

RESEARCH ARTICLE

Open Access



# Grafting or pruning in the animal tree: lateral gene transfer and gene loss?

Julie C. Dunning Hotopp<sup>1,2,3</sup>

## Abstract

**Background:** Lateral gene transfer (LGT), also known as horizontal gene transfer, into multicellular eukaryotes with differentiated tissues, particularly gonads, continues to be met with skepticism by many prominent evolutionary and genomic biologists. A detailed examination of 26 animal genomes identified putative LGTs in invertebrate and vertebrate genomes, concluding that there are fewer predicted LGTs in vertebrates/chordates than invertebrates, but there is still evidence of LGT into chordates, including humans. More recently, a reanalysis of a subset of these putative LGTs into vertebrates concluded that there is not horizontal gene transfer in the human genome. One of the genes in dispute is an N-acyl-aromatic-L-amino acid amidohydrolase (ENSG00000132744), which encodes ACY3. This gene was initially identified as a putative bacteria-chordate LGT but was later debunked as it has a significant BLAST match to a more recently deposited genome of *Saccoglossus kowalevskii*, a flatworm, Metazoan, and hemichordate.

**Results:** Using BLAST searches, HMM searches, and phylogenetics to assess the evidence for LGT, gene loss, and rate variation in ACY3/ASPA homologues, the most parsimonious explanation for the distribution of ACY3/ASPA genes in eukaryotes involves both gene loss and bacteria-animal LGT, albeit LGT that occurred hundreds of millions of years ago prior to the divergence of gnathostomes.

**Conclusions:** ACY3/ASPA is most likely a bacteria-animal LGT. LGTs at these time scales in the ancestors of humans are not unexpected given the many known, well-characterized, and adaptive LGTs from bacteria to insects and nematodes.

**Keywords:** Horizontal gene transfer, Lateral gene transfer, Gene loss, Rate variation, Evolution, Chordate, Vertebrate, Bacteria, ASPA, ACY3

## Background

“If all the trees were one tree, what a great tree that would be.” – from a children’s nursery rhyme [1].

We have one great tree of life that grows and is pruned by evolutionary processes. In 1859, Darwin published “On the Origin of Species” describing the role natural selection plays on the evolution of species, elucidating an interplay between competition and survival [2]. Seventy years later, Frederick Griffith discovered that traits, specifically virulence, can be directly transferred between bacteria in a process we now understand to be horizontal/lateral gene transfer (HGT/LGT) [3]. It was another 16 years before Avery, MacLeod, and McCarty demonstrated that DNA is

the molecule that encodes traits and is inherited [4]. Darwin’s theory predates the discovery of DNA and as such transcends any one specific molecular mechanism.

Today, the field of molecular evolution focuses on understanding Darwinian evolution of genomes along with Kimura’s neutral theory with an emphasis on using phylogenetic techniques to analyze nucleotide sequence variation in protein coding genes. With some exceptions, this research in eukaryotes focuses on nucleotide substitutions in conserved protein-coding regions from genes deemed a priori to be vertically inherited. But as Avery et al. discovered [4], traits can also be transferred horizontally or laterally via LGT. LGT has played a major role in the natural evolution and niche adaptation of bacteria, but the role of LGT in the evolution of eukaryotic genomes has been understudied and underappreciated.

When we started working on LGT of bacterial DNA into animal genomes more than a decade ago, the prevailing paradigm was that it was non-existent. Subsequently,

Correspondence: [jdhotopp@som.umaryland.edu](mailto:jdhotopp@som.umaryland.edu)

<sup>1</sup>Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA

<sup>2</sup>Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA

Full list of author information is available at the end of the article



instances of bacteria-animal LGT have been observed in multiple invertebrates [5–34], including many such integrations of genes that have at least some evidence for being functional [9–19, 23–26, 28, 30, 32–34]. The coffee berry borer acquired a bacterial mannanase gene that allows it to exploit coffee berries as a new ecological niche relative to its sister taxa [23]. The invasive brown marmorated stink bug that ravaged crops in the mid-Atlantic region is thought to have several LGTs from bacteria, including a mannanase gene [14]. Several plant parasitic nematodes have acquired cellulases, pectate lyases, and expansin-like proteins from bacteria that allow them to degrade plant material [26, 28]. In mealybugs, LGTs from at least three different bacterial lineages have resulted in hybrid biosynthetic pathways [13]. There have been numerous functional transfers of bacterial peptidoglycan remodeling genes [11, 12, 14, 32–34] to various eukaryotes that may indicate that eukaryotes can acquire bacterial genes that the eukaryotes then use against the bacteria [35].

