

Haute fréquence de conversions géniques entre les gènes codant pour la phosphatase du 2-déoxyglucose 6-phosphate chez trois espèces du genre *Saccharomyces*

Sara-Pier Piscopo et Guy Drouin

Résumé : Les conversions géniques sont des échanges non réciproques de séquences entre gènes. Elles sont relativement fréquentes chez *Saccharomyces cerevisiae*, mais peu d'études ont analysé le destin évolutif des conversions géniques ou leur impact fonctionnel chez les gènes de levures. Ici, nous analysons l'évolution et l'impact des conversions géniques entre les deux gènes codant pour la phosphatase du 2-déoxyglucose-6-phosphate chez *S. cerevisiae*, *Saccharomyces paradoxus* et *Saccharomyces mikatae*. Nos résultats démontrent que la dernière moitié de ces gènes est sujette à des conversions géniques chez ces trois espèces. La grande similarité et le plus grand pourcentage de nucléotides G et C dans les régions converties, ainsi que l'absence de longues régions de sites convertis communs adjacents, suggèrent que ces conversions géniques se produisent fréquemment et de façon indépendante chez ces trois espèces. La haute fréquence de ces conversions résulte probablement du fait qu'elles ont peu d'impact sur les séquences protéiques codées par ces gènes.

Mots-clés : *Saccharomyces cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, conversions géniques, phosphatase du 2-déoxyglucose 6-phosphate.

Abstract: Gene conversions are nonreciprocal sequence exchanges between genes. They are relatively common in *Saccharomyces cerevisiae*, but few studies have investigated the evolutionary fate of gene conversions or their functional impacts. Here, we analyze the evolution and impact of gene conversions between the two genes encoding 2-deoxyglucose-6-phosphate phosphatase in *S. cerevisiae*, *Saccharomyces paradoxus* and *Saccharomyces mikatae*. Our results demonstrate that the last half of these genes are subject to gene conversions among these three species. The greater similarity and the greater percentage of GC nucleotides in the converted regions, as well as the absence of long regions of adjacent common converted sites, suggest that these gene conversions are frequent and occur independently in all three species. The high frequency of these conversions probably result from the fact that they have little impact on the protein sequences encoded by these genes.

Key words: *Saccharomyces cerevisiae*, *Saccharomyces paradoxus*, *Saccharomyces mikatae*, gene conversions, 2-deoxyglucose-6-phosphate phosphatase.

Introduction

Les gènes dupliqués ont des séquences similaires qui sont souvent sujettes à des conversions géniques, c.-à-d., des échanges non réciproques de séquences. Ceci a pour conséquence de remplacer une partie de la séquence du gène accepteur par la partie correspondante du gène donneur. Ce phénomène a initialement été découvert chez les levures et a par la suite aussi été documenté chez une grande variété d'organismes, incluant les vertébrés (Slightom et al. 1980; Drouin 2002). La grande majorité du temps, ces conversions géniques ont peu d'impacts fonctionnels parce que celles qui ont un effet délétère sont éliminées par sélection naturelle (Noonan et al. 2004; Hiwatashi et al. 2011; Petronella et Drouin 2011, 2014). Par contre, chez les humains, plusieurs maladies génétiques sont néanmoins causées par des conversions géniques entre gènes dupliqués (Bischof et al. 2006; Chen et al. 2007).

Plusieurs groupes de recherche ont étudié les caractéristiques des conversions géniques entre gènes humains. La fréquence des conversions géniques entre gènes humains a été évaluée à 1%–13% dépendant des types de séquences étudiées et des méthodes utilisées (Jackson et al. 2005; Benovoy et Drouin 2009; McGrath et al. 2009; Ezawa et al. 2010). Chez les mammifères, il y a une corréla-

tion positive entre la fréquence et la longueur des conversions et la similarité des séquences (Waldman et Liskay 1988; Elliott et al. 1998; Lukacsovich and Waldman 1999; Jackson et al. 2005; Chen et al. 2007; Benovoy et Drouin 2009). De plus, les régions fréquemment converties deviennent riches en nucléotides G et C (Galtier et al. 2001; Birdsell 2002; Meunier et Duret 2004).

À l'inverse, peu d'études informatiques existent concernant les caractéristiques des conversions géniques chez les levures. À notre connaissance, l'étude de Drouin (2002) est la seule à avoir étudié l'ensemble des conversions géniques présentes dans le génome de *Saccharomyces cerevisiae*. Cette étude a démontré que les conversions se produisent à une fréquence moyenne de 7,8 %, que cette fréquence atteint 80 % pour les gènes adjacents et qu'il y a une corrélation positive entre la fréquence et la longueur des conversions et la similarité des séquences. Elle a aussi suggéré que l'excès de conversions dans la partie terminale des gènes non liés est probablement le résultat de conversions géniques avec des molécules d'ADN complémentaire incomplètes. De plus, une étude subséquente a démontré que les conversions géniques augmentent aussi le contenu en bases G et C des gènes de levures dupliqués (Benovoy et al. 2005). Ici, nous analysons l'évolution et l'impact des conversions géniques entre les deux gènes codant

Reçu le 17 April 2014. Accepté le 30 June 2014.

