Title:

The evolution of the mammalian ABCA6-like genes: analysis of phylogenetic, expression and population genetic data reveals complex evolutionary histories.

Authors and Affiliations

Martin W. Breuss\textsuperscript{a,b,\#}

Allen Mamerto\textsuperscript{c,\#}

Tanya Renner\textsuperscript{d,\#}

Elizabeth R. Waters\textsuperscript{c,e}

\textsuperscript{a}: Department of Neurosciences, University of California, San Diego, La Jolla, CA, USA

\textsuperscript{b}: Rady Children’s Institute for Genomic Medicine, San Diego, CA, USA

\textsuperscript{c}: Department of Biology and Program in Biological and Medical Informatics, San Diego State University, 5500 Campanile Drive, San Diego, CA 92182.

\textsuperscript{d}: Department of Entomology, The Pennsylvania State University, 501 Agricultural Sciences & Industries Building, University Park, PA 16802, USA

\textsuperscript{e}: corresponding author ewaters@sdsu.edu

\textsuperscript{\#}: these authors contributed equally.
Abstract

ABC membrane transporters are a large and complex super-family 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 Statement

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 multi-gene 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 (Dermauw and Van Leeuwen 2014; Kaminski, et al. 2006; Peelman, et al. 2003). 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 (Figure 1A). The entire transporter can be functionally divided into two halves that contain 6 transmembrane domains and one nucleotide binding domain
(NBD) (Figure 1B). 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 (Figure 1B). Subsequent ATP-hydrolysis and release of ADP and inorganic phosphate will revert the ABCA transporter to its ground state (Figure 1B).

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 (Annilo, et al. 2003; Kaminski, et al. 2001; Piehler, et al. 2002; Tsuruoka, et al. 2002; 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; Sasaki, et al. 2018; Tsuruoka, et al. 2002; Wenzel, et al. 2003). 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 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 \textit{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 \textit{ABCA6}-like clade (\textit{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 \textit{ABCA} genes in humans: are these genes evolving towards new functions via positive selection or are they evolving neutrally suggesting a loss of function?

We have found that the \textit{ABCA6}-like genes are absent in monotremes or marsupials. This suggests that a burst of gene duplications generated the \textit{ABCA6}-like gene families at the base of the Eutherian mammalian divergence (Foley, et al. 2016). The \textit{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 \textit{ABCA6}-like genes in eutherian mammals suggests high organismal tolerance for loss of individual \textit{ABCA6}-like genes and that many of the encoded proteins have overlapping functions. Finally, it appears that the \textit{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 was obtained from the 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.

**ExAC analysis.** Observed variants, constraint metrics, and the frequencies of loss of function variants for seven ethnic populations of human exomes were collected from the ExAC Browser. Expression of the ABCA member 10 gene was taken from the GTEx portal and used to map the loss of function variants for member 10. Bar plots for population loss of function variants were generated using R.

**Loss-of-function variant analysis.** The ExAC VCF and the pLI/pRec/pNull table were obtained from the Exome Aggregation Consortium repository on March 14th 2018 (Karczewski, et al. 2020; Lek, et al. 2016). 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 non-genic/non-functional regions of the genome and are likely functionally redundant or obsolete. This metric is based on the catalogue of all observed 'null' mutations (e.g. frameshift, stop-gain) which are compared to 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 loss-of-function variants,
LoF) 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 10 individuals or that had a clustering of all LoF variants within 500 bp. Finally, ambiguous multi-variant sites resulting from conversion by VEP were manually curated (McLaren, et al. 2016). All the data processing, analysis, and visualization was 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 non-redundant 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 less than 1e-179 with a minimum score of 9500 were selected for each species and used for analysis. Accession numbers for the sequences identified and used for further analysis are provided in Table S2. 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/) (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 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 (AIC) to identify the best fitting model for the number of branch ω (dN/dS) classes. The GA-branch program (Pond and Frost 2005) assigns individual branches within the tree to discreet dN/dS ratio classes. A model-averaged probability of positive selection (dN/dS > 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 (dN/dS) 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 (ω₀ = ω₁) versus two-ratio model 2 (ω₀, ω₁)), divergent selection (M3 versus clade model D), and positive selection (model Aₚ (ω₂ = 1) versus model A (0 < ω₀ < 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 (FDR = 0.2) (McDonald 2014). Sites identified as under positive selection (ω > 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 un-gapped 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. A multiple sequence alignment of human ABCA1 (PDB ID: 5XJY) and ABCA6-like protein products with amino sites identified as under significant positive selection by codeml (PAML v4.4) (Yang 2007) is available as Supplementary Data 2.

