The Evolution of the Mammalian *ABCA6*-like Genes: Analysis of Phylogenetic, Expression, and Population Genetic Data Reveals Complex Evolutionary Histories

Martin W. Breuss[1,2,†], Allen Mamerto[3,†], Tanya Renner[4,†], and Elizabeth R. Waters[3,*]

[1]Department of Neurosciences, University of California, San Diego

[2]Rady Children's Institute for Genomic Medicine, San Diego, California

[3]Department of Biology and Program in Biological and Medical Informatics, San Diego State University

[4]Department of Entomology, The Pennsylvania State University, University Park

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: ewaters@sdsu.edu.

Accepted: 21 August 2020

Abstract

ABC membrane transporters are a large and complex superfamily of ATP-binding cassette transporters that are present in all domains of life. Both their essential function and complexity are reflected by their retention across large expanses of organismal diversity and by the extensive expansion of individual members and subfamilies during evolutionary history. This expansion has resulted in the diverse ABCA transporter family that has in turn evolved into multiple subfamilies. Here, we focus on the *ABCA6*-like subfamily of ABCA transporters with the goal of understanding their evolutionary history including potential functional changes in, or loss of, individual members. Our analysis finds that *ABCA6*-like genes, consisting of *ABCA6*, *8*, *9*, and *10*, are absent from representatives of both monotremes and marsupials and thus the duplications that generated these families most likely occurred at the base of the Eutherian or placental mammals. We have found evidence of both positive and relaxed selection among the *ABCA6*-like genes, suggesting dynamic changes in function and the potential of gene redundancy. Analysis of the *ABCA10* genes further suggests that this gene has undergone relaxed selection only within the human lineage. These findings are complemented by human population data, where we observe an excess of deactivating homozygous mutations. We describe the complex evolutionary history of this ABCA transporter subfamily and demonstrate through the combination of evolutionary and population genetic analysis that *ABCA10* is undergoing pseudogenization within humans.

Key words: ABCA, ABC membrane transporters, mammalian gene family evolution, gene birth and death.

Significance

ABC transporters fulfill a critical and varied cellular role by transporting various compounds across cell membranes; reflecting this, through evolution they have formed an extensive and complex gene family. Here, we explore the evolutionary history and population genetic status of one subfamily, the *ABCA6*-like transporters, and find that they underwent complex gain and loss of family members, exemplified by one gene—*ABCA10*—in humans, which is in the process of pseudogenization. Our results help to understand how this large and essential family can dynamically adapt within an evolutionary context and can serve an example for other complex multigene families.

Introduction

The *ABCA6*-like ABC transporters are ATP-binding cassette (ABC) membrane transporters that are part of a large gene family present in all domains of life (Jones and George 2004; Moitra and Dean 2011; Srikant and Gaudet 2019). The main function of their various gene products is the translocation of substrates, mainly lipids, across biological membranes fueled by ATP hydrolysis (Locher 2016). Due to their diversity and ubiquity in biological processes, the ABC transporters have been implicated in various human diseases. Their upregulation underlies multidrug resistance in cancer (Mohammad et al. 2018) and rare and common genetic variants have been implicated in a range of phenotypes (Moitra and Dean 2011). For instance, a common variant in *ABCA7*, a transporter implicated in cholesterol regulation and amyloid processing, has been identified as a risk factor in Alzheimer's disease (MIM: 608907) (Hollingworth et al. 2011); Tangier disease, a rare disorder of cholesterol metabolism with a wide-range of phenotypic effects, is caused by a mutation in *ABCA1* (Ceccanti et al. 2016); and rare, recessive mutations inactivating *ABCA12* have been identified in congenital forms of ichthyosis (MIM: 601277 and 242500) (Annilo et al. 2002; Thomas et al. 2006).

The ABC transporters comprise several subfamilies, such as the ABCA transporters (Peelman et al. 2003; Kaminski et al. 2006; Dermauw and Van Leeuwen 2014). Members of this subfamily typically consist of 12 transmembrane helices, two nucleotide-binding motifs containing a Walker A and a Walker B motif on the cytoplasmic side, and several loops on the extracellular surface that can contain glycosylation sites (fig. 1*A*). The entire transporter can be functionally divided into two halves that contain six transmembrane domains and one nucleotide-binding domain (NBD) (fig. 1*B*). Upon interaction with the substrate, the NBDs can cooperatively bind two ATP molecules, which results in a conformational change that allows the substrate to transverse the lipid bilayer of the membrane (fig. 1*B*). Subsequent ATP-hydrolysis and release of ADP and inorganic phosphate will revert the ABCA transporter to its ground state (fig. 1*B*).

The ABCA subfamily in turn can be further divided into subgroups (Moitra and Dean 2011). One of these is the class of *ABCA6*-like transporters that, in humans, consists of four genes that lie within a cluster on chromosome 17 that also contains *ABCA5*, a gene that is more distantly related (Kaminski et al. 2001; Piehler et al. 2002; Tsuruoka et al. 2002; Annilo et al. 2003; Wenzel et al. 2003). The members of this subgroup, *ABCA6*, *ABCA8*, *ABCA9*, and *ABCA10*, are all implicated in cholesterol-related pathways, either directly or based on their dynamic regulation upon cholesterol application (Kaminski et al. 2001; Piehler et al. 2002; Tsuruoka et al. 2002; Wenzel et al. 2003; Sasaki et al. 2018). Furthermore, a rare missense mutation in *ABCA6* has been associated with cholesterol levels in the Dutch population (van Leeuwen et al. 2015). Of the four members, only *ABCA10* is absent from the mouse genome. Remarkably, a loss-of-function (LoF) variant in this gene segregated with autism spectrum disorder in multiplex families (Lim et al. 2013).

One of the challenges in understanding the roles of the *ABCA6*-like transporters is a lack of information on the roles of each of the distinct *ABCA* genes in humans. Prior analyses suggest that all are closely related but there is much to be learned about when these genes duplicated and how they are evolving in humans and other mammals. In hopes of furthering our understanding of the functional diversity of the *ABCA6*-like genes, evolutionary analysis, including phylogenetic and evolutionary rate analysis across mammalian species, as well as evaluation of variation within humans, has allowed us to examine both the origins and evolutionary dynamics of each of the members of the *ABCA6*-like clade (*ABCA6, 8, 9, 10*). In this study, we integrate evolutionary analysis with available information on protein structure and gene expression. Of particular interest is the question of elevated rates of evolution of some of the *ABCA* genes in humans: Are these genes evolving toward new functions via positive selection or are they evolving neutrally suggesting an LoF?

We have found that the *ABCA6*-like genes are absent in monotremes or marsupials. This suggests that a burst of gene duplications generated the *ABCA6*-like gene families at the base of the Eutherian mammalian divergence (Foley et al. 2016). The *ABCA6*-like genes display a diverse range of evolutionary rates from purifying selection, to positive evolution to relaxed selection. Further, the patterns of frequent loss with only a few new gains (duplications) of *ABCA6*-like genes in eutherian mammals suggest high organismal tolerance for loss of individual *ABCA6*-like genes and that many of the encoded proteins have overlapping functions. Finally, it appears that the *ABCA10* alleles in humans are evolving under relaxed selection, suggesting that this gene is already or may become a pseudogene.

Materials and Methods

Expression Analysis

Data were obtained from the Genotype-Tissue Expression (GTEx) portal and processed using Python (3.64) with the Pandas (0.22.0) (McKinney 2010) and Seaborn (0.8.1) (Waskom et al. 2017) packages.

Exome Aggregation Consortium Analysis

Observed variants, constraint metrics, and the frequencies of LoF variants for seven ethnic populations of human exomes were collected from the Exome Aggregation Consortium (ExAC) Browser. Expression of the ABCA member 10 gene was taken from the GTEx portal and used to map the LoF variants for member 10. Bar plots for population LoF variants were generated using R.

