour la substitution de nucléotides.

Les duplications sont considérées comme un des moteurs de l'évolution, mais une séquence, une fois, dupliquée peut être fixée mais également perdue au sein d'un génome. Une copie dupliquée ayant une valeur sélective neutre par rapport à la copie originelle aura une probabilité très faible d'être retenue. Ceci suggère qu'une proportion importante de duplications va être rapidement perdue. La probabilité de fixation d'un gène est liée à la fonction qu'il représente. En effet, les gènes impliqués dans des fonctions physiologiques (l'immunité, la reproduction...) qui varient entre les espèces ont probablement un taux de naissance et de mort plus important.

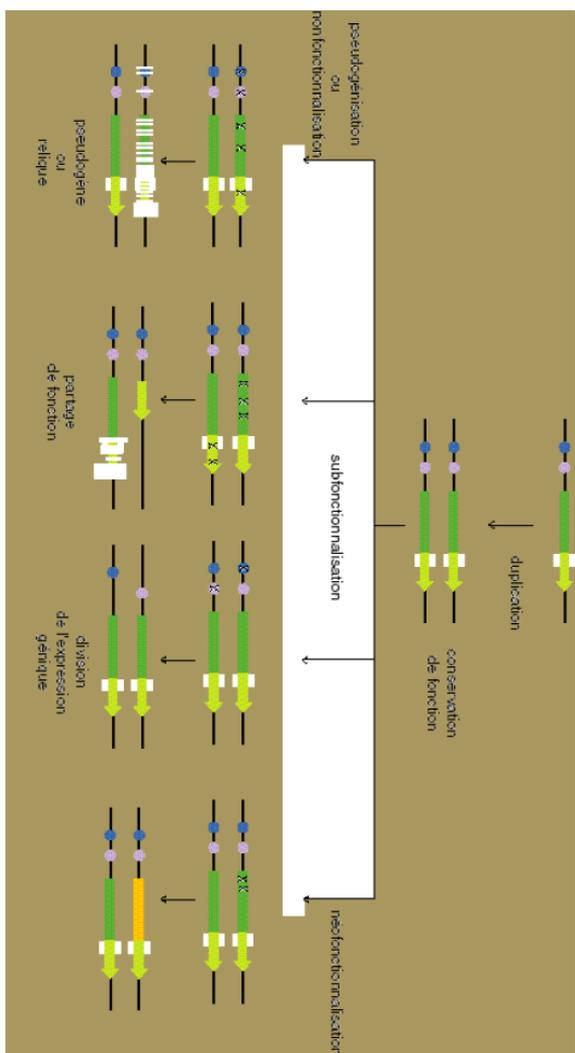


FIGURE 15 – Diverger des séquences dupliquées.
 Le gène code pour protéine avec deux domaines fonctionnels représentés en vert :
 domaine 1
 domaine 2
 Les cercles de couleurs en amont du gène représentent les régions régulières.

La copie dupliquée, une fois fixée, peut alors évoluer de différentes manières. Plusieurs scénarios évolutifs peuvent être proposés (Fig. I.5) :

- **la pseudogénéisation** : une des copies va être inactivée par l'accumulation de mutations ;
- **la conservation de la fonction** : la copie dupliquée conserve la même fonction que la copie originelle ;
- **la subfonctionnalisation** : on parle également de «partage de fonctions ancestrales». La subfonctionnalisation peut avoir lieu, par exemple, lors d'une duplication d'un gène codant deux domaines fonctionnels. Par ce processus, chaque copie va assurer une des fonctions suite à l'accumulation de mutations ;
- **la néofonctionnalisation** : les deux copies vont diverger. L'une des copies va conserver la fonction originelle alors que l'autre va acquérir une nouvelle fonction.

5.1. La pseudogénéisation

La duplication génère une redondance fonctionnelle et il n'est pas toujours avantageux d'avoir deux copies identiques d'un gène. En d'autres termes, les mutations destabilisant la structure et la fonction d'une des copies ne sont ni délétères, ni éliminées par sélection. Progressivement, l'accumulation de mutations va entraîner la formation d'un pseudogène c'est-à-dire d'une copie de gène qui n'est plus exprimée ou dont la protéine n'est plus fonctionnelle. Ce destin évolutif des duplications a été mis en évidence par des modèles de génétique de population (Lynch *et al.*, 2001), mais également par analyse génomique (Lynch and Conery, 2000 ; Harrison *et al.*, 2002). Par la suite, ces pseudogènes peuvent être délétés du génome ou alors diverger de telle manière qu'aucune similarité de séquence avec la copie originelle ne peut plus être mise en évidence. Par contre, si les pseudogènes ont été formés récemment, ils pourront être identifiés grâce à leur similarité de séquence. Par exemple, des analyses génomiques permettent d'identifier 2168 pseudogènes chez *C. elegans* c'est-à-dire environ un pseudogène pour huit gènes fonctionnels (Harisson *et al.*, 2001). Chez l'homme, il en existe encore davantage. Une analyse des chromosomes 21 et 22 montre un pseudogène pour deux gènes fonctionnels (Harisson *et al.*, 2002). On parle de pseudogénéisation pour qualifier le processus par lequel un gène fonctionnel devient un pseudogène. Il s'agit du processus le plus probable en l'absence d'avantages sélectifs.

Cependant, certains gènes dupliqués ont été maintenus dans le génome par leur spécificité fonctionnelle, mais sont devenus des pseudogènes par relaxation de la contrainte fonctionnelle. Par exemple, la taille de la famille de gènes olfactifs est similaire chez l'homme et la souris (environ 1000 gènes). Mais le pourcentage de pseudogènes présents dans cette famille avoisine les 60 % chez l'homme alors que chez la souris, il avoisine 40%. De nombreux gènes sont donc devenus des pseudogènes depuis l'origine des hominidés (Rouquier, 2000). Dans ces espèces, ceci est probablement dû à la réduction de l'utilisation du sens olfactif qui a été compensé par une meilleure vision.

5.2. La conservation de la fonction

La présence de deux copies de gènes peut parfois être un avantage simplement parce que cette redondance permet d'augmenter les quantités de protéines ou d'ARN synthétisés. Ceci s'applique largement aux gènes ayant une forte expression due à une demande importante du produit correspondant, comme pour les ARN ribosomiques, les ARN de transfert et les histones. Dans ces cas de figure, les séquences paralogues ont très peu divergé puisque les produits des gènes ont la même fonction. Les ADNs codant pour les ARN de transfert et ribosomiques constituant 90 % de l'activité de transcription d'une cellule eucaryote est l'exemple le plus évident. C'est également le cas, pour les gènes codant les histones chez *D. melanogaster* qui seraient nécessaires pour répondre aux besoins importants du développement embryonnaire précoce.

Ainsi, des déficits en dosage génique peuvent être observés suite à l'altération d'une copie dupliquée d'une famille de gènes chez certains mutants. Cette situation est due au fait que la ou les copies restantes ne peuvent pas compenser au niveau de la transcription ou de la traduction le déficit résultant de cette mutation. Un phénotype peut être associé à ce déficit de dosage génique. Chez *S. cerevisiae*, on trouve ainsi 59 familles de deux gènes codant pour des protéines ribosomiques (Payne and Garrels, 1997). Ces gènes sont très conservés (de l'ordre de 98% d'identité) et pour 28 familles au moins la délétion de l'un des deux gènes génère un phénotype de retard de croissance. Les phénotypes ne sont pas symétriques suivant que la délétion concerne l'un ou l'autre des gènes, mais le retard de croissance observé apparaît lié à l'activité transcriptionnelle du gène délété. On n'a donc pas de compensation transcriptionnelle, ni de régulation post-transcriptionnelle permettant de maintenir une

quantité de protéine normale si une copie est absente. Dans ces cas-là, la redondance s'explique par le besoin de maintenir une quantité optimale de transcrits.

Dans certains cas, il peut y avoir une régulation post-transcriptionnelle permettant le maintien de la proportion des produits de transcription si une copie est absente. Un exemple est constitué par la famille des protéines ribosomiques *CRY1*/*CRY2* de *S. cerevisiae* (Fewell and Woolford, 1999). La double délétion de ces gènes est létale. La quantité d'ARNm de *CRY1* est dix fois supérieure à celle de *CRY2* et donc la délétion de *CRY2* n'a que peu d'effets sur la croissance. Par contre, si *CRY1* est absent, la quantité d'ARNm de *CRY2* est augmentée d'un facteur cinq à dix. La délétion n'a donc que peu d'effets sur la croissance. La régulation de la quantité d'ARNm de *CRY2* dans la cellule sauvage s'effectue de manière post-transcriptionnelle et résulte de la fixation de la protéine Rps4p sur le pré-messager, empêchant son épissage et provoquant sa dégradation. Dans la cellule mutante, cette régulation apparaît modifiée. La duplication génique peut donc s'expliquer par le besoin de maintenir une certaine quantité de produit des gènes dans la cellule.

5.3. La subfonctionnalisation

Bien qu'une quantité plus importante de produit des gènes peut être avantageuses, deux gènes ayant la même fonction ne sont pas stables au sein d'un génome (Nowak *et al.*, 1997). En théorie, deux copies identiques peuvent être maintenues si la fonction correspondante diffère par certains points (Nowak *et al.*, 1997). Le processus de subfonctionnalisation, encore appelé « partage de fonctions ancestrales », permet justement aux deux copies d'acquérir une différence de fonction ou d'expression assurant leur maintien. La subfonctionnalisation peut avoir lieu lors d'une duplication d'un gène codant deux domaines fonctionnels. Chaque copie va conserver un des deux domaines fonctionnels par accumulation de mutations dans la séquence codant l'autre domaine fonctionnel. Ainsi, la fonction ancestrale sera partagée entre les deux copies. La subfonctionnalisation peut également correspondre à la division de l'expression génique après une duplication c'est-à-dire que les deux copies seront exprimées dans des conditions différentes. Dans les deux cas, les deux gènes se complètent fonctionnellement et seront conservés dans le génome.

En 1991, Piatigorsky et Wistow sont les premiers à mettre en avant ce modèle

en se basant sur une étude des protéines qui forment le cristallin de l'œil. Ils parlent alors de « partage de fonction ». Par la suite, à partir de ses recherches sur le Xénope, Hughes (1994) arrive aux mêmes conclusions. Le rôle de ces événements dans la conservation des gènes paralogues dans le génome a été développé et généralisé par Force *et al.* (1999). Ils qualifient ce processus de « DDC » c'est-à-dire Duplication-Dégénération-Complémentation.

Une forme de la subfonctionnalisation qui est potentiellement importante pour l'évolution correspond à la division de l'expression génique après une duplication (Force *et al.*, 1999). Il a été montré que de nombreux gènes dupliqués suivaient cette forme de subfonctionnalisation (Force *et al.*, 1999). Par exemple, chez le zébrafish, une paire de gènes paralogues *engrailed-1* et *engrailed-2* issue d'un événement de duplication segmentale ont suivi ce processus. En effet, ces deux gènes sont exprimés dans des tissus différents. Par contre, chez la souris, un seul gène *engrailed-1*, orthologue aux deux présents chez le zébrafish, est exprimé dans les deux tissus (Force *et al.*, 1999). Le changement d'expression après la duplication de gène semble être fréquent et apparaît rapidement après l'événement de duplication (Wagner, 2000 ; Gu *et al.*, 2002). Cependant, il est très difficile de déterminer le pourcentage de gènes dupliqués qui évoluent par un processus de subfonctionnalisation.

5.4. La néofonctionnalisation

L'un des processus le plus important sur le plan évolutif pour les séquences dupliquées est la formation d'une nouvelle fonction. Chez les primates, ce processus a été décrit pour deux gènes appartenant à une famille des Rnase A dont les produits sont présents au niveau des granules des leucocytes éosinophiles (Zhang *et al.*, 1998). Les copies résultant d'un événement de duplication ont divergé pour donner deux gènes : *EDN* (eosinophil-derived neurotoxin) et *ECP* (eosinophil cationic protein). Après l'événement de duplication, une activité anti-bactérienne a émergé au niveau de la protéine ECP. Cette activité n'est pas présente chez la protéine EDN. Plus surprenant, cette activité anti-bactérienne n'est pas liée à l'activité ribonucléase originelle du gène ancestral (Rosenberg and Dyer, 1995). Des analyses moléculaires suggèrent que cette nouvelle fonction résulte probablement de l'accumulation rapide, après la duplication, de substitutions d'arginine dans la séquence protéique (Zhang *et al.*, 1998). La protéine

ECPest toxique pour les bactéries car elle crée des pores au niveau de leur membrane cellulaire. Les résidus arginine chargés négativement doivent être importants pour établir un contact entre la protéine ECP et les membranes bactériennes chargées négativement afin d'engager le processus de formation de pores.

Cependant, dans de nombreux cas, la néofonctionnalisation correspond à une spécialisation de la fonction ancestrale. L'un des exemples incontournables est constitué par la superfamille des globines chez les eucaryotes supérieurs. Ces protéines à hème sont adaptées à la fixation réversible de l'oxygène. Chez l'homme, la super famille des globines est constituée, dans l'état actuel des connaissances, de quatre familles:

- la famille de la neuroglobine dont l'unique membre est localisé sur le chromosome 14 (Burmester *et al.*, 2000) ;
- la famille de la myoglobine dont l'unique membre est localisé sur le chromosome 22 ;
- la famille multigénique de γ -globine localisée sur le chromosome 16 ;
- la famille de la δ -globine localisée sur le chromosome 11.

Les quatre familles produisent trois types de protéines ayant une activité spécifique: la myoglobine, la neuroglobine et l'hémoglobine. Elles ont divergé à partir d'un ancêtre commun et sont devenues spécialisées sous plusieurs aspects. Ainsi, chaque protéine a une spécificité tissulaire. La myoglobine est devenue la protéine de stockage de l'oxygène dans les muscles, la neuroglobine celle des cellules nerveuses. L'hémoglobine, quant à elle, permet le transport de l'oxygène dans le sang. Par ailleurs, alors que la myoglobine et la neuroglobine ont gardé une structure monomérique, l'hémoglobine est devenue un hétérotétramère composé de deux sous-unités α et de deux sous-unités β .

Chez l'homme, la famille β comprend 4 gènes fonctionnels et trois pseudogènes alors que la famille α comprend 5 gènes et un pseudogène (Fig. I.6). Ces différents gènes ont divergé à la fois au niveau de la régulation et de la fonction. En effet, différents hétérotétramères d'hémoglobines apparaissent à différents stades du développement : $\alpha_2\beta_2$ et $\alpha_2\gamma_2$ dans l'embryon, $\alpha_2\beta_2$ dans le fœtus, puis $\alpha_2\beta_2$ et $\alpha_2\delta_2$ chez l'adulte. Des différences d'affinité pour l'oxygène sont à l'origine de la présence de ces différents hétérotétramères. Par exemple, l'hémoglobine fœtale ($\alpha_2\gamma_2$) possède une affinité à l'oxygène plus importante que celle de l'adulte. Ceci permet au fœtus de recevoir assez d'oxygène à partir du sang maternel. La base moléculaire de cette affinité provient de la présence d'une sérine en position 143 de la globine γ alors qu'il s'agit

d'une histidine dans la chaîne .

La création d'une nouvelle fonction à partir d'un gène dupliqué peut nécessiter un nombre important de substitutions d'acides aminés comme c'est le cas pour la protéine ECP(9 acides aminés) (Zhang *et al.*, 1998) ou une seule substitution comme pour les globines et .

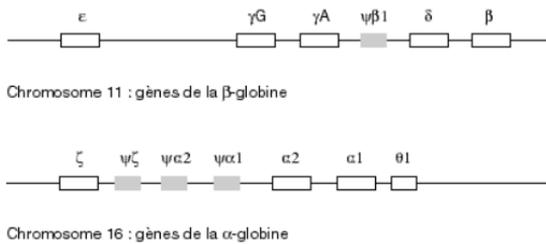


FIGURE 1.6 – Gènes de l'hémoglobine présents chez l'homme. Les rectangles gris correspondent aux pseudogènes.

6. SACCHAROMYCES CEREVISIAE : UN ORGANISME MODÈLE POUR L'ÉTUDE DES DUPLICATIONS

S. cerevisiae est un outil de choix pour l'étude de l'organisation et de la structure des duplications au sein d'un génome. En effet, la séquence complète de son génome est disponible depuis 1996 (Goffeau *et al.*, 1996). Il fut le premier eucaryote dont le génome fut entièrement séquencé. Seuls les génomes de deux organisme procaryotes : *Haemophilus influenza* (Fleischmann *et al.*, 1995) et *Mycoplasma genitalium* (Fraser *et al.*, 1995) furent achevés avant celui de cette levure. Son génome est remarquablement compact. On y trouve un gène environ tous les 2 kb et plus de 70% de l'ADN chromosomique correspond à une région codante (Goffeau *et al.*, 1996). De plus, le génome de *S. cerevisiae* est certainement le génome eucaryote pour lequel l'identification des ORFs est la plus précise. En effet, l'annotation, c'est-à-dire l'étape d'identification des ORFs, initiée en 1996 fut suivie de plusieurs étapes de modifications. L'annotation est

encore régulièrement mise à jour. Dans le cadre de l'étude des duplications de gène, l'intérêt de cette séquence complète a été de pouvoir comparer l'ensemble des ORFs présent au sein de *Saccharomyces cerevisiae* afin de caractériser leur redondance. Cette analyse a permis de donner une vision globale sur l'ensemble des séquences d'ADN dupliquées. De plus, d'autres levures hémiascomycètes ont été partiellement ou entièrement séquencées ces dernières années (Fig. I.7). Une analyse de comparaison des ORFs, notamment dupliquées, entre les différents génomes séquencés a été possible et a permis d'apporter un certain nombre de réponses sur le plan évolutif des hémiascomycètes.

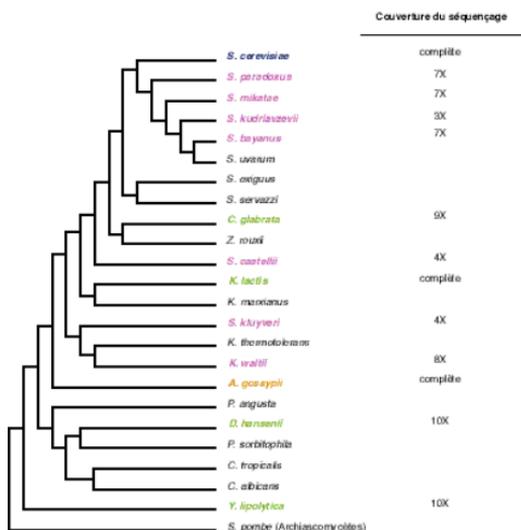


FIGURE I.7 – Levures hémiascomycètes partiellement ou entièrement séquencées.

6.1. Les catégories de duplications présentes chez *Saccharomyces cerevisiae*

La comparaison de l'ensemble du génome de *S. cerevisiae* a permis de prendre conscience de l'importance de la proportion de gènes paralogues. En effet, environ 40 % des gènes codant une protéine appartiennent à une famille de paralogues. Il existe environ 722 familles dont 457 comprennent deux gènes. Les 265 autres familles peuvent être constituées de 3 gènes (124 familles) à plus de 20 gènes (11 familles) (Blandin *et al.*, 2000). Les gènes paralogues de *S. cerevisiae* sont dispersés sur les 16 chromosomes et appartiennent à toutes les catégories fonctionnelles avec une sur-représentation de gènes codant des transporteurs membranaires.

On peut distinguer 4 types de gènes dupliqués (Fig. I.8):

les gènes dispersés,

les gènes répétés en tandem,

les gènes subtélomériques,

les blocs « fossiles » comprenant une série de gènes.

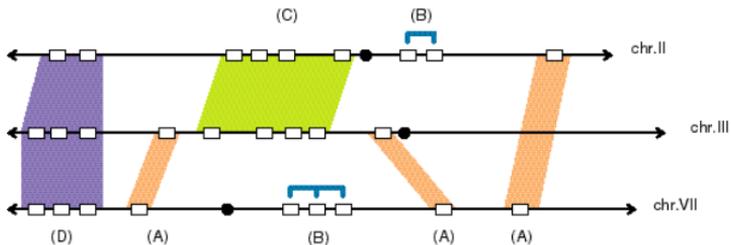


FIGURE I.8 – Types de duplications observées chez *Saccharomyces cerevisiae*.

(A) duplications dispersées,

(B) duplications en tandem,

(C) blocs fossiles,

(D) duplications subtélomériques.

6.1.1. Les duplications dispersées

La majorité des gènes dupliqués sont dispersés sur l'ensemble des chromosomes de *Saccharomyces cerevisiae*. Parmi ces gènes, on trouve les gènes de tRNA. Dans la souche séquencée, il y a 274 gènes codant pour des ARNt répartis sur les 16 chromosomes et se regroupant en 45 familles. La famille des gènes de tRNA^{Tyr} constitue un exemple typique. Il existe huit gènes identiques codant un seul type de tRNA^{Tyr} présent dans le cytoplasme. Les huit gènes sont dispersés sur cinq chromosomes et sont tous exprimés. Il faut déléter au moins cinq copies avant de détecter un effet sur la croissance cellulaire.

Parmi les séquences particulièrement répétées et dispersées dans le génome, on retrouve les éléments mobiles de type rétrotransposons. Chez *S. cerevisiae*, les rétrotransposons avec LTR sont les seuls éléments transposables présents et se répartissent en 5 classes (Ty1-Ty5) basées sur des similarités de séquence et sur l'organisation du gène *POL*. Dans la souche S288C, dont la séquence complète du génome a été publiée, 331 éléments Ty ont été retrouvés, se répartissant de la manière suivante : 217 Ty1, 34 Ty2, 41 Ty3, 32 Ty4 et 7 Ty5. Parmi ces éléments, 85% correspondent à des séquences LTR ou à des fragments de séquences LTR. Dans l'ensemble, les séquences des rétrotransposons constituent 3,1% du génome de *S. cerevisiae* (Kim *et al.*, 1998).

6.1.2. Les duplications en tandem

Le second type de duplications correspond aux gènes issus de duplications en tandem, directes ou inversées. Les ADN ribosomiques, situés sur le chromosome XII, constituent le plus grand type de séquences répétées dans le génome de la levure. Les deux unités 5S et 35S ont une taille de 9,1 kb (Fig. I.9). L'unité 35S transcrite par la polymérase I donne un ARN précurseur qui est ensuite mûré en ARN 18S ; 5,8S et 25S. Ce qui est le plus étonnant dans les répétitions d'ADN ribosomiques de *Saccharomyces cerevisiae* est l'existence d'un gène codant un ARN 5S à l'intérieur de l'unité de transcription par la polymérase I. Ce gène codant l'ARN 5S est, quant à lui, transcrit par la polymérase III. Le sens de transcription est contraire à celui de l'unité de transcription par la polymérase I. Ces deux unités forment un bloc de 9,1 kb qui est répété en tandem. Les répétitions sont jointes tête à queue. Le nombre de répétitions varie de 100 à 200 copies suivant la souche de levure, expliquant de ce fait le polymorphisme de taille du chromosome XII.

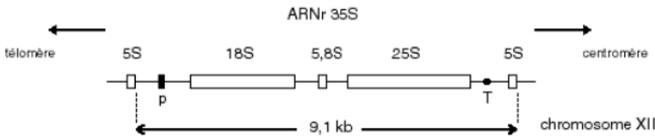


FIGURE 1.9 – Unité répétée d'ADN ribosomique chez *Saccharomyces cerevisiae*.

6.1.3. Les duplications subtélomériques

Les régions subtélomériques présentent une grande plasticité chez l'ensemble des génomes. Chez *Saccharomyces cerevisiae*, ce sont des régions riches en gènes paralogues. Elles contiennent des familles multigéniques de grandes tailles. De plus, elles sont hautement polymorphes et subissent des événements de recombinaison à haute fréquence (Louis, 1995). Le séquençage de nouveaux génomes de levure a également révélé la grande plasticité de ces régions subtélomériques (Dujon *et al.*, 2004 ; Kellis *et al.*, 2003).

Le locus *SUC* constitue un exemple de duplication subtélomérique. Les gènes *SUC* codent une invertase qui est une enzyme nécessaire à l'utilisation de saccharose (Carlson *et al.*, 1985). Ces gènes ont été cartographiés à six différents loci, tous proches des extrémités des chromosomes. Seul le locus *SUC2* est trouvé dans toutes les souches de laboratoire. Il est localisé à l'extrémité du bras gauche du chromosome IX à plus de 30 kb du télomère alors que les cinq autres loci *SUC* sont immédiatement adjacents aux télomères et entourés de répétitions subtélomériques.

6.1.4. Les blocs « fossiles »

La séquence complète du génome de *S. cerevisiae* a également révélé des blocs de duplication qui ont été qualifiés de « fossiles » c'est-à-dire correspondant aux traces d'un événement de duplication ancestrale. Wolfe et Shields (1997) ont mis en évidence 55 blocs de duplication ancestrale contenant 376 paires de gènes homologues. L'identité de séquence protéique entre les paires est comprise entre 24 % et 100 % avec une moyenne de 63%. Trois critères importants ont été définis pour rechercher ces régions dupliquées :

- le score attribué à l'alignement de séquence par le programme BLASTPest supérieur à 200 pour chaque paire de gènes homologues ;
- chaque bloc comprend au moins 3 paires de gènes homologues et la distance intergénique entre ces gènes doit être inférieure à 50 kb ;
- le même ordre et les mêmes orientations relatives des gènes sont gardés au niveau de ces deux régions.

Ainsi, 55 régions dupliquées avec une taille moyenne de 55 kb et contenant 6,9 paires de gènes homologues en moyenne ont été mises en évidence chez *S. cerevisiae*.

Deux hypothèses peuvent être formulées pour expliquer comment ces 55 régions se sont dupliquées. La première consisterait en une série de duplications segmentales successives et indépendantes menant à la situation actuelle. La seconde consisterait en une seule duplication complète du génome. Cet événement serait suivi par d'autres réarrangements comme des translocations réciproques entre les chromosomes et des délétions conduisant à la formation de blocs de duplications mosaïques. La polyploïdisation du génome de *S. cerevisiae* avait déjà été proposée par Smith en 1987.

Wolfe et Shield (1997) se prononcent en faveur du modèle de duplication complète du génome / translocations pour deux raisons. D'une part, pour 50 des 55 régions dupliquées, l'orientation des blocs par rapport au centromère est la même pour les deux segments, ce qui ne semble pas être aléatoire. En effet, l'orientation des blocs devrait être conservée s'ils sont issus de la duplication de chromosomes suivie d'événements de translocation réciproque. Par contre, si les blocs sont issus d'événements de duplication segmentale indépendant alors leur orientation devrait être aléatoire et on devrait s'attendre à des événements de triplication, ce qui n'est pas observé dans le génome de *S. cerevisiae*.

Cette duplication complète du génome se serait produite après la divergence entre *S. cerevisiae* et *K. lactis* (Fig. I.7) (Wong *et al.*, 2002). Le séquençage récent des levures *Kluyveromyces waltii* et *Ashbya gossypii* ont permis de confirmer le modèle de la duplication complète du génome (Kellis *et al.*, 2004 ; Dietrich *et al.*, 2004). Ces levures proviennent d'un ancêtre commun qu'elles partagent avec *S. cerevisiae* (Fig. I.7). Cependant, situées en aval de *K. lactis* sur l'arbre phylogénétique, elles n'ont pas été le siège de la duplication complète du génome. La comparaison de ces souches avec *S. cerevisiae* permet d'analyser la situation avant et après la duplication complète du

génomique. L'analyse montre que deux segments mosaïques formant un bloc de duplication chez *S. cerevisiae* correspondent effectivement à une région unique chez *K. waltii* et *A. gossypii* (Fig. 10). Les gènes homologues conservent l'ordre et l'orientation. Environ 10 % des gènes qui appartiennent aux blocs dupliqués se retrouvent dans les deux régions dupliquées alors qu'ils sont uniques dans les deux autres levures séquencées. Ces résultats soutiennent l'hypothèse qu'une duplication complète du génome de *S. cerevisiae* suivie d'une perte d'une copie sur l'un ou l'autre des segments s'est produite. Dans un certain nombre de cas (environ 10 %), les deux copies ont été retenues au sein du génome.

Bien que la présence des segments dupliqués chez *S. cerevisiae* soit issue d'une duplication complète du génome, des travaux récents montrent qu'il ne s'agit pas de la seule et unique source impliquée dans la formation de telles duplications (Dujon *et al.*, 2004). En effet, le séquençage systématique de 3 levures hémiascomycètes : *Kluyveromyces lactis*, *Debaryomyces hansenii* et *Yarrowia lipolytica* ont permis de mettre en évidence d'autres blocs de duplication :

- 8 blocs (un en tandem) chez *K. lactis*,
- 5 blocs (3 en tandem) chez *D. hansenii*,
- 2 blocs chez *Y.lipolytica*.

Ces blocs de duplication ne coïncident pas avec ceux retrouvés chez *S. cerevisiae*. De plus, ils sont trop peu nombreux pour suggérer un événement de duplication massive. Ces observations permettent donc de conclure que d'autres mécanismes peuvent être à l'origine de la formation de duplications segmentales chez les levures et illustrent la multiplicité des moyens par lesquels une séquence d'ADN peut se dupliquer.

6.2. Les duplications mises en évidence par des approches expérimentales

La séquence complète de *S. cerevisiae* a permis de mettre en évidence l'ensemble des duplications qui ont une forme stable dans un génome. Cependant, comme nous l'avons vu précédemment, d'autres types de duplication peuvent se produire, comme l'aneuploïdie, mais ne sont pas détectables par analyse des séquences. Ainsi par des approches expérimentales, il a été possible de mettre en évidence chez *S. cerevisiae* le processus d'amplification génique et l'importance de l'aneuploïdie.

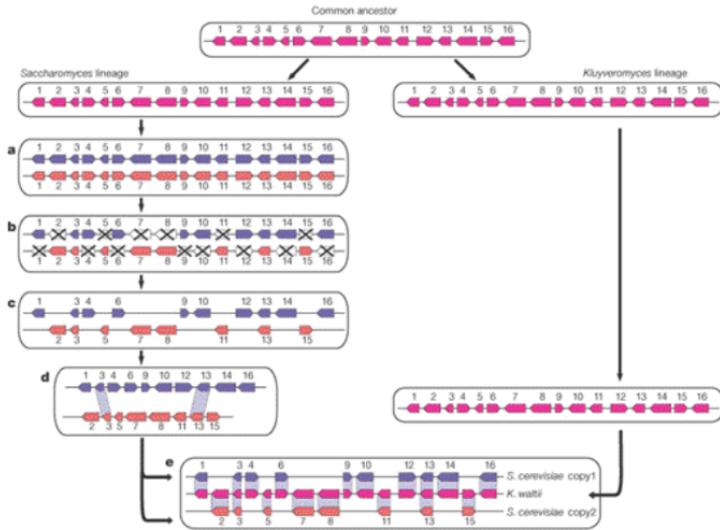


FIGURE I.10 –Modèle de la duplication totale du génome.

(a) l'événement de duplication totale dans le genre *Saccharomyces* entraîne la création d'une copie de chaque gène et chaque chromosome.

(b) la majorité des gènes dupliqués sont suivis un processus de nonfonctionnalisation et seront donc perdus,

(c) les régions dupliquées conservent de manière asymétrique certains nombre de gènes,

(d) certains gènes (environ 10%) sont conservés au niveau des deux régions dupliquées,

(e) la comparaison entre *S. cerevisiae* et *K. waltii* montre que deux segments mosaïques formant un bloc de duplication chez *S. cerevisiae*, correspondent effectivement à une région unique chez *K. waltii*, (d'après Kellis et al., 2004).

6.2.1. L'amplification génique

L'amplification génique caractérise une multiplication brutale de copies de gènes. Elle peut être spontanée, comme dans le cas du gène *CUPI* de *S. cerevisiae*. Ce gène code une protéine complexant les ions cuivre. Le nombre de gènes varie suivant le milieu de sélection. Il existe en effet une corrélation directe entre le nombre de répétitions en tandem et le niveau de résistance au cuivre. Ce locus est donc un modèle simple pour l'étude de l'amplification de gènes (Butt and Ecker, 1987). Dans la plupart des souches qui possèdent plusieurs copies *CUPI*, l'unité de répétition a une taille de 2 kb. Elle code deux transcrits, mais seule celle qui code la chélatine du cuivre est impliquée dans la résistance (Karin *et al.*, 1984). La sélection de souches ayant une résistance accrue aux ions cuivres provoque l'accroissement à haute fréquence du nombre de copies *CUPI* pendant la croissance mitotique, à condition qu'il y ait plus d'une copie dans la souche de départ. En effet, il n'a jamais été possible de sélectionner un événement initial de duplication dans une souche ne contenant au départ qu'une seule copie *CUPI* (Fogel and Welch, 1982).

6.2.2. L'aneuploïdie

Hughes *et al.* (2000) sont les premiers à avoir mis en évidence l'importance de l'aneuploïdie chez *S. cerevisiae*. L'analyse du transcriptome par microarray d'une collection de 300 mutants de délétion (issus du *Saccharomyces* Genome Deletion Consortium) montre que des gènes appartenant à un seul chromosome sont surexprimés dans environ 8 % des souches délétées. Ces données suggèrent fortement la présence d'un chromosome supplémentaire ce qui a été confirmé par hybridation génomique comparative sur microarray.

La ploïdie (le contenu génomique d'une cellule) est généralement maintenue à un niveau constant et les chromosomes sont généralement répartis avec un taux de fidélité élevé. Une mauvaise ségrégation des chromosomes conduit soit à la mort cellulaire soit à une maladie. Un certain nombre de maladies humaines sont liées à une aneuploïdie dont le principal exemple est le syndrome de Down causé par une trisomie du chromosome 21. Il est donc très surprenant de constater que l'aneuploïdie soit très répandue chez *S. cerevisiae*.

Plusieurs explications ont été avancées pour expliquer ce taux élevé d'aneuploïdies dans la collection des mutants de délétion. Premièrement, une mutation définie dans une souche peut augmenter le taux de mauvaise ségrégation des chromosomes. En effet, une mutation dans un gène impliqué dans la réparation et la ségrégation (comme *BUB1*, *BUB3* et *BIM1*) sont connus pour augmenter le taux d'événements d'aneuploïdies (Hartwell and Smith, 1985). Cependant, la majorité des mutants présentant une aneuploïdie ne sont pas directement impliqués dans la stabilité du génome. Une autre explication pour ce taux élevé est qu'une souche ayant acquis une délétion suivie d'une aneuploïdie peut avoir un avantage sélectif par rapport à une cellule ayant une ploïdie normale. Par exemple, dans leur étude, Hughes *et al.* ont mis en évidence un chromosome IX supplémentaire dans une souche diploïde portant une délétion des deux copies du gène codant la protéine ribosomique Rps24a localisé sur le chromosome V. Or le chromosome IX porte le gène *RPS24B* qui a 97 % d'identité de séquence avec le gène *RPS24A*. De plus, la délétion des deux copies du gène *RPS24A* entraîne un phénotype de croissance ralentie. Ainsi, les cellules qui ont gagné un chromosome IX permettent de compenser l'absence du gène *RPS24A* par un dosage génique de son paralogue *RPS24B*. Les cellules auront donc un avantage sélectif par le biais de l'aneuploïdie. Ces résultats donnent une explication biologique à la présence de tels événements.

7. SYSTÈME DE SÉLECTION UTILISÉ AU LABORATOIRE POUR L'ÉTUDE DES DUPLICATIONS

Ces dernières années, l'analyse des séquences complètes de génomes a permis de mettre en évidence l'importance quantitative des séquences dupliquées au sein des génomes. Malheureusement, leur analyse ne permet pas d'accéder aux mécanismes qui sont impliqués dans la genèse des duplications. C'est pourquoi, il est nécessaire d'élaborer des cribles de sélection permettant d'isoler puis d'analyser de tels remaniements.

Au laboratoire, nous possédons un système génétique, basé sur l'allèle muté *ura2 15-30-72* du gène *URA2*, qui permet la sélection positive de réarrangements chromosomiques chez la levure *Saccharomyces cerevisiae* : insertion de séquences d'ADN mobiles (rétrotransposons), délétion de fragments chromosomiques mais également duplication de régions chromosomiques (Roelants *et al.*, 1995).

Notre étude s'est focalisée sur les événements de duplication. Deux collections de révertants bien distinctes ont été analysées sur le plan moléculaire : l'une réalisée à partir d'une souche haploïde *ura2 15-30-72*, l'autre à partir d'une souche diploïde homozygote *a/a ura2 15-30-72*. L'objectif d'une telle étude était double. Dans un premier temps, nous avons voulu mettre en évidence les moyens par lesquels notre séquence d'intérêt était dupliquée et par cet intermédiaire apporter une preuve expérimentale quant à leur pluralité. Dans ce cadre, nous avons comparé les deux états : haploïde et diploïde afin de savoir si les moyens mis en œuvre pour dupliquer une séquence étaient similaires ou différents. Dans un second temps, l'objectif de l'analyse de ces révertants était d'apporter des réponses quant aux mécanismes pouvant être à l'origine de ces duplications et de proposer des modèles de formation de ces remaniements.

M MATÉRIEL ET MÉTHODES

1. ORGANISMES ET MÉTHODES DE CULTURE

1.1. *Saccharomyces cerevisiae*

1.1.1. Biologie de *Saccharomyces cerevisiae*

Saccharomyces cerevisiae est un champignon ascomycète hétérothallique dont le cycle cellulaire est haplodiplobiontique. Ainsi, les cellules peuvent se maintenir sous forme haploïde ou diploïde par bourgeonnement (Figure 1) (Kutzmann, 1998). En conditions de croissance défavorables (carence en source de carbone ou d'azote), les cellules diploïdes sporulent et donnent naissance à un asque contenant quatre spores non ordonnées dont deux de signe a et deux de signe A. Lorsque les conditions redeviennent plus favorables, ces spores germent et forment des cellules haploïdes. Deux cellules de signe opposé peuvent être croisées afin de générer un diploïde. Toutes ces étapes peuvent être provoquées en laboratoire.

1.1.2. Souches utilisées

Les souches de *S. cerevisiae* utilisées pour cette étude sont isogéniques aux souches FL100 (signe sexuel a, ATCC28383) et FL200 (isogénique à FL100, de signe sexuel A) construites par F. Lacroute. La souche *ura2 15-30-72* (a ou A) est une cellule haploïde de génotype *ura2 15-30-72 trp1D GAL1*. La souche homozygote **a/a** *ura2 15-30-72* est une cellule diploïde issue du croisement des souches *ura2 15-30-72* de signe opposé.

1.1.3. Milieux de culture

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

- milieu complet YPG: 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 de sporulation Sp1: yeast extract 2,5 g/l; glucose 1 g/l; acétate de potassium 9,8 g/l
- milieu de sporulation AcK: acétate de potassium 10 g/l.

Ces milieux (mis à part le milieu complet) peuvent être supplémentés en fonction des auxotrophies des souches. Ils peuvent être solidifiés en présence de 2% d'agar.

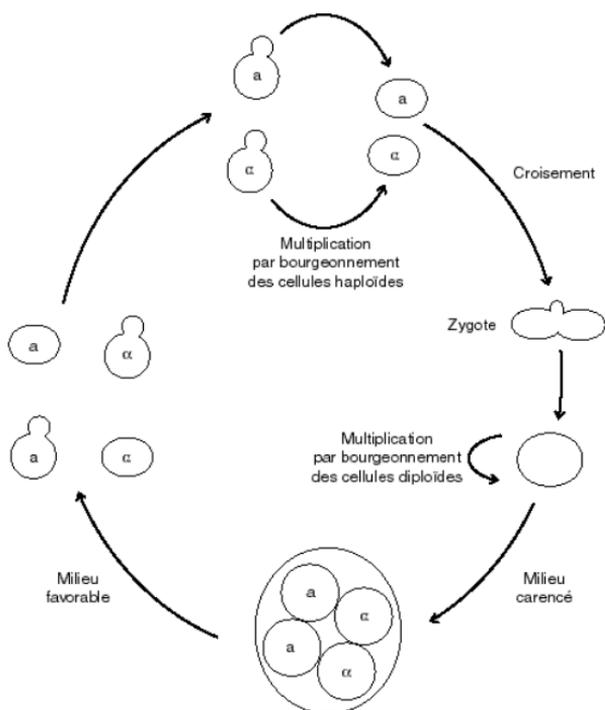


FIGURE M.1 – Cycle cellulaire de *Saccharomyces cerevisiae*.

1.1.4. Conditions de culture

Les cellules sont cultivées soit en milieu liquide sous agitation, soit sur 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.

1.2. *Escherichia coli*

1.2.1. Souche utilisée

Escherichia coli JM103 ((*lac-proAB*) *thi rpsL supE endA sbcB15 hsdR4 F'* [*traD36 proA⁺B⁺ lacF^ΔZ M15*]), hôte des plasmides de la série pUC (Messing *et al.*, 1981), a été utilisée pour la préparation de plasmides et lors de la détermination des régions flanquant une séquence d'ADN connue par la stratégie d'insertion / excision de plasmide.

1.2.2. Milieux de culture

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

- milieu complet LB : bactotryptone 10 g/l ; yeast extract 5g/l; NaCl 5g/l
- milieu SOC : bactotryptone 20 g/l ; yeast extract 5 g/l ; glucose 10 % ; NaCl (5 M) ; KCl (1 M) ; MgCl₂ (1 M); MgSO₄ (1 M).

Le milieu SOC liquide est utilisé lors de la transformation des bactéries, immédiatement après l'électroporation d'*E. coli* (Dower *et al.*, 1988).

Le milieu LB peut être solidifié en présence de 2 % d'agar . Par ailleurs, de l'ampicilline (100 µg / ml) peut être ajoutée au milieu de culture après autoclavage pour sélectionner les cellules résistantes.

1.2.3. Conditions de culture

En milieu liquide, les cultures d'*E. coli* sont réalisées à 37°C avec une agitation vigoureuse. Dans ces conditions, le temps de génération de la souche JM103 est d'environ 20 minutes dans le milieu complet LB.

1.3. Évaluation de la concentration des cultures

La concentration cellulaire d'une culture donnée a été déterminée en milieu

liquide par mesure de la turbidité de la culture à 600 nm dans un spectrophotomètre (l'appareil utilisé lors de ce travail est le BioPhotometer Eppendorf).

Une unité d'absorbance équivaut à :

- $7 \cdot 10^7$ cellules haploïdes de *S. cerevisiae*/ml
- $4 \cdot 10^7$ cellules diploïdes de *S. cerevisiae*/ml
- $9 \cdot 10^7$ cellules de *E. coli*/ml.

1.4. Conservation des souches

Les souches sont conservées par congélation à -80°C en présence d'une concentration finale de 15 à 20% de glycérol, qui permet de limiter la formation de cristaux à l'intérieur des cellules.

2. VECTEUR DE CLONAGE

Les vecteurs utilisés lors de cette étude sont des plasmides navettes *E. coli* / *S. cerevisiae* c'est-à-dire qu'il peuvent être utilisés chez les deux types d'organismes. Ces plasmides dérivent de pBR322 (Norrander *et al.*, 1983) et pUC19 (Bolivar *et al.*, 1977). Ce dernier contient l'origine de réplication de type ColE1 et le gène de résistance à l'ampicilline (Amp^{R}) du plasmide pBR322.

Les plasmides navettes se répliquent de manière indépendante chez *E. coli* et contiennent un gène de sélection de levure (*URA3*, *TRP1* ou *LEU2*) (Bonneaud *et al.*, 1991). Ils peuvent se répliquer chez *S. cerevisiae* s'ils contiennent l'origine de réplication 2μ ou une séquence de réplication autonome (ARS). Dans le premier cas, le vecteur sera présent à un haut nombre de copies (20 à 50 copies par cellule), tandis que dans le second, le nombre de copies sera réduit (1 à 3 par cellule). En l'absence de l'un de ces deux éléments, ces vecteurs seront de type intégratif.

Pour notre part, nous avons utilisé les plasmides pFL35 et pFL45 qui correspondent respectivement à un vecteur intégratif et à un vecteur réplcatif à haut nombre de copies (2μ). Au niveau de ces vecteurs, le marqueur de sélection chez *S. cerevisiae* est le gène *TRP1*.

Les vecteurs pFLD, pFLB et pFLE ont été construits au cours de ce travail à partir du plasmide pFL35. Les fragments de restriction *Bgl*III D, B et E du locus *URA2* (Fig. 2.2) ont été insérés dans le site *Bam*HI unique du plasmide pFL35.

3. TECHNIQUES GÉNÉTIQUES DE LA LEVURE

3.1. Obtention de cellules diploïdes

Deux souches de signe opposé, en phase de croissance exponentielle, sont mélangées par étalement sur milieu solide. Les zygotes de forme caractéristique apparaissent après 3-4 heures et sont facilement distinguables au microscope. Ils peuvent être isolés à l'aide d'un micromanipulateur Singer de type MSM.

3.2. Obtention de cellules haploïdes

Les cellules diploïdes en phase exponentielle de croissance sont étalées sur milieu SP1 et incubées 24 heures à 30°C avant d'être transférées sur milieu AcK. Les premiers asques apparaissent au bout de deux à trois jours à 30°C.

3.3. Détermination du signe sexuel

Les cellules dont on ignore le signe sont croisées avec des cellules haploïdes de signe connu a et a. L'apparition de zygotes dans l'un des cas permet d'identifier le signe de la souche testée.

3.4. Analyse des spores

Cette analyse a été réalisée par dissection de tétrades. Cette technique permet l'étude qualitative mais aussi quantitative des produits de la méiose (Sherman and Hicks, 1991). Une anse du mélange de sporulation est mise en suspension dans 100 µl d'eau stérile auxquels sont additionnés 100 µl de zymolyase à 5 mg/ml. Après 20 minutes à 30°C, la paroi des asques est totalement hydrolysée par l'enzyme. Les spores sont alors libérées, mais restent regroupées en tétrades. Elles sont ensuite alignées sur une boîte de milieu complet grâce à un micromanipulateur Singer de type MSM. Les clones issus des spores apparaissent au bout de 48 heures à 30°C et leur phénotype peut alors être analysé.

