FEBS Letters 487 (2000) 122–133 FEBS 24384

Genomic Exploration of the Hemiascomycetous Yeasts: 20. Evolution of gene redundancy compared to *Saccharomyces cerevisiae*

Bertrand Llorente[a,*], Pascal Durrens[b], Alain Malpertuy[a], Michel Aigle[b], François Artiguenave[c], Gaëlle Blandin[a], Monique Bolotin-Fukuhara[d], Elisabeth Bon[b,e], Philippe Brottier[f], Serge Casaregola[e], Bernard Dujon[a], Jacky de Montigny[f], Andrée Lépingle[e], Cécile Neuvéglise[e], Odile Ozier-Kalogeropoulos[a], Serge Potier[f], William Saurin[c], Fredj Tekaia[a], Claire Toffano-Nioche[d], Micheline Wésolowski-Louvel[g], Patrick Wincker[c], Jean Weissenbach[c], Jean-Luc Souciet[f], Claude Gaillardin[e]

[a] *Unité de Génétique Moléculaire des Levures (URA 2171 CNRS and UFR 927 Université Pierre et Marie Curie), Institut Pasteur, 25 rue du Docteur Roux, F-75724 Paris Cedex 15, France*
[b] *Laboratoire de Biologie Cellulaire de la Levure, IBGC, 1 rue Camille Saint-Säens, F-33077 Bordeaux Cedex, France*
[c] *Génoscope, Centre National de Séquençage, 2 rue Gaston Crémieux, BP 191, Evry Cedex, France*
[d] *Institut de Génétique Moléculaire (UMR 8621), Bat 400, Université Paris-Sud, F-91405 Orsay Cedex, France*
[e] *Collection de Levures d'Intérêt Biotechnologique, Laboratoire de Génétique Moléculaire et Cellulaire, INRA UMR 216, CNRS URA 1925, INA-PG, BP01, F-78850 Thiverval-Grignon, France*
[f] *Laboratoire de Génétique et Microbiologie, UPRES-A 7010 UPL/CNRS, Institut de Botanique, 28 rue Goethe, F-67000 Strasbourg, France*
[g] *Microbiologie et Génétique ERS 2009, CNRS/UCB/INSA, bat. 405 R2, Université Lyon I, F-69622 Villeurbanne Cedex, France*

Received 9 November 2000; accepted 11 November 2000

First published online 29 November 2000

Edited by Horst Feldmann

Abstract We have evaluated the degree of gene redundancy in the nuclear genomes of 13 hemiascomycetous yeast species. *Saccharomyces cerevisiae* singletons and gene families appear generally conserved in these species as singletons and families of similar size, respectively. Variations of the number of homologues with respect to that expected affect from 7 to less than 24% of each genome. Since *S. cerevisiae* homologues represent the majority of the genes identified in the genomes studied, the overall degree of gene redundancy seems conserved across all species. This is best explained by a dynamic equilibrium resulting from numerous events of gene duplication and deletion rather than by a massive duplication event occurring in some lineages and not in others.

Key words: Paralogue; Orthologue; Genome evolution; Gene duplication; Gene deletion

1. Introduction

Sequence duplication and the correlated emergence of gene families is believed to play a major role in molecular evolution. After duplication, the different copies of a gene can diverge and/or acquire novel regulations that may eventually lead to novel functions [1–6]. Paralogous genes are observed in all genomes sequenced so far [7] and constitute a large fraction of identified genes (ca. 40% in the case of *Saccharomyces cerevisiae* [8–10]). The present study of gene redundancy in the hemiascomycetous genomes deals both with conserved gene families present in their last common ancestor, and gene families that have emerged or disappeared after speciation (Fig. 1).

Recent studies [11,12] showed that the *S. cerevisiae* genome contains ca. 55 chromosomal segments, mostly in pairs, containing series of paralogous genes with conserved order and orientation. Such chromosomal segments may be the result of successive duplications of large chromosomal blocks, duplication of an entire chromosome or even of the whole genome, all of these events being followed by numerous single gene deletions and reciprocal translocations. Whatever the mechanisms, about one third of all the *S. cerevisiae* paralogs fall within the identified blocks. In the whole genome duplication hypothesis proposed by Wolfe et al. [11], a putative allo- or auto-tetraploidisation event has been precisely dated after the separation between the ancestors of *Saccharomyces kluyveri* and *S. cerevisiae*. The present sequencing programme contains yeast species supposed to have emerged after this event according to the phylogenetic tree (*Saccharomyces bayanus*) and others prior to this event (*S. kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis*, *Kluyveromyces marxianus*, *Pichia sorbitophila*, *Pichia angusta*, *Candida tropicalis*, *Debaromyces hansenii* and *Yarrowia lipolytica*). The comparison of the distribution of gene families in all yeast species should, therefore, be informative about the mechanisms involved in the molecular evolution of the hemiascomycetous yeasts.

2. Materials and methods

2.1. Defining paralogous sets of genes in partially sequenced yeast genomes

For each of the 13 yeast species studied, we first considered the genes having detectable homologues in *S. cerevisiae* and classified them according to the gene family classification of *S. cerevisiae* [10]. We then considered all genes without detectable homologues in *S. cerevisiae* but identified by similarity to the GPROTEOME database [13]. Such genes were compared to one another using *blast2*. In both cases, we assumed that two genes of our partially sequenced yeast

*Corresponding author. Fax: (33)-1-40 61 34 56.
E-mail: llorente@pasteur.fr

0014-5793/00/$20.00
PII: S0014-5793(00)02291-2

species are paralogous if their products are homologous to two proteins showing homology between them.

The total set of paralogues in each of the 13 species studied is defined as the union of the paralogues identified by homology to *S. cerevisiae* and to the GPROTEOME database.

2.2. Defining the subtelomeric regions of S. cerevisiae

Subtelomeric regions of *S. cerevisiae* are of lower gene density than internal parts of the chromosomes and contain a high occurrence of paralogous genes, influencing the present analysis because they show low levels of sequence divergence. The extent of such regions is difficult to define and varies between chromosomes. The coordinates of the regions we have considered here as subtelomeres are given in Table 1.

2.3. Rationale for the determination of the gene numbers in each yeast species

When comparing translated RSTs to the *S. cerevisiae* protein database, several situations are encountered (see [13]). In the simplest one, a single RST segment is homologous to only one *S. cerevisiae* gene. The corresponding alignment was validated as 'o' in the present programme (o for orthologue). The number of genes present in the RST was counted as 1 and the *S. cerevisiae* homologue considered as the orthologue with a probability of 1. When a single RST segment is homologous to several distinct *S. cerevisiae* genes with indistinguishable degrees of sequence similarity, the corresponding alignments were validated as 'oo' in the present programme (oo for multiple possible orthologues). In this case, the number of genes present in the RST of the studied species is still 1, but each of the *S. cerevisiae* homologues has a probability of $1/n$ to be the orthologue, if n is the total number of homologues. But there exists a more complex situation in which segments of different RSTs are either homologous to a unique *S. cerevisiae* gene ('o' validation) or homologous to n distinct *S. cerevisiae* genes with indistinguishable degrees of sequence similarity ('oo' validation). In this case, the number of genes present in the studied yeast species is m, and the probability for each of the n *S. cerevisiae* homologues to be their orthologue is m/n. The value of m can only be determined after examination of the contigs formed by the RSTs, since it is necessary to distinguish between partial sequences of a single gene (same contig) from several paralogous genes (distinct contigs). When segments of different non-contigable RSTs were aligned to non-overlapping regions of a *S. cerevisiae* gene, it was generally impossible to conclude whether they originated from the same gene or from different genes in the yeast species sequenced. We thus defined a minimum (min) and a maximum (max) value of m. A specific case giving a finite value for m is represented by tandemly duplicated genes present in the same contig.

The above calculations can be biased by the possible existence of identical segments between paralogous genes (a value of 1 is erronously attributed to m) or by the possibility of allelic divergence in diploid strains (a value of 2 is erronously attributed to m). We considered these two possibilities of minor quantitative incidence. All calculations were performed using min and max values of m but in order to simplify the presentation only the results with min values are given since identical conclusions were reached with max values. In the case of proteins identified by comparison to the GPROTEOME protein database, only the homologue with the best degree of similarity has been considered.

3. Results and discussion

3.1. Estimation of the overall degree of gene redundancy in each yeast species by comparison to S. cerevisiae

In a low coverage random genome sequencing project such as ours, we have only a statistical definition of the presence of gene families in the species studied, and the number of paralogues in a given family can only be defined between min and max values as described in Section 2. To circumvent this difficulty, we used *S. cerevisiae* as a reference since: (i) an exhaustive catalogue of singletons and paralogous gene families exists for that species [10]; (ii) only few genes identified are without homologues in *S. cerevisiae* (less than 6%, except in *Y. lipolytica*: 15%). We considered different genes of a given yeast species as paralogues when they are similar to either the same *S. cerevisiae* singleton, or to genes that are themselves paralogues in *S. cerevisiae*.

