View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by univOAK

Nucleic Acids Research, 2004, Vol. 32, No. 7 **2069–2078**
DOI: 10.1093/nar/gkh529

Differential evolution of the *Saccharomyces cerevisiae DUP240* paralogs and implication of recombination in phylogeny

V. Leh-Louis, B. Wirth, L. Despons, S. Wain-Hobson[1], S. Potier and J. L. Souciet*

Laboratoire de Microbiologie et Génétique, FRE 2326 Université Louis Pasteur/CNRS, Institut de Botanique, F-67083 Strasbourg Cedex, France and [1]Unité de Rétrovirologie Moléculaire, Institut Pasteur, F-75724 Paris Cedex 15, France

Received November 18, 2003; Revised February 5, 2004; Accepted March 16, 2004 DDBJ/EMBL/GenBank accession nos AJ585532–A585755

ABSTRACT

Multigene families are observed in all genomes sequenced so far and are the reflection of key evolutionary mechanisms. The *DUP240* family, identified in *Saccharomyces cerevisiae* strain S288C, is composed of 10 paralogs: seven are organized as two tandem repeats and three are solo ORFs. To investigate the evolution of the three solo paralogs, YAR023c, YCR007c and YHL044w, we performed a comparative analysis between 15 *S.cerevisiae* strains. These three ORFs are present in all strains and the conservation of synteny indicates that they are not frequently involved in chromosomal reshaping, in contrast to the *DUP240* ORFs organized in tandem repeats. Our analysis of nucleotide and amino acid variations indicates that YAR023c and YHL044w fix mutations more easily than YCR007c, although they all belong to the same multigene family. This comparative analysis was also conducted with five arbitrarily chosen Ascomycetes-specific genes and five arbitrarily chosen common genes (genes that have a homolog in at least one non-Ascomycetes organism). Ascomycetes-specific genes appear to be diverging faster than common genes in the *S.cerevisiae* species, a situation that was previously described between different yeast species. Our results point to the strong contribution, during DNA sequence evolution, of allelic recombination besides nucleotide substitution.

INTRODUCTION

The longstanding view of Ohno (1) that gene duplication is one of the driving forces of evolution is frequently cited in numerous papers about genome structure, organization and evolution (2–6). Gene duplication has long been occasionally observed by classical genetics (7,8) and biochemical approaches (9,10). However, the systematic sequencing of genomes reveals that gene duplication is indeed a widespread feature of eukaryotic and prokaryotic species. The percentage of genes present in two or more copies per strain is generally reported as high: 29% for *Haemophilus influenzae* (11), 30% for *Saccharomyces cerevisiae* (12) and up to 67% for the grass *Arabidopsis thaliana* (13), to cite a few examples.

In the eukaryotic kingdom, the origin of duplicated gene copies is mostly described by three mechanisms that can act independently, successively or in combination: (i) whole genome duplication (14,15), (ii) reiterative segmental duplication (16–18), or (iii) single gene unit duplication (19). The classical mechanisms of chromosomal rearrangements such as deletions, translocations, inversions and gene fusions can break the synteny observed between duplicated chromosomes or chromosomal fragments. Consequently, knowing the genomic organization in one species is not sufficient to make conclusions about the general mechanisms involved in generating genome redundancy. On the other hand, comparative genomic studies of phylogenetically related species can be deeply informative. Thus, a study of a homogenous group of hemiascomycete yeasts ('Génolevures'; 20) clarified the precise number of genes in the sequenced strain S288C of the *S.cerevisiae* species (21). It further allowed the distinction of two gene sets: (i) Ascomycetes-specific genes which are as yet only observed in genomes of species from this phylum; and (ii) common genes that have a homolog in at least one non-Ascomycete organism (prokaryote and/or eukaryote). This inter-species analysis also revealed that Ascomycetes-specific genes appear to be diverging faster than common genes (22). Multigene families are found in both gene sets, those belonging to the common gene set being more often functionally characterized (22). Members of the numerous Ascomycetes-specific multigene families, the most likely participants in the speciation processes, are less often characterized and need to be analyzed in more detail.

Gene families can comprise up to 20 members in the *S.cerevisiae* strain S288C (21). With few exceptions, such as the *PAU* and *OSBP* families (23,24), the large families identified by sequencing programs remain uninvestigated. One

*To whom correspondence should be addressed. Tel: +33 3 90 24 18 17; Fax: +33 3 90 24 20 28; Email: souciet@gem.u-strasbg.fr

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

Nucleic Acids Research, Vol. 32 No. 7 © Oxford University Press 2004; all rights reserved

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

Table 1. *Saccharomyces cerevisiae* strains used in this study

Strain	Origin	Area and/or country	Reference/source
Laboratory strains			
S288C			
Σ1278b (ATCC 42800)			47
Natural isolates and industrial strains			
CLIB95 (L1425-4B)[a]	Wine	France	48
CLIB219 (CBS 5287)	Grape berries	Russia	48
CLIB382 (CBS 1782)	Beer	Japan	49
CLIB388 (ATCC 10615)	Beer	Japan	INRA[b], Grignon
CLIB410	Sake	Japan	INRA, Grignon
CLIB413	Fermented rice	China	INRA, Grignon
K1	Wine	France	INRA, Colmar (Oeno-France)
R12	Grape berries	France (Alsace)	INRA, Colmar
R13	Grape berries	France (Alsace)	INRA, Colmar
CLIB556 (TL213[c])	Cheese	France	INRA, Grignon
CLIB630 (TL229[c])	Cheese	France	INRA, Grignon
YIIc12	Wine	France (Sauternes)	Bordeaux, France[d]
YIIc17	Wine	France (Sauternes)	Bordeaux

[a]Collection de Levures d'Intérêt Biotechnologique, INRA Grignon.
[b]Institut National de la Recherche Agronomique, France.
[c]Alternative name used in this study to better distinguish cheese strains.
[d]Laboratoire de Biologie Cellulaire de la Levure, UPR CNRS 9026.

of them, the *DUP240* gene family, specific to the *Saccharomyces sensu stricto* group (25) and comprising 10 members in the S288C strain, was subject to a first round of functional analysis to gain insight into the role of the proteins that they encode (25). The specific chromosomal organization of the *DUP240* ORFs is of particular interest since ORFs YAR027w, YAR028w, YAR029w, YAR031w and YAR033w are arranged as tandem repeats on chromosome I and ORFs YGL051w and YGL053w are tandemly repeated on chromosome VII. This unusual gene organization suggests that these loci could be more frequently involved in large genomic rearrangements than other parts of the yeast genome. A previous analysis of these tandem loci in 15 strains of different biological and geographical origins revealed the remarkable extent of the polymorphisms at these loci (V. Leh-Louis, B. Wirth, S. Potier, J.-L. Souciet and L. Despons, manuscript submitted for publication). In particular, they are characterized by the presence of new *DUP240* paralogs and specific DNA motifs that may be the target of recombination events.

