

Research

The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes

Jonathan A Sheps^{*}, Steven Ralph^{*‡}, Zhongying Zhao[†], David L Baillie[†] and Victor Ling^{*}

Addresses: ^{*}British Columbia Cancer Research Centre, BC Cancer Agency, 601 West 10th Avenue, Vancouver BC, V5Z 1L6 Canada. [†]Department of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby BC, V5A 1S6 Canada. [‡]Current address: Genome BC and the Departments of Botany and Forest Sciences, University of British Columbia, 6270 University Blvd., Vancouver BC, V6T 1Z4 Canada.

Correspondence: Victor Ling. E-mail: vling@bccancer.bc.ca

Published: 11 February 2004

Genome **Biology** 2004, **5**:R15

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2004/5/3/R15>

Received: 13 October 2003

Revised: 27 November 2003

Accepted: 13 January 2004

© 2004 Sheps et al.; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: Many drugs of natural origin are hydrophobic and can pass through cell membranes. Hydrophobic molecules must be susceptible to active efflux systems if they are to be maintained at lower concentrations in cells than in their environment. Multi-drug resistance (MDR), often mediated by intrinsic membrane proteins that couple energy to drug efflux, provides this function. All eukaryotic genomes encode several gene families capable of encoding MDR functions, among which the ABC transporters are the largest. The number of candidate MDR genes means that study of the drug-resistance properties of an organism cannot be effectively carried out without taking a genomic perspective.

Results: We have annotated sequences for all 60 ABC transporters from the *Caenorhabditis elegans* genome, and performed a phylogenetic analysis of these along with the 49 human, 30 yeast, and 57 fly ABC transporters currently available in GenBank. Classification according to a unified nomenclature is presented. Comparison between genomes reveals much gene duplication and loss, and surprisingly little orthology among analogous genes. Proteins capable of conferring MDR are found in several distinct subfamilies and are likely to have arisen independently multiple times.

Conclusions: ABC transporter evolution fits a pattern expected from a process termed 'dynamic-coherence'. This is an unusual result for such a highly conserved gene family as this one, present in all domains of cellular life. Mechanistically, this may result from the broad substrate specificity of some ABC proteins, which both reduces selection against gene loss, and leads to the facile sorting of functions among paralogs following gene duplication.

Background

ATP-binding cassette (ABC) transporters are one of the largest families of transport proteins constituting the single

largest gene family, comprising about 5% of the genome, in *Escherichia coli* [1]. ABC transporters are grouped into several structural classes, or subfamilies, on the basis of amino

acid sequence and domain organization [2] (Figure 1). The presence of a strongly conserved ATP-binding motif defines membership in the family and the basic functional organization of an ABC transporter in the membrane is the same from bacteria to humans, and in all subclasses [3-5]. A complex of at least two ATP-binding domains, coupled to two blocks of membrane-spanning helices, appears to be the minimum requirement for a functional transporter. Often these domains are found in tandem within a single molecule, but in many cases are distributed across separate proteins that must then assemble in the membrane. ABC transporters are collectively able to accommodate an unusually large array of different substrates. This diversity of function is manifest at the family level, but also in individual members of the family, for example those associated with multidrug resistance (MDR).

Decottignies and Goffeau [6] catalogued the entire ABC transporter family of the yeast *Saccharomyces cerevisiae* and in so doing delineated six of the major subgroups of eukaryotic ABC transporters. Allikmets *et al.* [7] catalogued all the then known 33 human ABC transporters, including those known only from partial expressed sequence tag (EST) sequences, and divided these into seven subfamilies. This scheme has been adopted, with a revised nomenclature, by the Human Genome Organisation (HUGO) [8] in order to provide a unified nomenclature for both human and mouse ABC transporters. Of these seven subfamilies, one, ABCA, has no exact equivalent in the yeast genome [9,10]. Genes considered to be part of subfamily ABCA have been identified in the slime mold *Dictyostelium discoideum*, as well as in malaria

parasites [11] and *Caenorhabditis elegans* (this paper). With the completion of the human and *Drosophila melanogaster* genomes, a joint summary of the ABC transporter complements of both genomes was published [12]. This identified a new subfamily, ABCH, which appears to be the most divergent yet. One, previously unclassified yeast ABC gene, YDRO61w [13], appears to be a structurally aberrant member of subfamily H.

The phenotypes of five ABC transporter knockouts have been reported in *C. elegans*. Four of these involve genes expected, by homology to mammalian genes, to be involved in drug resistance: three P-glycoproteins (*Pgp-1*, *Pgp-3* and *Pgp-4*) (subfamily B) and one multi-drug resistance protein (MRP) [14,15] (subfamily C). These ABC transporter mutants are associated with sensitivity to environmental insult [16]. *Pgp-3* mutant strains of *C. elegans* are more sensitive to the drugs chloroquine and colchicine. *Pgp-1* and *mrp-1* strains are hypersensitive to toxic pigments produced by some bacteria [17]. All the nematode P-glycoproteins examined so far seem to be highly expressed in intestinal cells [18], and in the excretory cell, which functions somewhat like a kidney in *C. elegans*. The *mrp-1*, *pgp-1* and *pgp-3* mutant strains have been reported to be hypersensitive to the heavy metals cadmium and arsenite [15]. The fifth reported knockout is of the product of the *ced-7* gene [19]. Mutant alleles of *ced-7* cause a defect in engulfment of the cell corpses left behind by apoptosis. *ced-7* is a member of the ABCA subfamily, and has a similar phenotype to the *abca1* gene in humans. ABCA1 protein is required for engulfment of apoptotic cells by macrophages and is thought to regulate membrane fluidity through an increase in phosphatidylserine exposure on the outer leaflet of the cell membrane [20].

The term orthology is used to describe genes separated from one another by speciation events while paralogy describes those separated by gene duplication events [21]. Of particular interest, from the point of view of functional annotation, are the cases where a pair of genes, one from each of a pair of organisms, are found. In these cases it is reasonable to presume that the orthologous genes may share a conserved function retained from the same single gene present in the common ancestor of the two organisms. However, where a single gene (or set of duplicated genes) in one genome is most closely related to a set of duplicated (paralogous) genes in another genome this is sometimes termed co-orthology [22], and then no particular orthologous pair can be unambiguously specified. In the case of co-orthologs the argument for retention of analogous functions between members of the sets of descendant genes is much weaker. Comparison of two complete genomes, those of *C. elegans* and *S. cerevisiae* [23], demonstrated a high fraction of ortholog pairs in gene families involved in core biological functions. Specifically, Chervitz *et al.* [23] found, when pairing conserved yeast genes with their most similar worm homologs (subject to a BLAST score cut-off of $< 10^{-100}$), 57% of these highly conserved gene pairs

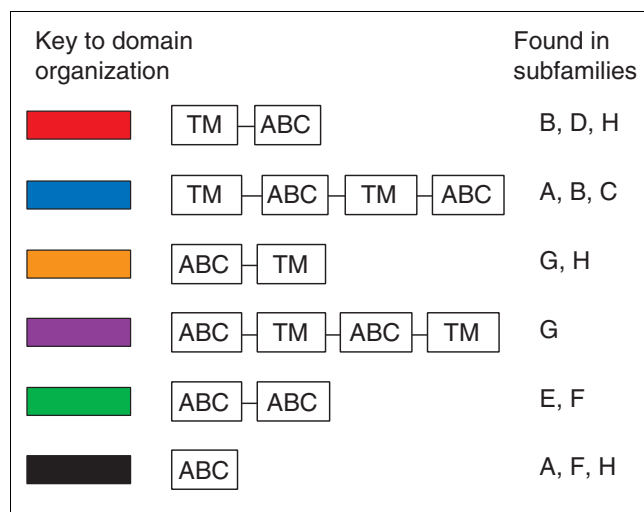


Figure 1
Structural diversity of ABC transporters. Illustration of the various domain organizations found among members of the ABC transporter family in *C. elegans*. TM indicates a transmembrane domain typically containing six predicted membrane-spanning helices. ABC indicates an ATP-binding cassette domain. The color codes for each structure are used throughout the figures to show the lack of concordance between structural categories and families defined on the basis of sequence similarity.

involved orthologous, rather than paralogous, pairs of genes. In this category of core functions they included trafficking, and, as possibly the largest family of trafficking genes in animal genomes, ABC transporters should be expected to share in this high level of one-to-one correspondence between genomes. We expected therefore that this would allow us to assign predicted functions to newly discovered *C. elegans* ABC proteins on the basis of their already-characterized mammalian orthologs. Following a comprehensive phylogenetic analysis of ABC transporters from four eukaryote genomes, we found that the frequency of orthologous pairs among ABC transporters was substantially lower than we expected. Particular domain organizations and substrate specificities seem to have evolved independently several times in multiple lineages. This is expected to complicate the functional analysis of ABC transporter function in newly characterized genomes.

Results and discussion

Here we present a classification of all ABC transporters encoded in the *C. elegans* genome, based on a phylogenetic analysis which includes the 49 currently known human ABC proteins for which there are reliable, public, sequence data. We took the approach of analyzing primarily the conserved ATP-binding cassettes from each protein, regardless of the structural class from which the domain is drawn. This allows evaluation of the evolutionary history of each protein in the family, without biases that might result from gene-fusion events resulting in convergent acquisition of similar domain structures by distantly related proteins. In addition, we re-evaluated the relationships of transporters within statistically reliable clusters whose members are closely related enough that structural variations do not lead to errors in alignment. We did this to capture additional phylogenetic information, which may be apparent in the less conservative transmembrane domains, at a level of analysis where it is less likely to be misleading.