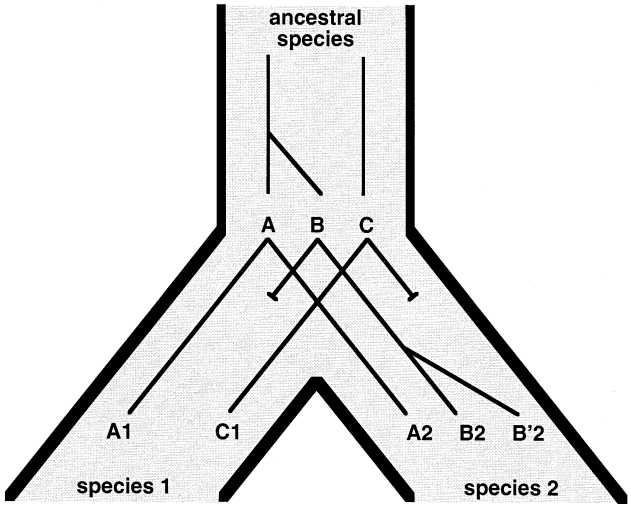

Fig. 1. Gene trees in species trees. This theoretical cartoon depicts the basic events affecting genes before or after speciation, namely conservation, loss or duplication. The species tree is grey coloured. The letters symbolise genes. Gene A from the ancestral species is conserved as A1 and A2 in species 1 and 2, respectively. Gene C from the ancestral species is conserved as C1 in species 1 but lost in species 2. Gene B from the ancestral species is lost in species 1 but conserved in species 2 where it has undergone a duplication event. In the example shown, A and B constitute a 2-gene family in the ancestral species, whereas C is a singleton (1-gene family). In species 1, the ancestral AB gene family is represented by a single gene (A1), while in species 2 it is represented by three genes (A2, B2 and B′2) forming a family of three. In this species, A2, B2 and B′2 are three paralogues. If the duplication event that led to the formation of the two A and B paralogs in the ancestral lineage is very ancient compared to the separation of the two species, the identification of A2 as the orthologue of A1 may be possible. In the opposite case, it may be difficult to distinguish which of A2, B2 or B′2 is the orthologue of A1.

In order to estimate the degree of gene redundancy in each of the 13 hemiascomycete species studied, we analysed the distributions of the homologues to *S. cerevisiae* singletons and to gene families of two or three members. These represent ca. 83% of all *S. cerevisiae* genes. Larger families were not considered here because their number per size class is too small to be significant (see [10]). We then computed the expected distributions if each genome of the other yeast species had the same degree of redundancy as *S. cerevisiae*. Such distributions are given by the probability $P(X=k)$ to obtain k homologues of a given n-gene family of *S. cerevisiae*, multiplied by the number of families of a given size class. This probability is given by the binomial law $(P(X=k)=C_n^k p^k (1-p)^{(n-k)} = n!/(k!(n-k)!)$, where p is the minimum fraction of each set of genes identified, calculated as the ratio of the minimum number of genes identified over the total number of *S. cerevisiae* genes (6213). For each yeast species, we finally compared the expected distribution to the observed one.

In all cases, there is a surprisingly good correlation between the expected and the observed distributions (Fig. 2A, B, C). The quality of the correlation increases with the number of families in each size class, i.e. it is better for singletons than

for 2- and, even more, for 3-gene families. This means that globally the distribution of gene family sizes is essentially conserved in the different yeast species studied (at least for families of two or three members). Yet, in each case, there exists a number of gene families that differ in size from their *S. cerevisiae* homologues, i.e. there occur expansions and contractions of some gene families. The proportion of families affected by the expansion or contraction increases from 4 to 13% in the case of singletons, from 2 to 18% for 2-gene families and from 5 to 31% for 3-gene families. The total number of genes involved in the family size variations represents only ca. 7% of the total. The expansion of some family statistically compensates the contraction of others, such that the overall distribution remains essentially unchanged. Note on Fig. 2 that homologues to *S. cerevisiae* singletons are systematically slightly less numerous than expected. This phenomenon is attributable to the presence of a number of questionable ORFs in the predicted set of singletons (see [14]).

The overall conservation of the *S. cerevisiae* gene family size distributions in other yeast species is represented by the near linear relationship between the mean number of homologues observed for each *S. cerevisiae* gene family size class and the *S. cerevisiae* family size (Fig. 3). In conclusion the overall degree of gene redundancy in each of the 13 hemiascomycete yeasts studied here is similar to that of the *S. cerevisiae* genome.

Considering that gene families showing expansion or contraction represent only 7% of the genes identified in each genome, and that gene families containing at least four genes represent 17% of *S. cerevisiae* genes, it appears that less than 24% of the genome in each yeast species studied is affected by expansion or contraction.

3.2. Genes without S. cerevisiae homologues

We have identified in all yeast species sequenced in this programme a small number of genes showing homology to the GPROTEOME database but not to *S. cerevisiae*. Such genes represent less than 6% of all genes identified except for *Y. lipolytica* (15%). Interestingly, most of the species studied (nine out of 13) contain at least one gene family of two or more genes without detectable homologue in *S. cerevisiae*. For instance, *K. marxianus* [15] and *K. thermotolerans* [16] each have two paralogues of the *Candida albicans SOU1* and *SOU2* genes, respectively. Similarly, *S. cerevisiae* contains multigene families for which we found no homologues in any of the 13 yeast species studied (see Table 1 in [17]). One of the most spectacular of them is the P26.1.f7.1 9-gene family of unclear function [18]. The existence of, apparently, species specific gene families raises an interesting problem with regard to the origin of such sequences and to their expansion or loss in some species.

Table 1
Size of the 32 subtelomeric regions of *S. cerevisiae*

Chromosome	Left subtelomere size (kbp)	Right subtelomere size (kbp)
I	32	28
II	10	15
III	17	27
IV	19	31
V	30	11
VI	33	1
VII	14	26
VIII	33	42
IX	33	15
X	37	37
XI	17	24
XII	47	20
XIII	10	14
XIV	18	48
XV	44	38
XVI	27	28
mean	26.3	25.3

The table indicates the size of the chromosome segments starting from the telomeres that have been considered as 'subtelomeric' in this work. Note the exceptionally small size of the right subtelomere of chromosome VI. Data kindly provided to us by A. Perrin.

3.3. Expansion and contraction of gene families

3.3.1. Rationale. Because of the partial nature of our genomic data, it was not straightforward to identify which genes and gene families had undergone expansion or contraction in *S. cerevisiae* or in any of the 13 yeasts studied since they diverged from their last common ancestor. We applied four criteria, of decreasing value, to try to identify these events.

First, we considered that when a given *S. cerevisiae* singleton had more than one homologue, or when a given *S. cerevisiae* gene family had more homologues than its own gene content, this gene or gene family had non-ambiguously undergone expansion in the considered species, or contraction in *S. cerevisiae*.

Second, we compared the number of *S. cerevisiae* homologues observed in each other yeast species to that expected assuming a gene family distribution identical to that of *S. cerevisiae*. We considered that a gene family was likely to have undergone either size expansion in one of the 13 species studied or contraction in *S. cerevisiae* when the number of homologues observed in that species was higher than the highest number of homologues expected in this species, regarding the genome coverage sequence even if lower or equal to the *S. cerevisiae* family size. Conversely, we considered that a gene family may have undergone size contraction in one of the species or expansion in *S. cerevisiae* when the number of homologues observed in this species was lower than the lowest number of homologues expected. Given the number of *S. cerevisiae* gene families per size class, this second criterion corresponds to probabilities lower than 0.05 to obtain the observed number of homologues for families of at least six members. Only a small number of 4- and 5-gene families follow this criterion.

Third, we considered that a gene family could have undergone either size expansion in *one* of the 13 species studied or contraction in *S. cerevisiae*, when the number of *S. cerevisiae* homologues observed in *at least two* species fitted the expected distributions but only at the upper borderline level, with probabilities of occurrence lower than 0.05. Conversely, we considered that a gene family could have undergone size contraction in *one* of the 13 species studied or expansion in *S. cerevisiae*, when the numbers of *S. cerevisiae* homologues observed in *at least two* species fitted the expected distributions but at the lower borderline level, with probabilities lower than 0.05.

