Emergence of Species-Specific Transporters During Evolution of the Hemiascomycete Phylum

Benoît De Hertogh,[*,1,2] **Frédéric Hancy,**[†,1] **André Goffeau**[‡,3] **and Philippe V. Baret**[*]

Unité de Génétique (GENA), †Unité de Génétique (GENE) and ‡Institut des Sciences de la Vie (ISV), Université Catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

ABSTRACT

We have traced the evolution patterns of 2480 transmembrane transporters from five complete genome sequences spanning the entire Hemiascomycete phylum: *Saccharomyces cerevisiae*, *Candida glabrata*, *Kluyveromyces lactis*, *Debaryomyces hansenii*, and *Yarrowia lipolytica*. The use of nonambiguous functional and phylogenetic criteria derived from the TCDB classification system has allowed the identification within the Hemiascomycete phylum of 97 small phylogenetic transporter subfamilies comprising a total of 355 transporters submitted to four distinct evolution patterns named "ubiquitous," "species specific," "phylum gains and losses," or "homoplasic." This analysis identifies the transporters that contribute to the emergence of species during the evolution of the Hemiascomycete phylum and may aid in establishing novel phylogenetic criteria for species classification.

THE Hemiascomycete yeasts are believed to have diverged from a common ancestral fungus at least 400 million years ago. This phylum comprises >1200 known species (KURTZMAN and FELL 1998; BOEKHOUT 2005). The first Hemiascomycete genome sequence, that of *Saccharomyces cerevisiae*, has unraveled the existence of >30% of duplicated genes. This observation led to the hypothesis of a recent whole-genome duplication of the Saccharomyces genome (PHILIPPSEN *et al.* 1997; WOLFE and SHIELDS 1997). A massive exploration of partially sequenced genomes from 13 Hemiascomycete species (SOUCIET *et al.* 2000) as well as the sequence comparison of a selected set of 40 duplicated gene sequences from six different species further investigated the origin of duplicated genes within the Hemiascomycete phylum (LANGKJAER *et al.* 2003). Analysis of several nearly complete genome sequences from Hemiascomycete species closely related to *S. cerevisiae* has dated the whole-genome duplication after the emergence of *Kluyveromyces waltii* (KELLIS *et al.* 2004), *Ashbya gossypii* (DIETRICH *et al.* 2004), and *K. lactis* (DUJON *et al.* 2004) but before that of all *sensus stricto* and *sensus lato* Saccharomyces species (CLIFTEN *et al.* 2003; KELLIS *et al.* 2003) and *Candida glabrata* (DUJON *et al.* 2004). These data established the high frequency of gene loss among the duplicated genes. The existence in fungi of numerous gene duplication events followed by differential

evolutionary drift of one of the two gene copies is believed to be a major force for adaptation to novel ecological niches and further speciation (OHNO 1970; KELLIS *et al.* 2003; LANGKJAER *et al.* 2003; DUJON *et al.* 2004).

The complete genome sequence of four species widely spread over the Hemiascomycete phylum became recently available: *C. glabrata*, the second most prominent causative agent of human fungal infection; *K. lactis*, a milk-loving yeast, believed to have diverged from the Saccharomyces clade at least 150 millions years ago; *Debaromyces hansenii*, a halotolerant species contaminating many dairy products; and *Yarrowia lipolytica*, a distantly related yeast that shares a number of properties with filamentous fungi (DUJON *et al.* 2004). In the context of this Genolevures project, a uniform nomenclature was designed to facilitate comparisons between species (DURRENS and SHERMAN 2005). These data established that in addition to a recent whole-genome duplication, variable levels of segmental duplication, tandem repeats, and other duplication events of still unknown mechanisms have occurred during evolution of the Hemiascomycete phylum. The analyses of more ancient genomes such as those from the major human Hemiascomycete pathogen *C. albicans* (JONES *et al.* 2004) and from the Euascomycete filamentous fungus *Neurospora crassa* (GALAGAN *et al.* 2003) are fully consistent with this view.

Most of the gene products and gene families analyzed so far concerned nonmembrane proteins or RNAs. We wish to focus our analysis here on the evolutionary fate of transmembrane transporter proteins that correspond to ~10% of the coding genes in the Hemiascomycete phylum. In fact, the transport in and out of the cell or

organelles of precursors and end products is expected to be an efficient checkpoint for emergence of complex metabolic pathways. An example of coevolution of a transporter and a related metabolic pathway is the loss or inactivation of the galactose transporter *Gal2p* gene and six other genes of the galactose catabolism that has been shown in several Hemiascomycete species (HITTINGER *et al.* 2004).

We have recently published a phylogenetic inventory of 402 established transporter proteins from *S. cerevisiae* (DE HERTOGH *et al.* 2002). Each of these proteins has been classified according to the transport classification (TC) system (SAIER 2000). This classification allocates five digits to each phylogenetic cluster of transporters. The two first digits ("class" and "subclass") identify the global transport mechanism. The third digit characterizes phylogenetic "families" or "superfamilies." The fourth digit identifies phylogenetic "subfamilies." The fifth digit ("clusters") corresponds to the transported substrate or range of substrates, as presumed by experimental data or stringent sequence identity (SAIER 2000). When the TC system did not provide the complete five-digit identification of *S. cerevisiae* transporters, we introduced X, Y, and Z (with numerical indexes such as X1, X2, X3, …) as the provisional last three digits, using only sequence homology criteria (DE HERTOGH *et al.* 2002). Moreover, we compiled a full inventory in *S. cerevisiae* of the class 9.B.X.Y.Z of "possible (or putative) transporters," which comprises membrane proteins of unknown function. This system termed yeast transporter information (YETI) provides a five-digit identifier to each *S. cerevisiae* transporter with more than two predicted transmembrane spans (DE HERTOGH *et al.* 2002).

As an example of the usefulness of our *S. cerevisiae* YETI database, its comparison with the 50,000 random sequence tags provided by the partial sequence of 13 Hemiascomycete species (TEKAIA *et al.* 2000) has unraveled 55 novel transporters, present in Hemiascomycete species but absent in *S. cerevisiae* (DE HERTOGH *et al.* 2003).