3.5. Détermination des phénotypes

Elle se fait en comparant la capacité de croissance de souches variées sur différents milieux sélectifs et ce 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 : une aliquote équivalente de chaque colonie est mise en suspension dans de l'eau stérile. Une goutte est alors déposée sur différents milieux sélectifs.

4. Techniques de transformation par électroporation

Lors de la transformation par électroporation, les cellules sont soumises à un champ électrique qui perturbe les structures membranaires, créant transitoirement des pores permettant l'entrée de l'ADN (Sukharev *et al.*, 1992).

4.1. Électroporation d'*E.coli* (Dower et al., 1988)

4.1.1. Préparation des cellules compétentes

250 ml de milieu LB sontensemencés avec 2,5 ml d'une préculture en phase stationnaire. La culture est incubée à 37°C avec agitation jusqu'à une absorbance à 660m de 0,5 à 1, puis elle est refroidie dans la glace pendant 15 à 30 minutes. Une centrifugation à 5500 rpm pendant 10 minutes à 0°C précède chacun des quatre lavages successifs des cellules dans des volumes décroissants : 250 ml d'eau froide, 125 ml d'eau froide, 80 ml de glycérol 10 % froid, puis finalement 0,3 ml de glycérol 10% froid.

4.1.2. Électroporation

40 µl de bactéries sont mélangés à un maximum de 4 µg d'ADN (dans un volume maximum de 2 µl de TE) dans une cuve d'électroporation froide, dont la distance entre électrodes est de 0,2 cm. Le mélange est soumis à une impulsion électrique (2,5 kV, 10 µF, 600 W) ; dans ces conditions, la constante de temps mesurée a une valeur de 4,6 à 4,8 ms. Les cellules sont immédiatement diluées dans 1 ml de milieu SOC, transférées en microtube Eppendorf stérile puis incubées à 37°C pendant 1 heure. Le mélange de transformation est éventuellement dilué ou concentré par centrifugation avant d'être étalé sur milieu sélectif. Les colonies apparaissent après 15 heures d'incubation à 37°C

4.2. Électroporation de *Saccharomyces cerevisiae*

Le principe est le même que celui de l'électroporation de bactéries, seuls les paramètres et les tampons utilisés changent.

4.2.1. Préparation des cellules compétentes

Les cellules sont cultivées dans 40 ml de milieu complet jusqu'en fin de phase exponentielle de croissance. Elles sont ensuite placées dans la glace pendant 10 min. Les cellules sont récupérées par centrifugation à 2500 rpm pendant 5 min à 4 °C. On lave ensuite les cellules avec 5 ml de sorbitol 1 M avant de reprendre le culot dans 200 µl de sorbitol 1 M. Les cellules sont placées dans la glace.

4.2.2. Électroporation

40 µl de cellules compétentes sont mélangés à un maximum de 4 µg d'ADN (dans un volume maximum de 2 µde TE) dans une cuve d'électroporation froide, dont la distance entre électrodes est de 0,2 cm. Le mélange est soumis à une impulsion (1,5kV, 10 µF, 600 W); dans ces conditions, la constante de temps mesurée a une valeur de 4,6 à 4,8 ms. Les cellules sont immédiatement diluées dans 1 ml de milieu complet Sorbitol, transférées en microtube Eppendorf stérile puis incubées à 30°C pendant 2 heures. Le mélange de transformation est éventuellement dilué ou concentré par centrifugation avant d'être étalé sur milieu sélectif. Les transformants sont incubés à 30 °C et les colonies apparaissent au bout de 2 à 3 jours.

5. PRÉPARATION D'ADN

5.1. ADN de levure

5.1.1. Extraction d'ADN génomique de levure (Hoffman and Winston, 1987)

L'ADN obtenu à l'aide de cette méthode n'est pas d'une pureté élevée, mais elle est néanmoins suffisante pour les analyses par hybridation ADN/ADN ainsi que pour les amplifications PCR.

Les cellules sont mises en culture dans 10 ml de milieu complet (YPG) liquide, à 30 °C sous agitation, jusqu'en début de phase stationnaire. Elles sont ensuite récoltées par centrifugation (5min à 3000rpm). Le culot est alors remis en suspension dans 1 ml d'eau distillée puis transféré dans un tube Eppendorf et centrifugé pendant 2 min à 12000 rpm. Les cellules sont ensuite reprises dans 250 µl de tampon de broyage (Tris-HCl pH 8 10 mM, EDTA 1 mM, NaCl 100mM, Triton 2%, SDS 1%) et on ajoute des billes de verre (de diamètre 0,5 mm) de manière à laisser environ 1 mm de milieu au-dessus des billes. On ajoute alors 250µl de phénol-chloroforme et le milieu est vortexé

pendant 3 min, puis centrifugé pendant 5 min à 12000 rpm. La phase aqueuse supérieure, transférée dans un nouveau tube Eppendorf, est additionnée d'un volume de phénol-chloroforme. Ce mélange est vortexé pendant 30 sec puis centrifugé à 12000 rpm pendant 2 min. La phase aqueuse supérieure est à nouveau transférée dans un nouveau tube Eppendorf et additionnée de 600 µl d'éthanol absolu froid et de 20 µl d'acétate d'ammonium 3 M. Cette solution est maintenue dans la glace pendant 15 min environ puis centrifugée pendant 5 min à 12000 rpm. Après séchage, le culot est repris dans 400 µl de TE (Tris-HCl 10 mM pH 8, EDTA 1 mM). On ajoute 30 µg de RNase A au milieu qu'on incube pendant 15 min à 37 °C. 1 ml d'éthanol absolu froid est alors mélangé à la solution qui est ensuite placée 15 min dans la glace, puis centrifugée pendant 10 min à 12000 rpm. Le surnageant est éliminé et le culot rincé avec 1 ml d'éthanol 70%. Après une nouvelle centrifugation, le culot est repris dans 50 µl de TE.

5.1.2. Préparation de chromosomes entiers de levure (Vézinhet et al., 1990)

L'analyse des caryotypes de souches de levure en électrophorèse en champs alternés nécessite d'obtenir des chromosomes intacts. Or les procédures habituelles de préparation d'ADN ne permettent pas d'obtenir des molécules de haute masse moléculaire. Ainsi, pour éviter les coupures dues aux forces mécaniques, les cellules entières incorporées dans des blocs d'agarose sont lysées et déprotéinées 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.

Les cellules sont cultivées dans 50ml de milieu YPG jusqu'en phase stationnaire et récoltées par centrifugation (5 min, 6000 rpm). Le culot est lavé dans de l'EDTA 50 mM pH8, puis les cellules sont incorporées dans des blocs d'agarose (BioRad). Ces derniers contenant les cellules et la zymolyase sont incubés 2 heures à 37°C dans un tampon CPE (acide citrique 40mM; Na₂HPO₄ 120mM; EDTA 20 mM) additionné de zymolyase à 100 µg/ml. Puis les cellules sont déprotéinées dans un tampon de lyse contenant la protéinase K (Tris 10mM pH8; EDTA 0,45 M; SDS 1 % ; 0,25-0,5mg/ml de protéinase K), une nuit à 50°C. Les blocs sont ensuite lavés 2fois 30min avec dH₂O à 50°C et 2 fois 30 min avec dH₂O à température ambiante. Les blocs sont alors repris dans 10 ml d'EDTA 0,25 M et stockés à 4°C.

5.2. ADN plasmidique d'*Escherichia coli*

Les souches d'*E. coli* possédant le gène *endA* (JM103 par exemple) produisent

une endonucléase pouvant détériorer l'ADN plasmidique. Cette endonucléase étant thermosensible, la méthode dite "de lyse des cellules à chaud" (Holmes and Quigley, 1981) est préconisée.

5 ml de milieu LB sont inoculés avec une colonie puis incubés à 37°C avec agitation pendant une vingtaine d'heures. Les cellules sont récoltées par centrifugation à 5000 rpm pendant 5 minutes et remises en suspension dans 350 µl de tampon STET (saccharose 8 %, Triton X-100 5 %, EDTA 50 mM, Tris-HCl 50 mM pH 8) auxquels on ajoute 30 µl d'une préparation extemporanée de lysozyme à 10 mg/ml. Les tubes sont immédiatement placés dans un bain-marie à 100°C pendant 50 secondes, puis centrifugés à 15000 rpm pendant 15 minutes. 200 µl de surnageant sont mélangés à 200µl d'isopropanol dans un microtube Eppendorf, incubés à -20°C pendant 10minutes, puis centrifugés à 15000 rpm pendant 5 minutes. Le culot est bien égoutté et remis lentement en solution dans 1 ml d'éther. Après une centrifugation à 15000 rpm pendant 5 minutes, le culot est séché et remis en solution dans 50 µl d'eau distillée stérile. Cette solution peut être conservée à -20°C. Dans le cas d'une digestion ultérieure par un enzyme de restriction, on peut ajouter 2,5 % (v/v) d'une solution de RNase H à 2,5mg/ml.

6. Analyse de l'ADN

6.1. Purification de l'ADN par extraction phénol-chloroforme

Le but de cette extraction est l'élimination des protéines contenues dans une solution d'ADN. Pour un volume de solution d'ADN, on ajoute un volume de phénol-chloroforme (le produit utilisé est commercialisé par Q-BIOgene et constitué d'un mélange d'AquaPhenol™, de chloroforme et d'alcool isoamylique dans les proportions respectives: 50 % / 47,5% / 2,5%). Le mélange est alors fortement agité à l'aide d'un vortex puis centrifugé 5 min à 12000rpm. Il apparaît alors deux phases: une phase phénolique inférieure et une phase aqueuse supérieure, les protéines se trouvant à l'interface. La phase aqueuse supérieure contenant l'ADN est récupérée par pipetage. L'extraction ainsi réalisée peut être répétée plusieurs fois afin d'augmenter la pureté de l'ADN.

6.2. Précipitation alcoolique

La solution d'ADN purifiée peut être concentrée par précipitation alcoolique. Pour ceci, on ajoute un dixième de volume d'acétate de sodium 3 M pH 7 et 2,5 volumes d'éthanol pur (stocké à -20°C); l'ADN précipite en 30minutes à 0°C, et une cen-

trifugation à 15000 rpm pendant 15 à 30 minutes permet de l'obtenir sous la forme d'un culot. Après lavage du culot avec de l'éthanol à 70 % (stocké à -20°C), l'ADN est dissous dans de l'eau distillée ou du tampon TE stérile. Cette solution peut être conservée pendant plusieurs mois à -20°C.

6.3. Purification de l'ADN par chromatographie

Cette méthode de purification, très rapide, a été utilisée pour séparer l'ADN double brin amplifié par PCR des autres substances présentes dans le milieu réactionnel après amplification (amorces en excès, dNTP, sels). L'ADN purifié peut ensuite être utilisé pour d'autres réactions : PCR, digestion par un enzyme de restriction, ligation, séquençage...

Cette méthode de chromatographie d'exclusion (gel-filtration) est basée sur la séparation des molécules selon leur taille dans une résine de Séphacryl sous l'action d'une force centrifuge. Les impuretés (dNTP, sels) et les petits fragments d'ADN (amorces de taille inférieure à 25-mer) sont retenus par la résine, alors que les grands fragments d'ADN (500-mer et davantage) passent rapidement à travers. Les colonnes utilisées (MicroSpin™ S-400 HR), préalablement équilibrées dans du tampon TE, sont commercialisées par Pharmacia.

La colonne est d'abord remise en suspension, placée dans un microtube Eppendorf et centrifugée à 2300 rpm pendant 1 minute pour éliminer l'excès de tampon. Après avoir placé la colonne dans un nouveau microtube Eppendorf, 25 à 50 µl d'échantillon sont déposés délicatement à la surface de la résine, puis centrifugés à 2300 rpm pendant 2 minutes. La solution d'ADN purifiée est récupérée dans le microtube.

6.4. Détermination de la concentration d'ADN

La concentration d'une solution d'ADN peut être estimée par la mesure de l'absorbance à 260 nm, une unité d'absorbance correspondant à 50 µg/ml d'ADN bicaténaire. Le calcul du rapport des absorbances à 260 nm et 280 nm permet d'estimer la pureté de la solution (absence de protéines), ce rapport devant être compris entre 1,8 et 2 (Sambrook *et al.*, 1989).

6.5. Analyse par électrophorèse sur gel d'agarose

6.5.1. Électrophorèse sur gel d'agarose classique

Les fragments d'ADN de taille inférieure à 50 kb sont séparés selon leur taille

par migration sur gel d'agarose 1%. Ces fragments sont visibles aux UV à 300 nm grâce au BET contenu dans le gel qui s'insère entre les plateaux de base de l'ADN. On peut en estimer la taille et la concentration en comparaison avec la comigration d'un marqueur de taille tel que de l'ADN de phage 1 digéré par *EcoRI* et *HindIII*. La migration s'effectue sous une tension pouvant varier de 20 à 100V dans un tampon d'électrophorèse, le TAE (Tris-Base 4,84 g ; acétate de sodium 0,68 g ; EDTA 0,336 g ; acide acétique 1,35 ml; eau qsp 1l).

6.5.2. Électrophorèse en champs alternés (ou PFGE)

Cette technique permet de séparer des molécules d'ADN dont la taille varie entre 200 et 1500 kb. Elle permet ainsi de visualiser les chromosomes de *S.cerevisiae*, dont la taille est comprise entre 240 kb et 1600 kb.

Le principe repose sur le fait qu'en alternant le champ électrique entre les paires d'électrodes spatialement distinctes, les molécules d'ADN se réorientent différemment à travers les réticulations d'un gel d'agarose et migrent en fonction de leur taille. Dans le système CHEF (Clamped Homogeneous Electric Field) (Chu *et al.*, 1986), 24 électrodes sont positionnées en hexagone. Nous avons utilisé un appareil CHEF-DRIII de BioRad. La migration des chromosomes est liée au type d'agarose et à sa concentration, à la tension appliquée, au temps de migration et au temps au cours duquel le champ électrique suit une certaine direction (« pulse »). Pour ce travail, nous avons appliqué des conditions permettant une bonne séparation globale des chromosomes. Ainsi, les migrations ont été effectuées dans un gel d'agarose 1% sur une durée allant de 22 à 26 heures à 6 V/cm avec des pulses de 40 et 120sec et un angle variant de 120°.

6.6. Digestion par des endonucléases de restriction

Les enzymes de restriction sont des endonucléases qui coupent l'ADN bicaténaire au niveau d'une séquence spécifique, généralement sous la forme d'un palindrome ; la coupure produit des extrémités franches ou des extrémités cohésives 5' ou 3' sortantes. Ces enzymes sont utilisés selon les conditions spécifiées par le fabricant (tampon, température, présence de BSA, concentration...).

La digestion est réalisée en microtube Eppendorf contenant, dans un volume final de 20 µl, un dixième de volume de tampon dix fois concentré, quelques unités d'enzyme, et de 0,1 à 1 µg d'ADN. Pour une digestion simultanée par plusieurs endonucléases, si les conditions sont différentes pour les enzymes utilisés, on peut faire

se succéder les tampons en commençant par le moins concentré en sels. Cependant, il est souvent préférable de purifier l'ADN par extraction phénolique et précipitation afin d'éliminer enzyme et sels ; les endonucléases peuvent en effet couper l'ADN de façon non spécifique dans des conditions non optimales.

6.7. Ligation des molécules d'ADN

Cette manipulation utilise la DNA ligase du phage T4. Elle permet la construction *in vitro* de molécules d'ADN recombinées.

Pour insérer un fragment d'ADN dans un plasmide linéarisé, il est recommandé de se placer en condition d'excès de l'insert par rapport au vecteur (rapport molaire 3:1). Certains paramètres de la réaction varient selon le type d'extrémités. Dans un volume réactionnel de 20µl, l'ADN total doit représenter 0,01 à 0,1µg dans le cas d'une ligation d'extrémités cohésives (0,1 à 1 µg dans le cas d'extrémités franches). On y ajoute 0,1 unité de ligase du bactériophage T4 (1 unité dans le cas d'extrémités franches) que l'on incube à 24°C pendant 1 heure (à 14°C pendant 24 heures dans le cas d'extrémités franches).

7. AMPLIFICATION D'UN FRAGMENT D'ADN PAR PCR (POLYMERASE CHAIN REACTION)

7.1. PCR classique

7.1.1. Principe

La PCR est une technique permettant l'amplification exponentielle et spécifique d'un fragment d'ADN *in vitro* par l'utilisation d'une polymérase thermostable, d'oligonucléotides amorces et des quatre dNTP, que l'on soumet à une série de cycles de température. Plusieurs cycles alternant une dénaturation de l'ADN, l'hybridation des oligonucléotides et la polymérisation de l'ADN sont ainsi effectués. La durée et la température des différentes étapes sont des paramètres essentiels qui sont définis pour chaque expérience.

7.1.2. Paramètres importants lors des différentes étapes

La dénaturation permet la formation d'ADN simple brin. La température utilisée varie peu selon les protocoles (92 à 94°C). Sa durée (30 secondes) est souvent plus longue que nécessaire car quelques secondes suffisent à obtenir une bonne dénaturation.

L'hybridation est l'étape déterminant la spécificité et le rendement de l'amplification. La température d'hybridation (T_m) se détermine en fonction de la taille et de la

séquence des oligonucléotides utilisés. Cette T_m est la température à laquelle la moitié des hybrides possibles est formée. Le temps d'hybridation est généralement de 30 secondes à une minute. La concentration saline du tampon agit sur la stringence de l'hybridation des oligonucléotides. La concentration de base en $MgCl_2$ est de 1mM. Une concentration supérieure augmente l'hybridation aspécifique des oligonucléotides.

La polymérisation est généralement effectuée à une température légèrement inférieure à la température optimale de l'ADN polymérase utilisée. La durée de polymérisation dépend de la taille du fragment d'ADN à amplifier. Il faut compter 30 secondes à 1 minute par kb à amplifier avec la *Taq* DNA polymérase.

Le nombre de cycles, quant à lui, dépend de la quantité de matrice de départ. Il faut en général entre 25 et 35 cycles pour obtenir 1 mg d'ADN pour 100 μ l de réaction à partir de quelques pg d'ADN. Une étape de post-élongation peut être effectuée après la fin des cycles pour terminer la synthèse de certains brins éventuellement inachevés.

7.1.3. Matériel

Les PCRs se font dans des tubes Eppendorf de 600 μ l de volume. Le volume du milieu de réaction varie de 50 à 100 μ l.

Appareil à PCR

La réaction se fait dans un appareil automatisé, le Mastercycler™ Gradient d'Eppendorf. Il dispose d'un couvercle chauffant qui est porté à 105°C lors de son utilisation.

Enzyme

La *Taq* DNA polymérase commercialisée par Q-BIOgene est fournie avec un tampon x10 dont la constitution est : Tris-HCl pH 8,3 100mM; KCl 500 mM; $MgCl_2$ 15 mM ; gélatine 0,01 %.

La composition du milieu de réaction est: 10 μ l de tampon 10x; 1 μ l d'une solution des 4 dNTPs à 20 mM ; 1 μ l de chaque oligonucléotide à 20 μ M; environ 100 ng d'ADN ; 0,2 unité d'enzyme; eau bidistillée qsp 100 μ l.

7.2. Tail - PCR (Thermal asymmetric interlaced –PCR)

La Tail-PCR est une technique permettant de cloner les séquences génomiques flanquant une séquence d'ADN connue. Elle a été développée pour déterminer les séquences flanquantes dans le cas des gènes étiquetés par une insertion d'ADN-T chez les plantes (Liu *et al.*, 1995).

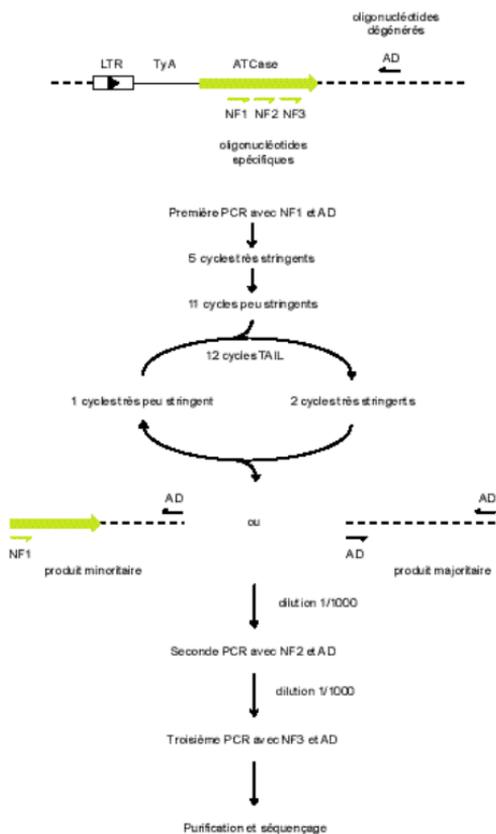


FIGURE M.2 – Diagramme schématisé de la technique de Tail-PCR.

(A)

Réactions	Nombre de cycles	Conditions
Première PCR	1	92°C (2min)
	5	94°C (15s), 88°C (8min)
	2	94°C (15s), 46°C (2min), 70°C (8min)
	12	94°C (15s), 88°C (8min)
		94°C (5s), 88°C (8min)
		94°C (5s), 46°C (30s)
	1	70°C (8min)
Deuxième PCR	1	94°C (30s)
	15	94°C (15s), 88°C (8min), 94°C (15s), 88°C (8min)
		94°C (15s), 46°C (1min)
	1	70°C (8min)
Troisième PCR	20	94°C (5s), 44°C (1min), 88°C (2min)
	1	88°C (5min)

(B)

Réactions	Nombre de cycles	Conditions
Première PCR	1	92°C (2min), 99°C (1min)
	5	94°C (15s), 63°C (1min), 72°C (2min)
	1	94°C (15s), 30°C (3min), 72°C (2min)
	10	94°C (5s), 44°C (1min), 72°C (2min)
	12	94°C (5s), 44°C (1min), 72°C (2min)
	1	72°C (8min)
	Deuxième PCR	1
16		94°C (15s), 63°C (1min), 72°C (2min), 94°C (15s), 63°C (1min), 72°C (2min)
		94°C (15s), 44°C (1min), 72°C (2min)
1		72°C (8min)
Troisième PCR	20	94°C (15s), 44°C (1min), 72°C (2min)
	1	72°C (8min)

TABLEAU M.1 – Paramètres d'amplification utilisés pour la technique de Tail-PCR.
(A) technique classique décrite par Liu et al. (1995).

(B) technique de Long-Range Tail-PCR décrite par Szabados et al. (2002).

Cette stratégie se base sur l'utilisation d'oligonucléotides spécifiques à la séquence d'ADN connue ainsi qu'une série d'oligonucléotides dégénérés (Figure 2). On parle de PCR « héli-spécifique ». Une telle PCR va obligatoirement entraîner la formation de produit aspécifique. Cependant, afin d'amplifier préférentiellement la région d'intérêt, 3réactions PCR successives sont réalisées avec des conditions de stringences différentes. De plus, pour chaque réaction, un autre oligonucléotide spécifique à la région connue est utilisée.

Dans le cadre de notre étude, cette technique nous a permis de déterminer les sites d'insertions 3' des événements de duplication génique. Pour l'amplification de fragments de petite taille (inférieur à 2 kb), la technique de Tail-PCR classique décrite par Liu *et al.* (1995) a été réalisée en utilisant la Taq DNApolymérase (Q-Biogene). Les différents paramètres d'amplification utilisés sont indiqués dans le tableau 1A. Par contre, pour l'amplification de fragments de grande taille (3 à 15kb), la technique de Long-Range Tail-PCR décrite par Szabados *et al.* (2002) a été appliquée en utilisant la Long Expand polymérase commercialisée par Roche et les conditions de réaction décrites dans le tableau 1B.

8. 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. À partir d'une amorce oligonucléotidique hybridée sur la matrice d'ADN à séquencer, une ADN polymérase synthétise un brin complémentaire. Cette synthèse est effectuée en présence des quatre 2'-désoxyribonucléoside-5'-triphosphates (dATP, dCTP, dGTPet dTTP) et des quatre 2',3'-didésoxyribonucléoside-5'-triphosphates (ddATP*, ddCTP*, ddGTP* ou ddTTP*) marqués chacun par un fluorochrome distinct. 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 désoxynucléotide ou un didésoxynucléotide. Dans le cas où elle incorpore un désoxynucléotide, la synthèse peut continuer, dans le cas contraire elle s'arrête. Ce choix étant aléatoire, chaque base de l'ADN matrice aura statistiquement vu un certain nombre de fois l'incorporation d'un didésoxynucléotide, si bien que le milieu réactionnel contient l'ensemble des molécules néosynthétisées possibles. Ces molécules sont ensuite dénaturées puis par électrophorèse capillaire afin d'être séparées 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 automatique ABI PRISM® 373XL (PE Applied Biosystems). Les réactions de séquençage ont été réalisées en utilisant le kit PRISM AmpliTaq® FS Big Dye Terminator.

9. HYBRIDATION ADN/ADN (SOUTHERN, 1975)

L'hybridation ADN-ADN appliquée à des fragments séparés par électrophorèse a été décrite initialement par Southern (1975) ; depuis, cette technique est couramment appelée Southern blot ou Southern. Une expérience d'hybridation comporte deux étapes : le transfert de l'ADN sur un support solide (généralement une membrane de nylon ou de nitrocellulose) et l'hybridation proprement dite.

9.1. Transfert

Dans la technique du chromo-blot, le transfert de l'ADN localisé sur le gel vers une membrane de nylon Hybond N⁺ (Amersham) est nécessaire. Un dispositif par l'intermédiaire du phénomène de capillarité permet de réaliser ce transfert. Mais avant de procéder au transfert proprement dit, les molécules d'ADN présentes dans le gel d'agarose sont soumises :

- à une dépurination par une solution HCl 0,25 M (pour fragmenter les molécules de hauts poids moléculaires)
- et à une dénaturation de l'ADN double brin par une solution NaOH 0,5 M; NaCl 1,5M

Dans le cadre du transfert de molécules de hautes masses moléculaires (>200kb) comme les chromosomes de *S.cerevisiae*, le transfert est réalisé avec du tampon de dénaturation sur la nuit. La membrane est ensuite neutralisée avec une solution Tris-HCl 0,5 M pH 7,5; 1MNaCl. L'ADN est ensuite fixé par incubation à 80°C pendant 1 heure.

9.2. Préparation d'une sonde froide

Cette technique présente l'avantage d'éviter l'utilisation de produits radioactifs. Nous avons employé un kit de marquage aléatoire utilisant des dNTP marqués à la digoxigénine (DIG Labeling and Detecting Kit, Roche). Le protocole utilisé est celui décrit par le fournisseur. Après précipitation alcoolique, le culot d'ADN marqué est repris dans 50 µl d'eau puis testé suivant le protocole indiqué.

9.3. Hybridation et lavages

La préhybridation se fait pendant une heure à 68°C dans du tampon d'hybridation (12,5ml de SSC 20fois concentré; 100µl de SDS 10 % ; 0,5 g de réactif bloquant ; 50 mg de N-lauryl sarcosine). L'hybridation se fait dans le même tampon, en présence de la sonde dénaturée, à 68°C pendant une nuit.

Afin d'éliminer l'excès de sonde, la membrane est lavée deux fois dans tampon 2xSSC ; SDS 1 % à température ambiante puis deux fois dans du tampon 0,1xSSC ; SDS 1 % à 68°C sous agitation douce.

9.4. Révélation au NBT/BCIP

Toutes les étapes se déroulent à température ambiante. La membrane est équilibrée pendant 1 min dans du tampon de lavage (Tris-HCl 100 mM pH 7,5 ; NaCl 150mM), puis incubée durant 30min dans un tampon de saturation (Tris-HCl 100mM pH 7,5 ; NaCl 150mM ; lait en poudre 1/2écrémé 5%) permettant de saturer les sites aspécifiques aux anticorps anti-DIG-AP. Elle est enfin placée dans un milieu de tampon de saturation (20 ml pour une membrane de 100 cm²) additionné d'anticorps (150mU/ml final) pendant 30min. Elle est ensuite lavée deux fois pendant 15 min dans le tampon de lavage, puis équilibrée 2 min dans la solution de révélation (Tris-HCl 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 NBT/BCIP, jusqu'à apparition de la coloration (en général de 2 à 16 heures). La révélation au CDP-Star se fait dans les mêmes conditions, le NBT/BCIP étant remplacé par ce substrat pour la révélation et la membrane étant exposée à l'obscurité contre un film d'autoradiographie.

10. HYBRIDATION GÉNOMIQUE COMPARATIVE SUR PUCE À ADN

La technique de la CGH (Comparative Genomic Hybridization) permet de détecter les variations du nombre de copies d'ADN telles que les duplications et les délétions de régions chromosomiques associées à des pathologies du développement ou à des cancers. Dans la CGH classique (Kallioniemi *et al.*, 1994), les ADN génomiques d'un échantillon à tester (tissu tumoral) et d'un témoin de référence (tissu normal) sont extraits, marqués par les colorants fluorescents et hybridés sur un étalement de chromosomes métaphasiques, afin de permettre l'observation de remaniements chromosomiques.

L'hybridation sur des puces à ADN représentant des portions connues du génome (appelées couramment « CGH-array ») a permis, ces dernières années, de cartographier les réarrangements génomiques avec une résolution beaucoup plus élevée (Pollack *et al.*, 1999; Pinkel *et al.*, 1998). Cette approche a été utilisée avec succès dans des études fondamentales chez la levure telles que :

- l'analyse de la distribution et de la dynamique d'activation des origines de réplication (Raghuraman *et al.*, 2001),
- la caractérisation des réarrangements génomiques au cours de l'évolution de souches de levure soumises à une forte pression de sélection (Dunham *et al.*, 2002).

Au cours de notre étude, les expériences d'hybridation génomique comparative sur puces à ADN ont été réalisées par la plateforme Affymetrix du Génopole Alsace-Lorraine (IGBMC Illkirch, France).

10.1. Puces à ADN utilisées

Les puces à ADN utilisées correspondent au GeneChip[®] Yeast Genome S98 Array commercialisé par Affymetrix. Ce sont des puces à oligonucléotides courts (20-25 mers) synthétisés *in situ*. Environ 6 400 ORFs de la levure *Saccharomyces cerevisiae* sont représentées sur cette puce.

10.2. Extraction de l'ADN génomique

L'extraction d'ADN des souches de levure analysées a été réalisée grâce au kit Qiagen Genomic-Tip 100/G. Le protocole utilisé est celui décrit par le fournisseur.

La concentration d'une solution d'ADN a été estimée par la mesure de l'absorbance à 260 nm sur le BioPhotometer (Eppendorf), une unité d'absorbance correspondant à 50 µg/ml d'ADN bicaténaire. Le calcul du rapport des absorbances à 260nm et 280nm permet d'estimer la pureté de la solution (absence de protéines), ce rapport devant être compris entre 1,8 et 2.

10.3. Marquage de l'ADN par amorçage aléatoire

Les ADNs extraits ont été soumis à un marquage fluorescent à la biotine-streptavidine à l'aide du kit BioPrime DNALabeling System (Invitrogen).

Le marquage est réalisée selon la technique d'amorçage aléatoire (random priming): en présence de dATP, dGTP, dUTP et biotine-14 - dCTP, la polymérase de Klenow synthétise de l'ADN double brin marqué à la biotine en utilisant comme matrice l'ADN de la

région à marquer et comme amorce un mélange d'hexanucléotides.

500ng d'ADN linéaire sont dénaturés par chauffage à 95°C pendant 10 minutes, puis incubés dans la glace pendant 5 minutes. Dans un volume final de 50 µl, on ajoute les différents composants du kit : 20 µl du mélange d'hexanucléotides, 5 µl de dNTP marqués et 1 µl de l'enzyme de Klenow (2 unités). La polymérisation est réalisée au cours d'une incubation à 37°C pendant 4 heures. La réaction est arrêtée par addition de 5 µl de tampon d'arrêt. L'ADN marqué est ensuite purifié par précipitation à -20°C en présence de 1/10 de volume d'acétate de sodium 3M et 2,5 volumes d'éthanol absolu. Le culot est lavé à l'éthanol 70 % et remis en solution dans 55 µl de TE.

10.4. Hybridation et lavages

La sonde marquée est incorporée dans 200 µl de milieu d'hybridation (0,005 % Triton X100, 6X SSPE) puis dénaturée par chauffage à 95°C et incubée dans la glace. Pendant ce temps, les puces à ADN sont préhybridées dans le tampon 6X SSPE. Pour l'hybridation, le tampon de préhybridation est remplacé par 200 µl de tampon d'hybridation à 68°C contenant la sonde dénaturée. Les puces sont incubées pendant 2 heures à 42°C sous rotation (60 rpm).

10.5. Lavage et coloration

Les lames sont par la suite lavées avec la station de lavage Affymetrix. Les 20 premiers lavages sont réalisés dans un tampon non stringent (6X SSPE, 0,01 % Tween 20) à 25°C. Alors que les 40 lavages suivants sont réalisés dans un tampon stringent (10 mM MES, 1MNA⁺, 0,01 % Tween 20) à 50%.

Une fois lavées, les puces sont colorées dans une première solution (Tampon de coloration Affymetrix, BSA acétylée 2 mg/ml, Streptavidine phycoérythrine 10 µg/ml) puis dans une deuxième solution (Tampon de coloration Affymetrix, BSA acétylée 2 mg/ml, IgG 0,1 mg/ml, Anticorps anti-streptavidine biotinylé 3 µg/ml) pendant 10 min à 25°C.

10.6. Scanning et analyse des données

Finalement, chaque puce est scannée deux fois avec le scanner GeneArray (Agilent Technologies). L'image finale correspond à la moyenne des 2 passages au scanner. L'intensité de l'hybridation mesurée pour chaque ORF est proportionnelle au nombre de copies dans une souche donnée.

Dans notre étude, cette technique a permis de détecter le changement du nombre de copies de gènes (délétion ou duplication) des révertants par rapport à la souche ancestrale ayant permis de les sélectionner. En effet, pour une ORF donnée, si le ratio de l'intensité du signal observé pour le révertant haploïde par rapport à celui observé pour la souche ancestrale est de 1 alors une seule copie est présente dans les deux souches. Par contre, si le ratio est de 2 ou de 1/2 alors cela signifie que le révertant possède respectivement une copie dupliquée ou une copie délétée de l'ORF concernée. Pour la simplification de la représentation graphique, les résultats ont été mis sous la forme du \log_2 du ratio. En effet, les \log_2 de 2 (duplication), 1/2 (délétion) et 1 (aucun changement) sont égaux à 1, -1 et 0.

11. INFORMATIQUE

La recherche de séquences d'acides nucléiques ou de protéines par similitude dans les banques de séquences sur Internet (EMBL, Genbank ou Swiss-Prot, par exemple) a été réalisée à l'aide des programmes BLAST (Altschul *et al.*, 1990) ou FASTA (Pearson and Lipman, 1988). Ces programmes utilisent une matrice de valeurs de substitution d'acides aminés ou de nucléotides (BLOSUM ou PAM, par exemple).

L'analyse et les comparaisons de séquences nucléotidiques ont été réalisées à l'aide du logiciel DNAStriderTM 1.3.

Les banques de données utilisées sont :

- SGDTM (*Saccharomyces* Genome Database, Stanford, USA)
<http://genome-www.stanford.edu/Saccharomyces/>
- GenBank (NCBI, Bethesda, Maryland, USA)
<http://www.ncbi.nlm.nih.gov/>
- Génolevures
<http://cbi.labri.fr/Genolevures/>
- MIT (Massachusetts Institute of Technology)
http://www.broad.mit.edu/annotation/fungi/comp_yeasts/
- *Saccharomyces* Genome Sequencing at the GSC (Genome Sequencing Center, St. Louis, USA)
<http://www.genome.wustl.edu/projects/yeast/>
- AGD (Ashbya Genome database, Suisse)
<http://agd.unibas.ch/>

1

SYSTÈME DE SÉLECTION ET CARACTÉRISATION MOLÉCULAIRE DES DUPLICATIONS

RÉSUMÉ

Ces dernières années, le séquençage systématique d'un certain nombre d'organismes a permis de mettre en évidence l'importance quantitative des séquences dupliquées au sein des génomes. Malheureusement, leur analyse ne permet pas d'accéder aux mécanismes qui sont impliqués dans la genèse de tels événements. C'est pourquoi, il est nécessaire d'élaborer des cribles de sélection permettant d'isoler puis d'analyser de tels remaniements. Au laboratoire, nous possédons un tel système, basé sur un allèle mutant du gène URA2, permettant la sélection positive et spontanée de remaniements chromosomiques dont des duplications. Ce système a permis la sélection de 8 révertants haploïdes qui possèdent une duplication d'une partie du gène URA2. Afin de caractériser plus finement sur le plan moléculaire les remaniements sélectionnés, nous avons entrepris la détermination de 2 caractéristiques importantes: la localisation chromosomique et la taille des régions dupliquées. Les résultats montrent que la copie dupliquée peut s'insérer aussi bien sur le chromosome X porteur de l'allèle mutant ura2 15-30-72 que sur un autre chromosome. La détermination de la taille des régions dupliquées par une stratégie de marche sur le chromosome met en évidence deux types d'événements de duplication. Chez 4 révertants, la duplication est limitée à une partie du gène URA2 et constitue de la duplication génique. Par contre, chez les 4 autres révertants, l'événement correspond à un segment englobant la séquence codant le domaine ATCase ainsi qu'une série de gènes contigus et constitue de la duplication segmentale.

INTRODUCTION

Depuis la fin des années 90, l'étude de l'évolution des génomes a pu être abordée de manière différente grâce aux programmes de séquençage systématique d'un certain nombre d'organismes. L'analyse de ces séquences met en évidence l'importance quantitative des séquences d'ADN dupliquées au sein des génomes. En effet, tous les organismes séquencés à ce jour montrent des traces de duplications ancestrales ou récentes (Rubin *et al.*, 2000). Ces observations suggèrent fortement que les événements de duplication ont un rôle clé dans l'évolution moléculaire.

Chez *Saccharomyces cerevisiae*, l'analyse de la séquence complète du génome a révélé la présence de blocs de duplication qualifiés de « fossiles », c'est-à-dire correspondant aux traces d'un événement de duplication ancestrale. Deux théories ont été proposées pour expliquer cette redondance de séquences. Wolfe *et al.* (1997) ont émis l'hypothèse qu'il se serait produit une duplication totale du génome, très ancienne, suivie d'événements de translocations et de délétions de certains fragments chromosomiques. Au contraire, Llorente *et al.* (2000) et Kozsul *et al.* (2004) proposent qu'une série d'événements indépendants de duplication localisés dans différentes régions du génome se serait déroulée au cours de l'évolution menant à la situation actuelle. Les données du séquençage récent des levures *Kluyveromyces waltii* et *Ashbya gossypii* ont permis de confirmer sans ambiguïté que les blocs de duplication présents chez *Saccharomyces cerevisiae* sont issus d'une duplication complète de son génome (Kellis *et al.*, 2004 ; Dietrich *et al.*, 2004). En effet, la comparaison de ces 2 espèces avec *S. cerevisiae* permet d'analyser la situation avant et après la duplication complète du génome. L'analyse montre que deux segments dupliqués et mosaïques formant un bloc de duplication chez *S. cerevisiae* correspondent effectivement à une région unique chez *K. waltii* et *A. gossypii*.

Cependant, toutes les séquences d'ADN dupliquées dans le génome de *Saccharomyces cerevisiae* ne sont pas issues de l'événement de polyploïdie. Il existe notamment des gènes paralogues n'appartenant pas à un bloc « fossile » de duplication comme par exemple les gènes *HXT* codant les transporteurs d'hexoses. De plus, expérimentalement, il a été mis en évidence qu'une duplication peut concerner un seul gène mais aussi un segment en portant plusieurs, un bras ou encore un chromosome

entier. L'apparition de ces gènes implique donc d'autres processus, différents d'une duplication complète de génome. Contrairement à cette dernière, les autres duplications sont beaucoup plus difficiles à détecter au sein des génomes séquencés. Par exemple, la copie dupliquée d'un gène est souvent effacée par l'accumulation de mutations ponctuelles qui peuvent la rendre indétectable. Ainsi, une duplication de gènes ou de segments ne portent plus forcément les « cicatrices » du remaniement ayant conduit à sa formation dans les génomes. L'analyse de ces séquences ne permet donc pas de proposer des modèles quant aux mécanismes impliqués dans de tels réarrangements chromosomiques.

C'est pourquoi, l'identification des mécanismes responsables de ces remaniements génomiques nécessite de disposer d'un système expérimental performant qui permet de les isoler facilement. Dans un premier temps, un tel outil permet d'analyser des événements de duplication récents ayant encore les « cicatrices » du remaniement et de proposer des hypothèses de mécanismes impliqués dans leur formation. Dans un deuxième temps, ces hypothèses peuvent être vérifiées par une sélection de nouveaux remaniements dans un contexte affectant les mécanismes proposés.

1. LE SYSTÈME DE SÉLECTION

Au laboratoire, nous possédons un système génétique permettant la sélection positive de réarrangements chromosomiques chez *Saccharomyces cerevisiae*. Ce système est basé sur l'utilisation d'un allèle mutant particulier du gène *URA2*. Situé sur le chromosome X, ce gène code une protéine multifonctionnelle de 2214 acides aminés, intervenant dans les deux premières étapes de la voie de biosynthèse des nucléotides pyrimidiques (Fig.1.1). Cette protéine est constituée, à partir de l'extrémité N-terminale, d'un domaine actif GATase-CPSase (glutamine amidotransférase – carbamylphosphate synthétase), d'un domaine cryptique DHOase-like et d'un domaine actif ATCase (aspartate transcarbamylase).

L'allèle mutant *ura2 15-30-72* utilisé pour les sélections porte deux mutations non-sens dans la région codant le domaine GATase et une insertion d'un nucléotide provoquant un décalage de la phase de lecture dans la région codant le domaine CPSase (Fig.1.2). Bien que la séquence ATCase soit sauvage dans une telle souche, ce domaine

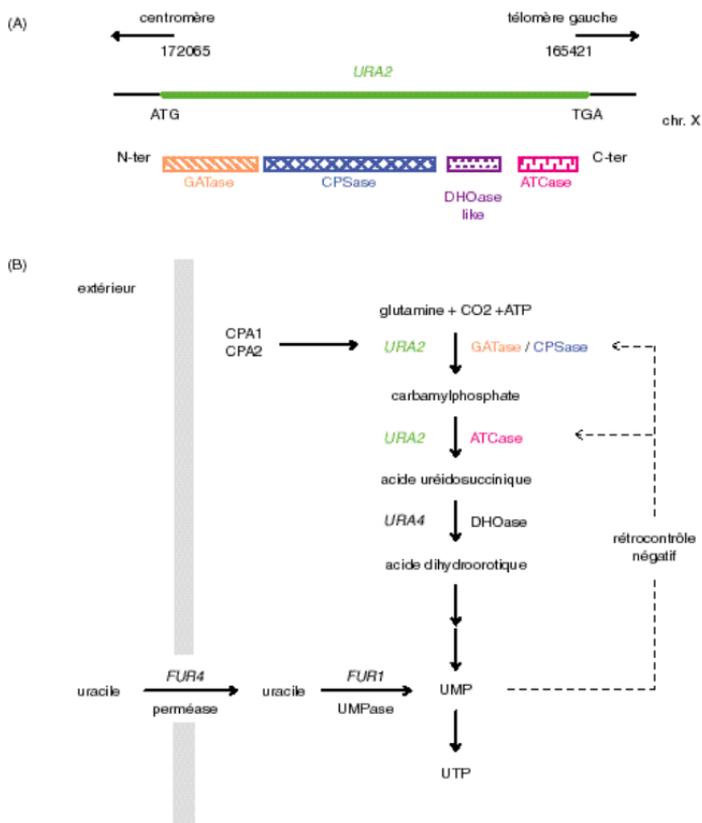


FIGURE 1.1 – Le gène *URA2* (A) et les étapes de la voie de biosynthèse de novo des nucléotides pyrimidiques (B). *GATase*: glutamine amidotransférase; *CPSase*: carbamyl-phosphate synthétase; *DHOase-like*: dihydroorotase-like; *ATCase*: aspartate transcarbamylase. Les chiffres indiquent les coordonnées du gène *URA2* (YJL130c) sur le chromosome X de la souche S288C. *CPA1* et *CPA2* sont les deux gènes qui assurent la synthèse du carbamylphosphate dans la chaîne de biosynthèse de l'arginine.

n'est pas actif en raison de la présence des trois mutations situées en amont. En effet, ces mutations conduisent à l'arrêt prématuré de la traduction de la protéine Ura2p. De plus, il n'y a pas de possibilité d'initiation interne de la transcription ou de la traduction permettant de contourner cet obstacle. Par contre, la fonction CPSase est complétée par l'activité CPSase intervenant dans la voie de biosynthèse de l'arginine. Ce mutant *ura2 15-30-72* est donc de phénotype CPSase⁺ ATCase⁻ ce qui le rend auxotrophe pour l'uracile (Ura⁻). Dans un tel contexte, tous les événements conduisant à la réactivation fonctionnelle du domaine ATCase rendent la souche prototrophe pour l'uracile (Ura⁺). La présence de trois mutations ponctuelles rend négligeable la probabilité de trois réversions spontanées.

La souche haploïde *ura2 15-30-72* a été utilisée pour la sélection de révertants spontanés Ura⁺ capables de se développer sur milieu minimum, contrairement à la souche de départ. La fréquence d'obtention de ces révertants est faible et peut être estimée à 10⁻⁹ (Roelants, 1996). L'analyse de ces révertants, par les techniques d'hybridation de type Southern et de PCR, a permis d'identifier trois types d'événements pouvant mener à la réactivation fonctionnelle du domaine ATCase (Fig. 1.2) :

- l'**insertion** d'un rétrotransposon Ty1 dans la phase codante du gène *URA2* (Roelants *et al.*, 1995) : un rétrotransposon Ty1 peut s'insérer au niveau de la région codant la seconde moitié de la CPSase. Cet élément est capable d'initier la transcription à partir de la séquence orientée en sens inverse par rapport à la région codant le domaine ATCase.