Fourth, we considered that if a gene or a gene family had undergone expansion in at least one of the 13 species or contraction in *S. cerevisiae*, it should be retrieved in more species than expected, and, conversely, in less species than expected in

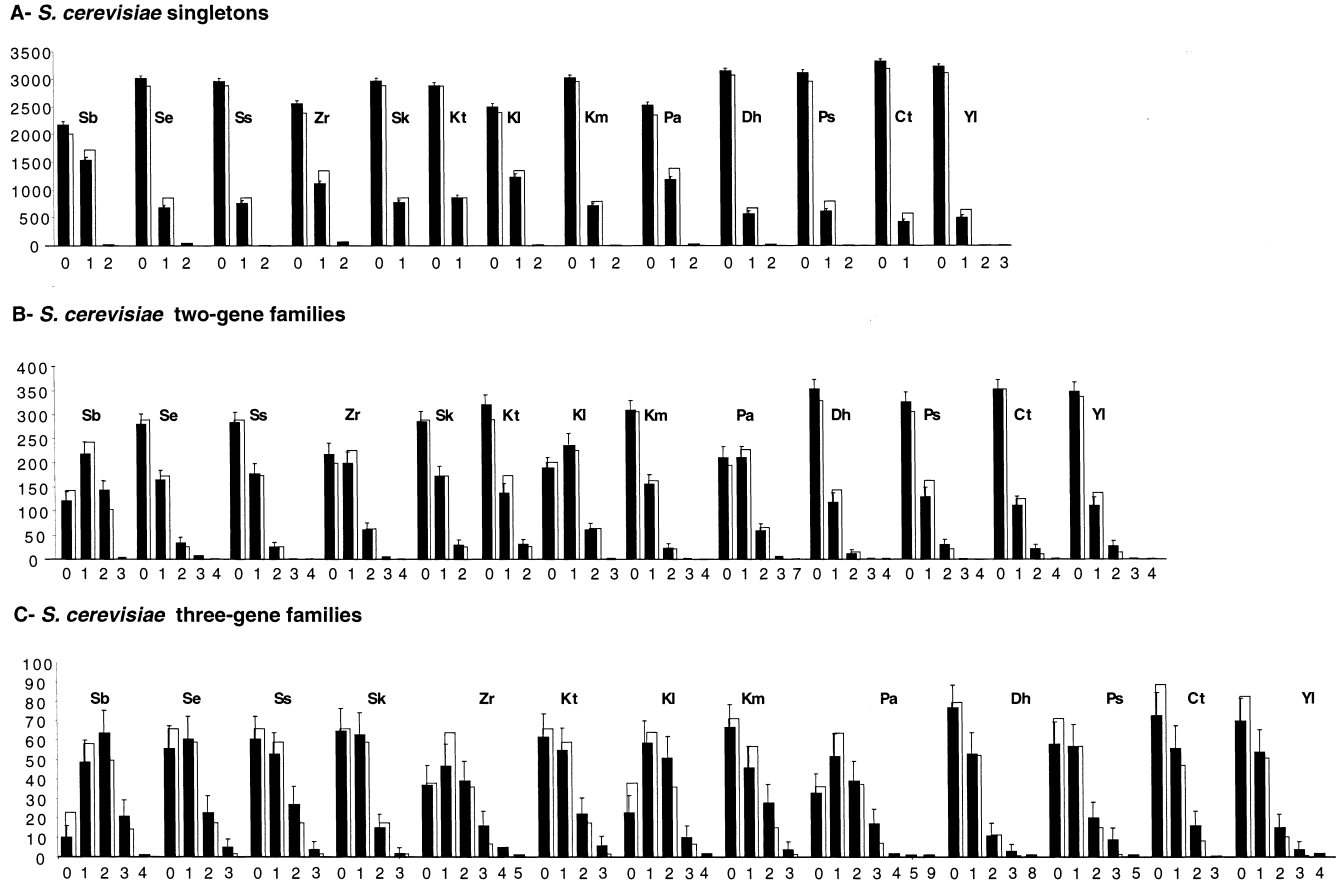

Fig. 2. A, B, C: Distribution of the number of homologues to *S. cerevisiae* singletons and two or three members gene families in the 13 yeast species studied. Figures on the horizontal axis refer to numbers of genes detected in each yeast species that are homologous to *S. cerevisiae* singletons (A), 2-gene families (B), and 3-gene families (C). Black bars represent observed distributions (standard deviations are calculated as $1.96(p(1-p)/n)^{1/2}$, with n being the total number of *S. cerevisiae* genes present in a given family size class, and p being the fraction of the n *S. cerevisiae* genes having the given number of homologues). Void bars represent expected distributions if the gene content of all the species studied was identical to that of *S. cerevisiae*. Species abbreviations are: *S. bayanus* (Sb), *S. exiguus* (Se), *S. servazzii* (Ss), *Z. rouxii* (Zr), *S. kluyveri* (Sk), *K. thermotolerans* (Kt), *K. lactis* (Kl), *K. marxianus* (Km), *P. angusta* (Pa), *P. sorbitophila* (Ps), *C. tropicalis* (Ct), *D. hansenii* (Dh), *Y. lipolytica* (Yl).

Table 2
Non-ambiguous gene duplications of *S. cerevisiae* homologues in the 13 hemiascomycetous yeast species

Species	Number of *S. cerevisiae* genes concerned		Total number of their homologues (3)	Number of expanded families (4)
	Total (1)	% In families (2)		
S. bayanus (Sb)	41	54	84	6
S. exiguus (Se)	82	57	167	9
S. servazzii (Ss)	29	59	63	2
Z. rouxii (Zr)	112	36	230	14
S. kluyveri (Sk)	19	79	38	0
K. thermotolerans (Kt)	7	100	10	0
K. lactis (Kl)	50	74	85	5
K. marxianus (Km)	20	70	38	2
P. angusta (Pa)	122	71	222	14
D. hansenii (Dh)	43	58	97	5
P. sorbitophila (Ps)	26	62	55	4
C. tropicalis (Ct)	13	54	27	2
Y. lipolytica (Yl)	45	73	106	6
Total in the 13 species	609		1224	71
Total in *S. cerevisiae*	523	60		59

Columns (1) and (2) refer respectively to the total number of *S. cerevisiae* genes having more than one homologue in the corresponding other yeast species and the percent of them belonging to multigene families. It concerns 523 *S. cerevisiae* genes, some of them having more than one homologue in at least two different species. Columns (3) and (4) refer respectively to the actual number of homologues found in the corresponding other yeast species and to the number of *S. cerevisiae* gene families that are unambiguously expanded in at least one other yeast species. 59 *S. cerevisiae* families are expanded, some of them in more than one species. Only min values are indicated (see Section 2).