We now have extended our YETI database to a total of 2480 transporters encoded by the genomes of five species covering the entire Hemiascomycete phylum, namely *S. cerevisiae, C. glabrata, K. lactis, D. hansenii,* and *Y. lipolytica.* The *N. crassa* species belonging to the Euascomycete phylum served as reference outgroup. We have classified the Hemiascomycete transporters into hierarchized phylogenetic families, subfamilies, and clusters and inferred their evolutionary history by tracing their expansion or contraction in each species.

Comparative analyses of these sequences have allowed us to identify the emergence and loss of "species-specific" transporter subfamilies within the Hemiascomycete phylum and to distinguish them from the "ubiquitous" subfamilies of transporters that are conserved throughout the entire phylum. Moreover the "phylum" transporter subfamilies that are "gained or lost" definitively during evolution have been identified as well as the "homoplasic"

subfamilies of transporters that are conserved transiently in the different Hemiascomycete species.

MATERIALS AND METHODS

A total of 24,165 proteins have been identified in *S. cerevisiae, C. glabrata, K. lactis, D. hansenii,* and *Y. lipolytica* genome sequences (DUJON *et al.* 2004).

A semiautomatic classification pipeline was developed in Perl programming language, with a MySQL database server for data storage and scripts to submit and retrieve data from remote prediction servers HMMTOP (TUSNADY and SIMON 2001) and TMHMM (KROGH *et al.* 2001). The Bioperl Blast wrapper was used to handle blast queries and parse blast reports (http://gemo28.gene.ucl.ac.be/bioflow/). This pipeline, illustrated by Figure 1, comprises three components: selection of putative transporters, exclusion of false positive proteins, and final annotation.

The selection of putative transporters comprises four steps:

Step 1: Selection of all proteins with at least two predicted transmembrane spans according to TMHMM or HMMTOP within the 24,165 Hemiascomycete proteins.
Step 2: Selection by BLAST of each of the 24,165 Hemiascomycete proteins against the Transporter Classification Database (TCDB) (version 09, March 2004) (SAIER 2000; BUSCH and SAIER 2002) that comprises ~1300 examples of transporter proteins from all species. Proteins with an *E*-value $<10^{-19}$ are retained.
Step 3: Selection by BLAST of each of the 24,165 Hemiascomycete proteins against a raw version of YETI that comprises the 1068 proteins from *S. cerevisiae* predicted to have at least two transmembrane spans according to either TMHMM or HMMTOP software. Proteins with an *E*-value $<10^{-19}$ are retained.
Step 4: Combination of proteins selected by steps 1, 2, and 3 in a nonredundant database of putative transporters.

The exclusion of false positive proteins in the putative transporter database comprises two extra steps:

Step 5: Exclusion of soluble proteins that BLAST with an *E*-value $<10^{-19}$ against a database that comprises soluble proteins such as some transcription factors and other proteins known to exhibit falsely predicted transmembrane spans in *S. cerevisiae.*
Step 6: Exclusion of membrane proteins that BLAST with an *E*-value $<10^{-19}$ against another database comprising membrane proteins from *S. cerevisiae* known to have nontransport activity.

The annotation of the transporters database comprises five steps:

Step 7: Allocation of a third digit (family) to each ORF according to the TC of the best BLAST hit in YETI and TCDB, respectively.
Step 8: Allocation of a fourth digit (subfamily) to each ORF according to the TC of the best BLAST hit in YETI and TCDB, respectively.
Step 9: Allocation of a fifth digit (cluster) to each ORF according to the TC of the best BLAST hit in YETI and TCDB, respectively.
Step 10: Manual allocation of digits X,Y, and Z to the proteins from step 4 that exhibit different BLAST values in YETI and TCDB, using the threshold *E*-value $>10^{-35}$ for X, between $1E^{-35}$ and $1E^{-65}$ for Y, and $<1E^{-65}$ for Z.
Step 11: Exclusion or manual allocation of digits to proteins from step 4 that have not been annotated by previous steps.
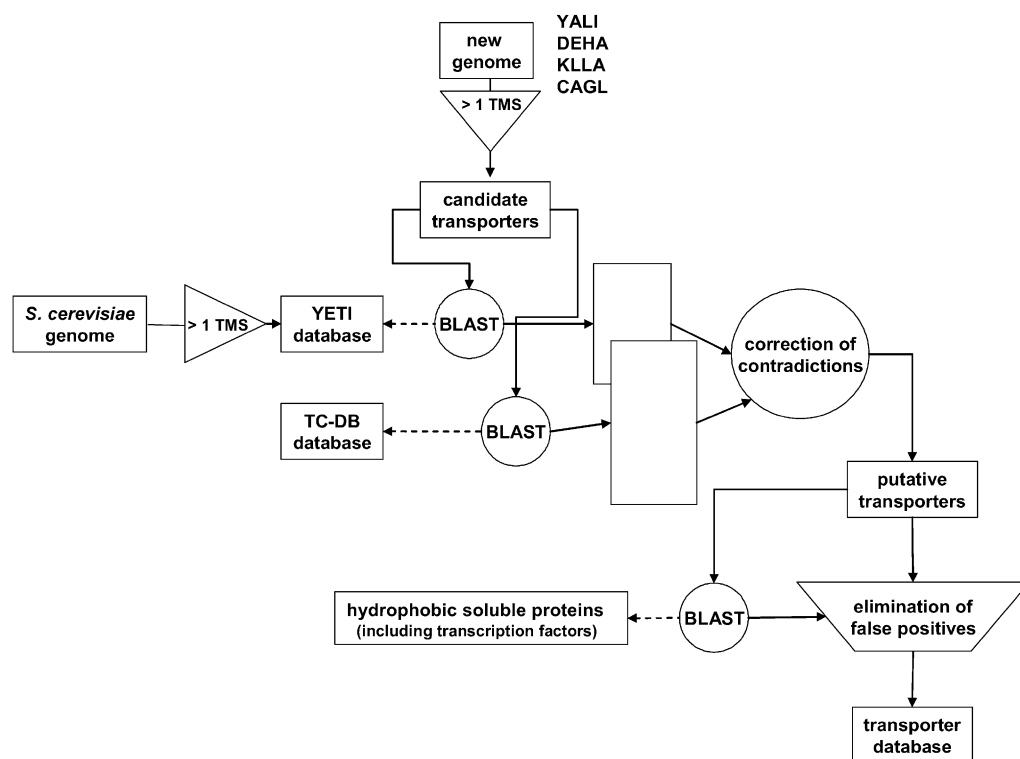
FIGURE 1.—Scheme for semiautomatic annotation.