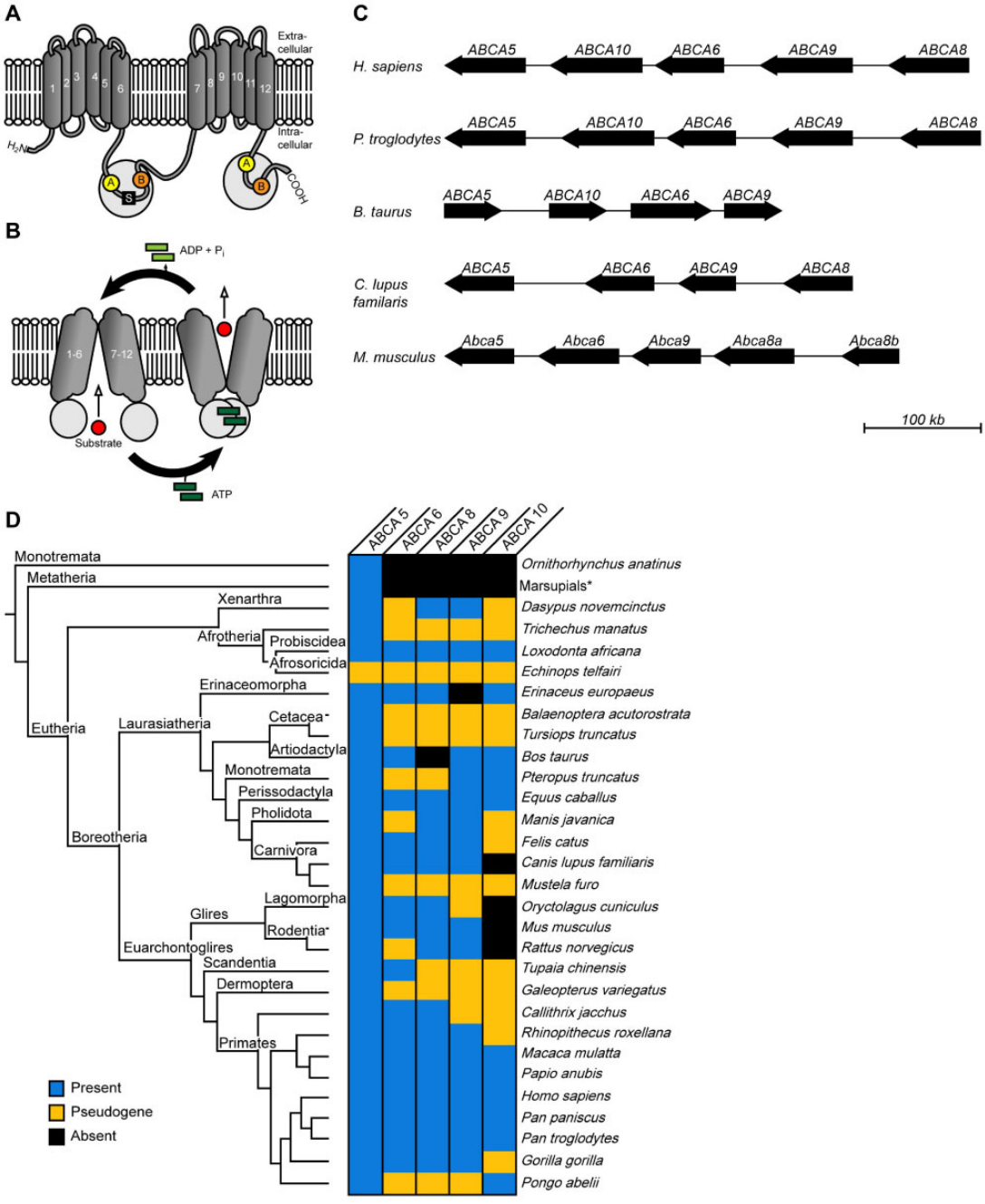


Fig. 1.—Structure, function, and genomic organization of the *ABCA6*-like family. (*A*) Schematic of ABCA subfamily members showing the 12 transmembrane domains and the two nucleotide-binding domains containing the Walker A (A) and Walker B (B) motifs for ATP binding, as well as the ABC signature motif (S). Schematic also shows the N-terminus (H_2N) and the C-terminus (COOH). (*B*) Schematic depicting an ABCA transporter and the ATP-dependent transformations needed to transport a substrate across the lipid bilayer. Following the conformational change, ATP is hydrolyzed and ADP and inorganic phosphate (P_i) are released to transition back to the ground state. Figure is modeled after schematics depicted in previous publications (Wenzel et al. 2003; Dermauw and Van Leeuwen 2014). (*C*) Synteny of *ABCA5* and *ABCA6*-like genes in five mammalian genomes. Genomic positions and orientation are based on analysis of data at the UCSC Genome Browser (https://genome.ucsc.edu/; last accessed September 14, 2020). Forward arrows indicate 5′–3′ orientation, reverse arrows indicate 3′–5′. Chromosome positions are as follows: *H. sapiens* Chr17: 68868085–69314415; *P. troglodytes* Chr17: 57633672–67708059; *B. taurus* Chr19: 61873253–62153618; *C. lupus* Chr9: 15421150–15767955; *M. musculus* Chr11: 109934510–110329238. (*D*) Gene loss and retention of ABCA genes across diverse mammalian lineages. Organismal tree is based on the current consensus of mammalian relationships (Pozzi et al. 2014; Foley et al. 2016). Full-length functional gene sequences are in blue; pseudogenes in yellow; and the absence homologous sequences are indicated in black. Sequence accession numbers are provided in supplementary table S2, Supplementary Material online.

LoF Variant Analysis

The ExAC VCF and the pLI/pRec/pNull table were obtained from the ExAC repository on March 14, 2018 (Exome Aggregation Consortium et al. 2016; Karczewski et al. 2020). For this analysis, mainly in order to reduce false-positive genes, only those that had a pNull score >0.9 were considered. Genes with a pNull score at this level show little to no functional constraint on the accumulation of heterozygous and homozygous deactivating mutations across the human population; thus, they behave as expected from nongenic/nonfunctional regions of the genome and are likely functionally redundant or obsolete. This metric is based on the catalog of all observed "null" mutations (e.g., frameshift, stop-gain) which are compared with an expected number of such variants in the absence of functional constraint. For our analysis, the ExAC VCF containing the remaining genes was filtered for sites that contained stop-gain and frameshift annotation (summarized as LoF variants) and combined with the pNull annotation. Furthermore, low-confidence sites and high-confidence sites with single-exon annotation were excluded from the analysis. To obtain genes with similar LoF patterns, we further excluded genes that had less than two sites with more than ten individuals or that had a clustering of all LoF variants within 500 bp. Finally, ambiguous multivariant sites resulting from conversion by VEP were manually curated (McLaren et al. 2016). All the data processing, analysis, and visualization were performed with Pandas and Seaborn (see above).

Identification of ABCA Homologs and Alignments

In order to identify mammalian and marsupial ABCA homologs, the program BlastN (using default parameters) was used to search the NCBI nonredundant nucleotide database using the human *ABCA5* (NM_018672.4) and *ABCA6*-like genes (NM_080284.2, BC130280.1, NM_080283.3, NM_080282.3) as query sequences. Database sequences that covered at least 95% of the related human *ABCA* genes and had an *E*-value of <1e-179 with a minimum score of 9,500 were selected for each species and used for analysis. Accession numbers for the sequences identified and used for further analysis are provided in supplementary table S2, Supplementary Material online. Sequences that shared significant similarity to the known *ABCA6*-like genes but possessed stop codons and truncated transcripts were deemed pseudogenes. All of the gene models were manually analyzed and annotated. In cases where homologs of the individual *ABCA6*-like genes were not found in a given genome, BlastP searches with default parameters were used with the human amino acid sequences to ensure that we did not miss any divergent homologs. Amino acid sequences were generated from the translation of the genomic sequences. The amino acid sequences were aligned with Clustal Omega (Sievers and Higgins 2018) using default parameters. Positions with gaps or that were unreliable were removed, resulting in a final amino acid alignment. This amino acid alignment was then used to align the DNA sequences using TranslatorX (Patricio et al. 2010). Comparative analysis of *ABCA* gene organization was examined in the UCSC Genome Browser (https://genome.ucsc.edu/; last accessed September 14, 2020) (Kent et al. 2002) (see the accompanying sequences [FASTA_list_ABCA5_ABCA6_like.fasta]).

Phylogenetic Reconstructions

Phylogenetic trees based on the DNA sequences were generated using RAxML (Stamatakis 2015) using default parameters and the general time reversible (GTR) model of nucleotide substitutions. Reliability of branches within each tree was evaluated with 1,000 bootstrap replicates. The DNA-based phylogenetic trees of each ABCA family were then used in the programs PAML (Yang 2007) and HyPhy (Weaver et al. 2018) to estimate levels and rates of natural selection. To further examine the evolution of the ABCA proteins, a phylogenetic tree of ABCA5 and ABCA6-like amino acid sequences was generated using RAxML and the reliability was evaluated with 1,000 bootstraps.