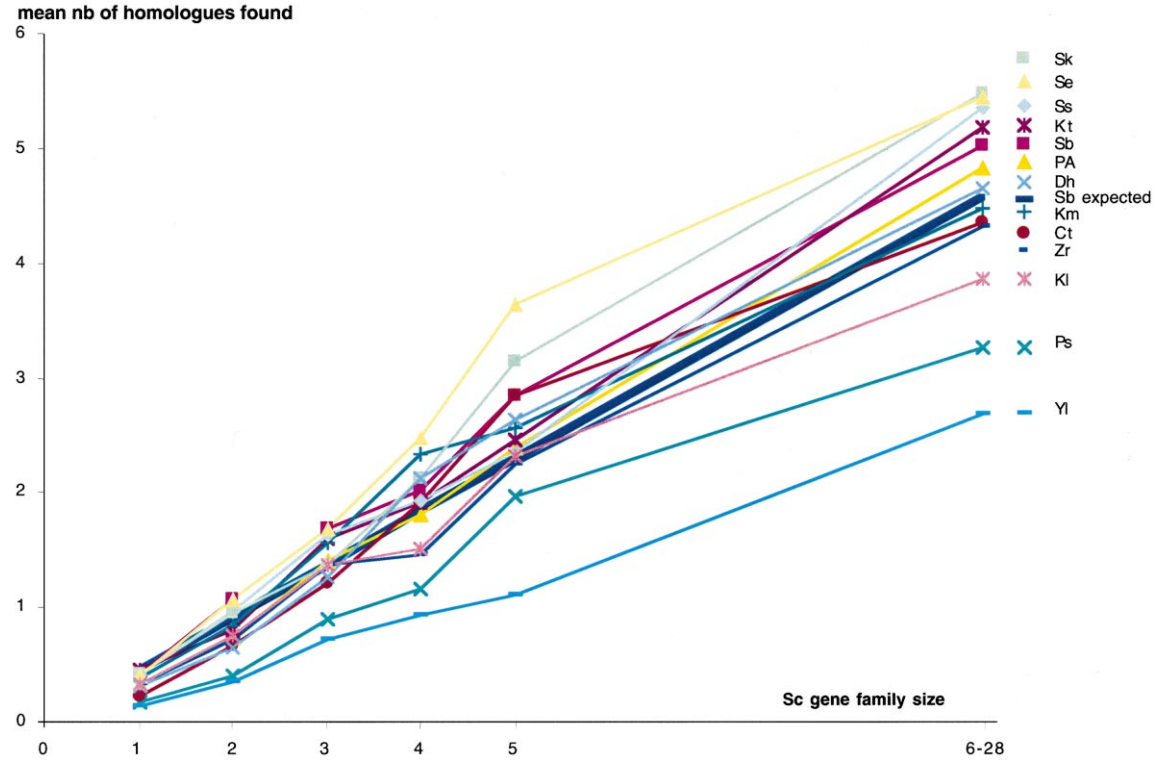

Fig. 3. Mean number of homologous genes found in each of the studied yeast species compared to *S. cerevisiae* genes classified by family size. Abscissae, number of genes in the *S. cerevisiae* families; ordinates, number of genes in the other yeast species partially sequenced in this programme. *S. cerevisiae* gene families containing six genes or more were grouped and considered as one large family whose mean size is 10 members. Figures obtained for species sequenced at ca. 0.2× genome coverage were multiplied by 2 in order to compare them with other species sequenced at ca. 0.4× genome coverage (see [17]). The bold line represents the theoretical curve for *S. bayanus*, with a 0.46 director coefficient. Same species name abbreviations as in Fig. 2.

the case of deletion in some of the 13 species or duplication in *S. cerevisiae*.

Note that the three last criteria do not take into account the sizes of genes although large genes may possibly be overestimated in our RST strategy.

In all these comparisons, it is impossible to conclude from the analysis of a single yeast species whether the putative expansion/contraction event occurred in *S. cerevisiae* or in the other yeast species considered. However, when a same discrepancy with respect to *S. cerevisiae* occurs in several other species, it is likely that the event took place in *S. cerevisiae* rather than several times independently. Conversely, if the discrepancy is observed only once, it is likely that the genetic rearrangement took place only in the concerned species.

3.3.2. Gene family size expansion in various hemiascomycete branches. Part of the variation in the number of homologs to the *S. cerevisiae* singletons or gene families in the 13 hemiascomycete species results from ancestral gene duplications that occurred in the phylogenetic branch leading to the studied species and not in that leading to *S. cerevisiae*. Unambiguous gene duplications could be identified when a given *S. cerevisiae* gene had more than one homolog in one species. This is the case for genes that are singletons in *S. cerevisiae* but present as at least 2-gene families in another species, like *VMA5* (*YKL080w*) which has two distinct homologues in *K. marxianus* [15]. Unambiguous expansion of family size with respect to *S. cerevisiae* is also detectable, as for the 2-gene family P2.168.f2.1 (*YHR123w* and *YNL130c*) encoding 1,2-diacylglycerol ethanolamine phosphotransferase and 1,2-diacylglycerol choline phosphotransferase, that has three detectable homologues in *S. bayanus* [19]. A total of 523 *S. cerevisiae* genes have more than one homolog in at least one of the 13 species, corresponding to 210 singletons and 313 members of 192 different gene families (Table 2). A total of 59 families out of the previous 192 are non-ambiguously larger in at least one of the 13 yeast species studied compared to *S. cerevisiae*. Note that the total number of 523 genes may be overestimated due to the existence of ambiguous matches (see Section 2 for gene number calculation). For example, each of the two genes of the *S. cerevisiae* P2.282.f2.1 (*YBL011w* and *YKR067w*) family of unknown function is represented by 1.5 gene in *P. angusta*, so that a total of three genes was counted in that species [20].

Surprisingly, the proportion of *S. cerevisiae* genes that have undergone expansion in at least one other species of the *Saccharomyces* group (*S. bayanus*, *S. servazzii*, *S. exiguus*, *Zygosaccharomyces rouxii*) is similar to those having undergone expansion in at least one species of the distant group (*D. hansenii*, *P. angusta*, *P. sorbitophila*, *C. tropicalis*, *Y. lipolytica*).

3.3.3. Gene family size variations in the various hemiascomycetous branches. In addition to the previous cases, a number of other gene family size expansions or contractions must have occurred as judged from the frequency distribution of the homologues to *S. cerevisiae* family size classes.

In the simplest case, some *S. cerevisiae* gene families have more (or less) homologues than the highest (or smallest) number expected. In the case of large gene families containing at

Table 3
List of *S. cerevisiae* gene families for which more (+) or less (−) homologues than expected are found in at least one other species

Family		Size	Su	Ss	Se	Zr	Sk	Kt	Kl	Km	Dh	Ps	Pa	Ct	Yl	A	Disc
P24.1.f23.1	(DAN1)	23		1	1	1	1	1	1	1	1	1	1	1	1	12	−
P26.1.f13.1	(COS1)	17	1	1	1	1	1	1	1	1	1	1	1		1	12	−
P17.1.f16.1	(CHD1)	17	1	1	1	1	1	1	1			1		1	1	10	+
P6.3.f6.1	(CDC46)	6			1		1		1	1	1	1	1	1		8	+
P28.1.f22.1	(DBP1)	28			1	1	1	1				1	1		1	7	+
P23.1.f18.1	(AGP1)	18	1			1	1		1		1		1			6	+
P12.2.f8.1	(DTR1)	9				1	1	1			1		1		1	6	+
P7.5.f7.1	(ECM16)	7		1		1	1			1		1			1	6	+
P10.3.f8.1	(CCT2)	10				1	1	1				1		1		5	+
P6.6.f5.1	(MSH1)	6	1	1	1	1						1				5	+
P21.1.f17.1	(AFG2)	21				1			1				1	1		4	+
P108.1.f16.1	(CBK1)	19		1	1			1					1			4	+
P26.1.f7.1	(YAR023c)	9	1			1			1				1			4	−
P9.4.f8.1	(YBL108w)	9	1			1			1				1			4	−
P7.9.f5.1	(ILV2)	7		1	1			1							1	4	+
P10.5.f4.1	(MDJ1)	6						1		1			1	1		4	+
P23.1.f3.1	(BIO5)	5	0			1					0		1	1		4	+
P5.17.f5.1	(DNA2)	5						1				1	1		1	4	+
P5.23.f4.1	(HUL4)	5	0				1		1		1			0		4	+
P2.1.f2.1	(OPT1)	2								1	1	0		1	1	4	+
P2.54.f2.1	(OYE2)	2		1	1						1			0	0	4	+
P108.1.f12.1	(FUS3)	12	1									1	1			3	+
P108.1.f9.1	(CHK1)	12		1		1							1			3	+
P37.1.f4.2	(IMH1)	12	1				1				1					3	+
P11.1.f10.1	(ATM1)	11										1		1	1	3	+
P11.2.f7.1	(FLO1)	11	1			1							1			3	−
P29.1.f7.1	(ASC1)	11		1		1							1			3	+
P37.1.f4.1	(NUF1)	9					1					1		1		3	+
P9.1.f7.1	(ALD2)	9										1	1	1		3	+
P10.2.f5.1	(FEN2)	7						1					1		1	3	+
P7.8.f4.1	(EFT1)	7										1	1	1		3	+
P5.22.f4.1	(MEC1)	5					1			1			1		0	3	+
P4.29.f2.1	(RVS161)	4			0			0		1		1	0			3	+
P10.2.f3.1	(DAL5)	3							0		1	0	1			3	+
P2.423.f2.1	(GDA1)	2			1							0		0	1	3	+
P33.1.f24.1[a]	(GAL2)	25						1	1							2	−
P33.1.f24.1[a]	(GAL2)	25									1				1	2	+
P14.1.f13.1	(ECM10)	14				1						1				2	+
P108.1.f4.1	(AKL1)	13		1	1											2	+
P33.2.f7.1	(DIC1)	11										1	1			2	+
P33.2.f4.1	(MIR1)	10							1				1			2	+
P8.1.f6.1[a]	(AAD10)	8	1						1							2	−
P8.5.f5.1	(BAR1)	8	1												1	2	+
P12.3.f5.1	(ADH1)	6											1		1	2	+
P37.1.f5.2	(CIN8)	6			1								1			2	+
P6.5.f6.1	(ARA1)	6										1			1	2	+
P6.7.f5.1	(GCN1)	6				1								1		2	+
P6.8.f5.1	(HBS1)	6	1		1											2	+
P16.2.f5.1	(DRS2)	5									1	1			0	2	+
P29.1.f3.1	(DOA1)	5							1	1						2	+
P37.1.f5.1	(MYO1)	5		1								1				2	+
P5.6.f3.1	(CSE1)	5					1					1				2	+
P4.12.f3.1	(AOS1)	4										1	0	0		2	+
P4.33.f3.1	(SUL1)	4								0	1			0		2	+
P4.45.f3.1	(ARG8)	4				0	0						1			2	+
P4.47.f4.1	(PLB1)	4			0				0					0	1	2	+
P4.9.f4.1	(ACO1)	4			1		0		0			0				2	+
P3.106.f3.1	(YDL001w)	3				1					0				0	2	+
P3.116.f2.1	(MBP1)	3		0		1			0							2	+
P3.16.f3.1	(MHP1)	3				1			1							2	+
P3.65.f3.1	(KEX1)	3				1									1	2	+
P2.128.f2.1	(CPS1)	2							1			0		0		2	+
P2.264.f2.1	(CLF1)	2								1		0			0	2	+
P2.377.f2.1	(ERC1)	2									1		1			2	+
P2.56.f2.1	(PXA1)	2								0		1		0		2	+
P2.73.f2.1	(ERG10)	2				1						0			0	2	+
P13.1.f9.1	(CDC34)	11												1		1	+
P10.1.f10.1	(ADP1)	10			1											1	+
P24.2.f9.1	(GSP1)	9									1					1	+
P9.2.f9.1	(KRE2)	9		1												1	+
P10.4.f6.1	(DOA4)	8			1											1	+
P33.2.f4.2	(AAC1)	8													1	1	+