RESULTS

Global analysis: In this study, membrane proteins are defined as proteins comprising at least two α-helical transmembrane spans predicted either by HMMTOP or by TMHMM. Because of their high number of false positive predictions, the proteins predicted to contain only one transmembrane span were not included. Predicted membrane proteins constitute on average 17.9% (data not shown) of all proteins encoded in the five Hemiascomycete genomes. Nearly half of the predicted membrane proteins are "established" or "possible" transporters (Table 1). Their frequency ranges from 7.5% in *C. albicans* to 9.0% in *Y. lipolytica*. Among the Hemiascomycete 2480 transporters, the secondary porters are by far the most prevalent, followed by transport ATPases

(Table 2). The 2480 transporters were classified into 425 phylogenetic clusters characterized by five digits, which were derived from 204 subfamilies (four digits) and 82 families (three digits). The YETI classification of these 2480 transporters is given in supplemental data (supplemental Table D1 at http://www.genetics.org/supplemental/) and includes the 107 9.B.X proteins that are homologous to the *S. cerevisiae* membrane proteins of which the function is still unknown and that are considered as possible transporters. However the 1296 membrane proteins of unknown function from the non-*S. cerevisiae* species that have no homolog in *S. cerevisiae* were not included. They contain a few interesting novel putative transporters but also numerous nontransport proteins that will be analyzed in further work.

TABLE 1

Global statistics of membrane proteins in Hemiascomycete species

Species: Code:	*Y. lipolytica* *YALI*	*D. hansenii* *DEHA*	*K. lactis* *KLLA*	*C. glabrata* *CAGL*	*S. cerevisiae* *SACE*	Total
Strain	CLIB122	CBS767	CLIB210	CBS138	S288c	
Database	Génolevures	Génolevures	Génolevures	Génolevures	SGD	
Release	May 22, 2004	May 22, 2004	May 22, 2004	May 22, 2004	May 22, 2004	
Natural substrate	Fats	Salted fish	Milk	Blood	Grapes	
ORFs	6,666	6,896	5,331	5,272	5,800	29,965
Classified transporters	597	538	439	398	508	2,480
%	9.0	7.8	8.2	7.5	8.8	8.3
Possible transporters (9.B.X.Y.Z)						
Still unannotated	296	295	226	236	243	1,296

SGD, Saccharomyces Genome Database.

TABLE 2

Functional distribution of the "established and putative" transporters in the Hemiascomycete phylum

	YALI	DEHA	KLLA	CAGL	SACE	Total
1.A, α-type channels	46	32	26	26	38	168
1.B, β-barrel porins	1	1	1	2	2	7
2.A, porters (uniporters, symporters, antiporters)	316	281	206	162	218	1183
3.A, P-P-bond-hydrolysis-driven transporters	94	89	91	87	107	468
3.B, decarboxylation-driven transporters	2	1	1	2	0	6
3.D, oxidoreduction-driven transporters	21	27	18	15	25	106
3.E, light absorption-driven transporters	0	0	0	3	3	6
8.A, auxiliary transport proteins	5	7	5	3	10	30
9.A, recognized transporters of unknown mechanism	82	75	76	81	85	399
9.B, putative uncharacterized transport proteins	30	25	15	17	20	107
Total	**597**	**538**	**439**	**398**	**508**	**2480**

Transporter superfamilies or families (as determined by three digits) are of low sequence similarity to each other. The gain or loss of a complete superfamily or family is not frequent during evolution of the Hemiascomycete phylum. On the other hand, the fifth digit distinguishes clusters that belong to the same phylogenetic subfamily and are likely to transport substrates of similar nature. Obviously, the evolution process involved in the extension or loss of members of phylogenetic clusters is more limited and of less physiological significance than that required to gain or lose subfamilies. Taking into account also that we are dealing with contemporary species that may have continued to evolve after speciation we focus primarily on the evolution of the subfamilies (as determined by four digits—see supplemental Table D2 at http://www.genetics.org/supplemental/) and mention only a few striking examples of cluster variation that we believe to be of particular physiological significance.

Variation of the size of phylogenetic subfamilies: To compare the relative variations of the size of the small as well as of the large subfamilies, their statistical variance was calculated. The variance is the sum of the squares of the differences between the number of transporters in each species and the average number of transporters across species. This sum is divided by 5 (the number of species). The variance value increases with the average value and takes into account both the size of the cluster and its variability. Table 3 lists the number of transporters of the subfamilies that have the highest variance across the five studied species. These subfamilies are those that have drifted most obviously during the evolution of the Hemiascomycete phylum.

The anion:cation symporter (ACS-2.A.1.14) subfamily takes up anionic vitamins (allantoate, nicotinate, panthotenate, biotin, . . .) in symport with protons or other cations. This subfamily is highly amplified in *Y. lipolytica* (39 transporters including 15 members of the "nicotinate cluster") and *D. hansenii* (27 transporters including 9 members of the "allantoate cluster") compared to the 6–13 transporters in the three more recent Hemiascomycete species.