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* (Figure 1C, Figure S1, Table S1) revealed conserved synteny. The ABCA6-like genes in both *H. sapiens* and *P. troglodytes* are found on chromosome 17 in the following 5’ to 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 thirty species representing the major placental mammalian lineages and the closely related monotremes and marsupials. In Figure 1D 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 (Table S2). 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 non-sense 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 (*R. norvegicus*) genome is also missing the ABCA10 gene. However, while rat lacks a functioning ABCA6 gene, it does possess an ABCA6 pseudogene. The rabbit (*O. cuniculus*) also lacks the ABCA10 gene and while the ABCA9 gene is present it is a pseudogene in this species. The orangutan (*P. abelii*) has ABCA10, but not 6, 8, or 9. Gorilla (*G. 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 (*H. sapiens*), bonobo (*P. paniscus*), and chimpanzee (*P. 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, Pan paniscus, Pan trioglodytes, and Homo 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 Figure S2 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 (Figure S2). 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, while 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 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 (Figure 2). The high bootstrap values for each of the ABCA6-like genes 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 Figure 1D 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). The estimated dN/dS (ω) ratios shown in Figure 3 (Tables S3-
reveal that the rates of positive, neutral, and purifying selection vary both within and between the \textit{ABCA6}-like gene families. For example, most \textit{ABCA6} genes are evolving under a mix of purifying and relaxed selection; however, the branch leading to \textit{Pan troglodytes} has evolved under positive selection (dN/dS 37.4) (Figure 3A, Table S3). In the \textit{ABCA8} tree, it is clear that the only branch evolving under positive selection (dN/dS 2.2) is the branch leading to the carnivores (Figure 3B, Table S4). In the \textit{ABCA9} tree, only the branch leading to \textit{Pan paniscus}, has evolved under positive selection (dN/dS 96.5) (Figure 3C, Table S5). The rest of the \textit{ABCA9} tree is evolving under strong purifying selection. It is worthy of note that there is evidence of relaxed selection (dN/dS 0.58) acting on the \textit{ABCA10} genes in the branch leading to \textit{Homo sapiens} (Figure 3D, Table S6). This is in contrast to the all the other \textit{ABCA6}-like genes 6, 8 and 9, where the branches leading to \textit{H. sapiens} are evolving under purifying selection (dN/dS 0.30, 0.30, 0.28) Tables S3, S4, S5 respectively.

In order to further refine our understanding of the evolutionary constraints acting on the \textit{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 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 \textit{ABCA6}-like genes are evolving under purifying selection. Our analysis shows that the \textit{ABCA9} genes are evolving under significant
divergent evolution. In the \textit{ABCA9} gene tree (Figure 4, Figure S3), five tips are under asymmetrical evolution: mouse, dog, cat, pangolin and armadillo (Supplemental Data 1). However, pangolin \textit{ABCA9} had one site under positive selection; but this site is not associated with a named structural or functional domain (Figure 4, Figure S3). In rhesus macaque \textit{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 (Figure 4, Figure S3, Supplemental Data 1). Although these rhesus macaque \textit{ABCA10} amino acid sites were only found to be significant following carrying out the Benjamini-Hochberg procedure for multiple comparisons on likelihood ratio test 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 \textit{ABCA10} has evolved under significant asymmetrical and divergent evolution (Supplementary Data 1).

It is notable that this analysis found four sites within armadillo \textit{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 (Figure 4, Supplementary Data 1). Armadillo \textit{ABCA8} was also found to be under asymmetrical and divergent evolution. This analysis also revealed that mouse \textit{ABCA8b} is under significant asymmetrical evolution and an internal node leading to rabbit \textit{ABCA8}, mouse \textit{ABCA8a}
and rat ABCA8a in the ABCA8 phylogeny to have a site under positive selection (and also under significantly 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 (Figure S4). 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. 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 while the expression of ABCA10 is similar to other family members, it might be heavily biased towards non-functional transcripts.

*Presence of ABCA6-like gene Loss of Function mutations in human populations.* The Exome Aggregation Consortium (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)
(Lek et al., 2016). This metric is based on the observed number of such losses
compared to the expected from a functionally non-constrained 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
(Table S7). While 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. While
the latter three appear to be similar in terms of their calculated sensitivity to
homozygous loss-of-function (LoF) mutations (e.g. frameshift or stop-gain mutations),
they differ significantly in terms of individuals that actually carry such a homozygous
LoF mutations (Table S7). Of all the ABCA6-like genes, ABCA10 has the greatest
number of these mutations (Figure 5A-B, Table S8-10).