An example of our first-pass approach is given in Figure 2, which shows an analysis of isolated ATP-binding cassette domains from the human ABC transporters only. In particular, we find that all seven subfamilies recognized by Allikmets *et al.* [7] are recovered with significant bootstrap support. Their finding, that subfamily B is more closely related to the carboxy-terminal component of subfamily C than the two halves of ABCC molecules are with one another, is supported by our results.

A collection of transporters

We found a total of 60 confirmed ABC transporters in the annotated protein set derived from the *C. elegans* genome sequence. This represents approximately 0.3% of the total number of genes (approximately 19,000) in the worm genome. Only 8 of the 60 predicted genes lack any corresponding mRNA (Table 1), and only one (F56F4.6) is

structurally aberrant in a way that would suggest it is likely to be a pseudogene.

Thirty ABC transporters are described in the yeast genome, or approximately 0.5% of its approximately 6,000 proteins [13]. At present 49 human ABC transporters have been identified and, at least partially, cloned. They are included here (Figures 3,4,5,6,7 and Table 2) to illustrate their relationships with nematode proteins, which might then shed light on their biological roles. Inclusion of human as well as *D. melanogaster* ABC transporters in our tree allows us to explicitly classify *C. elegans* ABC transporters according to the current eight-subfamily taxonomic scheme for ABC transporters [12].

Typing ABCs to subfamily

We define membership of a particular gene in an ABC transporter subfamily primarily on the basis of the position of its ATP-binding domains in our first phylogenetic tree (not shown). Genes that fell unambiguously within a clade containing genes already assigned to given subfamily, were included in that subfamily. Where we could not assign a gene to a particular clade with a significant bootstrap value, the assignment was made on the basis of which subfamily's members scored highest when that gene was used as query in a BLAST search. The subfamilies are sometimes named according to the well-characterized mammalian genes that typify each of them, for example, P-gp (P-glycoprotein), MRP, White gene homologs, RNase L inhibitor, GCN20 homologs, ABC1 and ALDP [7]. These correspond to the HUGO-defined subfamilies B, C, G, E, F, A and D, respectively. Re-analysis of the full-length sequences confirmed the placement all *C. elegans* genes within the preexisting subfamilies, with substantial bootstrap support (Figures 3,4,5,6,7).

Instances of orthology

In the set of worm and human ABC transporters, only 8 of 49 possible pairs (16%) of sister genes contained a single human protein and a nematode homolog (Table 3). Similarly, 10% of ABC transporters were found in orthologous pairs when the comparison is made between yeast and worm genomes. A more comprehensive comparison of worm and yeast genomes [23] came to the overall conclusion that 57% of genes in highly conserved gene families were found in orthologous pairs, and the study suggested that such gene families provide a conserved core proteome which forms the basis of eukaryote biochemistry. ABC transporters are conserved in all eukaryotic and prokaryotic genomes, so it is interesting to note that they are found in orthologous pairs much less frequently than most gene families that are roughly as well conserved. Clearly, ABC transporter evolution has not been typical of strongly conserved gene families, and while we might have inferred that ABC-transporter-mediated metabolism differs radically among eukaryotes, this seems improbable, given the broadly comparable set of substrates associated with ABC transporters in all eukaryotes where they have been studied.

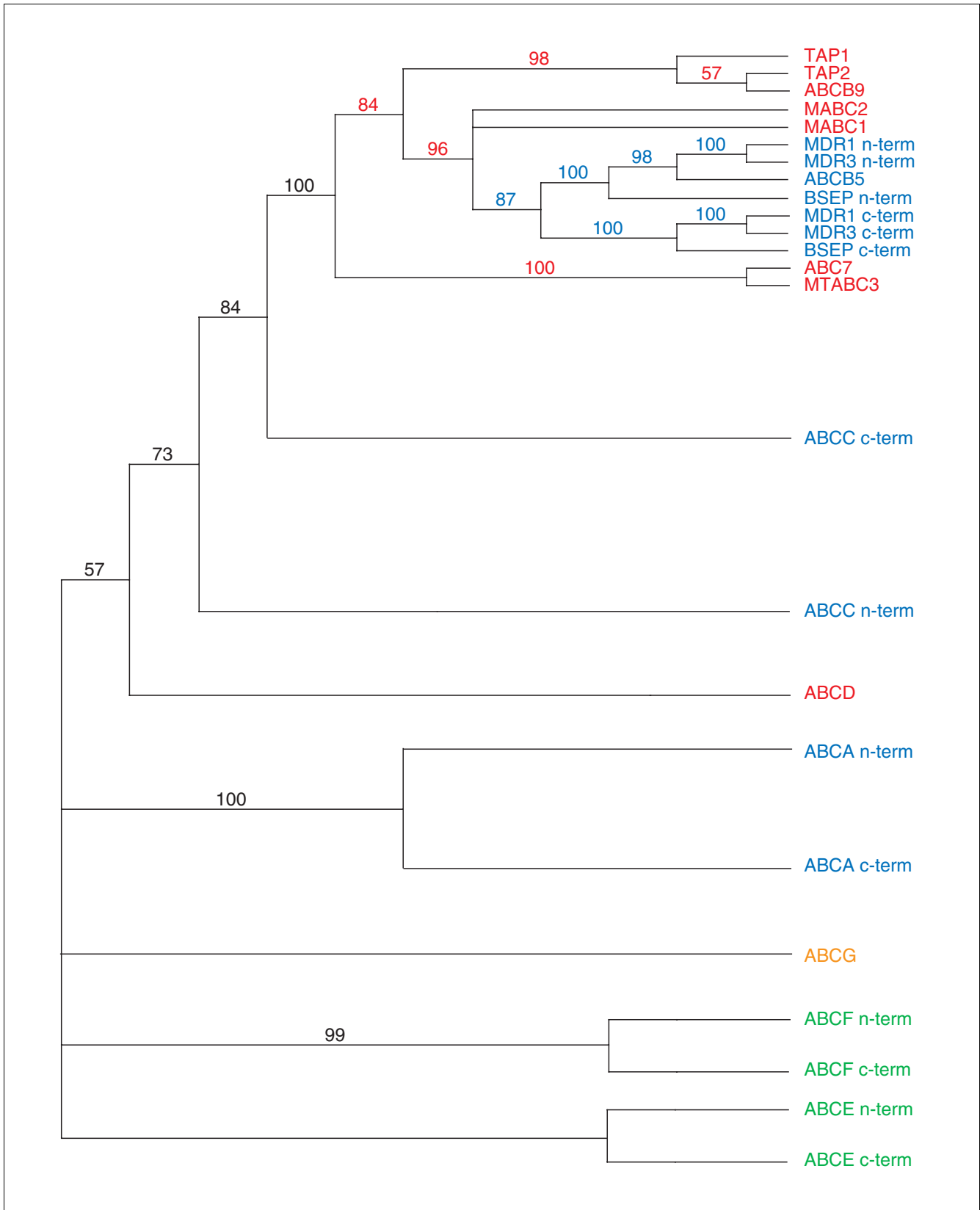


Figure 2 (see legend on next page)

Figure 2 (see previous page)

Tree of human ATP-binding cassette domains. The evolution of the ABCB subfamily from within the ABCC subfamily, and the structural diversity of subfamily B is shown here. Each cluster of ABC domains within each subfamily, except for subfamily B, is collapsed to form a single, representative, branch; n-term: amino-terminal ABC; c-term: carboxy-terminal ABC. The phylogeny of ATP-binding cassettes from human ABC transporters was produced according the following procedure. Predicted amino-acid sequences were aligned using ClustalX [54]. Aligned sequences were used to generate matrices of mean distances among proteins, and these matrices were used to generate a phylogenetic tree according to the neighbor-joining algorithm [55], refined using the SPR branch-swapping technique under the minimum evolution criterion, implemented by PAUP*4.0b10 [56]. Bootstrapping [57] was used to determine the relative support for the various branches of the tree (1,000 replicates), and nodes with less than 50% support were collapsed to form polytomies. The structures of the proteins in which the domains are embedded are indicated according to the color scheme in Figure 1. It should be noted that branch lengths in the figures are not to scale and do not represent distances between protein sequences. The original alignment files are available as Additional data files 1-8.

Within the P-gp-related ABCB subfamily, the only one-to-one pairings found between *C. elegans* and human genes are those of W09D6.6 (Haf-5) and MTABC3 (B6), and Y48G8AL.11 (Haf-6) and MABC1 (B8). These are both half-transporters localized to the mitochondria. MTABC3 (B6) is involved in iron homeostasis [24] and its rat ortholog, PRP, is overexpressed during hepatocarcinogenesis [25]. Two other mitochondrial ABC transporters in humans, MABC2 (B10) and ABCB7, have orthologs in flies and/or yeast, but not nematodes.

Among ABCC molecules, whose range of functions broadly overlaps with P-gps, only C18C4.2 (Cft-1) and CFTR (C7) are indicated as orthologs in our analysis. However, the bootstrap value on this pairing is very low (51%, see Figure 5), so we cannot attach much confidence to this observation. It may simply be that C18C4.2 (Cft-1) is a highly divergent member of subfamily C, and does not bear much functional similarity to CFTR (C7). Although not forming simple pairs with any nematode gene, human MRP5 (C5), a transporter of nucleotide analogs [26,27], and ABCC11 and ABCC12 appear to be co-orthologous to worm F14F4.3 (Mrp-5), which may provide some hint as to the function of the latter.