The sugar porter (SP-2.A.1.1) subfamily that comprises 17 transporters in *C. glabrata* is considerably amplified in *S. cerevisiae* (34 transporters) and rather unexpectedly so in *D. hansenii* (48 transporters). The

TABLE 3

The most variable subfamilies based on the statistical variance of the number of ORF, by subfamily

Subfamily	YALI	DEHA	KLLA	CAGL	SACE	Mean	Variance
2.A.1.14, anion cation symporter (ACS)	39	27	13	6	10	19.0	187.5
2.A.1.1, sugar porter (SP)	27	48	20	17	34	29.2	153.7
2.A.1.2, drug proton antiporter 1 (DHA1)	33	24	8	10	12	17.4	114.8
9.A.5.1, peroxisomal protein importer (PPI)	27	7	10	11	10	13.0	63.5
2.A.67.1, oligopeptide transporter (OPT)	17	4	3	0	2	5.2	45.7
2.A.1.16, ferrioxamine H+ symporter (SIT)	14	5	4	1	6	6.0	23.5
2.A.1.13, fructose uniporter (FTU)	5	8	12	3	0	5.6	21.3
9.B.17.1, putative fatty acid transporter (FAT-1)	14	3	3	5	5	6.0	21.0
3.D.1.2, NADH dehydrogenase 1 (NDH 1)	8	8	0	0	0	3.2	19.2
2.A.3.10, amino acid-polyamine-organocation yeast transporter (APC-YAT)	14	24	16	14	18	17.2	17.2
1.A.20.5, yeast metal channel (Cyt B-FRE)	11	7	5	1	7	6.2	13.2

The subfamily 2.A.1.1 sugar porter has been reclassified recently as 2.A.1.1.13.

TABLE 4

Mean and standard deviation of the subfamily size according to the different modes of evolution within the Hemiascomycete phylum

Mode of evolution	No. of subfamilies	Mean no. of ORFs per subfamily	Minimum no. of ORFs	Maximum no. of ORFs
Ubiquitous	107	19.9 ± 25.8	5	146
Species-specific unique	15	1.5 ± 1.6	1	7
Species-specific absent	20	7.9 ± 7.0	4	28
Phylum gained	13	3.2 ± 1.0	2	6
Phylum lost	36	2.6 ± 2.6	1	16
Homoplasic	13	3.1 ± 1.0	2	5

latter species gained 21 members belonging to the new clusters of undetermined substrates 2.A.1.1.Z1/Z10/Z14 that altogether constitute only one single member in *S. cerevisiae*. Surprisingly *Y. lipolytica* has no member of the prominent clusters 2.A.1.1.5/6/30/31, which comprise a total of 18 members (HXT) in *S. cerevisiae*.

The drug:proton antiporter-1 (12 transmembrane spans) (DHA1-2.A.1.2) subfamily pumps out a variety of hydrophobic drugs and is believed to be driven by the proton motive force. The DHA1 subfamily comprises 12 transporters in *S. cerevisiae* and is highly amplified in *Y. lipolytica* (33 transporters) and *D. hansenii* (24 transporters while *K. lactis* comprises only 8 transporters). The major expansion in *Y. lipolytica* corresponds to two novel clusters of undetermined substrates: 2.A.1.2.Z1 and more specifically 2.A.1.2.Z6 that comprises 8 members unique to *Y. lipolytica*. Note that a related subfamily of similar function, the DHA2 subfamily that comprises 14 transmembrane spans, is much more stable during evolution.

The peroxisomal protein importer (PPI-9.A.5.1) subfamily groups members of the peroximal membrane translocon. This subfamily is highly amplified in *Y. lipolytica* (27 members) compared to <11 members in the other species.

The oligopeptide transporter (OPT-2.A.67.1.) subfamily carries out the uptake of small oligopeptides such as the sexual peptides, glutathione, and others. This subfamily does not exist in *C. glabrata* and is extremely amplified in *Y. lipolytica* (17 transporters including 16 members of the 2.A.67.1.4 cluster named oligopeptide transporter type 4). Noteworthy is the rather parallel pattern of amplification within the five Hemiascomycete species of OPT and the proton-dependent oligopeptide transporter (POT-2.A.17.2) subfamily that also takes up oligopeptides.

The siderophore iron transporter subfamily (SIT-2.A.1.16) comprises a total of seven clusters, from which the substrates of only four have been determined in *S. cerevisiae*. In *Y. lipolytica*, a cluster of undetermined substrate is highly amplified (14 members) while *C. glabrata* has no SIT member.

The cluster of fructose uniporters (FTU-2.A.1.1.13) that has no member in *S. cerevisiae* and is highly am-

plified in *K. lactis* (12 members) is present in all other Hemiascomycete species.

The putative long chain fatty acid transporters subfamily (Fat1-9.B.17.1) is composed of fatty acyl coasynthases that have been proposed to control the uptake of long chain fatty acids. This subfamily is considerably amplified in *Y. lipolytica*, which contains a large species-specific cluster.

The proton-translocating NADH dehydrogenase (NDH1-3.D.1.2) subfamily that corresponds to site 1 of mitochondrial oxidative phosphorylation was known to be lost in *S. cerevisiae* (ONISHI *et al.* 1967). Our analysis dates the loss of phosphorylation site 1 between the emergence of *D. hansenii* (where it is present) and that of *K. lactis* (where it is absent).

The large amino acid-polyamine-organocation family (APC-2.A.3) contains the YAT-2.A.3.10 subfamily that transports a variety of amino acids in yeasts. The YAT subfamily is expanded in *D. hansenii* (24 members) compared to 14–18 members in the other species. Noteworthy is the different pattern of expansion of another APC subfamily (ACT-2.A.3.4) that transports choline as well as amino acids and is expanded in *Y. lipolytica* (9 members) compared to 2 members in *C. glabrata*.

The human phagocyte NADPH oxidase-associated cytochrome b558 H+-channel family (CytB-1.A.20) comprises electrogenic heme-binding proton channels that are believed to compensate for production of superoxide radicals. It includes the yeast metal transporter FRE-1.A.20.5 subfamily that has ferric and cupric reductase activity. The latter comprises only 1 member in *C. glabrata* compared to 7 members in *S. cerevisiae* and expands up to 11 members in *Y. lipolytica*.