Table 3 (*continued*)

Family		Size	Su	Ss	Se	Zr	Sk	Kt	Kl	Km	Dh	Ps	Pa	Ct	Yl	A	Disc
P8.1.f6.1[a]	(AAD10)	8		1												1	+
P8.2.f6.1	(PEP1)	8	1													1	−
P8.3.f5.1	(ACT1)	8				1										1	+
P8.4.f6.1	(CPH1)	8											1			1	+
P7.1.f7.1	(PRE10)	7													1	1	+
P7.2.f7.1	(PMT1)	7							1							1	−
P7.3.f6.1	(ARF1)	7													1	1	+
P7.4.f7.1	(FSP2)	7				1										1	−
P12.3.f5.2	(SFA1)	6											1			1	+
P6.2.f6.1	(CLB1)	6											1			1	+
P6.9.f5.1	(ACC1)	6										1				1	+
P8.7.f4.1	(FDH1)	6												1		1	+
P5.1.f3.1	(BRR2)	5							1							1	+
P5.26.f3.1	(CYS3)	5				1										1	+
P5.8.f4.1	(CYR1)	5										1				1	+
P4.19.f3.1	(BGL2)	4												1		1	+
P4.3.f3.1	(MLH1)	4			1		0									1	+
P4.38.f4.1	(FAA1)	4			1						0					1	+
P4.53.f4.1	(HDA1)	4										1	0			1	+
P16.1.f3.1	(ATR1)	3											1			1	+
P3.107.f3.1	(MET10)	3				1										1	+
P3.26.f3.1	(RPA190)	3			0	0				0						1	+
P3.40.f3.1	(PRB1)	3				0									1	1	+
P3.42.f3.1	(MDH1)	3							1							1	+
P3.7.f3.1	(RET1)	3				0						1				1	+
P3.72.f3.1	(NUP100)	3	1													1	+
P3.91.f3.1	(PPR1)	3					0						1			1	+
P11.3.f2.3	(AZF1)	2											1			1	+
P2.10.f2.1	(YDR060w)	2				1										1	+
P2.123.f2.1	(BUB3)	2											1			1	+
P2.13.f2.1	(AXL1)	2			1											1	+
P2.14.f2.1	(YLR454w)	2				1									0	1	+
P2.151.f2.1	(ARE1)	2		1												1	+
P2.168.f2.1	(CPT1)	2	1													1	+
P2.185.f2.1	(KAP123)	2				1										1	+
P2.193.f2.1	(PFK1)	2							1			0				1	+
P2.213.f2.1	(ERG11)	2				1										1	+
P2.214.f2.1	(GIT1)	2											1			1	+
P2.220.f2.1	(PIF1)	2													1	1	+
P2.25.f2.1	(ERV1)	2				1										1	+
P2.260.f2.1	(PSP1)	2	1													1	+
P2.270.f2.1	(SHM1)	2				1						0				1	+
P2.279.f2.1	(HSP104)	2				1									0	1	+
P2.282.f2.1	(SCT1)	2							1					0		1	+
P2.31.f2.1	(ICL1)	2			1							0				1	+
P2.315.f2.1	(AUR1)	2			1					0						1	+
P2.326.f2.1	(SNG1)	2												1		1	+
P2.329.f2.1	(MNN2)	2									0		1			1	+
P2.332.f2.1	(YBR162c)	2			1											1	+
P2.342.f2.1	(AYR1)	2									1	0				1	+
P2.343.f2.1	(MOB1)	2			1											1	+
P2.352.f2.1	(GIP2)	2	1													1	+
P2.353.f2.1	(YGL059w)	2											1			1	+
P2.355.f2.1	(NBP35)	2											1			1	+
P2.391.f2.1	(CTA1)	2				1						0				1	+
P2.398.f2.1	(YER037w)	2			1											1	+
P2.405.f2.1	(SDS23)	2			1											1	+
P2.409.f2.1	(YFR016c)	2									0	1				1	+
P2.432.f2.1	(DPP1)	2	1							0						1	+
P2.438.f2.1	(ARO3)	2										1			0	1	+
P2.51.f2.1	(YML117w)	2											1			1	+
P29.1.f2.5	(CDC20)	2											1			1	+
P6.13.f2.3	(CHA4)	2	1							0						1	+
P5.15.f5.1	(PRS1)	5												0	0	1	+
P4.13.f4.1	(AAP1)	4				0						0				1	+
P4.15.f3.1	(PRE1)	4									0	0	0			1	+
P4.16.f3.1	(FBP26)	4					0					0				1	+
P4.25.f2.1	(EUG1)	4						0					0			1	+
P4.28.f3.1	(ARG3)	4		0			0						0			1	+
P4.39.f4.1	(ESBP6)	4					0	0	0	0			0		0	1	+
P4.42.f4.1	(IDH1)	4					0			0						1	+
P4.44.f4.1	(GRE2)	4					0							0		1	+

Table 3 (*continued*)

Family		Size	Su	Ss	Se	Zr	Sk	Kt	Kl	Km	Dh	Ps	Pa	Ct	Yl	A	Disc
P4.46.f4.1	(YLL005c)	4											0	0		1	+
P4.5.f4.1	(BUL1)	4		0	0											1	+
P5.10.f2.1	(HAT2)	4			0	0					0			0		1	+
P6.12.f3.1	(DPS1)	4	0					0		0						1	+
P6.13.f3.1	(GAL4)	4	0				0					0	0			1	+
P8.6.f3.1	(OSH1)	4			0		0									1	+
P108.1.f3.5	(KNS1)	3				0		0								1	+
P3.108.f3.1	(YFR021w)	3			0	0				0						1	+
P3.115.f2.1	(CDC53)	3		0				0								1	+
P3.119.f3.1	(YBR239c)	3				0						0				1	+
P3.123.f2.1	(APL3)	3				0					0					1	+
P3.125.f2.1	(ACS1)	3				0									0	1	+
P3.14.f2.1	(CDC2)	3				0	0								0	1	+
P3.22.f3.1	(ILS1)	3				0			0							1	+
P3.23.f3.1	(PHM1)	3							0		0					1	+
P3.28.f3.1	(NOP58)	3							0						0	1	+
P3.43.f3.1	(SMF1)	3		0	0											1	+
P3.66.f3.1	(HEM1)	3							0			0				1	+
P3.71.f3.1	(FET3)	3				0		0		0						1	+
P108.1.f2.6	(GCN2)	2								0		0		0	0	1	+
P15.1.f2.4	(MRD1)	2								0	0					1	+
P16.2.f2.2	(SPF1)	2										0			0	1	+
P2.115.f2.1	(ALG6)	2								0		0				1	+
P2.126.f2.1	(SGT2)	2									0			0		1	+
P2.147.f2.1	(YER006w)	2								0		0				1	+
P2.152.f2.1	(FUR1)	2									0	0				1	+
P2.228.f2.1	(LEU3)	2								0				0		1	+
P2.255.f2.1	(SEC13)	2								0	0					1	+
P2.273.f2.1	(AAT1)	2										0			0	1	+
P2.297.f2.1	(MST1)	2									0				0	1	+
P2.382.f2.1	(YLF2)	2								0		0				1	+
P2.387.f2.1	(FUN11)	2								0					0	1	+
P2.408.f2.1	(YBR281c)	2										0			0	1	+
P2.414.f2.1	(BEM2)	2									0				0	1	+
P2.415.f2.1	(FTH1)	2												0	0	1	+
P2.454.f2.1	(SPB1)	2								0					0	1	+
P2.83.f2.1	(GDI1)	2								0					0	1	+
Total			19	15	25	36	14	12	20	10	16	29	42	17	23	340	

Underlined gene families are those with 50% of their members in subtelomeric location. 1 means that the number of homologues of the gene family in the corresponding species is different from that expected. 0 means that the probability to obtain the observed number of homologues of the corresponding family is lower than 0.05 in at least two species and concerns only the gene families of up to five members. In this case, the statistical significance of the discrepancy is given by the existence of at least two species in which the probability of finding the observed number of homologues is less than 0.05.