Tests of Selection and Protein Modeling

In order to identify and quantify heterogeneity in selection pressure within and among the *ABCA6*-like genes, we analyzed the alignments generated above with the GA Branch analysis tool using a GTR model of nucleotide substitutions. This was performed using Datamonkey on the HyPhy web server (Weaver et al. 2018) (datamonkey.org). GA Branch uses a genetic algorithm and the Akaike information criterion to identify the best-fitting model for the number of branch ω (d*N*/d*S*) classes. The GA Branch program (Pond and Frost 2005) assigns individual branches within the tree to discreet d*N*/d*S* ratio classes. A model-averaged probability of positive selection (d*N*/d*S* > 1) on any of these branches is used to test whether positive selection has occurred. The program codeml (PAML v4.4) was used to estimate ω (d*N*/d*S*) values (Yang 2007). Both branch-specific and branch-site models were used. Likelihood ratio tests (LRTs) were used to compare two nested models for asymmetric sequence evolution (one-ratio model 0 [$\omega_0 = \omega_1$] vs. two-ratio model 2 [ω_0, ω_1]), divergent selection (M3 vs. clade model D), and positive selection (model A_{null} [$\omega_2 = 1$] vs. model A [$0 < \omega_0 < 1$]). A chi-square test was performed with the log likelihood results for each tip and node in the phylogeny, and the resulting *P*-values were Bonferroni corrected for multiple comparisons. After Bonferroni correction, *P*-values ≤ 0.007 were considered significant. For tests of positive selection at tips in the phylogenies, we also carried out the Benjamini–Hochberg procedure for multiple comparisons on LRT *P*-values (False Discovery Rate [FDR] = 0.2) (McDonald 2014). Sites identified as under positive selection ($\omega > 1$) by Bayes empirical Bayes

(BEB) analysis correspond to amino residues in the multiple sequence alignments, following automatic removal of gaps by codeml. Amino acid sites identified as under positive selection in the ungapped alignments were matched to positionally homologous sites in the cryo-EM structure of human ABCA1 (PDB ID: 5XJY) via a multiple sequence alignment using MUSCLE (Edgar 2004) and subsequently visualized in MacPyMOL (2009, DeLano Scientific LLC). For complete information on codeml LRTs, chi-square tests, and *P*-values found to be significant under Bonferroni correction or following the Benjamini–Hochberg procedure, see supplementary data 1, Supplementary Material online. A multiple sequence alignment of human ABCA1 (PDB ID: 5XJY) and *ABCA6*-like protein products with amino acid sites identified as under significant positive selection by codeml (PAML v4.4) (Yang 2007) is available as supplementary data 2, Supplementary Material online.

Results

Evolution of ABCA6-like Genes in Mammals

Synteny of ABCA6-like Genes Is Conserved among Major Groups of Eutherian Mammals

Analysis of the chromosomal locations and orientations of the *ABCA5* and *ABCA6*-like genes in *Homo sapiens, Pan troglodytes, Bos taurus*, *Canis lupus*, and *Mus musculus* (fig. 1*C*, supplementary fig. S1 and table S1, Supplementary Material online) revealed conserved synteny. The *ABCA6*-like genes in both *H. sapiens* and *P. troglodytes* are found on chromosome 17 in the following 5′–3′ order: *ABCA 5*, *10*, *6*, *9*, *8*. When more distantly related mammals like the cow (*B. taurus*) are examined, some changes in synteny are found. Here, the *ABCA6*-like gene cluster lacks *ABCA8* and the other genes have the following order: *5*, *10*, *6*, *9*; however, the orientation had been reversed to the sense instead of the antisense strand. The *ABCA6*-like genes in dog (*C. lupus*) share synteny with *H. sapiens* on the antisense strand, but the cluster has lost *ABCA10*. The mouse (*M. musculus*) gene cluster shares synteny with *H. sapiens*, but has an additional *ABCA8* gene (*ABCA8a*, *ABCA8b*) and lacks the gene for *ABCA10*.

Gain and Loss of ABCA6-like Genes

In order to gain a deeper understanding of the evolutionary dynamics of the *ABCA6*-like genes, we expanded our analysis to include 30 species representing the major placental mammalian lineages and the closely related monotremes and marsupials. In figure 1*D*, we present an organismal tree that reflects the current understanding of the phylogenetic relationships of mammals (based on Foley et al. 2016). For each of the species included in the organismal tree, we indicate the presence or absence of the *ABCA* genes (*5*, *6*, *8*, *9*, and *10*) for that species (supplementary table S2, Supplementary Material online). We distinguish between the presence of a gene that codes for a full-length protein (blue) and pseudogenes (yellow). We define pseudogenes as genes with significant sequence similarity, but that possess disabling or nonsense mutations that prevent the production of full-length proteins. The mouse (*M. musculus*) genome has *ABCA6*, *8*, and *9* genes but lacks the *ACBA10* region. Analysis of the chromosomal region where, based on synteny, we would have expected to find *ABCA10* shows no homology to any coding region. The rat (*Rattus norvegicus*) genome is also missing the *ABCA10* gene. However, although rat lacks a functioning ABCA6 gene, it does possess an *ABCA6* pseudogene. The rabbit (*Oryctolagus cuniculus*) also lacks the *ABCA10* gene and although the ABCA9 gene is present it is a pseudogene in this species. The orangutan (*Pongo abelii*) has *ABCA10* but not *6*, *8*, or *9*. Gorilla (*Gorilla gorilla*) has *ABCA6*, *8*, *9* but not *10*. However, it has an *ABCA10* sequence with disabling mutations and is thus a pseudogene. Finally, both human (*Homo sapiens*), bonobo (*Pan paniscus*), and chimpanzee (*Pan troglodytes*) have functioning *ABCA6*, *8*, *9*, and *10* genes.

Our analysis further reveals that, of the 30 species examined, seven (*Loxodonta africana, Equus caballus, Macaca mulatta, Papio anubis, P. paniscus, P. trioglodytes*, and *H. sapiens*) have all five *ABCA6*-like genes, as well as the gene for *ABCA5*. The other 23 species have either a pseudogene or lack entirely at least one of the *ABCA6*-like genes. Only one species examined here, the tenrec (*E. telfairi*), lacks a fully functioning *ABCA5* gene. *Ornithorhynchus anatinus*, a monotreme, only has a single *ABCA5* gene and does not possess the *ABCA6*-like genes.

To further examine the origins of the *ABCA6*-like genes, we also searched the genomes of *Xenopus laevis* (an amphibian) and *Gallus gallus* (a bird) for homologs of *ABCA5* and the *ABCA6*-like families. Our phylogenetic analysis presented in supplementary figure S2, Supplementary Material online, indicates that *Xenopus* and *Gallus* do have *ABCA5* homologs and that these species lack any ABCA6-like homologs. We then reexamined the *Monodelphis domestica* genome and the genomes of two additional marsupial genomes, *Sarcophilus harrisii* (Tasmanian devil) and *Phascolarctos cinereus* (koala), for evidence of *ABCA5* and *ABCA6*-like gene homologs. *ABCA5* homologs were identified in each of these genomes, reconfirming that the *ABCA5* genes are shared among tetrapods. Notably each of the marsupial genomes contains two additional ABCA proteins (supplementary fig. S2, Supplementary Material online). Here, we term these ABCAs as "undefined." The marsupial ABCA-undefined proteins are not members of any of the *ABCA6*-like families. The phylogenetic relationships of the marsupial ABCA-undefined proteins do not reflect known organismal relationships, and it appears that these proteins have undergone either multiple duplications or gene conversion (i.e., the two *Phascolarctos ABCAs* are each other's closest relatives). Further, although

the genes that code for these proteins are found near the *ABCA5* locus, they do not share synteny with the mammalian *ABCA6*-like families. These data suggest that like the *ABCA6*-like proteins, the marsupial ABCA-undefined proteins evolved via gene duplication from an ABCA5 protein, but that marsupial ABCA-undefined proteins do not belong to any of the *ABCA6*-like families and thus duplicated independently.

In order to further understand the evolution of *ABCA6*-like genes, a maximum likelihood tree was constructed using *ABCA5* as the outgroup (fig. 2). The high bootstrap values for each of the *ABCA6*-like gene families indicate that the gene family evolved via independent gene duplications and has not undergone gene conversion across the individual *ABCA6*-like gene families after the early duplication events. From figure 2, it is clear that most of the gene trees reflect general organismal relationships. The branching patterns suggest that *ABCA8* and *ABCA9* shared a more recent common ancestor and that these genes evolved from a common gene duplication event. It also suggests that *ABCA6* is more closely related to *ABCA5*. There has been a gene duplication event in mouse for *ABCA8* and this species has two *ABCA8* genes (*ABCA8a* and *ABCA8b*). Taken together, the data presented in figures 1*D* and 2 suggest a rapid birth of the *ABCA6*-like genes, followed by variable loss via pseudogenization that has occurred across organismal lineages and across *ABCA6*-like gene families.