The ubiquitous subfamilies of transporters: Supplemental Table D1 (http://www.genetics.org/supplemental/) includes 2125 transporters classified in 107 subfamilies that are present in all Hemiascomycete species. Not surprisingly, most of them are large superfamilies (see Table 4) transporting essential substrates such as the mitochondrial carrier (21 subfamilies), several major facilitators (transporting sugars, allantoate, siderophores, drugs, CoA, and phosphate), ABC transporters (7 subfamilies), and P-type ATPases (6 subfamilies). The numerous

TABLE 5

Species-specific and phylum-evolving transporter subfamilies

	YALI	DEHA	KLLA	CAGL	SACE
Unique	2.A.6.Y1 2.A.29.Y19 3.A.1.211 9.A.11.Y1	2.A.37.Y1 2.A.43.Y1 9.A.11.Y2 9.A.11.Y3 9.A.12.Y1 9.B.2.Y1 9.B.24.Y3 9.B.4.Y1	9.A.12.Y2	2.A.29.Y15	8.A.9.Y1
Absent	2.A.29.10 2.A.29.9 2.A.29.Y2 2.A.37.4 2.A.39.4 9.A.26.1 9.B.26.Y1	1.A.1.7 2.A.1.24 2.A.38.2 2.A.66.3		2.A.1.12 2.A.20.2 2.A.54.1 2.A.59.1 2.A.67.1 8.A.5.1 9.B.12.3	2.A.1.13 3.B.1.1
Gained			1.A.8.5 2.A.29.Y8 2.A.43.2 2.A.5.2 9.A.10.Y1 9.A.11.1 9.A.9.1 9.B.1.1 9.B.24.Y1	2.A.29.Y11 3.E.1.4 9.A.12.1 9.B.2.1	
Lost	1.A.4.Y1 1.A.20.Y1 1.A.31.Y1 2.A.1.28 2.A.2.6 2.A.29.Y16 2.A.29.Y18 2.A.29.Y22 2.A.29.Y23 2.A.3.1 2.A.43.3 2.A.7.12 2.A.7.Y2 2.A.72.2 3.A.1.201 9.B.24.Y2 9.B.26.Y2	1.A.6.Y1 1.A.6.Y2 1.A.11.2 1.A.20.Y2 1.A.X1.Y1 2.A.1.Y3 2.A.18.Y1 2.A.29.Y20 2.A.29.Y21 2.A.41.2 2.A.5.Y2 3.D.1.2 9.B.1.Y1 9.B.17.Y2	2.A.18.4 2.A.29.6 2.A.29.Y17 2.A.40.5 2.A.X2.Y1		

Unique, present only in this species; Absent, absent only in this species; Gained, appears between the previous column species and the next column species; Lost, disappears between the previous column species and the next column species.

members of the nuclear pore complex (NPC) (5 subfamilies), the mitochondrial protein translocase (MPT) complexes, the outer mitochondrial pore (MPP), the oxidative phosphorylation complexes (F-ATPase, COB, COX), and the vacuolar-ATPase (V-ATPase) complex are fully conserved in all Hemiascomycete species as well as the calcium (VIC), chloride (ClC), or water (MIP) channels and amino acid (AAA, APC), cation (CDF), metal (MIT), nucleotide-sugar (DMT), or sugar-phosphate (TPT) transporters. The conservation of subfamilies of undetermined substrate such the MFS subfamily 2.1.Y1 or of the YaaH family (which is a controversed ammonium effluxer or acetate influxer) or of the lipid translocating efflux (LTE) family of unknown mechanisms is more unexpected.

The different evolution patterns of transporter subfamilies: Only 355 transporters are submitted to evolution drift in the Hemiascomycete phylum. They are
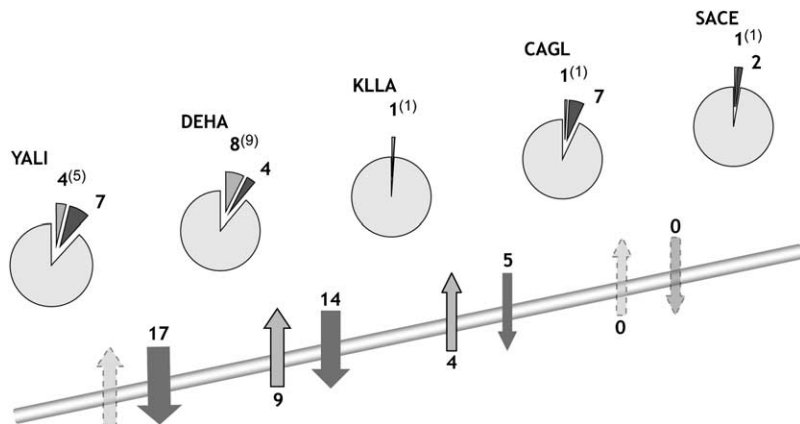
FIGURE 2.—A summary of the evolution of the subfamilies of transporters. The numbers in bold-face type refer to the numbers of gained or lost subfamilies. The numbers in parentheses indicate the total numbers of ORFs in these subfamilies. Dashed arrows refer to undocumented flows.

classified into 97 subfamilies that generally are of smaller size than the ubiquitous subfamilies (Table 4). Three different patterns of evolution were distinguished. They are listed in Table 5 and illustrated in Figure 2. The species-specific subfamilies (in the pie graphs at the top) are present (gain in light shading) or absent (gap in dark shading) in only one species. The phylum-gained-or-lost subfamilies are definitively gained or lost during intervals of evolution spanning the emergence of two distinct species (for instance, between the emergence of *D. hansenii* and *K. lactis*) and persisting in the species that emerged thereafter (Figure 2, vertical thick arrows). When subfamilies are gained and lost more than one time independently within the five species, they are considered as being submitted to homoplasic evolution (Table 7). These different evolution patterns were distinguished quite simply by using the matrix of presence or absence of transporters in the different species illustrated in Table 6.

The species-specific transporters: Table 5 identifies the transporter subfamilies that are unique or specifically absent in one of the five yeast species. A total of 17 transporters belonging to 15 subfamilies are unique to one Hemiascomycete species and a total of 21 subfamilies are absent in only one species (see Table 5 and supplemental Table D3 at http://www.genetics.org/ supplemental/ for detailed listings and comments).