In the present work, we address the fate and evolutionary divergence of the three remaining *DUP240* ORFs, or solo ORFs, YAR023c, YCR007c and YHL044w, in *S.cerevisiae*. Using sequencing and mapping approaches on the 15 strains mentioned above, we show that synteny is conserved in all cases and that solo DUP240 ORFs do not display any polymorphisms in their organization, in contrast to the results observed previously with the tandem repeats. We also observe that the three *DUP240* paralogs are not subject to the same degree of nucleotide replacement and that some of them are diverging faster than average compared with other genes. This conclusion has been proposed on the basis of a comparative analysis performed with common and Ascomycetes-specific genes. Finally, we discuss the impact of diploidy and allelic homologous recombination on the observed nucleotide variability for the different gene sets that were analyzed.

MATERIALS AND METHODS

Strains and media

The *S.cerevisiae* strains used in this study are listed in Table 1. They are homothallic, with the exception of the laboratory strains S288C and Σ1278b, which are heterothallic and haploid. Cells were grown at 30°C on YPD medium (1% yeast extract, 2% bactopeptone, 2% dextrose and 2% agar). Sporulation of diploids required overnight growth on YPD at 30°C prior to transfer onto sporulation plates (1% potassium acetate and 2% agar).

Molecular biology methods

Yeast genomic DNA was purified according to Hoffman and Winston (26). Primer sequences used for PCR amplification and sequencing were chosen on the basis of the published genomic sequence of S288C (27). DNA fragments smaller than 4 kb in size were obtained by PCR amplification using *Taq* DNA polymerase from Q-BIOgene. Amplification of longer fragments was achieved with the Expand Long Template PCR system (Roche). PCR conditions were those described by the manufacturers. PCR products were purified through MicroSpin™ S-400 HR columns (Amersham Pharmacia Biotech) and sequenced on one strand. The sequencing chemistry used was AmpliTaq FS DNA polymerase and BIGDYE TM terminators (version 1). Sequence reactions were analyzed with an Applied Biosystems 373XL sequencer. When heterozygous sites were identified for diploid strains, the nucleotide sequences of both alleles were determined from genomic DNA of haploid cells obtained after sporulation, with the exception of strain CLIB410 for which no spores were obtained under our growth conditions. Sequence data have been deposited with the EMBL/GenBank Data Libraries under accession nos AJ585532–AJ585755.

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

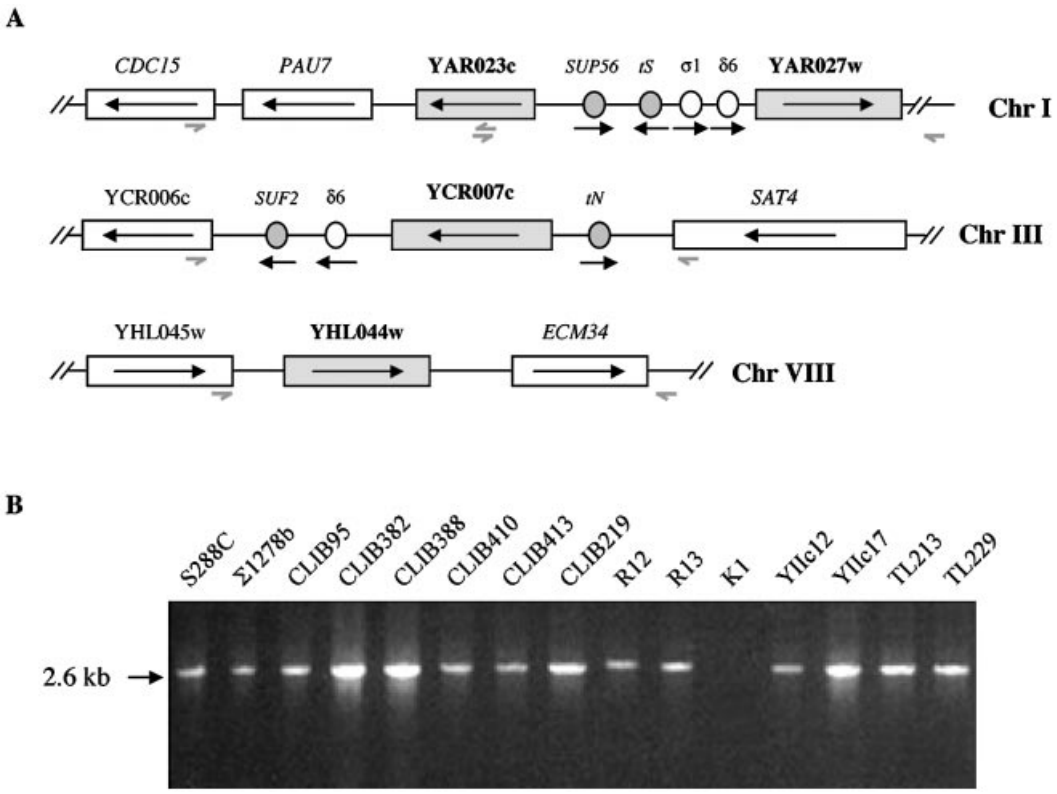

Figure 1. Localization of solo *DUP240* paralogs YAR023c, YCR007c and YHL044w. (**A**) Chromosomal maps of *DUP240* ORFs in S288C strain. The name and orientation of upstream and downstream ORFs (plain boxes) are indicated. Long terminal repeats and tRNA genes are represented by open and hatched circles, respectively. ORFs belonging to the *DUP240* multigene family are colored in gray. YAR027w corresponds to the first of the five *DUP240* ORFs tandemly repeated on chromosome I. Gray arrows indicate the position of primers used to amplify *DUP240* loci and to determine the synteny in all *S.cerevisiae* strains studied (the diagram is not drawn to scale). (**B**) Result of PCR amplification of the YHL044w locus using primers shown in (A). PCR fragments were separated by agarose electrophoresis and detected by ethidium bromide staining. The size of the PCR product obtained for S288C, used as reference, is indicated.

Sequence analysis

Nucleic acid sequences were aligned using the PILEUP (gap penalty = 5.00; gap extension penalty = 0.30) available in UWGCG package version 8.1 (28). The proportion of synonymous substitutions per potential synonymous sites (ds) and the proportion of non-synonymous substitutions per potential non-synonymous sites (dn) were calculated by the method of Nei and Gojobori (29) using the Synonymous Non-Synonymous Analysis Program (SNAP) available at http://hiv-web.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html. Sequence alignments used as input were performed with haplotypes to take into account identified heterozygous substitutions. Two ambiguous sequence positions remained for each allele of YHL044w in the case of CLIB410. Sequence alignments were also performed with diplotypes for phylogenetic analysis. The output format MSF was translated to PHYLIP format by the CLUSTAL X (1.8) program (30). Phylogenetic analyses were performed using the PHYLIP phylogeny inference package version 3.573 (31). Distance matrices were determined with Kimura 2-parameter and the Dnadist program. Unrooted phylogenetic trees were constructed using the neighbor-joining method and the Neighbor program. The final output file was generated by TreeViewPPC (1.6.6), and the stability of the individual branches was assessed using the bootstrap method (32). The multiple sequence alignment in NEXUS format generated for each gene was used as input for SplitsTree 2.4 (33), a network analysis program appropriate for describing homoplasies that may arise from recombination events. The stability of the network was tested by bootstrap analysis.