We have labelled the 10 identified ABCA10 LoF variation by order of their
position from 5' to 3' position, V10A to V10J respectively (Table S10). The ABCA10 LoF
variants are located on exons 5, 7, 9, 13, 14, 16, 19, 33, and 36 (Figure S5). The other
ABCA6-like genes (6, 8, and 9) show much lower rates of LoF homozygotes compared
to ABCA10, with ABCA6 showing none (Tables S8-S10). The ABCA9 gene has just one
LoF variant (Table S9). This is consistent with GA Branch analysis presented above
(Figure 3C) that indicated that ABCA9 is evolving under purifying selection. The ABCA8
gene has four LoF variants (Table S8). The high number of LoF homozygote mutations
in the human ABCA10 gene compared to 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 methods for details). We found a list of 24 genes that have multiple LoF variants with more than 10 homozygous individuals in the ExAC population (Figure 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 10 individuals.

While 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, while 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 a 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, i.e., 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 marsupials outside of the eutherian, or placental, mammals are well supported (Foley, et al. 2016; Pozzi, et al. 2014). 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 (Foley, et al. 2016; Pozzi, et al. 2014). It appears that the ABCA6-like divergence pre-dated the eutherian organismal diversification.

We are very much interested in the possibility that the ABCA6-like genes have diverged from each other in function. While 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 if
their evolutionary history is compatible with constrained function, via purifying selection,
or if 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 (Figure 3, Table S3).
However, there is one amino acid site in the ABCA9 gene of Pan 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 (Figure 4, Supplementary Data 1). These sites are in functionally
significant regions, three in the transmembrane regions, and one in the ABC subfamily
A region (Figure 4). Our analysis also found positive and asymmetrical evolution in
ABCA8 genes in both rat and mouse (Figure 4, Supplementary Data 1). 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 \textit{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. While our findings here are not definitive, it does suggest that further research into the role of \textit{ABCA8} and other \textit{ABCA} genes in the morphological evolution of the armadillo would be profitable. One possible mechanism for the \textit{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 \textit{ABACA10} genes. There is some evidence that in rhesus macaque, functionally significant sites are under positive selection (Figure 4, Supplementary Data 1). The organismal implications of these findings are unknown and would be worthy of future study. It is also noteworthy that within the \textit{ABCA10} tree, the branch leading to humans is not undergoing positive selection, but is evolving under relaxed selection (Figure 3D, Table S6). There have been suggestions that recessive mutations in \textit{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 (loss of function) or positive (new function) selection in human \textit{ABCA6}-like genes, especially \textit{ABCA10}, we turned to available datasets 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 \textit{ABCA6}-like genes, we found largely similar gene expression patterns (Figure S2). It is notable that \textit{ABCA6} and \textit{ABCA9} share the most similar gene expression patterns. Both of these genes are largely evolving under purifying selection among the mammalian lineages, and neither \textit{ABCA6} nor \textit{ABCA9} have low evidence of homozygous LoF variants within human populations (Table S7). In fact, of all the \textit{ABCA} genes examined, \textit{ABCA6} has the lowest tolerance for loss of function mutations. This suggests that within humans it has the highest functional constraints and that any loss of function results in significant loss of organismal fitness. The other \textit{ABCA} genes, 8, 9, and 10 had much higher calculated tolerances for loss of fitness, however individuals homozygous for LoF mutations for \textit{ABCA8} and \textit{ABCA9} are rare compared to \textit{ABCA10} (Tables S8-S10).

The expression patterns of \textit{ABCA10} are somewhat difficult to interpret due to the large number of isoforms present in \textit{ABCA10}. Many of these isoforms are truncated and their functional significance is unknown. \textit{ABCA10} also has a high number of individuals with LoF mutations (Table S10, Figure 5C). This is consistent with the gene expression data that many of the common isoforms of \textit{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 \textit{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 (Ji, et al. 2015; Poliseno, et al. 2010) 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. 2019).

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 ($N_e$) (Henn, et al. 2016; Lohmueller 2014). 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 has been possible (Ohta 1992). The recent rapid increase in the human census population has important long-term implications (Peischl and Excoffier 2015; Simons, et al. 2014); 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 \textit{ABCA} genes, especially \textit{ABCA10}, is that even if a given allele in \textit{ABCA10} had a low but positive selection coefficient it could have been lost due to drift. Conversely, even if a given \textit{ABCA10} allele had a negative selection coefficient, it could be maintained or even fixed within the human population, by drift alone.