All four of the *C. elegans* members of subfamilies E and F (Figure 6) form strongly supported and unambiguous pairs with their homologs in *D. melanogaster*, *Homo sapiens*, and yeast. This unusually strong conservation, compared to the other subfamilies of ABC genes, argues for involvement in something indispensable, at least on an evolutionary timescale. The three genes in subfamily F, which lack transmembrane domains, are generally regarded as forming ribosome associated proteins involved in regulation of mRNA translation, rather than transporters. The RNase L inhibitor (E1), also known as the oligoadenylate-binding protein (OABP), is thought to be involved in the regulation of the interferon-induced antiviral response [28] that bears some similarities to the mechanism thought to underlie the now common molecular biology technique of double-stranded RNA-directed interference (RNAi). It also seems to have a role in muscle differentiation [29] in mammals. The critical role of the RNase L inhibitor is underlined by its conservation even in a highly reduced genome. In the rather minimal genome of the endosymbiotic *Guillardia theta* nucleomorph (302 genes)

the RNase L inhibitor is the only ABC protein found [30]. The yeast ortholog of the RNase L inhibitor protein, YDR091c, is essential for growth, as is YER036c, the yeast ortholog of T27E9.7/ABCF2 [31]. On the other hand GCN20, the yeast version of F42A10.1/ABCF3, is not essential, although mutants do have specific defects in translation.

Processes of gene duplication and loss

While the conservation of simple orthologous gene pairs is a rare observation in our study, the numbers of genes in most ABC transporter subfamilies are about the same, despite numerous instances of gene duplication and loss. For example, within ABCB the number of half-transporters in each genome is almost identical. Furthermore, most mammalian half-transporters in subfamily B are found in clusters of functionally related, or at least co-localized, genes (the TAP (B2 and B3) genes, and the four mitochondrial ABCB genes, MABCs1 and 2 (B8 and B10), MTABC3 (B6) and ABCB7 [32]), paired with similarly sized groups of *C. elegans* genes. Likewise the number of genes in subfamilies A, C and D is much the same between genomes. However, it does appear that *C. elegans*, relative to humans, has undergone a massive expansion in the P-gp (full or pseudo-dimer configuration) subclass of subfamily B, and subfamily G, the 'White-like' genes. The likelihood that ABC transporter lineages have been lost repeatedly in evolution is evident from the phylogeny. The single group of P-gps in mammals contains only four members, while *C. elegans* has 15 P-gps, of which only three are closely related to their mammalian homologs. A literal reading of the tree (Figure 4) would suggest the presence of five additional P-gp lineages in the common ancestor of nematodes, flies and mammals that have been lost, independently, in both mammals and flies. These losses, and the species-specific expansion of the remaining lineages of genes, underlines the peculiarly dynamic composition of this group of multifunctional transport proteins.

Conclusions

The completion of the *C. elegans* and *D. melanogaster* genome projects [33,34] make it possible to analyze entire gene families in metazoans. The advantage of performing a combined analysis of all known ABC proteins from two organisms is that it allows unambiguous identification of

Table 1**Characterization of the 60 *C. elegans* ABC proteins**

Subfamily	ORF name/CGC name	Chromosome	GenBank accession number	Size (amino acids)	Predicted topology	cDNA if known	RNAi phenotype	
A Abt	C24F3.5/ Abt-1	IV	CAA18775	1,429	(6TM-ABC)2		None	
	C48B4.4/ Ced-7	III	NP_499115	1,704	(8TM-ABC)2	Complete	None	
	F12B6.1/ Abt-2	I	AAB54153	1,547	(6TM-ABC)2	Partial	None	
	F55G11.9/ Abt-3	IV	CAB05222	1,431	(8/4TM-ABC)2		None	
	F56F4.6	I	AAB54203	260	ABC		None	
	Y39D8C.1/ Abt-4	V	AAC69223	1,802	(6/8TM-ABC)2	Partial	None	
	Y53C10A.9/ Abt-5	I	CAA22142	1,564	(6TM-ABC)2	Partial	None	
B Pgp (full molecules)	C05A9.1/ Pgp-5	X	CAA94202	1,283	(6TM-ABC)2	Partial	None	
	C34G6.4/ Pgp-2	I	AAB52482	1,265	(6TM-ABC)2	Partial	None	
	C47A10.1/ Pgp-9	V	CAB03973	1,294	(6TM-ABC)2	Partial	None	
	C54D1.1/ Pgp-10	X	AAC48149	1,283	(4TM-ABC)2	Partial	None	
	DH11.3/ Pgp-11	II	CAA88940	1,270	(6TM-ABC)2	Partial	None	
	F22E10.1/ Pgp-12	X	CAA91799	1,318	(6TM-ABC)2	Partial	None	
	F22E10.2/ Pgp-13	X	CAA91800	1,291	(6TM-ABC)2		None	
	F22E10.3/ Pgp-14	X	CAA91801	1,327	(6TM-ABC)2	Partial	None	
	F22E10.4/ Pgp-15	X	CAA91802	1,270	(6TM-ABC)2		None	
	F42E11.1/ Pgp-4	X	CAA91463	1,266	(6TM-ABC)2	Partial	None	
	K08E7.9/ Pgp-1	IV	CAB01232	1,321	(6TM-ABC)2	Partial	None	
	T21E8.1/ Pgp-6	X	CAA94220	1,225	(6TM-ABC)2	Partial	None	
	T21E8.2/ Pgp-7	X	CAA94219	1,269	(6TM-ABC)2		None	
	T21E8.3/ Pgp-8	X	CAA94203	1,243	(6TM-ABC)2	Partial	None	
	ZK455.7/ Pgp-3	X	CAA91467	1,268	(6TM-ABC)2	Partial	None	
	Haf (half molecules)	C30H6.6/ Haf-1	IV	CAB02812	586	4TM-ABC	Partial	None
		F43E2.4/ Haf-2	II	AAC71121	761	8TM-ABC	Partial	None
		F57A10.3/ Haf-3	V	CAB09418	733	6TM-ABC	Partial	None
		W04C9.1/ Haf-4	I	AAC68724	787	8TM-ABC	Complete	Weak embryonic lethality, slow growth
		W09D6.6/ Haf-5	III	CAB04947	801	8TM-ABC	Complete	None
Y48G8AL.11/ Haf-6		I	AAK29911	565	4TM-NBF	Partial		
Y50E8A.16/ Haf-7		V	CAB60586	807	6TM-ABC	Partial		
Y57G11C.1/ Haf-8		IV	CAB16503	633	4TM-ABC		None	
ZK484.2/ Haf-9		I	AAK39394	815	8TM-ABC	Complete	None	
C Mrp/Cft		C18C4.2/ Cft-1	V	AAK52175	1247	(5/6TM-NBF)2	Partial	None
	E03G2.2/ Mrp-3	X	CAA92148	1,398	(6TM-ABC)2	Partial	None	
	F14F4.3/ Mrp-5	X	CAB54225	1,427	(6TM-ABC)2	Partial	Slow growth, Clear	
	F20B6.3/ Mrp-6	X	AAA82317	1,396	(6TM-ABC)2	Partial	Egg laying defect	
	F21G4.2/ Mrp-4	X	CAB02667	1,573	(10/6TM-ABC)2	Partial	None	