- la **délétion** de la région du gène *URA2* contenant les trois mutations ponctuelles (Welcker *et al.*, 2000) : la délétion de la région portant les 3 mutations permet à la séquence codant le domaine ATCase d'avoir un promoteur conduisant à sa transcription suivie de sa traduction.

- la **duplication** de la région codant le domaine ATCase suivie de sa fusion en aval d'une séquence promotrice.

Notre étude s'est focalisée sur les révertants dans lesquels la réversion phénotypique est due à une duplication. Cet événement de duplication correspond à la formation d'une

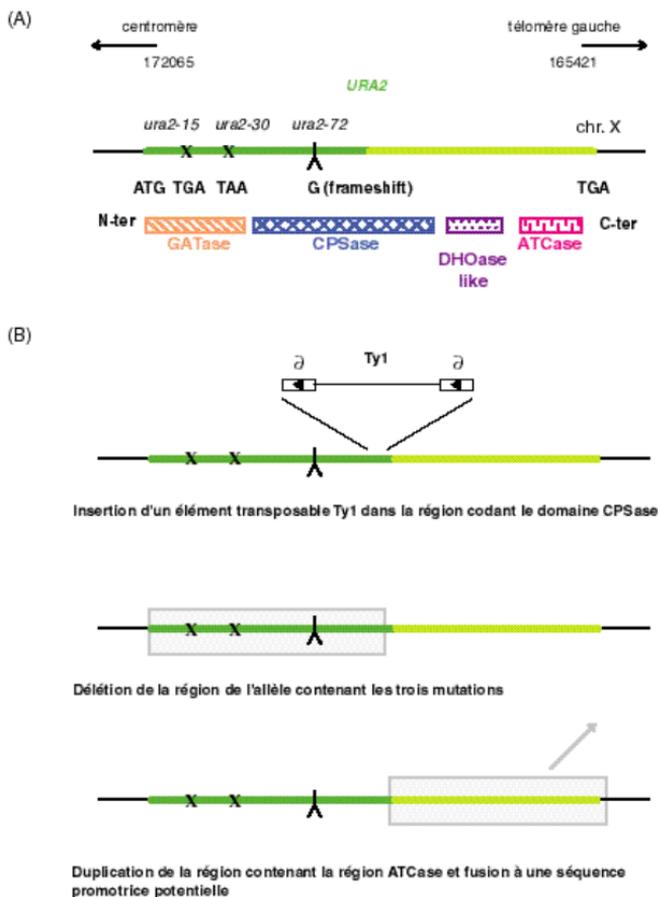


FIGURE 1.2 --Allèle mutant *ura2* 15-30-72 (A) et types d'événements identifiés pouvant mener à la réactivation fonctionnelle du domaine ATCase (B).

copie dupliquée de la région codant le domaine ATCase suivie de son insertion ailleurs dans le génome. Dans notre système de sélection, les événements de duplication doivent permettre l'expression du domaine ATCase. Cette insertion peut potentiellement se réaliser soit dans une région intergénique, soit dans une phase ouverte de lecture. Si la copie s'insère dans un gène, 2 conditions doivent être respectées. La sélection étant réalisée au niveau haploïde, l'insertion ne doit pas interrompre un gène essentiel ni provoquer une auxotrophie. De plus, elle doit se faire de telle manière que la fusion entre le gène receveur et la région ATCase soit en phase. Si la copie s'insère dans une région intergénique, cette dernière doit comporter des séquences permettant l'expression de l'ADN dupliquée.

2. CARACTÉRISATION MOLÉCULAIRE DES DUPLICATIONS

Le système de sélection positive basé sur l'utilisation de l'allèle mutant du gène *URA2* a permis d'isoler 44 révertants spontanés portant l'un des remaniements génomiques mentionnés précédemment : insertions de Ty1, délétions ou duplications (Roelants *et al.*, 1996). Parmi ces souches, 8 révertants ont été identifiés comme porteurs d'une duplication. Ces remaniements ont été caractérisés grâce à une analyse du profil de restriction *Bam*HI du locus *URA2*. La relation entre l'événement de duplication et la réversion phénotypique a été vérifiée par une analyse génétique.

2.1. Analyse du profil de restriction *Bam*HI au locus *URA2*

Une analyse du profil de restriction du locus *URA2* a été réalisée sur les 8 révertants sélectionnés. L'ADN génomique de ces souches a été digéré par l'enzyme de restriction *Bam*HI puis hybridé avec une sonde correspondant à la région codant le domaine ATCase. Pour la souche *ura2 15-30-72* ayant permis de sélectionner ces révertants, l'hybridation révèle une seule bande de 6,8 kb qui correspond au locus *URA2*. Par contre, l'hybridation de l'ADN génomique des 8 révertants révèle un profil de restriction *Bam*HI différent de celui de la souche *ura2 15-30-72* (Fig. 1.3). Ce profil est constitué de deux bandes : l'une de 6,8 kb correspondant au locus *URA2* et l'autre de taille variable correspondant à la partie dupliquée qui s'est insérée dans une région différente du génome et qui se caractérise par un environnement *Bam*HI différent.

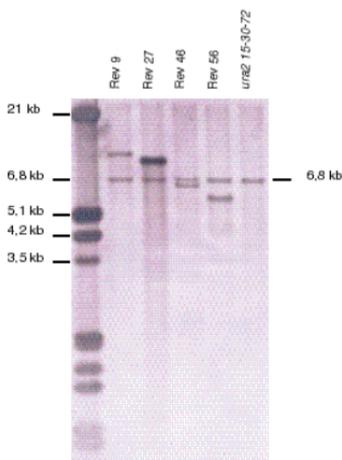


FIGURE 1.3 – Exemple de profil de restriction *Bam*HI dans le cas d'un événement de duplication.

Hybridation de l'ADN génomique digéré par l'enzyme de restriction *Bam*HI avec une sonde spécifique de la séquence codant le domaine *ATCase*.

2.2. Analyse génétique

Afin de vérifier le lien entre la réversion phénotypique et la présence de la copie dupliquée, nous avons croisé les 8 révertants avec la souche *ura2 15-30-72* et réalisé une analyse des tétrades après sporulation.

Après sporulation des diploïdes issus des croisements précédents, quatre types de spores peuvent théoriquement être obtenus (Fig. 1.4) :

- deux spores de type parental, l'une de phénotype Ura^+ et l'autre de phénotype Ura^- ;
- deux spores de type recombiné, l'une de phénotype Ura^+ et l'autre de phénotype Ura^- .

L'analyse des tétrades disséquées après sporulation montre bien une ségrégation de type 2 $Ura^+ / 2 Ura^-$ pour les révertants analysés. Une analyse du profil de restriction *Bam*HI, comme décrit précédemment, met en évidence que les spores Ura^+ sont effectivement porteuses d'une duplication alors que les spores Ura^- ne le sont pas. Ces données montrent que la duplication de la région codant l'*ATCase* est bien à l'origine de la réversion phénotypique.

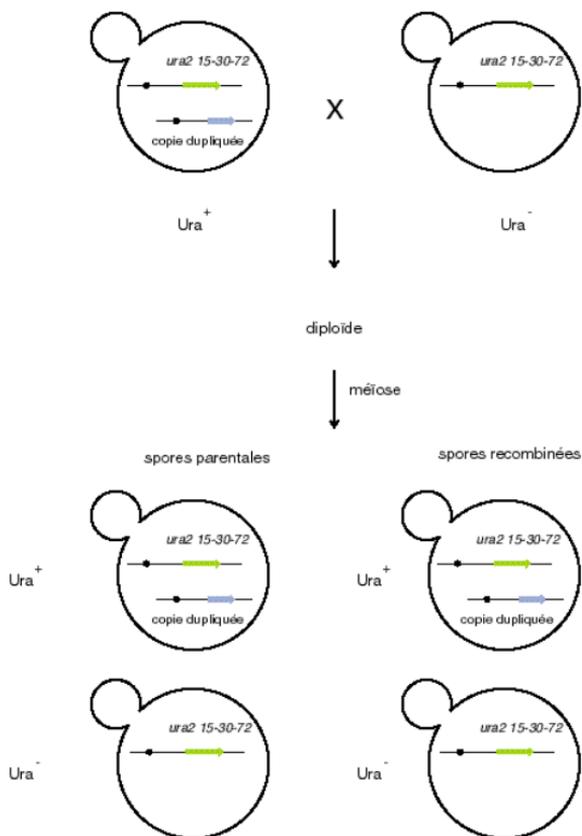


FIGURE 1.4 – Ségrégation observée dans le croisement d'un révertant par la souche *ura2 15-30-72*.

3. LOCALISATION CHROMOSOMIQUE ET ESTIMATION DE LA TAILLE DE L'ÉLÉMENT DUPLIQUÉ

L'identification du mécanisme ayant conduit à ces duplications nécessite dans une première étape de caractériser plus précisément les régions dupliquées. Nous nous sommes intéressés à deux caractéristiques moléculaires importantes pouvant donner des informations sur le type de duplications sélectionnées : la localisation chromosomique et la taille des éléments dupliqués.

Les séquences dupliquées peuvent être insérées sur le chromosome X porteur de la copie originelle ou sur un autre chromosome. Ces duplications sont qualifiées respectivement d'intrachromosomique et d'interchromosomique.

La duplication de séquence d'ADN peut se faire par plusieurs processus bien distincts : la duplication de gène, de segment, de bras de chromosome (Sankoff, 2001). Il est très difficile d'envisager qu'un seul et unique mécanisme conduise à la fois à ces différents types de duplications. À l'heure actuelle, les différentes duplications décrites dans la littérature sont classées en fonction de leur taille, étant donné qu'une classification basée sur les mécanismes impliqués n'est pas possible. La mise en évidence de la taille de la duplication est donc une caractéristique essentielle permettant de les différencier. Dans notre système, deux hypothèses peuvent être envisagées. La première serait que la duplication soit limitée à la séquence du gène *URA2* et aurait donc une taille d'environ 3 kb. La seconde impliquerait la duplication d'un fragment plus grand pouvant englober la région codant l'ATCase ainsi que des gènes contigus qui seraient par conséquent codupliqués.

3.1. Localisation chromosomique des régions dupliquées

La première caractéristique à laquelle nous nous sommes intéressés est la détermination du chromosome porteur de la séquence dupliquée. La séparation des chromosomes des 8 révertants par électrophorèse en champs alternés suivie d'une hybridation de type Southern a permis de déterminer la localisation des régions dupliquées (Fig. 1.5A).

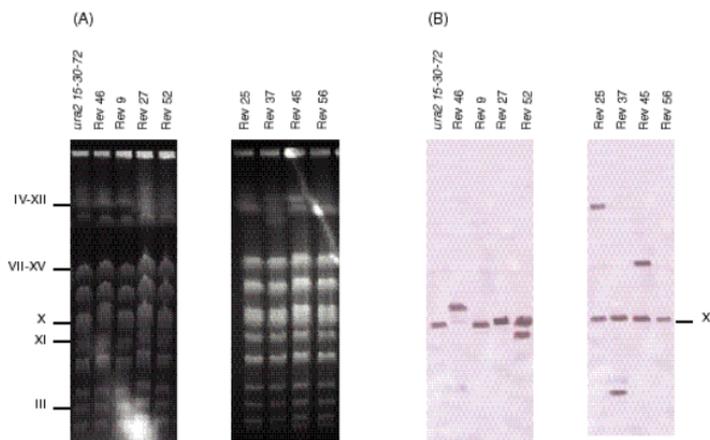


FIGURE 1.5 – Analyse du caryotype des révertants haploïdes porteurs d'un événement de duplication.

(A) Migration des chromosomes par électrophorèse en champs alternés et coloration au BET.

(B) Hybridation ADN/ADN de type Southern avec une sonde spécifique de la séquence codant le domaine ATCase.

Révertants	Chromosome
Riev 9	X
Riev 25	IV - XII
Riev 27	X
Riev 37	III
Riev 45	VII - XV
Riev 46	X
Riev 52	IX
Riev 56	X

TABLEAU 1.1 – Localisation chromosomique des régions dupliquées présentes chez les 8 révertants spontanés.

L'hybridation des chromosomes a été réalisée avec une sonde correspondant à la région codant l'ATCase. Dans le cas de 4 souches (Rev 25, Rev 37, Rev 45 et Rev 52), l'hybridation révèle le chromosome X porteur de l'allèle *URA2* mutant ainsi qu'un autre chromosome porteur de la copie dupliquée (Fig. 1.5B). Le remaniement observé dans ces 4 révertants correspond donc à une duplication interchromosomique localisée sur un autre chromosome que celui portant l'allèle mutant *ura2 15-30-72* (Tableau 1.1). La détermination précise du chromosome n'est pas évidente à déterminer car certains chromosomes, de taille voisine, comigrent lors de leur séparation par électrophorèse en champs alternés. Pour les 4 autres révertants (Rev 9, Rev 27, Rev 46 et Rev 56), l'hybridation du caryotype ne révèle qu'un seul chromosome indiquant que l'événement correspond à une duplication intrachromosomique sur le chromosome X (Fig. 1.5B).

L'analyse de ces résultats montre donc que la copie dupliquée peut s'insérer aussi bien sur le chromosome X porteur de l'allèle mutant *ura2 15-30-72* que sur un autre chromosome.

De plus, l'hybridation des chromosomes avec la sonde ATCase montre également une augmentation de la taille du chromosome X porteur de l'élément dupliqué chez les révertants Rev 27 et Rev 46. Cette augmentation de taille peut être estimée à environ 20 kb pour Rev 27 et environ 100 kb pour Rev 46 par comparaison avec la souche de référence *ura2 15-30-72*.

3.2. Estimation de la taille des régions dupliquées

Comme mentionnées précédemment, deux hypothèses peuvent être envisagées quant à la taille des éléments dupliqués. La première serait que la duplication soit limitée à la séquence du gène *URA2* et aurait donc une taille d'environ 3 kb. La seconde impliquerait la duplication d'un fragment plus grand pouvant englober la région codant l'ATCase ainsi que des gènes contigus.

Pour déterminer la taille des séquences dupliquées présents chez les 8 révertants, nous avons développé une stratégie de marche sur le chromosome par hybridation de type Southern en utilisant une série de sondes correspondant aux gènes adjacents au locus *URA2*. Par une telle procédure, si l'événement a conduit à une duplication génique, la sonde correspondant à la région ATCase sera la seule à révéler le chromosome porteur de la duplication. Par contre, si l'événement correspond à la duplication d'un segment

englobant un ou plusieurs gènes contigus au locus *URA2*, alors les sondes correspondant à ces gènes vont également permettre de révéler le chromosome porteur de la duplication.

3.2.1. Événements de duplication interchromosomique

Pour 4 souches (Rev 25, Rev 37, Rev 45 et Rev 52) porteuses d'une duplication interchromosomique, la marche sur le chromosome a été réalisée par hybridation de type Southern directement sur les chromosomes séparés par électrophorèse en champs alternés.

Pour 3 souches (Rev 25, Rev 37, Rev 45), deux sondes spécifiques des ORFs *YJL131c* et *YJL132w* (*MRS3*), localisées respectivement à 0,375 kb et 4,16 kb en aval du gène *URA2* (*YJL130c*), ont été testées. L'hybridation des caryotypes avec ces sondes ne révèle plus le chromosome porteur de la duplication mais uniquement le chromosome X porteur de l'allèle *ura2 15-30-72* (Fig. 1.6A).

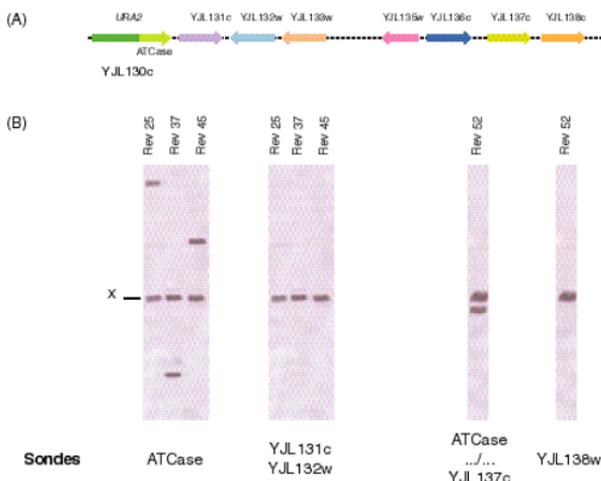


FIGURE 1.6 – Marche par hybridation de type Southern directement sur les chromosomes séparés par électrophorèse en champs alternés.

(A) Carte chromosomique de la région située en aval du gène *URA2*.

(B) Hybridation des chromosomes avec une sonde spécifique des ORFs situées en aval du gène *URA2*.

Ces résultats indiquent que seule la partie du gène *URA2* englobant le domaine ATCase est dupliquée.

Pour le révertant Rev 52, l'hybridation du caryotype en utilisant une série de sondes spécifiques de chacune des ORFs située entre l'ORF YJL130c et YJL137c, montre le chromosome X porteur de l'allèle *URA2* mutant et le chromosome XI porteur de la copie dupliquée (Fig. 1.6B). Par contre, l'hybridation du caryotype avec une sonde spécifique de l'ORF YJL138w ne révèle plus que le chromosome X (Fig. 1.6B). Ces résultats indiquent clairement qu'un segment d'environ 10 kb couvrant la région comprise entre les ORFs YJL130c à YJL137c s'est dupliquée chez le révertant Rev 52.

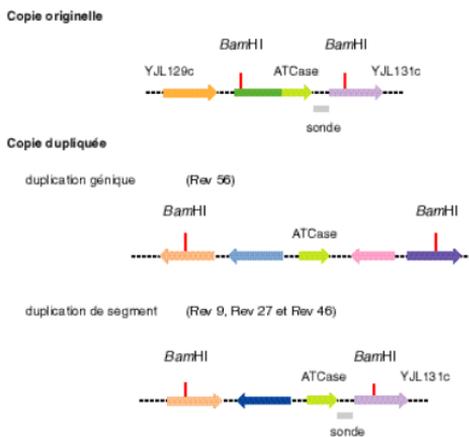
3.2.2. Événements de duplication intrachromosomique

Pour les 4 révertants: Rev 9, Rev 27, Rev 46 et Rev 56, les copies résidentes et dupliquées sont localisées soit sur le chromosome X, soit sur deux chromosomes différents (chromosome X et un autre) qui comigrent. Une marche directe sur les chromosomes par hybridation de type Southern n'est donc pas envisageable car les copies originelles et dupliquées ne peuvent être séparées par électrophorèse en champs alternés. Dans ce cas, la marche sur le chromosome a alors été réalisée sur l'ADN génomique par l'analyse de profils de restriction.

3.2.2.1. Marche sur le fragment de restriction *Bam*HI du locus *URA2*

Pour ces révertants, une marche a été réalisée sur le fragment de restriction *Bam*HI du locus *URA2* (Fig. 1.7A). Ce fragment de restriction révélé par une sonde spécifique de l'ATCase contient en plus du gène *URA2*, la partie 5' de l'ORF YJL131c ainsi que la région intergénique localisée entre ces 2 ORFs. Si le segment dupliqué se limite au gène *URA2*, une hybridation avec une sonde spécifique de la région intergénique révélera uniquement le fragment de 6,8 kb. Par contre si la duplication va au-delà de cette région, cette même sonde révélera le même profil que la sonde ATCase : une bande de 6,8 kb correspondant au locus *URA2* et une autre de taille variable correspondant à la partie dupliquée.

(A)



(B)

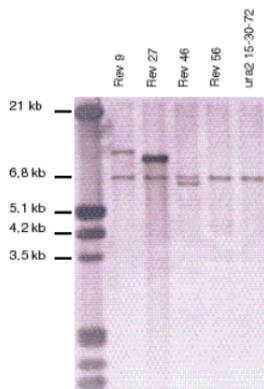


FIGURE 1.7 – Stratégie de marche sur le fragment de restriction *Bam*HI du locus *URA2*.

(A) Exemple théorique de carte de restriction *Bam*HI de la région environnant la copie originelle et dupliquée. (B) Hybridation de l'ADN génomique digéré par l'enzyme *Bam*HI avec une sonde spécifique de la région intergénique située entre les ORFs *YJL130c* et *YJL131c*.

Pour la souche Rev 56, le profil d'hybridation avec la sonde intergénique est constitué d'un fragment unique de 6,8 kb montrant ainsi que dans cette souche, la duplication est limitée à la région ATCase (Fig. 1.7B).

Par contre pour les 3 révertants restants (Rev 9, Rev 27, Rev 46), le profil obtenu avec la sonde ATCase et intergénique sont identiques montrant que dans ces révertants la duplication ne se limite pas au gène *URA2*.

3.2.2.2. Marche sur les fragments de restriction *Bam*HI suivants

Pour ces 3 révertants (Rev 9, Rev 27, Rev 46), la caractérisation exacte de la taille de la région dupliquée nécessite de marcher sur les fragments de restriction *Bam*HI contigus à celui portant le gène *URA2*. La mise en œuvre de cette stratégie n'est pas aussi évidente.

Le segment dupliqué est inséré dans un environnement *Bam*HI différent. Le site *Bam*HI localisé en amont de l'élément dupliqué va conduire à un fragment ayant une taille différente par rapport à celui du locus initial. Il en est de même du site situé en aval de l'élément dupliqué (Fig. 1.8A). Une hybridation de type Southern de l'ADN génomique digéré par l'enzyme *Bam*HI avec les sondes spécifiques à ces 2 fragments va donc révéler deux bandes. Par contre, les fragments *Bam*HI situés à l'intérieur de l'élément dupliqué conservent leur taille car les sites de restriction sont invariants. L'hybridation de l'ADN génomique avec une sonde spécifique de ces fragments va donc révéler une seule bande.

Chez les 3 révertants, une hybridation du fragment *Bam*HI suivant avec une sonde spécifique de l'ORF YJL132w a été réalisée (Fig. 1.8B). Pour le révertant Rev 9, le profil de restriction révèle deux bandes indiquant que le site *Bam*HI situé en 3' est variant par rapport à la copie originelle. La région dupliquée présente chez ce révertant a donc une taille inférieure ou égale à 6kb. Par contre, chez les révertants Rev 27 et Rev 46, le profil de restriction révèle une seule bande indiquant que l'ensemble du fragment *Bam*HI est encore dupliqué. Ces données indiquent donc que la région dupliquée présente dans ces 2 révertants à une taille supérieure à 6 kb.

Pour avoir une taille plus précise, il faudrait continuer à marcher de fragment de restriction en fragment de restriction jusqu'au rétablissement d'un profil à deux bandes indiquant la fin de la région dupliquée. Néanmoins, cette stratégie est laborieuse ce qui nous a conduit à envisager une autre approche décrite dans le chapitre III.

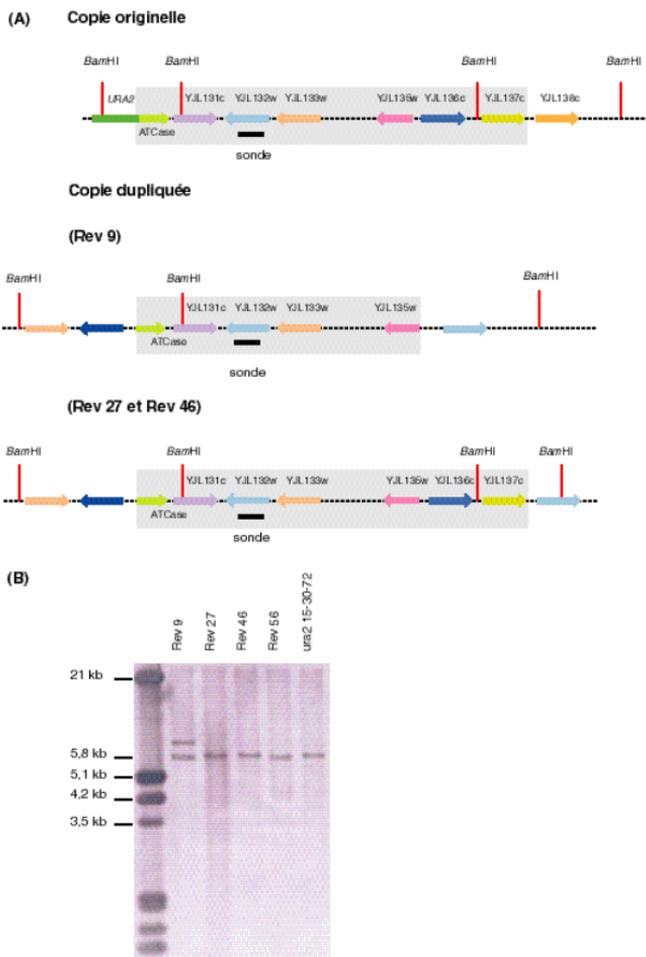


FIGURE 1.8 – Stratégie de marche sur les fragments de restriction *Bam*HI contigus au locus *URA2*.

(A) Exemple théorique de carte de restriction *Bam*HI de la région environnant la copie originelle et dupliquée.

(B) Hybridation de l'ADN génomique digéré par l'enzyme *Bam*HI avec une sonde spécifique de l'ORF *YJL132w*.

CONCLUSION

Les résultats décrits dans ce chapitre montrent que la copie dupliquée de la région codant l'ATCase peut s'insérer aussi bien sur le chromosome X porteur de l'allèle mutant *ura2 15-30-72* que sur un autre chromosome. De plus, l'estimation de la taille des séquences dupliquées a permis de mettre en évidence deux types de duplication parmi les 8 révertants. Chez 4 souches (Rev 25, Rev 37, Rev 45 et Rev 56), la duplication est limitée à une partie du gène *URA2* et peut alors être considérée comme de la duplication génique. Par contre, chez les 4 autres révertants (Rev 9, Rev 27, Rev 46 et Rev 52), la duplication correspond à un segment englobant la séquence codant le domaine ATCase ainsi qu'une série de gènes contigus. Ces remaniements constituent de la duplication segmentale.

Ces réarrangements chromosomiques constituent la première preuve expérimentale que la duplication d'une séquence d'ADN donnée dans une cellule haploïde de *Saccharomyces cerevisiae* peut se faire par au moins deux types de remaniements bien distincts : la duplication génique ou la duplication segmentale.

Il est difficile de concevoir qu'un mécanisme unique puisse conduire à la formation de ces deux types de réarrangement. C'est dans cette optique que la caractérisation moléculaire des révertants porteurs d'une duplication génique (chapitre II) sera abordée indépendamment de celle des révertants porteurs d'une duplication segmentale (chapitre III).

2

DUPLICATIONS GÉNIQUES PAR RÉTROPOSITION EN PHASE HAPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

RÉSUMÉ

*Le système de sélection basé sur un allèle mutant du gène URA2 a permis d'isoler 8 souches présentant un événement de duplication. Chez 4 révertants sélectionnés, la duplication est limitée à une partie du gène URA2 et peut alors être considéré comme de la duplication génique. Une analyse des sites d'insertion en amont et en aval de la région dupliquée a été entreprise afin d'obtenir des renseignements sur les mécanismes impliqués dans ces réarrangements. Dans les 4 révertants, la région dupliquée est généralement ponctuée en 3' par une séquence poly(A) et localisée entre des séquences de rétrotransposons de type Ty1. La présence de la région poly(A) en 3' de la copie dupliquée suggère fortement un mécanisme médié par la réverse transcription de l'ARNm correspondant. Chez *Saccharomyces cerevisiae*, le rétrotransposon de type Ty1 est une source évidente d'activité de transcriptase inverse. Afin de tester le rôle du rétrotransposon Ty1, 23 révertants portant un événement de duplication ont été sélectionnés à partir d'une souche dans laquelle un élément Ty1 a été surexprimé. Dans un tel contexte, la fréquence des duplications géniques ponctuées en 3' par une séquence poly(A) est considérablement augmentée. L'ensemble de ces résultats suggère fortement que les événements de duplication sélectionnés sont issus d'un processus de rétroposition.*

INTRODUCTION

Les événements de duplication concernant un seul gène sont qualifiés de duplication génique. Suite au séquençage systématique de *Saccharomyces cerevisiae*, la comparaison de l'ensemble des ORFs a permis de mettre en évidence les gènes dupliqués (Tekaiia and Dujon, 1999). Les résultats montrent que la majorité des familles multigéniques comprennent seulement deux ou trois gènes et peu de familles de grande taille sont observées. Ces familles couvrent approximativement 40 % des gènes de *Saccharomyces cerevisiae*. On peut distinguer 2 types de duplications géniques. Le premier type, majoritaire, correspond généralement à des familles dont les deux membres sont dispersés aléatoirement sur l'ensemble des chromosomes. Le second type, minoritaire, correspond aux gènes issus de duplications en tandem.

Les mécanismes impliqués dans la duplication génique sont difficiles à mettre en évidence. La nature de l'événement qui est à l'origine des duplications de séquences n'est pas facile à établir, notamment lorsque la duplication est ancienne car les « cicatrices » de cette origine sont souvent effacées. Deux mécanismes hypothétiques ont été proposés pour expliquer la formation de duplications géniques: un crossing-over inégal et un processus de rétroposition (Zhang *et al.*, 2003). Cependant, à ce jour, aucun de ces mécanismes n'a été confirmé expérimentalement.

La présence de duplications en tandem dans les génomes pourrait effectivement être expliquée par des événements de crossing-over inégaux, mitotiques ou méiotiques (Ohta, 1976; Smith, 1976). Ces événements consistent en une recombinaison ectopique entre des séquences répétées, comme les rétrotransposons, localisées de part et d'autre de la région dupliquée. Les duplications dispersées, quant à elles, pourraient être expliquées par un processus de rétroposition. Cette hypothèse est issue de la découverte dans la plupart des génomes eucaryotes de séquences appelées pseudogènes processés (c'est-à-dire copies dupliquées ayant perdu leur intron). Ces séquences correspondent à des copies ADNc de gènes qui ne codent plus de protéines fonctionnelles. Des études ont montré que l'activité transcriptase inverse des rétrotransposons sans LTR de type LINE est capable de rétrotranscrire des ARNm dont les produits forment des pseudogènes chez les eucaryotes supérieurs (Esnault *et al.*, 2000).

Dans les deux types de mécanismes, crossing-over inégal et rétroposition, les rétrotransposons peuvent être impliqués structurellement en tant que séquences d'ADN répétées ou fonctionnellement par l'intermédiaire de l'une de leurs activités.

Les rétrotransposons sont des composants ubiquitaires des génomes eucaryotes. Comme les rétrovirus, les rétrotransposons sont d'abord transcrits en une copie d'ARN, qui est ensuite rétrotranscrite en un ADNc capable de s'intégrer à un nouvel emplacement dans le génome (Boeke *et al.*, 1985). Les rétrotransposons sont divisés en 2 classes, basées sur la présence ou l'absence de LTRs (Long Terminal Repeat).

Les rétrotransposons ne possédant pas de séquences LTR ont été très largement étudiés chez les mammifères, comme par exemple les éléments LINE1 ou L1 (Long Interspersed Nuclear Element) présents chez l'homme et la souris. Il y a environ 850000 éléments LINEs dans le génome humain ce qui représente environ 20% de la séquence totale (IHGSC, 2001). Chaque élément mesure 6,1 kb et possède deux ORFs (cadre ouvert de lecture), ORF1 et ORF2, codant respectivement une protéine de fonction indéterminée et une protéine possédant un domaine endonucléase et transcriptase inverse. Toutefois, il existe peu d'éléments entiers et actifs car la plupart sont tronqués dans la région 5' et ont une taille d'environ 0,8 kb. Les éléments SINE (Short Interspersed Nuclear Element) font également partie des rétrotransposons dépourvus de séquences LTR, mais ne codent pas de protéines. L'exemple le plus connu est constitué par les éléments Alu, longs de 130 à 300 pb. Ils possèdent un promoteur interne pour l'ARN polymérase III. Ils peuvent ainsi être transcrits et recrutés par le mécanisme de rétrotransposition des éléments L1.

La seconde classe de rétrotransposons est flanquée de séquences en répétition directe : les LTRs. Ces transposons avec LTR codent des protéines équivalant aux protéines Gag et Pol des rétrovirus. Deux groupes bien distincts de rétrotransposons sont retrouvés parmi les eucaryotes: les éléments Ty1-copia et Ty3-gypsy. Ces deux groupes se distinguent par l'organisation des domaines de leur gène *POL* et la similarité de la séquence codant la reverse transcriptase (Xiong and Eickbush, 1990). La levure *Saccharomyces cerevisiae* est un modèle important pour l'étude des rétrotransposons avec LTR. Chez *S. cerevisiae*, les rétrotransposons avec LTR sont les seuls éléments transposables présents et se répartissent en 5 catégories (Ty1-Ty5). Dans la souche S288C, dont la séquence complète du génome a été publiée en 1996 (Goffeau *et al.*, 1996), un total de 331 éléments Ty ont été retrouvés (Tableau 2.1). Parmi ces éléments, 85 % correspondent à des séquences LTR ou à des fragments de séquences LTR. Dans l'ensemble, les séquences des rétrotransposons constituent 3,1% du génome de *S. cerevisiae* (Kim *et al.*, 1998).

Type d'élément	Nombre d'éléments Ty	
	Éléments entiers	Ensemble des structures (LTRs sob...)
Ty1	32	217
Ty2	34	13
Ty3	2	41
Ty4	3	32
Ty5	1	7

TABLEAU 2.1 – Distribution des éléments Ty dans le génome de *Saccharomyces cerevisiae*.

Les rétrotransposons Ty1 constituent les éléments transposables les plus abondants chez *S. cerevisiae* avec 217 éléments dont 32 transposons Ty1 entiers. Ces rétrotransposons ont une taille comprise entre 5 et 6 kb. Ils sont encadrés des séquences en répétition directe (LTR) de 300 à 400 pb appelées séquence . La séquence comprise entre les deux LTR présente deux cadres de lecture ouverts : *TYA* et *TYB*. *TYA* est l'équivalent du gène *GAG* des rétrovirus et code les protéines de la nucléocapside. Chez *S. cerevisiae*, l'assemblage des protéines TYA conduit à la formation de VLPs (virus-like particles). *TYB* correspond au gène *POL* des rétrovirus et porte plusieurs activités enzymatiques : protéase, reverse transcriptase, ribonucléase H et intégrase.

Les différentes étapes intracellulaires de la réplication et de l'insertion des rétrotransposons sont semblables à celles observées chez les rétrovirus (Fig. 2.1). Les LTRs possèdent les séquences nécessaires à l'initiation et la terminaison de la transcription. L'ARNm obtenu par transcription des éléments Ty commence dans le LTR localisé en 5' et s'achève dans le LTR localisé en 3'. L'ARNm, transporté dans le cytoplasme, permet la synthèse des protéines TYA et TYB. Par ailleurs, il est rétrotranscrit en ADN complémentaire dans les VLPs qui contiennent la reverse transcriptase et l'intégrase. Cette dernière porte un signal de localisation nucléaire à son extrémité C-terminale (Kenna *et al.*, 1998 ; Moore *et al.*, 1998). Ceci permettrait à cette protéine et à l'ADNc d'entrer dans le noyau de la cellule et de catalyser ainsi l'intégration de l'ADNc dans le génome en provoquant des coupures simple brin décalées dans l'ADN génomique.

La majorité des 32 éléments Ty1 endogènes complets semblent être compétents pour la transposition (Kim *et al.*, 1998). L'ARNm synthétisé à partir des éléments Ty1 et Ty2 peut représenter, selon les conditions de croissance, jusqu'à 1 % de l'ARN total et 10 % des ARN polyadénylés dans des cellules haploïdes (Elder *et al.*, 1983; Curcio *et al.*, 1990). Cependant, les protéines TYB ne sont pas abondantes (Curcio and Garfinkel, 1992), ce qui limite la rétrotranscription et entraîne un taux de transposition faible, de l'ordre de 10^{-5} à 10^{-7} événements/cellule/génération (Curcio and Garfinkel, 1991).

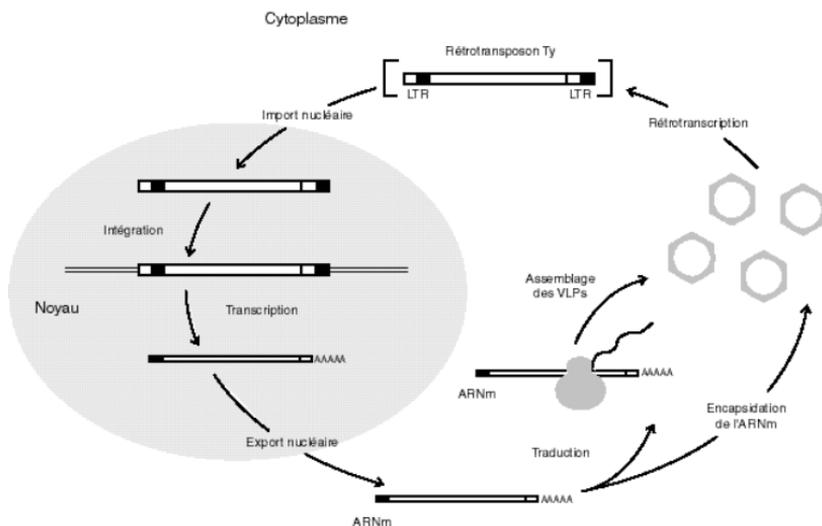


FIGURE 2.1 – Cycle réplcatif des éléments Ty.

Le fond gris délimite le noyau. L'ADN double brin est représenté par un rectangle épais et son ARN par un rectangle mince.

Les rétrotransposons Ty1 peuvent également être à l'origine de remaniements telles les inversions, translocations, délétions... Les séquences répétées sont, en théorie, de très bons substrats pour la recombinaison homologe. Les rétrotransposons Ty constituent une part importante de la redondance génique au sein du génome de *S. cerevisiae*. De plus, ces éléments sont dispersés dans le génome, entiers ou sous forme de LTR isolé (LTR solo). Potentiellement, ils peuvent donc provoquer des recombinaisons ectopiques. Cette recombinaison entre éléments Ty et/ou séquences LTR a été étudiée par de nombreux groupes (Liebman *et al.*, 1981; Sugawara and Szostak, 1983 ; Kupiec and Petes, 1988).

Récemment, plusieurs travaux ont montré que les recombinaisons ectopiques entre éléments Ty1 peuvent jouer un rôle important dans l'évolution de *S. cerevisiae*. Tout d'abord, les souches œnologiques, dont le métabolisme est adapté à la fermentation alcoolique, présentent plusieurs translocations chromosomiques en comparaison aux souches de laboratoires. L'analyse des jonctions entre les régions transloquées montre que dans tous les cas des éléments Ty1 sont présents (Rachidi *et al.*, 1999 ; Pérez-Ortín *et al.*, 2002). Des résultats similaires ont été obtenus avec des souches industrielles adaptées à d'autres métabolismes (Codon *et al.*, 1998). Chez des souches de laboratoire, dans des conditions de croissance limitante et sur un grand nombre de générations, les changements caryotypiques associés à des éléments Ty1 sont nombreux (Camasses, 1996 ; Dunham *et al.*, 2002).

Le système de sélection positive basé sur l'utilisation de l'allèle mutant du gène *URA2* a permis d'isoler 8 révertants spontanés correspondant à une duplication de la région codant l'ATCase à partir d'une souche haploïde *ura2 15-30-72* (Roelants *et al.*, 1995).

Dans le chapitre I, l'analyse de ces 8 révertants spontanés met en évidence deux processus bien distincts conduisant à la duplication d'une séquence d'ADN : la duplication génique et la duplication segmentale.

Dans ce premier chapitre, notre étude s'est focalisée sur les 4 révertants (Rev 25, Rev 37, Rev 45 et Rev 56) dans lesquels l'événement de duplication correspond à la duplication d'une partie du gène *URA2*. Afin de caractériser les mécanismes qui pourraient être impliqués dans la formation de ces remaniements, nous nous sommes attachés à déterminer les sites d'insertion en amont et en aval de la séquence

dupliquée. La détermination de ces points de jonction est une étape indispensable car les jonctions constituent les « cicatrices » de l'événement de duplication. Ces séquences « cicatrices » peuvent apporter des indications clés pour élucider les mécanismes conduisant aux duplications géniques sélectionnées.

RÉSULTATS

1. Caractérisation des sites d'insertions de la région dupliquée

L'identification des sites d'insertion de la copie dupliquée a été réalisée pour les 4 révertants (Rev 25, Rev 37, Rev 45 et Rev 56) par le séquençage des régions flanquantes. Pour des raisons de commodité, nous avons qualifié de jonction 5' le site d'insertion situé en amont de la région dupliquée. Dans notre crible de sélection, cette région 5' doit apporter une séquence promotrice permettant la réactivation fonctionnelle de l'ATCase. Par analogie, la région 3' correspond aux séquences situées en aval de la région dupliquée.

1.1. Détermination du site d'insertion en amont de la duplication

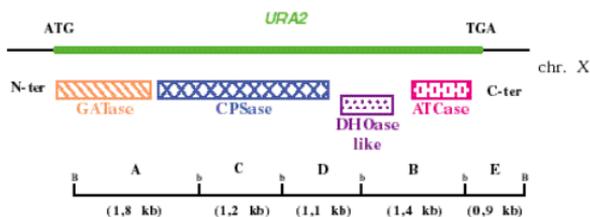
1.1.1. Stratégie d'insertion-excision de plasmide

La jonction 5' de la région dupliquée a été clonée en utilisant une stratégie d'insertion et d'excision de plasmide (Fig. 2.3). La première étape consiste à insérer un plasmide intégratif dans la région dupliquée en aval du point de départ de la duplication. Dans une deuxième étape, le plasmide recombinant portant la jonction inconnue est isolé puis caractérisé.

◦ Estimation du point de départ de la région dupliquée

La première étape consiste à déterminer approximativement le point de départ de la région dupliquée afin de pouvoir insérer un plasmide en aval de celui-ci. Pour ceci, nous avons développé une approche basée sur la carte de restriction *Bgl*III du locus *URA2* (Fig. 2.2A). L'extrémité 5' de la région dupliquée est caractérisée par un environnement *Bgl*III différent par rapport au locus *ura2* 15-30-72. Ce point de départ est forcément localisé entre les 3 mutations et la séquence codant le domaine ATCase. Si l'on considère la carte de restriction *Bgl*III du gène *URA2*, le point de départ sera donc situé dans l'un des 3 fragments *Bgl*III suivants : C, D ou B. Ainsi,

(A)



(B)

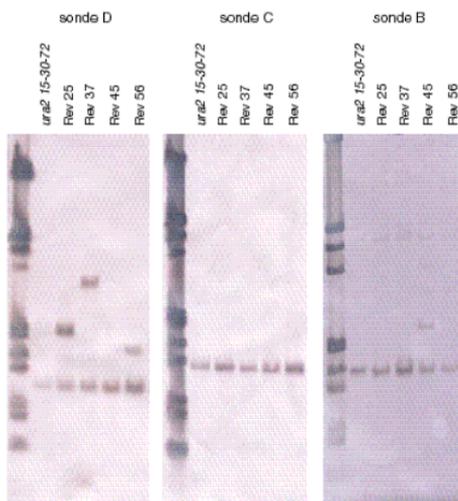


FIGURE 2.2 – Détermination de la région de départ de la partie du gène URA2 dupliquée.

(A) Carte de restriction BglIII du locus URA2.

(B) Hybridation de l'ADN génomique digéré par l'enzyme de restriction BglIII avec une sonde spécifique des fragments D, C et B.

par analyse du profil de restriction *Bgl*III avec des sondes spécifiques des 3 fragments (C, D et B), il sera donc possible de déterminer lequel des sites *Bgl*III est variant et par conséquent savoir à partir de quel fragment *Bgl*III la duplication va commencer. Ainsi, l'ADN génomique des 4 révertants (Rev 25, Rev 37, Rev 45 et Rev 56) a été digéré par l'enzyme de restriction *Bgl*III puis hybridés avec les 3 sondes spécifiques. Pour une sonde donnée, la présence de deux bandes montre que le début de la duplication est localisé dans le fragment correspondant. Par contre, si une seule bande est révélée, le début de la duplication n'est pas localisé dans ce fragment. Les résultats obtenus indiquent que les régions dupliquées débutent à partir du fragment D chez 3 révertants (Rev 25, Rev 37 et Rev 45) alors que chez le révertant Rev 56, le segment dupliqué prend son origine dans le fragment B (Fig. 2.2B).

◦ Insertion du plasmide et sélection

Les 3 plasmides portant les fragments D, B et E (respectivement appelés pFLD, pFLB et pFLE) ont été construits à partir du plasmide intégratif pFL35. La localisation du point de départ de la duplication nous a permis de choisir le plasmide à utiliser pour l'intégration en vue du clonage. Par exemple, pour les 3 souches dont les segments dupliqués commencent dans le fragment D, le vecteur intégratif pFLB est inséré dans la région correspondante du segment dupliqué (Fig. 2.3). L'intégration du plasmide est sélectionnée par l'utilisation du marqueur *Trp1*. L'insertion du vecteur peut se faire dans 2 régions différentes : dans la copie originale ou dupliquée. Si l'insertion se fait dans la copie dupliquée et fonctionnelle, la séquence codant le domaine ATCase va être interrompu, ce qui rend les transformants auxotrophes à l'uracile. Par contre, dans le cas d'une intégration dans la copie *ura2 15-30-72*, les révertants possédant toujours leur copie fonctionnelle intègre restent prototrophes à l'uracile. Ainsi, parmi les transformants *Trp*⁺, nous avons retenu les clones auxotrophes à l'uracile (Fig. 2.3).