Henn and colleagues reported that when large datasets 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 (Henn, et al. 2016). 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 and colleagues compared patterns of synonymous and non-synonymous substitutions between Late Neolithic or Bronze Age individuals and modern Europeans (Chekalin, et al. 2019). They identified metabolic pathways that were either enriched for non-synonymous substitutions (Ka or dN) or that had fewer nonsynonymous substitutions than expected. They reported that in modern Europeans the ABC transporter pathway (KEGG hsa02010) was enriched in non-synonymous substitutions when compared to 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 a loss of function 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 to Bronze Aged 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 if 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.
Data Availability Statement

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

Acknowledgements.

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 03/03/2018. M.W.B. was supported by an EMBO Long-Term Fellowship (no. ALTF 174-2015), which is co-funded 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. J 4197-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)
References

Annilo T, Chen ZQ, Shulenin S, Dean M 2003. Evolutionary analysis of a cluster of ATP-binding cassette (ABC) genes. Mamm Genome 14: 7-20. doi: 10.1007/s00335-002-2229-9
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: 169-176. doi: 10.1159/000069811
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. doi: 10.3389/fneur.2016.00185
Cheetham SW, Faulkner GJ, Dinger ME 2019. Overcoming challenges and dogmas to understand the functions of pseudogenes. Nat Rev Genet. doi: 10.1038/s41576-019-0196-1
Chekalin E, et al. 2019. Changes in Biological Pathways During 6,000 Years of Civilization in Europe. Mol Biol Evol 36: 127-140. doi: 10.1093/molbev/msy201
Chen IH, et al. 2011. Armadillo armor: mechanical testing and micro-structural evaluation. J Mech Behav Biomed Mater 4: 713-722. doi: 10.1016/j.jmbbm.2010.12.013
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. doi: 10.1016/j.ibmb.2013.11.001
Edgar RC 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113. doi: 10.1186/1471-2105-5-113
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. doi: 10.1098/rstb.2015.0140
Gerdts J, Brace EJ, Sasaki Y, DiAntonio A, Milbrandt J 2015. SARM1 activation triggers axon degeneration locally via NAD(+) destruction. Science 348: 453-457. doi: 10.1126/science.1258366
Gilad Y, Man O, Paabo S, Lancet D 2003. Human specific loss of olfactory receptor genes. Proceedings of the National Academy of Sciences of the United States of America 100: 3324-3327. doi: 10.1073/pnas.0535697100
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: E440-449. doi: 10.1073/pnas.1510805112
Hollingworth P, et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPB1A1, CD33 and CD2AP are associated with Alzheimer's disease. Nat Genet 43: 429-435. doi: 10.1038/ng.803
Ji Z, Song RS, Regev A, Struhl K 2015. Many IncRNAs, 5' UTRs, and pseudogenes are translated and some are likely to express functional proteins. Elife 4. doi: ARTN e08890 10.7554/eLife.08890
Jones PM, George AM 2004. The ABC transporter structure and mechanism: perspectives on recent research. Cell Mol Life Sci 61: 682-699. doi: 10.1007/s00018-003-3336-9
Kaminski WE, Pichler A, Wenzel JJ 2006. ABC A-subfamily transporters: structure, function and disease. Biochim Biophys Acta 1762: 510-524. doi: 10.1016/j.bbadis.2006.01.011

29
Kaminski WE, Wenzel JJ, Piehler A, Langmann T, Schmitz G 2001. ABCA6, a novel a subclass ABC transporter. Biochem Biophys Res Commun 285: 1295-1301. doi: 10.1006/bbrc.2001.5326

Karczewski KJ, et al. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 581: 434-443. doi: 10.1038/s41586-020-2308-7

Kent WJ, et al. 2002. The human genome browser at UCSC. Genome Research 12: 996-1006. doi: 10.1101/gr.229102

Kim DW, et al. 2012. Inactivation of the MSLNL gene encoding mesothelin-like protein during African great ape evolution. Gene 496: 17-21. doi: 10.1016/j.gene.2012.01.005

Lek M, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. Nature 536: 285-291. doi: 10.1038/nature19057

Lim ET, et al. 2013. Rare complete knockouts in humans: population distribution and significant role in autism spectrum disorders. Neuron 77: 235-242. doi: 10.1016/j.neuron.2012.12.029