Table I (Continued)**Characterization of the 60 *C. elegans* ABC proteins**

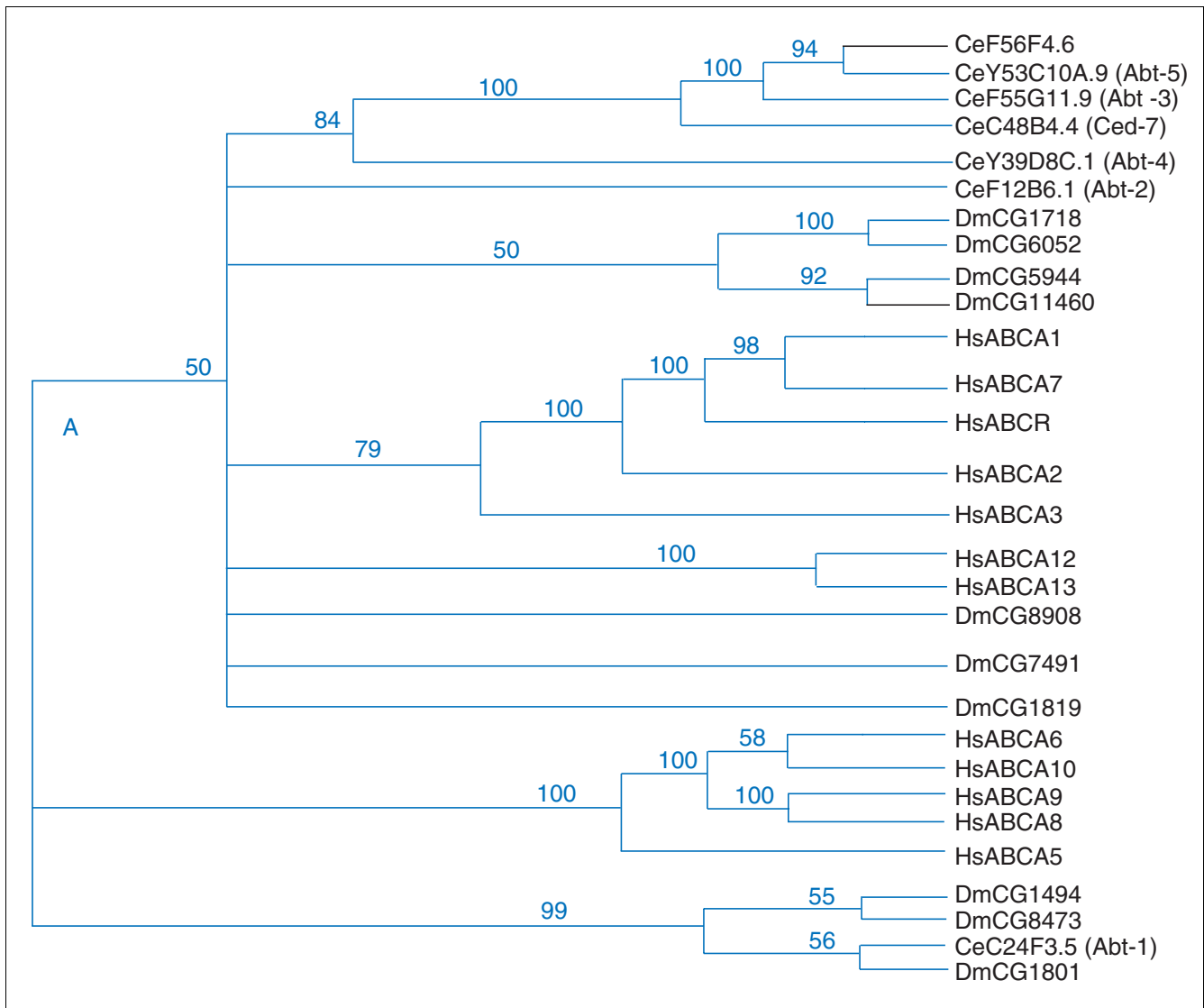
	F57C12.4/ Mrp-2	X	AAB07022	1,525	(10/6TM-ABC) ₂	Complete	None
	F57C12.5/ Mrp-1	X	AAD31550	1,528	(12/6TM-ABC) ₂	Complete	None
	Y43F8C.12/ Mrp-7	V	CAA21622	1,119	(12/2TM-ABC) ₂	Partial	
	Y75B8A.26/ Mrp-8	X	CAA22110	1,144	(4/6TM-ABC) ₂	Partial	
D	C44B7.8	II	AAA68339	665	4TM-ABC	Partial	
	C44B7.9	II	AAA68340	661	4TM-ABC	Partial	None
	C54G10.3	V	CAA99810	660	6TM-ABC	Complete	None
	T02D1.5	IV	CAB0590	734	6TM-ABC	Partial	None
	T10H9.5	V	AAC19238	598	6TM-ABC	Complete	None
E	Y39E4B.1	III	CAB54424	610	ABC-ABC	Partial	Embryonic lethality
F	F18E2.2	V	CAA99835	622	ABC-ABC	Partial	None
	F42A10.1	III	AAA19072	712	ABC-ABC	Partial	None
	T27E9.7/ GCN20-2	III	CAB04880	622	ABC-ABC	Complete	None
G	C05D10.3	III	AAA20989	598	ABC-4TM	Partial	None
	C10C6.5	IV	CAB05682	610	ABC-6TM	Partial	None
	C16C10.12	III	CAA86750	610	ABC-4TM	Partial	None
	F02E11.1	II	AAB66050	658	ABC-4TM	Partial	None
	F19B6.4	IV	CAA93461	695	ABC-6TM	Partial	None
	T26A5.1	III	AAC77504	608	ABC-4TM	Partial	None
	Y42G9A.6	III	AAF60554	684	ABC-6TM	Partial	None
	Y47D3A.11	III	CAB57891	547	ABC-6TM	Partial	None
	Y49E10.9	III	CAB11549	454	ABC-4TM		None
H	C56E6.1	II	AAA81093	1,667	ABC-12TM		Larval arrest
	C56E6.5	II	AAA81094	595	ABC-6TM	Partial	None

Subfamily names given are according to the HUGO nomenclature (A-H) [33] as well as the CGC (*Caenorhabditis* Genetics Centre [61]) gene names for each subfamily. TM, transmembrane domain, where the number preceding it is the predicted number of membrane spanning helices or the number in the amino-terminal/carboxy-terminal TM domains, respectively. ABC, ATP-binding cassette. The existence of known cDNAs, whether complete or partial, is listed according to information in WormBase release WS112 [62]. RNAi phenotypes of genes on chromosome I are given according to [63], those on chromosomes II, IV, V and X are from [64], those of genes on chromosome III are from [65].

orthologous pairs of genes, as well as allowing the pattern of evolution by a process of gene duplication, lineage sorting, and functional convergence to be explicitly modeled.

Saurin *et al.* [35] surveyed the ABC transporters, considering both eukaryotic and prokaryotic systems, and found that there is a fundamental phylogenetic division among ABC transporters involved in import versus export processes. The importer class of ABCs is found only in prokaryotes, whereas exporters are found in all domains of life [35]. However, that survey, while covering all classes of ABC transporter, was not

comprehensive with respect to any of the organisms surveyed. Most recently, Schriml and Dean [10] compared the human ABC family to that of the mouse *Mus musculus*, and found almost perfect identity between the two genomes. We have integrated previous information with the complete inventory of ABC transporters from the genome of the nematode worm *C. elegans*. We find that most of the ABC transporters in the worm can be classified into the existing human transporter taxonomy. We find 60 ABC transporters in the worm genome, representing an overall doubling in size of the ABC transporter family relative to yeast, whose genome

**Figure 3**

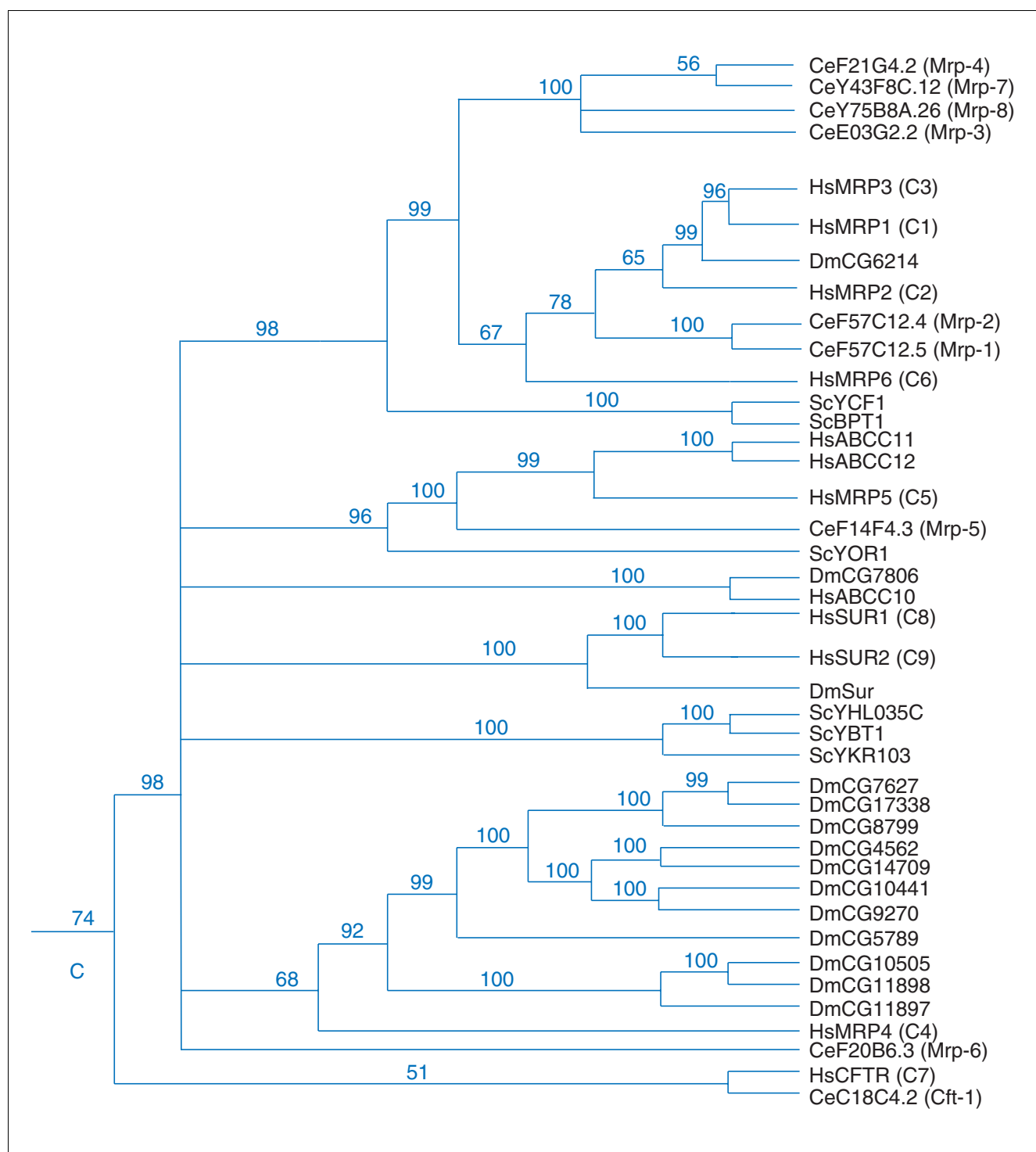
Phylogenetic tree of ABCA proteins in three eukaryote genomes. A phylogeny derived and displayed according to the procedure outlined in the legend to Figure 2, except that complete protein sequences were used, not just those of the ATP-binding cassettes. The genome of origin for each protein is indicated by prefixes before each gene name, according the following scheme: Ce, *C. elegans*; Dm, *D. melanogaster*; Hs, *H. sapiens*; Sc, *S. cerevisiae*.

contains one third as many protein-coding genes. No ABC genes were found that could be classified among the bacterial import proteins.

At least three subfamilies of ABC transporter contain members capable of a conferring an MDR phenotype, and transporters from at least two different subfamilies cause

Figure 4 (see following page)

Phylogenetic tree of ABCB proteins in four eukaryote genomes. A phylogeny derived and displayed according to the procedure outlined in the legend to Figure 3. Shown here is the division between the half transporters, which are most of the ABCB genes in mammals, and the full-transporters (called P-glycoproteins (P-gps)) that have evolved from them. Four lineages of P-gps (exemplified by genes F22E10.1-4, T21E8.1-3, C47A10.1 and C54D1.1) have been lost in both flies and mammals, and of the two remaining P-gp lineages, one has been lost in each of the fly and human lines of descent. Subsequent duplications within the single remaining P-gp lineage in both flies and mammals have not been sufficient to keep pace with continuing P-gp duplications in the worm genome.