ABCA6-like Genes Have a Wide Range of Evolutionary Rates

In order to understand the levels and types of selective pressures acting on the individual *ABCA6*-like gene families, we used a number of rate analysis methods. First, we used the programs within the package HyPhy (Pond et al. 2005) on the Datamonkey web site (http://datamonkey.org; last accessed September 14, 2020). The estimated d*N*/d*S* (ω) ratios shown in figure 3 (supplementary tables S3–S6, Supplementary Material online) reveal that the rates of positive, neutral, and purifying selection vary both within and between the *ABCA6*-like gene families. For example, most *ABCA6* genes are evolving under a mix of purifying and relaxed selection; however, the branch leading to *P. troglodytes* has evolved under positive selection (d*N*/d*S* 37.4) (fig. 3*A*, Supplementary table S3, Supplementary Material online). In the *ABCA8* tree, it is clear that the only branch evolving under positive selection (d*N*/d*S* 2.2) is the branch leading to the carnivores (fig. 3*B*, supplementary table S4, Supplementary Material online). In the *ABCA9* tree, only the branch leading to *P. paniscus*, has evolved under positive selection (d*N*/d*S* 96.5) (fig. 3*C*, supplementary table S5, Supplementary Material online). The rest of the *ABCA9* tree is evolving under strong purifying selection. It is worthy of note that there is evidence of relaxed selection (d*N*/d*S* 0.58) acting on the *ABCA10* genes in the branch leading to *H. sapiens* (fig. 3*D*, supplementary table S6, Supplementary Material online). This is in contrast to the all the other *ABCA6*-like genes *6*, *8*, and *9*, where the branches leading to *H. sapiens* are evolving under purifying selection (d*N*/d*S* 0.30, 0.30, 0.28; supplementary tables S3, S4, and S5, Supplementary Material online, respectively).

In order to further refine our understanding of the evolutionary constraints acting on the *ABCA6*-like genes, we performed additional evolutionary rate analysis using codeml in the PAML package of programs (Yang 2007). The results of this analysis are presented in figure 4 (see supplementary data 1, Supplementary Material online, for codeml LRTs, chi-square tests, and *P*-values found to be significant under Bonferroni correction [corrected *P*-value $\leq$ 0.007] or following the Benjamini–Hochberg procedure). Again, we found a range of evolutionary rates with some genes in some organismal lineages evolving under positive selection, but overall most of the *ABCA6*-like genes are evolving under purifying selection. Our analysis shows that the *ABCA9* genes are evolving under significant divergent evolution. In the *ABCA9* gene tree (fig. 4, supplementary fig. S3, Supplementary Material online), five tips are under asymmetrical evolution: mouse, dog, cat, pangolin, and armadillo (supplementary data 1, Supplementary Material online). However, pangolin *ABCA9* had one site under positive selection; but this site is not associated with a named structural or functional domain (fig. 4, supplementary fig. S3, Supplementary Material online). In rhesus macaque *ABCA10* (XM_015120204.1), two amino acid sites in the first ABC subfamily A domain, one site in the second ABC subfamily domain that overlaps with a Walker A/P-loop, and one site in the DUF4162 domain were found to be under positive selection (fig. 4, supplementary fig. S3 and supplementary data 1, Supplementary Material online). Although these rhesus macaque ABCA10 amino acid sites were only found to be significant following carrying out the Benjamini–Hochberg procedure for multiple comparisons on LRT *P*-values (FDR = 0.2) as opposed to Bonferroni correction, it is noteworthy that many fall within functionally important domains, with the first ABC subfamily A domain essentially marking the ATP-binding (or nucleotide-binding) domain and the Walker A/P-loop being involved in phosphate binding (Locher 2016). Our analysis also revealed that rhesus macaque *ABCA10* has evolved under significant asymmetrical and divergent evolution (supplementary data 1, Supplementary Material online).

It is notable that this analysis found four sites within armadillo *ABCA8* (XM_12525536.1) that are evolving under positive selection: Three are amino acid sites in transmembrane domains, and one site in the first ABC subfamily A domain (fig. 4, supplementary data 1, Supplementary Material online). Armadillo *ABCA8* was also found to be under asymmetrical and divergent evolution. This analysis also revealed that mouse *ABCA8b* is under significant asymmetrical evolution and an internal node leading to rabbit *ABCA8*, mouse

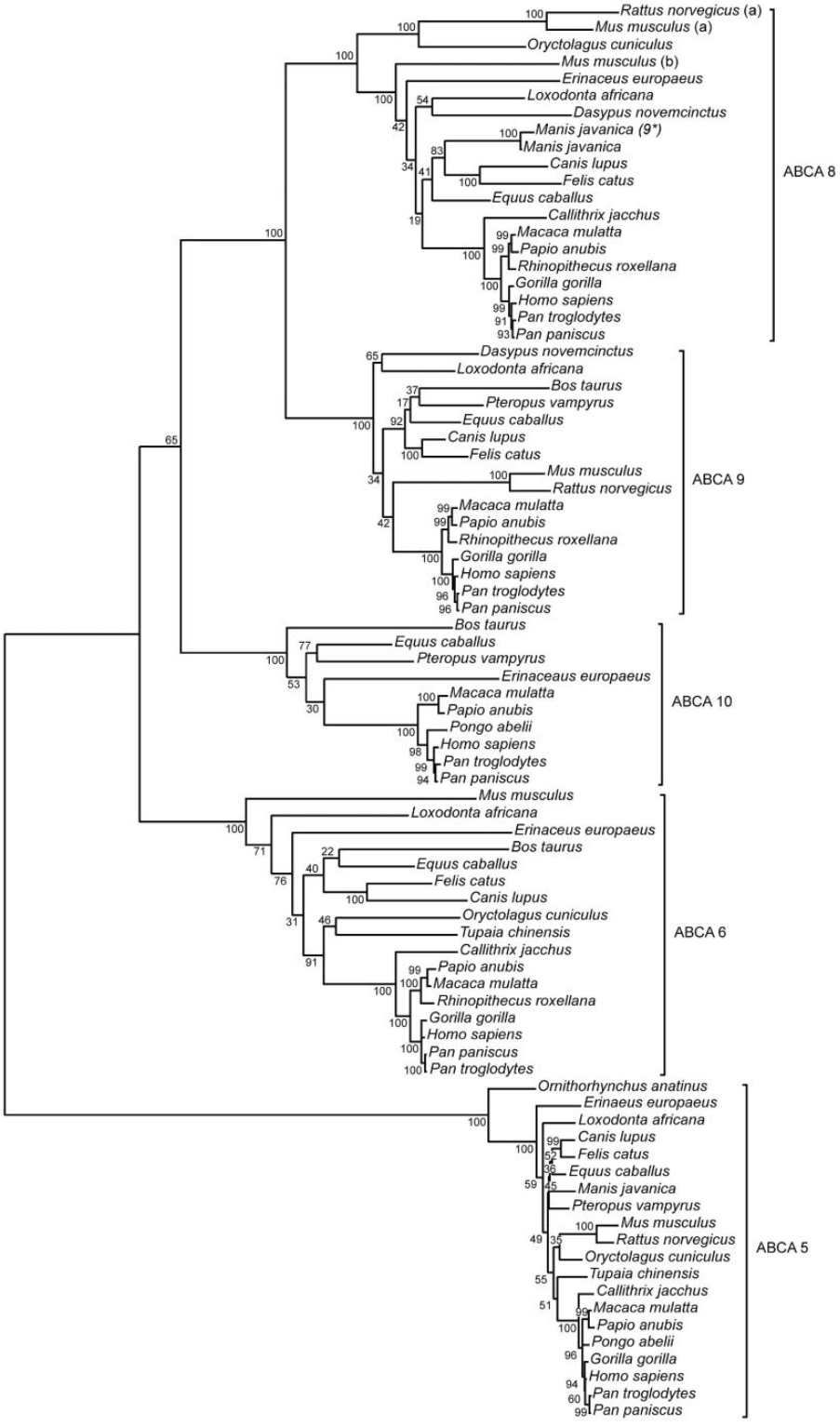


Fig. 2.—Gene tree of *ABCA6*-like genes. Tree was constructed using maximum likelihood implemented in RAXml using the following parameters: the GTR model with empirical base frequencies and GAMMA distribution. Bootstrap support for individual branches is indicated below each branch and reflects the results of 1,000 bootstrap replicates. *ABCA5* is used as the outgroup for the *ABCA6*-like lineages.