Despite this, LGT in multicellular eukaryotes with differentiated tissues, particularly gonads, continues to be met with skepticism. For example, Crisp et al. conducted a detailed examination of 26 animal genomes in order to identify putative LGTs in invertebrate and vertebrate genomes, including the human genome [36]. They found that there are fewer predicted LGTs in vertebrates/chordates than invertebrates, but there is still evidence of LGT into chordates, including humans [36]. Some people might not find LGT to chordates to be unusual, since chordates are known to have co-opted endogenous retroviral *env* genes multiple times during the evolution of placental mammals [37]. However, LGT from bacteria is thought to pose a higher barrier than acquisition of new functions from endogenous retroviruses.

More recently, Salzberg re-analyzed a subset of the putative LGTs in vertebrates that were proposed by Crisp et al. [36] and concluded that “horizontal gene transfer is not a hallmark of the human genome” [38]. One of the genes Crisp proposed to be a bacteria-chordate LGT [36], but Salzberg attempts to debunk [38], is an *N*-acyl-aromatic-L-amino acid amidohydrolase (ENSG00000132744), which encodes ACY3. ACY3 can convert *N*-acyl-aromatic-L-amino acid to the corresponding aromatic-L-amino acid and a carboxylate, or alternatively ACY3 can convert *N*-acetyl-L-cysteine-S-conjugate to L-cysteine-S-conjugate and acetate [39]. ACY3 has an important role in humans, catalyzing the deacetylation of mercapturic acids in kidney proximal tubules [39]. It is highly expressed in the gastrointestinal tract, the endocervix of women, and the kidneys [40, 41]. BLASTP searches of non-redundant protein database (NR) with ACY3 returns matches to the human ASPA protein. ASPA is a protein that converts *N*-acetylaspargate to aspartate and acetate [42]. In humans, mutations in

ASPA are responsible for Canavan disease, an autosomal recessive disease leading to brain defects and subsequently early death in children [42]. It is expressed in the central nervous system [40, 41]. Given that ACY3 and ASPA are homologues and our BLASTP searches, and therefore likely the BLASTP searches by Crisp et al. [36] and Salzberg [38], return both homologues, we will refer to them as the ACY3/ASPA homologues.