S.-P. Piscopo et G. Drouin. Département de biologie et Centre de recherche avancée en génomique environnementale, Université d'Ottawa, 30 Marie Curie, Ottawa, ON K1N 6N5, Canada.

Auteur correspondant : Guy Drouin (courriel : gdrouin@science.uottawa.ca).

Tableau 1. Gènes de la phosphatase du 2-déoxyglucose-6-phosphate utilisés.

Espèces	Gène	Identification
<i>Saccharomyces cerevisiae</i>	<i>DOG1</i>	YHR044C
	<i>DOG2</i>	YHR043C
<i>Saccharomyces paradoxus</i>	<i>DOG1</i>	MIT_Spar_c37_10226
	<i>DOG2</i>	MIT_Spar_c37_10228
<i>Saccharomyces mikatae</i>	<i>DOG1</i>	MIT_Smik_c317_9791
	<i>DOG2</i>	MIT_Smik_c317_9789
<i>Saccharomyces kudriavzevii</i>	<i>DOG2</i>	WashU_Skud_Contig1151.2
<i>Saccharomyces bayanus</i>	<i>DOG2</i>	WashU_Sbay_Contig572.25
<i>Saccharomyces castellii</i>	<i>DOG2</i>	WashU_Scas_Contig690.31
	<i>DOG2.2</i>	AACF01000031,1B

pour la phosphatase du 2-déoxyglucose-6-phosphate chez *S. cerevisiae*, *Saccharomyces paradoxus* et *Saccharomyces mikatae*.

Chez les levures du genre *Saccharomyces*, les gènes *DOG* codent pour la phosphatase de 2-déoxyglucose-6-phosphate. Le rôle de cette enzyme est d'empêcher l'accumulation de 2-déoxyglucose-6-phosphate dans les cellules et ainsi permettre aux cellules de croître en sa présence (Randez-Gil et al. 1995). Le génome de *S. cerevisiae* contient deux gènes *DOG*, les gènes *DOG1* (YHR044C) et *DOG2* (YHR043C). Ces deux gènes sont situés côte à côte sur le chromosome 8 à une distance de 524 paires de bases (pb). Ces deux gènes ont eu leur origine à partir d'un locus original unique (yeastgenome.org). Cette duplication s'est vraisemblablement produite dans l'ancêtre commun de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* parce que le génome de ces trois espèces contient deux gènes *DOG* alors que ceux de *S. kudriavzevii* et *S. bayanus*, leurs espèces sœurs, en contiennent un seul (tableau 1). Le fait que le génome de *S. castellii* contient aussi deux gènes *DOG* est vraisemblablement dû à une duplication indépendante (tableau 1). En fait, le haut degré de divergence entre les deux gènes *DOG* de *S. castellii* suggère que cette duplication est très ancienne (fig. 1). De plus, le fait que ces deux gènes sont situés sur des contigs différents suggère qu'ils ne sont pas côte à côte sur le même chromosome (tableau 1).

Chez *S. cerevisiae*, les protéines Dog1 et Dog2 ont toutes deux un pH optimal de 6. Par contre, Dog1 a une température optimale de 30 °C, un K_M de 17 mmol/L et un V_{max} de 29 nmol/min sur le 2-déoxyglucose-6-phosphate tandis que Dog2 a une température optimale de 40 °C, un K_M de 41 mmol/L et un V_{max} de 69 nmol/min sur le 2-déoxyglucose-6-phosphate (Randez-Gil et al. 1995). Dog2 est aussi capable d'agir sur le fructose mono-phosphate, mais Dog1 a peu d'effet sur ce sucre (Randez-Gil et al. 1995). De plus, l'expression du gène *DOG2* est stimulée par le stress oxydatif, le stress osmotique et l'absence de glucose (Tsujiyama et al. 2000). Le fait que les régions non traduites 5' de ces deux gènes ne partagent aucune similarité est peut être responsable du fait que la régulation de ces deux gènes diffère (Randez-Gil et al. 1995).

Matériels et méthodes

Séquences utilisées

Nous avons utilisé les gènes de phosphatase du 2-déoxyglucose-6-phosphate (*DOG*) de six espèces du genre *Saccharomyces* : *S. cerevisiae*, *S. paradoxus*, *S. mikatae*, *Saccharomyces kudriavzevii*, *Saccharomyces bayanus* et *Saccharomyces castellii* (récemment renommé *Naumovozyma castellii*). Notez que cet ordre correspond aux relations phylogénétiques de ces espèces (Rokas et al. 2003). *S. kudriavzevii* est donc le groupe sœur de l'ancêtre commun de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* alors que *S. castellii* est le groupe sœur des cinq autres espèces. Les séquences furent obtenues à partir du site Internet du *Saccharomyces Genome Database* (yeastgenome.org; Cherry et al. 1998). Les séquences utilisées sont énumérées dans le tableau 1. Elles furent obtenues soit directement à partir de ce site, soit à l'aide des programmes BLAST disponibles sur ce site (Altschul et al.

1997). Seulement les gènes dont les protéines avaient une valeur E réciproque plus petite que 10^{-20} furent utilisés.