RESULTS

Chromosomal localization and conservation of synteny for YAR023c, YCR007c and YHL044w

We adopted a systematic PCR mapping approach to determine if the three paralogs YAR023c, YCR007c and YHL044w of the *DUP240* multigene family are present in the 14 tested strains (Table 1) with the same neighboring genes as seen in the S288C reference strain. The primers used corresponded to coding or non-coding sequences specifying the ORFs flanking *DUP240* paralogs (Fig. 1A). With the exception of YHL044w in strain K1 (Fig. 1B, lane 11), the size of the PCR products was identical or nearly identical to the expected size for S288C. The ORFs YAR023c, YCR007c and YHL044w were identified in the corresponding amplified DNA fragments without exception. Observed size variations of the PCR products correspond to significant deletion(s) or insertion(s) in intergenic areas and not in the coding sequence of the tested *DUP240* ORFs. For example, in strains R12 and R13, a

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

A

S288C	A	C	T	C	A	A	T	G	C	G	C	G	A	A	G	A	A	A	C	T	G	C	G	G	T	C
Σ1278b	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CLIB219	G	T	.	A	.	.	.	A	.	.	.	.	.	G	.	.	C	.	.	.	C	T	.	.	A	.
CLIB95	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
CLIB382	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
K1	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
R12	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	A/G	C	.	T	G	C	T	.	.	.	.
R13	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
YIIc12	G	.	T/C	.	A/G	.	.	A	.	G/C	C/T	.	A/G	G	G/T	.	C	.	C/T	T/G	C	T	.	.	T/A	.
YIIc17	G	.	T/C	.	A/G	.	.	A	.	G/C	C/T	.	A/G	G	G/T	.	C	.	C/T	T/G	C	T	.	.	T/A	.
TL213	G	.	C	.	A/G	.	.	A	.	G/C	.	.	.	G	.	.	C	A/G	.	.	C	T	.	.	A	.
TL229	G	.	C	.	.	.	.	A	.	C	.	.	.	G	.	.	C	.	.	.	C	T	.	Δ	A	.
CLIB410	G	.	C	.	.	.	.	A	G/C	C	.	.	.	G	.	.	C	.	.	.	C	T	G/A	.	A	.
CLIB413	G	.	C	.	.	.	.	A	.	.	.	A/G	.	G	.	.	C	.	.	.	C	T	.	.	A	C/T
CLIB388	G	.	C	.	.	C	C	A	.	C	.	.	G	G	.	.	C	.	.	.	C	T	.	.	A	.
	N	**I**	**F**	**F**	**I**	**L**	**L**	**S**	**S**	**R**	**L**	**A**	**T**	**M**	**R**	**I**	**Y**	**Q**	**Y**	**Y**	**G**	**N**	**L**	**G**	**V**	**T**
	9	**27**	**47**	**47**	**75**	**77**	**86**	**96**	**98**	**103**	**113**	**115**	**119**	**124**	**131**	**145**	**157**	**160**	**170**	**180**	**189**	**194**	**222**	**224**	**233**	**234**
	D		**L**	**L**	**V**	**F**	**P**	**N**	*	**T**				**V**		**M**	**S**	**R**		**D**	**A**				**D**	**I**

B

S288C	A	C	T	C	A	A	T	G	C	G	C	G	A	A	G	A	A	A	C	T	G	C	G	G	T	C
R13	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
YIIc12 all1	G	.	.	.	G	.	.	A	.	.	T	.	.	G	T	.	C	.	T	G	C	T	.	.	.	.
YIIc12	G	.	T/C	.	A/G	.	.	A	.	G/C	C/T	.	A/G	G	G/T	.	C	.	C/T	T/G	C	T	.	.	T/A	.
YIIc12 all2	G	.	C	.	.	.	.	A	.	C	.	.	G	G	.	.	C	.	.	.	C	T	.	.	A	.
CLIB388	G	.	C	.	.	C	C	A	.	C	.	.	G	G	.	.	C	.	.	.	C	T	.	.	A	.

Figure 2. Description of variable positions within YHL044w for the 15 yeast strains tested. Only differences with respect to the sequence of the reference strain S288C are shown. (**A**) The nature of the corresponding amino acid in S288C, its position in the peptide sequence and the new amino acid if the substitution is non-synonymous are described in bold below each variable position. Two nucleotides separated by a slash indicate that the corresponding strain is heterozygous at this position. (**B**) Both allele sequences of YHL044w from YIIc12 strain (all1 and all2) are compared with the corresponding sequences of the homozygous R13 and CLIB388 strains.

duplication of 76 and 46 nt, respectively, exists between YHL044w and *ECM34* (Fig. 1B, lanes 9 and 10). Strain K1 was subject to further experiments to explain the lack of amplification for the YHL044w locus. It was shown by PCR mapping and Southern analysis that YHL044w is still present in the K1 strain. However, if this ORF is still linked to YHL045w on chromosome VIII, the *ECM34* gene is now located on chromosome XII (data not shown). In conclusion, ORFs YAR023c, YCR007c and YHL044w identified in the S288C strain are present in the 14 other strains studied and the gene synteny of *DUP240* paralogs with flanking pairs of genes is systematically recovered with the exception of the YHL044w locus for strain K1.

Identification of nucleotide variation in the coding sequence of solo *DUP240* ORFs

The coding sequences of YAR023c, YCR007c and YHL044w of the 14 strains were compared with the corresponding sequences of the S288C strain. No deletion, insertion or nonsense substitutions were detected for YAR023c and YCR007c. Two events leading to changes in the length of the coding sequence were observed in YHL044w, one in strain TL229 and one in strain CLIB410. A single nucleotide deletion at position 224 in the peptide sequence of YHL044w in strain TL229 induces a frameshift that leads to a C-terminal extension of 20 extra amino acids (Fig. 2A). This deletion was identified on both chromosomes VIII of the diploid strain TL229. In strain CLIB410, a heterozygous situation is found, with a nonsense codon at position 98 in YHL044w on one of the two chromosomes VIII (Fig. 2A). All other observed mutations in YHL044w correspond to single nucleotide replacements. Therefore, nucleotide changes in the coding sequences of ORFs YAR023c, YCR007c and YHL044w are mainly nucleotide substitutions, whereas changes in flanking intergenic regions are frequently sequence deletion or duplication.