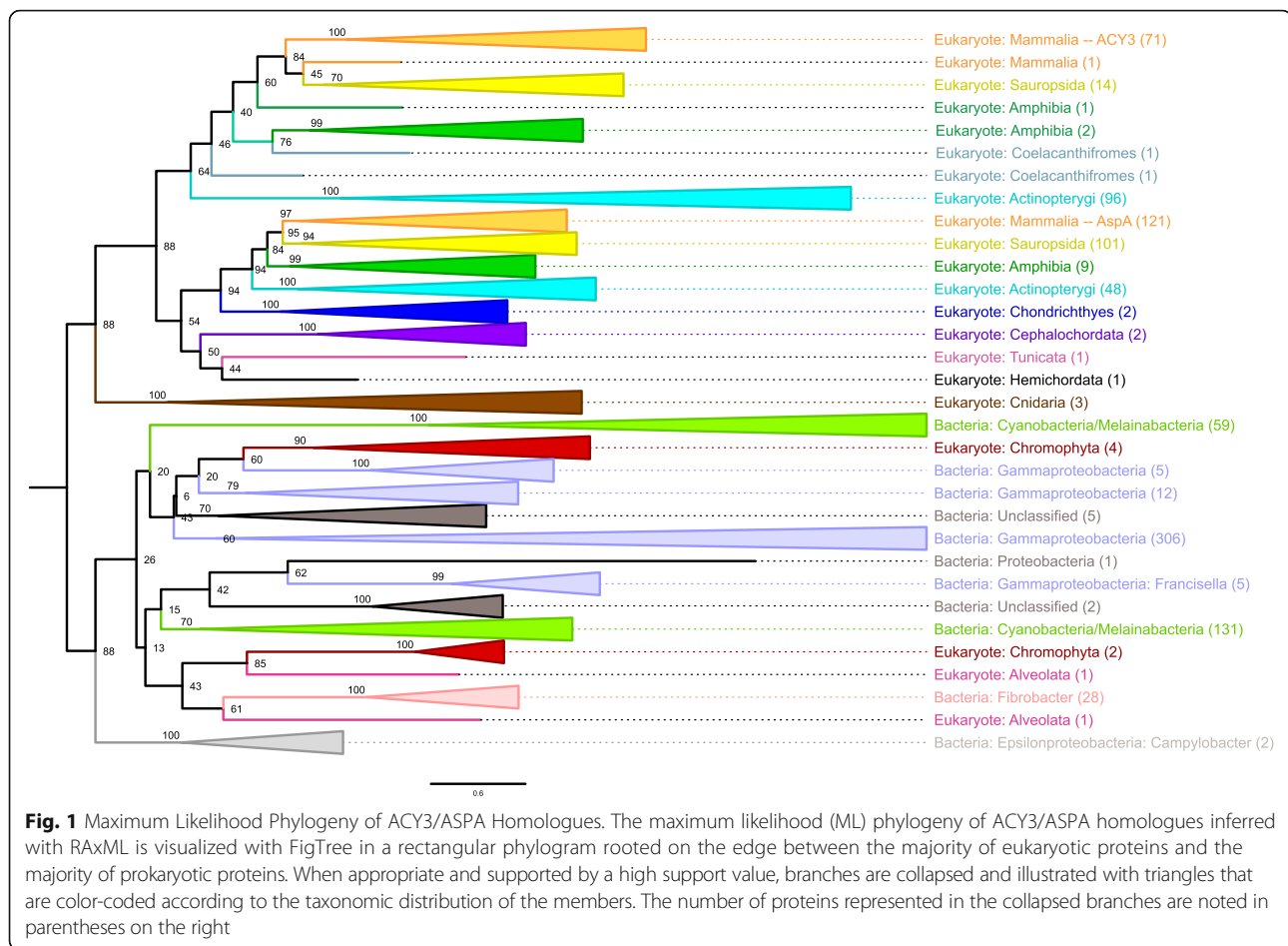
Salzberg discounted the ACY3/ASPA homologues as “no HGT” [38] because it no longer passes the test Crisp et al. [36] devised for bacteria-chordate LGT since it has a significant BLAST match to the recently deposited genome of *Saccoglossus kowalevskii* [43], a flatworm, Metazoan, and hemichordate. We sought to examine the evolutionary history of the ACY3/ASPA homologues further in an effort to better understand the evidence for LGT, gene loss, and rate variation.

## Results

### Phylogeny of human aspartoacylase

BLASTP was used to identify homologues of the human aspartoacylase gene (ENSG00000132744; NP\_542389; ACY3) in NR using the NCBI website. This search largely confirmed the BLAST-based results from Crisp [36] and Salzberg [38] demonstrating a large number of matches from bacteria and chordates, but no significant matches (e-value <1e-5) from arthropods, nematodes, plants, fungi, or apicomplexa, among others. This BLAST search identified the proteins encoding human ACY3 and the human ASPA (NP\_000040; ASPA), as well as their homologues in other animal genomes.