◦ Récupération de la borne inconnue et séquençage

Après l'étape d'intégration, l'ADN génomique des transformants a été digéré par l'enzyme de restriction *Hind*III. Le fragment *Hind*III contenant la jonction inconnue porte également l'origine de répllication d'*E. coli* ainsi que le gène de résistance à l'ampicilline du plasmide initial. Une circularisation par ligation de ce fragment conduit

à un plasmide pouvant être sélectionné chez *E. coli*. Le produit de digestion a alors été ligué. Les clones résistants à l'ampicilline sélectionnés après transformation ont été récupérés et analysés. Les plasmides ont été récupérés par extraction et les jonctions ont été séquencées.

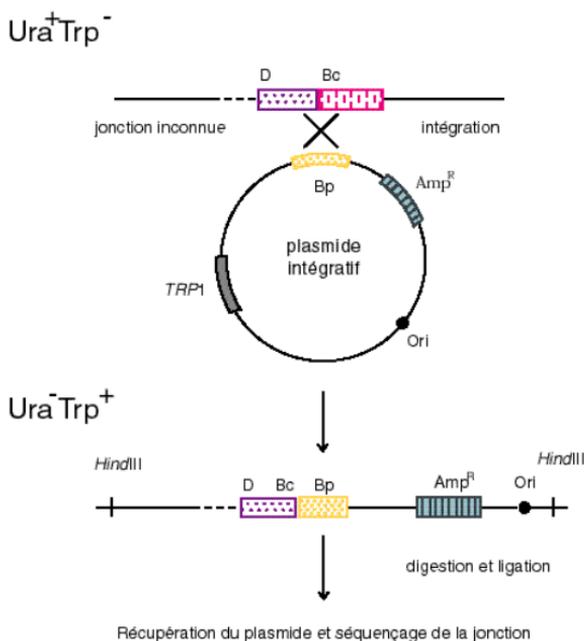


FIGURE 2.3 – Stratégie de clonage de l'extrémité 5' de la région dupliquée.

D et B correspondent aux fragments BglII du locus *URA2*.

Bc et Bp sont respectivement les fragments B chromosomique et plasmidique.

1.1.2. Analyse des séquences à la jonction

Les séquences ont alors été utilisées pour faire une comparaison contre la banque de données Saccharomyces Genome Database à l'aide du programme BLAST (<http://seq.yeastgenome.org/cgi-bin/SGD/nph-blast2sgd>). L'analyse des séquences montre que le point de départ des copies dupliquées est effectivement localisé en aval des trois mutations de l'allèle *ura2 15-30-72*. Ce point de départ n'est cependant pas unique car il peut être localisé à la fin de la région codant la CPSase ou dans la région codant le domaine cryptique DHOase-like (Fig. 2.4). Dans les 4 révertants, la région dupliquée est fusionnée en 5' à une séquence de rétrotransposon Ty1 ayant la structure suivante : LTR suivi d'une partie de l'ORF TYA du rétrotransposon (Fig. 2.4). La partie correspondant à l'ORF TYA est de taille variable chez les différents révertants. La fusion ainsi créée entre la séquence TyA et l'extrémité 5' de la région dupliquée est en phase, ce qui conduit à la formation d'un gène chimère TyA-ATCase (Fig. 2.4). On peut également noter que les 4 jonctions présentent de courtes régions de similarité (2 à 11pb) entre la séquence TyAet l'extrémité 5' de la partie dupliquée du gène *URA2* (Fig. 2.5).

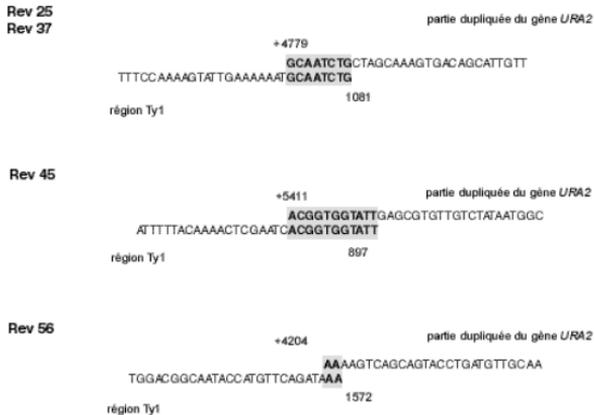
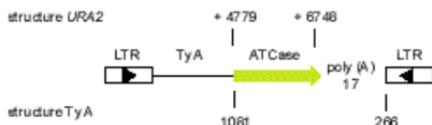


FIGURE 2.5 –Présence de microhomologie à la jonction 5' des duplications géniques.

La typographie en gras indique les courtes séquences de similarité présente entre la séquence TyAdu rétrotranspo - son Ty1 et la partie dupliquée du gène *URA2*.

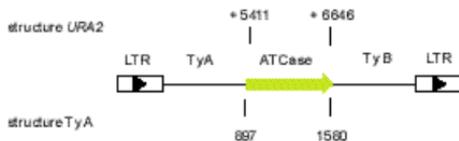
Rev 25



Rev 37



Rev 45



Rev 56

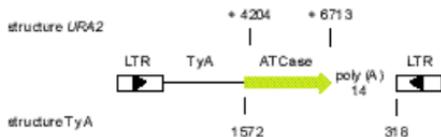


FIGURE 2.4 – Site d'insertion chromosomique en 5' et 3' des duplications géniques présentes chez 4 révertants spontanés.

Pour la structure URA2, les coordonnées nucléotidiques au début et à la fin de la région dupliquée sont données par rapport à la position +1 du nucléotide A du codon initiateur ATG du gène URA2. Pour la structure Ty1, les nombres situés en-dessous montrent la variation que l'on peut observer entre les différents révertants.

Pour préciser la localisation chromosomique de ces éléments Ty, nous avons utilisé la technique de Long Expand Tail (terminal asymmetric interlaced) PCR décrite par Szabados *et al.* (2002). Cette technique a permis de déterminer la région 5' située en amont des structures Ty pour 2 révertants (Rev 25 et Rev 56) mais a été infructueuse dans le cas des 2 autres. Pour les 2 souches analysées (Rev 25 et Rev 56), les séquences en amont de la région dupliquée correspondent respectivement au gène de tRNA-Ile [I(UAU)D] localisé sur le chromosome IV et au gène de tRNA-Ala [tA(AGC)P] localisé sur le chromosome XVI (Fig. 2.6). Dans la souche *ura2 15-30-72*, ces deux gènes sont situés en amont de Ty1 entiers. L'ensemble de ces données indiquent donc que la région dupliquée s'est insérée dans un Ty résident (Fig. 2.6).

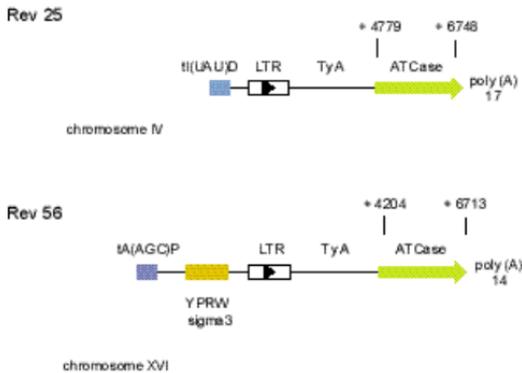


FIGURE 2.6 – Région située en amont des séquences Ty1 chez les révertants Rev 25 et Rev 56. Pour la structure URA2, les coordonnées nucléotidiques au début et à la fin de la région dupliquée sont données par rapport à la position +1 du nucléotide A du codon initiateur ATG du gène URA2.

1.2. Détermination du site d'insertion en aval de la duplication

La technique classique de PCR asymétrique (Tail-PCR) décrit par Liu *et al.* (1995) a été utilisée pour déterminer la région située en aval de la copie dupliquée de la région codant l'ATCase. Les résultats montrent que cette région se termine par une séquence poly(A) de 10 à 14nucléotides de long chez trois des 4souches (Rev 25, Rev37 et Rev 56) (Fig.2.4). Dans la souche *ura2 15-30-72*, cette séquence poly(A) est absente aussi bien au niveau du locus *URA2* que du rétrotransposon receveur. De plus, l'analyse des séquences montre que la région codant l'ATCase est complète dans toutes les souches étudiées. La partie située entre le codon de terminaison et la séquence poly(A) varie de 68 à 103 nucléotides selon les révertants. La présence d'une séquence poly(A) apporte des informations supplémentaires sur les mécanismes qui peuvent être impliqués dans la formation de ces événements de duplication. En effet, elle suggère fortement la présence d'un intermédiaire ARN lors de la duplication de ces séquences d'ADN.

Les séquences de la jonction 3' ont été utilisées pour faire une comparaison contre la banque de donnée Saccharomyces Genome Database à l'aide du programme BLAST. Les résultats montrent que pour tous les événements de duplication, l'extrémité 3' de la région dupliquée est attachée après la séquence poly(A) à des séquences de rétrotransposons de type Ty1 (Fig.2.4). Cependant, ces séquences Ty1 n'ont pas la même orientation que celles localisées à la jonction 5'et proviennent par conséquent d'autres structures Ty1.

2. LA SUREXPRESSION D'UN RÉTROTRANSPON TY1 AUGMENTE LA FRÉQUENCE DES ÉVÉNEMENTS DE DUPLICATION DE LA RÉGION CODANT L'ATCASE

La présence de la séquence poly(A) à l'extrémité 3'de la copie dupliquée chez les trois révertants (Rev 25, Rev 37 et Rev 56) suggère fortement un mécanisme impliquant la formation d'un ADNc par rétrotranscription de l'ARNm suivie de son intégration dans le génome. Les régions dupliquées analysées pourraient donc être issues d'un processus de rétroposition (Zhang, 2003). Ainsi, la formation de telles duplications géniques nécessite une activité transcriptase inverse. Une source évidente d'activité transcriptase inverse chez *S. cerevisiae* est constituée par le rétrotransposon Ty1. Afin de tester ces hypothèses, l'analyse moléculaire d'une collection de révertants, sélectionnés à partir de la souche *ura2 15-30-72* dans laquelle la transposition a été augmentée, a été réalisée.

2.1. Sélection des révertants

Pour faciliter l'étude de la transposition des éléments Ty1, Boeke *et al.* (1985) ont développé une méthode simple permettant d'augmenter la fréquence de la transposition en augmentant le taux d'ARN messenger de ces éléments dans les souches étudiées. Cette méthode fait intervenir un élément Ty1 en aval du promoteur inductible du gène *GALI*. Cet élément est porté par un plasmide contenant l'origine de répllication 2 μ (plasmide pGTy1). Lorsque l'expression de l'élément contenu sur ce plasmide est induite par croissance sur milieu galactose, la fréquence de la transposition est augmentée d'une manière significative. Ce système a été utilisé pour induire la transposition dans la souche *ura2 15-30-72*.

Le plasmide pFJ2 (Roelants, 1996) a été construit à partir du plasmide pJEF (Eichinger et Boeke, 1988). Ce plasmide contient l'élément TyH3 (Boeke *et al.*, 1988) dont l'expression est sous la dépendance du promoteur *GALI*. Il a été modifié en introduisant le marqueur *TRP1* au niveau du site *Bam*HI unique du plasmide afin d'avoir un marqueur de transformation pour la souche réceptrice *ura2 15-30-72*. Les sélections réalisées dans ce contexte ont permis d'isoler 35 révertants.

Dans un tel contexte, l'analyse des révertants montre que la fréquence de l'ensemble des remaniements sélectionnés est augmentée par rapport à une sélection sans surexpression de l'élément Ty c'est-à-dire sur milieu glucose (Tableau 2.2). De plus, cette augmentation est plus importante pour les événements de duplication que pour les insertions d'éléments Ty et les délétions. L'induction de la transposition dans une souche *ura2 15-30-72* provoque donc une augmentation de la fréquence de réversion et favorise de manière importante les duplications. Cette constatation va dans le sens de l'implication de la transposition ou de l'expression des éléments Ty1 dans le mécanisme conduisant aux événements de duplication. Afin de préciser et de vérifier les hypothèses émises sur l'implication de la transcriptase inverse dans la genèse des duplications géniques, l'analyse des 23 révertants porteurs d'une duplication a été entreprise.

2.2. Analyse des révertants porteurs d'un événement de duplication

Dans un premier temps, les 23 révertants ont été identifiés comme porteurs d'une duplication. Ces remaniements ont été caractérisés grâce à une analyse du profil de restriction *Bam*HI du locus *URA2*. Par la suite, comme pour les 4 révertants spontanés porteurs d'une duplication génique, nous avons déterminé la localisation, la taille et les sites d'insertion de l'élément dupliqué.

Rearrangements	Sélections				
	FRJ1 (pFJ2) sur glucose ^{a b}		FRJ1 (pFJ2) sur galactose ^{a b}		
Insertion Ty1	25	$0,57 \times 10^{-9}$	6	$2,1 \times 10^{-9}$	X 3,6
Délétion	10	$0,11 \times 10^{-9}$	6	$2,1 \times 10^{-9}$	X 19
Duplication	5	$0,23 \times 10^{-9}$	23	$8,05 \times 10^{-9}$	X 35

TABLEAU 2.2 – Révertants sélectionnés à partir de la souche *ura2 15-30-72* dans laquelle la transposition a été augmentée.

(a) Le premier chiffre indique le nombre de révertants isolés dans chaque cas,

(b) le taux de réversion calculé pour chaque type de réarrangement d'après Roelants (1996).

2.2.1. Analyse du profil de restriction *Bam*HI au locus *URA2*

L'analyse du profil de restriction *Bam*HI du locus *URA2* a été réalisée sur les 23 révertants sélectionnés. L'hybridation de l'ADN génomique digéré par l'enzyme *Bam*HI avec une sonde spécifique de la région codant le domaine ATCase diffère de celui de la souche ancestrale par la présence d'une bande supplémentaire. Dans le cas des 23 révertants, le profil de restriction *Bam*HI révèle deux bandes : l'une de 6,8 kb correspondant au locus *URA2* et l'autre de taille variable correspondant à la partie dupliquée.

2.2.2. Localisation chromosomique de la région dupliquée

La séparation des chromosomes des 23 révertants par électrophorèse en champs alternés suivie d'une hybridation de type Southern a permis de déterminer la localisation des régions dupliquées (Fig.2.7A). Contrairement aux révertants spontanés, une modification du profil électrophorétique a été mise en évidence. En effet, on peut observer pour certains révertants une bande supplémentaire par rapport à celui de la souche ayant permis de les sélectionner.

L'hybridation a été réalisée avec une sonde correspondant à la région codant l'ATCase (Fig.2.7B). Pour 19 souches, l'hybridation révèle le chromosome X porteur de l'allèle mutant *ura2 15-30-72* ainsi qu'un autre chromosome porteur de la copie

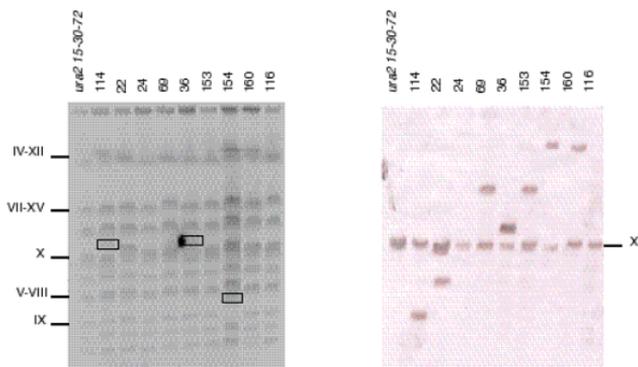


FIGURE 2.7 – Exemples de caryotypes après électrophorèse en champs alternés de quelques révertants sélectionnés à partir de la souche FRJ1.

(A) Migration des chromosomes par PFGE et coloration au BET.

(B) Révélation par hybridation ADN/ADN de type Southern avec une sonde spécifique de la région codant le domaine ATCase.

Chromosomes	Révertants
IV - XII	38 - 72 - 74 - 154 - 160
VII - XV	16 - 69 - 153 - 156
XIII - XVI	63 - 68
II - XIV	36 - 67
X	24 - 116
XI	152 - 158
V - VIII	22 - 149
IX	/
III	114
I - VI	40

TABLEAU 2.3 – Localisation chromosomique des régions dupliquées présentes chez les 23 révertants isolés à partir de la souche FRJ1.

dupliquée. Le remaniement observé dans ces 19 révertants correspond donc à une duplication interchromosomique dans laquelle un autre chromosome que le X (Tableau 2.3) est porteur de la copie dupliquée. Pour 4 autres souches, l'hybridation du caryotype ne révèle qu'un seul chromosome, suggérant que l'événement correspond à une duplication intrachromosomique sur le chromosome X. L'ensemble de ces données montre que la majorité des chromosomes de *S. cerevisiae* peuvent être le siège de la copie dupliquée. Ces données indiquent que le site d'intégration de la copie dupliquée n'est pas limité à une région chromosomique précise.

2.2.3. Détermination de la taille de la région dupliquée

Afin de distinguer les duplications géniques et segmentales parmi les événements sélectionnés, nous avons déterminé la taille des éléments dupliqués en suivant la stratégie de marche sur le chromosome utilisée dans le chapitre I.

Pour les 19 souches dans lesquelles une duplication interchromosomique s'est produite, la marche sur le chromosome a été réalisée par hybridation directe des chromosomes séparés par électrophorèse en champs alternés. Pour les 4 révertants dans lesquelles une duplication intrachromosomique s'est produite, cette procédure n'a pas pu être utilisée car la copie résidente et dupliquée comigrent lors de la séparation des chromosomes par électrophorèse en champs alternés. Dans ce cas, la marche sur le chromosome a été réalisée sur l'ADN génomique par l'analyse du profil de restriction *Bam*HI.

Dans le cas des 23 révertants, les résultats obtenus montrent que les duplications sont limitées en taille à la séquence du gène *URA2*. On peut donc en conclure que la sur-expression d'un élément Ty1 permet d'augmenter la fréquence des duplications et plus particulièrement la sélection de duplications géniques.

2.2.4. Détermination des sites d'insertion de la région dupliquée

La détermination des points de jonction a été réalisée dans le but de voir si on trouvait les mêmes « cicatrices » et notamment la présence de la séquence poly(A), située en aval de la région dupliquée, qui apporte des informations supplémentaires sur le mécanisme pouvant être impliqué dans la formation de ces événements de duplication.

2.2.4.1. Site d'insertion en amont de la duplication

L'analyse des 4 révertants spontanés avait mis en évidence que la région dupliquée est rattachée en 5' à une séquence de rétrotransposon Ty1 ayant la structure suivante : LTR suivi d'une partie variable de l'ORF TYA du rétrotransposon (Fig. 2.4). Ainsi, avant d'utiliser la stratégie d'insertion et d'excision de plasmide, nous avons décidé de vérifier par amplification PCR si ce type de structure Ty se retrouvait au niveau des 23 révertants sélectionnés. Une amplification PCR a alors été réalisée en utilisant l'oligonucléotide spécifique B41R de la région du rétrotransposon et un second oligonucléotide B1.1 spécifique de la séquence codant le domaine ATCase (Fig. 2.8).

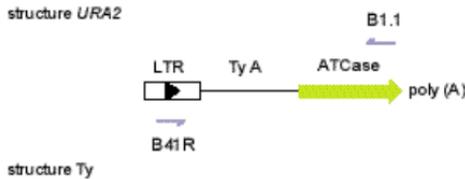


FIGURE 2.8 – Site d'insertion chromosomique en 5' des duplications géniques présentes chez les 23 révertants isolés à partir de la souche FRJ1. Les flèches bleues indiquent les oligonucléotides ayant permis d'amplifier la jonction 5'.

Chez 19 révertants, un fragment a effectivement pu être amplifié puis séquencé. L'analyse des séquences montre que l'insertion en 5' dans ces 19 révertants a une structure similaire à celle observée chez les 4 révertants spontanés. La région dupliquée est toujours rattachée en 5' à une séquence de rétrotransposon Ty1 ayant la structure suivante : LTR suivi d'une partie variable de l'ORF TYA du rétrotransposon (Fig. 2.9). La fusion entre la séquence TyA et l'extrémité 5' de la région dupliquée est en phase, ce qui conduit à la formation d'un gène chimère TyA-ATCase (Fig. 2.9). Le point de départ des régions dupliquées varie d'un révertant à

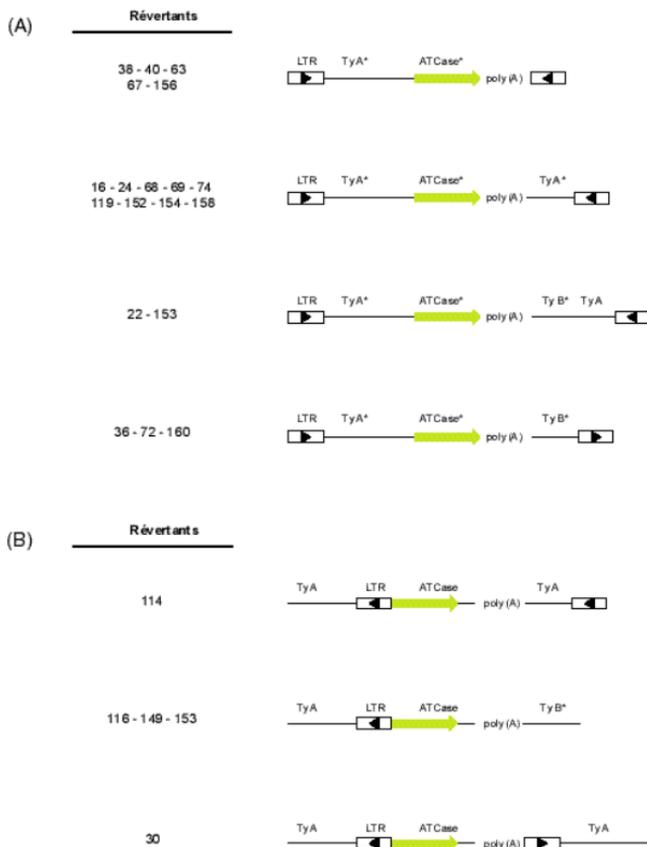


FIGURE 2.9 – Représentation schématique des sites d'insertion 5' et 3' des duplications géniques présentes chez les 23 révertants isolés à partir de la souche *FRJ1*.

Le symbole * indique que la séquence est variable d'un révertant à l'autre.

Les révertants 16, 72 et 74 n'ont pas de séquence poly(A).

l'autre. Cependant, il est toujours localisé en aval de la mutation de décalage du cadre de lecture. De plus, il est à noter que les 19 jonctions présentent de courtes régions d'identité (de 2 à 9 pb) entre la séquence TyA et l'extrémité 5' de la partie dupliquée du gène *URA2*.

Pour les 4 révertants restants (30, 114, 116 et 149), aucun fragment n'a pu être amplifié. La stratégie d'insertion et d'excision de plasmide décrite précédemment a donc été entreprise. Les résultats obtenus après clonage et séquençage montrent que la région ATCase est associée à un élément Ty1 qui se trouve en orientation inverse par rapport aux structures observées chez les 19 autres révertants (Fig. 2.9). La situation est donc comparable au remaniement observé lors d'une réactivation fonctionnelle du domaine ATCase par insertion d'un rétrotransposon Ty1 au niveau du locus *ura2 15-3072* (Roelants *et al.*, 1995).

2.2.4.2. Site d'insertion en aval de la duplication

Comme pour les 4 révertants, la technique de PCR asymétrique décrite par Liu *et al.* (1995) a été utilisée pour déterminer la région située en aval de la copie dupliquée. Les résultats montrent que la fin de la région dupliquée varie d'un révertant à l'autre. Elle comprend toujours le codon de fin de traduction suivi d'une région variable allant de 56 à 122 nucléotides en aval de celui-ci. De plus, chez 20 révertants, la séquence dupliquée est également ponctuée par une séquence poly(A) de 11 à 73 nucléotides de long. Chez trois révertants (16, 72 et 74), cette séquence poly(A) est absente.

Toutes les séquences situées en aval de la région ATCase sont de type Ty1 et diffèrent d'une souche à l'autre (Fig. 2.9). Majoritairement, cette région est composée d'une partie variable de l'ORF TYA suivie d'un LTR. Mais elle peut également être composée d'un LTR solo ou d'une partie variable de l'ORF TYB suivie d'un LTR. Ces séquences Ty n'ont généralement pas la même orientation que celles localisées à la jonction 5' et proviennent par conséquent d'autres structures Ty1.

L'analyse des 23 révertants, sélectionnés à partir d'une souche dans laquelle un élément Ty1 a été surexprimé, apporte deux informations importantes. Premièrement, l'induction de la transposition dans une souche *ura2 15-30-72* conduit à une augmentation de la fréquence des duplications. Deuxièmement, les remaniements sélectionnés correspondent tous à de la duplication génique ayant les mêmes caractéristiques que

ceux isolés à partir de la souche *ura2 15-30-72*. En effet, les éléments dupliqués sont généralement ponctués en aval par une séquence poly(A) et localisés entre des séquences de rétrotransposons de type Ty1. Ces données mettent en évidence un lien entre les éléments Ty et les mécanismes impliqués dans la formation des duplications géniques. Par la suite, nous allons développer les différentes hypothèses concernant les mécanismes possibles conduisant à ces réarrangements.

DISCUSSION

Le crible génétique basé sur l'allèle mutant du gène *URA2* a permis de mettre en évidence deux processus bien distincts par lesquels une séquence d'ADN peut se dupliquer : la duplication génique et la duplication segmentale. Dans ce chapitre, nous nous sommes plus particulièrement attachés à l'analyse moléculaire des révertants portant une duplication génique. Les résultats montrent clairement un lien entre les rétrotransposons Ty1 et la formation des duplications géniques. Deux hypothèses peuvent être formulées pour expliquer le rôle de ces rétrotransposons dans la formation des duplications géniques.

La première des hypothèses serait l'implication de la recombinaison homologue entre des éléments Ty dans la formation de ces duplications. Dans ce cas, l'événement peut s'expliquer par un processus en deux étapes. Dans un premier temps, un élément Ty s'insère dans le gène *URA2* au niveau de la région codant la seconde moitié de la CPSase. Cette première étape correspond donc à l'un des événements observés au niveau des révertants sélectionnés à partir de la souche *ura2 15-30-72* (Roelants *et al.*, 1995). Dans un second temps, un transfert d'information non-réciproque de type conversion génique va avoir lieu entre une séquence donneuse (rétrotransposon nouvellement inséré) et une séquence receveuse (un autre élément Ty ou un LTR localisés ailleurs dans le génome). Cette conversion pourra englober la séquence codant le domaine ATCase et conduire ainsi aux réarrangements observés. Si cette hypothèse de transposition-duplication est correcte, des séquences Ty1 devraient être trouvées près de la copie originelle et de la copie dupliquée. Cette dernière est effectivement insérée dans des séquences Ty, mais aucune séquence de ce type n'est observée à proximité de la séquence *ura2 15-30-72*. Cette hypothèse n'apporte donc pas une explication valable à la formation des duplications géniques sélectionnées.

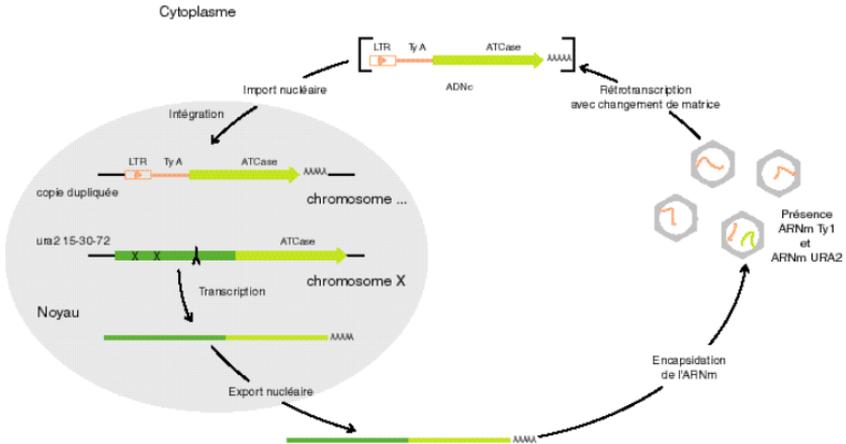


FIGURE 2.10 – Duplication génique via un mécanisme de rétroposition. Le fond gris délimite le noyau. L'ADN double brin est représenté par une flèche épaisse et son ARN par un rectangle mince.

La deuxième hypothèse serait qu'une fonction codée par l'élément Ty1 joue un rôle dans la formation des duplications géniques. La présence d'une séquence poly(A) à l'extrémité 3' de la copie dupliquée chez 3 révertants spontanés suggère fortement que cette copie a été formée par rétrotranscription à partir d'un ARNm du gène *URA2*. Les rétrotransposons constituent une source évidente d'activité transcriptase inverse. L'induction de la transposition dans une souche *ura2 15-30-72* conforte cette hypothèse car elle conduit à une augmentation de la fréquence des duplications géniques. En effet, l'analyse des 23 révertants montre que les événements sélectionnés correspondent à des duplications géniques ayant une structure similaire à celles trouvées chez les 4 révertants spontanés. Les duplications géniques sont localisées au voisinage de séquences Ty1 et ponctuée à leur extrémité 3' d'une séquences poly(A) chez la majorité des révertants.

Toutes ces données nous conduisent à proposer un mécanisme de rétroposition pour expliquer la formation des duplications sélectionnées (Fig. 2.10) (Deer *et al.*, 1991; Zhang, 2003). Le mécanisme le plus probable serait donc la formation d'un ADNe issu

de la rétrotranscription de l'ARNm correspondant à une partie du gène *URA2*. Pour être rétrotranscrit, l'ARNm doit vraisemblablement être empaqueté dans les particules pseudovirales (VLP). En effet, il a été montré que lors de la purification de la transcriptase réverse, cette dernière se trouvait dans la fraction contenant les particules pseudovirales de Ty (Mellor *et al.*, 1990). De plus, Xu et Boeke (1990) ont montré que certains transcrits cellulaires (*TRP1* et *HIS3*) pouvaient se trouver associés aux particules isolées d'éléments Ty. Ainsi, des ARN messagers du gène *ura2 15-30-72* peuvent théoriquement être présents au niveau des particules pseudovirales.

Une fois rétrotranscrit, l'ADNc correspondant doit s'insérer dans un des chromosomes du génome. Cette intégration ne semble pas être aléatoire car dans l'ensemble des révertants, les copies dupliquées sont insérées au niveau de séquences Ty1. Cette observation montre donc que les éléments Ty1 apportent plus que l'activité transcriptase inverse. La présence des courtes régions de similarité (2 à 11 pb) entre la séquence TyA et l'extrémité 5' de la partie dupliquée du gène *URA2* constitue un indice quant au mécanisme et à la présence de séquence Ty1. La rétrotranscription de l'ARNm doit dans ce cas être initiée au niveau de la région poly(A) allant vers la séquence codant l'ATCase puis un changement sur une matrice Ty doit s'effectuer au niveau de la jonction 5' observée. Ce changement de matrice conduit à la formation de la structure chimère observée: un LTR suivi d'une partie variable de l'ORF TYA du rétrotransposon. Cette possibilité est cohérente avec l'hypothèse de la formation des transposons hybrides Ty1/Ty2. Ces transposons sont générés par une rétrotranscription avec un changement de matrice (Jordan et McDonald, 1998).

L'intégration de cette structure dans un chromosome peut être expliquée par deux processus distincts. L'ADNc ainsi formé peut être intégré à un nouveau site chromosomique via l'intégrase codée par le transposon. Cependant, on constate que les séquences Ty encadrant la copie dupliquée ne sont pas dans la même orientation et ne forment donc pas une structure similaire à un rétrotransposon Ty1 qui pourrait être un substrat pour l'intégrase. Le second processus qui permettrait l'intégration de cette structure est un événement de recombinaison homologue résultant d'une conversion génique entre un élément Ty résident et l'extrémité 5' de l'ADNc. En effet, le site d'insertion précis déterminé chez deux révertants (Rev 25 et Rev 56) correspond à un élément Ty1 résident et non pas à un rétrotransposon nouvellement inséré. Ce processus semble donc plus conforme aux résultats obtenus. Cette hypothèse concernant l'intégration de l'ADNc sera rediscutée dans le chapitre III.

Dans notre crible de sélection, rendant les souches prototrophes pour l'uracile, les duplications doivent permettre l'expression de l'activité ATCase. Pour être transcrite, la copie dupliquée doit être insérée en aval d'un promoteur fonctionnel. Chez 27 révertants étudiés dans ce chapitre, la séquence codant l'ATCase est fusionnée en phase avec une partie variable de l'ORF TYA suivie d'un LTR (Fig. 2.9). Dans ces souches, le promoteur localisé dans la structure LTR permet vraisemblablement d'activer la transcription du gène mosaïque. Dans les 4 révertants restants, la séquence codant l'ATCase est en aval d'une structure Ty présente en sens opposé (Fig. 2.9). La situation est, ici, semblable au remaniement correspondant à l'insertion d'un rétrotransposon Ty1 au niveau du locus *ura2 15-30-72*. Cet élément est capable d'initier la transcription à partir de la séquence orientée en sens inverse par rapport à la région codant le domaine ATCase (Roelants *et al.*, 1997).

Par ailleurs, pour le révertant Rev 56, on peut noter qu'il n'y a pas de corrélation entre les résultats obtenus par Tail-PCR et par hybridation des chromosomes (chapitre I §1.3). En effet, l'hybridation indique que le chromosome X est porteur de l'élément dupliqué alors qu'il se trouve inséré sur le chromosome XVI. Ce résultat suggère qu'en plus de la duplication génique un autre réarrangement chromosomique s'est produit chez Rev56. De plus, dans un grand nombre de révertants sélectionnés dans un contexte de surexpression d'un Ty, une modification du profil électrophorétique a été mise en évidence. En effet, on peut observer pour certaines souches une bande supplémentaire par rapport à celui de la souche *ura2 15-30-72* ayant permis leur sélection. L'ensemble de ces observations peut être mis en corrélations avec les résultats obtenus par Levine *et al.* (1990) qui ont montré que l'insertion d'un ADNc est souvent associée à un réarrangement chromosomique. Pour l'ensemble de ces souches, il serait intéressant de procéder à une analyse globale du nombre de copie de gène (copie dupliquée ou délétée) grâce à la technique d'hybridation génomique comparative par microarray (Winzeler *et al.*, 1999) afin d'identifier quels sont les remaniements génomiques responsables des modifications du caryotype.

Ainsi, les résultats présentés dans ce chapitre nous conduisent à proposer un modèle montrant que des gènes peuvent provenir d'un processus de rétroposition (Fig.2.10) et soutient l'hypothèse formulée par Fink (1987) pour expliquer le peu d'introns présents chez *S. cerevisiae*. La levure *S. cerevisiae* se différencie des cellules mammifères par de nombreux points dont la pauvreté en introns et des processus de recombinaison homologue très efficaces. En se fondant sur ces observations, Fink a proposé

que les rétrotransposons auraient remplacé les gènes initiaux par recombinaison homologue, entraînant la perte des introns.

Finalement, le processus à l'origine des événements de duplication génique sélectionnés semble être similaire à celui responsable de la formation des pseudogènes chez les eucaryotes supérieurs (Moran *et al.*, 1999; Esnault *et al.*, 2000). Les réarrangements chromosomiques décrits dans ce chapitre constituent le premier exemple expérimental de duplication génique par rétroposition chez *S. cerevisiae* et indiquent que l'activité transcriptase inverse des rétrotransposons à LTR de type Ty1 est capable de rétrotranscrire des ARNm cellulaires dont les produits forment des copies dupliquées de gène.

3

DUPLICATIONS SEGMENTALES EN PHASE HAPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

RÉSUMÉ

Des événements de duplication segmentale ont été mis en évidence chez 4 révertants haploïdes spontanés. La détermination de la taille de la région dupliquée montre que celle-ci comprend la séquence codant le domaine ATCase ainsi qu'une série de gènes contigus et par conséquent codupliqués. Les segments dupliqués ont une taille variable comprise entre 5 et 90 kb. Afin de connaître la localisation précise et l'orientation de ces segments dupliqués, les régions situées en amont et en aval ont été déterminées. Les résultats montrent que les réarrangements correspondent à des duplications segmentales en tandem direct. Par ailleurs, la partie dupliquée du gène URA2 est insérée dans une région intergénique ou dans un gène non essentiel. Dans ce dernier cas, la fusion entre le gène et le fragment dupliqué s'est produite en phase, ce qui entraîne la formation d'un gène mosaïque.

*Plusieurs mécanismes hypothétiques peuvent être avancés afin d'expliquer la formation de duplications segmentales en tandem direct. Bailey et al. (2003) ont montré que ces remaniements peuvent résulter d'un crossing-over inégal impliquant une recombinaison homologue. Afin de vérifier cette hypothèse, le rôle du gène RAD52, indispensable à la recombinaison homologue, dans la formation des duplications a été testé. À partir d'une souche *ura2 15-30-72 rad52Δ*, 6 révertants portant un événement de duplication ont été sélectionnés et analysés. L'ensemble des réarrangements sélectionnés correspond à des duplications segmentales ce qui démontre que le mécanisme est indépendant de la recombinaison homologue. La seconde hypothèse (Kozsul et al., 2004) serait une erreur lors de la réplication impliquant un mécanisme de type BIR (break-induced replication). Ce modèle semble être cohérent avec les duplications segmentales sélectionnées. Cependant, une sélection de révertants dans un contexte affectant la réplication serait nécessaire pour confirmer ce modèle.*

INTRODUCTION

Comme nous l'avons déjà mentionné à plusieurs reprises, l'analyse des séquences complètes de génomes donne une vision globale des séquences d'ADN dupliquées et permet plus particulièrement de déterminer leur proportion, leur taille, leur localisation... L'une des surprises majeures a été l'observation d'une proportion importante de duplications segmentales, c'est-à-dire de duplications de fragments comportant plusieurs gènes au sein de nombreux organismes eucaryotiques.

Du point de vue évolutif, ces événements de duplications segmentales peuvent non seulement avoir une répercussion au niveau des gènes dupliqués mais aussi au niveau du génome. Dans le cadre de l'évolution des gènes, ces segments dupliqués constituent un réservoir de gènes qui peuvent être soumis à différents processus. Ces gènes dupliqués codent des protéines qui ne sont pas essentielles pour la viabilité des organismes mais peuvent être à l'origine de la mise en place d'une spécificité fonctionnelle (subfonctionnalisation) ou encore d'une nouvelle fonction (néofonctionnalisation). Ces séquences sont également de bons substrats pour la recombinaison homologue. En effet, les duplications segmentales sont des régions très similaires et par conséquent peuvent être des sites préférentiels pour les réarrangements chromosomiques. Dans ce cadre, ces régions peuvent être impliquées dans l'évolution des génomes.

Chez l'homme, 5 % du génome est constitué de duplications segmentales récentes, c'est-à-dire de segments d'ADN dupliqués ayant une similarité de séquence supérieure à 90 % (Bailey *et al.*, 2001). La séquence complète du génome de *S. cerevisiae* a également révélé des blocs de duplication mais qualifiés de « fossiles » c'est-à-dire correspondant aux traces d'un événement de duplication ancien. Ces blocs ont une similarité de séquence bien plus faible que les duplications « récentes ». Wolfe et Shields (1997) ont ainsi mis en évidence 55 blocs de duplication de taille moyenne de 55 kb dans le génome de *S. cerevisiae*.

Cependant, bien que la présence des segments dupliqués qualifiés de « fossiles » chez *S. cerevisiae* soit attribuée à un événement de duplication complète de son génome, des travaux récents montrent qu'il ne s'agit pas de la seule et unique source impliquée dans la formation de telles duplications. En effet, en utilisant un crible de sélection basé sur le dosage des gènes *RPL20A* et *RPL20B* chez *Saccharomyces cerevisiae*, Koszul *et al.* (2004) ont sélectionné des duplications segmentales spontanées indépendamment de tout événement de polyploïdie. De plus, le séquençage systématique de 3 levures hémi-

ascomycètes : *Kluyveromyces lactis*, *Debaryomyces hansenii* et *Yarrowia lipolytica* a permis de mettre en évidence que le génome de ces espèces n'a pas été le siège d'un événement de duplication totale mais comporte cependant des blocs de duplication (Dujon *et al.*, 2004) :

- 8 blocs (un en tandem) chez *K. lactis*,
- 5 blocs (3 en tandem) chez *D. hansenii*,
- 2 blocs chez *Y. lipolytica*.

Ces blocs ne coïncident pas avec ceux retrouvés chez *S. cerevisiae*. De plus, ils sont trop peu nombreux pour suggérer qu'ils soient issus d'un événement de polypléidie. Ces observations permettent donc de conclure que d'autres mécanismes peuvent être à l'origine de la formation d'une duplication segmentale chez les levures et illustrent la multiplicité des moyens par lesquels une séquence d'ADN peut se dupliquer.

Dans le cadre de notre étude, le crible de sélection positive basé sur l'utilisation de l'allèle mutant du gène *URA2* nous a permis d'aborder l'analyse des duplications segmentales. En effet, parmi 8 révertants spontanés portant une duplication de la région codant l'ATCase, sélectionnés à partir de la souche haploïde *ura2 15-30-72* (Roelants *et al.*, 1995), 4 d'entre eux (Rev 9, Rev 27, Rev 46 et Rev 52) portent une duplication correspondant à un segment d'une taille supérieure à 6kb (chapitre I). Ces remaniements sont donc issus d'un événement de duplication segmentale.

Dans ce troisième chapitre, nous nous sommes attachés à caractériser plus précisément ces duplications présentes dans ces 4 souches. La détermination de la structure des segments dupliqués ainsi que de leur environnement chromosomique constitue une étape essentielle qui va nous permettre de formuler des hypothèses quant aux mécanismes qui sont impliqués dans ce type de remaniements. Dans cette optique, nous nous sommes intéressés à déterminer deux caractéristiques physiques: la taille et la localisation chromosomique précise des segments dupliqués.

Dans le premier chapitre, la stratégie utilisée avait permis de différencier les duplications géniques des duplications segmentales avec pour ces derniers une estimation très grossière de la taille. Ainsi, nous avons développé une autre stratégie reposant sur la technique d'hybridation génomique comparative sur microarray pour caractériser plus précisément la taille des segments dupliqués. Cette stratégie, basée sur l'hybridation de l'ADN génomique sur microarray, permet de mettre en évidence le changement du nombre de copies de gènes (délétion ou duplication) intervenu chez les révertants par

rapport à la souche ayant permis de les sélectionner. Cette approche permet également de déceler d'éventuels autres remaniements qui se seraient produits simultanément à la duplication de la région codant le domaine ATCase.

Nous nous sommes également attachés à déterminer la localisation chromosomique précise de ces segments ainsi que leur orientation relative. En effet, dans le premier chapitre, l'analyse de ces révertants a permis de localiser le chromosome porteur du segment dupliqué. Chez 3 révertants (Rev 9, Rev 27, Rev 46), la région dupliquée s'est insérée dans le chromosome X porteur de l'allèle mutant *ura2 15-30-72*. Les remaniements présents dans ces souches correspondent donc à des duplications intrachromosomiques. Par contre chez le révertant Rev 52, la région dupliquée s'est insérée dans le chromosome XI. La duplication est alors qualifiée d'interchromosomique. Afin d'avoir une localisation plus précise, nous avons identifié les sites d'insertion de la copie dupliquée par le séquençage des régions flanquantes en 5' et 3'.

L'ensemble de ces résultats nous a permis de formuler des hypothèses de mécanismes pouvant être à l'origine de telles duplications. De plus, nous allons également examiner et discuter de l'implication de la recombinaison homologue dans les événements de duplications géniques et segmentales.

RÉSULTATS

1. Détermination de la taille des régions dupliquées

Dans le chapitre I, une première série d'expériences a permis de montrer que 4 révertants portent une duplication segmentale. La souche Rev 52 porte une duplication interchromosomique d'environ 10 kb couvrant la région comprise entre les ORFs YJL130c à YJL137c. Ce segment a été caractérisé par une série d'hybridations successives de type Southern directement sur les chromosomes séparés par électrophorèse en champs alternés. Les souches Rev 9, Rev 27 et Rev 46 portent également une duplication segmentale, mais localisée sur le chromosome X. Une série d'hybridations de type Southern, réalisée sur l'ADN génomique par une analyse de profil de restriction, a montré que la région dupliquée a une taille au moins supérieure à 6kb.

Pour déterminer de manière précise la taille des régions dupliquées présentes chez les 4 révertants, nous avons utilisé la technique d'hybridation génomique comparative par microarray (array CGH) (Winzeler *et al.*, 1999). Basée sur l'hybridation d'ADN

génomique sur microarrays, cette technique permet de détecter le changement du nombre de copies de gènes (délétion ou duplication) d'une souche par rapport à une souche de référence. En effet, l'intensité de l'hybridation de chaque ORF est proportionnelle au nombre de copies dans une souche donnée. Dans notre étude, cette technique a permis de comparer le changement du contenu génomique entre les révertants et la souche ancestrale ayant permis de les sélectionner. En effet, pour une ORF donnée, si le ratio de l'intensité du signal observé pour le révertant par rapport à celui observé pour la souche ancestrale est de 1 alors une seule copie est présente dans les deux souches. Par contre, si le ratio est de 2 ou de 1/2 alors cela signifie que le révertant possède respectivement une copie dupliquée ou une copie délétée de l'ORF concernée. Pour la simplification de la représentation graphique, les résultats seront sous la forme du \log_2 du ratio. En effet, les \log_2 de 2 (duplication), 1/2 (délétion) et 1 (aucun changement) sont égaux à 1, -1 et 0.

L'ADN génomique de chacun des 4 révertants a donc été hybridé sur microarray et comparé à la souche *ura2 15-30-72*. Le ratio de l'intensité du signal de chaque révertant par rapport à la souche *ura2 15-30-72* a été déterminé pour chaque ORF. Dans les 4 révertants, l'analyse globale du nombre de copies de gènes montre que les seuls gènes dupliqués appartiennent au chromosome X et sont localisés en aval de l'ORF YJL130c (*URA2*). Aucune autre région chromosomique n'a été le siège d'un changement du nombre de copies de gènes (délétion ou duplication).