Détection des conversions géniques

Les séquences furent alignées à l'aide du programme ClustalW inclus dans le logiciel BioEdit version 2.1.3.0 (Thompson et al. 1994; Hall 1999). Le programme GENECONV version 1.81 fut utilisé pour identifier les conversions géniques présentes dans chaque espèce ayant plus d'un gène *DOG* (Sawyer 1999). Ces analyses furent effectuées en utilisant l'option g2 pour permettre la présence d'un certain niveau de différences dans les régions converties.

Identité et phylogénies des séquences

Le logiciel BioEdit version 2.1.3.0 fut utilisé pour calculer les pourcentages de nucléotides identiques entre différentes séquences (Hall 1999). Les phylogénies de relation de voisin des séquences furent construites à l'aide du logiciel MEGA6 (Tamura et al. 2013). La phylogénie des séquences d'ADN fut construite en utilisant le modèle de vraisemblance composée maximale et une distribution uniforme des taux de substitutions des sites. La phylogénie des séquences protéiques fut construite en utilisant le modèle JTT et une distribution uniforme des taux de substitutions des sites.

Substitutions synonymes et nonsynonymes

Les nombres de substitutions synonymes par site synonyme (dS) et les nombres de substitutions nonsynonymes par site nonsynonyme (dN) furent calculés avec la méthode de Pamilo, Bianchi et Li à l'aide du logiciel MEGA 6 (Tamura et al. 2013).

Résultats et discussion

Une région convertie d'une longueur moyenne de 469 paires de base (pb) est présente entre les gènes *DOG1* et *DOG2* de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* (tableau 2). Cette région couvre la dernière moitié de ces gènes qui ont tous 741 pb de longueur, sauf le gène *DOG1* de *S. mikatae* qui a une longueur de 774 pb. Étant donné que cette région est commune aux trois paires de gènes, il serait parcimonieux de conclure que cette région a été convertie dans l'ancêtre commun de ces trois espèces. Par contre, l'analyse de cette région n'appuie pas cette hypothèse. Si cette région avait été héritée d'un ancêtre commun, chaque séquence aurait depuis dû acquérir des mutations indépendantes. Ce n'est pas le cas. En fait, la région convertie commune de ces six séquences, qui a une longueur de 411 pb, contient 78 sites variables. De ces 78 sites, 70 sont identiques entre les gènes *DOG* de chaque espèce (fig. 2). Le fait que 90 % des sites variables sont identiques entre les deux gènes de chaque espèce démontre que les mutations sont rapidement homogénéisées à l'intérieur de chaque espèce. De ces 70 sites, 24 sont des sites où *S. cerevisiae* et *S. paradoxus* ont les mêmes nucléotides, 20 sont des sites où *S. cerevisiae* et *S. mikatae* ont les mêmes nucléotides, 20 sont des sites où *S. paradoxus* et *S. mikatae* ont les mêmes nucléotides et 6 sont des sites où les deux gènes de chaque espèce ont un nucléotide différent (fig. 2). Le fait que *S. cerevisiae* et *S. paradoxus* n'aient pas plus de sites en commun que *S. cerevisiae* et *S. mikatae* (chi-carré par rapport à un nombre attendu de $22 = 0,36$, $p = 0,55$) suggère que les conversions géniques se produisent fréquemment à l'intérieur de chaque espèce. En effet, si elles avaient été héritées d'un ancêtre commun, les nucléotides identiques seraient plus abondants entre espèces ayant un ancêtre commun plus récent (*S. cerevisiae* et *S. paradoxus*) par rapport à des espèces ayant un ancêtre commun plus distant (*S. cerevisiae* et *S. mikatae*).

Ces conversions géniques ont eu pour effet d'augmenter la similarité des séquences nucléotidiques dans la région convertie commune. Dans cette région convertie commune, les gènes *DOG1* et *DOG2* de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* sont 98,3 %, 99,0 % et 99,8 % identiques, respectivement. Par contre, ils ne sont que 77,7 %, 83,3 % et 81,3 % identiques, respectivement, à l'extérieur de

Fig. 1. Phylogénies des séquences *DOG*. (a) Phylogénie basée sur les séquences des gènes (744 nucléotides alignés). (b) Phylogénie basée sur les séquences protéiques (258 acides aminés alignés). Les nombres aux nœuds représentent le pourcentage des valeurs de « bootstrap ». Les échelles représentent la proportion de sites différents.