Differential evolution between YAR023c, YCR007c and YHL044w

Our first goal was to assess whether the frequency and nature of observed nucleotide replacements were comparable among the three paralogs YAR023c, YCR007c and YHL044w. Secondly, we analyzed the corresponding values with regard to what could be expected for genes belonging to the Ascomycetes-specific gene class (22), since the *DUP240* ORFs have not so far been identified in genomes of other classes of fungi and are probably restricted to the *Saccharomyces sensu stricto* group (25). Nucleotide divergence among *S.cerevisiae* strains is still poorly documented and could be subject to considerable variation depending on the origin and history of a given strain. We therefore decided to determine, for these 15 strains, the nucleotide polymorphism in the coding sequences of genes belonging to the common and to the Ascomycetes-specific gene sets. Five Ascomycetes-specific genes and five common genes were selected arbitrarily and submitted to sequencing after PCR amplification. The 10 genes sequenced in the 14 strains, and in strain S288C as a control, are listed in Table 2. Criteria used for selecting these genes were: (i) each of them is present in only one copy in the S288C genome; (ii) they share a coding

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

Table 2. Sequenced genes and ORFs

Classification	ORF	Gene	Size (bp)	Biological process[a]
Saccharomyces	YAR023c[b]		540	Unknown
sensu stricto specific	YCR007c[b]		720	Unknown
	YHL044w[b]		708	Unknown
Ascomycetes specific	YKL096w	*CWP1*	735	Cell wall organization and biogenesis
	YBR040w	*FIG1*	897	Mating
	YEL009c	*GCN4*	846	Amino acid biosynthesis
	YER149c	*PEA2*	1263	Cytoskeleton organization during mating
	YFL026w	*STE2*	1296	Pheromone response
Common	YFL045c	*SEC53*	765	Phosphomannomutase, essential gene
	YOR202w	*HIS3*	663	Histidine biosynthesis
	YDR392w	*SPT3*	1014	Chromatin modification, histone acetylation
	YDR007w	*TRP1*	675	Tryptophan biosynthesis
	YKL216w	*URA1*	945	Pyrimidine base biosynthesis

[a]From *Saccharomyces* Genome Database.
[b]Belongs to the *DUP240* multigene family.

sequence close to 240 codons as observed for the *DUP240* ORFs; (iii) the 10 genes are distributed on different chromosomes and at various positions on these chromosomes.

For each gene and *DUP240* ORF, we first calculated the ratio between synonymous substitutions per potential synonymous sites and non-synonymous substitutions per potential non-synonymous sites (ds/dn) using the method described by Nei and Gojobori (29). As shown in Table 3, none of the ds/dn values obtained for the selected genes and for the *DUP240* paralogs was smaller than 1, indicating that they are subject to negative selection for mutation. The smallest ds/dn values were obtained for *DUP240* paralogs YAR023c and YHL044w, suggesting that they could evolve more rapidly than other genes, including those belonging to the Ascomycetes-specific gene set. In contrast, the value of ds/dn for YCR007c was twice that obtained for YAR023c and YHL044w and comparable to that of reference genes (common and Ascomycetes-specific genes). Thus, the three solo *DUP240* ORFs seem to evolve with different evolutionary constraints even though they are part of the same multigene family. For YAR023c and YHL044w, and in contrast to the situation observed for all other genes investigated, the selection pressure to preserve protein sequence seems less stringent. None of the values obtained for genes belonging to either the common or the Ascomycetes-specific gene set allowed us to distinguish between these two gene classes, in contrast to what was previously demonstrated for different species of the hemiascomycetous class (22). Furthermore, ds/dn values for these two gene sets do not seem to reflect the results from sequence alignments observed at the nucleotide and amino acid levels, as illustrated by *SEC53* and *PEA2* genes. *SEC53* is an essential conserved gene belonging to the common gene set. Sequence alignment revealed seven variable sites for 765 nt, only two of them being heterozygous and non-synonymous. In contrast, there are 32 variable positions for the Ascomycetes-specific gene *PEA2* (1263 nt), half of them being non-synonymous substitutions (15 positions, 14 of them homozygous). Evolutionary constraints for *SEC53* therefore seem more stringent than for *PEA2* (although potential substitution sites were not taken into account as in the Nei and Gojobori method). Strikingly, the values of ds/dn indicate the opposite trend. This discrepancy is probably due to the fact that very few mutations affect the sequences of the investigated genes. Indeed, the nucleotide variability is analyzed at the intra-species level and we observe that the percentage of nucleotide identity between two strains is always higher than 98%. Therefore, we focused on the variable nucleotide positions (VP) to analyze our highly similar sequences. The VP value corresponds to the sum of all positions where there is a nucleotide substitution for at least one of the 15 strains (14 studied strains and strain S288C) and is normalized for a gene size of 100 nt. Its calculation is restricted to the condition that there are not two different mutations at the same variable position (a situation never

Table 3. Fixed mutations and effects on protein sequence

ORF/gene	ds/dn[a]	Variable positions (%)[b]
DUP240 multigene family		
YAR023c	2.6	3.0
YCR007c	6.0	1.1
YHL044w	2.6	3.5
Ascomycetes specific		
CWP1	8.5	1.6
FIG1	8.6	1.8
GCN4	6.3	1.3
PEA2	8.4	2.5
STE2	13.4	1.9
Common		
SEC53	4.6	0.9
HIS3	7.1	1.4
SPT3	9.5	1.7
TRP1	3.8	0.9
URA1	28.0	1.7

[a]Ratio between synonymous substitutions per potential synonymous sites and non-synonymous substitutions per potential non-synonymous sites calculated using the SNAP program (http://hiv-web.lanl.gov/content/hiv-db/SNAP/WEBSNAP/SNAP.html).
[b]Correspond to the sum of all positions in the sequence where there is a nucleotide substitution in one or both alleles of a gene for at least one of the 15 yeast strains (14 studied strains and S288C). Deletions or insertions were not considered. The indicated value is normalized for a theoretical gene size of 100 nt and was obtained by (number of mutated positions × 100 / size of the gene), allowing a direct comparison of all studied genes irrespective of their size.

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

encountered in our analysis). As an example, all variable positions identified in YHL044w are shown in Figure 2A. Analysis of obtained VP values (Table 3) confirms that ORFs YAR023c and YHL044w fix mutations more easily than other genes since their VP values (3.0 and 3.5, respectively) are twice those of common genes and higher than that of Ascomycetes-specific genes. Furthermore, this analysis allows us to distinguish Ascomycetes-specific genes from common genes.