Un graphique du \log_2 du ratio en fonction de la localisation chromosomique a été tracé pour chaque ORF du bras gauche du chromosome X (Fig. 3.1). Ces données montrent que les segments dupliqués sont différents d'un révertant à l'autre. Ces régions sont formées d'un nombre variable d'ORFs contigues avec la région codant le domaine ATCase et par conséquent codupliquées avec celle-ci. Le premier gène dupliqué à partir du centromère correspond toujours à la région codant le domaine ATCase (YJL130c). Dans notre système de sélection, les événements de duplication doivent permettre la réactivation du domaine ATCase, c'est-à-dire l'insertion de la séquence correspondante en aval d'un promoteur fonctionnel. Ceci ne peut se faire que si cette séquence se retrouve à l'une des extrémités du segment dupliqué. L'autre extrémité du segment est variable. En effet, les différentes duplications identifiées couvrent une région située entre l'ORF YJL130c et l'ORF YJL189w. La détermination de la taille exacte montre effectivement une grande variabilité entre ces éléments dupliqués (Tableau 3.1). Ils ont

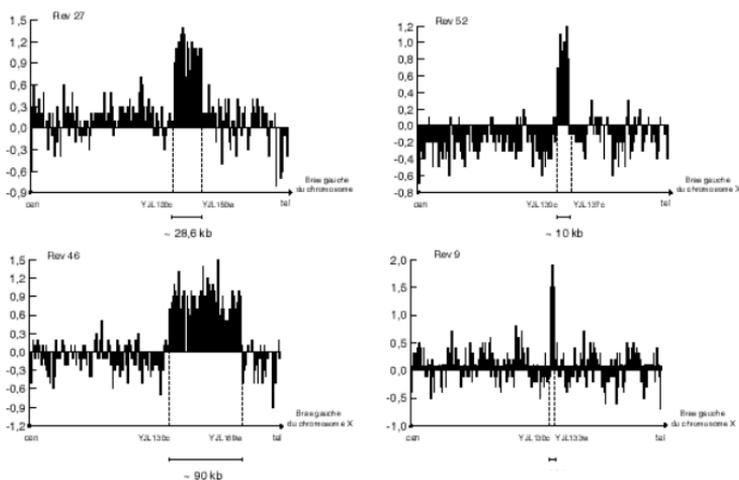


FIGURE 3.1 – Identification des segments dupliqués chez les 4 révertants spontanés.

Représentation graphique des résultats de l'hybridation génomique comparative sur microarray.

L'axe des abscisses correspond au bras gauche du chromosome X c'est-à-dire aux ORFs de YJL001w à YJL204c.

L'axe des ordonnées représente le \log_2 du ratio de l'intensité du signal observé pour le révertant par rapport à celui observé pour la souche ancestrale. La zone entre pointillés correspond à l'ensemble des ORFs codupliquées avec la région codant le domaine ATCase.

Révertants	Localisation	Limite en 5' *	Limite en 3' *	Taille du segment
Rev 9	X	YJL130c	YJL130w	6 kb
Rev 27	X	YJL130c	YJL150w	29 kb
Rev 46	X	YJL130c	YJL189w	90 kb
Rev 52	IX	YJL130c	YJL137c	10 kb

TABLEAU 3.1 – Taille des segments dupliqués présents chez les révertants spontanés.

* ORFs localisées aux deux extrémités du segment dupliqué déterminées par hybridation génomique comparative sur microarray.

une taille variant entre 6 kb et 90 kb.

Les résultats obtenus par la technique d'hybridation génomique comparative ont été validés par une série d'hybridation ADN/ADN de type Southern.

Pour les souches Rev 9, Rev 27 et Rev 46, l'événement sélectionné correspond à une duplication intrachromosomique. La marche sur le chromosome a alors été réalisée sur l'ADN génomique par l'analyse de profils de restriction. Le segment dupliqué se retrouve dans un nouvel environnement *Bam*HI (Fig. 3.2). Il diffère par rapport à la région originelle par les deux sites *Bam*HI localisés en amont et en aval.

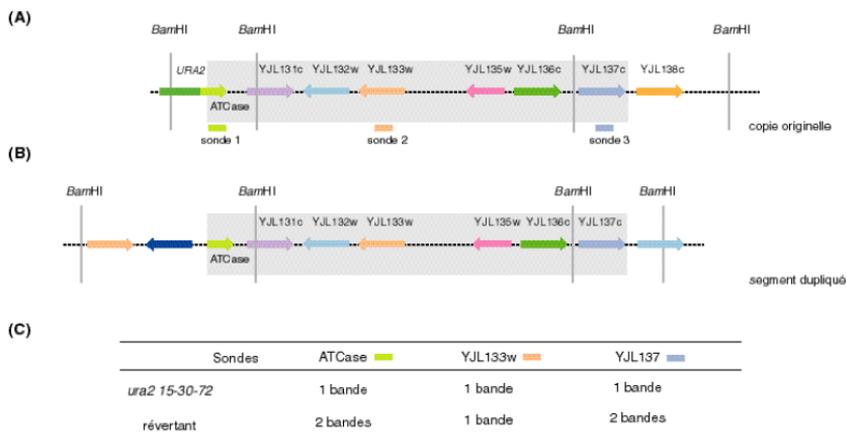


FIGURE 3.2 – Stratégie de marche sur le chromosome par analyse des profils de restriction.

(A) Carte de restriction *Bam*HI de la région *ura2* 15-30-72.

(B) Carte de restriction théorique de la région dupliquée.

(C) Hybridation de type Southern de l'ADN génomique digéré par *Bam*HI.

Ainsi, les fragments *Bam*HI localisés en 5' et 3' de l'élément dupliqué varient en taille par rapport aux fragments originels. Une hybridation de type Southern de l'ADN génomique digéré par l'enzyme *Bam*HI avec les sondes spécifiques à ces 2 fragments va donc révéler deux bandes. L'hybridation des ADN génomiques des révertants Rev 9, Rev 27 et Rev 46 avec la sonde correspondant au dernier gène dupliqué (respectivement YJL133w, YJL150w, YJL189w) révèle effectivement deux bandes confirmant les résultats issus de l'hybridation comparative sur microarray.

Pour le révertant Rev 52 porteur d'une duplication interchromosomique, la marche sur le chromosome, réalisée par hybridation de type Southern directement sur les chromosomes séparés par électrophorèse en champs alternés, montre que le segment dupliqué englobe la région codant le domaine ATCase ainsi que les ORFs YJL131c à YJL137c. Les résultats de la marche sur le chromosome et de l'hybridation génomique comparative sont donc concordants.

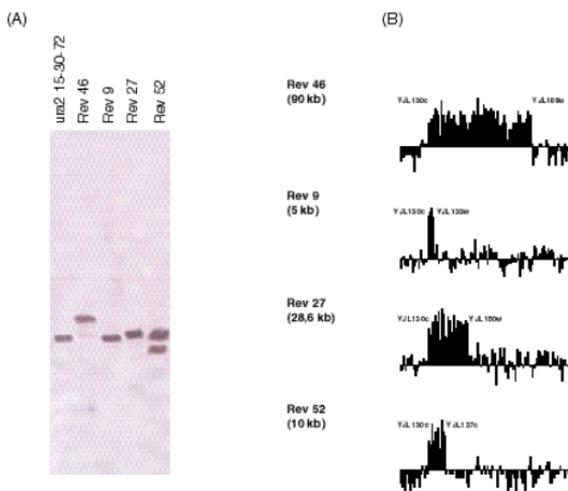


FIGURE 3.3 – Corrélation entre la migration du chromosome X et la taille des segments dupliqués.

(A) Révélation par hybridation de type Southern avec une sonde spécifique de l'ATCase des chromosomes séparés par électrophorèse en champs alternés.

(B) Représentation graphique des résultats de l'hybridation génomique comparative sur microarray.

Par ailleurs, l'hybridation des chromosomes avec la sonde ATCase, dans le chapitre I, a montré une augmentation de la taille du chromosome X porteur de l'élément de duplication chez les révertants Rev 27 et Rev 52 (Fig. 3.3). Cette augmentation de taille peut être estimée à environ 20 kb pour Rev 27 et environ 100 kb pour Rev 52 par comparaison avec la souche de référence *ura2 15-30-72*. Ces augmentations correspondent effectivement à la taille de la région dupliquée et confirment donc que la variation de migration du chromosome X observée chez ces révertants est liée à l'événement de duplication.

L'ensemble de ces données montrent que les remaniements présents dans les révertants analysés correspondent donc à des duplications de segments ayant une taille variable. Il est intéressant de souligner qu'aucun autre remaniement chromosomique de type délétion ou duplication n'a été mis en évidence au sein de leur génome.

2. Analyse de la séquence des sites d'insertions de la région dupliquée

Afin de connaître la localisation précise ainsi que l'orientation relative des segments dupliqués, l'identification des sites d'insertion de la copie dupliquée a été réalisée par le séquençage des régions flanquantes en 5' et 3'. La détermination de ces points de jonction peut également être primordiale car ces séquences constituent les « cicatrices » de l'événement de duplication.

2.1. Détermination du site d'insertion en amont de la duplication

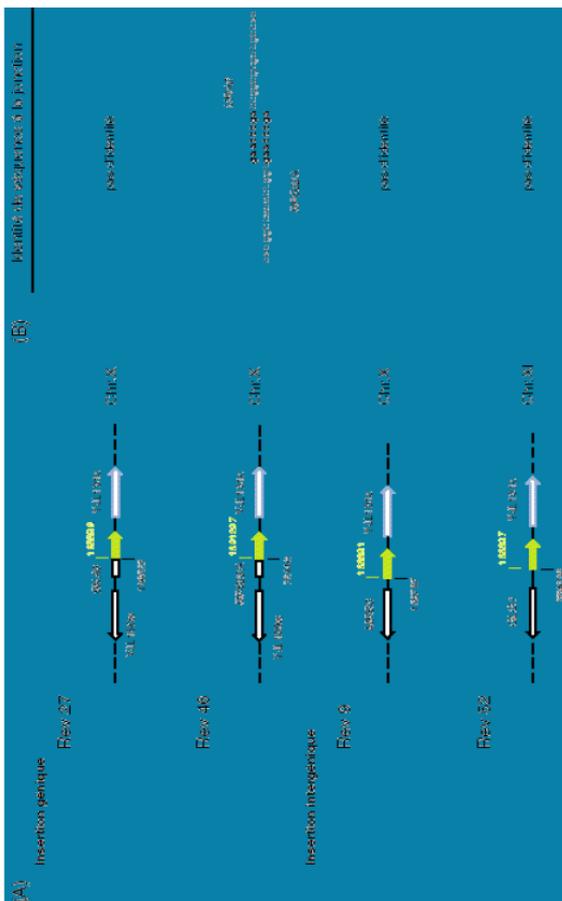
La jonction 5' de la région dupliquée a été clonée en utilisant une stratégie d'insertion et d'excision de plasmide (procédure détaillée dans le chapitre I). Pour rappel, la première étape consiste à insérer un plasmide intégratif en aval du point de départ de la duplication. Après l'étape d'intégration, l'ADN génomique des transformants a été digéré par l'enzyme de restriction *HindIII*. Le produit de digestion a alors été ligué. Afin de récupérer le plasmide contenant la jonction inconnue, des souches *E. coli* ont été transformées avec le produit de ligation. Les plasmides ont été récupérés par extraction et les jonctions ont été séquencées. Les séquences ont alors été utilisées pour faire une comparaison contre la banque de données Saccharomyces Genome Database à l'aide du programme BLAST (<http://seq.yeastgenome.org/cgi-bin/SGD/nph-blast2sgd>).

L'analyse des jonctions montre que le point de départ des copies dupliquées est variant d'une souche à l'autre. Mais il est toujours localisé en aval des trois mutations présentes dans l'allèle mutant. Il peut être localisé à la fin de la région codant la CPSase ou dans la région codant le domaine cryptique DHOase-like (Fig. 3.4A).

La partie du gène *URA2* dupliquée peut être insérée dans un gène ou dans une région intergénique (Fig. 3.4A). Dans les révertants Rev 27 et Rev 46, les segments dupliqués se sont insérés dans un gène. L'extrémité 5' consiste en une fusion en phase avec le gène *SNA3* (YJL151c) pour le révertant Rev 27 et le gène *RPS22A* (YJL190c) pour le révertant Rev 46. Cette fusion conduit à la formation d'un gène mosaïque composé de l'extrémité 5' du gène tronqué (*SNA3* ou *RPS22A*) et de la partie dupliquée du gène *URA2* codant le domaine ATCase. Ces deux gènes receveurs (*SNA3* ou *RPS22A*) sont des gènes non-essentiels. Ce résultat n'est pas surprenant dans la mesure où la sélection a été réalisée au niveau haploïde et par conséquent l'insertion dans un gène essentiel ne serait pas viable. Dans les deux autres révertants (Rev 9 et Rev 52), la séquence dupliquée est insérée dans une région intergénique qui est respectivement localisée en aval des ORF YJL133w (*MRS3*) et YKR058w (*GLG1*). Dans ces deux souches, la fusion n'entraîne pas la formation de gène mosaïque. La séquence codant le domaine ATCase est donc vraisemblablement transcrite à partir de cette région intergénique.

Nous avons recherché la présence d'identité de séquence entre la région receveuse et la partie dupliquée du gène *URA2*. Chez le révertant Rev 46, la jonction présente une courte région d'identité de séquence (7 pb) entre le gène *RPS22A* et la partie dupliquée du gène *URA2* (Fig. 3.4B). Par contre, dans les 3 autres souches (Rev 9, Rev 52 et Rev 27), les jonctions correspondantes ne présentent aucune similarité de séquence.

L'ensemble de ces données confirme que les remaniements présents dans les souches Rev 9, Rev 27 et Rev 6 correspondent à une duplication intrachromosomique alors que dans le révertant Rev 52, il s'agit d'une duplication interchromosomique. Cette dernière présente une particularité intéressante. En effet, le segment dupliqué est localisé sur le chromosome XI et plus particulièrement au niveau d'une des deux régions dupliquées du bloc 41 défini par Wolfe et Shields (1997). Le bloc 41 est composé de deux régions dupliquées. La première est située sur le chromosome X entre les ORFs YJL078c (*YURI*) et YJL139c (*PRY3*) et contient le gène *URA2* (YJL130c). La seconde, qui dans la souche Rev 52 est porteuse du segment dupliqué, est localisée sur le chromosome XI entre les ORFs YKR061w (*KTR2*) et YKR013w (*PRY2*). Cette région ne contient pas de paralogue du gène *URA2* dans la souche *ura2 15-30-72*. Par contre, le gène *GLG1* (YKR058w) qui correspond à la région d'insertion forme avec le gène *GLG2* (YJL137c) une paire de gènes homologues. L'insertion de l'élément dupliqué ne semble donc pas aléatoire car elle se fait dans une région similaire au chromosome X.



2.2. Détermination du site d'insertion en aval du segment dupliqué

La connaissance du site d'insertion en 5' ainsi que de la taille de la région dupliquée nous a permis de déterminer le site d'insertion en 3'. Nous avons émis l'hypothèse que l'insertion du segment n'avait été accompagnée d'aucun autre remaniement aux bornes. Dans ce cas de figure, l'intégration de la région dupliquée interrompant un gène ou une région intergénique va déplacer la deuxième partie de ce gène ou de la région intergénique directement en aval de la duplication. La jonction 3' devrait donc correspondre à l'union de l'extrémité 3' du segment et la région chromosomique contigue au site d'insertion 5' dans la souche *ura2 15-30-72*. Afin de tester cette hypothèse et de caractériser les jonctions 3', nous avons choisi pour chaque révertant un oligonucléotide situé dans la dernière ORF dupliquée et un autre situé dans la région contigue au site d'insertion 5' dans la souche *ura2 15-30-72* (Fig. 3.5).

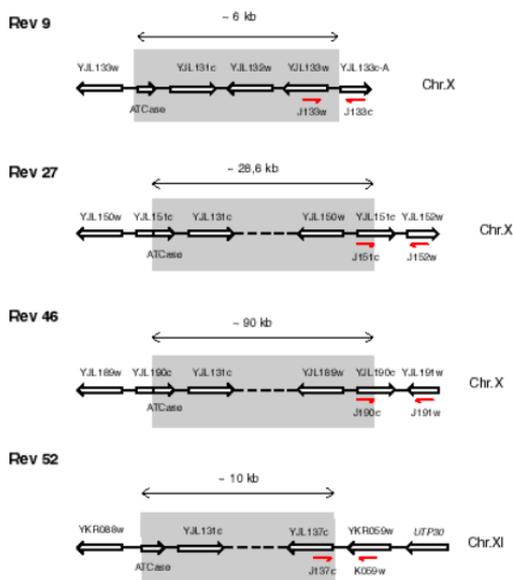


FIGURE 3.5 – Site d'insertion chromosomique en 3' des segments dupliqués. Les flèches rouges correspondent aux oligonucléotides ayant permis d'amplifier la jonction 3'.

Par exemple, dans la souche Rev 9, le segment dupliqué est inséré dans la région intergénique située en aval de l'ORF YJL133w (*MRS3*). Il s'étend de la partie dupliquée du gène *URA2* (YJL130c) jusqu'à l'ORF YJL133w. Deux oligonucléotides ont été choisis: J133w spécifique à l'ORF YJL133w (correspondant la fin du segment) et J133c spécifique de l'ORF YJL133c-A (l'ORF contigu au site d'insertion en 5' dans la souche *ura2 15-30-72*) afin d'amplifier la jonction. Nous avons procédé de la même manière pour les 3 autres souches.

Dans chaque cas, les oligonucléotides choisis ont permis d'amplifier un fragment qui a été séquencé. Les séquences ont alors été utilisées pour faire une comparaison contre la banque de données Saccharomyces Genome Database à l'aide du programme BLAST (<http://seq.yeastgenome.org/cgi-bin/SGD/nph-blast2sgd>). Ces comparaisons montrent que le site d'insertion en 3' correspond effectivement à la région chromosomique contigue au site d'insertion 5' dans la souche *ura2 15-30-72* (Fig. 3.5).

Chez les révertants Rev 9, Rev 27 et Rev 46, les duplications intrachromosomiques présentent une particularité très intéressante (Fig. 3.6). En effet, en regardant la structure globale du chromosome X porteur d'une duplication, les résultats montrent que le segment dupliqué correspond à la région chromosomique comprise entre le gène *URA2* et le site d'insertion. De plus, l'élément dupliqué a la même orientation que la copie originelle. L'ensemble de ces données permettent de conclure que l'événement présent dans ces 3 souches correspond à une duplication en tandem direct (Fig. 3.6).

Dans le révertant Rev 52, la région dupliquée est insérée entre les ORFs YKR058w (*GLG1*) et YKR059w (*TIF1*) sur le chromosome XI. Le dernier gène dupliqué correspond au gène *GLG2* paralogue de *GLG1* (région d'insertion). On constate donc une similarité de structure avec les autres révertants : le segment dupliqué peut être considéré comme étant en tandem avec le segment ancestral « homologue » appartenant au bloc 41 de duplication ancestrale défini par Wolfe et Shields (1997). Ces données renforcent l'hypothèse selon laquelle l'insertion n'est pas aléatoire.

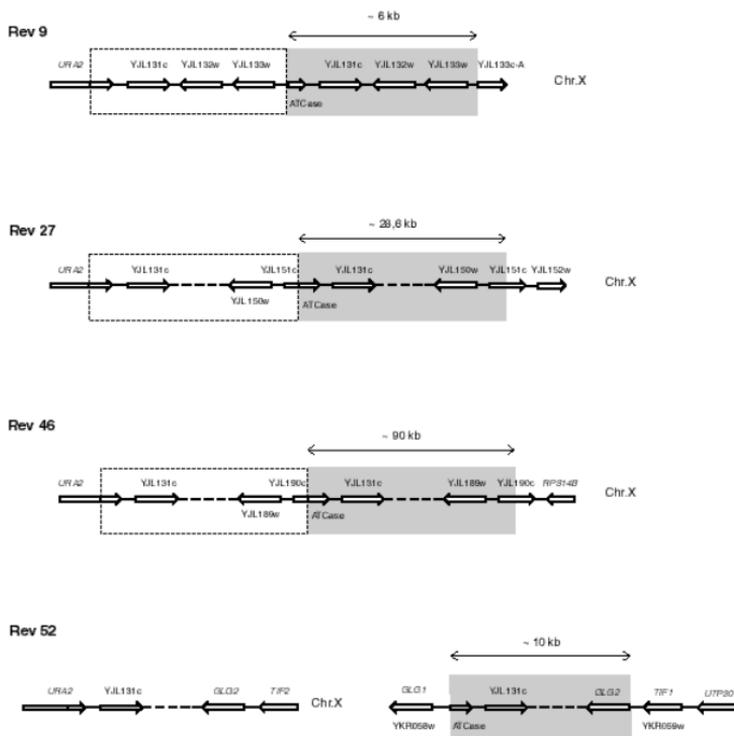


FIGURE 3.6 – Duplications segmentales en tandem direct.
 Les rectangles gris correspondent aux régions dupliquées.
 Les rectangles en pointillés correspondent aux régions originales.

3. Influence de la recombinaison homologue sur les événements de duplication segmentale et de duplication génique

L'analyse moléculaire des révertants haploïdes a permis de définir deux processus par lesquels une séquence d'ADN peut se dupliquer : la duplication segmentale et la duplication génique. Pour les événements de duplication segmentale, le mécanisme généralement évoqué pour expliquer leur formation correspond à un crossing-over inégal entre deux séquences similaires non-allèles localisées à des sites différents (Zhang, 2003 ; Bailey *et al.*, 2003). Concernant la duplication génique, les résultats décrits dans le chapitre II suggèrent fortement que la duplication limitée à une partie du gène *URA2* est issue d'un mécanisme de rétroposition c'est-à-dire de la formation d'un ADNc suite à la réverse transcription de l'ARNm correspondant suivie de son intégration dans le génome. Comme la région dupliquée est encadrée de part et d'autre par des séquences de rétrotransposon Ty1, on peut logiquement se demander si l'intégration chromosomique se fait par recombinaison homologue entre de telles séquences ou via l'intégrase codée par les éléments Ty1.

Pour comprendre dans quelle mesure la recombinaison est impliquée lors de l'intégration de l'ADNc et plus généralement dans la genèse des duplications segmentales, nous avons examiné l'influence du gène de recombinaison *RAD52* sur la sélection de tels remaniements. Ce gène, ainsi que les gènes *RAD51* et *RPA*, sont impliqués dans la recombinaison homologue. Le produit du gène *RAD52* (*YML032c*), une protéine de 471 aa, se fixe sur les extrémités d'une cassure d'ADN double brin et permet l'appariement de deux régions d'ADN complémentaires de grande taille (de l'ordre de 200 bp) pour sa réparation. La protéine Rad52p est indispensable pour la recombinaison homologue. Ainsi, nous avons regardé s'il y a une dépendance des deux types d'événements de duplication par rapport au gène *RAD52*.

3.1. Sélection

Dans le cadre de l'étude des événements de délétion au laboratoire, une collection de révertants a été isolée à partir de la souche *ura2 15-30-72 rad52Δ :: kanMX4* dans laquelle le gène *RAD52* a été remplacé par le gène *kanMX4* (Welcker *et al.*, 2000).

Les sélections de révertants Ura⁺ dans ce contexte génétique ont permis d'isoler 19 révertants indépendants dans lesquels un remaniement chromosomique a conduit à la réactivation fonctionnelle du gène *URA2*. Dans 6 révertants, une première analyse a

montré un profil de restriction *Bam*HI conforme à un événement de duplication (Welcker, 2000). Ce profil révèle deux bandes : l'une de 6,8 kb correspondant au locus *URA2* et l'autre de taille variable correspondant à une partie dupliquée du gène *URA2* qui s'est insérée dans une région différente du génome qui se caractérise par un environnement *Bam*HI différent.

Nous avons repris ces 6 révertants pour une étude fine de la structure de l'élément dupliqué. Cette caractérisation moléculaire va permettre de tester l'implication de la recombinaison homologue dans la genèse des duplications géniques ou segmentales.

3.2. Caractérisation moléculaire

Comme précédemment, la détermination de la localisation chromosomique ainsi que la taille des éléments dupliqués a été entreprise.

3.2.1. Localisation chromosomique de la région dupliquée

La séparation des chromosomes des 6 révertants par électrophorèse en champs alternés suivie d'une hybridation de type Southern avec une sonde spécifique de l'ATCase a permis de montrer que les régions dupliquées sont localisées uniquement sur le chromosome X (Fig. 3.7B). L'analyse du caryotype électrophorétique des 6 révertants ne montre pas de variation du profil de migration par rapport à celui de la souche de départ ayant permis de les sélectionner (Fig. 3.7A).

De plus, l'hybridation des chromosomes avec la sonde ATCase ne montre pas d'augmentation de la taille du chromosome X porteur de la duplication. Ces données permettent de conclure que le remaniement présent dans ces souches correspond à une duplication génique ou à une duplication d'un segment de taille inférieure à 20 kb.

3.2.2. Détermination de la taille de la région dupliquée

Pour préciser la nature des événements de duplication présents chez les 6 révertants, nous avons déterminé la taille de ces éléments dupliqués grâce à la technique d'hybridation génomique comparative par microarray comme décrit précédemment (§3.1).

L'ADN génomique de chacun des 6 révertants a été hybridé sur microarray et comparé à la souche ancestrale *ura2 15-30-72 rad52Δ :: kanMX4*. Le ratio de l'intensité du signal entre chaque révertant et la souche ancestrale a été déterminé pour chaque ORF. Dans les 6 révertants, les résultats montrent que les seules régions dupliquées

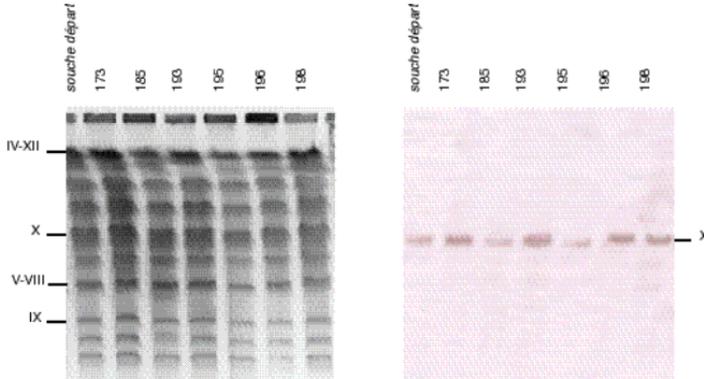


FIGURE 3.7 – Caryotypes après électrophorèse en champs alternés des révertants sélectionnés à partir de la souche *ura2 15-30-72 rad52 :: kanMX4*.

(A) Migration des chromosomes par PFGE et coloration au BET.

(B) Révélation par hybridation ADN/ADN de type Southern avec une sonde spécifique de la région codant le domaine ATCase.

appartiennent au chromosome X et sont localisées en aval du gène *URA2* (YJL130c) (Fig. 3.8). Les résultats montrent que les remaniements correspondent à des duplications de segment chromosomique qui sont formés par un nombre variable d'ORFs localisées en aval et codupliquées avec la région codant le domaine ATCase. Les duplications couvrent une région située entre l'ORF YJL130c et l'ORF YJL137c et ont une taille comprise entre 6 et 10 kb (Tableau 3.2).

L'ensemble de ces données indiquent que la délétion du gène *RAD52*, indispensable à la recombinaison homologue, n'a pas d'incidence sur la sélection de duplications segmentales. Par contre, il est intéressant à noter qu'aucune duplication génique n'a été sélectionnée dans un tel contexte génétique. Bien que l'observation ne porte que sur 6 révertants, il est étonnant qu'aucune duplication génique n'ait été sélectionnée. En effet, la sélection à partir de la souche *ura2 15-30-72* avait permis d'isoler 8 révertants dont 4 (50%) porteurs d'une duplication génique. Ces données suggèrent fortement que la recombinaison homologue a un rôle dans la genèse des duplications géniques.

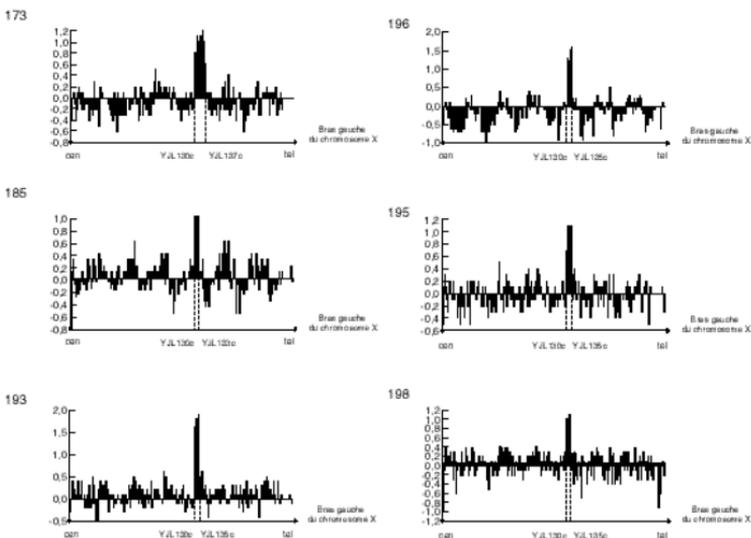


FIGURE 3.8 – Identification des segments dupliqués chez les révertants *rad52*. Représentation graphique des résultats de l'hybridation génomique comparative sur microarray. L'axe des abscisses correspond au bras gauche du chromosome X c'est-à-dire aux ORFs de *YJL001w* à *YJL204c*. L'axe des ordonnées représente le \log_2 du ratio de l'intensité du signal observé pour le révertant par rapport à celui observé pour la souche ancestrale.

Révertants	Localisation	Limite en 5' *	Limite en 3' *	Taille du segment
173	X	Y.JL130c	Y.JL137c	10 kb
185	X	Y.JL130c	Y.JL133c	6 kb
193	X	Y.JL130c	Y.JL135c	7,5 kb
196	X	Y.JL130c	Y.JL135c	7,5 kb
198	X	Y.JL130c	Y.JL135c	7,5 kb
199	X	Y.JL130c	Y.JL133c	6 kb

TABLEAU 3.2 – Taille des segments dupliqués présents chez les révertants *rad52*.

* ORFs localisées aux deux extrémités du segment dupliqué déterminées par hybridation génomique comparative sur microarray.

DISCUSSION

La duplication de segment d'ADN joue un rôle très important dans l'évolution moléculaire par l'apport de matériel génétique pouvant évoluer vers de nouvelles fonctions. Chez *S. cerevisiae*, il a été montré qu'un événement de polyploïdie peut conduire à la formation de segments dupliqués (Wolfe and Shield, 1997 ; Kellis *et al.*, 2004). Cependant, les résultats de notre étude montrent une fois de plus qu'il ne s'agit pas de la seule et unique source entraînant l'apparition de telles duplications. En effet, dans ce troisième chapitre, nous nous sommes d'abord focalisés sur 4 révertants dans lesquels un événement de duplication segmentale conduit à la réactivation fonctionnelle de l'ATCase. Ces données apportent une preuve expérimentale qu'une souche haploïde de *S. cerevisiae* peut dupliquer une partie de son génome par une duplication segmentale indépendamment de tout événement de polyploïdie.

Parmi les révertants analysés dans ce chapitre, trois correspondent à une duplication intrachromosomique alors qu'un seul correspond à une duplication interchromosomique. Les segments dupliqués ont une taille variable allant de 5 à 90 kb couvrant une région située entre les ORFs YJL131c et YJL190c. Afin de réactiver le domaine ATCase, le segment doit s'insérer en aval d'un promoteur fonctionnel. Nous avons montré qu'il peut s'insérer au niveau d'une région intergénique ou interrompre un gène non essentiel. Dans le cas d'une insertion « génique », la séquence ATCase est fusionnée en phase avec la partie 5' d'un gène non essentiel. La phase de lecture est préservée lors de la fusion ce qui permet la formation d'un gène mosaïque. Dans ce cas, la région dupliquée codant le domaine ATCase peut ainsi bénéficier du promoteur du gène receveur. Cette formation de gènes mosaïques peut aussi constituer un exemple de fusion de domaines menant à une protéine multifonctionnelle (Eichler, 2001). Dans le cas de deux révertants, l'élément dupliqué est inséré dans une région intergénique qui doit certainement agir comme promoteur fonctionnel. Ces événements de duplication constituent, quant à eux, un exemple de la modification de la région promotrice qui peut avoir un rôle clé dans l'évolution moléculaire (Lynch, 2002).

L'analyse des révertants portant une duplication génique (chapitre II) a fortement suggéré un mécanisme via la rétroposition. Pour les duplications décrites dans ce chapitre, l'hypothèse d'une transcription réverse ne peut être retenue car la région dupliquée est beaucoup plus grande que l'ARNm correspondant au gène *URA2*. De plus, aucun événement de duplication segmentale n'a été isolé lors de la sélection réalisée à

partir de la souche *ura2 15-30-72* dans laquelle un élément Ty1 a été surexprimé (chapitre II §2). Or si les éléments Ty1 intervenaient fonctionnellement dans leur formation, il aurait été logique d'avoir une proportion non négligeable de duplications segmentales parmi les 23 révertants porteurs d'un événement de duplication.

D'autres hypothèses peuvent être formulées pour expliquer la genèse de duplications segmentales. La première serait que ces segments soient issus d'un crossing-over inégal résultant d'une recombinaison homologe entre deux séquences paralogues. Dans ce cas, l'événement peut s'expliquer par un processus en deux étapes (Fig. 3.9).

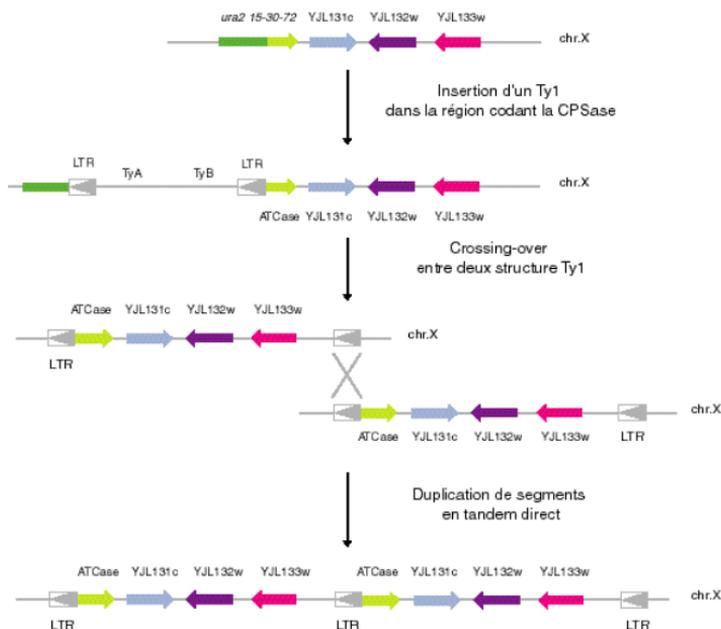


Figure 3.9: Modèle de formation du segment dupliqué en tandem direct par crossing-over inégal.

Dans un premier temps, un élément Ty1 s'insère dans le gène *URA2* au niveau de la région codant la seconde moitié de la CPSase. Cette première étape correspond à un des événements qui ont été observés au niveau des révertants sélectionnés à partir de la souche *ura2 15-30-72* (Roelants *et al.*, 1995). Dans un second temps, ce rétrotransposon nouvellement inséré recombine avec un autre élément Ty1 ou un LTR localisé ailleurs dans le génome, conduisant alors aux réarrangements observés. Récemment, Bailey *et al.* (2003) ont montré que des éléments Alu sont localisés au niveau des jonctions de la copie originelle et dupliquée dans le génome humain. Ils pensent que ces séquences répétées doivent être en partie responsables de l'explosion récente des duplications segmentales chez les primates et ceci par l'intermédiaire de crossing-overs inégaux. Chez *S. cerevisiae*, de nombreux éléments Ty comme des LTR ont été retrouvés au niveau des jonctions de réarrangements chromosomiques décrits à ce jour comme les translocations, (Fischer *et al.*, 2000; Dunham *et al.*, 2002).

Cependant, la présence des duplications segmentales, décrites dans ce chapitre, peut difficilement être expliquée par un tel processus qui implique la recombinaison homologue. La détermination des sites d'insertion en aval et en amont des segments dupliqués ne montre pas la présence de séquences répétées comme des séquences Ty qui pourraient être à l'origine d'un crossing-over inégal. Seule une courte identité de séquence (7pb) a été mise en évidence à la jonction 5' chez le révertant Rev 46. Pour les autres souches (Rev 9, Rev 52 et Rev 27), les jonctions ne présentent aucune similarité de séquence. De plus, la délétion du gène *RAD52* (indispensable à la recombinaison homologue) n'empêche pas la sélection de duplications segmentales. Ces données indiquent clairement que les mécanismes responsables de la formation des duplications sélectionnées sont indépendants de la recombinaison homologue.

La seconde hypothèse qui peut être proposée pour la formation de segments dupliqués en tandem direct correspond à une erreur lors de la réplication impliquant un mécanisme de type BIR (break-induced replication) (Malkova *et al.*, 1996 ; Morrow *et al.*, 1997; Ira and Haber, 2002 ; Davis and Symington, 2004). Le mécanisme peut être expliqué par deux étapes (Fig. 3.10). Dans un premier temps, la fourche de réplication va être perturbée par une cassure double brin située en aval de la région codant le domaine ATCase. Dans un second temps, le brin néosynthétisé coupé envahit une nouvelle zone de réplication qui va servir de matrice et entraîne la formation d'un segment dupliqué si cette région correspond à la séquence codant le domaine ATCase (Fig. 3.10).

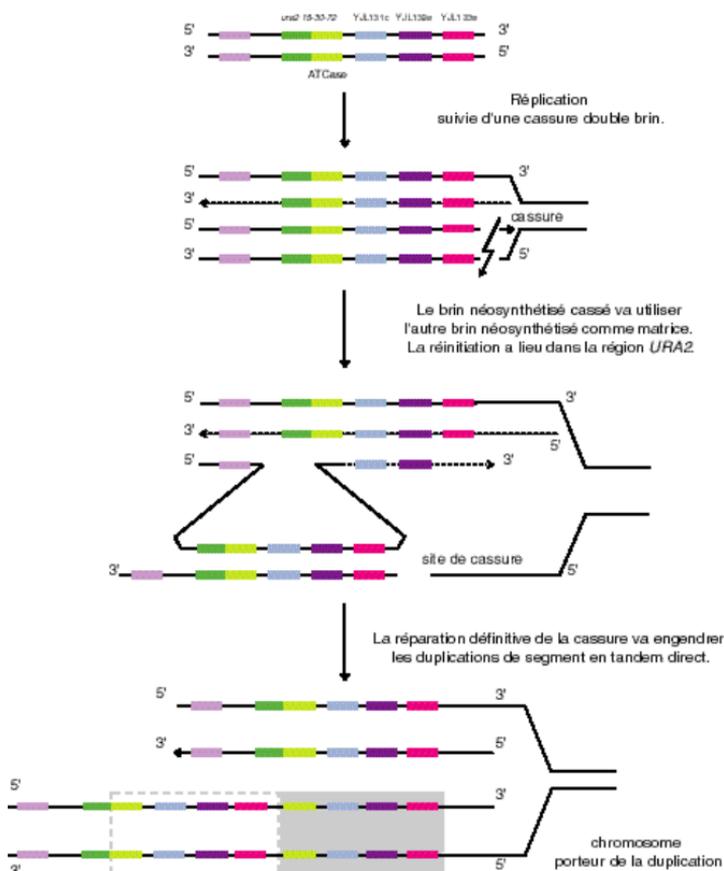


FIGURE 3.10 –Modèle de formation du segment dupliqué en tandem direct par erreur lors de la réplication impliquant un mécanisme de type BIR (break induced replication).

Ce modèle hypothétique est cohérent avec la formation des duplications segmentales en tandem direct décrits dans ce chapitre. Dans les révertants Rev 9, Rev 27 et Rev 46, seul le site d'insertion situé en amont du segment dupliqué est modifié. L'extrémité 3' correspond à une séquence sauvage ne portant aucune « cicatrice » du remaniement. Dans notre modèle, la réparation de la cassure double brin conduirait effectivement à une extrémité 5' modifiée alors que l'extrémité 3' correspondrait à une région sauvage.

Par contre, pour le révertant Rev 56, ce modèle n'est pas suffisant pour expliquer l'événement de duplication interchromosomique observé. Le remaniement pourrait résulter de deux réarrangements successifs. En effet, la première étape consisterait à la formation d'une duplication segmentale en tandem sur le chromosome X de la même manière que dans les 3 souches précédentes. Dans une deuxième étape, cette duplication a du être suivie d'un événement de recombinaison homologue entre les régions paralogues situées dans le bloc 41 situées sur les chromosomes X et XI qui a ainsi mené à cette duplication segmentale interchromosomique. Ces données indiquent que ces blocs de duplications ancestraux résultant d'une duplication complète du génome chez *S. cerevisiae* peuvent être sur le plan de leur structure à la source de réarrangements chromosomiques et ainsi jouer un rôle évolutif. Afin de tester le modèle proposé, il serait nécessaire de voir dans quelle mesure certains gènes impliqués dans la réplication le sont également dans la formation des duplications de segments en tandem. Le gène *RAD27* serait un bon candidat car il intervient dans l'élimination des fragments d'Okazaki intervenant dans la réplication. Ce gène n'étant pas essentiel à la survie cellulaire, il sera possible de construire un mutant *ura2 15-30-72 rad 27Δ* afin de sélectionner des révertants dans un tel contexte.

Au cours de ce chapitre, nous avons également testé l'implication de la recombinaison homologue dans les événements de duplication. Concernant les duplications segmentales, nous avons vu précédemment que leur formation est indépendante de la recombinaison homologue. Par contre pour les événements de duplication génique, les résultats semblent indiquer le contraire. Comme la région dupliquée est encadrée de part et d'autre par des séquences de rétrotransposon Ty1 (chapitre II), on pouvait logiquement se demander si l'intégration chromosomique de l'ADNc, formé après transcription inverse, se fait par recombinaison homologue ou via l'intégrase codée par les éléments Ty1. Or la délétion du gène *RAD52*, indispensable à la recombinaison homologue, conduit à l'absence de sélection de duplications géniques. Ces résultats montrent donc

que les mécanismes impliqués sont dépendants du gène *RAD52* et que l'insertion de l'ADNc après transcription inverse peut se faire par recombinaison homologue.

L'ensemble de ces données soutiennent l'hypothèse selon laquelle les mécanismes impliqués dans la duplication de séquences d'ADN sont pluriels. À partir de la souche haploïde *ura2 15-30-72*, deux types bien distincts de duplication ont pu être sélectionnés : la duplication génique et la duplication segmentale. Les résultats de cette étude apportent une preuve expérimentale que la duplication spontanée d'un gène particulier chez *S. cerevisiae* peut se faire par au moins deux moyens à l'état haploïde. Cependant, en utilisant une cellule haploïde, seule une partie des événements de duplication peut être isolée. L'insertion d'un élément dupliqué dans un gène essentiel ne va pas être viable pour un révertant haploïde. De même, l'insertion dans un gène entraînant une auxotrophie sera contre sélectionnée dans notre système. L'étude des événements de duplication trouvés au sein d'un génome diploïde permettra d'avoir une vue d'ensemble sur ces remaniements.

4

ANEUPLOÏDIES ET TRANSLOCATIONS NON-RÉCIPROQUES À L'ÉTAT DIPLOÏDE CHEZ *SACCHAROMYCES CEREVISIAE*

RÉSUMÉ

L'analyse de 14 révertants diploïdes portant un événement de duplication a permis de mettre en évidence deux autres moyens par lesquels une séquence d'ADN peut se dupliquer: l'aneuploïdie et la translocation non-réciproque.

Chez 6 révertants sélectionnés à partir d'une souche diploïde homozygote a/a ura2 15-30-72, la détermination de la taille de la région dupliquée montre que le bras gauche du chromosome X couvrant la région située entre le gène URA2 et les gènes subtélomériques a été dupliqué. Une analyse globale du nombre de copies de chaque gène a alors été réalisée grâce à la technique d'hybridation génomique comparative par microarray (aCGH) afin de préciser la nature du réarrangement présent dans ces souches. Les résultats suggèrent la formation d'un chromosome chimère surnuméraire issu de la fusion de deux régions dupliquées. La détermination du site d'insertion en amont de la partie dupliquée du gène URA2 confirme la fusion entre les deux structures chromosomiques.

Chez 2 révertants sélectionnés à partir d'une souche diploïde homozygote a/a ura2 15-30-72, la détermination de la taille montre également que le bras gauche du chromosome X couvrant la région située entre le gène URA2 et les gènes subtélomériques a été dupliqué. Cependant, les résultats de l'hybridation génomique comparative par microarray (aCGH) suggèrent que la région dupliquée a remplacé la région délétée formant ainsi un chromosome chimère. La détermination du site d'insertion en amont de la partie dupliquée du gène URA2 confirme la formation de ce chromosome chimère. La duplication observée chez ces révertants correspond donc à une translocation non-réciproque qui permet la réactivation de la région codant l'ATCase dans notre système.

Dans le cas des deux types de remaniements sélectionnés à partir d'une souche diploïde homozygote a/a ura2 15-30-72, aucune longue séquence homologue n'est trouvée à la jonction suggérant que les mécanismes impliqués dans leur formation sont indépendants de la recombinaison homologue. Par contre, un mécanisme de type BIR (break-induced replication) pourrait permettre la formation de chromosomes chimériques en l'absence de longue séquence homologue.