As shown in Figure 2 and Table 5, the species that exhibit the highest number of unique Hemiascomycete transporters are *D. hansenii* (nine specific transporters) and *Y. lipolytica* (five specific transporters) whereas *S. cerevisiae*, *K. lactis*, and *C. glabrata* have each only one unique transporter. As shown by the frequent use of Y as the fourth digit (see Table 5), the nature of the substrates transported by the unique transporters is often undetermined. This may be related to the particular nature of the growth substrates of *Y. lipolytica* (lipids and hydrocarbons) and of *D. hansenii* (protein hydrolysates).

TABLE 6

Identification principles of the different evolution patterns distinguished in Figure 2

Symbols used in Figure 2	Mode of evolution	Species A	Species B	Species C	Species D
	Ubiquitous	≥1 ORF	≥1 ORF	≥1 ORF	≥1 ORF
	Unique	No ORF	≥1 ORF	No ORF	No ORF
	Absent	≥1 ORF	No ORF	≥1 ORF	≥1 ORF
	Gained	No ORF	No ORF	≥1 ORF	≥1 ORF
	Lost	≥1 ORF	≥1 ORF	No ORF	No ORF
Not shown	Homoplasic	≥1 ORF	No ORF	≥1 ORF	No ORF

TABLE 7

Homoplasic evolution

Subfamily	YALI	DEHA	KKLA	CAGK	SACE
1.A.8 1.A.8.7, undefined subfamily of the major intrinsic protein (MIP) family	2	0	0	0	1
2.A.1.Y2, undefined subfamily of the major facilitator (MFS) superfamily	1	1	0	0	1
2.A.5.Y1, undefined subfamily of the zinc–iron permease (ZIP) family	0	0	1	0	1
2.A.7.10, the UDP-*N*-acetylglucosamine: UMP antiporter (UAA) family	1	0	2	0	1
2.A.19.Y1, undefined subfamily of the calcium cation antiporter (CaCA) family	0	1	1	0	1
2.A.29.14, undefined subfamily of the mitochondrial carrier (MC) family	1	0	0	0	1
2.A.29.Y14, undefined subfamily of the mitochondrial carrier (MC) family	0	1	1	1	0
2.A.X1.Y, undefined subfamily of the porters	2	0	2	1	0
8.A.9.1, undefined subfamily of the rBAT transport accessory protein (rBAT) family	0	2	2	0	0
8.A.15.Y1, undefined subfamily of the K+ channel accessory protein (KChAP) family	1	0	0	1	0
9.B.12.1, undefined subfamily of the stress-induced hydrophobic peptide (SHP) family	0	1	0	0	1
9.B.4.Y2, undefined subfamily of the putative efflux transporter (PET) family	1	0	1	1	0
9.B.4.Y3, undefined subfamily of the putative efflux transporter (PET) family	1	0	1	1	0

The species-specific losses are most abundant in *Y. lipolytica* and *C. glabrata*. For the latter a general high loss of genes has been noted (DUJON *et al.* 2004) and believed to reflect the pathogenic status of the species. Curiously, the specific gains (and losses) of transporters are very limited in the milk-loving *K. lactis*.

The phylum-gains-and-losses of transporters subfamilies: Table 5 lists the 15 subfamilies that are gained in *D. hansenii* and *K. lactis* and are maintained in the more recent species during the evolution of the Hemiascomyces phylum. In contrast, numerous subfamilies are present in *Y. lipolytica* (17 subfamilies), *D. hansenii* (14 subfamilies), and *K. lactis* (5 subfamilies) but absent in the younger species. They are therefore considered as being definitively lost after the emergence of *Y. lipolytica*, *D. hansenii*, or *K. lactis*, respectively. It has to be noted that most of the transporters lost after emergence of *Y. lipolytica* were also present in *N. crassa*. The nature of the transporters gained or lost during phylum evolution is commented upon in supplemental Table D4 (http://www.genetics.org/supplemental/).

The homoplasic transporters: Table 7 shows that 12 transporter subfamilies comprise members that have emerged two or even three times during the speciation of the five Hemiascomycete species. They belong to a variety of families and the physiological driving forces of these evolution patterns are difficult to comprehend especially as most of the transported substrates are of undetermined nature.

The mitochondrial carrier: The mitochondrial carrier (MC) family is unusual by its large size: 172 Hemiascomycete transporters distributed in 36 subfamilies. The transported substrates of 27 of the 35 *S. cerevisiae* carriers are known so far. They include monocarboxylic, dicarboxylic, or tricarboxylic acids; sulfate, phosphate, iron, manganese, basic, or acid amino acids; ATP/ADP or GTP/GDP nucleotides; and carnitine, SAM, CoA, FAD, or TPP cofactors. The MC family comprises mitochondrial and peroxisomal members. Even though each carrier is believed to comprise six transmembrane spans, their hydrophobic properties are atypical and their number is underestimated by most software and especially by TMHMM.

Despite a large spread of phylogenetic subfamilies and clusters among the five fungal species listed, each species contains a similar total number of individual transporters (32–39) that can be classified in 36 phylogenetic subfamilies (supplemental Table D1 at http://www.genetics.org/supplemental/). Twenty subfamilies are present in all five species but several species-specific gains and losses are identified. Table 8 lists the 21 MC transporters that have no homologs in *S. cerevisiae* (as determined by an *E*-value $<10^{-35}$). All of them correspond to novel MC subfamilies of undetermined substrates. Compared to *S. cerevisiae*, a total of 11 novel subfamilies of mitochondrial carriers are identified in the four other Hemiascomycete species. The species that have lost the site 1 of oxidative phosphorylation (*K. lactis*, *C. glabrata*, and *S. cerevisiae*) do not seem to have been submitted to particular gain or loss of carriers. Noteworthy is the presence of 10 mitochondrial carriers in *Y. lipolytica*; most are also present in *N. crassa* but are lost in *D. hansenii* and afterward. All have undetermined substrates. Both *Y. lipolytica* (2.A.29.Y19) and *C. glabrata* (2.A.29.Y15) contain a species-specific MC subfamily, also of undetermined substrate.