Implication of different evolutionary constraints and recombination events for phylogenetic relationships

We have demonstrated that the three gene sets, common, Ascomycetes-specific and *DUP240*, are subject to different pressures for nucleotide replacement. Therefore, it is interesting to evaluate the impact of this factor on phylogenetic analyses of these data. To this end, all coding sequences of genes that belong to the same set were fused to generate a unique sequence per strain. Genes were fused following an arbitrary but identical order for all the strains. Subsequent sequence comparisons were performed using PILEUP, CLUSTAL X and PHYLIP. The same approach was also followed for the *DUP240* set since the number of nucleotide replacements observed for YCR007c and the two other *DUP240* paralogs (YAR023c and YHL044w) shows clear differences. In this case, identical dendrograms were obtained using a reconstruction either with YAR023c, YCR007c and YHL044w, or with YAR023c and YHL044w only (data not shown). Therefore, differences in nucleotide replacements within the *DUP240* family have only minor effects on the determination of phylogenetic relationships among the *S.cerevisiae* strains that were investigated. On the other hand, the three gene-class specific dendrograms all differ (Fig. 3). Indeed, comparison between dendrograms obtained for the *DUP240* ORFs (Fig. 3A) and the common gene set (Fig. 3B) shows that strains CLIB382 and CLIB388 are closely related to each other in the phylogeny established for common genes, whereas they appear to have a different ancestor in the *DUP240* dendrogram (Fig. 3). This type of inconsistency is also observed for YIIc12 and YIIc17 strains. Nevertheless, phylogenetic relationships between strains are otherwise identical in both trees. We note in particular the close relationships between CLIB410, CLIB413 and CLIB219, between the two laboratory strains S288C and Σ1278b and between R12, R13, K1 and CLIB95, which presumably reflect common origins (Table 1). Indeed, R12, R13 and CLIB95 have been isolated from grapes or wine from eastern France, and CLIB410 and CLIB413 from fermented rice. Some inconsistencies are also observed in the case of the dendrogram generated for Ascomycetes-specific genes (Fig. 3C). For example, the two strains CLIB410 and CLIB413 are closer to the strain Σ1278b than to CLIB219 (Fig. 3A and B). In conclusion, comparative analysis of dendrograms generated per class of genes indicates that the phylogenetic relationships between strains can be influenced by differential selective pressure.

However, in our analysis we cannot exclude that the inconsistencies observed between trees are not due to another phenomena, combined or not with substitution events. Notably, most of the yeast strains tested are diploid, which constitutes a propitious situation for allelic homologous recombination events. The effect of recombination would be to generate discrepancies between strictly bifurcating evolutionary trees (34,35) since establishing relationships between strains requires networks of sequences (represented by a polygon) as described in Figure 4. To determine if recombination events are likely to have taken place in the studied genes, the sequence alignments obtained for each gene were subject to a SplitsTree decomposition, a method which depicts all the shortest pathways linking sequences, including those that produce an interconnected network (33). As shown in Figure 4 for YHL044w (A), *URA1* (B) and *PEA2* (C), some relationships between strains are indeed best described by polygons, suggesting that recombination events between different alleles could explain part of the observed nucleotide variation. Similar results were obtained for YAR023c and *FIG1* (data not shown). *URA1*, YAR023c, YHL044w, *PEA2* and *FIG1* are located on chromosomes XI, I, VIII, V and II, respectively, and belong either to the common genes, the Ascomycetes-specific genes or the *DUP240* ORFs. Thus, recombination events appear to be independent of the chromosomal position and affect all three gene sets.

Two strains, YIIc12 and YIIc17, exhibit an unusual level of heterozygosity at the variable positions identified. Numerous heterozygous positions have been found for six genes in YIIc12 (YHL044w, *TRP1*, *PEA2*, *SPT3*, *FIG1* and *GCN4*) and five genes in YIIc17 (YHL044w, *TRP1*, *HIS3*, *FIG1* and *GCN4*), located on different chromosomes and belonging to all three classes. In the case of YHL044w in strains YIIc12 and YIIc17, nine of the 15 variable positions identified are in a heterozygous state (Fig. 2A and B). The sequence of each allele was therefore determined after induction of sporulation of diploids. For both strains, tetrad analysis indicated that one allele is strictly identical to the one present in R13, the second being very similar to the one of CLIB388 with only two non-equivalent positions (Fig. 2B). More generally, whenever heterozygous positions were detected, the sequence determined for each allele was identical or very close to the one present in a homozygous state in another strain studied. The probability of generating such a situation of homoplasy by independent mutations is very low. They are likely the result of mating of two haploid cells that display different allele combinations. Again, the more adequate way for describing the phylogenetic relationships among the *S.cerevisiae* tested strains is the network diagram. Taken together, our phylogenetic analyses clearly demonstrate that nucleotide variability arises from two distinct mechanisms involved in genome dynamics and evolution: single nucleotide substitution, and mating and subsequent allelic recombination events.

DISCUSSION

Systematic sequencing of genomes has revealed the importance of gene duplication among the eukaryotic kingdom (3,36). Gene amplification generating paralogous gene copies appears to be one of the driving forces for genome evolution and for enhancement of functional diversity. The homogeneous group of Hemiascomycetes yeast species, with their compact genomes, constitutes a good model for deciphering, by comparative genomics, the molecular mechanisms underlying duplication events (20). The availability of a large variety of well characterized strains of species belonging to