[a]Indicates a gene family showing both expansion and contraction in the various yeast species. Column A: sum of the corresponding line. Column disc (for discrepancy): more (+) or less (−) homologues than expected are found. The genetic name, when available, or the systematic name of a gene representative of the family described in [10] is given between parentheses.

least six members, the probability to obtain such numbers of homologues is less than 0.05. For smaller families, this probability is lower and only four such families (three of five genes and one of four genes) fit this criterion at this threshold level. We considered these cases as putative expansions or contractions of the corresponding gene families in either *S. cerevisiae* or the other yeast species. As an example of family expansion, the *S. cerevisiae* 11-gene family P33.2.f7.1 encoding mitochondrial carriers is represented by seven and eight homologues in *P. sorbitophila* [21] and *P. angusta* [20], respectively (the probability this situation occurred by chance is less than 2%). An example of family contraction is found with the *S. cerevisiae* 23-gene seripauperin family P24.1.f23.1 that has homologues only in *S. bayanus*, while several homologues were expected in all species. The likely conclusion is that a rapid and massive expansion of this family occurred in the last common ancestor of the closely related species *S. cerevisiae* and *S. bayanus*.

Other gene families containing less than six members do not fit the above criterion, but the probability of obtaining the observed number of homologues in at least two species is very low (i.e. less than 0.05). For example, the *S. cerevisiae* P3.14.f2.1 3-gene family encoding DNA polymerases has three homologues in each of the three species *P. angusta* [20], *Y. lipolytica* [22] and *S. kluyveri* [23]. It is thus predicted to have undergone expansion in at least one of these three species, or contraction in *S. cerevisiae*. Note, however, that the unusually long length of these proteins could explain that they were more frequently recognised in our partial sequencing data.

Table 3 lists all the *S. cerevisiae* gene families believed to have undergone expansion or contraction according to the two latter criteria. The table comprises 179 different gene families. 171 are suspected to have undergone either size expansion in at least one of the 13 studied species, or size contraction in *S. cerevisiae*. Ten are suspected to have undergone either size contraction in at least one of the 13 studied species, or size expansion in *S. cerevisiae*. The hexose transporter family P33.1.f24.1 and the P8.1.f6.1 family encoding aryl-alcohol dehydrogenase-like proteins appear to have undergone size expansion in some species and size contraction in others. Out of these 179 families, nine contain more than half of their

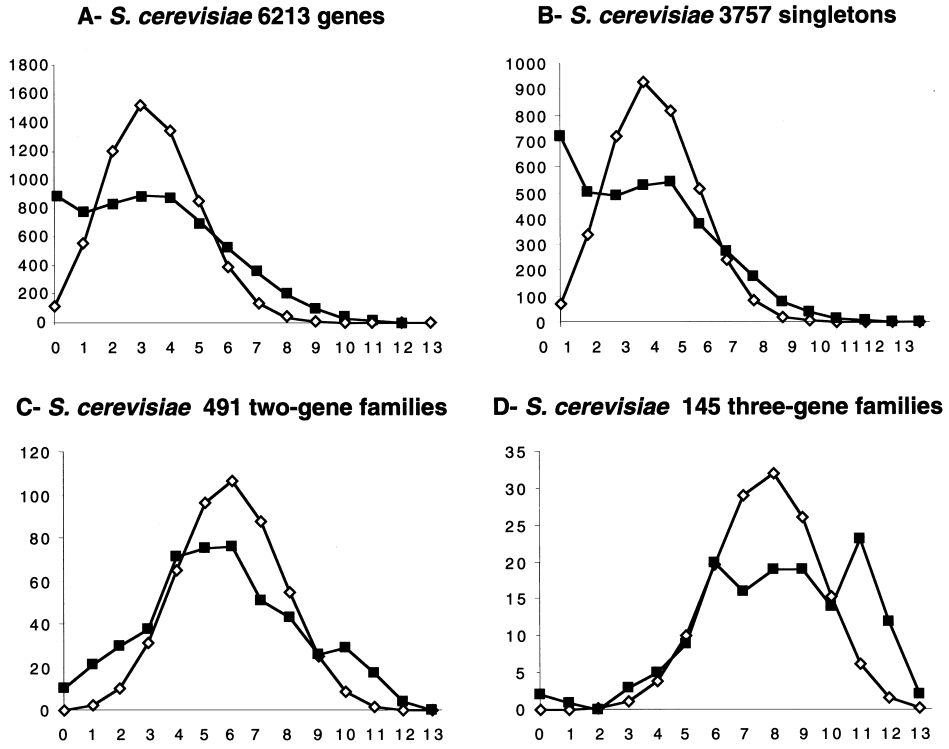

Fig. 4. Distributions of the number of *S. cerevisiae* genes according to the total number of yeast species in which homologues were found. Number of *S. cerevisiae* genes (vertical axis) for which homologues were found in 0,1,...,13 other yeast species (horizontal axis). Squares, observed values; diamonds, expected values. A: The entire set of *S. cerevisiae* predicted ORFs (see [14]). B: The 3757 singletons. C: The 982 genes of the 491 2-gene families. D: The 435 genes of the 145 3-gene families. To calculate expected distributions, we considered that the 13 yeast genomes were exact copies of that of *S. cerevisiae*. We considered that the probability to obtain k homologues of a given *S. cerevisiae* gene among the 13 species followed the binomial law, with a mean p probability of presence into a given genome of 0.26. This value corresponds to the mean of the 13 ratios between the min number of genes in each species over the 6213 *S. cerevisiae* genes.

members located in subtelomeric regions, six of which are suspected of either size contraction in at least one of the 13 species or size expansion in *S. cerevisiae* (this will be discussed below).

As previously noted, the proportion of *S. cerevisiae* gene families that have undergone size variation compared to the *Saccharomyces* group (*S. bayanus*, *S. servazzii*, *S. exiguus*, *Z. rouxii*) is similar to those having undergone size variation compared to the distant group (*D. hansenii*, *P. angusta*, *P. sorbitophila*, *C. tropicalis*, *Y. lipolytica*).

3.3.4. Conservation of S. cerevisiae gene families throughout evolution. Given the partial nature of our sequencing data, some *S. cerevisiae* gene or gene family that have undergone expansion or contraction during evolution may have escaped our statistical analysis because of the limited sample available. We therefore analysed how *S. cerevisiae* homologues were conserved among the entire set of 13 species. To do this, we counted as 1 each *S. cerevisiae* gene having at least one homologue in a given species and as 0 when no homologue was detected in that species. We then compared the expected and the observed distributions of the *S. cerevisiae* singletons and gene families of two or three members in the 13 species (Fig. 4). When singletons are considered (Fig. 4B) the most striking feature is the over-representation of the *S. cerevisiae* genes without any detectable homologue in any of the 13 species. The same is also visible when all *S. cerevisiae* genes are considered together (Fig. 4A). This bias is due to the presence of a significant number of 'questionable' ORFs among the 6213 predicted ones (see [14]). Correlatively, we also note on the distributions an excess of genes having homologues in six or more species. When two members (Fig. 4C) or three members (Fig. 4D) gene families are considered, the discrepancies between observed and expected distributions seem to increase (but not the smaller numbers) but the bias of the zero class disappears. We used these distributions to identify genes and families of *S. cerevisiae* that are more (or less) frequently present in other yeasts than expected. If we consider only the cases in which homologues are found among other species, while none was expected (Table 4), we identify ten families of two genes, three of three genes, six of four genes, two of five genes, two of nine genes, one of 11 genes, one of 17 and one of 23 genes clearly under-conserved among the 13 other species. On the opposite, 11 genes, among which 7 singletons, in addition to four families of two genes and two of three genes are clearly over-conserved among the 13 species. If a five fold ratio between the observed and the expected frequencies of occurrence is taken as an indicator of variation, this adds to the previous list 720 singletons without homologue in any of the 13 species, 52 singletons represented in at least 9 species, 21 2-gene families present in only one species and 17 present in at least 11 species, 12 3-gene families present in at least 12 species, and six 5-gene families present in the 13 species.