INTRODUCTION

Les modifications génétiques conduisant à un changement phénotypique chez les cellules diploïdes sont souvent plus complexes que celles rencontrées chez les cellules haploïdes. Les études de réarrangements chromosomiques chez la levure ont été effectuées pour la plupart sur des cellules haploïdes lors de la mitose ou diploïdes après méïoses. En effet, pour des commodités d'expérimentation et d'interprétation, la détection de ces remaniements a été réalisée préférentiellement dans un tel contexte. Il existe ainsi de nombreux exemples de modifications génomiques spontanées chez la levure à l'état haploïde. En plus des duplications présentées dans ce travail, des événements de délétion entre des séquences répétées comme les éléments Ty ou les ADN ribosomiques ont été mis en évidence et caractérisés (Rothstein *et al.*, 1987 ; Christman *et al.*, 1988 ; Kupiec and Petes, 1988 ; Keil and Mc Williams, 1993 ; Klein, 1995). De même, des événements de translocation et d'inversion résultant de recombinaison entre des éléments Ty ont été observés (Camasses, 1996; Rachidi *et al.*, 1999). Plus récemment, des analyses de remaniements chromosomiques révèlent la possibilité à l'état haploïde de différents types de remaniements qui ne nécessitent pas la présence de longues séquences d'identité à la jonction (Chen *et al.*, 1998; Chen and Kolodner, 1999; Myung *et al.*, 2001). Des translocations et délétions médiés par des courtes séquences similaires ont été observées chez des cellules sauvages ou des révertants (Welcker *et al.*, 2000 ; Perez-Ortin *et al.*, 2002 ; Yu and Gabriel, 2004). Cette liste d'exemples non exhaustive indique que différents mécanismes de recombinaison sont impliqués dans la formation de réarrangements chromosomiques spontanés au sein du génome. Malgré l'ensemble de ces analyses, il est difficile d'avoir une vision globale car très peu d'études ont été réalisées à l'état diploïde. La recombinaison mitotique entre les loci alléliques est beaucoup plus fréquente que les réarrangements chromosomiques dans une cellule diploïde ce qui rend plus difficile la détection de tels remaniements (Hiraoka *et al.*, 2000). De plus, lorsque l'on utilise une cellule haploïde, seule une partie des événements peut être observée car ceux qui entraînent la perte d'un gène ou d'un segment d'ADN essentiel rendent la cellule non viable. En conséquence, la ploïdie induit forcément un biais dans la sélection des événements et doit être prise en compte. L'étude des événements de duplications trouvés au sein d'un génome diploïde devrait donc permettre d'obtenir des informations supplémentaires et surtout d'avoir une vision globale sur la formation de ces remaniements.

Dans le cadre de notre étude, deux événements distincts de duplication ont pu être sélectionnés : la duplication génique et la duplication segmentale à partir de la souche haploïde *ura2 15-30-72*. Dans un premier temps, il serait intéressant de voir si les événements rencontrés à l'état haploïde, le sont également à l'état diploïde. Concernant les duplications géniques, il serait, cependant, très étonnant d'en sélectionner à partir d'une cellule diploïde *a'*. En effet, dans une telle cellule, la transposition des Ty est fortement réprimée (Errede *et al.*, 1987 ; Xu and Boeke, 1991). Ainsi, le mécanisme de duplication génique médié par la transcription inverse de l'ARNm sera également fortement réprimé. Concernant les duplications segmentales, il est tout à fait envisageable d'en sélectionner à l'état diploïde. De plus, les segments dupliqués pourraient être insérés dans un gène essentiel sans que cela n'affecte la viabilité de la cellule.

Il est également probable que d'autres événements de duplication puissent être sélectionnés dans un tel contexte. En effet, plusieurs autres moyens, non isolés à l'état haploïde, peuvent potentiellement conduire à la duplication de la séquence codant le domaine ATCase: la duplication d'un bras de chromosome ou encore d'un chromosome entier. Bien que les événements d'aneuploïdie semblent très répandus chez *S. cerevisiae* (Hughes *et al.*, 2000), aucun événement de ce type n'a été observé avec notre crible de sélection à l'état haploïde. Les sélections à l'état diploïde permettront donc de savoir si l'absence de ce type d'événement est liée ou non à la ploïdie.

Le crible de sélection positive basé sur l'utilisation de l'allèle mutant du gène *URA2* a été adapté pour la sélection de remaniements chromosomiques à l'état diploïde et 14 révertants portant une duplication de la région codant l'ATCase ont été isolés (Tourrette, 2004).

Une étude moléculaire et génétique de ces souches a été entreprise en déterminant les mêmes caractéristiques que pour les révertants haploïdes, à savoir :

- la localisation chromosomique de la région dupliquée,
- la taille de la région dupliquée,
- et la détermination des sites d'insertion de l'élément dupliqué.

L'ensemble de ces données va permettre d'apporter des réponses quant aux événements sélectionnés et de savoir si les duplications présentes chez ces révertants diploïdes correspondent à des remaniements différents ou déjà sélectionnés à l'état haploïde.

RÉSULTATS

1. Caractérisation moléculaire et génétique des événements de duplication

1.1. Analyse du profil de restriction *Bam*HI au locus *URA2*

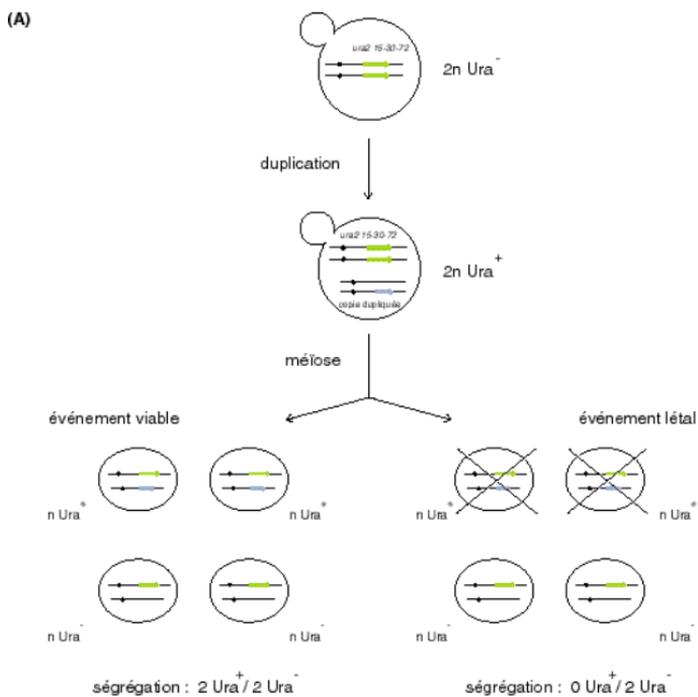
Une analyse du profil de restriction du locus *URA2* a été réalisée sur les 14 révertants sélectionnés. L'ADN génomique de ces souches a été digéré par l'enzyme de restriction *Bam*HI puis hybridé avec une sonde correspondant à la région codant le domaine ATCase. Pour la souche *ura2 15-30-72* ayant permis de sélectionner ces révertants, l'hybridation révèle une seule bande de 6,8 kb qui correspond au locus *URA2*. Par contre, l'hybridation de l'ADN génomique des 14 révertants révèle un profil de restriction *Bam*HI qui diffère de celui de la souche *ura2 15-30-72* par la présence d'une bande supplémentaire. Dans le cas des 14 révertants, le profil de restriction *Bam*HI révèle deux bandes : l'une de 6,8 kb correspondant au locus *URA2* et l'autre de taille variable correspondant à la partie dupliquée qui s'est insérée dans une région différente du génome qui se caractérise par un environnement *Bam*HI différent.

1.2. Analyse génétique

L'analyse des produits de la méiose a été réalisée pour les 14 souches diploïdes (Fig. 4.1A). Pour chaque révertant, la méiose a conduit à l'obtention de tétrades contenant toujours quatre spores. La capacité de germination des quatre spores sur milieu complet a été testée afin de déterminer si l'événement sélectionné est viable à l'état haploïde. L'analyse de la ségrégation des caractères Ura^+ et Ura^- permet de déterminer si les événements non viables observés sont liés ou indépendants de la réactivation du domaine ATCase.

Dans un premier temps, l'analyse de la viabilité des produits de la méiose des révertants diploïdes a été réalisée sur milieu complet. Deux types de ségrégation peuvent être observés. En effet, chez 7 révertants, les 4 spores sont capables de germer indiquant la sélection d'un événement viable en contexte haploïde (Fig. 4.1B). Par contre, chez les 7 autres, 2 spores seulement germent (Fig. 4.1B). Ceci indique la présence d'un événement létal à l'état haploïde.

Dans un second temps, la ségrégation Ura^+/Ura^- propre à chaque souche a été analysée. Pour les tétrades ayant 4 spores viables, la ségrégation observée est de type 2 spores Ura^+ et 2 spores Ura^- . Pour les tétrades ayant deux spores viables, ces deux spores sont de phénotype Ura^- c'est-à-dire que les spores Ura^+ porteur de l'événement



(B)

Ségrégation	Révertants
$2 \text{ Ura}^+ / 2 \text{ Ura}^-$	B038, B066/1, B199, B219, B284, B259 et B286
$0 \text{ Ura}^+ / 2 \text{ Ura}^-$	B025, B132, B177, B264, B267, B326 et B328

FIGURE 4.1 – Analyse génétique des révertants diploïdes porteurs d'une duplication.

de duplication ne sont pas viables à l'état haploïde. Par ailleurs, une analyse par hybridation de type Southern, comme décrit précédemment, met en évidence que les spores Ura^+ sont porteuses de la duplication alors que les spores Ura^- ne le sont pas. Ces données mettent en avant que la duplication de la région codant l'ATCase est bien à l'origine de la réversion phénotypique et que ces duplications sont viables chez une souche haploïde.

2. LOCALISATION CHROMOSOMIQUE DE LA RÉGION DUPLIQUÉE

Les chromosomes issus des révertants diploïdes ont été séparés par électrophorèse en champs alternés de façon à observer le caryotype de ces souches. Par la suite, ils ont été soumis à une hybridation ADN/ADN à l'aide d'une sonde spécifique de la région ATCase dans le but de déterminer la localisation chromosomique la région dupliquée (Fig. 4.2).

Dans le cas de 10 révertants, l'hybridation révèle le chromosome X porteur de l'allèle *ura2 15-30-72* ainsi qu'un autre chromosome porteur de la copie dupliquée indiquant une duplication interchromosomique (Tableau 4.1). Pour les 4 autres révertants (B259, B264, B282 et B286), l'hybridation du caryotype ne révèle que le chromosome X suggérant que l'événement correspond à une duplication intrachromosomique (Tableau 4.1).

De plus, pour les 7 souches pour lesquelles une ségrégation de type $2 Ura^+ / 2 Ura^-$ est observée, une étude du caryotype des spores a été entreprise. L'hybridation des chromosomes séparés en champs alternés avec une sonde spécifique de l'ATCase montre que contrairement aux spores Ura^- , les spores Ura^+ sont porteuses d'un événement de duplication qui est localisé sur le même chromosome que le révertant diploïde correspondant. Pour les 7 souches pour lesquelles une ségrégation $2 Ura^- / 0 Ura^+$ est observée, le caryotype des spores et le profil d'hybridation avec une sonde ATCase est similaire à celui de la souche diploïde homozygote $a/\alpha \text{ } ura2 \text{ } 15-30-72$. Ces données indiquent, une fois de plus, que la duplication de la région codant l'ATCase est bien à l'origine de la réversion phénotypique.

3. DÉTERMINATION DE LA TAILLE DE LA RÉGION DUPLIQUÉE

L'analyse des révertants isolés à partir de la souche haploïde *ura2 15-30-72* a permis de mettre en évidence deux types de duplications : la duplication génique et la

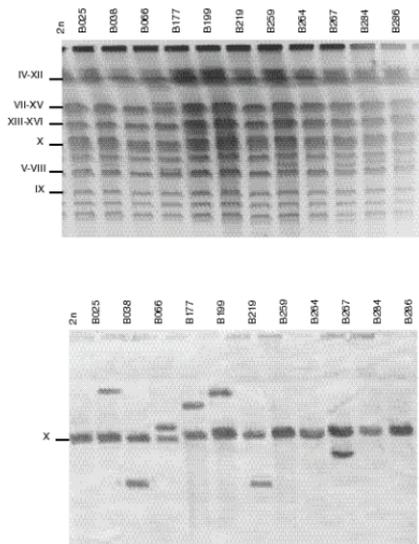


FIGURE 4.2 – Exemples de caryotypes après électrophorèse en champs alternés de quelques révertants sélectionnés à partir de la souche diploïde *ura2 15-30-72*. (A) Migration des chromosomes par PFGE et coloration au BET. (B) Révélation par hybridation ADN/ADN de type Southern avec une sonde spécifique de la région codant le domaine ATCase.

Localisation chromosomique		Souches
	VII - XV*	B025
	XII - XVI*	B177
Duplication interchromosomique	II - XIV*	B066, B132
	XI	B328
	V - VII*	B199, B267, B326
	IX	B038, B219
Duplication intra-chromosomique	X	B259, B264, B284, B286

TABLEAU 4.1 – Localisation chromosomique des régions dupliquées présentes chez les 8 révertants diploïdes.

* Les conditions électrophorétiques ne permettent pas la séparation optimale des chromosomes afin de les différencier.

duplication segmentale qui se distinguent, entre autres, par la taille de la région dupliquée. Afin de déterminer si ces deux types de remaniements peuvent également être sélectionnés à l'état diploïde, nous avons estimé la taille des régions dupliquées présentes chez les 14 révertants diploïdes en utilisant dans un premier temps la stratégie de marche sur le chromosome par hybridation de type Southern décrite dans le premier chapitre.

3.1. Événements de duplication interchromosomique

Pour les 10 révertants dans lesquels un événement de duplication interchromosomique s'est produit, la marche sur le chromosome a été réalisée par hybridation directe des chromosomes séparés par électrophorèse en champs alternés. Une série de 14 sondes correspondant aux gènes contigus au locus *URA2* a été choisie à partir de la séquence complète de la souche S288C de *Saccharomyces cerevisiae* (Fig. 4.3). Une marche chromosomique de 10 kb en 10 kb à partir du gène *URA2* jusqu'à atteindre un fragment d'une taille de 140 kb (voisin des régions télomériques) a ainsi pu être réalisée. L'hybridation du caryotype des 10 révertants avec les 14 sondes spécifiques montre un profil similaire à celui observé avec la sonde correspondant à la région ATCase. Ces sondes s'hybrident à la fois au chromosome X et au chromosome porteur de la copie dupliquée (Fig. 4.3). Ces données indiquent que l'ensemble gènes situés entre les ORFs YJL130c et YJL216c sur le bras gauche du chromosome X sont dupliqués. Les duplications correspondent donc à des régions d'une taille égale ou supérieure à 140 kb. Pour les régions subtélomériques se situant au-delà de l'ORF YJL216c, il a été impossible de trouver des séquences spécifiques pouvant être utilisées comme sonde car elles sont fortement redondantes. La marche sur le chromosome n'a donc pas pu être menée à terme pour ces 10 révertants.

3.2. Événements de duplication intrachromosomique

Pour les révertants B259, B264, B282 et B286, cette procédure n'a pas pu être utilisée car les copies résidente et dupliquée comigrent lors de la séparation des chromosomes par électrophorèse en champs alternés. Dans ce cas, la marche sur le chromosome a été réalisée sur l'ADN génomique par l'analyse du profil de restriction *Bam*HI (procédure détaillée chapitre I §1.3.2.2.2).

Les résultats pour la souche B286 montrent que seules les sondes YJL130c et YJL133w permettent d'obtenir un profil de restriction constitué de deux bandes.

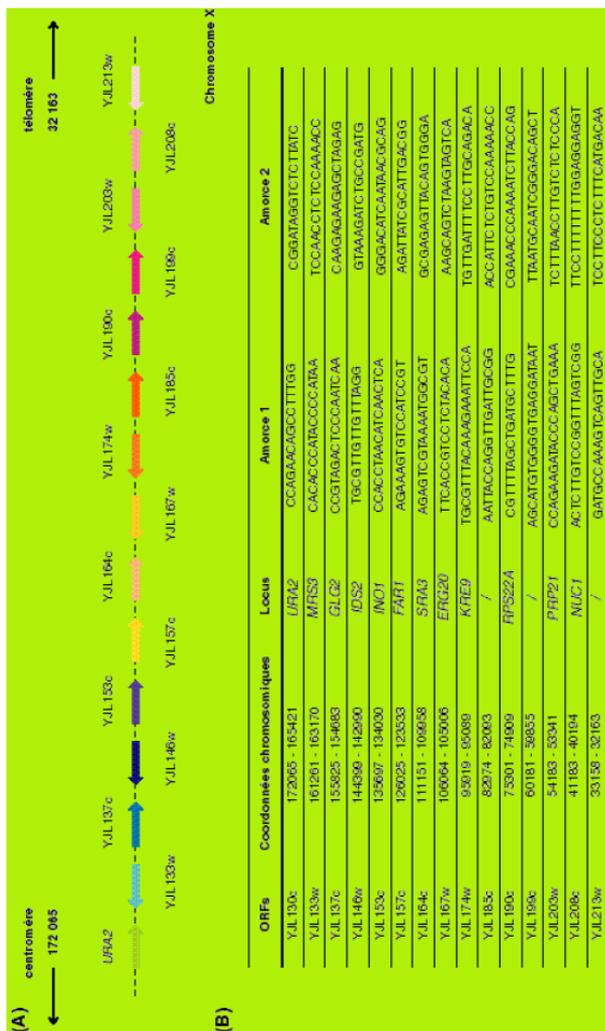


FIGURE 4.3 – Stratégie de marche sur le chromosome par hybridation de type Southern.
 (A) Carte chromosomique de la région située au nord du gène URA2.
 (B) Séquences et coordonnées des sondes utilisées pour marcher sur le chromosome porteur de la duplication.

Par contre la sonde correspondant à l'ORF YJL134c ne révèle qu'une seule bande. Ces résultats indiquent que le segment dupliqué englobe la région codant le domaine ATCase ainsi que les ORFs YJL130c à YJL133w. L'élément dupliqué présent chez le révertant B286 a donc une taille d'environ 6 kb.

Par contre, pour les 3 autres souches (B259, B264, B282), l'hybridation avec une sonde spécifique à l'ORF YJL134c montre par un profil à une bande que cette ORF est encore élément de la duplication. Par conséquent, ces événements correspondent à une duplication d'une taille supérieure à 6 kb.

Ainsi, la détermination de la taille des régions dupliquées présentes chez les révertants diploïdes permet d'apporter des éléments de réponse quant aux remaniements sélectionnés. En effet, d'une part, aucune duplication génique n'a été mise en évidence. D'autre part, la majorité des duplications correspondent à des régions de grande taille supérieure à 140 kb. Finalement, comme pour les souches haploïdes, la stratégie utilisée permet uniquement de distinguer les duplications géniques des duplications segmentales. Ainsi, la technique d'hybridation génomique comparative sur microarray sera nécessaire afin de déterminer précisément cette caractéristique physique.

4. Analyse de la séquence des sites d'insertion en amont de la région dupliquée

Pour les 10 révertants porteurs d'une duplication interchromosomique d'une taille égale ou supérieure à 140kb, deux hypothèses peuvent être envisagées. La première impliquerait la duplication d'une grande partie du bras gauche du chromosome X couvrant la région située entre le gène *URA2* et incluant le télomère gauche. La seconde serait que la duplication corresponde à un segment de 140kb englobant la région codant l'ATCase ainsi qu'une série de gènes contigus n'incluant pas le télomère.

Si l'on émet l'hypothèse que ces événements correspondent à une duplication de segment, la viabilité à l'état haploïde est dépendante du site d'insertion de l'élément dupliqué. Parmi ces souches diploïdes, deux types de ségrégation sont observés mettant en avant des événements de duplication viables mais également létaux au niveau des descendants haploïdes. Si la duplication s'insère dans un gène essentiel, cet événement va interrompre le gène ce qui n'aura pas de conséquence dans une souche diploïde mais entraînera la létalité à l'état haploïde. Par contre, si la duplication s'insère dans un gène non essentiel ou dans une région intergénique, l'événement sera viable dans les deux états : haploïde et diploïde. Afin de vérifier s'il y a une corrélation entre la viabilité et le

site d'insertion des remaniements sélectionnés, l'identification du point de jonction en amont des copies dupliquées a été réalisée par séquençage. Par ailleurs, la détermination de ces points de jonction peut également apporter des indices permettant de proposer des mécanismes conduisant à la duplication de séquences d'ADN dans une souche diploïde.

La jonction 5' de la région dupliquée a été clonée en utilisant la stratégie d'insertion et d'excision de plasmide (procédure détaillée dans le chapitre I). La première étape consiste à insérer un plasmide intégratif dans la région dupliquée en aval du point de départ de la duplication. Dans une deuxième étape, le plasmide recombinant portant la jonction inconnue est isolée puis caractérisée.

L'analyse des jonctions montre que pour l'ensemble des révertants, la partie dupliquée du gène *URA2* s'est insérée dans une ORF (Tableau 4.2). Les gènes receveurs sont dispersés dans le génome de *Saccharomyces cerevisiae*. De manière générale, l'insertion se fait préférentiellement dans un gène non essentiel. Cette observation n'est pas étonnante puisque seuls 17 % des gènes de la souche S288C de *S. cerevisiae* sont essentiels à la survie de la cellule sur milieu complet (Winzeler *et al.*, 1999).

Ségrégation	Révertants	ORF receveuse*	Locus	Localisation chromosomique par clonage	Taille de la duplication	Localisation chromosomique déterminée par PFGE
2 Ura ^r / 2 Ura ^s	B038	YNL064c (ne)	<i>YDJ1</i>	XIV (789kb)	≥140 kb	IX (450 kb)
	B066/1	YMR186w (ne)	<i>HSC2</i>	XIII (915kb)	≥140 kb	II – XIV (800 kb)
	B199	YLR109w (ne)	<i>AHP1</i>	XIV (785 kb)	≥140 kb	V – VIII (800 kb)
	B219	YNL079c (ne)	<i>TFM1</i>	XII (1900 kb)	≥140 kb	IX (450 kb)
	B259	YGL007w (ne)	/	VII (1120 kb)	≥ 6 kb	X (745 kb)
	B284	YLR249w (e)	<i>YEF3</i>	XII (1900 kb)	≥ 6 kb	X (745 kb)
	B286	YJL133c-A (/)	/	X (745 kb)	6 kb	X (745 kb)
	B025	YGL205w (ne)	<i>POX1</i>	VII (1100 kb)	≥140 kb	VII – XV (1100 kb)
0 Ura ^r / 2 Ura ^s	B132	YMR186w (ne)	<i>HSC2</i>	XIII (915kb)	≥140 kb	II – XIV (800 kb)
	B177	YNL299w (ne)	<i>TRF5</i>	XIV (785 kb)	≥140 kb	XIII – XVI (900 kb)
	B284	YEL076c (e)	<i>ILS1</i>	II (815 kb)	≥140 kb	XIII – XVI (900 kb)
	B267	YHR171w (ne)	<i>ATG7</i>	VIII (555 kb)	≥140 kb	V – VIII (800 kb)
	B326	YHR174w (e)	<i>ENC2</i>	VIII (555 kb)	≥140 kb	V – VIII (800 kb)
	B328	YMR119w-A (ne)	/	XIII (915 kb)	≥140 kb	XI (680 kb)

TABLEAU 4.2 – Caractéristiques des régions dupliquées présentes chez les 14 révertants diploïdes.

* Les gènes essentiels et non-essentiels sont indiqués entre parenthèses par respectivement (e) et (ne).

Pour les révertants présentant une ségrégation de type $2 \text{Ura}^- / 0 \text{Ura}^+$, l'insertion dans un gène essentiel permettrait d'expliquer la non-viabilité des spores Ura^+ porteuses de l'événement de duplication segmentale à l'état haploïde. Or, sur 7 souches ayant une ségrégation de ce type, seuls les révertants B264 et B326 présentent une duplication interrompant un gène essentiel respectivement *ILS1* et *ENO2* (Tableau 4.2). La létalité observée dans la descendance pour les 5 autres révertants ne peut être expliquée par un événement de duplication segmentale.

De même, on peut observer dans la souche B284 (présentant une ségrégation de type $2 \text{Ura}^+ / 2 \text{Ura}^-$) l'interruption du gène *YEF3* qui est essentiel à la survie cellulaire (Tableau 4.2). Or, si l'événement correspond à une duplication de segment, l'insertion de la région dupliquée doit se faire dans un gène non essentiel pour les révertants diploïdes présentant ce type de ségrégation.

L'ensemble de ces données indiquent très clairement qu'il n'y a pas de corrélation stricte entre la létalité à l'état haploïde et le site d'insertion. Par conséquent, les remaniements sélectionnés ne semblent pas correspondre à des événements de duplication segmentale.

Il est également intéressant à noter qu'il n'y a pas de corrélation entre la localisation chromosomique déterminée par hybridation des chromosomes séparés par électrophorèse en champs alternés et le site d'insertion chromosomique déterminé ci-dessus sauf pour la souche B286 (Tableau 4.2). Par exemple, pour la souche B038, la région dupliquée d'une taille supérieure à 140 kb s'est insérée dans le chromosome XIV ayant une taille de 785 kb. Suite à une duplication segmentale, le chromosome résultant devrait donc avoir une taille supérieure à 925 kb. Or le chromosome porteur de la séquence dupliquée a une taille équivalente au chromosome IX, c'est-à-dire 450 kb. Ces données montrent à nouveau que les remaniements présents chez les révertants diploïdes (à l'exception de la souche B286) ne peuvent pas se résumer à une duplication d'un segment d'ADN suivi de son insertion.

À ce jour, le site d'insertion en 5' a été déterminé au nucléotide près pour 5 révertants (Fig. 4.4A), il est en cours de détermination pour les autres. Dans les souches analysées, la jonction consiste en une fusion en phase entre la partie dupliquée du gène *URA2* et le gène receveur. Comme la fusion préserve la phase ouverte de lecture, un gène mosaïque composé de l'extrémité 5' du gène tronqué et de l'extrémité 3' dupliquée du gène *URA2* est formé. Nous avons recherché la présence d'identité de séquence entre

la région receveuse et la partie dupliquée du gène *URA2* à la jonction. Les jonctions présentent des courtes régions de similarité de séquence comprises entre 3 pb et 17 pb (Fig. 4.4B). Par ailleurs, comme pour les remaniements observés à l'état haploïde, l'analyse des jonctions montre que le point de départ des copies dupliquées est variant d'une souche à l'autre. Il est toujours localisé en aval des trois mutations de l'allèle *ura2 15-30-72*. Cependant, il n'est pas unique et peut être localisé soit à la fin de la région codant la CPSase ou dans la région codant le domaine cryptique DHOase-like.

En conclusion, l'analyse des sites d'insertion des régions dupliquées indique très clairement qu'il n'y a pas de corrélation stricte entre la viabilité et le site d'insertion. Ces données ne semblent pas être en faveur de l'hypothèse de la duplication segmentale mais plutôt en faveur de la duplication de bras de chromosome. Cependant, un tel événement doit forcément impliquer d'autres éléments chromosomiques comme un centromère et un autre télomère pour que cette région dupliquée puisse être préservée au sein du génome. Ainsi, une analyse globale du nombre de copies de gène par l'utilisation de la technique d'hybridation génomique comparative permettra de vérifier d'une part s'il s'agit bien d'une duplication de bras de chromosome et d'autre part de voir quelles sont les chromosomes impliqués dans ces remaniements génomiques.

5. Analyse globale du nombre de copies de gènes par hybridation génomique comparative par microarray

Dans une première étape, une analyse globale du nombre de copies de chaque gène a été entreprise chez 6 souches ayant une ségrégation de type $2 \text{ Ura}^+ / 2 \text{ Ura}^-$ et chez 2 souches (B025 et B132) ayant une ségrégation de type 2 Ura^- . Pour cela, nous avons utilisé la technique d'hybridation génomique comparative par microarray (array CGH) déjà utilisée pour la détermination de la taille des événements de duplication segmentale chez les révertants haploïdes (Winzeler *et al.*, 1999). Basée sur l'hybridation d'ADN génomique sur microarrays, cette technique permet de détecter le changement du nombre de copies de gènes (délétion ou duplication) d'une souche par rapport à une souche de référence. En effet, l'intensité de l'hybridation de chaque ORF est proportionnelle au nombre de copies dans une souche donnée. Dans notre étude, cette technique nous a permis de comparer le changement du contenu génomique entre les révertants et la souche diploïde ayant permis de les sélectionner (chapitre II).

L'ADN génomique de chacun des 8 révertants a été hybridé sur microarray et

comparé à la souche ancestrale diploïde homozygote *a/α ura2 15-30-72*. Le ratio de l'intensité du signal entre chaque révertant et la souche diploïde homozygote *a/α ura2 15-30-72* a été déterminé pour chaque ORF. Un graphique du ratio en fonction de la localisation chromosomique a été tracé pour chacun des révertants.

Chez les 8 révertants, les résultats montrent que tout le segment chromosomique localisé en aval du gène *URA2* (YJL130c) sur le chromosome X est dupliqué (Fig. 4.5). Les éléments dupliqués ont donc une taille d'environ 168 kb couvrant la région allant de l'ORF YJL130c jusqu'à la région télomérique gauche. Ces données montrent que les événements de duplication présents chez les 8 révertants analysés correspondent à une duplication du bras gauche du chromosome X à partir du gène *URA2*.

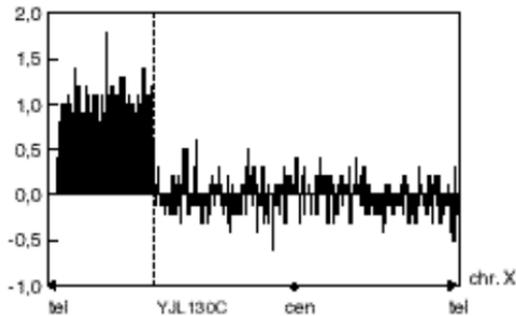


FIGURE 4.5 –Caractérisation des gènes dupliqués du chromosome X chez les révertants diploïdes. Représentation graphique des résultats de l'hybridation génomique comparative sur microarray pour l'ensemble du chromosome X de la souche B038. Les révertants B066, B199, B219, B259, B284, B025 et B132 présentent exactement le même profil.

De plus, dans les 6 souches ayant une ségrégation de type $2 \text{Ura}^+ / 2 \text{Ura}^-$, l'hybridation génomique comparative a permis de mettre en évidence la duplication d'une région localisée sur un autre chromosome que le chromosome X. Cette région comprend un bras chromosomique complet, le centromère ainsi qu'une partie incomplète de l'autre bras (Fig. 4.6). En fonction du révertant, la duplication implique les chromosomes VII, XII, XIII ou XIV. La caractéristique génomique de ces 6 souches est donc d'avoir d'une part la duplication du bras gauche du chromosome X (avec *URA2* comme origine) et d'autre part la duplication d'une partie d'un autre chromosome (comprenant un bras entier et un centromère). L'explication la plus simple, rendant compte de ces résultats, est la formation d'un nouveau chromosome chimérique qui réunit ces deux segments chromosomiques dans chacune de ces souches. Les sites d'insertion déterminés précédemment permettent de vérifier cette hypothèse. En effet, les points de jonction correspondent dans toutes les souches à la réunion des deux régions dupliquées. La taille de ces chromosomes chimères obtenue en additionnant les deux régions dupliquées est égale à celle du chromosome porteur de la région dupliquée révélée lors de l'hybridation de caryotype électrophorétique (Tableau 4.3).

Ségrégation	Révertants	Taille de bras du chromosome X dupliqué (kb)	Taille de la partie chromosomique receveuse de la région dupliquée du chromosome X (kb)	Taille déterminée par PFGE (kb)
$2 \text{Ura}^+ / 2 \text{Ura}^-$	B038	~168	277	450
	B0661	~168	633	800
	B199	~168	669	800
	B219	~168	306	450
	B259	~168	637	745
	B284	~168	636	745
$0 \text{Ura}^+ / 2 \text{Ura}^-$	B025	~168	1015	1100
	B132	~168	634	800
	B177	~168	718	900
	B264	~168	731	900
	B267	~168	447	600
	B326	~168	637	800
B328	~168	636	680	

TABLEAU 4.3 – Estimation de la taille des chromosomes porteurs de la séquence *ATCase* dupliquée.

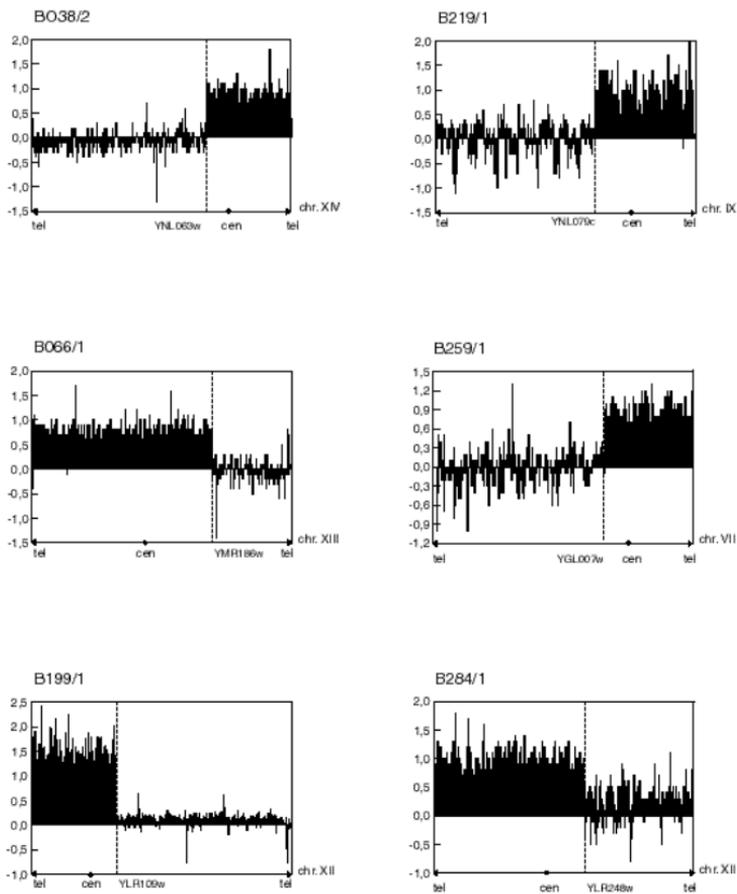


FIGURE 4.6 – Duplication d'une région localisée sur un autre chromosome que le chromosome X. Représentation graphique des résultats de l'hybridation génomique comparative sur microarray.

L'ensemble de ces données mettent en évidence que ces 6 souches sont des aneuploïdes possédant un chromosome chimère supplémentaire (Fig. 4.8A).

Dans les deux souches (B025 et B132) présentant une ségrégation de type 2 Ura^- , l'hybridation génomique comparative met également un autre remaniement en évidence : la délétion d'un bras d'un autre chromosome, respectivement du chromosome VII et XII (Fig. 4.7). La caractéristique génomique de ces 2 souches est donc d'avoir d'une part la duplication du bras gauche du chromosome X (avec *URA2* comme origine) et d'autre part la délétion d'un bras d'un autre chromosome. Là encore, l'explication la plus simple serait la formation d'un chromosome chimère par une translocation non-réciproque. Les points de jonction permettent à nouveau de vérifier cette hypothèse car ils correspondent à la réunion du bras gauche du chromosome X et du chromosome ayant perdu une partie d'un de ces bras. Nous avons également estimé la taille des chromosomes chimères formés en additionnant la région dupliquée à celle du chromosome ayant été délété (Tableau 4.3). Dans les 2 cas (B025 et B132), la taille ainsi obtenue est égale à celle du chromosome porteur de la copie dupliquée révélée lors de l'hybridation de caryotype électrophorétique (Tableau 4.3). L'ensemble de ces données mettent en évidence que les 2 souches analysées présentent un chromosome chimère issu d'un événement de translocation non-réciproque (Fig. 4.8B).

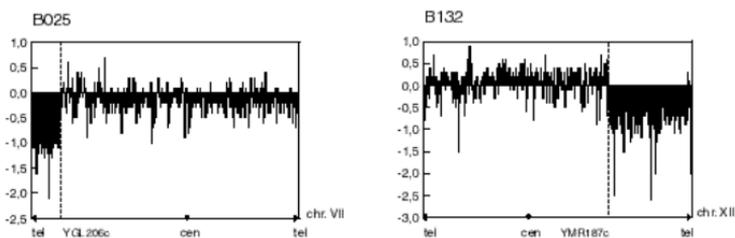


FIGURE 4.7 – Délétions présentes chez les révertants B025 et B132. Représentation graphique des résultats de l'hybridation génomique comparative sur microarray.

5 souches (B177, B264, B267, B326 et B328) restent à être analysées par microarray CGH. Ces révertants présentent une ségrégation de type 2 Ura^- et semblent être le siège d'un événement de translocation non-réciproque comme B025 et B132. En effet, en se basant sur les points de jonction, nous avons estimé la taille des chromosomes chimères formés en additionnant la région dupliquée au chromosome dont une partie serait déléetée. Dans tous les cas, la taille estimée est en corrélation avec la taille du chromosome porteur de la copie dupliquée du domaine codant l'ATCase déterminée précédemment (Tableau 4.3). Une analyse globale des copies de gènes par array CGH permettra de confirmer cette hypothèse.

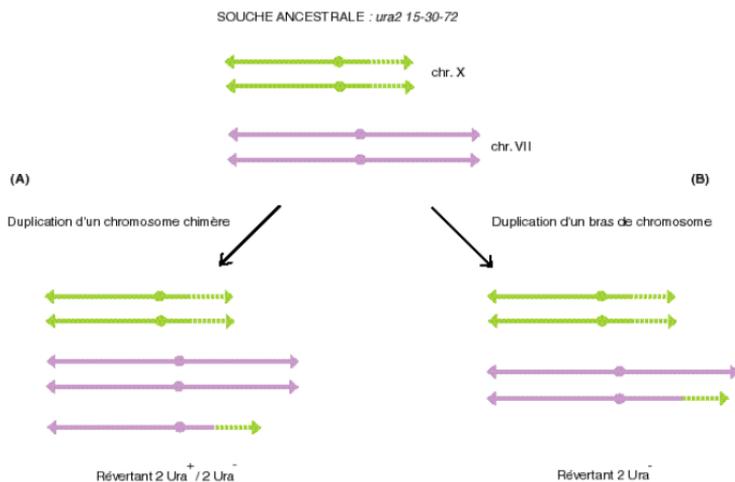


FIGURE 4.8 – Remeiements présents chez les révertants diploïdes.
 (A) Duplication d'un chromosome chimère.
 (B) Translocation non-réciproque.

6. Détermination du site d'insertion en aval du segment dupliqué pour le révertant B286

Concernant le révertant B286, la détermination de la taille montre que l'événement correspond à une duplication d'un segment d'environ 6 kb. Ce révertant n'a pas été analysé de la même manière que les autres souches portant un chromosome chimère. Nous nous sommes donc intéressés à la détermination de la jonction située en aval du segment dupliqué. Connaissant l'extrémité 3' de la région dupliquée et le site d'insertion chromosomique en 5', nous avons choisi des oligonucléotides pour amplifier et séquencer la jonction 3' du segment dupliqué (Fig. 4.9). Le segment s'étend de la partie dupliquée du gène *URA2* (YJL130c) jusqu'à l'ORF YJL133w. De plus, ce segment est inséré dans la région intergénique située dans l'ORF YJL133c-A (Fig. 4.9). Ainsi, deux oligonucléotides ont été choisis: l'un spécifique à l'ORF YJL133w (correspondant à la fin du segment) et l'autre spécifique à l'ORF YJL134w (l'ORF contigu au site d'insertion en 5' dans la souche *ura2 15-30-72*) afin d'amplifier la jonction (Fig. 4.9). Ces oligonucléotides ont effectivement permis d'amplifier un fragment qui a été séquencé. La comparaison de séquences montre que la jonction située en 3' correspond à la région sauvage comprise entre les ORFs YJL133w et YJL133c-A chez la souche ancestrale *ura2 15-30-72*. L'ensemble de ces données permettent de conclure que l'événement correspond à une duplication en tandem direct similaire à celles isolées à partir de la souche haploïde *ura2 15-30-72* (chapitre III).

B286

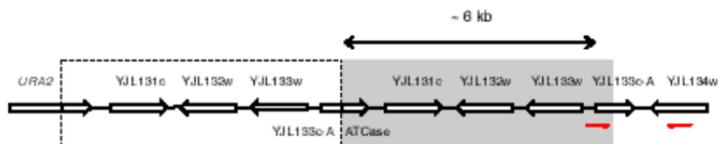


FIGURE 4.9 – Duplication segmentale en tandem direct chez le révertant B286. Les flèches rouges correspondent aux oligonucléotides ayant permis d'amplifier la jonction 3'.

DISCUSSION

Dans ce chapitre, nous nous sommes focalisés sur l'étude des événements de duplication sélectionnés à partir la souche diploïde homozygote a/α *ura2 15-30-72* (Tourrette, 2004) dans le but d'avoir une vision globale sur la formation des duplications au sein du génome de *Saccharomyces cerevisiae*. Une telle étude se justifie car la ploïdie induit forcément un biais dans la sélection des événements et doit être prise en compte. Par exemple, lorsque l'on utilise une cellule haploïde, les événements entraînant la perte d'un gène ou d'un segment d'ADN essentiel rendent la cellule non viable et ne peuvent donc pas être sélectionnés.

Au niveau haploïde, nous avons mis en évidence deux types de duplication bien distincts : la duplication génique et la duplication segmentale. Parmi la collection des 14 révertants diploïdes analysés, seul un révertant (souche B286) présente un événement de duplication segmentale similaire à ceux observés à l'état haploïde. L'analyse des 13 autres révertants a permis de mettre en évidence deux autres moyens par lesquels une séquence d'ADN peut se dupliquer : la duplication de bras de chromosome et la duplication de chromosome chimère entier. Ces données apportent une preuve expérimentale quant à l'importance de la ploïdie dans l'émergence de réarrangements chromosomiques.

L'analyse du révertant B286 montre que l'événement de duplication correspond à une duplication segmentale intrachromosomique en tandem direct. Le segment dupliqué a une taille d'environ 5 kb couvrant une région située entre les ORFs YJL130c et YJL133w. Nous avons montré que l'élément dupliqué s'est inséré au niveau de l'ORF YJL133c-A. La séquence ATCase est fusionnée en phase avec la partie 5' de ce gène non essentiel et doit ainsi bénéficier de son promoteur. De plus, seule une courte identité de séquence (7pb) a été mise en évidence à la jonction. Cette duplication est similaire en tout point à celles trouvées chez les souches haploïdes analysées. Sa formation est donc indépendante de la ploïdie et peut s'expliquer par le modèle proposé dans le chapitre III impliquant un mécanisme de type BIR (break-induced replication).

Par contre, aucune duplication génique médiée par un mécanisme de rétroposition n'a été mise en évidence dans les 14 révertants analysés. Ce résultat n'est cependant pas surprenant. En effet, les cellules diploïdes a/α produisent dix fois moins d'ARNm Ty1 que les cellules haploïdes en raison de la fixation du répresseur hétérodimérique $a1/2$, spécifique des cellules diploïdes, sur un site de liaison situé en

aval du démarrage de la transcription dans la séquence Gag (Errede *et al.*, 1987 ; Company and Errede, 1987). Cette répression entraîne une diminution importante de la synthèse de reverse transcriptase, inhibant ainsi fortement la transposition (Xu and Boeke, 1991). Ainsi, dans une cellule diploïde, un mécanisme de duplication génique médié par la transcription reverse de l'ARNm est très peu envisageable.

L'analyse moléculaire des révertants diploïdes a permis de mettre en évidence d'autres moyens permettant la duplication de séquences d'ADN : l'aneuploïdie et la translocation non-réciproque. Dans les deux cas de figure, il y a formation d'un chromosome chimère qui permet la réactivation fonctionnelle du domaine ATCase.

Pour les deux souches B025 et B132, l'analyse moléculaire montre que ces souches ont été le siège d'une translocation non-réciproque. En effet, les résultats de l'hybridation génomique comparative montrent d'une part la duplication du bras gauche du chromosome X à partir du gène *URA2* et d'autre part la délétion d'un bras d'un autre chromosome, respectivement du chromosome VII et XII. La détermination des points de jonction permet de vérifier que le bras gauche du chromosome X est réuni au chromosome ayant perdu une partie d'un de ces bras. Par ailleurs, la ségrégation observée pour ces deux souches est de type 2 spores viables et 2 spores non-viables. Les deux spores viables ont un phénotype Ura^- c'est-à-dire que les spores Ura^+ porteuses de l'événement de duplication ne sont pas viables à l'état haploïde. La ségrégation observée est en accord avec le remaniement qui s'est produit. En effet, les spores Ura^+ sont porteuses de la copie fonctionnelle c'est-à-dire du chromosome chimère et ont par conséquent perdu une partie de l'information génétique (Fig. 4.8). La partie délétée contenant des gènes essentiels, les spores Ura^+ ne sont pas viables. Dans ce cas, il est évident que ce type de remaniement n'aurait pas pu être sélectionné à partir d'une souche haploïde. Pour les 5 autres révertants présentant la même ségrégation, une analyse d'hybridation génomique comparative permettra vraisemblablement d'arriver aux mêmes conclusions.

Pour les 6 autres souches ayant une ségrégation de type 2 Ura^+ / 2 Ura^- , l'analyse moléculaire montre que le génome contient un chromosome chimère sur-numéraire. L'hybridation génomique comparative a permis de mettre en évidence en plus de la duplication du bras gauche du chromosome X, la duplication d'une région d'un autre chromosome comprenant un bras chromosomique complet et une structure centromérique. Dans ce cas de figure, la réunion permet également l'apport d'un promoteur à la séquence dupliquée codant le domaine ATCase. Là encore, la ségrégation

observée reflète le remaniement sélectionné. En effet, les spores Ura^+ possèdent le chromosome chimère surnuméraire, et contrairement à l'événement de translocation, il n'y a pas de perte de matériel génétique et par conséquent, elles sont viables (Fig. 4.8).