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

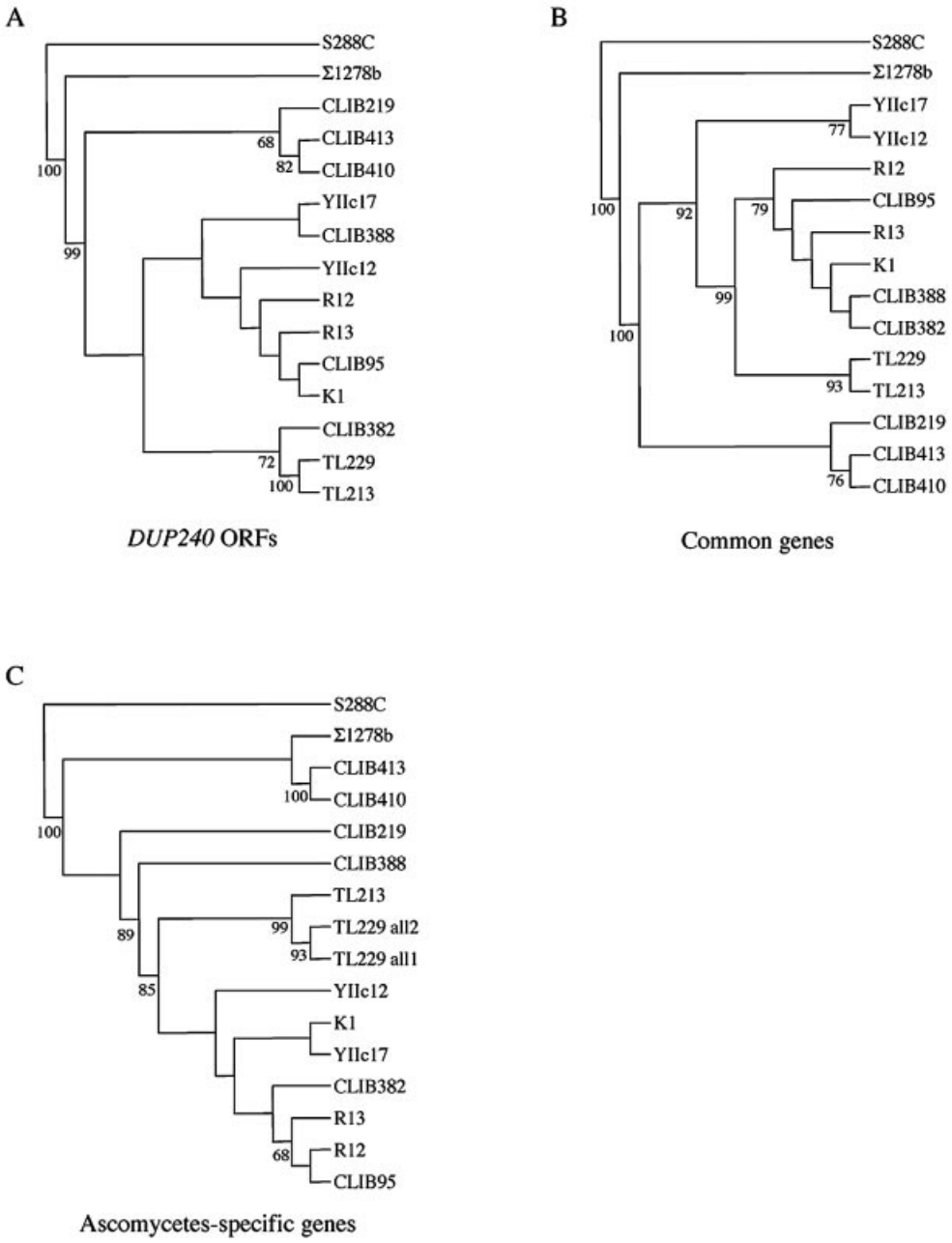

Figure 3. Phylogenetic relationships between strains obtained for solo *DUP240* ORFs (**A**), the common gene set (**B**) and the Ascomycetes-specific gene set (**C**), using the PHYLIP phylogeny inference package. Phylogenetic reconstructions (unrooted trees) were performed with diplotypes except for three strains: S288C and Σ1278b that are haploid and TL229 in the case of the Ascomycetes-specific gene set, where sequences of both alleles were used due to a 15 nt insertion in one allele of the CWP1 gene. This does not affect phylogenetic relationships between strains since TL229 all1 and all2 are linked together. Bootstrap values for 100 replicates are indicated.

this group also offers the opportunity to estimate the intra-species evolution of genomes. In the present study, we addressed the intra-species variability of genes and more specifically the evolution of duplicated copies by a comparative analysis of solo *DUP240* ORFs in 15 *S.cerevisiae* strains.

Our investigation shows that the three *DUP240* paralogs YAR023c, YCR007c and YHL044w are present in all strains tested. With only one exception (strain K1), the solo copies of the *DUP240* family were identified at the same loci as those described in the reference strain S288C and are flanked by the same neighboring genes identical in directions of transcription. Conservation of synteny indicates that the corresponding loci are not frequently involved in chromosomal reshaping events. These results contrast with the previous analysis performed with the *DUP240* tandem repeats and conducted with the same 15 *S.cerevisiae* strains (V. Leh-Louis, B. Wirth, S. Potier, J.-L. Souciet and L. Despons, manuscript submitted for publication). This latter study clearly showed that the number of paralogs per strain is subject to important variations, with the appearance of new paralogs and the