DISCUSSION

Transmembrane transporters make up to 10% of the proteins encoded by the yeast genomes. They may be considered as metabolic checkpoints for complex anabolic or catabolic pathways. The gain or loss of transporters during speciation is an essential, and possibly initial, step in the differentiation of large intracellular metabolic pathways. Therefore the tracing of transporters during evolution may provide a vital, but rather

TABLE 8

Hemiascomycete mitochondrial carrier subfamilies that are absent in *S. cerevisiae*

Subfamily	*Y. lipolytica*	*D. hansenii*	*K. lactis*	*C. glabrata*
2.A.29.6	YALI0A20944g	DEHA0G14454g	KLLA0E02750g	
2.A.29.Y14		DEHA0E08349g	KLLA0A09383g	CAGL0F08305g
2.A.29.Y15				CAGL0B03883g
2.A.29.Y16	YALI0A16863g			
2.A.29.Y17	YALI0A20988g	DEHA0G19437g	KLLA0E09680g	
2.A.29.Y18	YALI0B05852g			
2.A.29.Y19	YALI0E33341g			
	YALI0F00418g			
2.A.29.Y20	YALI0F20262g	DEHA0E11022g		
2.A.29.Y21	YALI0F15609g	DEHA0B16401g		
		DEHA0E09691g		
2.A.29.Y22	YALI0E06897g			
2.A.29.Y23	YALI0D06798g			
ORF no.	10	6	3	2

neglected so far, panorama of the physiological and biochemical speciation processes.

To trace the evolution of transporters from the Hemiascomycete phylum we constructed a database (termed YETI) comprising all transporters within five complete genome sequences spanning ancient and more recent Hemiascomycete species, *Y. lipolytica, D. hansenii, K. lactis, C. glabrata,* and *S. cerevisiae,* using *N. crassa* as a Euascomycete reference outgroup. The YETI database uses the TCDB nomenclature and distinguishes the phylogenetic families, subfamilies, and clusters according to the TCDB five-digit system that we have adapted to cover all established transporters from the five studied species. Due to the large evolution span of the species used, we expect that the YETI database annexed in supplemental Table D1 (http://www.genetics.org/supplemental/) will provide a reference data set for the nonambiguous semiautomatic annotation of the vast majority of transporters from the novel Hemiascomycete genome sequences that are presently accumulating.

Numerous phylogenetic studies have established that the chronological order of emergence of the different species in the Hemiascomycete phylum is the following: *Y. lipolytica, D. hansenii, K. lactis, C. glabrata,* and *S. cerevisiae.* To study the speciation process under ideal conditions, one should compare the ancestral genomes of each of these species as they may have continued to evolve after separation from the phylum. These ancestral genomes are not available; we thus consider that the genome sequences of the presently available species provide a valid approximation of speciation events.

A total of 2480 transporters have been identified within the five species and classified into 204 phylogenetic subfamilies, characterized by the fourth digit, which has been adopted as the most significant level of analysis. We found that only 97 subfamilies comprising 355 individual transporters were submitted to the evolution drift. The statistical "variance" value of the size of these phylogenetic subfamilies within each species allowed us to identify those that had shown considerable expansion or contraction during speciation. The most variable subfamilies of transporters are listed in Table 3.

We also established that the emergence and loss of transporter subfamilies during speciation as well as their expansion or contraction in size can follow five different evolution patterns schematized in Table 6. The ubiquitous transporters are conserved in all Hemiascomycete species. The species-specific transporters are gained or lost in only one of the studied species. Other transporters are definitively lost or gained at a given step of the phylum evolution and these phylum gains or losses are definitively maintained thereafter. Homoplasic transporters have emerged and are lost more than one time independently during the evolution span under study.

The molecular mechanisms of emergences or losses of transporter genes during evolution are not understood and are probably diverse. It is, however, reasonable to assume that they generally result from physiological selection of diverging duplicated genes during adaptation to new ecological niches.

When more genome sequences become available, some transporter subfamilies that are considered here as species specific may be found shared by other species or subspecies. Similarly, transporters considered here as being definitively gained or lost only once may become homoplasic. Nevertheless we believe that the general methodology developed here and schematized in Table 6 will remain valid and may lead to the development of a new set of phylogenetic and classification criteria based on the presence or absence of a transporters subfamily. Such phylogenetic criteria may complement the classical systematic physiological parameters such as growth or fermentation on different substrates.

Full physiological analysis of our data is not possible here but some comments on a few striking results can be made. The emergence of the *Y. lipolytica* and *D. hansenii* species is accompanied by drastic gains of transporters compared to those identified during emergence of *K. lactis*, *C. glabrata*, and *S. cerevisiae*. For instance, a total of 154 transporters from the three large families, ACS (anion transporters), OPT (peptide transporters), and SP (sugar transporters), are encoded in *Y. lipolytica* and *D. hansenii* genomes whereas only 101 transporters from the same families are found in *C. glabrata* and *S. cerevisiae*. We may assume that physiological constraints produced by the nature of the available growth substrates may have driven the emergence of transporters. The drastic definitive loss of transporters before and after emergence of *D. hansenii* is also surprising and may be related to still undefined metabolic and physiological specificities. Another striking observation is the very limited number of species-specific transporters in *K. lactis* and the specific loss of several transporter subfamilies during emergence of *C. glabrata* or *Y. lipolytica*.

Some of our genomic findings could have been expected from known physiological traits such as the loss of the mitochondrial site 1 of oxidative phosphorylation between *D. hansenii* and *K. lactis*. Similarly, the proliferation of peroxisomal transporters and fatty acid-related proteins in *Y. lipolytica* may have been expected as this species is known to grow on fatty acids. The emergence of oligopeptide transporters in *Y. lipolytica* and *D. hansenii* may reflect the ability of species to grow in the absence of other nitrogen sources (AHEARN *et al.* 1968). The physiological drives underlining more unexpected genomic amplifications are puzzling, such as the amplification in *Y. lipolytica* and *D. hansenii* of anion:cation symporters (transporting anionic vitamins such as nicotinate and allantoate), of drug:H+ antiporter transporters, and of quinate:H+ symporters. The specific gain in *Y. lipolytica* of ABC transporters of the ABCA1 subfamily that so far had members only in mammalian cells is unexpected. Also the specific scarcity in *S. cerevisiae* of monocarboxylate porters of the sialate:H+ symport family (comprising only the Jen1p monocarboxylate transporter in *S. cerevisiae* compared to six transporters in *K. lactis*) is surprising but is congruent with some pioneering work (CASSIO *et al.* 1987). The scarcity of SITs in *C. glabrata* could be explained by the likely absence of siderophores in blood. The unexpected broad sequence divergence of mitochondrial transporters in *Y. lipolytica* and *D. hansenii* merits further scrutiny.