**Figure 5**

Phylogenetic tree of ABCC proteins in four eukaryote genomes. A phylogeny derived and displayed according to the procedure outlined in the legend of Figure 3.

MDR in human tumors [36]. A multi-drug transporter is a single protein capable of specifically recognizing several structurally distinct classes of compounds, and which catalyzes their efflux from the cell or sequestration in a

subcellular compartment. Proteins of the P-glycoprotein (P-gp) group (ABCB) transport hydrophobic compounds and function in transport of lipids and bile from the liver as well as generally defending the body from toxic natural products

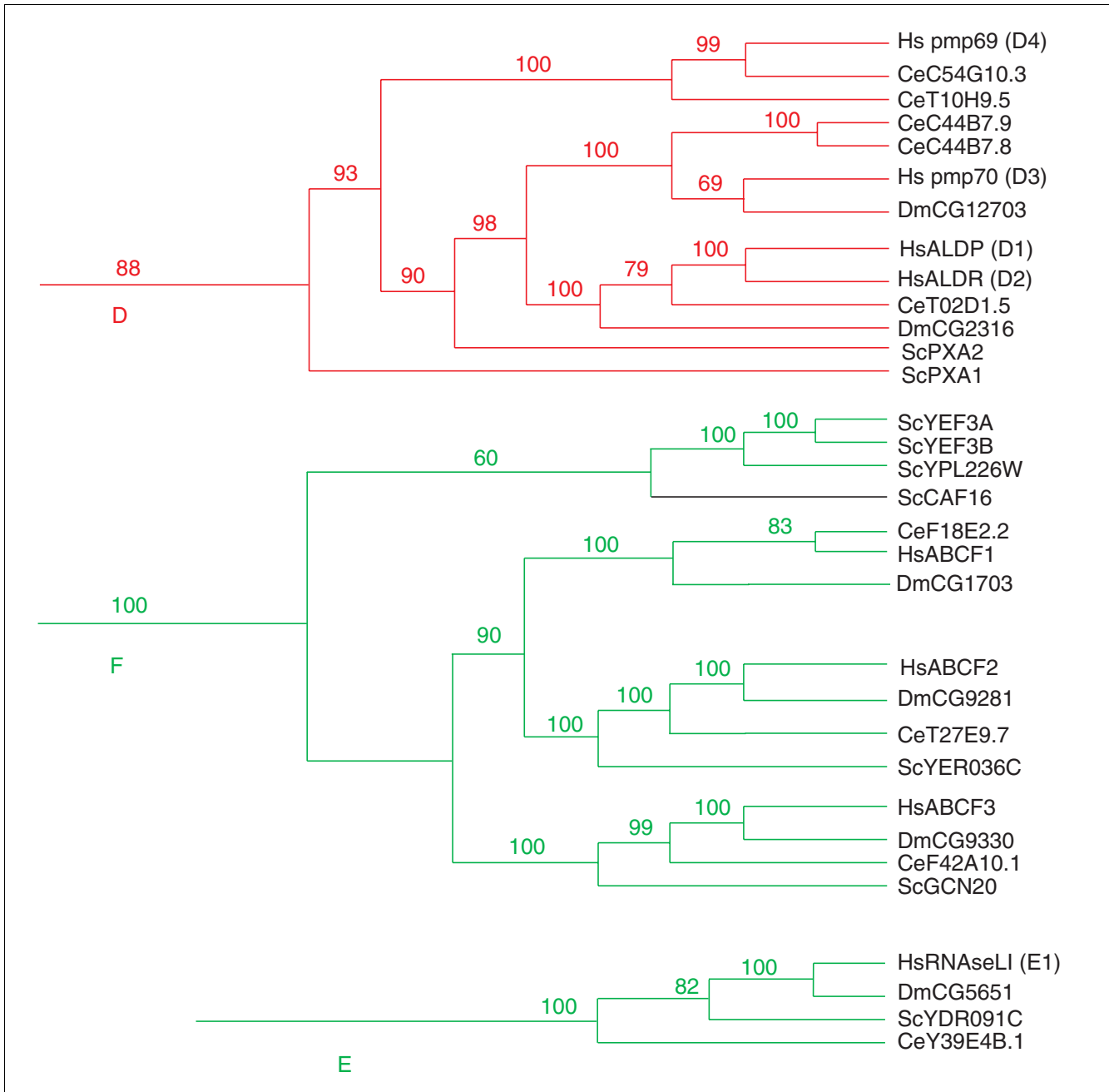


Figure 6
Phylogenetic trees of ABCD, ABCE, and ABCF proteins in four eukaryote genomes. Phylogenies derived and displayed according to the procedure outlined in the legend of Figure 3.

in the diet [37]. P-gps are also a component of the blood-brain barrier and function in tolerance of drugs normally minimally toxic to mammals, such as ivermectin [38]. Multi-drug resistance mediated by MRP group (ABCC) proteins depends on a slightly different mechanism. MRPs seem to function by co-transporting toxic compounds with glutathione, or as glutathione conjugates [36]. An MDR phenotype is also associated with some members of the ABCG group of transporters, in both yeast [39] and humans [40]. The MDR phenotype

appears to have evolved not just once, but at least three times in the history of ABC transporters. Given the distribution of MDR-causing and non-MDR genes among mammalian P-gps; it seems reasonable to infer that MDR genes may well have arisen more than once among the P-gps themselves. It has been observed [41,42] that the entire ABC transporter family is characterized by a highly adaptable common mechanism for coupling substrate binding to ATP hydrolysis and extrusion. It has been pointed out that, because P-gp

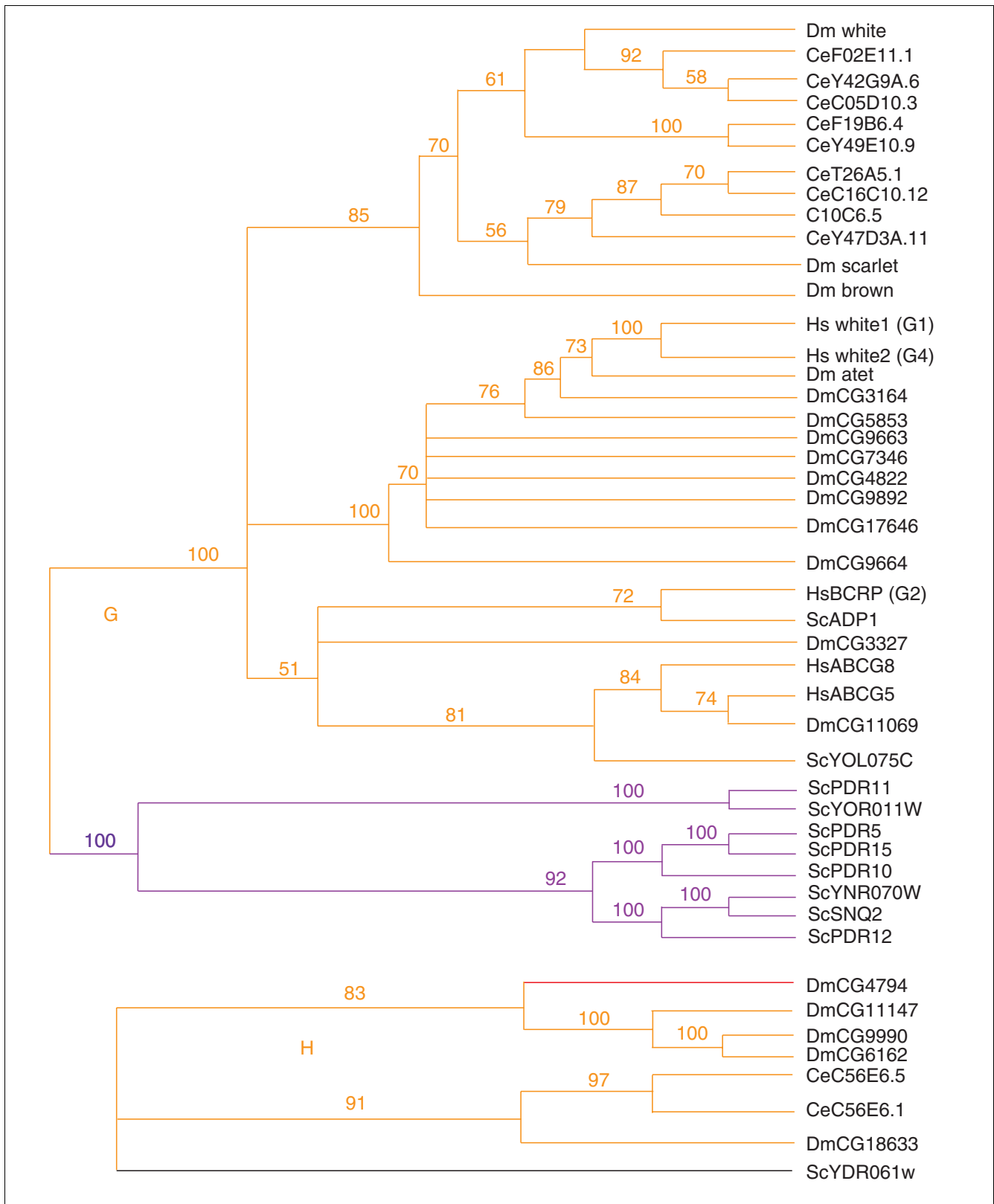


Figure 7
Phylogenetic trees of ABCG and ABCH proteins in four eukaryote genomes. Phylogenies derived and displayed according to the procedure outlined in the legend of Figure 3.