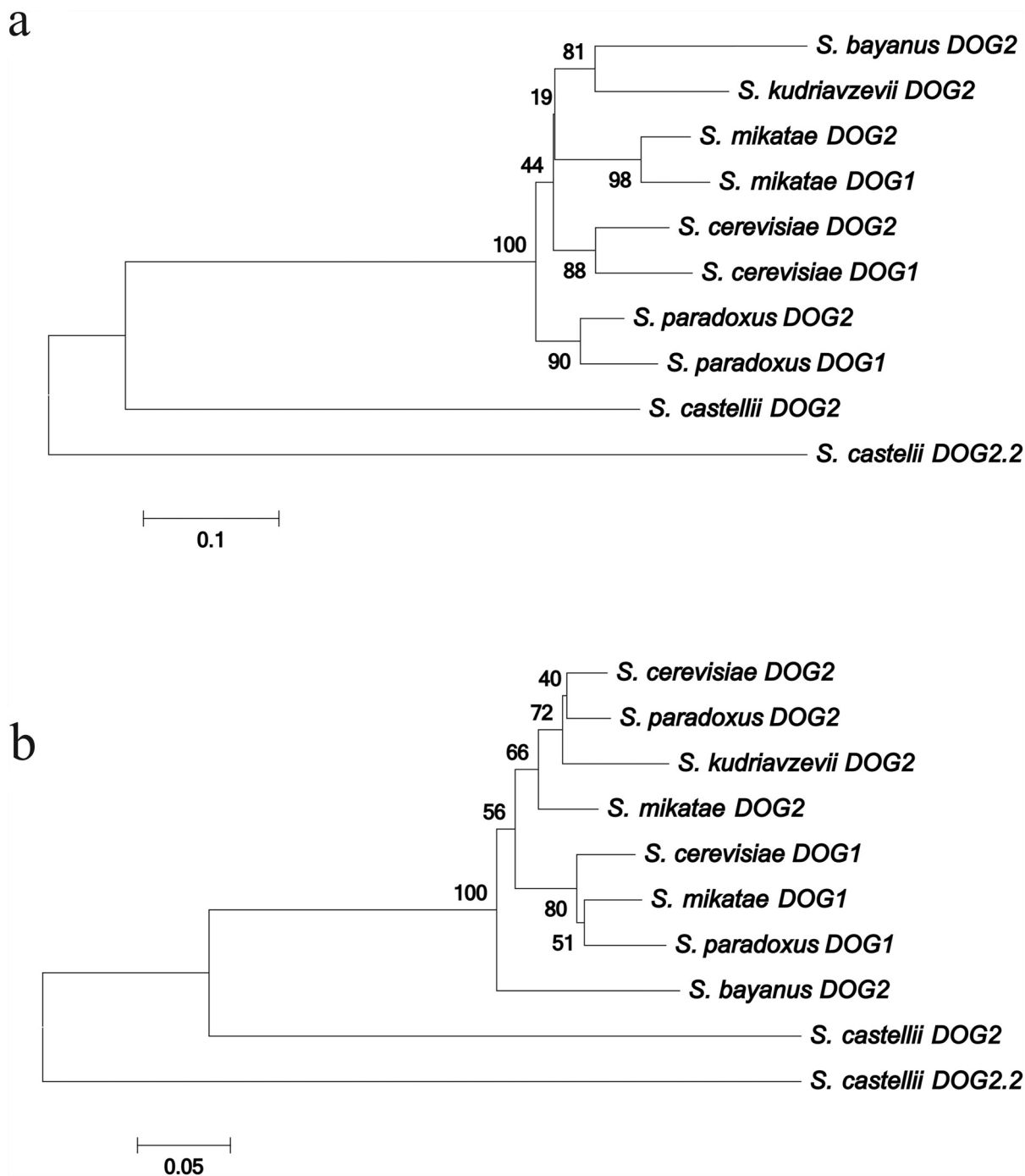


Tableau 2. Conversions géniques détectées chez les gènes de phosphatase du 2-déoxyglucose-6-phosphate.

Espèces	Longueur (pb)	Début (pb)	Fin (pb)	<i>p</i>
<i>Saccharomyces cerevisiae</i>	412	284	695	0,0000
<i>Saccharomyces paradoxus</i>	492	240	734	0,0000
<i>Saccharomyces mikatae</i>	507	235	741	0,0155

Note : La position des conversions géniques est indiquée par rapport au gène *DOG2* de *S. cerevisiae*. La longueur des conversions géniques n'inclut pas les insertions-délétions de l'alignement des séquences. Les valeurs *p* ont été calculées à partir de 10 000 permutations.

cette région. Cette différence est significative (test de *t*, $p = 0,004$). Le contenu en bases G et C dans la région convertie commune (moyenne \pm erreur type : $42,34 \% \pm 0,33\%$) est aussi significativement plus grand qu'à l'extérieur de cette région ($40,58 \% \pm 0,36\%$; test de *t*, $p = 0,004$).

Les conversions géniques ne semblent pas avoir eu d'effet significatif sur les séquences protéiques pour lesquelles elles codent. En fait, dans cette région, les acides aminés des protéines de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* sont en moyenne 90,4 % identiques à ceux de leur espèce sœur, *S. kudriavzevii*. De plus, 12 des 21 sites variables contiennent des acides aminés similaires selon la matrice de

Fig. 2. Séquences des régions converties des gènes *DOG*. S. c., *Saccharomyces cerevisiae*; S. p., *Saccharomyces paradoxus*; S. m., *Saccharomyces mikatae*. Les chiffres correspondent aux types de sites: 1, nucléotides identiques chez *S. cerevisiae* et *S. paradoxus*; 2, nucléotides identiques chez *S. cerevisiae* et *S. mikatae*; 3, nucléotides identiques chez *S. paradoxus* et *S. mikatae*; 4, nucléotides différents chez les trois espèces; •, sites variables non identiques à l'intérieur d'une ou deux des espèces.