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

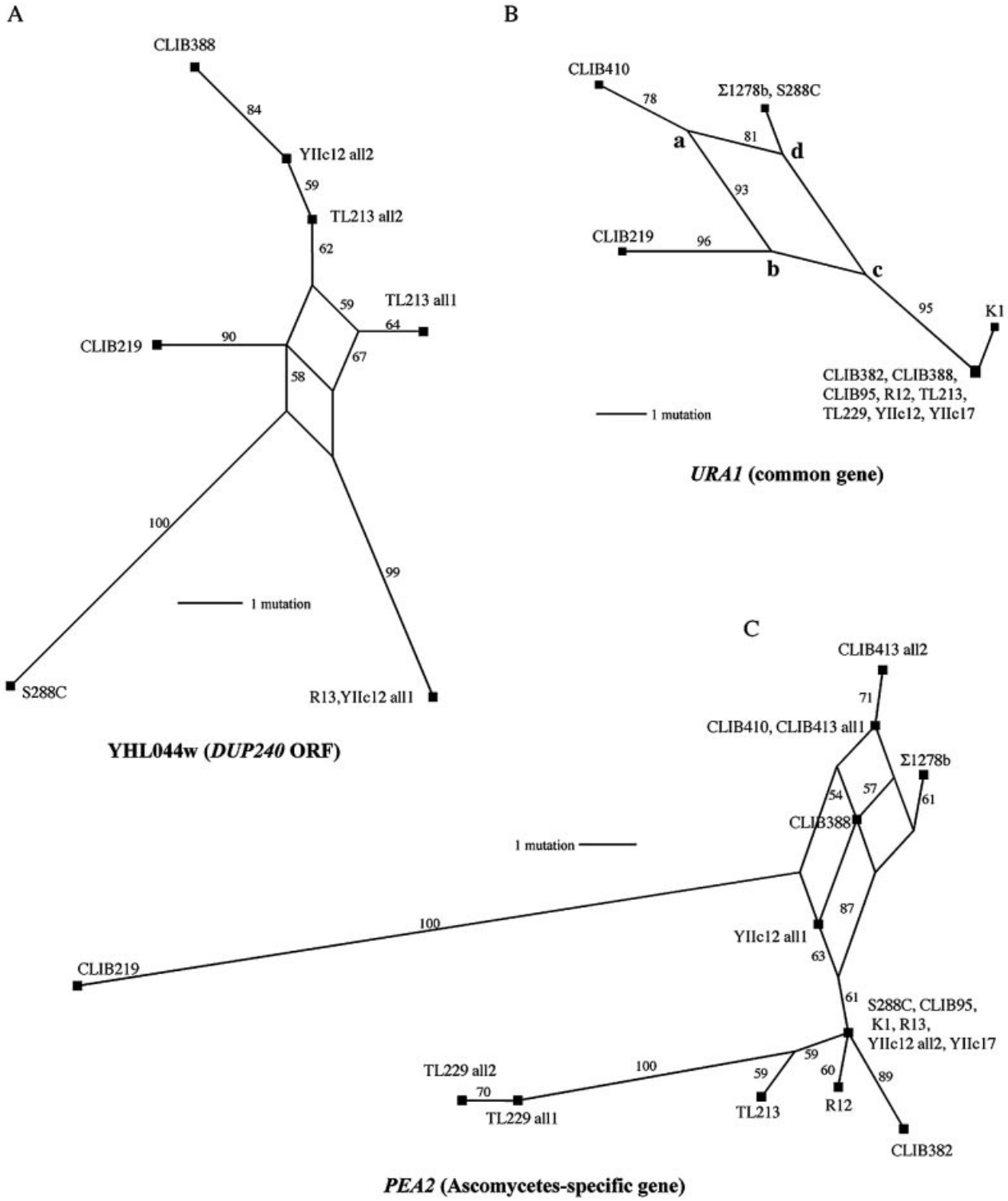

Figure 4. Examples of SplitsTree decomposition, obtained for YHL044w (**A**), *URA1* (**B**) and *PEA2* (**C**) that belong to the *DUP240*, common and Ascomycetes-specific gene sets, respectively. All networks generated have a fit of 100 (34). Bootstrap values for 100 replicates are indicated. In each case, polygon(s) are obtained suggesting that some of the nucleotide variations are not only due to point mutations but also to recombination events. For instance, in the case of *URA1* (B), a, b, c and d, each located at one apex of the polygon, represent the gene sequence of the ancestor strains. Parallel edges of the polygon indicate homoplasies, i.e. identical mutations present in different genomes. Thus, the same two nucleotide substitutions separate a from b and d from c. Similarly, one identical mutation differentiates a from d and b from c. Therefore, three point mutations separate a from c. If we suppose that c stems from a, three paths are possible to generate c: a → b → c or a → d → c, both involving three substitution events, or a → b and a → d followed by a recombination event between b and d. This latter situation seems the most parsimonious solution, even if it cannot be excluded that the resulting identical mutations arise from independent substitution events in the different genomes.

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

absence of others. Therefore, the *DUP240* family undergoes expansion and contraction of the number of paralogous genes within the *S.cerevisiae* species. This situation also holds true at the inter-species level since preliminary BLAST searches against recently available genome sequences of *Saccharomyces sensu stricto* species (37,38) did not allow us to identify all *DUP240* orthologs.

Systematic sequencing of the coding sequences of five common genes and five Ascomycetes-specific genes represents a first attempt to compare the nucleotide variability between these two gene sets in the *S.cerevisiae* species. Taking into account the number of variable positions, the Ascomycetes-specific genes tend to diverge faster than the common genes. Therefore, the situation at the intra-species level reflects that observed previously between different species of yeast (22). Using these data for common and Ascomycetes-specific gene sets, it is now possible to appreciate the level of nucleotide replacement for genes belonging to a defined multigene family within the *S.cerevisiae* group. For instance, the number of variable positions for YCR007c is similar to that of common genes, while values obtained for YAR023c and YHL044w are superior to all others. This discrepancy between the three paralogs has also been demonstrated by an analysis at the protein level (ds/dn) and confirms that different evolutionary constraints are operative. These results illustrate one aspect of the gene catalog enhancement. In some cases, gene amplification and sequence conservation in both copies are linked to regulation of the level of protein required, as observed for some ribosomal proteins in various organisms (39). In other cases, selection pressure maintains the same structure and function for duplicated copies, but may allow specialization, such as modulation of transcription level, subcellular localization or substrate specificity (40,41). Finally, sequence divergence of one of the duplicated copies may allow the emergence of new functions. This last possibility probably applies to solo *DUP240* ORFs since they are subject to different selection pressures. This situation seems at first paradoxical because it indicates that genes belonging to the same family could be subject to different evolution pressures, but is in accordance with the experimental results obtained for the *DUP240* ORF products. Poirey *et al*. (25) demonstrated that, depending on the paralog, *DUP240* gene products could have different cellular localizations and different types of protein–protein interactions, suggesting that they have different functions. As the *DUP240* ORFs are as yet only found in the *Saccharomyces sensu stricto* group, their associated functions are probably species (or group) specific and may be linked to the speciation process applied to this group (42,43).

Finally, we underscore the important role of allelic homologous recombination in addition to that of nucleotide substitution, in the process of DNA sequence evolution. *Saccharomyces cerevisiae* has a diploid life style in nature, unlike numerous other yeast species and the vast majority of fungi. Analysis of nucleotide replacements at this diploid level reveals that heterozygosity is frequently found at various loci, a genome feature rarely observed with laboratory strains since they are very often homozygous. The trace of possible allelic recombination has been observed for at least five of the tested loci, irrespective of the gene set they belong to. In this situation, the phylogenetic history of the corresponding ORFs in the *S.cerevisiae* species, is more adequately represented by a network than by a bifurcating tree. The impact of recombination events on nucleotide variability has already been reported for viruses (44,45) as well as for eukaryotic cells using the CMH gene complex (46).

In conclusion, this study provides evidence that paralogs of a multigene family can show different evolutionary pathways. If ectopic non-homologous and homologous recombinations have a great impact on the birth-and-death of tandemly repeated *DUP240* ORFs (V. Leh-Louis, B. Wirth, S. Potier, J.-L. Souciet and L. Despons, manuscript submitted for publication), here we observed another aspect of the enhancement of protein diversity through the effect of allelic recombinations. Our results provide new insights into the evolution of duplicated copies for ORFs that are present in a limited group of species. Such ORFs are probably involved in some aspects of the speciation mechanism in the *Saccharomyces sensu stricto* group.

ACKNOWLEDGEMENTS

We are grateful to Michel Aigle, Claude Gaillardin and Huu-Vang Nguyen for providing *S.cerevisiae* strains, and to Philippe Hammann and Malek Alioua for their help with DNA sequencing in the Strasbourg IBMP/CNRS facilities. We thank Mikael Dubow, David Sherman and Stéphane Vuilleumier for comments and suggestions on the manuscript. This work was supported in part by a EU Grant Comprehensive Yeast Genome Database CYGD (QLRI CT 1999 01333) and by the Génolevures-2 sequencing consortium GDR CNRS 2354. Bénédicte Wirth is supported by a grant from the French Ministère de l'Education Nationale, de la Recherche et de la Technologie.

REFERENCES

1. Ohno,S. (1970) *Evolution by Gene Duplication.* Springer Verlag, New York.
2. Meyer,A. and Schartl,M. (1999) Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr. Opin. Cell Biol.*, **11**, 699–704.
3. Lynch,M. and Conery,J.S. (2000) The evolutionary fate and consequences of duplicate genes. *Science*, **290**, 1151–1155.
4. Nei,M., Rogozin,I.B. and Piontkivska,H. (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc. Natl Acad. Sci. USA*, **97**, 10866–10871.
5. Ohta,T. (2000) Evolution of gene families. *Gene*, **259**, 45–52.
6. Mazet,F. and Shimeld,S.M. (2002) Gene duplication and divergence in the early evolution of vertebrates. *Curr. Opin. Genet. Dev.*, **12**, 393–396.
7. Sherman,F., Taber,H. and Campbell,W. (1965) Genetic determination of iso-cytochromes c in yeast. *J. Mol. Biol.*, **13**, 21–39.
8. Filer,D. and Furano,A.V. (1981) Duplication of the *tuf* gene, which encodes peptide chain elongation factor Tu, is widespread in Gram-negative bacteria. *J. Bacteriol.*, **148**, 1006–1011.
9. Pfleiderer,G., Neufahrt-Kreiling,A., Kaplan,R.W. and Fortnagel,P. (1968) Biochemical, immunological and genetic investigations of the multiple forms of yeast enolase. *Ann. N. Y. Acad. Sci.*, **151**, 78–84.
10. Holland,J.P. and Holland,M.J. (1980) Structural comparison of two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase genes. *J. Biol. Chem.*, **255**, 2596–2605.
11. 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* Rd. *Science*, **269**, 496–512.