Table 2**Alphabetic list, by taxon, of protein sequences used in this study**

<i>S. cerevisiae</i>	Accession number	<i>D. melanogaster</i>	Accession number	<i>C. elegans</i>	Accession number	<i>H. sapiens</i>	Accession number
ADPI	NP_009937	I7ID11.2	AAF45509	C05A9.1	CAA94202	ABCA1	NP_005493
ATMI	NP_014030	Atet	AAF51027	C05D10.3	AAA20989	ABCA2	NP_001597
BPT1	NP_013086	Brown	AAF47020	C10C6.5	CAB05682	ABCA3	CAA65825
CAF16	NP_116625	CG10226	AAF50670	C16C10.12	CAA86750	ABCA5	NP_061142
GCN20	NP_116664	CG10441	AAF53737	C18C4.2	AAK52175	ABCA6	NP_525023
MDL1	NP_013289	CG10505	AAF46706	C24F3.5	CAA18775	ABCA7	AF328787
MDL2	NP_015053	CG11069	AAF56361	C30H6.6	CAB02812	ABCA8	AB020629
PDR10	NP_014973	CG11147	AAF52284	C34G6.4	AAB52482	ABCA9	NP_525022
PDR11	NP_012252	CG11460	AAF55727	C44B7.8	AAA68339	ABCA10	XP_085647
PDR12	NP_015267	CG11897	AAF56869	C44B7.9	AAA68340	ABCA12	NP_056472
PDR15	NP_010694	CG11898	AAF56870	C47A10.1	CAB03973	ABCA13	NP_689914
PDR5	NP_014796	CG12703	AAF49018	C48B4.4	CAA82384	ABCB5	AAO73470
PXA1	NP_015178	CG14709	AAF54656	C54D1.1	AAC48149	ABCB7	AB005289
PXA2	NP_012733	CG1494	AAF50838	C54G10.3	CAA99810	ABCB9	AC002486
SNQ2	NP_010294	CG1703	AAF48069	C56E6.1	AAA81093	ABCC10	NP_258261
Ste6	NP_012713	CG1718	AAF50837	C56E6.5	AAA81092	ABCC11	NP_149163
YBT1	NP_013052	CG17338	AAF53736	DH11.3	CAA88940	ABCC12	NM_033226
YCF1	NP_010419	CG17646	AAF51341	E03G2.2	CAA92148	ABCC13	NP_742021
YDR061w	NP_010346	CG1801	AAF50836	F02E11.1	AAB66050	ABCF1	AAH34488
YDR091C	NP_010376	CG1819	AAF50847	F12B6.1	AAB54153	ABCF2	NP_005683
yEF3	NP_013350	CG1824	AAF48177	F14F4.3	CAB54225	ABCF3	NP_060828
yEFB	P53978	CG18633	AAF56360	F18E2.2	CAA99835	ABCG5	AF320293
YER036C	NP_010953	CG2316	AAF59367	F19B6.4	CAA93461	ABCG8	AF320294
YHL035C	NP_011828	CG3164	AAF51548	F20B6.3	AAA82317	ABCR (A4)	AF001945
YKR103VW	NP_013030	CG3327	AAF51122	F21G4.2	CAB02667	ALDP (D1)	CAA79922
YNR070w	NP_014468	CG4225	AAF55241	F22E10.1	CAA91799	ALDR (D2)	NP_005155
YOL075C	NP_014567	CG4562	AAF55707	F22E10.2	CAA91800	BCRP (G2)	XP_032425
YOR011w	NP_878167	CG4794	AAF55726	F22E10.3	CAA91801	BSEP (B11)	AF091582
YORI	NP_011797	CG4822	AAF51552	F22E10.4	CAA91802	CFTR (C7)	AAC13657
YPL226W	S65245	CG5651	AAF50342	F42A10.1	AAA19072	MABC1 (B8)	AF047690
		CG5789	AAF56312	F42E11.1	CAA91463	MABC2 (B10)	XP_001871
		CG5853	AAF52835	F43E2.4	AAC71121	MDR1 (B1)	4505769
		CG5944	AAF49305	F55G11.9	CAB05222	MDR3 (B4)	AAA36207
		CG6052	AAF49312	F56F4.6	AAB54203	MRP1 (C1)	AAB46616
		CG6162	AAF56584	F57A10.3	CAB09418	MRP2 (C2)	CAA65259
		CG6214	AAF53223	F57C12.4	AAB07022	MRP3 (C3)	AB010887
		CG7346	AAF50035	F57C12.5	AAD31550	MRP4 (C4)	NP_005836
		CG7491	AAF53328	K08E7.9	CAB01232	MRP5 (C5)	AAB71758
		CG7627	AAF52648	T02D1.5	CAB05909	MRP6 (C6)	AF076622
		CG7806	AAF52639	T10H9.5	AAC19238	MTABC3 (B6)	NP_005680
		CG7955	AAF47525	T21E8.1	CAA94220	PMP69 (D4)	AF009746
		CG8473	AAF48511	T21E8.2	CAA94219	PMP70 (D3)	CAA41416
		CG8799	AAF58947	T21E8.3	CAA94203	RNAse LI (E1)	CAA53972
		CG8908	AAF57490	T26A5.1	AAC77504	SUR1 (C8)	AAB02278
		CG9270	AAF53950	T27E9.7	CAB04880	SUR2 (C9)	AF061323
		CG9281	AAF48493	W04C9.1	AAC68724	TAPI (B2)	CAA40741
		CG9330	AAF49142	W09D6.6	CAB04947	TAP2 (B3)	AAA59841

Table 2 (Continued)**Alphabetic list, by taxon, of protein sequences used in this study**

CG9663	AAF51130	Y39D8C.1	AAC69223	WHITE 1 (G1)	AAC51098
CG9664	AAF51131	Y39E4B.1	CAB54424	WHITE 2 (G4)	NP_071452
CG9892	AAF51223	Y42G9A.6	NP_498332		
CG9990	AAF56807	Y43F8C.12	CAA21622		
Mdr49	AAF58437	Y47D3A.11	CAB57891		
Mdr50	AAF58271	Y48G8AL.11	AAK29911		
Mdr65	AAF50669	Y49E10.9	CAB11549		
Scarlet	AAF49455	Y50E8A.16	CAB60586		
Sur	AAF52866	Y53C10A.9	CAA22142		
White	AAF45826	Y57G11C.1	CAB16503		
		Y75B8A.26	CAA22110		
		ZK455.7	CAA91467		
		ZK484.2	AAK39394		

recognizes substrate directly within the cytoplasmic leaflet of the plasma membrane [43], it does so at a much higher effective substrate concentration than would be the case if it recognized aqueous substrate. As a result, P-gp drug-binding sites can operate at relatively low affinity, and this, in turn, facilitates recognition of multiple substrates. This flexibility may be the key to explaining the range of tasks performed by ABC transporters, but also their apparently anomalous evolutionary history.

The mammalian P-gps include proteins capable of producing an MDR phenotype (MDR1 (B1)), as well as members with, apparently, specificity restricted to single physiological substrates such as phosphatidylcholine (MDR3 (B4)). As none of these have simple, orthologous, relationships with any of the *C. elegans* P-gps, no detailed predictions of function in nematode P-gps can be drawn on the basis of phylogeny alone. *C. elegans* P-gps do differ from one another in their ability to cause resistance to various environmental toxins [16], with no apparent correlation between phenotype and genetic distance from their mammalian homologs. Both human *abca1* and nematode *ced-7* mutants present similar apoptotic phenotypes, despite their rather distant relationship (Figure 3). *ABCA1* mutations also cause defects in high-density lipoprotein cholesterol transport, and it is still an open question as to whether the analogous function of these two homologs in apoptosis accurately predicts a sharing of other functions. Similar limitations on the extent to which function may be predicted from sequence alone are likely to obtain in those subfamilies whose members are noted for variability and multiplicity of function, that is, subfamilies A, B, C and G.

Schriml and Dean [10] speculated that the distinct clustering of amino- and carboxy-terminal halves of ABCA proteins suggests that full ABC transporters have generally evolved from half-transporters. The pattern of structural change within the

closely related subfamilies ABCD, ABCC and ABCB does suggest that the half-transporter configuration was the ancestral one for at least these three subfamilies (Figure 2). It also reveals instances where half-transporters have evolved from duplicated genes, as in the origination of ABCB from a fragment of an ABCC gene, and that, in turn, some ABCB genes have duplicated again, in giving rise to the P-gp genes.

A comprehensive comparison of worm and yeast genomes [23] noted that while most of the nematode genome did not closely resemble that of yeast, there was a strongly conserved 20% of the nematode genome that had a high degree of homology to a corresponding 40% of the yeast genome. Within this highly conserved subset of genes, there was a very frequent finding of orthology between members of the two genomes. As many as 57% of the most closely related gene pairs contained exactly one worm and one yeast gene. The obvious inference is that one corresponding gene was present in the common ancestor of the two species. Their overall picture of genome evolution is one in which a conserved cadre of proteins performs core biological functions required by all eukaryotes. These would remain essentially invariant throughout eukaryotes, and one expects analogous functions to be carried out by orthologous genes across large evolutionary distances. These gene families are presumably protected over the long run by their essential and irreplaceable roles in basic biochemical functions required by all organisms. However, as Chervitz *et al.* [23] point out, only a minority of gene families fit this mode, with most genes belonging to poorly conserved or taxonomically restricted families.