ABCA8a, and rat *ABCA8a* in the *ABCA8* phylogeny to have a site under positive selection (and also under significantly

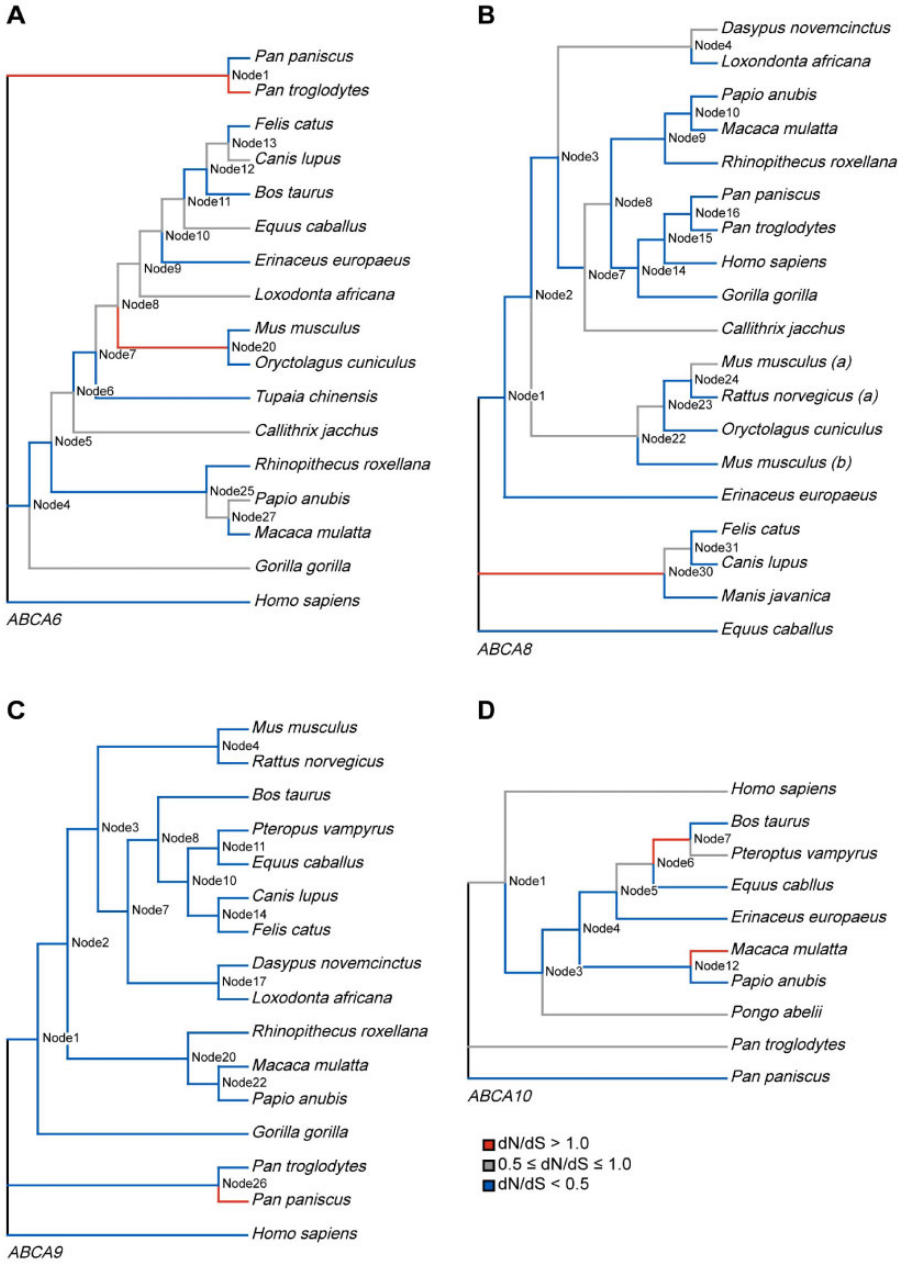


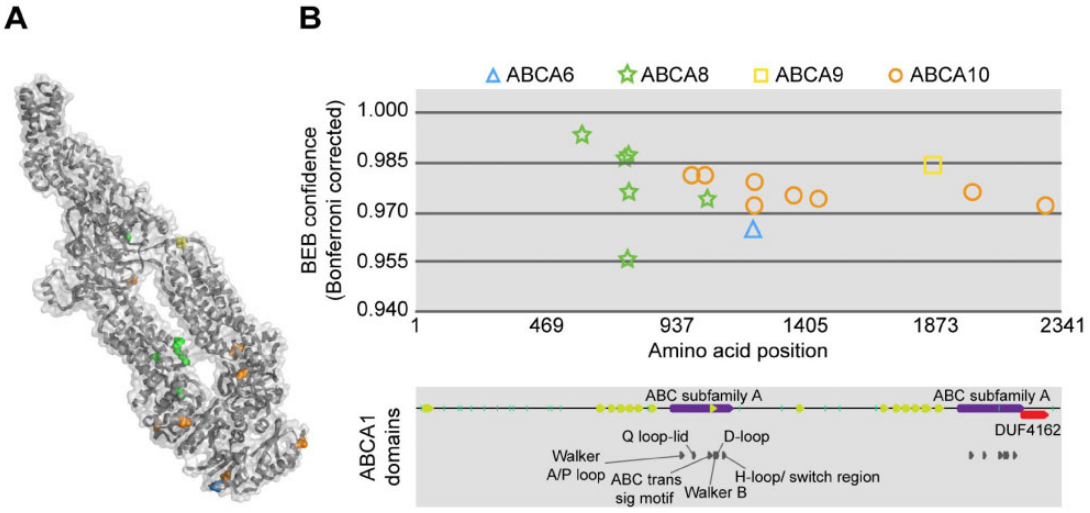
Fig. 3.—GA branch analysis of *ABCA6*-like genes reveals heterogeneity of evolutionary rates across genes family members and across mammalian lineages. Branch analysis for *ABCA6* (*A*), *ABCA8* (*B*), *ABCA9* (*C*), and *ABCA10* (*D*). Evolutionary rates (d*N*/d*S*) for each branch are denoted by branch color. Red branches indicate d*N*/d*S* ratios >1.0 and positive selection. Gray branches indicate d*N*/d*S* ratios between 0.5 and 1.0 and neutral selection. Blue branches indicate d*N*/d*S* ratios <0.5 and purifying selection. Evolutionary rates for each branch correspond to supplementary tables S3–S6, Supplementary Material online, for their respective *ABCA* gene.

asymmetrical evolution).

Gene Expression and Genetic Variation in Humans

Expression of Human ABCA6-like Transporters

Using data derived from the GTEx database, we explored the expression of the four members of the *ABCA6*-like subgroup in various human tissues (supplementary fig. S4, Supplementary Material online). Each exhibited a similar pattern with matching high expression in tibial nerve, ovary, mammary tissue, and adipose tissue. It is important to note that the expression profile of *ABCA10*, as annotated in the GTEx database, was entirely derived from short truncated transcripts or those that contain premature stop codons.



C

Gene	Coordinates (AA)	BEB	Region/Domain
ABCA8	603	0.993	-
ABCA8	758	0.986	Transmembrane region
ABCA8	768	0.956	-
ABCA8	772	0.987	Transmembrane region
ABCA8	773	0.976	Transmembrane region
ABCA10	1001	0.981	ABC subfamily A
ABCA10	1050	0.981	ABC subfamily A
ABCA8	1058	0.974	ABC subfamily A
ABCA10	1229	0.972	-
ABCA10	1230	0.979	-
ABCA6	1235	0.964	-
ABCA10	1372	0.975	-
ABCA10	1462	0.974	-
ABCA9	1878	0.984	-
ABCA10	2022	0.976	Walker A/P-loop
ABCA10	2287	0.972	DUF4162

Fig. 4.—Molecular evolutionary analyses reveal signatures of adaptive sequence evolution within *ABCA6*-like protein products. (*A*) Amino acid sites identified as under positive selection ($\omega > 1$) by codeml within marmoset ABCA6 (XM_008997646.2), armadillo ABCA8 (XM_12525536.1), pangolin ABCA9 (XM_0176755120.1), and rhesus macaque ABCA10 (XM_015120204.1) mapped to positionally homologous sites in the cryo-EM structure of human ABCA1 (PDB ID: 5XJY) in blue, green, yellow, and orange, respectively. Following Bonferroni correction for multiple comparisons, the marmoset ABCA6, armadillo ABCA8, and pangolin ABCA9 amino acid sites were considered significantly under positive selection, whereas the sites under positive selection in rhesus macaque ABCA10 were considered significant following the Benjamini–Hochberg procedure for multiple comparisons (FDR = 0.2). For complete information on LRTs, chi-square tests, and *P*-values found to be significant under Bonferroni correction [corrected *P*-values $\leq$ 0.007] or following the Benjamini–Hochberg procedure, see supplementary data 1, Supplementary Material online. For the multiple-sequence alignment human ABCA1 (PDB ID: 5XJY) and *ABCA6*-like protein products, see supplementary data 2, Supplementary Material online. (*B*) Plot of amino acid sites under positive selection for each protein encoded by an *ABCA6*-like member, indicating BEB confidence. Sites identified as under positive selection by codeml were matched to positionally homologous sites in the cryo-EM structure of human ABCA1 (PDB ID: 5XJY). Also shown are structurally and functionally important domains in the human ABCA1 protein. Within armadillo ABCA8 (XM_12525536.1), three amino acid sites in transmembrane domains and one site in the first ABC subfamily A domain were found to be under positive selection (LRT *P*-value significant following Bonferroni correction for multiple comparisons). In rhesus macaque ABCA10 (XM_015120204.1), two amino acid sites in the first ABA subfamily A domain, one site in the second ABC subfamily domain that overlaps with a Walker A/P-loop, and one site in the DUF4162 domain were found to be under positive selection (LRT *P*-value significant following the Benjamini–Hochberg procedure for multiple comparisons [FDR = 0.2]). (*C*) Table of all selected variants and their location within the protein.