Finalement, dans les deux cas : translocation ou aneuploïdie, la réactivation est engendrée par la présence d'un chromosome chimère. La réunion du bras gauche du chromosome X à un autre chromosome permet l'apport d'un promoteur à la région codant le domaine ATCase. Dans ces révertants, la jonction consiste en une fusion en phase entre la partie dupliquée du gène *URA2* et le gène receveur (essentiel ou non). Comme la fusion préserve la phase ouverte de lecture, un gène mosaïque composé de l'extrémité 5' du gène tronqué et de l'extrémité 3' dupliquée du gène *URA2* est formé. Ces fusions constituent un exemple d'émergence de nouvelles fonctions, comme nous l'avons mentionné pour les duplication de segments. Cependant, ce type de réarrangements est également impliqué dans la progression tumorale chez l'homme comme par exemple dans la formation du chromosome Philadelphie impliqué dans les leucémies myéloïdes (Stam *et al.*, 1985). Ce chromosome chimère résulte d'une translocation réciproque entre un chromosome 22 et un chromosome 9. Ce réarrangement conduit à une fusion entre les gènes *ABL* et *BCR*. Contrairement à la protéine normale, cette protéine de fusion a une forte activité tyrosine-kinase. Il en résulte une division cellulaire incontrôlée dans les cellules affectées et un développement tumoral.

Si on regarde de plus près les deux événements sélectionnés : chromosome chimérique surnuméraire et translocation non-réciproque, les mécanismes qui doivent être à l'origine de ces remaniements sont identiques. En effet, le point de départ pour ces deux types de remaniements est la formation d'un chromosome chimère par translocation non-réciproque. Si au cours de la mitose, la rétention du chromosome chimère en plus du jeu normal de chromosomes se fait, on aboutit à une aneuploïdie. La cause de rétention est certainement la non-disjonction au cours de la mitose comme c'est le cas dans la majorité des cas d'aneuploïdie. Cette non-disjonction correspond à un défaut du processus de ségrégation des chromosomes.

Ainsi, la question est de savoir comment ces chromosomes chimères ont pu se former. Dans les deux types de remaniements sélectionnés à partir d'une souche diploïde homozygote *a/α ura2 15-30-72*, aucune longue séquence homologue n'est trouvée à la jonction suggérant que les mécanismes impliqués dans leur formation sont indépendants de la recombinaison homologue.

Plusieurs hypothèses peuvent être formulées pour expliquer la formation de chromosomes chimères. Dans le cadre des translocations réciproques, deux cassures double-brin sont nécessaires selon les modèles décrits (revue de Pâques & Haber, 2000) pour la formation de chromosomes chimères. La réparation par échange de bras de chromosomes peut se faire soit par recombinaison homologue (Potier *et al.*, 1982; Fasullo and Davis, 1988), soit par recombinaison non-homologue (Yu and Gabriel, 2004). Ces événements de translocations réciproques entraînent la présence de deux chromosomes chimères au sein du génome. Cependant, un tel mécanisme n'est pas envisageable pour les révertants sélectionnés à partir du système *URA2* car leur génome ne possède qu'un seul chromosome chimère.

Un mécanisme de type BIR « break induced replication » peut expliquer la formation d'un seul chromosome chimère au sein d'un génome c'est-à-dire qu'il peut être à l'origine d'un événement de translocation non-réciproque (Malkova *et al.*, 1996 ; Morrow *et al.*, 1997; Ira and Haber, 2002; Davis and Symington, 2004). Ce processus permet la réparation d'une cassure double brin via les mécanismes impliqués dans la réplication (Fig. 4.10). La fourche de réplication du chromosome ayant subi une cassure va envahir un autre chromosome afin de reformer une structure chromosomique stable. Il va se servir du chromosome envahi comme matrice depuis la cassure jusqu'au télomère du chromosome matrice. De cette façon, le chromosome ayant subi la cassure double-brin perd le segment allant de la cassure jusqu'à son extrémité et reçoit en échange le segment de chromosome qu'il a copié sur le chromosome matrice qui dès lors est dupliqué. Ce mécanisme peut ainsi conduire à une duplication de bras de chromosome couplé à la perte d'un bras du chromosome receveur du bras dupliqué.

Un mécanisme de type BIR (break-induced replication) pourrait donc permettre la formation d'un seul chromosome chimérique en l'absence de longue séquence homologue. Les jonctions présentent de courtes régions de similarité de séquence (7-15 pb). Une sélection de révertants dans un contexte génétique affectant ce mécanisme serait nécessaire pour confirmer cette hypothèse comme par exemple *rad27*.

En conclusion, la sélection à partir de la souche diploïde homozygote *a/α ura2 15-30-72* permet effectivement d'avoir une vision globale sur les événements de duplication et de compléter les résultats obtenus à l'état haploïde. Pour les révertants haploïdes, la duplication génique et la duplication segmentale sont les médiateurs de la duplication de séquences d'ADN. Par contre, pour les révertants diploïdes, ce rôle est assuré par la formation d'un chromosome chimère conduisant à l'aneuploïdie ou à la translocation non-réciproque.

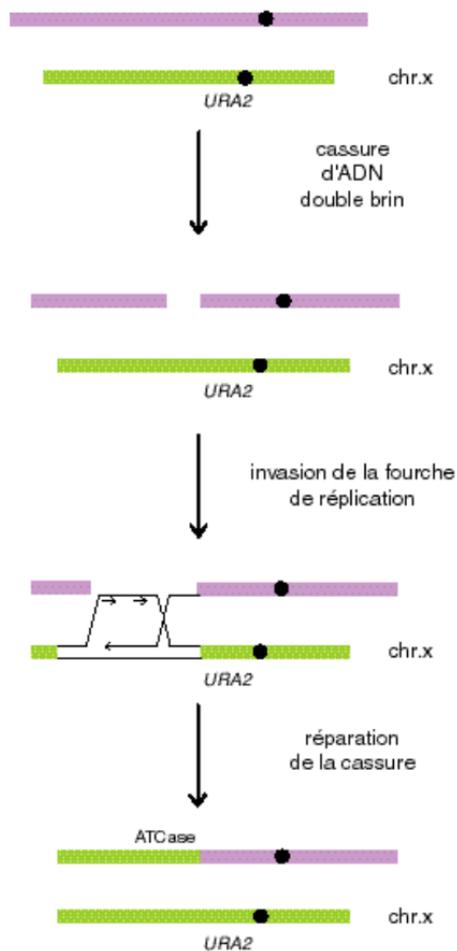


FIGURE 4.10 – Modèle de réparation d'une cassure double brin par un mécanisme de type BIR (Break Induced Replication).

5

LES PARALOGUES *URA5* ET *URA10* : ÉVOLUTION DE DEUX GÈNES ISSUS DE LA DUPLICATION COMPLÈTE DU GÉNOME DE *SACCHAROMYCES CEREVISIAE*

RÉSUMÉ

*L'accumulation de séquences partielles ou complètes de génomes a permis d'ouvrir de nouveaux horizons. En effet, le séquençage n'est pas un but final, mais permet d'initier de nombreuses études et notamment de génomique comparative. Suite au séquençage de nombreuses levures hémiascomycètes, nous nous sommes intéressés à l'évolution des gènes paralogues *URA5* et *URA10*. Ces gènes appartiennent au bloc ancestral 36 de duplication défini par Wolfe et Shield (1997) et sont donc issus de la duplication totale du génome. L'ensemble des séquences disponibles nous a donné l'opportunité de suivre l'évolution de ces copies de gènes au sein de génomes couvrant l'ensemble de l'arbre phylogénétique des hémiascomycètes. Les résultats ont permis de mettre en évidence que ces gènes paralogues présents chez *Saccharomyces cerevisiae* ont pu évoluer différemment après la duplication au sein des autres espèces hémiascomycètes. Le gène *URA5* semble correspondre à la copie garante de la fonction originelle alors que le gène *URA10* correspond à la copie ayant évolué sans pression de sélection. De plus, la présence des deux gènes *URA5* et *URA10* peut servir de marqueur de la polyploïdisation et permettre de dater cet événement.*

INTRODUCTION

La comparaison est une stratégie très utilisée depuis bien longtemps pour l'étude et la compréhension du vivant. L'anatomie comparée a permis depuis l'antiquité l'élaboration de classifications du monde vivant conduisant par la suite aux premières théories de l'évolution. À l'heure actuelle, l'accumulation de séquences complètes de génomes permet d'utiliser cette stratégie de comparaison au niveau moléculaire. On parle alors de génomique comparative. Elle repose sur le fait que les génomes d'espèces phylogénétiquement proches présentent une similarité de séquence importante. Deux organismes avec un ancêtre commun proche phylogénétiquement vont avoir des génomes qui vont être construits sur des fondations communes avec différentes spécificités pour chaque espèce. Plus les organismes seront proches sur l'échelle évolutive, plus leurs génomes seront architecturalement proches (Nadeau and Sankoff, 1998). Après un événement de spéciation, les génomes des nouvelles espèces vont accumuler des réarrangements au cours du temps. La colinéarité entre les chromosomes homologues des différentes espèces va être rompue par des événements de translocation, de duplication, de délétion ou d'inversion (Nadeau and Taylor, 1984). Ainsi, la comparaison des génomes de deux espèces très distantes sur l'échelle évolutive va être très difficile, voire impossible, car l'accumulation de réarrangements va conduire à la disparition de toute base commune c'est-à-dire à la perte de toute synténie entre ces deux espèces. Pour que la comparaison de l'architecture génomique puisse dévoiler les mécanismes de dynamique chromosomique et à plus long terme de spéciation, il faut pouvoir comparer des espèces relativement proches pour lesquelles une conservation au moins partielle de synténie peut être observée.

Cependant, après le séquençage de *S. cerevisiae*, les programmes se sont tournés vers le séquençage d'autres organismes modèles : *C. elegans*, *D. melanogaster*, *A. thaliana*, *H. sapiens* appartenant à des branches évolutives assez éloignées. La comparaison intergénomique sur le plan de l'architecture des génomes s'avère très difficile. Ces dernières années, le séquençage d'organismes plus proches, notamment dans le groupe des levures hémiascomycètes, a permis d'initier la comparaison de génomes phylogénétiquement plus proches. Les analyses comparatives concernant le groupe des hémiascomycètes initiées par le programme Génolevures sont dans ce domaine exemplaires (Souciet *et al.*, 2000 ; Kellis *et al.*, 2003 ; Cliften *et al.*, 2003 ; Dujon *et al.*, 2004).

Les génomes des levures hémiascomycètes présentent, comme celui de *S. cerevisiae*, une compacité remarquable, un faible nombre d'éléments répétés non codants et un faible nombre d'introns. Ces caractéristiques facilitent l'identification des ORFs dans ces levures. Bien que ces génomes diffèrent par leur taille avec ceux des eucaryotes supérieurs, tous ces organismes partagent un grand nombre de structures et de fonctions biologiques essentielles comme le cycle cellulaire, le cytosquelette, un grand nombre de voies métaboliques... De plus, la comparaison de l'architecture génomique des hémiascomycètes permet de suivre le processus évolutif au sein de ce groupe.

Le programme Génolevures-I fut le premier projet d'exploration génomique de 13 levures (Fig. 5.1) couvrant l'arbre évolutif des hémiascomycètes, l'espèce la plus éloignée phylogénétiquement de *Saccharomyces cerevisiae* étant *Yarrowia lipolytica* (Souciet *et al.*, 2000). Cette exploration génomique avait pour but d'obtenir des signatures d'un nombre de gènes de chaque espèce étudiée et, ceci, par un séquençage aléatoire avec une couverture 0,2× ou 0,4× équivalent génome, par la stratégie de RST (Random Sequence Tag) (Ozier-Kalogeropoulos *et al.*, 1998). Ces signatures ont permis de mettre en évidence :

- les principales classes fonctionnelles de gènes présents dans chaque espèce ;
- la conservation de synténie avec *S. cerevisiae* ;
- la proportion de duplication au sein de chaque espèce ;
- et finalement les gènes spécifiques des ascomycètes.

Plus récemment, le séquençage avec une faible couverture (3× à 7× équivalent génome) d'espèces (Fig. 5.1) très proches de *S. cerevisiae* a été réalisé (Kellis *et al.*, 2003; Cliften *et al.*, 2004). Parmi les levures séquencées, on en trouve 4 faisant parties du groupe des *Saccharomyces sensu stricto* (*S. paradoxus*, *S. mikatae*, *S. kudrivzenii* et *S. bayanus*). La majorité des gènes de *S. cerevisiae* a un orthologue dans chaque espèce et les blocs synténiques couvrent plus de 90 % de leur génome. La comparaison de ces espèces a également permis de mettre en évidence les séquences régulatrices très conservées. Par ailleurs, la comparaison de l'architecture des génomes montre l'ensemble des réarrangements comme les translocations, les inversions, les insertions et les duplications qui sont responsables de la dynamique chromosomique au niveau de ces espèces.

En parallèle, d'autres levures plus distantes de *Saccharomyces cerevisiae* ont été séquencées : 4 complètement (Dujon *et al.*, 2004) et 2 avec une faible couverture (Cliften *et al.*, 2004 ; Kellis *et al.*, 2004). L'ensemble de ces données va permettre de compléter avantageusement les analyses de génomique comparative déjà initiées pour ce groupe d'espèces phylogénétiquement proches. Ils vont également pouvoir être le point de départ de nombreuses expériences en laboratoire. Actuellement, les expérimentations génomiques génèrent beaucoup de données sur les organismes vivants par des approches systématiques telles que l'analyse de transcriptome, protéome... qui peuvent être réalisées à grande échelle, donc à haut débit, et qui ne sont pas forcément dépendantes d'une hypothèse de travail (Brent, 2000). La collection de données issues de ces expérimentations n'est pas le but ultime mais une étape nécessaire à la connaissance du vivant.

Comme nous l'avons déjà mentionné (introduction § 6.1.4), l'analyse de la séquence complète du génome de *Saccharomyces cerevisiae* a révélé la présence de blocs de duplication qualifiés de « fossiles » c'est-à-dire correspondant aux traces d'un événement de duplication ancestrale. Wolfe *et al.* (1997) ont émis l'hypothèse qu'il se serait produit une duplication totale du génome, ancienne, suivie d'événements de translocations et de délétions de certains fragments chromosomiques. Les données du séquençage récent des levures *Kluyveromyces waltii* et *Ashbya gossypii* ont permis de confirmer sans ambiguïté que les blocs de duplication sont issus d'une duplication complète de génome (Kellis *et al.*, 2004 ; Dietrich *et al.*, 2004). Après l'événement de duplication, le génome aurait subi une centaine de translocations réciproques (Seoighe, 1998). De plus, près de 90 % des gènes dupliqués auraient été perdus après l'événement de polyploïdie. Seuls 500 homologues auraient été conservés dans le génome actuel de *S. cerevisiae*. Ces 500 gènes dupliqués survivants sont répartis en 2 groupes. Le premier est constitué par les gènes paralogues dont les séquences n'ont pas divergé. Ils ont été retenus pour leur avantage sélectif lié au dosage génique. Ce groupe comporte, par exemple, les gènes codant les protéines ribosomiques ainsi que les facteurs d'élongation. Le second groupe, plus important quantitativement et plus intéressant, comporte les gènes qui ont divergé en séquence. Ces paralogues ayant divergé asymétriquement peuvent potentiellement évoluer de différentes manières : pseudogénéisation, subfonctionalisation ou néofonctionalisation (introduction § 5).

Dans ce chapitre, nous nous sommes intéressés à l'évolution d'une paire de gènes paralogues: *URA5* (YML106w) et *URA10* (YMR271c) qui font partis du groupe des 500 gènes conservés après l'événement de duplication totale. L'ensemble des séquences disponibles (Fig. 5.1) permet de suivre l'évolution de ces copies de gènes au sein de génomes couvrant l'ensemble de l'arbre phylogénétique des hémiascomycètes. Cette étude est intéressante à plusieurs titres. D'une part, ces gènes paralogues présents chez *Saccharomyces cerevisiae* pourraient permettre de suivre l'évolution différentielle de séquences d'ADN dupliquées au sein du groupe des levures hémiascomycètes. D'autre part, la présence des deux gènes paralogues comme *URA5* et *URA10* pourrait servir de marqueur de la polyploïdisation et permettre de dater cet événement.

1. LES GÈNES *URA5* (YML106w) ET *URA10* (YMR271c) DE *SACCHAROMYCES CEREVISIAE*

1.1. Fonction et régulation

Chez *Saccharomyces cerevisiae*, le gène *URA5* intervient dans la voie de biosynthèse des pyrimidines (Fig. 5.2). Il code l'OMP pyrophosphorylase (OMPpase) qui catalyse la transformation de l'acide orotique en OMP. Une mutation dans ce gène conduit à un phénotype bradytrophe pour l'uracile caractérisé par une croissance lente sur milieu minimum. Ce résultat a permis de conclure qu'il existait une autre enzyme chez *S. cerevisiae* qui portait une activité OMPpase (Jund and Lacroute, 1972). Le clonage de l'ADN codant pour l'OMPpase, par complémentation d'auxotrophie chez la levure, a permis de mettre en évidence un second gène *URA10* qui code une isoenzyme OMPpase (de Montigny *et al.*, 1990). L'analyse des séquences montre qu'il existe 75 % de similitude entre les protéines codées par les gènes *URA5* et *URA10*. Un mutant de délétion *ura10* n'a pas de phénotype particulier mais présente une activité OMPpase plus faible qu'une souche sauvage. L'enzyme codée par le gène *URA10* intervient pour 20 % dans l'activité OMPpase totale. Par contre, le double mutant *ura5 ura10* est auxotrophe à l'uracile.

Les études de la transcription de ces deux gènes montrent que *URA5* n'est pas régulé par un des produits de la chaîne de biosynthèse. Par contre *URA10* semble être régulé par le complexe formé par la protéine PPR1 et l'acide dihydroorotique qui induit également l'expression des gènes *URA4*, *URA1* et *URA3* (Fig. 5.2).

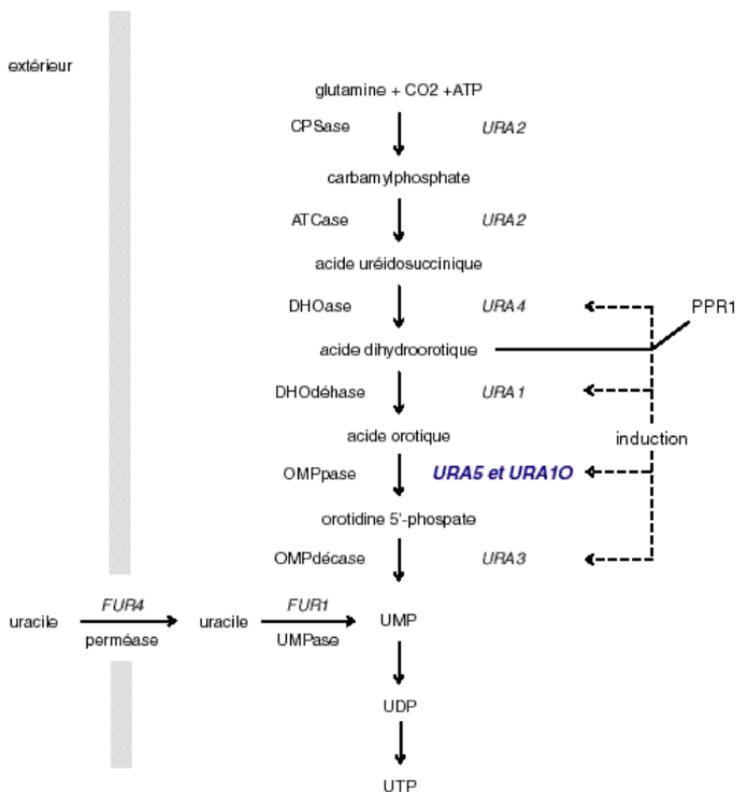


FIGURE 5.2 – Voie de biosynthèse de novo des pyrimidines.

Les gènes paralogues *URA5* et *URA10* sont un exemple de gènes dupliqués mis en évidence par de Montigny *et al.* (1990) avant le séquençage complet de *S. cerevisiae* (Goffeau *et al.*, 1996). De plus, ces deux gènes présentant des différences de régulation de leur expression constituent un exemple de subfonctionnalisation.

1.2. *URA5* et *URA10* sont éléments du bloc de duplication 36

Les gènes *URA5* (YML106w) et *URA10* (YMR271c) sont situés respectivement sur les bras gauche et droit du chromosome XIII et plus particulièrement dans le bloc de duplication 36 défini par Wolfe et Shield (1997) (Fig. 5.3). Ce bloc composé de deux régions dupliquées contient 6 paires de gènes paralogues. La première région de 40 kb contenant le gène *URA5* possède en plus des 6 gènes ayant un paraglogue, 13 gènes non dupliqués. La seconde région, quant à elle, possède en plus des 6 gènes, 19 gènes non dupliqués et a une taille d'environ 50kb.

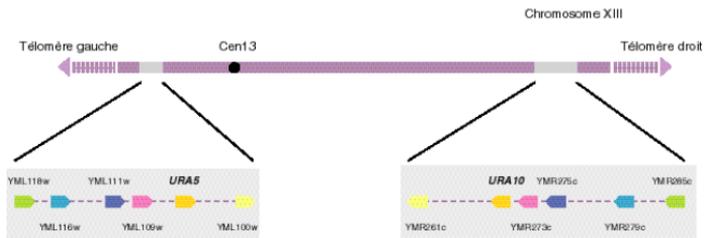


FIGURE 5.3 – Localisation chromosomique des gènes paralogues *URA5* et *URA10*. Représentation des deux régions dupliquées appartenant au bloc 36 défini par Wolfe et Shield (1997).

La duplication complète du génome a engendré deux copies d'une région ancestrale contenant au moins 38 gènes: les 32 gènes non dupliqués et les 6 paralogues. Cet événement a été suivi d'une série de délétions asymétriques dans les deux régions conduisant à la présence de 32 gènes non dupliqués. Ces deux régions initialement situées sur deux chromosomes différents après la duplication se retrouvent chacune sur un bras du chromosome XIII. De plus, l'orientation des deux régions par rapport au centromère est conservée. L'ensemble de ces données suggère qu'en plus des délétions asymétriques un événement de type translocation a conduit à la situation actuelle.

2. LES GÈNES *URA5* ET *URA10* CHEZ LES AUTRES LEVURES HÉMIASCOMYCÈTES SÉQUENCÉES

2.1. Recherche des orthologues des gènes *URA5* et *URA10*

Les gènes homologues ont été identifiés par similarité de séquence des protéines codées prédites avec celles de *S. cerevisiae* à l'aide de méthodes heuristiques comme le fait le programme BLAST (Altschul et al., 1990). Ces algorithmes ont permis de rechercher des éléments similaires aux séquences des protéines Ura5p et Ura10p grâce aux banques de données issues des séquençages partiels ou complets de génomes de levure hémiascomycètes. Dans un premier temps, cette recherche a permis de voir si les deux copies de gènes : *URA5* et *URA10* étaient présentes chez toutes les levures hémiascomycètes (Tableau 5.1). Les résultats mettent en évidence l'absence d'une des deux copies chez un certain nombre d'espèces. Chez le genre *Saccharomyces*, deux espèces : *S. mikatae* et *C. glabrata* ne possèdent qu'une seule copie de gène. Il en est de même chez l'ensemble des espèces situées sur l'arbre phylogénétique après l'espèce *S. castellii*.

Espèces		Nombre d'homologues de <i>URA5</i> et <i>URA10</i>
<i>Saccharomyces sensu stricto</i>	<i>S. cerevisiae</i>	2
	<i>S. paradoxus</i>	2
	<i>S. mikatae</i>	1
	<i>S. kudriavzevii</i>	2
	<i>S. bayanus</i>	2
<i>Saccharomyces sensu lato</i>	<i>C. glabrata</i>	1
	<i>Z. rouxii</i>	2
	<i>S. castellii</i>	2
<i>Kluyveromyces</i> genus	<i>K. lactis</i>	1
	<i>S. kluyveri</i>	1
	<i>K. waltii</i>	1
Espèces plus distantes	<i>A. gossypii</i>	1
	<i>D. hansenii</i>	1
	<i>Y. lipolytica</i>	1
	<i>S. pombe</i>	1

TABLEAU 5.1 – Nombre d'homologues aux gènes *URA5* et *URA10* de *S. cerevisiae* dans le génome de autres levures hémiascomycètes séquencés.

L'absence d'une des copies peut s'expliquer par deux hypothèses bien distinctes. La première hypothèse serait que le génome de l'espèce considérée ne soit pas issu de l'ancêtre commun avec *Saccharomyces cerevisiae* ayant dupliqué son génome. Par conséquent, le génome de ces espèces ne posséderait pas les deux régions dupliquées portant chacune un des gènes homologues. Ces espèces possèderaient donc une seule copie. Cette hypothèse pourrait effectivement expliquer la situation observée dans l'ensemble des espèces situées sur l'arbre phylogénétique après *S. castellii* (Tableau 5.1). En effet, l'événement de polyploïdie a eu lieu, d'après Wolfe et Shield, après la divergence entre les genres *Saccharomyces* et *Kluyveromyces* (Wolfe and Shield, 1997). De plus, les données issues du séquençage permettent à l'heure actuelle d'affirmer que *S. cerevisiae* est issu d'un ancêtre ayant dupliqué son génome alors que *K. waltii* (Kellis *et al.*, 2004) et *A. gossipy* (Dietrich *et al.*, 2004) ne le sont pas permettant d'expliquer la différence du nombre de copies.

La seconde hypothèse serait que l'espèce considérée soit effectivement issue de l'ancêtre commun avec *Saccharomyces cerevisiae* ayant dupliqué son génome mais que l'une des deux copies ait suivi un processus de nonfonctionnalisation. Cette absence peut alors s'expliquer par un événement simple de délétion au niveau d'une des régions dupliquées ou par un réarrangement chromosomique plus complexe telle une translocation ou une inversion conduisant à la perte d'un des deux gènes. Cette hypothèse, quant à elle, pourrait expliquer la situation observée chez les deux espèces du genre *Saccharomyces* : *S. mikatae* et *C. glabrata*.

Afin de vérifier ces hypothèses, il est indispensable de regarder si les deux régions dupliquées du bloc 36, signatures de l'événement de polyplôïdie, sont présentes ou absentes dans le génome de ces différentes espèces hémiascomycètes. Pour cela, nous nous sommes attachés à déterminer l'ensemble des gènes situés de part et d'autre des homologues des gènes *URA5* et *URA10* présentes chez les hémiascomycètes séquencés. Cela revient à rechercher la conservation de synténie.

2.2. Synténie au niveau des régions contenant *URA5* et *URA10*

Rechercher la conservation de synténie au niveau des régions dupliquées consiste à regarder si les gènes homologues présentent le même ordre et la même orientation que chez *S. cerevisiae* dans le bloc de duplication 36. Pour cela, nous avons cartographié ces régions à l'aide des mêmes banques de données que celles utilisées précédemment.

Dans un premier temps, la recherche de conservation de synténie a été réalisée pour les levures du genre *Saccharomyces* (Fig. 5.4 et Fig. 5.5). Les résultats montrent que toutes les levures analysées possèdent les deux régions dupliquées formant le bloc 36. De plus, cette recherche de synténie a permis de distinguer les orthologues du gène *URA5* d'une part et celles du gène *URA10* d'autre part dans les espèces du genre *Saccharomyces*.

L'analyse de la conservation de synténie au niveau de la région dupliquée comportant le gène *URA5* réalisée pour les espèces du genre *Saccharomyces* montre que la région environnante de ce gène est très conservée au niveau des *Saccharomyces sensu stricto* : *S. paradoxus*, *S. mikatae*, *S. kudriavzevii* et *S. bayanus* (Fig. 5.4). La conservation de l'ordre et l'orientation des gènes est quasi-parfaite. Seule la délétion des ORFs YML104c et YML103c (gènes non-essentiels chez *S. cerevisiae*) respectivement chez les espèces *S. mikatae* et *S. bayanus* engendre une rupture de synténie. Par contre, chez les espèces un peu plus distantes comme *C. glabrata* et *S. castellii*, les données montrent au niveau de ces régions dupliquées une rupture de synténie associée à un réarrangement chromosomique plus important. La présence de tels remaniements témoigne une fois de plus de la fluidité et de la dynamique des génomes. Enfin, il est important à noter que le gène *URA5* est présent dans toutes les espèces analysées.

L'analyse de la conservation de synténie au niveau de la région dupliquée comportant le gène *URA10* montre des résultats similaires à ceux obtenus pour le gène *URA5* (Fig. 5.5). Chez les *Saccharomyces sensu stricto*, la conservation est quasi-parfaite alors que pour les deux espèces un peu plus distantes de *S. cerevisiae*, en plus des délétions, des réarrangements chromosomiques plus importants vont entraîner une rupture de synténie. De plus, chez *S. mikatae* et *C. glabrata*, le seul gène est délété au niveau de cette région est le gène *URA10*. Ces données permettent de confirmer l'hypothèse que nous avons émise. En effet, la présence des régions dupliquées montre que les génomes *S. mikatae* et *C. glabrata* sont bien issus de l'ancêtre commun avec *Saccharomyces cerevisiae* ayant dupliqué son génome. L'absence du gène *URA10* résulte donc d'un simple événement de délétion qui s'est produit postérieurement à l'événement de polyploïdie.

Plus généralement, l'ensemble de ces données mettent en évidence une évolution différentielle des duplications. En effet, les copies de gènes dupliqués : *URA5* et *URA10* ont eu un destin différent chez *S. cerevisiae*, *S. mikatae* et *C. glabrata*. Alors que chez *S. cerevisiae*, les copies ont suivi un processus de subfonctionalisation, chez les deux autres espèces, le processus peut être qualifié de nonfonctionalisation.

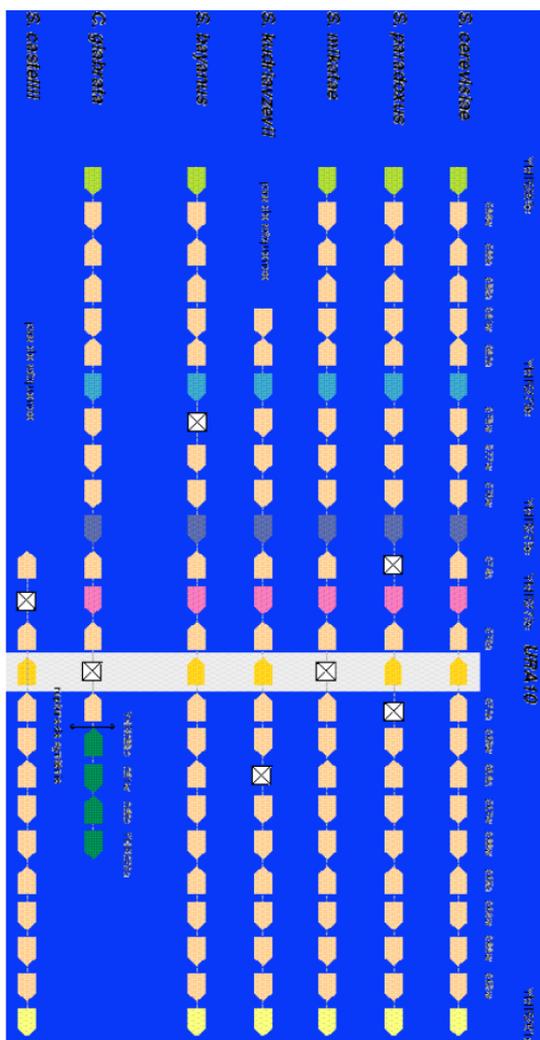


FIGURE S.5 – Synténie au niveau de la région dupliquée contenant le gène *URA10* chez les Saccharomycetes.

La recherche de conservation de synténie a également été réalisée pour les levures situées à partir du genre *Kluyveromyces* sur l'arbre phylogénétique. Pour cela, nous avons identifié les gènes situés de part et d'autre du seul homologue aux gènes *URA5* et *URA10* présent (Fig. 5.6). Dans le cas des levures *K.lactis* et *K. waltii* du genre *Kluyveromyces*, la région dans laquelle se situe le gène homologue correspond à une mosaïque des deux segments dupliqués composant le bloc 36 présent chez *S. cerevisiae*. Cette situation se voit particulièrement bien entre *K. waltii* et *S. cerevisiae* (Fig. 5.7). Chaque ORF comprise dans les deux régions dupliquées du bloc 36 se retrouve dans une seule et unique région. De plus, il y a une conservation de l'orientation des gènes. Ces données confirment l'hypothèse selon laquelle le génome de ces espèces n'est pas issu d'un ancêtre commun avec *Saccharomyces cerevisiae* ayant dupliqué son génome ce qui explique la présence d'un seul des paralogues. De plus, les espèces *K. waltii* et *S. cerevisiae* permettent de comparer clairement les génomes avant et après la duplication complète (Fig. 5.6). Dans le cadre de cette étude, *K. waltii* correspond au génome de référence à partir duquel la duplication totale aurait pu se réaliser. Les espèces *K. lactis* et *A. gossypii* présentent en plus de *K. waltii* des remaniements chromosomiques qui entraînent une rupture de synténie (Fig. 5.6). Ces réarrangements correspondent à des délétions, insertions et même inversions.

Par contre, pour les souches très distantes de *S. cerevisiae* comme *D. hansenii* et *Y. lipolytica*, la conservation de synténie est inexistante (Fig. 5.6). Il s'agit d'un exemple de comparaison de génomes d'espèces très distantes sur l'échelle évolutive. Cette comparaison va être très difficile, voire impossible, car l'accumulation de réarrangements va conduire à la disparition de toute base commune c'est-à-dire de blocs synténiques entre ces espèces.

Nous venons ainsi de voir que la présence des paralogues *URA5* et *URA10* ou la conservation des deux régions dupliquées du bloc 36 peuvent servir de marqueur de l'événement de polyploïdie et permettent de préciser sa position sur l'arbre phylogénétique (Fig. 5.1). À ce jour, les résultats montrent donc que cet événement a eu lieu entre *K. lactis* et *S. castellii*. L'analyse des données Génolevures I montre également la présence de 2 orthologues aux gènes *URA5* et *URA10* chez *Z. rouxii* soutenant l'hypothèse précédente (Tab. 5.1).

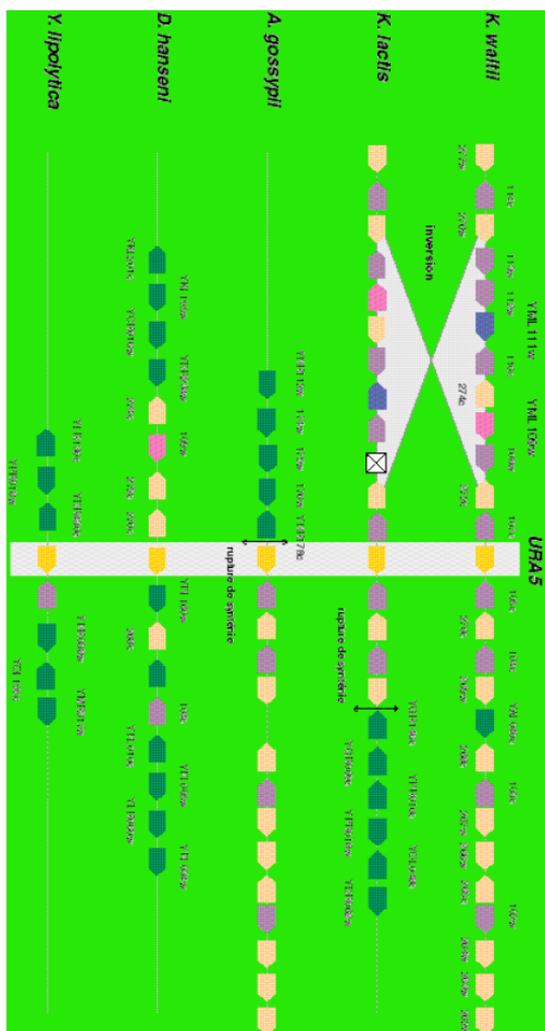


FIGURE S6 – Synténie au niveau de la région dupliquée contenant le gène *URA5* à partir des espèces du genre *Kluyveromyces*.

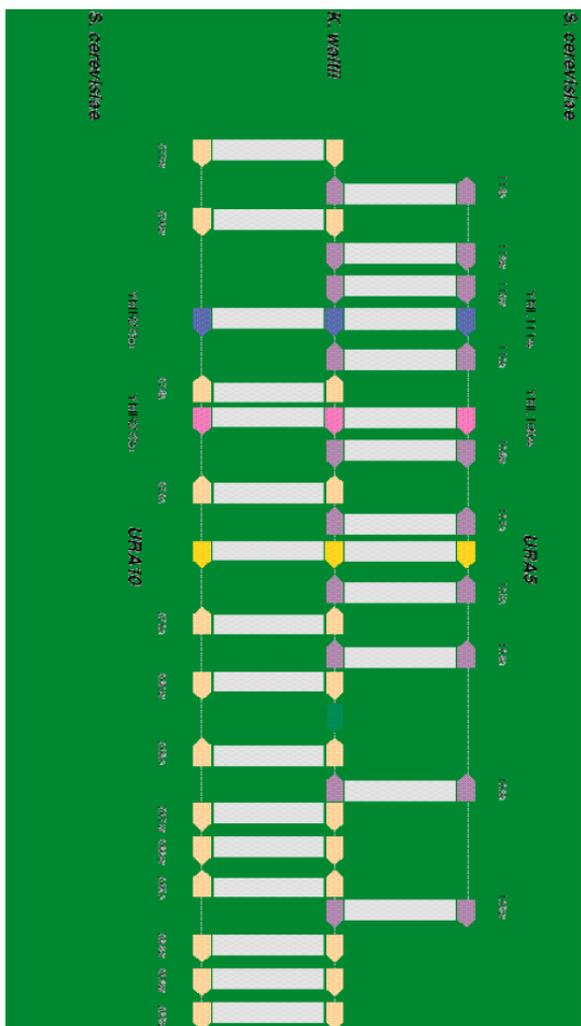


FIGURE S7 – Régions orthologues chez *S. cerevisiae* et *K. waltii*.

2.3. Évolution des deux copies de gène

Dans cette partie, nous nous sommes intéressés à l'évolution relative des deux copies de gène. En effet, Ohno (1970) proposait que la duplication était le principal mécanisme de l'évolution en se basant sur le principe que la présence de deux copies permet une flexibilité génique. Dans son modèle, une des deux copies va évoluer plus rapidement que l'autre aboutissant souvent à un pseudogène. La duplication conduit au relâchement de la pression de sélection naturelle sur les deux copies et autorise l'accumulation de mutations. Cependant, la modification d'une des deux copies restaure la pression de sélection sur l'autre qui doit rester fonctionnelle pour permettre la conservation de la fonction initiale.

Afin de pouvoir comparer l'évolution des gènes paralogues, le pourcentage d'identité de séquence protéique des homologues par rapport aux séquences des protéines *Ura5p* et *Ura10p* de *S. cerevisiae* a été déterminé (Tableau 5.2). Il en a été fait de même avec la protéine *Ura3p* car le gène correspondant, à copie unique, a servi de témoin.

Les résultats de comparaison des séquences protéiques montrent que le seul homologue présent chez les espèces situées à partir du genre *Kluyveromyces* sur l'arbre phylogénétique sont plus proches du gène *URA5* que du gène *URA10*.

De plus, les pourcentages d'identité de séquence protéique (Tableau 5.1) permettent de mettre en évidence une évolution différentielle entre les copies dupliquées. Les gènes *URA5* et *URA3* présentent des pourcentages d'identité de séquence équivalents au cours de l'évolution des espèces hémiascomycètes. Par contre, si on regarde l'évolution du gène *URA10*, les séquences semblent moins conservées. Pour *S. castellii*, par exemple, l'identité de séquence de la protéine *Ura10p* par rapport à celle de *S. cerevisiae* est de 72 % alors qu'elle est de 89 % pour *Ura5p*.

L'ensemble de ces données permettent de proposer que le gène *URA5* soit la copie garante de la fonction originelle et *URA10* la copie ayant évolué de manière indépendante. Cela est confirmé par les données fonctionnelles chez *S. cerevisiae*. En effet, l'enzyme codée par le gène *URA10* n'intervient que pour 20 % dans l'activité OMPase totale. La perte du gène *URA10* chez *S. mikatae* et *C. glabrata* indique également que le gène *URA5* a conservé la fonction originelle et renforce encore davantage cette hypothèse.

Espèces		<i>URA5</i> % d'identité	<i>URA10</i> % d'identité	<i>URA3</i> % d'identité
<i>Saccharomyces sensu stricto</i>	<i>S. cerevisiae</i>	100	100	100
	<i>S. paradoxus</i>	99	92	99
	<i>S. mikatae</i>	96	abs.	96
	<i>S. kudriavzevii</i>	98	84	96
	<i>S. bayanus</i>	97	88	95
<i>Saccharomyces sensu lato</i>	<i>C. glabrata</i>	85	abs.	82
	<i>Z. rouxii</i>	78	67	nd
	<i>S. castellii</i>	89	72	85
<i>Kluyveromyces</i> genus	<i>K. lactis</i>	81	abs.	83
	<i>S. kluyveri</i>	77	abs.	
	<i>K. waltii</i>	77	abs.	81
Espèces plus distantes	<i>A. gossypii</i>	79	abs.	72
	<i>D. hansenii</i>	66	abs.	73
	<i>Y. lipolytica</i>	58	abs.	
	<i>S. pombe</i>	46	abs.	52

TABLEAU 5.2 – Pourcentages d'identité de séquence protéique des homologues par rapport aux séquences protéiques *Ura5p* et *Ura10p* de *S. cerevisiae*.

CONCLUSION

L'analyse de l'évolution des gènes paralogues *URA5* et *URA10* au sein des hémiascomycètes a été possible suite au séquençage partiel ou complet d'un certain nombre de levures appartenant à ce groupe. La comparaison des séquences a permis de mettre en évidence que la duplication totale du génome s'est produite après la divergence de *K.lactis* et *S. castellii*. En effet, les espèces situées après *S. castellii* ne possèdent plus de gène *URA10* mais seulement la copie du gène *URA5*. Ainsi, la présence de ces paralogues peut servir de marqueur de la polyploïdisation et permettre de dater cet événement.

D'autre part, la comparaison entre les séquences des gènes *URA5* et *URA10* au sein du genre *Saccharomyces* met en évidence une évolution différentielle entre les deux copies dupliquées. Les résultats suggèrent que le gène *URA5* est la copie garante de la fonction originelle alors que le gène *URA10* a évolué de manière indépendante sans pression de sélection. En effet, les copies de gènes dupliqués : *URA5* et *URA10* ont eu un destin différent chez les espèces du genre *Saccharomyces*. Chez *S. cerevisiae*, ces deux gènes présentent des différences de régulation de leur expression et constituent dans ce cas un exemple de subfonctionnalisation. Par contre chez *S. mikatae* et *C. glabrata*, le gène *URA10* a été délété et a ainsi suivi un processus de nonfonctionnalisation. Pour les autres espèces séquencées (*S. paradoxus*, *S. kudriavzevii*, *S. bayanus* et *S. castellii*), le gène *URA10* est toujours présent. Dans ces espèces, il serait intéressant de savoir si cette copie code une protéine ayant une activité OMPpase. Dans ce cas, la situation serait similaire à celle rencontrée chez *S. cerevisiae*. Par contre, il n'est pas exclu que ces copies aient suivi un processus de nonfonctionnalisation c'est-à-dire qu'elle portent des mutations conduisant à une perte de la fonction OMPpase. De même, ces copies peuvent également avoir suivi un processus de néofonctionnalisation et portent des mutations conduisant à une spécificité fonctionnelle nouvelle.

C

CONCLUSION GENERALE ET PERSPECTIVES

Si « la nécessité est la mère de toute invention », il est cependant nécessaire d'avoir un vecteur permettant de conduire à toute évolution. L'innovation nécessite donc d'une part un motif et d'autre part une occasion. Chez les organismes vivants, un des aspects les plus importants de l'évolution sur le plan moléculaire est l'émergence de nouvelles fonctions. Dans ce cas de figure, il est clair que le motif prédominant correspond à un avantage sélectif et l'occasion est créée par l'ensemble des réarrangements chromosomiques auquel une cellule peut être soumise. Dans un premier temps, il y a formation d'un remaniement, au sens large, puis dans un deuxième temps, il peut y avoir sélection et fixation de ces remaniements. L'évolution des organismes vivants implique la modification, l'acquisition, la délétion ainsi que le réarrangement du matériel génétique qui les constituent.

L'étude de l'évolution des génomes passe donc par la compréhension des mécanismes impliqués dans la formation de remaniements génomiques. Afin de pouvoir étudier les facteurs intervenant dans l'apparition de tels remaniements, il est nécessaire de disposer de systèmes de sélection efficaces permettant d'analyser ce qui se passe spontanément dans une cellule. Au laboratoire, nous possédons un système génétique, basé sur un allèle particulier du gène *URA2*, qui permet la sélection positive de réarrangements chromosomiques chez la levure *Saccharomyces cerevisiae*: insertion de séquences d'ADN mobiles (rétrotransposons), délétion de fragments chromosomiques mais également duplication de régions chromosomiques.

Notre étude s'est focalisée sur les duplications. Deux collections de révertants bien distinctes ont été analysées sur le plan moléculaire : l'une réalisée à partir d'une souche haploïde *ura2 15-30-72*, l'autre à partir d'une souche diploïde homozygote *a/a ura2 15-30-72*. Une telle étude avait un double objectif. Le premier était de voir par quels moyens une séquence d'ADN peut se dupliquer et par cet intermédiaire d'apporter une preuve expérimentale quant à leur pluralité. Dans ce cadre, il a été intéressant de comparer les deux états: haploïde et diploïde afin de savoir si les moyens mis en œuvre pour dupliquer une séquence étaient similaires ou différents. Le deuxième objectif de

l'analyse de ces révertants était d'apporter des réponses quant aux mécanismes pouvant être à l'origine de ces duplications et de proposer des modèles de formation de ces remaniements.