A maximum likelihood phylogeny was inferred with RAxML after model testing with PROTTEST on an alignment that included all well-aligning sequences from hundreds of bacteria and chordates as well as alveolates ( $n = 2$ ), chromophytes ( $n = 6$ ), cnidaria ( $n = 2$ ), and hemichordates ( $n = 1$ ) (Fig. 1). This phylogeny reveals 88% support for a clade of mostly vertebrate proteins and a clade of mostly bacteria, which initially gives an impression of LGT, with a gene moving from bacteria to vertebrates, or vice versa (Fig. 1). Further refinement of the tree, collapsing branches down to the class designation reveals that the vast majority of the eukaryotic proteins (Fig. 1) are evolving in a manner consistent with our understanding of eukaryote evolution (Fig. 2). The human ACY3 and ASPA are paralogs that likely arose following duplication. This duplication may have occurred after the divergence of bilateria (88% support) in the ancestor of deuterostomia, which includes chordates, hemichordates, and echinodermata. Alternatively, given the poor support for the position of hemichordate, tunicate, and cephalochordate ACY3/ASPA proteins (<60%) (Fig. 1), this duplication may have occurred as recently as the ancestor of gnathostomes, which includes the majority of vertebrate animals.



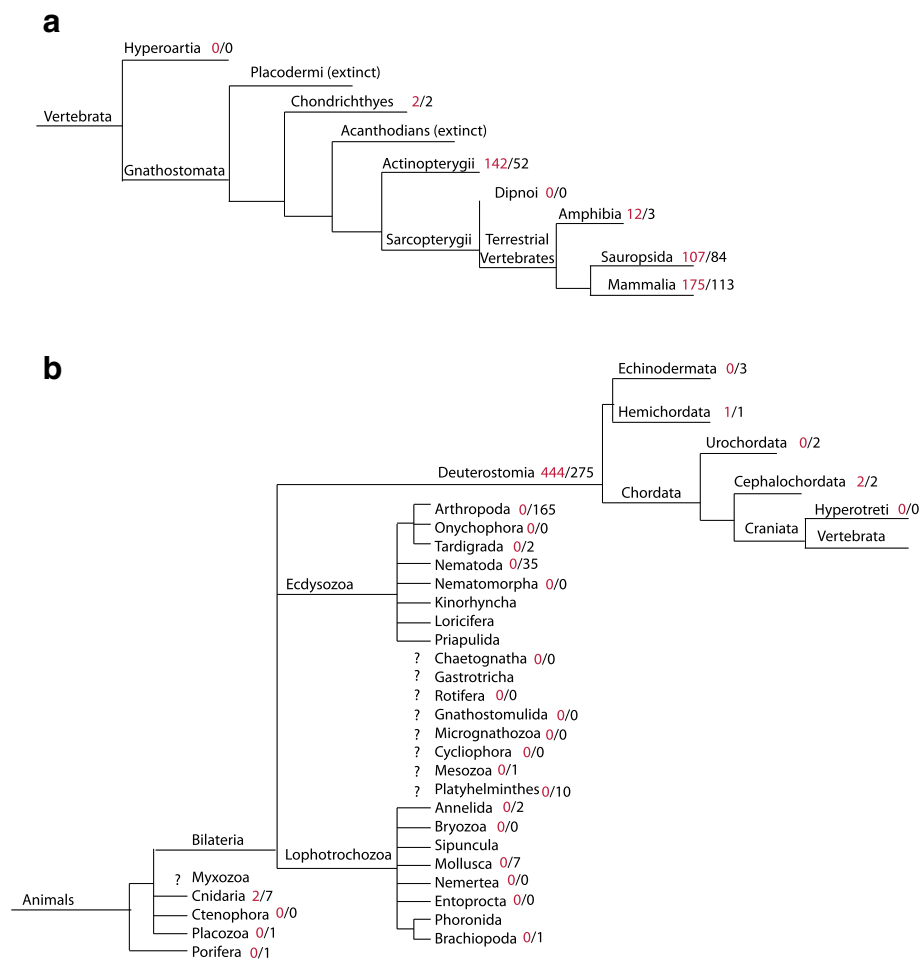
The latter is probably more likely as it is consistent with the 1R or the 2R whole genome duplications, which are predicted to have occurred in the ancestor of hyperoartia(lampreys)/hyperotreti(hagfishes) and the ancestor of gnathostomes, respectively [44]. Given the poor support values, it cannot be ruled out that ASPA/ACY3 was acquired by hemichordates, tunicates, and cephalochordates from another animal early in animal evolution.