		2	2	2	1	2	4	3	4	1	2	3	1
S. c.	DOG1	GTTATCGTTA	GATGTAGATA	CTGAGACTCA	AAAAAAGTTA	CCTGAAAGGA							
S. c.	DOG2	GTTATCGTTA	GATGTAGATA	CTGAGACTCA	AAAAAAGTTA	CCTGAAAGGA							
S. p.	DOG1	ATTATCCTTA	GATATAGACA	CCGACAGTGG	AAAGAAGTTG	CCTCAAAGGA							
S. p.	DOG2	ATTATCCTTA	GATATAGACA	CCGACAGTGG	AAAGAAGTTG	CCTCAAAGGA							
S. m.	DOG1	GTTATCGTTA	GATGTGGATA	CAGACAACAA	AAAGAAGTTA	CCTCAAAGAA							
S. m.	DOG2	GTTATCGTTA	GATGTGGATA	CAGACAACAA	AAAGAAGTTA	CCTCAAAGAA							
			3				4		1				
S. c.	DOG1	AATGGGCTAT	CGTTACCTCT	GGTTCTCCAT	ATTTGGCATT	TTCATGGTTC							
S. c.	DOG2	AATGGGCTAT	CGTTACCTCT	GGTTCTCCAT	ATTTGGCATT	TTCATGGTTC							
S. p.	DOG1	AATGGGCCAT	TGTTACCTCT	GGTTCTCCTT	ATTTGGCATT	TTCATGGTTC							
S. p.	DOG2	AATGGGCCAT	TGTTACCTCT	GGTTCTCCTT	ATTTGGCATT	TTCATGGTTC							
S. m.	DOG1	AATGGGCCAT	TGTTACCTCT	GGTTCTCCGT	ACTTGGCATT	TTCATGGTTC							
S. m.	DOG2	AATGGGCCAT	TGTTACCTCT	GGTTCTCCGT	ACTTGGCATT	TTCATGGTTC							
			2		13		3		3			2	
S. c.	DOG1	GAGACAATAT	TGAAAAATGT	TGGAAAGCCC	AAAGTTTTC	TTACTGGGTT							
S. c.	DOG2	GAGACAATAT	TGAAAAATGT	TGGAAAGCCC	AAAGTTTTC	TTACTGGATT							
S. p.	DOG1	GAGACAATAC	TGAAAAAAGT	TGGGAAGCCT	AAAGTTTTC	TTACCGGATT							
S. p.	DOG2	GAGACAATAC	TGAAAAAAGT	TGGGAAGCCT	AAAGTTTTC	TTACCGGATT							
S. m.	DOG1	GAGACAATAT	TGAGAACAGT	TGGGAAGCCT	AAAGTTTTC	TTACTGGTTT							
S. m.	DOG2	GAGACAATAT	TGAAAAACAGT	TGGGAAGCCT	AAAGTTTTC	TTACTGGTTT							
		3	3	1	1	2	1	4	3	2	3		
S. c.	DOG1	TGACGTGAAG	AACGGTAAGC	CTGATCCCGA	GGGTTATTC	AGAGCTCGTG							
S. c.	DOG2	TGACGTGAAG	AACGGTAAGC	CTGATCCCGA	GGGTTACTCA	AGAGCTCGTG							
S. p.	DOG1	CGATGTGAAG	AATGGTAAGC	CGGATCCTGA	AGGTTATGCA	AGAGCTCGTG							
S. p.	DOG2	CGATGTGAAG	AATGGTAAGC	CGGATCCTGA	AGGTTATGCA	AGAGCTCGTG							
S. m.	DOG1	CGATGTCAAA	AACGGCAAGC	CAGATCCTGA	GGGTTATGCA	AGAGCTCGTG							
S. m.	DOG2	CGATGTCAAA	AACGGCAAGC	CAGATCCTGA	GGGTTATGCA	AGAGCTCGTG							
			1				1	2	3	21	1	12	1
S. c.	DOG1	ATTTATTGCG	TCAAGATTG	CAATTAACCTG	GTAACAGGA	TCTGAAGTAT							
S. c.	DOG2	ATTTATTGCG	TCAAGATTG	CAATTAACCTG	GTAACAGGA	TCTGAAGTAT							
S. p.	DOG1	ATTTATTGCG	TCAAGATTG	CAACTAGCTG	ATAACAGGA	TTTGAATAC							
S. p.	DOG2	ATTTATTGCG	TCAAGATTG	CAACTAGCTG	ATAACAGGA	TTTGAATAC							
S. m.	DOG1	ATTTATTGCG	TCAAGATTG	AAATTAGCTG	GCAAAAAGGA	CCTGGAATAC							
S. m.	DOG2	ATTTATTGCG	TCAAGATTG	AAATTAGCTG	GCAAAAAGGA	CCTGGAATAC							
			2		2	1			1	2	2		
S. c.	DOG1	GTTGTCTTCG	AAGATGCACC	CGTGGGCATA	AAGCCGGCA	AAGCAATGGG							
S. c.	DOG2	GTTGTCTTCG	AAGATGCACC	CGTGGGCATA	AAGCCGGCA	AAGCAATGGG							
S. p.	DOG1	GTGGTCTTCG	AAGATGCACC	CGTGGGCATA	AAGCCGGTA	AATCAATGGG							
S. p.	DOG2	GTGGTCTTCG	AAGATGCACC	CGTGGGCATA	AAGCCGGTA	AGTCAATGGG							
S. m.	DOG1	GTTGTTTTCG	AGGATGCACC	TGTGGGCATA	AAGGCTGGCA	AAGCAATGGG							
S. m.	DOG2	GTTGTTTTCG	AGGATGCACC	TGTGGGCATA	AAGGCTGGCA	AAGCAATGGG							
		1	.	1	4	1		3	2	.	1	3	1
S. c.	DOG1	CGCCATTACT	GTGGGTATAA	CATCCTCGTA	TGACAAGAGC	GTTTTATTTG							
S. c.	DOG2	CGCAATTACT	GTGGGTATAA	CATCCTCGTA	TGATAAGAGC	GTTTTATTTG							
S. p.	DOG1	CGCCATTACC	GTGGGTATAA	CATCTTCCTA	TGATAAGTGT	GTTTTATTTG							
S. p.	DOG2	CGCCATTACC	GTGGGTATAA	CATCTTCCTA	TGATAAGAGT	GTTTTATTTG							
S. m.	DOG1	TGCTATAACA	GTGGGAATAA	CATCTTCGTA	TGACAAAAGT	GTTCTGTTCG							
S. m.	DOG2	TGCTATAACA	GTGGGAATAA	CATCTTCGTA	TGACAAAAGT	GTTCTGTTCG							
			3	.	.	2	.	3	2		1	2	3
S. c.	DOG1	ACGCAGGAGC	AGATTATGTG	GTCTGTGATT	TGACACAGGT	TTCCGTGGTT							
S. c.	DOG2	ACGCAGGAGC	AGATTATGTG	GTCTGTGATT	TGACACAGGT	TTCCGTGGTT							
S. p.	DOG1	ACGCTGGCGC	AGACTATGTG	GTTTGTGACT	TGACACAGGT	TTCCGTGATC							
S. p.	DOG2	ACGCTGGTGC	GGACTATGTG	GTTTGTGACT	TGACACAGGT	TTCCGTGATC							
S. m.	DOG1	ACGCTGGTGC	GGATTATGTG	GTTTGTGATT	TGACACAGGT	TTCTGTGGTC							
S. m.	DOG2	ACGCTGGTGC	GGATTATGTG	GTTTGTGATT	TGACACAGGT	TTCTGTGGTC							
			3										
S. c.	DOG1	AAGAACAATG	A										
S. c.	DOG2	AAGAACAATG	A										
S. p.	DOG1	AAGAACAATG	C										
S. p.	DOG2	AAGAACAATG	C										
S. m.	DOG1	AAGAACAATG	C										
S. m.	DOG2	AAGAACAATG	C										