In brief, the tracing of the genes encoding transporters within related yeast species genomes has established the involvement of transporter protein in the evolution mechanism of speciation. We hope that the rigorous identification and annotation of transporter proteins through the YETI classification system and further elaboration of the different evolution patterns initi-ated here for the Hemiascomycete phylum will contribute to such investigations.

LITERATURE CITED

AHEARN, D. G., S. P. MEYER and R. A. NICHOLS, 1968 Extracellular proteinases of yeasts and yeast-like fungi. Appl. Microbiol. **16:** 1370–1374.

BOEKHOUT, T., 2005 Biodiversity gut feeling for yeasts. Nature **434:** 449–451.

BUSCH, W., and M. H. SAIER, 2002 The transporter classification (TC) system. Crit. Rev. Biochem. Mol. Biol. **37:** 287–337.

CASSIO, F., C. LEAO and N. VANUDEN, 1987 Transport of lactate and other short-chain monocarboxylates in the yeast Saccharomyces-cerevisiae. Appl. Environ. Microbiol. **53:** 509–513.

CLIFTEN, P., P. SUDARSANAM, A. DESIKAN, L. FULTON, B. FULTON *et al.*, 2003 Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Gene **301:** 71–76.

DE HERTOGH, B., E. CARVAJAL, E. TALLA, B. DUJON, P. V. BARET *et al.*, 2002 Phylogenetic classification of transporters and other membrane proteins from Saccharomyces cerevisiae. Funct. Integr. Genomics **2:** 154–170.

DE HERTOGH, B., E. TALLA, F. TEKAIA, E. BEYNE, D. SHERMAN *et al.*, 2003 Novel transporters from hemiascomycete yeasts. J. Mol. Microbiol. Biotechnol. **6:** 19–28.

DIETRICH, F. S., S. VOEGELI, S. BRACHAT, A. LERCH, K. GATES *et al.*, 2004 The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome. Gene **304:** 304–307.

DUJON, B., D. SHERMAN, G. FISCHER, P. DURRENS, S. CASAREGOLA *et al.*, 2004 Genome evolution in yeasts. Nature **430:** 35–44.

DURRENS, P., and D. J. SHERMAN, 2005 A systematic nomenclature of chromosomal elements for hemiascomycete yeasts. Yeast **22:** 337–342.

GALAGAN, J. E., S. E. CALVO, K. A. BORKOVICH, E. U. SELKER, N. D. READ *et al.*, 2003 The genome sequence of the filamentous fungus Neurospora crassa. Nature **422:** 859–868.

HITTINGER, C. T., A. ROKAS and S. B. CARROLL, 2004 Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proc. Natl. Acad. Sci. USA **101:** 14144–14149.

JONES, T., N. A. FEDERSPIEL, H. CHIBANA, J. DUNGAN, S. KALMAN *et al.*, 2004 The diploid genome sequence of Candida albicans. Proc. Natl. Acad. Sci. USA **101:** 7329–7334.

KELLIS, M., N. PATTERSON, M. ENDRIZZI, B. BIRREN and E. S. LANDER, 2003 Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature **423:** 241–254.

KELLIS, M., N. PATTERSON, B. BIRREN, B. BERGER and E. S. LANDER, 2004 Methods in comparative genomics: genome correspondence, gene identification and regulatory motif discovery. J. Comput. Biol. **11:** 319–355.

KROGH, A., B. LARSSON, G. VON HEIJNE and E. L. L. SONNHAMMER, 2001 Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. **305:** 567–580.

KURTZMAN, C. P., and J. W. FELL, 1998 *The Yeasts: A Taxonomic Study.* Elsevier Science, Amsterdam.

LANGKJAER, R. B., P. F. CLIFTEN, M. JOHNSTON and J. PISKUR, 2003 Yeast genome duplication was followed by asynchronous differentiation of duplicated genes. Nature **421:** 848–852.

OHNO, S., 1970 *Evolution by Gene Duplication.* Springer Verlag, Berlin/Heidelberg, Germany/New York.

ONISHI, T., A. KROGER, H. W. HELDT, E. PFAFF and M. KLINGENBERG, 1967 The response of the respiratory chain and adenine nucleotide system to oxidative phosphorylation in yeast mitochondria. Eur. J. Biochem. **1:** 303–311.

PHILIPPSEN, P., K. KLEINE, R. POHLMANN, A. DUSTERHOFT, K. HAMBERG *et al.*, 1997 The nucleotide sequence of Saccharomyces

cerevisiae chromosome XIV and its evolutionary implications. Nature **387:** 93–98.

Saier, M. H., 2000 A functional-phylogenetic classification system for transmembrane solute transporters. Microbiol. Mol. Biol. Rev. **64:** 354.

Souciet, J. L., M. Aigle, F. Artiguenave, G. Blandin, M. Bolotin-Fukuhara et al., 2000 Genomic exploration of the Hemiascomycetous yeasts. 1. A set of yeast species for molecular evolution studies. FEBS Lett. **487:** 3–12.

Tekaia, F., G. Blandin, A. Malpertuy, B. Llorente, P. Durrens et al., 2000 Genomic exploration of the Hemiascomycetous yeasts: 3. Methods and strategies used for sequence analysis and annotation. FEBS Lett. **487:** 17–30.

Tusnady, G. E., and I. Simon, 2001 The HMMTOP transmembrane topology prediction server. Bioinformatics **17:** 849–850.

Wolfe, K. H., and D. C. Shields, 1997 Molecular evidence for an ancient duplication of the entire yeast genome. Nature **387:** 708–713.

Communicating editor: F. Winston