The relationship between bacteria and eukaryote proteins is less clear in non-animals. Proteins from chromophytes and alveolates are nested amongst proteins from disparate bacterial taxa (88% support), predominantly cyanobacteria, fibrobacteria, and gamma-proteobacteria, but also two *Campylobacter*, which are epsilon proteobacteria (Fig. 1). Chromophytes and alveolates are both Chromista, a group of non-animal/metazoan eukaryotic photosynthetic organisms that likely acquired their chloroplasts from red algae. The structure of the phylogeny suggests that LGT may have occurred between the bacterial and these non-animal/metazoan eukaryotic ASPA/ACY3 homologues, although the support values for the tree topology do not allow for any further elucidation of the relationship. On one extreme, the phylogeny supports that ACY3/ASP may

have been present in the ancestor of all eukaryotes and was acquired by bacteria via LGT from the ancestor of chromophyta and alveolata, and on the other extreme, ACY3/ASP may have been acquired by animals, chromophyta, and/or alveolates from bacteria.

#### Gene loss or lateral gene transfer?

Protein phylogenies only examine the relationship between extant proteins that have been sequenced. Two alternate hypotheses to consider when examining evidence for/against LGT are gene loss and rate variation. To consider gene loss, information about lineages lacking these proteins is required since some taxa are more abundant on earth, and genome sequencing has been unevenly applied across taxa. For example, despite many arthropod and nematode sequences in NR, arthropod and nematode homologues of ACY3/ASP were not identified in the BLASTP searches. To account for this, the numbers of ACY3/ASP homologues for given taxonomic levels were compared to the number of organisms at that taxonomic level that have > 5000 protein sequences in NR. If an organism has > 5000 protein sequences in NR, any ACY3/ASP homologues are likely to have been sequenced and



**Fig. 2** Gene Loss Analysis. Phylogenies from the Tree of Life Web Project were concatenated and used to interpret gene loss. However, it is important to consider that some of the older branches in the tree of life are still disputed. The animal phylogeny was broken out into two panels illustrating: **a** the evolution of mammals from vertebrates and **b** the evolution of vertebrates from animals. To assess gene loss, the number of ACY3/ASP homologues in a given taxonomic lineage were compared to the number of organisms with > 5000 proteins deposited in public databases. ACY3/ASP homologues are consistently found in the deuterostome lineage, but are missing from some well-sequenced sister taxa like arthropods and nematodes. There are inadequate levels of genome sequence data at key taxonomic levels to enable the delineation of the relative contribution of LGT, gene loss, and rate variation for ASPA/ACY3 homologues

identified through the BLASTP searches of NR. Nearly identical results were obtained for thresholds between 5000 and 10,000 proteins, giving confidence that a threshold of 5000 proteins was neither too stringent nor lenient. However, it is important to note that while it is likely that ACY3/ASP homologues have been sequenced, genome and transcriptome assemblies can be incomplete and as such absence may be over-predicted. Contamination is also a concern, which would lead to under-predicting absence. However, in most cases at least two organisms of a taxa were sequenced and had concordant results.

Among the chordates, ACY3/ASP homologues are distributed among all of the vertebrate lineages that were sequenced sufficiently (Fig. 2a). In all vertebrate lineages, more ACY3/ASP homologues were identified in NR from that taxon than there were organisms with > 5000

protein sequences in NR for that taxon suggesting that many of the vertebrate organisms contain at least one ACY3/ASP homologue. This strongly supports the conclusion that the aspartoacylase was likely present in the ancestor of all vertebrates, or at least gnathostomata, and was duplicated.