Genome Downloaded from www.nrcresearchpress.com by Depository Services Program on 09/30/14 For personal use only.

Fig. 3. Conservation des séquences protéiques dans les régions converties des gènes *DOG*. *S. c.*, *Saccharomyces cerevisiae*; *S. p.*, *Saccharomyces paradoxus*; *S. m.*, *Saccharomyces mikatae*; *S. k.*, *Saccharomyces kudriavzevii*; *S. b.*, *Saccharomyces bayanus*. Les « - » indiquent des acides aminés identiques à la séquence Dog2 de *S. kudriavzevii*. Les « s » indiquent des acides aminés similaires selon la matrice de substitutions BLOSUM 62.

<i>S. k.</i>	DOG2	LSLDVDVEDE	KKLPKRKWAI	VTSGSPYLAF	SWFETILKKV	GMPKIFITGF
<i>S. c.</i>	DOG1	-----T-TQ	----E-----	-----	-----N-	-K--V-----
<i>S. c.</i>	DOG2	-----T-TQ	----E-----	-----	-----N-	-K--V-----
<i>S. p.</i>	DOG1	----I-TDSG	----Q-----	-----	-----	-K--V-----
<i>S. p.</i>	DOG2	----I-TDSG	----Q-----	-----	-----	-K--V-----
<i>S. m.</i>	DOG1	-----TDNK	----Q-----	-----	-----RT-	-K--V-----
<i>S. m.</i>	DOG2	-----TDNK	----Q-----	-----	-----T-	-K--V-----
		s s	s		s	s
<i>S. k.</i>	DOG2	DVKNGKPDPE	GYARARDLLR	QDLQLTDKGD	LKYVVFEDAP	VGIKAGKAMG
<i>S. c.</i>	DOG1	-----	--S-----	-----G-Q-	-----	-----
<i>S. c.</i>	DOG2	-----	--S-----	-----G-Q-	-----	-----
<i>S. p.</i>	DOG1	-----	-----	-----A-Q-	-----	-----S--
<i>S. p.</i>	DOG2	-----	-----	-----A-Q-	-----	-----S--
<i>S. m.</i>	DOG1	-----	-----C	---K-AG---	-E-----	-----
<i>S. m.</i>	DOG2	-----	-----C	---K-AG---	-E-----	-----
			s	s s s		s
<i>S. k.</i>	DOG2	AITVGITSSY	DKSVLFDAGA	DYVVCDLTQI	SVVKNN	
<i>S. c.</i>	DOG1	-----	-----	-----V	-----	
<i>S. c.</i>	DOG2	-----	-----	-----V	-----	
<i>S. p.</i>	DOG1	-----	--C-----	-----V	--I---	
<i>S. p.</i>	DOG2	-----	-----	-----V	--I---	
<i>S. m.</i>	DOG1	-----	-----	-----V	-----	
<i>S. m.</i>	DOG2	-----	-----	-----V	-----	
				s s		