Locher KP 2016. Mechanistic diversity in ATP-binding cassette (ABC) transporters. Nat Struct Mol Biol 23: 487-493. doi: 10.1038/nsmb.3216

Lohmueller KE 2014. The distribution of deleterious genetic variation in human populations. Current Opinion in Genetics & Development 29: 139-146. doi: 10.1016/j.gde.2014.09.005

McDonald JH. 2014. Handbook of Biological Statistics, 3rd ed. Baltimore, Maryland: 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: 122. doi: 10.1186/s13059-016-0974-4

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. doi: 10.1016/j.biopha.2018.02.038

Moitra K, Dean M 2011. Evolution of ABC transporters by gene duplication and their role in human disease. Biol Chem 392: 29-37. doi: 10.1515/BC.2011.006

Ohta T 1992. The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. 23: 263-286.

Patricio M, Abascal F, Zardoya R, Posada D 2010. Accurate Selection of Models of Protein Evolution. Advances in Bioinformatics 74: 117-+. doi: Doi 10.1007/978-3-642-13214-8_15

Peelman F, et al. 2003. Characterization of the ABCA transporter subfamily: identification of prokaryotic and eukaryotic members, phylogeny and topology. J Mol Biol 325: 259-274.

Peischl S, Excoffier L 2015. Expansion load: recessive mutations and the role of standing genetic variation. Mol Ecol 24: 2084-2094. doi: 10.1111/mec.13154

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: 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. Molecular Phylogenetics and Evolution 66: 558-564. doi: 10.1016/j.ympev.2012.07.032
Poliseno L, et al. 2010. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. Nature 465: 1033-U1090. doi: 10.1038/nature09144

Pond SL, Frost SD 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. Mol Biol Evol 22: 478-485. doi: 10.1093/molbev/msi031

Pond SL, Frost SD, Muse SV 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679. doi: 10.1093/bioinformatics/bti079

Pozzi L, et al. 2014. Primate phylogenetic relationships and divergence dates inferred from complete mitochondrial genomes. Mol Phylogenet Evol 75: 165-183. doi: 10.1016/j.ympev.2014.02.023

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: 343-355. doi: 10.1021/acs.molpharmaceut.7b00679

Sievers F, Higgins DG 2018. Clustal Omega for making accurate alignments of many protein sequences. Protein Sci 27: 135-145. doi: 10.1002/pro.3290

Simons YB, Turchin MC, Pritchard JK, Sella G 2014. The deleterious mutation load is insensitive to recent population history. Nature Genetics 46: 220-+. doi: 10.1038/ng.2896

Somel M, et al. 2013. A Scan for Human-Specific Relaxation of Negative Selection Reveals Unexpected Polymorphism in Proteasome Genes. Molecular Biology and Evolution 30: 1808-1815. doi: 10.1093/molbev/mst098

Srikant S, Gaudet R 2019. Mechanics and pharmacology of substrate selection and transport by eukaryotic ABC exporters. Nat Struct Mol Biol 26: 792-801. doi: 10.1038/s41594-019-0280-4

Stamatakis A 2015. Using RAxML to Infer Phylogenies. Curr Protoc Bioinformatics 51: 6 14 11-16 14 14. doi: 10.1002/0471250953.bi0614s51

Subramanian S, Kumar S 2004. Gene expression intensity shapes evolutionary rates of the proteins encoded by the vertebrate genome. Genetics 168: 373-381. doi: 10.1534/genetics.104.028944

Thomas AC, et al. 2006. ABCA12 is the major harlequin ichthyosis gene. J Invest Dermatol 126: 2408-2413. doi: 10.1038/sj.jid.5700455

Tsuruoka S, et al. 2002. Functional analysis of ABCA8, a new drug transporter. Biochim Biophys Res Commun 298: 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: 6065. doi: 10.1038/ncomms7065

Waskom M, et al. 2017. mwaskom/seaborn: v0.8.1 (September 2017). In.

Weaver S, et al. 2018. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol Biol Evol 35: 773-777. doi: 10.1093/molbev/msx335

Wenzel JJ, et al. 2003. ABCA10, a novel cholesterol-regulated ABCA6-like ABC transporter. Biochim Biophys Res Commun 306: 1089-1098.