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018

12. Dujon,B. (1998) European Functional Analysis Network (EUROFAN) and the functional analysis of the *Saccharomyces cerevisiae* genome. *Electrophoresis*, **19**, 617–624.
13. Vision,T.J., Brown,D.G. and Tanksley,S.D. (2000) The origins of genomic duplications in *Arabidopsis*. *Science*, **290**, 2114–2117.
14. Wolfe,K.H. and Shields,D.C. (1997) Molecular evidence for an ancient duplication of the entire yeast genome. *Nature*, **387**, 708–713.
15. Lander,E.S., Linton,L.M., Birren,B., Nusbaum,C., Zody,M.C., Baldwin,J., Devon,K., Dewar,K., Doyle,M., FitzHugh,W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
16. Clark,A.G. (1994) Invasion and maintenance of a gene duplication. *Proc. Natl Acad. Sci. USA*, **91**, 2950–2954.
17. Llorente,B., Malpertuy,A., Neuveglise,C., de Montigny,J., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 18. Comparative analysis of chromosome maps and synteny with *Saccharomyces cerevisiae*. *FEBS Lett.*, **487**, 101–112.
18. Emanuel,B.S. and Shaikh,T.H. (2001) Segmental duplications: an 'expanding' role in genomic instability and disease. *Nature Rev. Genet.*, **2**, 791–800.
19. Carlton,M.B., Colledge,W.H. and Evans,M.J. (1995) Generation of a pseudogene during retroviral infection. *Mamm. Genome*, **6**, 90–95.
20. Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de Montigny,J., Dujon,B. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies. *FEBS Lett.*, **487**, 3–12.
21. 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.
22. Malpertuy,A., Tekaia,F., Casaregola,S., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., de Montigny,J. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 19. Ascomycetes-specific genes. *FEBS Lett.*, **487**, 113–121.
23. Rachidi,N., Martinez,M.J., Barre,P. and Blondin,B. (2000) *Saccharomyces cerevisiae PAU* genes are induced by anaerobiosis. *Mol. Microbiol.*, **35**, 1421–1430.
24. Beh,C.T., Cool,L., Phillips,J. and Rine,J. (2001) Overlapping functions of the yeast oxysterol-binding protein homologues. *Genetics*, **157**, 1117–1140.
25. Poirey,R., Despons,L., Leh,V., Lafuente,M.J., Potier,S., Souciet,J.L. and Jauniaux,J.C. (2002) Functional analysis of the *Saccharomyces cerevisiae DUP240* multigene family reveals membrane-associated proteins that are not essential for cell viability. *Microbiology*, **148**, 2111–2123.
26. Hoffman,C.S. and Winston,F. (1987) A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene*, **57**, 267–272.
27. Goffeau,A., Aert,R., Agostini-Carbone,M.L., Ahmed,A., Aigle,M. and Alberghina,L. (1997) The yeast genome directory. *Nature Suppl.*, **387**, 5–105.
28. Devereux,J., Haeberli,P. and Smithies,O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.
29. Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.
30. Thompson,J.D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, **25**, 4876–4882.
31. Felsenstein,J. (1989) PHYLIP–Phylogeny Inference Package (version 3.2). *Cladistics*, **5**, 164–166.
32. Saitou,N. and Nei,M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, **4**, 406–425.
33. Huson,D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics*, **14**, 68–73.
34. Wain-Hobson,S., Renoux-Elbe,C., Vartanian,J.P. and Meyerhans,A. (2003) Network analysis of human and simian immunodeficiency virus sequence sets reveals massive recombination resulting in shorter pathways. *J. Gen. Virol.*, **84**, 885–895.
35. Holmes,E.C., Urwin,R. and Maiden,M.C. (1999) The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Mol. Biol. Evol.*, **16**, 741–749.
36. Wagner,A. (2001) Birth and death of duplicated genes in completely sequenced eukaryotes. *Trends Genet.*, **17**, 237–239.
37. Cliften,P., Sudarsanam,P., Desikan,A., Fulton,L., Fulton,B., Majors,J., Waterston,R., Cohen,B.A. and Johnston,M. (2003) Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science*, **301**, 71–76.
38. Kellis,M., Patterson,N., Endrizzi,M., Birren,B. and Lander,E.S. (2003) Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature*, **423**, 241–254.
39. Barakat,A., Szick-Miranda,K., Chang,I.F., Guyot,R., Blanc,G., Cooke,R., Delseny,M. and Bailey-Serres,J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiol.*, **127**, 398–415.
40. Gille,C., Goede,A., Schloetelburg,C., Preissner,R., Kloetzel,P.M., Gobel,U.B. and Frommel,C. (2003) A comprehensive view on proteasomal sequences: implications for the evolution of the proteasome. *J. Mol. Biol.*, **326**, 1437–1448.
41. Zhang,J., Zhang,Y.P. and Rosenberg,H.F. (2002) Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.*, **30**, 411–415.
42. Gaillardin,C., Duchateau-Nguyen,G., Tekaia,F., Llorente,B., Casaregola,S., Toffano-Nioche,C., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M. *et al.* (2000) Genomic exploration of the hemiascomycetous yeasts: 21. Comparative functional classification of genes. *FEBS Lett.*, **487**, 134–149.
43. 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.
44. Jung,A., Maier,R., Vartanian,J.P., Bocharov,G., Jung,V., Fischer,U., Meese,E., Wain-Hobson,S. and Meyerhans,A. (2002) Multiply infected spleen cells in HIV patients. *Nature*, **418**, 144.
45. Twiddy,S.S. and Holmes,E.C. (2003) The extent of homologous recombination in members of the genus *Flavivirus*. *J. Gen. Virol.*, **84**, 429–440.
46. Yeager,M. and Hughes,A.L. (1999) Evolution of the mammalian MHC: natural selection, recombination and convergent evolution. *Immunol. Rev.*, **167**, 45–58.
47. Sumbu,Z.L., Thonart,P. and Bechet,J. (1983) Action of patulin on a yeast. *Appl. Environ. Microbiol.*, **45**, 110–115.
48. Nguyen,H.V. and Gaillardin,C. (1997) Two subgroups within the *Saccharomyces bayanus* species evidenced by PCR amplification and restriction polymorphism of the non-transcribed spacer 2 in the ribosomal DNA unit. *Syst. Appl. Microbiol.*, **20**, 286–294.
49. Ryu,S.L., Murooka,Y. and Kaneko,Y. (1996) Genomic reorganization between two sibling yeast species, *Saccharomyces bayanus* and *Saccharomyces cerevisiae*. *Yeast*, **12**, 757–764.

Downloaded from https://academic.oup.com/nar/article-abstract/32/7/2069/1095936
by UNIVERSITE LOUIS PASTEUR SERVICE COMMUN DE DOCUMENTATION user
on 20 March 2018