This was not the case for the other three members where major isoforms either completely or mostly contribute to the expression signal. Therefore, it is unclear whether the *ABCA10* expression correlates with the presence of actual protein in these tissues. Collectively, these data suggest that although the expression of *ABCA10* is similar to other family members, it might be heavily biased toward nonfunctional transcripts.

Presence of ABCA6-like Gene LoF Mutations in Human Populations

The ExAC offers a unique insight into the distribution of genetic variants with the human population. It is clear that, as with the analysis of the evolutionary rate across species, the *ABCA6*-like genes differ in their evolutionary patterns within humans. One analysis enabled by this resource is the assessment of a gene's sensitivity, or lack thereof, to heterozygous and homozygous loss. This information has been represented by various scores, but one of particular interest is pNull, which represents the probability of a gene to be neutral to heterozygous and homozygous loss of its function (with 1 being the highest probability and 0 the lowest) (Exome Aggregation Consortium et al. 2016). This metric is based on the observed number of such losses compared with the expected from a functionally nonconstrained area; thus, a high pNull score of a gene suggests functional redundancy or absence of function. When assessing this score for the *ABCA6*-like transporters, we observed a curious pattern (supplementary table S7, Supplementary Material online). Although *ABCA6* (pNull = 0.02) was highly sensitive to homozygous loss, *ABCA8* (pNull = 1.00), *ABCA9* (pNull = 0.91), and *ABCA10* (pNull = 0.99) were not. Although the latter three appear to be similar in terms of their calculated sensitivity to homozygous LoF mutations (e.g., frameshift or stop-gain mutations), they differ significantly in terms of individuals that actually carry such a homozygous LoF mutations (supplementary table S7, Supplementary Material online). Of all the *ABCA6*-like genes, *ABCA10* has the greatest number of these mutations (fig. 5*A* and B, supplementary tables S8–S10, Supplementary Material online).

We have labeled the ten identified *ABCA10* LoF variation by order of their position from 5′ to 3′ position, V10A to V10J, respectively (supplementary table S10, Supplementary Material online). The *ABCA10* LoF variants are located on exons 5, 7, 9, 13, 14, 16, 19, 33, and 36 (supplementary fig. S5, Supplementary Material online). The other *ABCA6*-like genes (*6*, *8*, and *9*) show much lower rates of LoF homozygotes compared with *ABCA10*, with *ABCA6* showing none (supplementary tables S8–S10, Supplementary Material online). The *ABCA9* gene has just one LoF variant (supplementary table S9, Supplementary Material online). This is consistent with GA Branch analysis presented above (fig. 3C) that indicated that *ABCA9* is evolving under purifying selection. The *ABCA8* gene has four LoF variants (supplementary table S8, Supplementary Material online). The high number of LoF homozygote mutations in the human *ABCA10* gene compared with the other *ABCA6*-like genes suggests that these related genes are evolving under different types of functional constraints within humans.

We were interested to explore whether this large number of individuals that harbor LoF mutations in *ABCA10* would allow any conclusions regarding its functional role. Thus, we accessed the ExAC VCF that contains information on the number of homozygous individuals, the functional annotation of variants, and information on the affected gene. We focused our attention on those variants that result in a gained stop or a frameshift variant and applied several filters to reduce noise and to obtain genes with similar LoF frequencies as *ABCA10* (see Materials and Methods for details). We found a list of 24 genes that have multiple LoF variants with more than ten homozygous individuals in the ExAC population (fig. 5C). Among these *ABCA10* has the sixth highest total number of individuals with homozygous LoF mutations, and the highest number of homozygous sites present in at least ten individuals.

Although this indicated that this gene might have lost its function in humans, this is not conclusively supported by the other genes found in this list. For instance, the gene with the highest number of individuals with homozygous LoF mutations, *SARM1*, has an established role in activating a local destruction program following axonemal injury (Gerdts et al. 2015). *MSLNL*, on the other hand, has been inactivated during evolution and represents a pseudogene (Kim et al. 2012). Yet, although the pNull value close to 1 and the extensive number of homozygous LoF mutation in the human population do not necessarily exclude a cellular function for *ABCA10*, its loss by itself is most likely neutral in humans. Thus, enrichment of an LoF variant in *ABCA10* in a recessive autism cohort was likely one of the expected false positives and does not reflect a critical function of *ABCA10* in brain development or function (Lim et al. 2013). This is further supported by the absence of functional full-length transcript in the GTEx database.

Discussion

Here, we report our findings on the evolution of the *ABCA6*-like genes. Based on our phylogenetic analysis of the ABCAs, it appears that the *ABCA6*-like genes evolved in a burst of gene duplications at the base of the eutherian mammalian tree. We have found *ABCA6*-like gene families in all major eutherian mammalian lineages. However, platypus, a monotreme, and the three marsupial species examined lack the *ABCA6*-like gene families, that is, *ABCA6*, *8*, *9*, and *10*. Our analyses also revealed that *ABCA5* is the most likely ancestor to *ABCA6*-like families. The monophyly of the eutherian mammals and the placement of both monotremes and

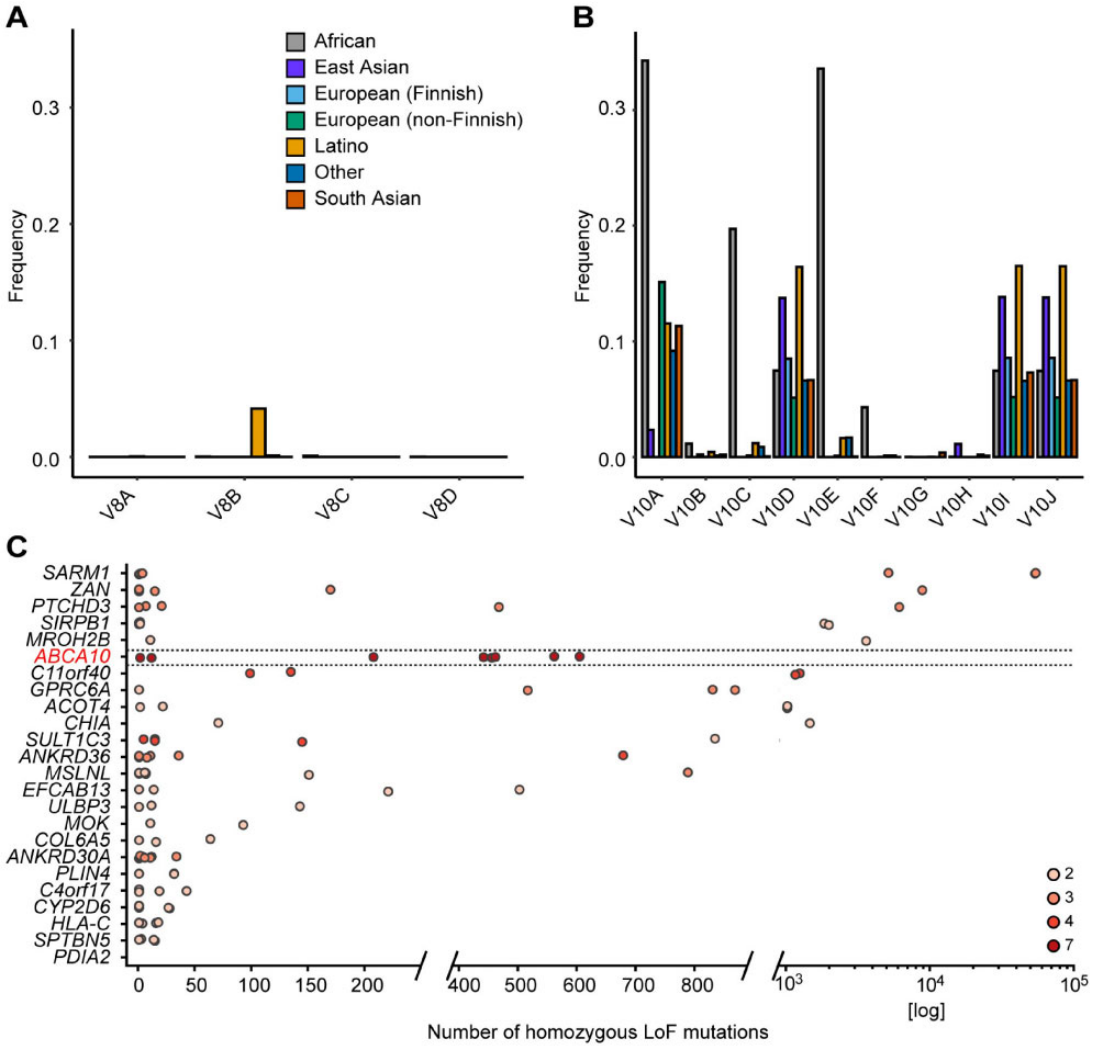


Fig. 5.—*ABCA10* is highly enriched for homozygous LoF homozygous mutations. Frequency of homozygous LoF alleles in seven human population groups based on the ExAC database for *ABCA8* (*A*) and *ABCA10* (*B*). (*C*) Plot showing all genes with similar number and distribution of high-frequency LoF mutations (see Materials and Methods for filter criteria). Shown are all LoF variants for each gene with the corresponding number of individuals that are homozygous for this mutation on the *x*-axis. In addition, all genes are color-coded with the number of homozygous LoF variants that are present in more than ten individuals. *x*-Axis is interrupted twice for better representation of the data and the third part changes to a logarithmic plot to capture all the frequencies. Genes are sorted by the total number of homozygous LoF mutations found.