Finally, we have plotted the mean number of yeast species containing at least one homologue of a *S. cerevisiae* gene family of a given size against the corresponding size of that family (Fig. 5). This number varies from ca. 2.5 for singletons to 13 for families larger than 10 members. The observed distribution fits the calculated one for singletons and 2- to 5-

Table 4
Under- and over-conservation of *S. cerevisiae* singletons and gene families throughout the 13 yeast species studied

Gene	Family	Size	Nb species	Presence	Nb homologues
YKL182w	singleton	1	12	+	
YHR005c	P2.380.f2.1[a]	2	12	+	
YBL022c	singleton	1	11	+	
YDL140c	P3.26.f3.1[a]	3	11	+	++
YDL141w	singleton	1	11	+	
YKL215c	singleton	1	11	+	
YOR093c	singleton	1	11	+	
YOR141c	singleton	1	11	+	
YPL231w	singleton	1	11	+	
YER013w	P7.5.f7.1	7	11	+	+
YJL095w	P108.1.f4.1	13	11	+	+
(*YIL029c*)	P2.21.f2.1	2	0	−	
(*YJL043w*)	P2.293.f2.1	2	0	−	
(*YIL171wa*)	P2.327.f2.1	2	0	−	
(*YIL102c*)	P2.344.f2.1	2	0	−	
(*YAR068w*)	P2.356.f2.1	2	0	−	
(*YAR060c*)	P2.357.f2.1	2	0	−	
(*CUP1-1*)	P2.374.f2.1	2	0	−	
(*YBR300c*)	P2.383.f2.1	2	0	−	
(*LRE1*)	P2.428.f2.1	2	0	−	
(*RPL35A*)	P2.446.f2.1	2	0	−	
(*CPS1*)	**P2.128.f2.1**	2	12	+	++
(*YBR281c*)	**P2.408.f2.1**	2	12	+	+
(*DIS3*)	P2.142.f2.1	2	12	+	
(*GPA1*)	P2.380.f2.1[a]	2	12	+	
(*GIN11*)	P3.49.f3.1	3	0	−	
(*YLR156w*)	P3.87.f3.1	3	0	−	
(*YNL033w*)	P3.6.f2.1	3	1	−	
(*RPA190*)	**P3.26.f3.1**[a]	3	13	+	++
(*RET1*)	**P3.7.f3.1**	3	13	+	++
(*YAR066w*)	**P4.43.f4.1**	4	1	−	
(*MAL13*)	P4.31.f4.1	4	2	−	
(*YDR533c*)	P4.34.f4.1	4	3	−	
(*AZR1*)	P16.1.f4.1	4	4	−	
(*OAF1*)	P4.37.f4.1	4	4	−	
(*THI11*)	P4.52.f4.1	4	4	−	
(*ASP1*)	P5.19.f5.1	5	3	−	
(*TIP1*)	P5.25.f5.1	5	3	−	
(*FSP2*)	P7.4.f7.1	7	7	−	−
(*AAD10*)	**P8.1.f6.1**	8	7	−	+ and −
(*YBL108w*)	P9.4.f8.1	9	0	−	−
(*YAR023c*)	P26.1.f7.1	9	1	−	−
(*FLO1*)	P11.2.f7.1	11	4	−	−
(*COS1*)	P26.1.f13.1	17	2	−	−
(*DAN1*)	P24.1.f23.1	23	1	−	−

The upper panel of the table deals with *S. cerevisiae* genes considered on their own. The gene family they belong to is indicated. The lower panel deals with *S. cerevisiae* gene families. The genetic name, when available, or the systematic name of a gene representative of the family described in [10] is given between parentheses.
The 'nb species' column indicates the number of species in which at least one homologue of the corresponding *S. cerevisiae* gene or gene family is found. + and − in the 'presence column' indicate that homologues of the corresponding *S. cerevisiae* gene or gene family are found in more or less species than expected, respectively. + and − in the 'nb homologues' column indicate that the number of homologues of the corresponding *S. cerevisiae* gene or gene family is higher or lower than expected in at least one particular species, as indicated in Table 3. ++ indicates a non-ambiguous expansion of the size of the corresponding family in at least one particular species. Bold genes indicate that they have more than one homologue in at least one species. Bold families indicate that at least one of the members of the family has more than one homologue in at least one species (see Table 2).
[a]both the two singletons *YHR005c* and *YDL140c* and their corresponding families P2.380.f2.1 and P3.26.f3.1 are present in this table because they both have homologues in more species than expected.

members families which represent 90% of the *S. cerevisiae* genome.

3.3.5. Conservation of the S. cerevisiae subtelomeres throughout evolution. Genes located in subtelomeric regions of *S. cerevisiae* must represent some of the most recently amplified families as judged from their high degree of sequence conservation. We showed above that subtelomeric gene families tend to be larger in *S. cerevisiae* than in most of the other species studied. We now show (Fig. 5) that such genes tend to be significantly under-represented compared to other genes in other yeast species. This under-representation is true for both singletons and members of gene families suggesting that subtelomeric regions may be a preferential location for the acquisition of novel genes in a given species or for expansion of gene families. Such a role for subtelomeric regions is consistent with the fact that double strand break repair in subtelo-

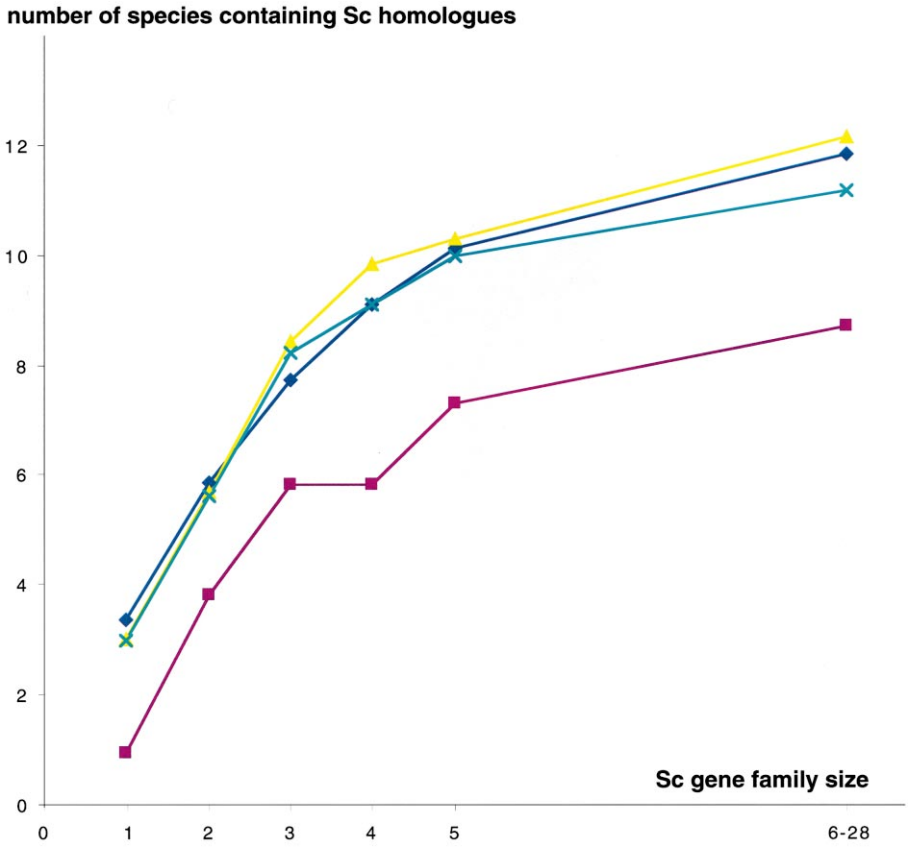

Fig. 5. Conservation of the *S. cerevisiae* gene families in the 13 yeast species studied. The mean number of yeast species containing at least one member of all gene families of a given size class (vertical axis) is plotted against the *S. cerevisiae* gene family size classes (horizontal axis). Crosses, observed numbers; diamonds, expected numbers; triangles, non-subtelomeric gene families; squares, subtelomeric gene families. *S. cerevisiae* gene families containing six genes or more were grouped and considered as one large family whose mean size is 10 members.