substitutions BLOSUM62 (fig. 3). Il est intéressant de noter que le faible impact des conversions géniques sur les séquences protéiques se reflète aussi au niveau des arbres phylogénétiques. Alors que la phylogénie basée sur les séquences d'ADN groupe les séquences paraloues *DOG1* et *DOG2* de *S. cerevisiae*, *S. paradoxus* et *S. mikatae* ensemble (fig. 1a), la phylogénie basée sur les séquences protéiques groupe les séquences orthologues de ces trois espèces ensemble (fig. 1b). Donc, malgré un effet significatif au niveau des séquences d'ADN, cet effet n'est pas détectable au niveau des séquences protéiques. Par contre, ces conversions sont probablement responsables du fait que les valeurs de «bootstrap» sont plus élevées dans la phylogénie des séquences d'ADN que dans celle des séquences protéiques.

En conclusion, les gènes *DOG* des trois espèces du genre *Saccharomyces* qui ont hérité ces deux gènes de leur ancêtre commun partagent aussi une région terminale qui est souvent sujette à des conversions géniques (tableau 2). Cette conclusion est appuyée par le fait que cette région est toujours très similaire entre les deux gènes des trois espèces (identités de 98,3 %, 99,0 % et 99,8 % pour *S. cerevisiae*, *S. paradoxus* et *S. mikatae*, respectivement). Elle est aussi appuyée par le fait que les nucléotides identiques ne sont pas plus abondants entre espèces ayant un ancêtre commun plus récent qu'entre espèces ayant un ancêtre commun plus distant (fig. 2). Finalement, elle est aussi appuyée par le fait que cette région convertie commune est considérablement plus riche en

bases G et C, une caractéristique des régions sujettes à des recombinaisons fréquentes (Gerton et al. 2000; Galtier et al. 2001; Birdsell 2002; Meunier et Duret 2004; Benovoy et al. 2005). Le fait que ces conversions géniques soient fréquentes dans la région terminale de ces gènes reflète probablement le fait que cette région est interchangeable entre les deux protéines codées par ces gènes et que les quelques mutations présentes dans cette région n'ont pas d'impacts fonctionnels importants (fig. 3). Ceci suggère que ces conversions sont un autre exemple de conversions géniques qui ne sont pas éliminées par sélection naturelle parce qu'elles n'ont pas d'effet délétère (Noonan et al. 2004; Hiwatashi et al. 2011; Petronella et Drouin 2011, 2014; Zid et Drouin 2013). Au contraire, les régions initiales de ces gènes, composées de leurs 234 premiers nucléotides, semblent immunisées contre de telles conversions géniques parce qu'elles sont sujettes à de fortes pressions de sélection purificatrice (ratios dN/dS de 0,110 (0,143/1,300), 0,162 (0,122/0,752) et 0,131 (0,118/0,897) pour *S. cerevisiae*, *S. paradoxus* et *S. mikatae*, respectivement).

Remerciements

Cette recherche a été réalisée grâce à une bourse de premier cycle du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) à S.-P.P. et une subvention à la découverte du CRSNG à G.D.