marsupials outside of the eutherian, or placental, mammals are well supported (Pozzi et al. 2014; Foley et al. 2016). When the gene family phylogenies are examined in light of these known organismal relationships we can conclude that the *ABCA6*, *8*, *9*, and *10* families, while clearly related to the other ABCA transporters, are found only in the placental mammals. Organismal and molecular analyses suggest that the placental mammals underwent significant organismal divergence with many major lineages evolving over a short period of time (Pozzi et al. 2014; Foley et al. 2016). It appears that the *ABCA6*-like divergence predated the eutherian organismal diversification.

We are very much interested in the possibility that the *ABCA6*-like genes have diverged from each other in function. Although all of the *ABCA6*-like genes share sufficient sequence similarity to other ABCAs, to indicate that they are all transporter proteins, the role that each of these distinct ABCA protein families plays in mammalian physiology and diversity is still unclear. The purpose of this study was to explore patterns of evolution both across major mammalian lineages and within humans to determine whether their evolutionary history is compatible with constrained function, via purifying selection, or whether the *ABCA6*-like genes have diverged in function and have undergone positive or diversifying selection.

Our results suggest that the selective and thus functional constraints acting on the *ABCA6*-like genes have varied across the gene families and within individual gene families across organismal lineages. Rate analysis across the mammals indicates that *ABCA9* has been evolving under consistent purifying selection (fig. 3, supplementary table S3, Supplementary Material online). However, there is one amino acid site in the *ABCA9* gene of *P. paniscus* with evidence of positive selection, but this site is not in a functionally significant region (fig. 4). The other *ABCA* genes examined had much more variable rates of evolution and this suggests that the functions of these proteins may have changed significantly across organismal lineages. Of particular interest are the *ABCA8* and the *ABCA10* genes.

Our analysis of the *ABCA8* gene found four sites in armadillo that are under positive selection (fig. 4, supplementary data 1, Supplementary Material online). These sites are in functionally significant regions, three in the transmembrane regions, and one in the ABC subfamily A region (fig. 4). Our analysis also found positive and asymmetrical evolution in *ABCA8* genes in both rat and mouse (fig. 4, supplementary data 1, Supplementary Material online). The positive selection in armadillo is intriguing. Armadillos are morphologically distinct among mammals for having a hard shell that is made of plates from dermal bone with epidermal scales made of keratin (Chen et al. 2011). It has been reported that in humans, mutations in *ABCA12* causes ichthyosis (MIM: 601277 and 242500) (Annilo et al. 2002; Thomas et al. 2006). Ichthyosis is a genetic disease that is characterized by a scaling and hardness of the skin. Although our findings here are not definitive, it does suggest that further research into the role of *ABCA8* and other *ABCA* genes in the morphological evolution of the armadillo would be profitable. One possible mechanism for the *ABCA*s to play a role in the scaling of the skin or epidermal bone-like structures in armadillo would be in the transport of keratin to the outer epidermal cells where scales form. This hypothesis will require further work to evaluate.

Our rate analysis also indicated a range of evolutionary rates for the *ABACA10* genes. There is some evidence that in rhesus macaque, functionally significant sites are under positive selection (fig. 4, supplementary data 1, Supplementary Material online). The organismal implications of these findings are unknown and would be worthy of future study. It is also noteworthy that within the *ABCA10* tree, the branch leading to humans is not undergoing positive selection, but is evolving under relaxed selection (fig. 3*D*, supplementary table S6, Supplementary Material online). There have been suggestions that recessive mutations in *ABCA10* in humans may be related to brain abnormalities (e.g., autism) (Lim et al. 2013). However, based on the expression of mainly truncated transcripts in the human, together with the rarely encountered homozygous LoF rate across the human population, this appears unlikely and represents one of the rare, expected false-positive results.

One of the challenges of distinguishing positive selection from relaxed selection is that the signal of episodic positive selection can be transient. Our analyses using HyPhy and PAML are based on phylogenetic trees across mammalian diversity. In order to examine the possibility of either relaxed (LoF) or positive (new function) selection in human *ABCA6*-like genes, especially *ABCA10*, we turned to available data sets on population level variation and gene expression in humans.

It has been long established that gene evolution is often correlated to rates of gene expression (Subramanian and Kumar 2004). When gene expression patterns were analyzed for the *ABCA6*-like genes, we found largely similar gene expression patterns (supplementary fig. S2, Supplementary Material online). It is notable that *ABCA6* and *ABCA9* share the most similar gene expression patterns. Both of these genes are largely evolving under purifying selection among the mammalian lineages, and neither *ABCA6* nor *ABCA9* has low evidence of homozygous LoF variants within human populations (supplementary table S7, Supplementary Material online). In fact, of all the *ABCA* genes examined, *ABCA6* has the lowest tolerance for LoF mutations. This suggests that within humans it has the highest functional constraints and that any LoF results in significant loss of organismal fitness. The other *ABCA* genes, *8,9*, and *10* had much higher calculated tolerances for loss of fitness; however, individuals homozygous for LoF mutations for *ABCA8* and *ABCA9* are rare compared with *ABCA10* (supplementary tables S8–S10, Supplementary Material online).

The expression patterns of *ABCA10* are somewhat difficult to interpret due to the large number of isoforms present in *ABCA10*. Many of these isoforms are truncated and their functional significance is unknown. *ABCA10* also has a high number of individuals with LoF mutations (supplementary table S10, Supplementary Material online, fig. 5C). This is consistent with the gene expression data that many of the common isoforms of *ABCA10* are truncated and do not possess important functional regions of the full-length protein, suggesting a lack of function. All of these data together suggest that *ABCA10* has undergone relaxation of function in humans and is now evolving as a pseudogene in human populations. There is considerable evidence that pseudogenes can be expressed (Poliseno et al. 2010; Ji et al. 2015) and it is still possible for pseudogenes to impact fitness in some individuals and to interact via recombination with other closely related genes (Cheetham et al. 2020).

It is very important to put human population variation data within a robust population genetic framework. Human populations have an extremely low effective population size (N_e) (Lohmueller 2014; Henn et al. 2016). Because of this extremely low N_e natural selection has been inefficient. Thus, the loss of alleles with positive selection coefficients and the fixation of alleles with negative selection coefficients have been possible (Ohta 1992). The recent rapid increase in the

human census population has important long-term implications (Simons et al. 2014; Peischl and Excoffier 2015); however, in terms of understanding the current standing rate of diversity, the long-term N_e for humans is more important than the current census population size. What this means for the *ABCA* genes, especially *ABCA10*, is that even if a given allele in *ABCA10* had a low but positive selection coefficient it could have been lost due to drift. Conversely, even if a given *ABCA10* allele had a negative selection coefficient, it could be maintained or even fixed within the human population, by drift alone.

Henn et al. (2016) reported that when large data sets were analyzed evidence of purifying selection acting within African populations was found, but that even deleterious amino acid mutations have evolved as if they are neutral with respect to natural selection during the "Out of Africa" bottleneck. What implications do these large-scale human-population genetic processes have for the evolution of the *ABCA6*-like genes?

In a thought-provoking and thorough study, Chekalin et al. (2019) compared patterns of synonymous and nonsynonymous substitutions between Late Neolithic or Bronze Age individuals and modern Europeans. They identified metabolic pathways that were either enriched for nonsynonymous substitutions (Ka or d*N*) or that had fewer nonsynonymous substitutions than expected. They reported that in modern Europeans the ABC transporter pathway (KEGG hsa02010) was enriched in nonsynonymous substitutions when compared with individuals from the Bronze Age. Other pathways that were also enriched included those involved in drug metabolism (P450s) and antigen processing (Chekalin et al. 2019). This analysis did not distinguish between the distinct ABC families, as we have done here. As these authors point out, excess amino acid substitutions can reflect either positive selection to change function or an LoF due to pseudogenization. In their analysis, the authors were unable to distinguish between these two possible evolutionary mechanisms, but it is notable that they also found olfactory genes to have an excess of Ka in modern humans compared with Bronze Age individuals. Others have found evidence that the olfactory genes are undergoing pseudogenization in humans and primates (Gilad et al. 2003; Pierron et al. 2013; Somel et al. 2013).