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24: 1586-1591. doi: 10.1093/molbev/msm088
Figures Legends

Figure 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$_2$N) 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 (Dermauw and Van Leeuwen 2014; Wenzel, et al. 2003). (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/). Forward arrows indicate 5’ to 3’ orientation, reverse arrows indicate 3’ to 5’. Chromosome positions are as follows: *H.sapiens* Chr17: 68,868,085 - 69,314,415; *P.troglodytes* Chr17: 57,633,672 - 67,708,059; *B.taurus* Chr19: 61,873,253 - 62,153,618; *C.lupus* Chr9: 15,421,150 - 15,767,955; *M.musculus* Chr11: 109,934,510 - 110,329,238. (D) Gene loss and retention of ABCA genes across diverse mammalian lineages. Organismal tree is based on the current consensus of mammalian relationships (Foley, et al. 2016; Pozzi, et al. 2014). 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 Table S2.
Figure 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.

Figure 3. GA Branch analysis of ABCA6-like genes reveals heterogeneity of evolutionary rates across genes family members and across mammalian lineages. (A-D) Branch analysis for ABCA6 (A), ABCA8 (B), ABCA9 (C), and ABCA10 (D). Evolutionary rates (dN/dS) for each branch are denoted by branch color. Red branches indicate dN/dS ratios greater than 1.0 and positive selection. Gray branches indicate dN/dS ratios between 0.5 and 1.0 and neutral selection. Blue branches indicate dN/dS ratios less than 0.5 and purifying selection. Evolutionary rates for each branch correspond to Tables S3 to S6 for their respective ABCA gene.

Figure 4. Molecular evolutionary analyses reveal signatures of adaptive sequence evolution within ABCA6-like protein products. (A) Amino acid sites identified as under positive selection (ω > 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 sites were considered significantly under positive selection, while 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 ≤ 0.007] or following the Benjamini-Hochberg procedure see Supplementary Data 1. For the multiple sequence alignment human ABCA1 (PDB ID: 5XJY) and ABCA6-like protein products, see Supplementary Data 2. (B) Plot of amino acid sites under positive selection for each protein encoded by an ABCA6-like member, indicating Bayes empirical Bayes (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 (likelihood ratio test 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 (likelihood ratio test 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.
Figure 5. **ABCA10** is highly enriched for homozygous loss-of-function (LoF) homozygous mutations. (A-B) 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 loss-of-function mutations (LoF; see 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 10 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.
Figure 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 (H2N) 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 (Pi) are released to transition back to the ground state. Figure is modeled after schematics depicted in previous publications (Dermauw and Van Leeuwen 2014; Wenzel, et al. 2003). (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/). Forward arrows indicate 5’ to 3’ orientation, reverse arrows indicate 3’ to 5’. Chromosome positions are as follows: H.sapiens Chr17: 68,868,085 - 69,314,415; P.troglodytes Chr17: 57,633,672 - 67,708,059; B.taurus Chr19: 61,873,253 - 62,153,618; C.lupus Chr9: 15,421,150 - 15,767,955; M.musculus Chr11: 109,934,510 - 110,329,238. (D) Gene loss and retention of ABCA genes across diverse mammalian lineages. Organismal tree is based on the current consensus of
mammalian relationships (Foley, et al. 2016; Pozzi, et al. 2014). 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 Table S2.
Figure 2. Gene tree of ABCA6-like genes. Tree was constructed using maximum likelihood implemented in RAXml using the following parameters: the GTR (general time reversible) 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.
Figure 3. GA Branch analysis of ABCA6-like genes reveals heterogeneity of evolutionary rates across genes family members and across mammalian lineages. (A-D) Branch analysis for ABCA6 (A), ABCA8 (B), ABCA9 (C), and ABCA10 (D). Evolutionary rates (dN/dS) for each branch are denoted by branch color. Red branches indicate dN/dS ratios greater than 1.0 and positive selection. Gray branches indicate dN/dS ratios between 0.5 and 1.0 and neutral selection. Blue branches indicate dN/dS ratios less than 0.5 and purifying selection. Evolutionary rates for each branch correspond to Tables S3 to S6 for their respective ABCA gene.
Figure 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 sites were considered significantly under positive selection, while 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 $0.007$] or following the Benjamini-Hochberg procedure see Supplementary Data 1. For the multiple sequence alignment human ABCA1 (PDB ID: 5XJY) and ABCA6-like protein products, see Supplementary Data 2. (B) Plot of amino acid sites under positive selection for each protein encoded by an ABCA6-like member, indicating Bayes empirical Bayes (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 (likelihood ratio test p-value 0.007).
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 (likelihood ratio test 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.

472x509mm (96 x 96 DPI)
Figure 5. ABCA10 is highly enriched for homozygous loss-of-function (LoF) homozygous mutations. (A-B) 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 loss-of-function mutations (LoF; see 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 10 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.