We expected that the frequency of simple orthologous gene pairs typical of highly conserved gene families shared by both yeast and worm would hold true for our comparison between nematode and human versions of such a highly conserved gene family as ABC transporters. However, this generality

Table 3**Frequency of orthologous pairs among ABC transporters**

	Sc	Ce	Dm	Hs
Sc		10%	17%	10%
Ce	3		14%	16%
Dm	5	8		22%
Hs	5	8	11	

Numbers below the diagonal represent the number of orthologous pairs of ABC transporters, according to our phylogeny, found in pairwise comparisons between each of the four genomes in this study. Percentages above the diagonal are calculated from the corresponding number given for that pair, divided by the smaller of the two counts of ABC transporters in that pair of genomes. Ce, *C. elegans*; Dm, *D. melanogaster*; Hs, *H. sapiens*; Sc, *S. cerevisiae*.

clearly does not apply to ABC transporters, despite their strong conservation across all domains of life. It seems reasonable to suppose that the rather loose relationship between substrate specificity and amino acid sequence that characterizes ABC transporters allows for much more potential exchange and sorting of biological functions among homologous genes than is typical. In turn, this pervasive pre-adaptation for functional overlap enables organisms to survive the occasional loss of substantial numbers of ABC transporters and to rapidly re-evolve lost functionality by co-opting homologous genes.

The evolutionary dynamic we propose here is reminiscent of an explanation put forward by Huynen *et al.* [44] to explain a pattern observed in a comparative analysis of 11 microbial genomes. They found that the frequency distribution of gene-family sizes within each completely sequenced genome tended to follow a power-law distribution across a 30-fold range of genome sizes. Their model is one in which genes are duplicated or deleted randomly in time, but the gene families are coherent with respect to the probability of duplication or deletion in each time unit in the simulation. In other words, the probability of duplicating or deleting a gene may change over time, but every member of a gene family always has the same probability of duplication or deletion as every other member of the family. So, whereas a given family can be either favored for expansion or targeted for deletion in a given time period, all members of the family are equally favored or disfavored by selection at the same time. Huynen *et al.* argued that this property of 'dynamic coherence' in a gene family could arise if all gene-family members have more or less the same function, so that they are all favored or disfavored by selection at the same time, depending on how much that function is needed.

Under a power-law distribution, gene families would tend to be subject to fluctuations of a size on the same order as the gene-family size itself [44]. We should then expect that

typical gene families will have undergone very substantial episodes of expansion and near-extinction, and in Huynen *et al.*'s model all gene families do become extinct within a finite time. It is evident that ABC transporters are highly atypical for a strongly conserved gene family, in that the family as a whole is highly conserved across genomes despite being subject to the same large fluctuations in size, which would tend to eventually eliminate gene families whose members are not individually indispensable. It should be noted that the ABC family does not seem uniformly subject to one or the other mode of evolution. Subfamilies E and F, which are not involved with transport, but rather have roles in translation and gene regulation, fit the 'strongly conserved' [23] model very well, retaining simple orthologous relationships over long spans of time. Only the transporter subfamilies themselves, because of their highly adaptable substrate-recognition capability, are subject to large fluctuations in size. We propose that finding large sets of paralogous genes, and infrequently conserved orthologs, in a gene family reflects ongoing cycles of gene loss and reacquisition of analogous functions in distantly related, newly expanded, lineages. Furthermore, we suggest that this is in fact the expected outcome of dynamic coherence, a mode shared, perhaps, by most of the less-conservative gene families, as well as the ABC genes.

We expect that future functional studies, to determine the extent of parallel and convergent evolution among ABC transporters, will eventually allow us to discern the fundamental roles of ABC transporters that ensure their long-term survival as a group. Also of interest will be whether the functional suites of genes fulfilling these roles are bounded in any way that resembles the phylogenetic subdivisions into which we presently categorize these proteins.

Materials and methods

Identification of ABC transporter genes

A computer file, WormPep16 [45], containing 16,332 protein sequences predicted from the completed *C. elegans* genome was searched using the FASTA program [46]. Our initial query sequences were those of known *C. elegans* ABC proteins (for example, Pgp-1, the *D. melanogaster white* gene homolog T26A5.1, and so on). Matching protein sequences returned by FASTA were checked by BLAST [47], using either the NCBI [48] or Baylor College of Medicine (BCM) servers [49]. Only those with highly significant matches to annotated ABC proteins in the sequence database were retained. The most poorly matched, verified ABC protein from each FASTA run was used as the query sequence for an additional FASTA search, and this process was repeated until no new ABC proteins were found. At a later stage in the analysis, representative members of different ABC transporter subfamilies were used as query sequences to search the updated WormPep81 file using a BLAST server at the Sanger Centre [45]. Searches were conducted using multiple queries until all proteins already included in our dataset were found. No additional

ABC proteins were identified, though some sequences were found to have been included in our dataset twice under different names. These redundant sequences were eliminated. FASTA searches were run on a SUN Microsystems UltraSPARC 5 computer. All other computer operations were carried out on an Apple Power Macintosh G3. Yeast and human ABC transporter sequences were obtained from NCBI and are described in the literature [10,13].

Identification of ABC protein features

BLAST + Beauty searches on the BCM server identified the location of the conserved Walker A and ABC signature motifs (Prosite motifs [50] PS00017 and PS00211, respectively) associated with the ATP-binding cassette(s) of each protein. The number and positions of transmembrane domains in each ABC protein were predicted by using TopPred II v1.3 [51] and then vetting the program's results by eye to exclude spurious transmembrane segments. Chromosomal locations of each ABC protein in the *C. elegans* genome were looked up in the *C. elegans* database AceDB [52].

Phylogenetic analyses

Using the information derived from each protein sequence (as above) we extracted only the sequence of each predicted ATP-binding cytoplasmic domain. These domains were assembled into a single file using the SeqApp1.9 multiple sequence editor [53], and aligned using ClustalX [54]. In those cases where two ATP-binding cassettes (ABCs) are present in a single protein with no intervening transmembrane domains (Subfamilies E and F, see Figure 1), the entire sequence was divided into two at an arbitrary point halfway between the two predicted ABC domains. As a result, 'two-domain' proteins are represented twice in our initial analysis. Once this approach had been used to assign genes to particular well-supported subgroups, we realigned the sequences and reanalyzed the relationships within each group using full-length amino acid sequence data.

Aligned sequences were used to generate matrices of mean distances between proteins, and these matrices were used to generate phylogenetic trees according to the neighbour-joining algorithm [55], refined using the SPR branch-swapping technique under the minimum evolution criterion, implemented by PAUP*4.0b10 [56]. Bootstrapping (1,000 replicates) was done according to the method of Felsenstein [57], using the same parameters described above. Phylogenetic trees were visualized and manipulated using TreeView 1.6.2 [58] and MacClade 3.0.4 [59].

Additional data files

The following additional data are included with the online version of this article: the protein sequence alignments for the ABCA subfamily (Additional data file 1), the ABCB subfamily (Additional data file 2), the ABCC subfamily (Additional data file 3), the ABCD subfamily (Additional data file 4), the ABCE

and ABCF subfamilies (Additional data file 5), the ABCG subfamily (Additional data file 6), the ABCH subfamily (Additional data file 7), and the protein sequences from the nucleotide-binding folds only (Additional data file 8). In addition to the four genomes discussed in this paper, mouse (*M. musculus*) ABC transporter genes are included in some of these alignments. All eight files are in Nexus format, which is a plain-text format designed for use with the programs PAUP [56] and MacClade [59]. A Nexus Data Editor for Windows is also available [60].

Acknowledgements

We thank Fang Zhang, whose insight and curiosity were essential, on more than one occasion, to the initiation and completion of this work. We are grateful to Yuji Kohara for the elucidation of *C. elegans* cDNAs. The helpful comments of anonymous reviewers made a substantial contribution to the final draft.

References

- Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, et al.: **The complete genome sequence of *Escherichia coli* K-12.** *Science* 1997, **277**:1453-1474.
- Croop JM: **Evolutionary relationships among ABC transporters.** *Methods Enzymol* 1998, **292**:101-116.
- Higgins CF: **ABC transporters: from microorganisms to man.** *Annu Rev Cell Biol* 1992, **8**:67-113.
- Childs S, Ling V: **The MDR superfamily of genes and its biological implications.** In *Important Advances in Oncology* Edited by: DeVita VT, Hellman S, Rosenberg SA. Philadelphia: J.B. Lippincott; 1994:21-36.
- Linton KJ, Higgins CF: **The *Escherichia coli* ATP-binding cassette (ABC) proteins.** *Mol Microbiol* 1998, **28**:5-13.
- Decottignies A, Goffeau A: **Complete inventory of the yeast ABC proteins.** *Nat Genet* 1997, **15**:137-145.
- Allikmets R, Gerrard B, Hutchinson A, Dean M: **Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database.** *Hum Mol Genet* 1996, **5**:1649-1655.
- Nomenclature for Human ABC-Transporter Genes** [<http://www.gene.ucl.ac.uk/nomenclature/genefamily/abc.html>]
- Klein I, Sarkadi B, Varadi A: **An inventory of the human ABC proteins.** *Biochim Biophys Acta* 1999, **1461**:237-62.
- Schriml LM, Dean M: **Identification of 18 mouse ABC genes and characterization of the ABC superfamily in *Mus musculus*.** *Genomics* 2000, **64**:24-31.
- Broccardo C, Luciani M, Chimini G: **The ABCA subclass of mammalian transporters.** *Biochim Biophys Acta* 1999, **1461**:395-404.
- Dean M, Rzhetsky A, Allikmets R: **The human ATP-binding cassette (ABC) transporter superfamily.** *Genome Res* 2001, **11**:1156-1166.
- Bauer BE, Wolfer H, Kuchler K: **Inventory and function of yeast ABC proteins: about sex, stress, pleiotropic drug and heavy metal resistance.** *Biochim Biophys Acta* 1999, **1461**:217-236.
- Lincke CR, The I, Vangroen M, Borst P: **The P-glycoprotein gene family of *Caenorhabditis elegans*-cloning and characterization of genomic and complementary DNA sequences.** *J Mol Biol* 1992, **228**:701-711.
- Broeks A, Gerrard B, Allikmets R, Dean M, Plasterk RHA: **Homologues of the human multidrug resistance genes MRP and MDR contribute to heavy metal resistance in the soil nematode *Caenorhabditis elegans*.** *EMBO J* 1996, **15**:6132-6143.
- Broeks A, Janssen HWRM, Calafat J, Plasterk RHA: **A P-glycoprotein protects *Caenorhabditis elegans* against natural toxins.** *EMBO J* 1995, **14**:1858-1866.
- Mahajan-Miklos S, Tan M-W, Rahme LG, Ausubel FM: **Molecular mechanisms of bacterial virulence elucidated using a *Pseudomonas aeruginosa*-*Caenorhabditis elegans* pathogenesis model.** *Cell* 1999, **96**:47-56.