Ainsi, dans la première partie de notre étude, nous nous sommes attachés à l'analyse moléculaire des duplications isolées à l'état haploïde. Le crible de sélection, développé au laboratoire, a permis d'isoler 8 révertants porteurs d'une duplication à partir de la souche haploïde *ura2 15-30-72*. L'analyse moléculaire, et plus particulièrement la détermination de la taille des régions dupliquées, a mis en évidence deux types d'événements de duplication. Chez 4 révertants, la duplication est limitée à une partie du gène *URA2* et constitue de la duplication génique. Par contre, chez les 4 autres révertants, l'événement correspond à un segment englobant la séquence codant le domaine ATCase ainsi qu'une série de gènes contigus et constitue de la duplication segmentale. Ces données montrent qu'une même séquence peut se dupliquer par au moins deux processus distincts et apportent une preuve expérimentale sur la formation de tels remaniements.

Pour les 4 révertants porteurs d'une duplication génique, l'ensemble des résultats de l'analyse moléculaire nous ont permis de proposer un mécanisme expliquant la formation des duplications. En effet, dans ces révertants, la région dupliquée est généralement ponctuée en 3' par une séquence poly(A) et se retrouve entre des séquences de rétrotransposons de type Ty1. La présence de la région poly(A) en 3' de la copie dupliquée nous a orientés vers un mécanisme médié par la réverse transcription de l'ARNm correspondant c'est-à-dire un mécanisme de rétroposition.

Afin de vérifier cette hypothèse, l'analyse de révertants dans deux contextes affectant un processus de rétroposition a été entreprise. Les rétrotransposons Ty1 étant une source évidente d'activité de transcriptase inverse, 23 révertants portant un événement de duplication ont été sélectionnés à partir d'une souche dans laquelle un élément Ty1 a été surexprimé. Dans un tel contexte, la fréquence des duplications géniques ponctuées en 3' par une séquence poly(A) est considérablement augmentée. Par contre, aucune duplication génique de ce type n'a été sélectionnée à l'état diploïde dans lequel la transposition est inhibée. L'ensemble de ces résultats apportent des arguments incontestables plaidant en faveur de notre hypothèse à savoir que les événements de duplication sélectionnés sont issus d'un processus de rétroposition.

Dans la suite de cette étude, il serait intéressant de réaliser deux expériences qui permettraient de confirmer définitivement la relation entre les duplications géniques et la rétroposition (Fig. C.1). La première consisterait à construire une souche haploïde *ura2 15-30-72* dans laquelle une structure intronique aurait été insérée au niveau de la séquence codant le domaine ATCase. L'objectif serait de suivre la perte de l'intron au cours de l'événement de duplication génique. En effet, si la genèse de ces duplications passe par la formation d'un ADNc issu de la rétrotranscription de l'ARNm correspondant à la séquence codant le domaine ATCase, la copie dupliquée sera dépourvue de l'intron. Ainsi, la sélection de révertants possédant une copie limitée à une partie du gène *URA2* dépourvu d'intron et ponctuée d'une séquence poly(A) serait une preuve irréfutable de l'implication d'un mécanisme de rétroposition. Une seconde possibilité serait de construire une souche dans laquelle le promoteur de l'allèle haploïde *ura2 15-30-72* aurait été délété abolissant ainsi toute transcription de l'ARNm correspondant. Dans un mécanisme par rétroposition, aucun ADNc ne pourrait être formé ce qui abolirait toute sélection de duplication génique. Ainsi, la sélection à partir d'une telle souche permettrait uniquement d'isoler des événements de duplication segmentale.

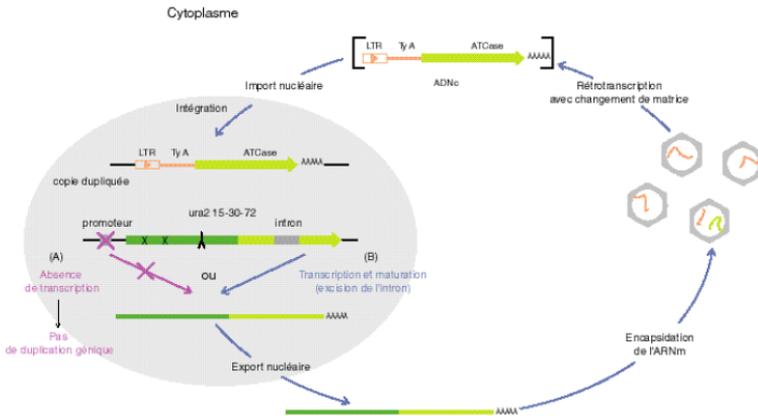


FIGURE C.1 – Stratégie d'étude du mécanisme de rétroposition. Le fond gris délimite le noyau.

L'ADN double brin est représenté par une flèche épaisse et son ARN par un rectangle mince.

(A) Absence de transcription du gène *ura2 15-30-72*

(B) Suivi de la perte d'un intron inséré dans la séquence codant le domaine ATCase.

À côté de ces duplications géniques, des événements de duplication segmentale ont été mis en évidence chez les 4 autres révertants haploïdes spontanés. La taille, variant entre 5 et 90kb, a été déterminée grâce à la technique d'hybridation génomique comparative sur microarray. L'analyse moléculaire a montré que ces réarrangements correspondent à des duplications segmentales en tandem direct. La partie dupliquée du gène *URA2* est insérée dans une région intergénique ou dans un gène non essentiel. Dans ce second cas, la fusion entre le gène et le fragment dupliqué s'est produite en phase ce qui entraîne la formation d'un gène mosaïque.

Afin d'apporter des réponses quant aux mécanismes impliqués dans la formation de tels remaniements, nous avons testé l'importance de la recombinaison homologue. Pour cela, le rôle du gène *RAD52*, dont le produit est indispensable à la recombinaison homologue, dans la formation des duplications a été testé. À partir d'une souche *ura2 15-30-72rad52Δ*, 6 révertants portant un événement de duplication ont été sélectionnés et analysés. L'ensemble des réarrangements sélectionnés correspond à des duplications segmentales ce qui démontre clairement que le mécanisme est indépendant de la recombinaison homologue. Ces données ainsi que l'architecture des duplications sélectionnées : duplication en tandem direct nous ont amené à proposer un modèle de mécanisme par erreur de réplication de type BIR (break-induced replication).

Dans la suite de cette étude, il serait intéressant de tester le modèle proposé et d'apporter des arguments permettant de le valider. À cet effet, il serait nécessaire de voir dans quelle mesure certains gènes impliqués dans la réplication le sont également dans la formation des duplications de segment en tandem. Dans ce cadre, une sélection à partir de la souche *ura2 15-30-72* délétée des gènes *CLB5* ou *RAD27* permettrait d'apporter des réponses. Si le mécanisme est dépendant de la réplication, la formation des duplications segmentales devrait être affectée dans de tels contextes. Le gène *RAD27* code une endo/exonucléase nécessaire pour la maturation des fragments d'Okazaki au cours de la réplication (Lieber, 1997). Le gène *CLB5* est quant à lui impliqué, entre autres, dans l'initiation de la réplication (Schwob and Nasmyth, 1996).

De plus, dans notre modèle (chapitre IV), nous proposons qu'une cassure double brin puisse être à l'origine des événements de duplication segmentale sélectionnés. Ainsi, il pourrait également être intéressant d'introduire un site HO qui permet d'induire des cassures double brin en amont de la séquence codant le domaine ATCase. Selon notre modèle, cette induction devrait favoriser la formation de duplications segmentales en tandem.

Dans la deuxième partie de notre étude, nous nous sommes attachés à l'analyse moléculaire des duplications isolées à l'état diploïde. La sélection de 14 révertants isolés à partir de la souche diploïde homozygote *a/a ura2 15-30-72* a permis d'avoir une vision globale sur les événements de duplication et de compléter les résultats obtenus à l'état haploïde. Les résultats ont mis en évidence deux autres moyens par lesquels une séquence d'ADN peut se dupliquer : l'aneuploïdie et la translocation non-réciproque. Parmi la collection des 14 révertants diploïdes analysés, seul un révertant présente un événement de duplication segmentale en tandem similaire à ceux observés à l'état haploïde.

Chez 6 révertants diploïdes, la détermination de la taille de la région dupliquée montre que la partie du bras gauche du chromosome X couvrant la région située entre le gène *URA2* et le télomère a été dupliquée. L'analyse globale du nombre de copies de chaque gène réalisée grâce à la technique d'hybridation génomique comparative par microarray (aCGH) montre la formation d'un chromosome chimère surnuméraire.

Chez 2 révertants, la détermination de la taille montre également que le bras gauche du chromosome X, couvrant la région située entre le gène *URA2* et le télomère, a été dupliqué. Cependant, l'hybridation génomique comparative par microarray (aCGH) montre que la région dupliquée a remplacé une région délétée formant ainsi un chromosome chimère. La duplication observée chez ces révertants correspond à une translocation non-réciproque.

Dans la suite de cette étude, il serait intéressant d'achever l'analyser les 5 souches: B177, B264, B267, B326 et B328 par microarray CGH. L'analyse génétique de ces révertants suggère un événement de translocation non-réciproque. De plus, en se basant sur les sites d'insertion déterminés par clonage, nous avons estimé la taille des chromosomes chimères potentiels en additionnant la région dupliquée au chromosome dont une partie serait délétée. Dans tous les cas, la taille estimée est en corrélation avec la taille du chromosome porteur de la copie déterminée par l'analyse du caryotype. Une analyse globale des copies de gènes par array CGH permettra donc de confirmer cette hypothèse.

Dans le cas des deux types de remaniements sélectionnés (aneuploïdie et translocation non-réciproque) à partir d'une souche diploïde homozygote *a/a ura2 15-30-72*, aucune longue séquence homologue n'est trouvée à la jonction suggérant que les mécanismes impliqués dans leur formation sont indépendants de la recombinaison

homologue. Par contre, un mécanisme de type BIR (break-induced replication) pourrait permettre la formation de chromosomes chimériques en l'absence de longues séquences homologues.

Afin de vérifier cette hypothèse, il serait indispensable de tester le modèle proposé par des sélections de révertants dans un contexte affectant la réplication. Comme pour les événements de duplication segmentale, des sélections avec des souches mutées au niveau du gène *RAD27* ou du gène *CLB5* devraient permettre de confirmer ce modèle.

De plus, dans le cadre des chromosomes surnuméraires chimères, il serait intéressant de voir si la situation « aneuploïde » est stable chez la levure *S. cerevisiae*. L'aneuploïdie est très répandue chez *Saccharomyces cerevisiae* (Hughes *et al.*, 2000), il existe cependant très peu de données sur la stabilité d'une telle situation génomique. Ainsi, dans nos révertants, la perte de la copie dupliquée (conférant la prototrophie à l'uracile) pourrait être suivie et permettrait d'étudier la stabilité lors de culture dans un milieu sans pression de sélection. Cette stabilité pourrait être étudiée à l'état diploïde (révertant sélectionné) mais également à l'état haploïde sur le descendant du révertant porteur d'un chromosome surnuméraire.

Plus généralement, les deux premières parties de notre étude ont permis d'apporter des éléments de réponses quant aux questions que nous nous étions initialement posées. En effet, l'analyse des révertants à l'état haploïde et diploïde a mis en évidence la pluralité des moyens permettant à une même séquence de se dupliquer. Cette étude constitue la première preuve expérimentale qu'une même séquence d'ADN peut suivre au moins quatre processus différents afin de se dupliquer : la duplication de gène, de segment, de bras de chromosome ou encore de chromosome entier. Le crible génétique basé sur un allèle mutant du gène *URA2* est un système qui nous a permis d'avoir une vision globale sur les événements de duplication. De plus, notre étude avait également pour but d'apporter des réponses quant aux mécanismes pouvant être à l'origine de ces duplications. Là encore, l'analyse des révertants a permis de mettre en évidence la pluralité des mécanismes impliqués et de proposer des modèles.

Lors de cette étude, nous avons également démontré que les événements de duplication sélectionnés sont dépendants de la ploïdie. Ce travail a en outre montré que la phase diploïde permet une grande souplesse génomique notamment avec la possibilité

de perte de fragment d'ADN ou de gènes essentiels à la survie cellulaire à l'état haploïde. La ploïdie induit un biais dans la sélection des événements et doit donc être prise en compte.

Enfin, les remaniements génomiques observés dans notre étude en réponse à une pression de sélection peuvent également être mis en corrélation avec ceux retrouvés lors du développement des cellules cancéreuses. La progression tumorale est étroitement liée à l'accumulation de nombreuses modifications au niveau du génome des cellules cancéreuses correspondant souvent à des réarrangements chromosomiques et notamment des événements de duplication, de translocation non-réciproque et d'aneuploïdie. L'analyse de ces réarrangements permet de formuler des hypothèses quant aux mécanismes impliqués dans leur formation. Ces hypothèses peuvent être expérimentalement testées en conditions de laboratoire. La levure se prête particulièrement bien à de telles études et les résultats expérimentaux obtenus chez cet eucaryote peuvent apporter des informations sur les mécanismes impliqués dans le même type de réarrangement retrouvés chez les eucaryotes supérieurs.

BIBLIOGRAPHIE

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., Eichler, E.E. 2001. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Research* **11**:1005-1017.

Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., Eichler, E.E. 2002. Recent segmental duplications in the human genome. *Science* **297**: 1003-1007.

Bailey, J.A., Liu, G., Eichler, E.E. 2003. An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* **73**: 823-834.

Bailey, J.A., Church, D.M., Ventura, M., Rocchi, M., Eichler, E.E. 2004. Analysis of segmental duplications and genome assembly in the mouse. *Genome Research* **14**: 789-801.

Blandin, G., Durrens, P., Tekaia, F., Aigle, M., Bolotin-Fukuhara, M., Bon, E., Casaregola, S., de Montigny, J., Gaillardin, C., Lepingle, A. et al. 2000. Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett* **487**: 31-36.

Boeke, J.D., Eichinger, D., Castrillon, D., Fink, G.R. 1988. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. *Mol Cell Biol* **8**: 1432-1442.

Boeke, J.D., Garfinkel, D.J., Styles, C.A., Fink, G.R. 1985. Ty elements transpose through an RNA intermediate. *Cell* **40**: 491-500.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W. 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* **2**: 95-113.

- Bonneaud, N., Ozier-Kalogeropoulos, O., Li, G.Y., Labouesse, M., Minvielle-Sebastia, L., Lacroute, F. 1991. A family of low and high copy replicative, integrative and single-stranded *S. cerevisiae*/*E. coli* shuttle vectors. *Yeast* **7**: 609-15.
- Bridges, C. 1936. The bar "gene" duplication. *Science* **83**: 210-211.
- Burmester, T., Weich, B., Reinhardt, S., Hankeln, T. 2000. A vertebrate globin expressed in the brain. *Nature* **407**: 520-523.
- Butt, T.R., Ecker, D.J. 1987. Yeast metallothionein and applications in biotechnology. *Microbiol Rev* **51**: 351-364.
- Camasses, A. 1996. Natural translocation of a large segment of chromosome III to chromosome I in a laboratory strain of *Saccharomyces cerevisiae*. *Curr Genet* **30**: 218-223.
- Carlson, M., Celenza, J.L., Eng, F.J. 1985. Evolution of the dispersed *SUC* gene family of *Saccharomyces* by rearrangements of chromosome telomeres. *Mol Cell Biol* **5**: 2894-2902.
- Chen, C., Umezu, K., Kolodner, R.D. 1998. Chromosomal rearrangements occur in *S. cerevisiae rfa1* mutator mutants due to mutagenic lesions processed by double-strand-break repair. *Mol Cell* **2**: 9-2.
- Chen, C., Kolodner, R.D. 1999. Gross chromosomal rearrangements in *Saccharomyces cerevisiae* replication and recombination defective mutants. *Nat Genet* **23**: 81-85.
- Christman, M.F., Dietrich, F.S., Fink, G.R. 1988. Mitotic recombination in the rDNA of *S. cerevisiae* is suppressed by the combined action of DNA topoisomerases I and II. *Cell* **55**: 413-425.
- Chu, G. 1989. Pulsed field electrophoresis in contour-clamped homogeneous electric fields for the resolution of DNAs by size or topology. *Electrophoresis* **10**: 290-295.
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., Johnston, M. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**: 71-76.
- Codon, A.C., Benitez, T., Korhola, M. 1998. Chromosomal polymorphism and adaptation to specific industrial environments of *Saccharomyces* strains. *Appl Microbiol Biotechnol* **49**: 154-163.

- Company, M., Errede, B. 1987. Cell-type-dependent gene activation by yeast transposon Ty1 involves multiple regulatory determinants. *Mol Cell Biol* **7**: 3205-3211.
- Copley, R.R., Goodstadt, L., Ponting, C. 2003. Eukaryotic domain evolution inferred from genome comparisons. *Curr Opin Genet Dev* **13**: 623-628.
- Curcio, M.J., Garfinkel, D.J. 1992. Posttranslational control of Ty1 retrotransposition occurs at the level of protein processing. *Mol Cell Biol* **12**: 2813-2825.
- Curcio, M.J., Garfinkel, D.J. 1991. Regulation of retrotransposition in *Saccharomyces cerevisiae*. *Mol Microbiol* **5**:1823-1829.
- Davis A.P., Symington L.S. 2004. RAD-51 dependent break induced replication in yeast. *Mol Cell Biol* **24**: 2344-2351.
- De Montigny, J., Kern, L., Hubert, J.C., Lacroute, F. 1990. Cloning and sequencing of *URA10*, a second gene encoding orotate phosphoribosyl transferase in *Saccharomyces cerevisiae*. *Curr Genet* **17**: 105-111.
- Derr, L.K., Strathern, J.N., Garfinkel, D.J. 1993. RNA-mediated recombination in *S.cerevisiae*. 1991. *Cell* **18**: 355-364.
- Dehal, P., Predki, P., Olsen, A.S., Kobayashi, A., Folta, P., Lucas, S., Land, M., Terry, A., Ecale Zhou, C.L., Rash et al. 2001. Human chromosome 19 and related regions in mouse: conservative and lineage-specific evolution. *Science* **293**: 104-111.
- Dhellin, O., Maestre, J., Heidmann, T. 1997. Functional differences between the human LINE retrotransposon and retroviral reverse transcriptases for in vivo mRNA reverse transcription. *EMBO J* **16**: 6590-6602.
- Dietrich, F.S., Voegeli, S., Brachat, S., Lerch, A., Gates, K., Steiner, S., Mohr, C., Pohlmann, R., Luedi, P., Choi, S., Wing, R.A., Flavier, A., Gaffney, T.D., Philippsen, P. 2004. The *Ashbya gossypii* genome as a tool for mapping the ancient *Saccharomyces cerevisiae* genome. *Science* **304**: 304-307.
- Dorsey, M.J., Hoeh, P., Paquin, C.E. 1993. Phenotypic identification of amplifications of the ADH4 and CUP1 genes of *Saccharomyces cerevisiae*. *Curr Genet* **12**: 263-396.
- Dower, W.J., Miller, J.F., Ragsdale, C.W. 1988. High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res* **16**: 6127-6145.

- Duda, T.F., Palumbi, S.R. 1999. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci USA* **96**: 6820-6823.
- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E. et al. 2004. Genome evolution in yeasts. *Nature* **430**: 35-44.
- Dunham, M.J., Badrane, H., Ferea, T., Adams, J., Brown, P.O., Rosenzweig, F., Botstein, D. Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc Natl Acad Sci USA* **99**:16144-16149.
- Eichler, E.E. 2001. Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* **17**: 661-669.
- Eichinger, D.J., Boeke, J.D. 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* **54**: 955-966.
- Elder, R.T., Loh, E.Y., Davis, R.W. 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. *Proc Natl Acad Sci USA* **80**: 2432-2436.
- Emanuel, B.S., Shaikh, T.H. 2001. Segmental duplications: an 'expanding' role in genomic instability and disease. *Nat Rev Genet* **2**: 791-800.
- Errede, B., Cardillo, T.S., Sherman, F., Dubois, E., Deschamps, J., Wiame J.M. 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. *Cell* **22**: 427-436.
- Errede, B., Company, M., Hutchison, C.A. 1987. Ty1 sequence with enhancer and mating-type-dependent regulatory activities. *Mol Cell Biol* **7**: 258-265.
- Esnault, C., Maestre, J., Heidmann, T. 2000. Human LINE retrotransposons generate processed pseudogenes. *Nat Genet* **24**: 363-367.
- Fasullo, M.T., Davis, R.W. Direction of chromosome rearrangements in *Saccharomyces cerevisiae* by use of his3 recombinational substrates. *Mol Cell Biol* **8**: 4370-4380.

- Fewell, S.W., Woolford, J.L. 1999. Ribosomal protein S14 of *Saccharomyces cerevisiae* regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Mol Cell Biol* **19**: 826-834.
- Fink, GR. 1987. Pseudogenes in yeast? *Cell* **49**: 5-6.
- Fischer, G., James, S.A., Roberts, I.N., Oliver, S.G., Louis, E.J. 2000. Chromosomal evolution in *Saccharomyces*. *Nature* **405**: 451-454.
- Fleischmann, R.D., Adams, M.D., White, O., Clayton, R.A., Kirkness, E.F., Kerlavage, A.R., Bult, C.J., Tomb, J.F., Dougherty, B.A., Merrick, J.M. et al. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. *Science* **269**: 496-512.
- Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M. et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**: 397-403.
- Fogel, S., Welch, J.W. 1982. Tandem gene amplification mediates copper resistance in yeast. *Proc Natl Acad Sci USA* **79**: 5342-5346.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**: 1531-1545.
- Garrels, J.I., McLaughlin, C.S., Warner, J.R., Fitcher, B., Latter, G.I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D.S., Mesquita-Fuentes, R., Payne, W.E. 1997. Proteome studies of *Saccharomyces cerevisiae*: identification and characterization of abundant proteins. *Electrophoresis* **18**: 1347-1360.
- Gibbs et al. (Rat Genome Sequencing Project Consortium). 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**: 493-521.
- 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., Oliver, S.G. 1996. Life with 6000 genes. *Science* **274**: 563-567.
- Greer, H., Fink, G.R. 1979. Unstable transpositions of *his4* in yeast. *Proc Natl Acad Sci USA* **76**: 4006-4010.

- Grutzner, F., Crollius, H.R., Lutjens, G., Jaillon, O., Weissenbach, J., Ropers, H.H., Haaf T. 2002. Four-hundred million years of conserved synteny of human Xp and Xq genes on three Tetraodon chromosomes. *Genome Research* **12**: 1316-1322.
- Gu, Z., Cavalcanti, A., Chen, F.C., Bouman, P., Li, W.H. 2002. Extent of gene duplication in the genomes of *Drosophila*, nematode, and yeast. *Mol Biol Evol* **19**: 256-262.
- Hansche, P.E., Beres, V., Lange, P. 1978. Gene duplication in *Saccharomyces cerevisiae*. *Genetics* **88**: 673-687.
- Harrison, P.M., Echols, N., Gerstein, M.B. 2001. Digging for dead genes: an analysis of the characteristics of the pseudogene population in the *Caenorhabditis elegans* genome. *Nucleic Acids Research* **29**: 818-830.
- Harrison, P.M., Hegyi, H., Balasubramanian, S., Luscombe, N.M., Bertone, P., Echols, N., Johnson, T., Gerstein, M. 2002. Molecular fossils in the human genome: identification and analysis of the pseudogenes in chromosomes 21 and 22. *Genome Research* **12**: 272-280.
- Hartwell, L.H., Smith, D. Altered fidelity of mitotic chromosome transmission in cell cycle mutants of *S. cerevisiae*. *Genetics* **110**: 381-395.
- Himmelreich, R., Hilbert, H., Plagens, H., Pirkl, E., Li, B.C., Herrmann, R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res* **24**: 4420-4449.
- Hiraoka, M., Watanabe, K., Umezu, K., Maki, H. 2000. Spontaneous loss of heterozygosity in diploid *Saccharomyces cerevisiae* cells. *Genetics* **156**: 1531-1548.
- Hoffman, C.S., Winston, F. 1987. A ten-minute DNApreparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**: 267-272.
- Holmes, D.S., Quigley, M. 1981. A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114**: 193-197.
- Hughes, A.L. 1994. The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lond B Biol Sci* **256**:119-124.

Hughes, T.R., Roberts, C.J., Dai, H., Jones, A.R., Meyer, M.R., Slade, D., Burchard, J., Dow, S., Ward, T.R., Kidd, M.J., Friend, S.H., Marton, M.J. 2000. Widespread aneuploidy revealed by DNAmicroarray expression profiling. *Nat Genet* **25**: 333-337.

International Human Genome Mapping Consortium.2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921

International Human Genome Mapping Consortium.2001. A physical map of the human genome. *Nature* **409**: 934-941.

Ira G., Haber J.E. 2002. Characterization of RAD51 independent break-induced replication that acts preferentially with short homologous sequences. *Mol Cell Biol* **22**: 6384-6392.

Jordan, I.K. and McDonald, J.F. 1998. Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. *J Mol Evol* **47**: 14-20.

Jund, R., Lacroute, F. 1972. Regulation of orotidylic acid pyrophosphorylase in *Saccharomyces cerevisiae*. *J Bacteriol* **109**: 196-202.

Kallioniemi, O.P., Kallioniemi, A., Piper, J., Isola, J., Waldman, F.M., Gray, J.W., Pinkel, D. 1994. Optimizing comparative genomic hybridization for analysis of DNA sequence copy number changes in solid tumors. *Genes Chromosomes Cancer* **10**: 231-243.

Karin, M., Najarian, R., Haslinger, A., Valenzuela, P., Welch, J., Fogel, S. 1984. Primary structure and transcription of an amplified genetic locus: the *CUP1* locus of yeast. *Proc Natl Acad Sci USA* **81**: 337-341.

Kim, J.M., Vanguri, S., Boeke, J.D., Gabriel, A., Voytas, D.F. 1998. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Research* **8**: 464-478.

Keil, R.L., McWilliams, A.D. 1993. A gene with specific and global effects on recombination of sequences from tandemly repeated genes in *Saccharomyces cerevisiae*. *Genetics* **135**: 711-718.

Kellis, M., Patterson, N., Endrizzi, M., Birren, B., Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241-254.

- Kellis, M., Birren, B.W., Lander, E.S. 2004. Proof and evolutionary analysis of ancient genome duplication in the yeast *Saccharomyces cerevisiae*. *Nature* **428**: 617-624.
- Kenna, M.A., Brachmann, C.B., Devine, S.E., Boeke, J.D. 1998. Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition in vivo. *Mol Cell Biol* **18**:1115-1124.
- Klein, H.L. 1995. Examination of mitotic recombination by means of hyper-recombination mutants in *Saccharomyces cerevisiae*. *Prog Nucleic Acid Res Mol Biol* **51**: 271-303.
- Klenk, H.P., Clayton, R.A., Tomb, J.F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D. et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**: 364-370.
- Kozul, R., Caburet, S., Dujon, B., Fischer, G. 2004. Eucaryotic genome evolution through the spontaneous duplication of large chromosomal segments. *EMBO J* **23**: 234-243.
- Kupiec, M., Petes, T.D. 1988. Allelic and ectopic recombination between Ty elements in yeast. *Genetics* **119**: 549-559.
- Kurtzman C.P. & J.W. Fell (1998) *The Yeasts, a taxonomic study* (fourth edition). Elsevier.
- Lengauer, C., Kinzler, K.W., Vogelstein, B. 1998. Genetic instabilities in human cancers. *Nature* **396**: 643-649.
- Lespinet, O., Wolf, Y.I., Koonin, E.V., Aravind, L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research* **12**:1048-1059.
- Levine, K.L., Steiner, B., Johnson, K., Aronoff, R., Quinton, T.J., Linial, M.L. 1990. Unusual features of integrated cDNAs generated by infection with genome-free retroviruses. *Mol Cell Biol* **10**: 1891-1900.
- Li, W.H. 1997. *Molecular Evolution*, Sinaur.
- Li, W.H. 2001. Evolutionary analyses of the human genome. *Nature* **409**: 847-849.
- Lieber, M.R. 1997. The FEN-1 family of structure-specific nucleases in eukaryotic DNA replication, recombination and repair. *Bioessays* **19**: 233-240.

- Liebman, S., Shalit, P., Picologlou, S. 1981. Ty elements are involved in the formation of deletions in DEL1 strains of *Saccharomyces cerevisiae*. *Cell* **26**: 401-409.
- Liu, Y.G., Mitsukawa, N., Oosumi, T., Whittier, R.F. 1995. Efficient isolation and mapping of *Arabidopsis thaliana* T-DNAinsert junctions by thermal asymmetric interlaced PCR. *Plant J* **8**: 457-63.
- Llorente B, Durrens P, Malpertuy A, Aigle M, Artiguenave F, Blandin G, Bolotin-Fukuhara M, Bon E, Brottier P, Casaregola S et al. 2000. Genomic exploration of the hemiascomycetous yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*. *FEBS Lett.* **487**: 122-133.
- Louis, E.J. 1995. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* **11**: 1553-1573.
- Lupski, J.R. 1998. Genomic disorders: structural features of the genome can lead to DNA rearrangements and human disease traits. *Trends Genet* **14**: 417-422.
- Lynch, M., Conery, J.S. 2000. The evolutionary fate and consequences of duplicate genes. *Science* **290**: 1151-1155.
- Lynch, M., O'Hely, M., Walsh, B., Force, A. 2001. The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**: 1789-1804.
- Lynch, M. 2002. Genomics. Gene duplication and evolution. *Science* **297**: 945-947.
- Maeda, N., Smithies, O. 1986. The evolution of multigene families: human haptoglobin genes. *Annu Rev Genet* **20**: 81-108.
- Markert, C.L., Moller F. 1959. Multiple Forms of Enzymes: Tissue, Ontogenetic, and Species Specific Patterns. *Proc Natl Acad Sci USA* **45**: 753-763.
- Malkova, A., Ivanov, E.L., Haber J.E. 1996. Double-strand break repair in the absence of RAD51 in yeast: a possible role for break-induced DNA replication. *Proc Natl Acad Sci USA* **93**: 7131-7136.
- Mazzarella, R., Schlessinger, D. 1998. Pathological consequences of sequence duplications in the human genome. *Genome Research* **8**: 1007-1021.
- McClintock B. 1956. Controlling elements and the gene. *Cold Spring Harbor Symposia on Quantitative Biology* **21**: 197-216.

- Messing, J., Crea, R., Seeburg, P.H. 1981. A system for shotgun DNA sequencing. *Nucleic Acids Res* **9**: 309-321.
- Moore, S.P., Rinckel, L.A., Garfinkel, D.J. 1998. A Ty1 integrase nuclear localization signal required for retrotransposition. *Mol Cell Biol* **18**:1105-1114.
- Moran, J.V., DeBerardinis, R.J., Kazanian, H.H. 1999. Exon shuffling by L1 retrotransposition. *Science* **283**: 1530-1544.
- Morrow D.M., Connelly C., Hieter P. 1997. "Break copy" duplication: a model for chromosome fragment formation in *Saccharomyces cerevisiae*. *Genetics* **147**: 371-382.
- Mighell, A.J., Smith, N.R., Robinson, P.A., Markham, A.F. 2000. Vertebrate pseudogenes. *FEBS Lett* **468**: 109-114.
- Mombaerts, P. 2001. The human repertoire of odorant receptor genes and pseudogenes. *Annu Rev Genomics Hum Genet* **2**: 493-510.
- Myung, K., Chen, C., Kolodner, R.D. 2001. Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature* **411**: 1073-1076.
- Nadeau, J.H., Taylor, B.A. 1984. Lengths of chromosomal segments conserved since divergence of man and mouse. *Proc Natl Acad Sci USA* **81**: 814-818.
- Needleman, S.B., Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* **48**: 443-453.
- Nickerson, E., Nelson, D.L. 1998. Molecular definition of pericentric inversion breakpoints occurring during the evolution of humans and chimpanzees. *Genomics* **50**: 368-372.
- Norrander, J., Kempe, T., Messing, J. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**: 101-106.
- Nowak, M.A., Boerlijst, M.C., Cooke, J., Smith, J.M. 1997. Evolution of genetic redundancy. *Nature* **388**: 167-171.
- Ohno, S. 1970. *Evolution by Gene Duplication*, Springer.
- Ohta, T. Simple model for treating evolution of multigene families. 1976. *Nature* **263**: 74-76.

- Ozier-Kalogeropoulos, O., Malpertuy, A., Boyer, J., Tekaia, F., Dujon, B. 1998. Random exploration of the *Kluyveromyces lactis* genome and comparison with that of *Saccharomyces cerevisiae*. *Nucleic Acids Res* **26**: 5511-5524.
- Paques, F., Haber J.E. 1999. Multiple pathways of recombination induced by double strand breaks in *Saccharomyces cerevisiae*. *Microbiol Mol Biol Rev* **63**: 349-404.
- Paquin, C.E., Dorsey, M., Crable, S., Sprinkel, K., Sondej, M., Williamson, V.M. 1992. A spontaneous chromosomal amplification of the *ADH2* gene in *Saccharomyces cerevisiae*. *Genetics* **130**: 263-71.
- Pearson, W.R., Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* **85**: 2444-2448.
- Perez-Ortin, J.E., Querol, A., Puig, S., Barrio, E. 2002. Molecular characterization of a chromosomal rearrangement involved in the adaptive evolution of yeast strains. *Genome Research* **12**: 1533-1539.
- Petes, T.D. 1980. Molecular genetics of yeast. *Annu Rev Biochem* **49**: 845-876.
- Piatigorsky, J., Wistow, G. 1991. The recruitment of crystallins: new functions precede gene duplication. *Science* **252**: 1078-1079.
- Pihan, G., Doxsey, S.J. 2003. Mutations and aneuploidy: co-conspirators in cancer? *Cancer Cell* **4**: 89-94.
- Pinkel, D., Seagraves, R., Sudar, D., Clark, S., Poole, I., Kowbel, D., Collins, C., Kuo, W.L., Chen, C., Zhai, Y., Dairkee, S.H., Ljung, B.M., Gray, J.W., Albertson, D.G. 1998. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat Genet* **20**: 207-211.
- Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., Brown, P.O. 1999. Genome-wide analysis of DNACopy-number changes using cDNAmicroarrays. *Nat Genet* **23**: 41-46.
- Potier, S., Winsor, B., Lacroute, F. 1982. Genetic selection for reciprocal translocation at chosen chromosomal sites in *Saccharomyces cerevisiae*. *Mol Cell Biol* **2**:1025-1032.

- Rachidi, N., Barre, P., Blondin, B. 1999. Multiple Ty-mediated chromosomal translocations lead to karyotype changes in a wine strain of *Saccharomyces cerevisiae*. *Mol Gen Genet* **261**: 841-850.
- Raghuraman, M.K., Winzeler, E.A., Collingwood, D., Hunt, S., Wodicka, L., Conway, A., Lockhart, D.J., Davis, R.W., Brewer, B.J., Fangman, W.L. 2001. Replication dynamics of the yeast genome. *Science* **294**:115-121.
- 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**: 767-773.
- Roelants, F. 1996. Rôle du transposon Ty1 dans la réactivation fonctionnelle d'un allèle mutant du gène *URA2* et dans les remaniements chromosomiques associés chez *Saccharomyces cerevisiae*. Thesis, pp 130-136. Université Louis Pasteur, Strasbourg I.
- 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.
- Rothstein, R., Helms, C., Rosenberg, N. 1987. Concerted deletions and inversions are caused by mitotic recombination between delta sequences in *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**:1198-1207.
- Rosenberg, H.F., Dyer, K.D. 1995. Eosinophil cationic protein and eosinophil-derived neurotoxin. Evolution of novel function in a primate ribonuclease gene family. *J Biol Chem* **270**: 302.
- Rouquier, S., Blancher, A., Giorgi, D. 2000. The olfactory receptor gene repertoire in primates and mouse: evidence for reduction of the functional fraction in primates. *Proc Natl Acad Sci USA* **97**: 2870-2874.
- Rubin GM, Yandell MD, Wortman JR, Gabor Miklos GL, Nelson CR, Hariharan IK, Fortini ME, Li PW, Apweiler R, Fleischmann W et al. 2000. Comparative genomics of the eukaryotes. *Science* **287**: 2204-2215.

Sambrook, J., Fritsch, E.F., Maniatis, T. Molecular cloning: a laboratory manual. 2ème édition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N. Y., 1989

Samonte, R.V., Eichler, E.E. 2002. Segmental duplications and the evolution of the primate genome. *Nat Rev Genet* **3**: 65-72.

Sanger, F., Nicklen, S., Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**: 5463-5467.

Sankoff, D. Gene and genome duplication. 2001. *Curr Opin Genet Dev* **11**: 681-684.

Schwob, E., Nasmyth, K. 1993. CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. *Genes Dev* **7**: 1160-1175.

Seoighe, C., Wolfe, K.H. Extent of genomic rearrangement after genome duplication in yeast. *Proc Natl Acad Sci USA* **95**: 4447-4452.

Sherman, F., Hicks, J. 1991. Micromanipulation and dissection of asci. *Methods Enzymol* **194**: 21-37.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., Repping, S., Pyntikova, T., Ali, J., Bieri, T. et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**: 825-837.

Smith, G.P. 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* **191**: 528-535.

Smith, M.M. 1987. Molecular evolution of the *Saccharomyces cerevisiae* histone gene loci. *J Mol Evol* **24**: 252-259.

Smith, T.F., Waterman, M.S., Fitch, W.M. 1981. Comparative biosequence metrics. *J Mol Evol* **18**: 38-46.

Souciet, J., Aigle, M., Artiguenave, F., Blandin, G., Bolotin-Fukuhara, M., Bon, E., Brottier, P., Casaregola, S., de Montigny, J., Dujon, B. et al. Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett* **487**: 3-12.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98**: 503-517.

- Stam, K., Heisterkamp, N., Grosveld, G., de Klein, A., Verma, R.S., Coleman, M., Dosik, H., Groffen, J. 1985. Evidence of a new chimeric bcr/c-abl mRNA in patients with chronic myelocytic leukemia and the Philadelphia chromosome. *N Engl J Med* **313**: 1429-1433.
- Stankiewicz, P., Lupski, J.R. 2002. Molecular-evolutionary mechanisms for genomic disorders. *Curr Opin Genet Dev* **12**: 312-319.
- Stephens, S.G. 1951. Possible significances of duplication in evolution. *Adv Genet* **4**: 247-265.
- Sukharev, S.I., Klenchin, V.A., Serov, S.M., Chernomordik, L.V., Chizmadzhev, Yu.A. 1992. Electroporation and electrophoretic DNAtransfer into cells. The effect of DNA interaction with electropores. *Biophys J* **63**: 1320-1327.
- Szabados, L., Kovacs, I., Oberschall, A., Abraham, E., Kerekes, I., Zsigmond, L., Nagy, R., Alvarado, M., Krasovskaja, I., Gal, M., Berente, A., Redei, G.P., Haim, A.B., Koncz, C. 2002. Distribution of 1000 sequenced T-DNA tags in the Arabidopsis genome. *Plant J* **32**: 233-242.
- Szostak, J.W., Wu, R. 1980. Unequal crossing over in the ribosomal DNA of *Saccharomyces cerevisiae*. *Nature* **284**: 426-430.
- Tekaia, F., Dujon, B. 1999. Pervasiveness of gene conservation and persistence of duplicates in cellular genomes. *J Mol Evol* **49**: 591-600.
- Tchénio, T., Ségal-Bendirdjian, E., Heidmann, T. 1993. Generation of processed pseudogenes in murine cells. *EMBO J* **12**: 1487-1497.
- The Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796-815.
- Tomb, J.F., White, O., Kerlavage, A.R., Clayton, R.A., Sutton, G.G., Fleischmann, R.D., Ketchum, K.A., Klenk, H.P., Gill, S., Dougherty, B.A. et al. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388**: 539-547.
- 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.

- Tuzun, E., Bailey, J.A., Eichler, E.E. 2004. Recent segmental duplications in the working draft assembly of the brown Norway rat. *Genome Research* **14**: 493-506.
- Vézinhel, F., Blondin, B., Hallet, J.N. 1990. Chromosomal DNA patterns and mitochondrial DNA polymorphism as tool for identification of oenological strains of *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **32**: 568-571.
- Wagner, A. 1998. The fate of duplicated genes: loss or new function? *Bioessays* **20**: 785-788.
- Weiner, A.M., Deininger, P.L., Efstratiadis, A. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information. *Annu Rev Biochem* **55**: 631-661.
- Welcker, A.J., de Montigny, J., Potier, S., Souciet, J.L. 2000. Involvement of very short DNAtandem repeats and the influence of the RAD52 gene on the occurrence of deletions in *Saccharomyces cerevisiae*. *Genetics* **156**: 549-557.
- Winzeler, E.A., Lee, B., McCusker, J.H., Davis, R.W. 1999. Whole genome genetic-typing in yeast using high-density oligonucleotide arrays. *Parasitology* **118**: 73-80.
- Winzeler, E.A., Shoemaker, D.D., Atromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H. et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901-906.
- Wolfe, K.H., Shields, D.C. 1997. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**: 708-713.
- Wong, S., Butler, G., Wolfe, K.H. 2002. Gene order evolution and paleopolyploidy in hemiascomycete yeasts. *Proc Natl Acad Sci USA* **99**: 9272-9276.
- Xiong, Y., Eickbush, T.H. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J* **9**: 3353-3362.
- Xu, H., Boeke, J.D. 1990. Localization of sequences required in cis for yeast Ty1 element transposition near the long terminal repeats: analysis of mini-Ty1 elements. *Mol Cell Biol* **10**: 2695-2702.

- Xu, H., Boeke, J.D. 1991. Inhibition of Ty1 transposition by mating pheromones in *Saccharomyces cerevisiae*. *Mol Cell Biol* **11**: 2736-2743.
- Yu, X., Gabriel, A. Reciprocal translocations in *Saccharomyces cerevisiae* formed by nonhomologous end joining. *Genetics* **166**: 741-751.
- Zhang, X., Firestein, S. 2002. The olfactory receptor gene superfamily of the mouse. *Nature Neurosci* **5**:124-133.
- Zhang, J., Rosenberg, H.F., Nei, M. 1998. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci USA* **95**: 3708-3713.
- Zhang, J. 2003. Evolution by gene duplication: an update. *Trends in Ecology and Evolution* **18**: 292-298.

La duplication des séquences d'ADN contribue à la plasticité des génomes et représente un des moteurs de l'évolution. Ces dernières années, le séquençage systématique de plusieurs organismes a montré l'importance quantitative des régions dupliquées au sein des génomes. Cependant, l'analyse de ces séquences n'a pas permis d'élucider les mécanismes impliqués dans la genèse de tels événements. Au laboratoire, un système expérimental basé sur l'utilisation de l'allèle mutant du gène *URA2* a permis de sélectionner positivement des duplications d'ADN chez *S. cerevisiae*. Afin de caractériser les duplications et de comprendre les mécanismes impliqués, une analyse moléculaire fine de 2 collections de révertants (haploïdes et diploïdes) porteurs d'une duplication a été réalisée.

L'analyse des révertants haploïdes a permis de mettre en évidence deux processus par lesquels une séquence d'ADN peut se dupliquer : la duplication génique et la duplication segmentale. Les duplications géniques sont issues d'un processus de rétroposition c'est-à-dire de la formation d'un ADNc suite à la transcription inverse de l'ARNm correspondant suivie de son intégration dans le génome. Les duplications segmentales sont des régions dupliquées en tandem direct de taille comprise entre 5 et 90 kb. Le mécanisme impliqué dans leur formation correspond probablement à une erreur lors de la réplication.

L'analyse des révertants diploïdes a permis de mettre en évidence deux autres moyens par lesquels une séquence d'ADN peut se dupliquer : l'aneuploïdie et la translocation non-réciproque. Ces deux types de remaniements conduisent à la formation d'un chromosome chimère obtenu probablement par un mécanisme de type BIR (break-induced replication).

Cette étude apporte une preuve expérimentale sur la pluralité des moyens et des mécanismes permettant à une même séquence de se dupliquer. De plus, elle souligne l'importance de la ploïdie dans la sélection de tels remaniements.

The duplication of DNA sequences contributes to genomic plasticity and is known to be one of the key factors responsible for evolution. Recent genomic sequence data provide substantial evidence for the abundance of duplicated genes in all organisms surveyed. However, the mechanisms underlying these events have not yet been completely elucidated.

In this study, using a positive selection screen based on a particular mutated allele of the *URA2* gene, duplication events were selected in the yeast *Saccharomyces cerevisiae*. In order to characterize and understand the mechanism implicated, a molecular analysis of two collections of revertants (haploid and diploid) was performed.

The analyze of the haploid revertants revealed two different ways in which duplicate gene can arise: single gene duplication and segmental duplication. Important features show that single gene duplication result from retroposition that means from the reverse transcription of the corresponding mRNA and the subsequent integration of the cDNA in the genome. The segmental duplication correspond to direct tandem duplicated region ranging from 5 to 90 kb in size. A replication error mechanism could explain the formation of such duplication events.

The analyze of the diploid revertants revealed two other ways to duplicate a chromosomal region: aneuploidy and non-reciprocal translocation. These two types of events chimeric generate the formation of a chromosome which is probably obtained by a BIR process.

The genomic rearrangements described constitute an experimental evidence that spontaneous duplications of a peculiar gene can occur in *S. cerevisiae* by a multiplicity of ways.

Discipline : biologie moléculaire et cellulaire

Mots-clés : *Saccharomyces cerevisiae*, remaniements chromosomiques, duplications, rétroposition, aneuploïdie

Laboratoire de Dynamique, Évolution et Expression de génomes de micro-organismes
FRE 2326 ULP / CNRS
28, rue Goethe
67083 Strasbourg Cedex