Bibliographie

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. et Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402. doi:10.1093/nar/25.17.3389. PMID:9254694.
- Benovoy, D. et Drouin, G. 2009. Ectopic gene conversions in the human genome. *Genomics*, **93**: 27–32. doi:10.1016/j.ygeno.2008.09.007. PMID:18848875.
- Benovoy, D., Morris, R.T., Morin, A. et Drouin, G. 2005. Ectopic gene conversions increase the G + C content of duplicated yeast and *Arabidopsis* genes. *Mol. Biol. Evol.* **22**: 1865–1868. doi:10.1093/molbev/msi176. PMID:15917495.
- Birdsell, J.A. 2002. Integrating genomics, bioinformatics, and classical genetics to study the effects of recombination on genome evolution. *Mol. Biol. Evol.* **19**: 1181–1197. doi:10.1093/oxfordjournals.molbev.a004176. PMID:12082137.
- Bischof, J.M., Chiang, A.P., Scheetz, T.E., Stone, E.M., Casavant, T.L., Sheffield, V.C. et Braun, T.A. 2006. Genome-wide identification of pseudogenes capable of disease-causing gene conversion. *Hum. Mutat.* **27**: 545–552. doi:10.1002/humu.20335. PMID:16671097.
- Chen, J.-M., Cooper, D.N., Chuzhanova, N., Férec, C. et Patrinos, G.P. 2007. Gene conversion: mechanisms, evolution and human disease. *Nature Rev. Genet.* **8**: 762–775. doi:10.1038/nrg2193. PMID:17846636.
- Cherry, J.M., Adler, C., Ball, C., Chervitz, S.A., Dwight, S.S., Hester, E.T., et al. 1998. SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* **26**: 73–79. doi:10.1093/nar/26.1.73. PMID:9399804.
- Drouin, G. 2002. Characterization of the gene conversions between the multi-gene family members of the yeast genome. *J. Mol. Evol.* **55**: 14–23. doi:10.1007/s00239-001-0085-y. PMID:12165839.
- Elliott, B., Richardson, C., Winderbaum, J., Nickoloff, J.A. et Jasin, M. 1998. Gene conversion tracts from double-strand break repair in mammalian cells. *Mol. Cell. Biol.* **18**: 93–101. PMID:9418857.
- Ezawa, K., Ikeo, K., Gojobori, T. et Saitou, N. 2010. Evolutionary pattern of gene homogenization between primate-specific paralogs after human and macaque speciation using the 4-2-4 method. *Mol. Biol. Evol.* **27**: 2152–2171. doi:10.1093/molbev/msq109. PMID:20430861.
- Galtier, N., Piganeau, G., Mouchiroud, D. et Duret, L. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics*, **159**: 907–911. PMID:11693127.
- Gerton, J.L., DeRisi, J., Shroff, R., Lichten, M., Brown, P.O. et Petes, T.D. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* **97**: 11383–11390. doi:10.1073/pnas.97.21.11383. PMID:11027339.
- Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.* **41**: 95–98.
- Hiwatashi, T., Mikami, A., Katsumura, T., Suryobroto, B., Perwitasari-Farajallah, D., Malaijintond, S., et al. 2011. Gene conversion and purifying selection shape nucleotide variation in gibbon L/M opsin genes. *BMC Evol. Biol.* **11**: 312. doi:10.1186/1471-2148-11-312. PMID:22017819.
- Jackson, M.S., Oliver, K., Loveland, J., Humphray, S., Dunham, I., Rocchi, M., et al. 2005. Evidence for widespread reticulate evolution within human duplicons. *Am. J. Hum. Genet.* **77**: 824–840. doi:10.1086/497704. PMID:16252241.
- Lukacsovich, T. et Waldman, A.S. 1999. Suppression of intrachromosomal gene conversion in mammalian cells by small degrees of sequence divergence. *Genetics*, **151**: 1559–1568. PMID:10101177.
- McGrath, C.L., Casola, C. et Hahn, M.W. 2009. Minimal effect of ectopic gene conversion among recent duplicates in four mammalian genomes. *Genetics*, **182**: 615–622. doi:10.1534/genetics.109.101428. PMID:19307604.
- Meunier, J. et Duret, L. 2004. Recombination drives the evolution of GC-content in the human genome. *Mol. Biol. Evol.* **21**: 984–990. doi:10.1093/molbev/msh070. PMID:14963104.
- Noonan, J.P., Grimwood, J., Schmutz, J., Dickson, M. et Myers, R.M. 2004. Gene conversion and the evolution of protocadherin gene cluster diversity. *Genome Res.* **14**: 354–366. doi:10.1101/gr.2133704. PMID:14993203.
- Petronella, N. et Drouin, G. 2011. Gene conversions in the growth hormone gene family of primates: stronger homogenizing effects in the Hominidae lineage. *Genomics*, **98**: 173–181. doi:10.1016/j.ygeno.2011.06.001.
- Petronella, N. et Drouin, G. 2014. Purifying selection against gene conversions in the folate receptor genes of primates. *Genomics*, **103**: 40–47. doi:10.1016/j.ygeno.2013.10.004. PMID:24184359.
- Randez-Gil, F., Blasco, A., Prieto, J.A. et Sanz, P. 1995. *DOGRI* and *DOGRI2*: two genes from *Saccharomyces cerevisiae* that confer 2-deoxyglucose resistance when overexpressed. *Yeast*, **11**: 1233–1240. doi:10.1002/yea.320111303. PMID:8553694.
- Rokas, A., Williams, B.L., King, N. et Carroll, S. 2003. Genome-scale approach to resolving incongruence in molecular phylogenies. *Nature*, **425**: 798–804. doi:10.1038/nature02053. PMID:14574403.
- Sawyer, S.A. 1999. GENECONV: a computer package for the statistical detection of gene conversion. [Online.] Available from <http://www.math.wustl.edu/~sawyer> [accessed 3 June 2014].
- Slightom, J.L., Blechl, A.E. et Smithies, O. 1980. Human fetal γ - and Λ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell*, **21**: 627–638. doi:10.1016/0092-8674(80)90426-2. PMID:7438203.
- Tamura, K., Stecher, G., Peterson, D., Filipiński, A. et Kumar, S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**: 2725–2729. doi:10.1093/molbev/mst197. PMID:24132122.
- Thompson, J.D., Higgins, D.G. et Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific GAP penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680. doi:10.1093/nar/22.22.4673. PMID:7984417.
- Tsujimoto, Y., Izawa, S. et Inoue, Y. 2000. Cooperative regulation of *DOG2*, encoding 2-deoxyglucose-6-phosphate phosphatase, by Snf1 kinase and high-osmolarity glycerol-mitogen-activated protein kinase cascade in stress response of *Saccharomyces cerevisiae*. *J. Bacteriol.* **182**: 5121–5126. doi:10.1128/JB.182.18.5121-5126.2000. PMID:10960096.
- Waldman, A.S. et Liskay, R.M. 1988. Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Mol. Cell. Biol.* **8**: 5350–5357. PMID:2854196.
- Zid, M. et Drouin, G. 2013. Gene conversions are under purifying selection in the carcinoembryonic antigen immunoglobulin gene families of primates. *Genomics*, **102**: 301–309. doi:10.1016/j.ygeno.2013.07.003. PMID:23867109.