18. Lincke CR, Broeks A, The I, Plasterk RHA, Borst P: **The expression of two P-glycoprotein (pgp) genes in transgenic *Caenorhabditis elegans* is confined to intestinal cells.** *EMBO J* 1993, **12**:1615-1620.
19. Wu Y, Horvitz HR: **The *C. elegans* cell corpse engulfment gene *ced-7* encodes a protein similar to ABC transporters.** *Cell* 1998, **93**:951-960.
20. Hamon Y, Broccardo C, Chambenoit O, Luciani M-F, Toti F, Chaslin S, Freyssinet J-M, Devaux PF, McNeish J, Marguet D, Chimini G: **ABC1 promotes engulfment of apoptotic cells and transbilayer redistribution of phosphatidylserine.** *Nat Cell Biol* 2000, **2**:399-406.
21. Fitch W: **Distinguishing homologous from analogous proteins.** *Syst Zool* 1970, **19**:99-113.
22. Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Philos Trans R Soc London B Biol Sci* 2001, **356**:1661-1679.
23. Chervitz SA, Aravind L, Sherlock G, Ball CA, Koonin EV, Dwight SS, Harris MA, Dolinski K, Mohr S, Smith T, et al.: **Comparison of the complete protein sets of worm and yeast; orthology and divergence.** *Science* 1998, **282**:2022-2028.
24. Mitsuhashi N, Miki T, Senbongi H, Yokoi N, Yano H, Miyazaki M, Nakajima N, Iwanaga T, Yokoyama Y, Shibata T, Seino T: **MTABC3, a novel mitochondrial ATP-binding cassette protein involved in iron homeostasis.** *J Biol Chem* 2000, **275**:17536-17540.
25. Furuya KN, Bradley G, Sun D, Schuetz EG, Schuetz JD: **Identification of a new P-glycoprotein-like ATP-binding cassette transporter gene that is overexpressed during hepatocarcinogenesis.** *Cancer Res* 1997, **57**:3708-3716.
26. Wijnhouds J, Mol CA, van Deemter L, de Haas M, Scheffer GL, Baas F, Beijnen JH, Scheper RJ, Hatse S, De Clercq E, et al.: **Multidrug-resistance protein 5 is a multispecific organic anion transporter able to transport nucleotide analogs.** *Proc Natl Acad Sci USA* 2000, **97**:7476-7481.
27. Jedlitschky G, Burchell B, Keppler D: **The multidrug resistance protein 5 functions as an ATP-dependent export pump for cyclic nucleotides.** *J Biol Chem* 2000, **275**:30069-30074.
28. Le Roy F, Laskowska A, Silhol M, Salehzada T, Bisbal C: **Characterization of RNase L, an RNA binding protein that associates with RNase L.** *J Interferon Cytokine Res* 2000, **20**:635-644.
29. Bisbal C, Silhol M, Laubenthal H, Kaluza T, Carnac G, Milligan L, Le Roy F, Salehzada T: **The 2'-5' oligoadenylate/RNase L/RNase L inhibitor pathway regulates both MyoD mRNA stability and muscle cell differentiation.** *Mol Cell Biol* 2000, **20**:4959-4969.
30. Douglas S, Zauner S, Fraunholz M, Beaton M, Penny S, Deng LT, Wu X, Reith M, Cavalier-Smith T, Maier UG: **The highly reduced genome of an enslaved algal nucleus.** *Nature* 2001, **410**:1091-1096.
31. **Saccharomyces Genome Database** [<http://genome-www.stanford.edu/Saccharomyces>]
32. Allikmets R, Raskind WH, Hutchinson A, Schueck ND, Dean M, Koeller DM: **Mutation of a putative mitochondrial iron transporter gene (ABC7) in X-linked sideroblastic anemia and ataxia (XLSA/A).** *Hum Mol Genet* 1999, **8**:743-749.
33. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PV, Hoskins RA, Galle RF, et al.: **The genome sequence of *Drosophila melanogaster*.** *Science* 2000, **287**:2185-2195.
34. The *C. elegans* Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282**:2012-2018.
35. Saurin W, Hofnung M, Dassa E: **Getting in or out: early segregation between importers and exporters in the evolution of ATP-binding cassette (ABC) transporters.** *J Mol Evol* 1999, **48**:22-41.
36. Cole SPC, Deeley R: **Multidrug resistance mediated by the ATP-binding cassette transporter protein MRP.** *BioEssays* 1998, **20**:931-940.
37. van Helvoort A, Smith AJ, Sprong H, Fritzsche I, Schinkel AH, Borst P, van Meer G: **MDR1 P-glycoprotein is a lipid translocase of broad specificity, while MDR3 P-glycoprotein specifically translocates phosphatidylcholine.** *Cell* 1996, **87**:507-517.
38. Schinkel AH: **The physiological function of drug-transporting P-glycoproteins.** *Semin Cancer Biol* 1997, **8**:161-170.
39. Balzi E, Wang M, Leterme S, Van Dyck L, Goffeau A: **PDR5, a novel yeast multidrug resistance conferring transporter controlled by the transcription regulator PDRI.** *J Biol Chem* 1994, **269**:2206-2214.
40. Doyle LA, Yang W, Abruzzo LV, Krogmann T, Gao Y, Rishi AK, Ross DD: **A multidrug resistance transporter from human MCF-7 breast cancer cells.** *Proc Natl Acad Sci USA* 1998, **95**:15665-15670.
41. Loo TW, Clarke DM: **Functional consequences of phenylalanine mutations in the predicted transmembrane domain of P-glycoprotein.** *J Biol Chem* 1993, **268**:19965-19972.
42. Zhang F, Sheps JA, Ling V: **Complementation of transport-deficient mutants of *Escherichia coli* α -hemolysin by second-site mutation in the transporter hemolysin B.** *J Biol Chem* 1993, **268**:19889-19895.
43. Shapiro AB, Ling V: **Transport of LDS-751 from the cytoplasmic leaflet of the plasma membrane by the rhodamine-123-selective site of P-glycoprotein.** *Eur J Biochem* 1998, **254**:181-188.
44. Huynen MA, van Nimwegen E: **The frequency distribution of gene family sizes in complete genomes.** *Mol Biol Evol* 1998, **15**:583-589.
45. **The Sanger Institute: *Caenorhabditis* genome sequencing projects** [http://www.sanger.ac.uk/Projects/C_elegans]
46. Pearson WR: **Rapid and sensitive sequence comparison with FASTP and FASTA.** *Methods Enzymol* 1990, **183**:63-98.
47. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403-410.
48. **NCBI BLAST home page** [<http://www.ncbi.nlm.nih.gov/BLAST>]
49. **BCM search launcher** [<http://searchlauncher.bcm.tmc.edu>]
50. Hofmann K, Bucher P, Flaquet L, Bairoch A: **The PROSITE database, its status in 1999.** *Nucleic Acids Res* 1999, **27**:215-219.
51. Claros MG, von Heijne G: **TopPred II: an improved software for membrane protein structure predictions.** *Comput Appl Biosci* 1994, **10**:685-686.
52. **AceDB home page** [<http://www.acedb.org>]
53. **UIBio archive: SeqApp** [<http://iubio.bio.indiana.edu/soft/molbio/seqapp/>]
54. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res* 1997, **25**:4876-4882.
55. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
56. Swofford DL: *Paup* (Phylogenetic analysis using parsimony and other methods)*. Sunderland, MA: Sinauer; 1997.
57. Felsenstein J: **Confidence limits on phylogenies: an approach using the bootstrap.** *Evolution* 1985, **39**:783-791.
58. Page RDM: **TreeView: an application to display phylogenetic trees on personal computers.** *Comput Appl Biosci* 1996, **12**:357-358.
59. Maddison WP, Maddison DR: *MacClade: analysis of phylogeny and character evolution*. 3.0th edition. Sunderland, MA: Sinauer; 1992.
60. **NEXUS data editor for Windows** [<http://taxonomy.zoology.gla.ac.uk/rod/NDE.html>]
61. **Genetic nomenclature for *Caenorhabditis elegans*** [<http://biosci.umn.edu/CGC/Nomenclature/nomenguid.htm>]
62. **WormBase home page** [<http://www.wormbase.org>]
63. Fraser AG, Kamath RS, Zipperlen P, Martinez-Campos M, Sohrmann M, Ahringer J: **Functional genomic analysis of *C. elegans* chromosome I by systemic RNA interference.** *Nature* 2000, **408**:325-330.
64. Kamath RS, Fraser AG, Dong Y, Poulin G, Durbin R, Gotta M, Kanapin A, Le Bot N, Moreno S, Sohrmann M, et al.: **Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi.** *Nature* 2003, **421**:231-237.
65. Gönczy P, Echeverri C, Oegema K, Coulson A, Jones SJ, Copley RR, Duperon J, Oegema J, Brehm M, Cassin E, et al.: **Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III.** *Nature* 2000, **408**:331-336.