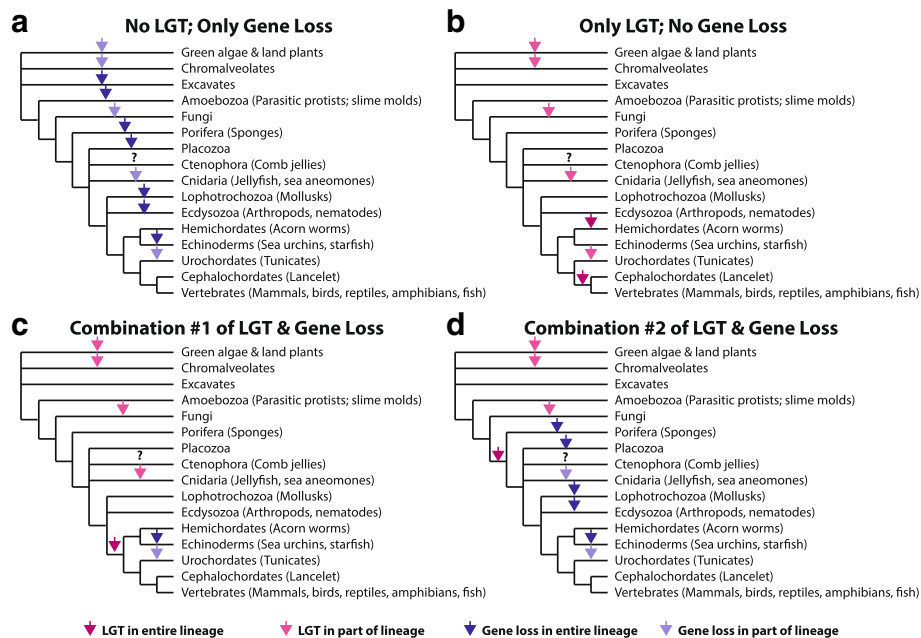
Among the deuterostomes, there is very limited sequencing outside the chordates such that only 3 echinodermata, 1 hemichordata, 2 urochordata, 2 cephalochordata, and no hyperotreti have > 5000 proteins characterized in NR (Fig. 2b). Of those, the hemichordate and the cephalochordate have ACY3/ASP homologues, and it is the BLAST match to the hemichordate homologue that led to the re-assignment of this as “not HGT” by Salzberg. If ACY3/ASP were present in the ancestor of all deuterostomes, then ACY3/ASP proteins were lost or unsequenced in the 2

urochordata and 3 echinodermata sequenced. Of note, the phylogenetic analysis of BLAST-identified homologues shows presence of this protein in at least one urochordate.

While ACY3/ASPA homologues are prevalent among the deuterostomes, they are noticeably absent in the 202 ecdysozoa with > 5000 proteins in NR, including arthropods, nematodes, and tardigrades. Likewise, they are also absent from the 10 lophotrochozoa with > 5000 proteins in NR, including annelids, brachiopods, and mollusks. They are also missing in unplaced bilateria taxa, including platyhelminths and mesozoa (Fig. 2b). Therefore, this protein has a limited taxonomic distribution within bilateria such that if the ACY3/ASPA homologue was present in the ancestor of bilateria, it would have needed to be lost from numerous lineages, including at least ecdysozoa, lophotrochozoa, urochordata, and echinodermata, given our current understanding of genomics/transcriptomics and the tree of life.

In animals, there were two ACY3/ASPA homologues in cnidaria, which are not bilateria (Fig. 2b). While there are lots of bilaterian species that have > 5000 proteins in NR, only 9 non-bilateria species have been sequenced

and had data deposited in NR across the other five taxa – 7 cnidaria, 1 placozoa, and 1 porifera – and only 2 cnidaria have an ACY3/ASPA homologue (Fig. 2b). If the ACY3/ASPA homologue was present in the ancestor of all animals, it would have had to have been lost from some cnidaria as well as placozoa and porifera, in addition to the four bilateria lineages discussed above; if it was present in the ancestor of all eukaryotes even more gene loss events are needed (Fig. 3a). Given that extensive LGTs occur in other bilateria, namely insect and nematodes, LGT alone might actually be a more parsimonious explanation than gene loss (Fig. 3b) or a combination of LGT or gene loss may be responsible for the distribution seen today (Fig. 3c, d). In other words, two LGTs of a bacterial gene into animals, one in cnidaria and one in deuterostomes, could be more likely than a half dozen gene loss events across diverse animal taxa. Unfortunately, without rate estimates for both gene loss and LGT, robust resolution of the relatedness of key unresolved taxa, and more genomic data from key animal taxa, it is not possible to be definitive or calculate the probabilities of the events.