Table 5
Number of homologues to the *S. cerevisiae* ancestral block pairs of genes

	Sc	Sb	Se	Ss	Zr	Sk	Kt	Kl	Km	Pa	Dh	Ps	Ct	Yl
Number of homologues to Sc ancestral block pairs	887	478	211	218	279	152	136	271	154	306	137	174	139	138
(1) Ratio versus the number of Sc ancestral block pairs	1	0.54	0.24	0.25	0.31	0.17	0.15	0.31	0.17	0.35	0.15	0.20	0.16	0.16
Homologues to Sc other genes	5326	2404	1240	1192	1966	1254	1297	1964	1147	2014	981	1131	789	940
(2) Ratio versus the 5326 other Sc genes	1	0.45	0.23	0.22	0.37	0.24	0.24	0.37	0.22	0.38	0.18	0.21	0.15	0.18
(3) Ratio versus the 4670 other Sc genes	1	0.51	0.27	0.26	0.42	0.27	0.28	0.42	0.25	0.43	0.21	0.24	0.17	0.20
(1)/(2)	1	1.19	1.02	1.10	0.85	0.73	0.63	0.83	0.81	0.91	0.84	0.92	1.06	0.88
(1)/(3)	1	1.05	0.90	0.96	0.75	0.64	0.55	0.73	0.71	0.80	0.74	0.81	0.93	0.77

Pairs of *S. cerevisiae* paralogues (designed here as ancestral block pairs) are those contained in the ancestral blocks defined in [11] and believed to have arisen by a whole genome duplication event occurred after the ancestors of *S. kluyveri* and *S. cerevisiae* became separated. The predicted 6213 ORFs for *S. cerevisiae* can be broken down into 887 genes of the ancestral block parts (1) and 5326 other genes (2). Note that this last class contains ca. 700 questionable ORFs identified in [14]. The total number of actual *S. cerevisiae* genes being estimated at ca. 5651, a probable figure for this last class is ca. 4670 (3). For each yeast species, the number of homologues to each class of *S. cerevisiae* genes is calculated as indicated in text (Section 3.1). The relative representation of homologues to the ancestral block pairs versus other genes is indicated by the ratios (1)/(2) and (1)/(3).

meric regions can be repaired by insertions of foreign sequences such as mitochondrial DNA [24].

3.4. What about the whole genome duplication hypothesis?

It has been proposed [11,25] that a whole genome duplication occurred in the *Saccharomyces* lineage after the separation of the *S. kluyveri* branch (see phylogenetic cladogram in [17]). This hypothesis was based on the comparison of chromosome numbers among hemiascomycetous yeasts, and on the existence of blocks corresponding to ancestral duplicated regions in the *S. cerevisiae* chromosomes. Such regions cover about half of the *S. cerevisiae* genome and are defined by ca. 400 paralogous gene pairs. In support of the whole genome duplication hypothesis, it has been pointed out that several such pairs of *Saccharomyces* genes correspond to unique genes in *K. lactis* [11].

We have now reexamined this problem in a much more extensive manner using the data from the 13 yeast species. To do so, we have examined the number of homologues per each *Saccharomyces* gene that is part of an ancestral block pair and compared it to the mean number of homologues per each other *S. cerevisiae* gene (Table 5, line a/b). A ratio of 1 indicates that *S. cerevisiae* genes remained in pairs in the other species. Now, for species that have emerged prior to the postulated whole genome duplication, the expected ratio should be 0.5 because homologues of such ancestral block pairs should be unique.

Observed ratios are clearly larger than 0.5 for *P. sorbitophila*, *P. angusta*, *C. tropicalis*, *D. hansenii*, *Y. lipolytica*, *K. lactis* and *K. marxianus*, all believed to have emerged prior to the whole genome duplication. Observed ratios of one or even above are obtained for *S. bayanus*, *S. exiguus*, *S. servazzii* and *C. tropicalis*. Figures above 1 are due to the presence of questionable ORFs in the predicted set of *S. cerevisiae* ORFs. If one eliminates the ca. 700 questionable ORFs from *S. cerevisiae* genes [14], this ratio becomes 1.03 for *S. bayanus*, but still remains clearly larger than 0.5 for *P. sorbitophila*, *P. angusta*, *C. tropicalis*, *D. hansenii*, *Y. lipolytica*, *K. lactis* and *K. marxianus*.

Furthermore, the mean degree of gene redundancy of the homologues to the *Saccharomyces* ancestral block pairs of genes is similar in all the Hemiascomycetes, with the exception of *K. thermotolerans* and *S. kluyveri* where it is smaller. In addition, we showed above that all the 13 yeast species studied have a quite similar degree of gene redundancy. If a whole genome duplication affected the *Saccharomyces* lineage, our data require compensating genetic events to have occurred in the non-*Saccharomyces* lineages. At any rate, it seems unlikely that such a convergence in gene redundancy levels would be the output of independent whole genome duplication events [26].

Acknowledgements: We are very grateful to our colleagues of Unité de Génétique Moléculaires des Levures, especially Arnaud Perrin for sharing unpublished results, to Cécile Fairhead for fruitful discussions and to H. Feldmann and A. Goffeau for careful reading of the manuscript. B.D. is a member of Institut Universitaire de France.

References

[1] Ohno, S. (1970) Evolution by gene duplication, Allen and Unwin, London.
[2] Thomas, J.H. (1993) Trends Genet. 9, 395–399.
[3] Hughes, A.L. (1994) Proc. R. Soc. of Lond. Ser. B Biol. Sci. 256, 119–124.
[4] Pickett, F.B. and Meeks-Wagner, D.R. (1995) Plant Cell 7, 1347–1356.
[5] Nowak, M.A., Boerlijst, M.C., Cooke, J. and Smith, J.M. (1997) Nature 388, 167–171.
[6] Hughes, A.L. (1999) in: Adaptive Evolution of Genes and Genomes (Hughes, A.L., Ed.), pp. 143–179, Oxford University Press, New York.
[7] Koonin, E.V. and Galperin, M.Y. (1997) Curr. Opin. Genet. Dev. 7, 757–763.
[8] Goffeau, A. et al. (1996) Science 274, 563–567.
[9] Dujon, B. (1996) Trends Genet. 12, 263–270.
[10] Blandin, G. et al. (2000) FEBS Lett. 487, 31–36 (this issue).
[11] Wolfe, K.H. and Shields, D.C. (1997) Nature 387, 708–713.
[12] Coissac, E., Maillier, E. and Netter, P. (1997) Mol. Biol. Evol. 14, 1062–1074.
[13] Tekaia, F. et al. (2000) FEBS Lett. 487, 17–30 (this issue).
[14] Malpertuy, A. et al. (2000) FEBS Lett. 487, 61–65 (this issue).
[15] Llorente, B., Malpertuy, A., Blandin, G., Wincker, P., Artiguenave, F. and Dujon, B. (2000) FEBS Lett. 487, 71–75 (this issue).
[16] Malpertuy, A., Llorente, B., Blandin, G., Artiguenave, F., Wincker, P. and Dujon, B. (2000) FEBS Lett. 487, 113–121 (this issue).
[17] Souciet, J.-L. et al. (2000) FEBS Lett. 487, 3–12 (this issue).
[18] Feuermann, M., de Montigny, J., Potier, S. and Souciet, J.L. (1997) Yeast 13, 861–869.
[19] Bon, E., Neuvéglise, C., Casaregola, S., Artiguenave, F., Wincker, P., Aigle, M. and Durrens, P. (2000) FEBS Lett. 487, 37–41 (this issue).
[20] Blandin, G., Llorente, B., Malpertuy, A., Wincker, P., Artiguenave, F. and Dujon, B. (2000) FEBS Lett. 487, 76–81 (this issue).
[21] De Montigny, J., Spehner, C., Souciet, J.-L., Tekaia, F., Dujon, B., Wincker, P., Artiguenave, F. and Potier, S. (2000) FEBS Lett. 487, 87–90 (this issue).
[22] Casaregola, S., Neuvéglise, C., Lépingle, A., Bon, E., Feynerol, C., Wincker, P., Artiguenave, F. and Gaillardin, C. (2000) FEBS Lett. 487, 95–10 (this issue).
[23] Neuvéglise, C., Bon, E., Lépingle, A., Wincker, P., Artiguenave, F., Gaillardin, C. and Casaregola, S. (2000) FEBS Lett. 487, 56–60 (this issue).
[24] Ricchetti, M., Fairhead, C. and Dujon, B. (1999) Nature 402, 96–100.
[25] Smith, M.M. (1987) J. Mol. Evol. 24, 252–259.
[26] Llorente, B. et al. (2000) FEBS Lett. 487, 71–75 (this issue).