Taken together, all of this suggests that *ABCA10* has undergone a relaxation of selection in the human lineage and either is now or is becoming a pseudogene. We have found that this gene is evolving as if neutral and is accumulating LoF mutations at a much higher rate than the other *ABCA6*-like genes. Not all humans possess faulty or truncated *ABCA10* genes, but these genes are accumulating nonsense mutations at a faster rate than are the other *ABCA6*-like genes. It is still unclear whether there are any fitness consequences for the presence of these variants in humans. We do know that transcripts are made and could, if translated, have negative fitness consequences. What we can say is that due to the very small human N_e, even alleles that have negative fitness consequences could have been maintained in recent human populations, most especially in the Out of Africa human populations.

Future questions to pursue include: What function(s) does *ABCA10* have in other mammals? Why are those functions no longer required in humans? Have the other *ABCA6*-like transporters replaced it in function? If so, this appears to have happened without leaving evidence of positive selection in other human *ABCA6*-like genes.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on March 3, 2018. M.W.B. was supported by an EMBO Long-Term Fellowship (No. ALTF 174-2015), which is cofunded by the Marie Curie Actions of the European Commission (Nos. LTFCOFUND2013 and GA-2013-609409), and an Erwin Schrödinger Fellowship by the Austrian Science Fund (No. J4197-B30). E.R.W. and A.M. were supported by a California State University Program for Education and Research in Biotechnology Research Development Award. T.R. was supported by the National Science Foundation (DEB 1556931/1762760).

Data Availability

All data used for this manuscript was available through publicly available databases as indicated in Materials and Methods.

Literature Cited

Annilo T, Chen ZQ, Shulenin S, Dean M. 2003. Evolutionary analysis of a cluster of ATP-binding cassette (ABC) genes. Mamm Genome. 14(1):7–20.

Annilo T, et al. 2002. Identification and characterization of a novel ABCA subfamily member, ABCA12, located in the lamellar ichthyosis region on 2q34. Cytogenet Genome Res. 98(2–3):169–176.

Ceccanti M, et al. 2016. A novel mutation in ABCA1 gene causing Tangier disease in an Italian family with uncommon neurological presentation. Front Neurol. 7:185.

Cheetham SW, Faulkner GJ, Dinger ME. 2020. Overcoming challenges and dogmas to understand the functions of pseudogenes. Nat Rev Genet. 21(3):191–201.

Chekalin E, et al. 2019. Changes in biological pathways during 6,000 years of civilization in Europe. Mol Biol Evol. 36(1):127–140.

Chen IH, et al. 2011. Armadillo armor: mechanical testing and microstructural evaluation. J Mech Behav Biomed Mater. 4(5):713–722.

Dermauw W, Van Leeuwen T. 2014. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol. 45:89–110.
Edgar RC. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5(1):113.
Exome Aggregation Consortium, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536(7616):285–291.
Foley NM, Springer MS, Teeling EC. 2016. Mammal madness: is the mammal tree of life not yet resolved? Philos Trans R Soc Lond B Biol Sci. 371(1699):20150140.
Gerdts J, Brace EJ, Sasaki Y, DiAntonio A, Milbrandt J. 2015. SARM1 activation triggers axon degeneration locally via NAD(+) destruction. Science 348(6233):453–457.
Gilad Y, Man O, Paabo S, Lancet D. 2003. Human specific loss of olfactory receptor genes. Proc Natl Acad Sci U S A. 100(6):3324–3327.
Henn BM, et al. 2016. Distance from sub-Saharan Africa predicts mutational load in diverse human genomes. Proc Natl Acad Sci U S A. 113(4):E440–E449.
Hollingworth P, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet. 43(5):429–435.
Ji Z, Song RS, Regev A, Struhl K. 2015. Many lncRNAs, 5′ UTRs, and pseudogenes are translated and some are likely to express functional proteins. Elife 4:e08890.
Jones PM, George AM. 2004. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci. 61(6):682–699.
Kaminski WE, Piehler A, Wenzel JJ. 2006. ABC A-subfamily transporters: structure, function and disease. Biochim Biophys Acta. 1762(5):510–524.
Kaminski WE, Wenzel JJ, Piehler A, Langmann T, Schmitz G. 2001. ABCA6, a novel a subclass ABC transporter. Biochem Biophys Res Commun. 285(5):1295–1301.
Karczewski KJ, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581(7809):434–443.
Kent WJ, et al. 2002. The human genome browser at UCSC. Genome Res. 12(6):996–1006.
Kim DW, et al. 2012. Inactivation of the MSLNL gene encoding mesothelin-like protein during African great ape evolution. Gene 496(1):17–21.
Lim ET, et al. 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77(2):235–242.
Locher KP. 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol. 23(6):487–493.
Lohmueller KE. 2014. The distribution of deleterious genetic variation in human populations. Curr Opin Genet Dev. 29:139–146.
McDonald JH. 2014. Handbook of biological statistcs. 3rd ed. Baltimore (MD): Sparky House Publishing.
McKinney W, editor. Proceedings of the 9th Python in Science Conference. 2010.
McLaren W, et al. 2016. The Ensembl variant effect predictor. Genome Biol. 17(1):122.
Mohammad IS, He W, Yin L. 2018. Understanding of human ATP binding cassette superfamily and novel multidrug resistance modulators to overcome MDR. Biomed Pharmacother. 100:335–348.
Moitra K, Dean M. 2011. Evolution of ABC transporters by gene duplication and their role in human disease. Biol Chem. 392(1–2):29–37.
Ohta T. 1992. The nearly neutral theory of molecular evolution. Annu Rev Ecol Syst. 23(1):263–286.
Patricio M, Abascal F, Zardoya R, Posada D. 2010. Accurate selection of models of protein evolution. Adv Bioinformatics. 74:117.
Peelman F, et al. 2003. Characterization of the ABCA transporter subfamily: identification of prokaryotic and eukaryotic members, phylogeny and topology. J Mol Biol. 325(2):259–274.
Peischl S, Excoffier L. 2015. Expansion load: recessive mutations and the role of standing genetic variation. Mol Ecol. 24(9):2084–2094.
Piehler A, Kaminski WE, Wenzel JJ, Langmann T, Schmitz G. 2002. Molecular structure of a novel cholesterol-responsive A subclass ABC transporter, ABCA9. Biochem Biophys Res Commun. 295(2):408–416.
Pierron D, Cortes NG, Letellier T, Grossman LI. 2013. Current relaxation of selection on the human genome: tolerance of deleterious mutations on olfactory receptors. Mol Phylogenet Evol. 66(2):558–564.
Poliseno L, et al. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465(7301):1033–1038.
Pond SL, Frost SD. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol. 22(3):478–485.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21(5):676–679.
Pozzi L, et al. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol Phylogenet Evol. 75:165–183.
Sasaki K, et al. 2018. ATP-binding cassette transporter A subfamily 8 is a sinusoidal efflux transporter for cholesterol and taurocholate in mouse and human liver. Mol Pharm. 15(2):343–355.
Sievers F, Higgins DG. 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. 27(1):135–145.
Simons YB, Turchin MC, Pritchard JK, Sella G. 2014. The deleterious mutation load is insensitive to recent population history. Nat Genet. 46(3):220–224.
Somel M, et al. 2013. A scan for human-specific relaxation of negative selection reveals unexpected polymorphism in proteasome genes. Mol Biol Evol. 30(8):1808–1815.
Srikant S, Gaudet R. 2019. Mechanics and pharmacology of substrate selection and transport by eukaryotic ABC exporters. Nat Struct Mol Biol. 26(9):792–801.
Stamatakis A. 2015. Using RAxML to infer phylogenies. Curr Protoc Bioinformatics. 51:6.14.1–6.14.14.
Subramanian S, Kumar S. 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 168(1):373–381.
Thomas AC, et al. 2006. ABCA12 is the major harlequin ichthyosis gene. J Invest Dermatol. 126(11):2408–2413.
Tsuruoka S, et al. 2002. Functional analysis of ABCA8, a new drug transporter. Biochem Biophys Res Commun. 298(1):41–45.
van Leeuwen EM, et al. 2015. Genome of the Netherlands population-specific imputations identify an ABCA6 variant associated with cholesterol levels. Nat Commun. 6(1):6065.
Waskom M, et al. 2017. mwaskom/seaborn: v0.8.1 (September 2017). Available from: https://doi.org/10.5281/zenodo.883859.
Weaver S, et al. 2018. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. Mol Biol Evol. 35(3):773–777.
Wenzel JJ, et al. 2003. ABCA10, a novel cholesterol-regulated ABCA6-like ABC transporter. Biochem Biophys Res Commun. 306(4):1089–1098.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.

Associate editor: Mary O'Connell