**Fig. 3** Schematics Illustrating Possible Paths to Explain the Current Distribution of ACY3/ASPA Homologues. Phylogenies from the Tree of Life Web Project were concatenated and LGT and gene loss events were overlaid in four possible scenarios: **a** presence of ASPA/ACY3 in the last common ancestor of eukaryotes and only gene loss, **b** absence of ASPA/ACY3 in the last common ancestor of eukaryotes and only LGT, **c** combination of LGT and gene loss where LGT occurred in the ancestor of all deuterostomes, **d** combination of LGT and gene loss where LGT occurred in the ancestor of all animals. Maroon arrows are used to indicate LGT in an entire lineage, while pink arrows are used to indicate LGT in a subset of taxa represented here. Dark purple arrows are used to indicate gene loss in an entire lineage, while lavender arrows are used to indicate gene loss in a subset of taxa represented here. It is not possible at this time to determine the likelihood of all of these possible scenarios, without better resolution of the eukaryotic tree of life, more sequence data from non-animal lineages, and a better understanding of the rates of gene loss and LGT in eukaryotes, which likely vary by lineage. However, it seems improbable that gene loss alone explains these results, which suggests that some LGT from bacteria to eukaryotes, and most likely, animals is responsible for the distribution of ACY3/ASPA homologues observed today

### Rate variation

Another alternate explanation to LGT and/or gene loss is rate variation. When considering rate variation, some proteins are under accelerated rates of evolution relative to other proteins. For example, following duplication, two proteins may diverge at different rates. Following LGT, genes might be expected to undergo different rates of evolution as they enter a new environment. As such rate variation and LGT are not mutually exclusive. However, when considering rate variation as an alternative to LGT we are looking for signatures that might suggest that one lineage has vertically inherited genes that are evolving at a different rate confounding BLAST- and phylogeny-based methods.

To examine this, we relied on the results of two large pre-computed datasets, eggNOG and PFAM. EggNOG is an algorithm that currently uses graph-based unsupervised clustering to identify orthologous genes in 2031 eukaryote and prokaryote genomes. When eggNOG is interrogated with ACY3 or ASPA, an orthologous group is identified that is found to contain proteins from bacteria and metazoans, making it largely similar to the results returned with the BLAST-based results described above (Additional file 1). However, key taxa were not recovered like the hemichordate homologues (Additional file 1), likely because the genome was only recently reported.

PFAM uses hidden Markov model (HMM) searches to find functionally-related, but substantially-diverged, proteins. HMM searches rely on the use of probabilistic, hidden Markov models to identify protein homologues with great sensitivity and specificity; these models quickly and efficiently find homologues based on the presence of protein features shared between homologues (e.g. catalytic residues) not identified through traditional BLAST-based searches. These HMM results can be overlaid on a species tree (e.g. for Acy3/ASPA: [http://pfam.xfam.org/family/AstE\\_AspA#tabview=tab7](http://pfam.xfam.org/family/AstE_AspA#tabview=tab7)).

Among the metazoa, the HMM searches yielded the same results as the gene loss analysis above, except for a match to *Acyrtosiphon pisum* (pea aphid) which is an arthropod. The match is not to a protein in NR, but instead to a nearly 1 kbp region that has closest similarity to proteins annotated as succinylglutamate desuccinylase from *Pantoea* endosymbionts that is on a 7.6 kbp contig from the whole genome sequencing project (NW\_003385628.1) (Additional file 2). It is likely that this is a contig from a contaminant bacterial endosymbiont given that the contig contains multiple regions with homology to almost exclusively bacterial sequences; there is only one non-bacterial match that is to an ascomycetes in a region that also matches *Salmonella enterica* (Additional files 3 and 4).

However, in more distant eukaryotic lineages, the HMMs gave different results from eggNOG and BLAST.

Two taxonomically disparate plant taxa, *Cajanus cajan* (pigeon pea) and *Monoraphidium neglectum* (single-cell green alga) as well as 49 fungi across many diverse fungal lineages contain proteins with the ACY3/ASPA domain. Unfortunately, and similar to the problem with phylogenetic trees, no information can be gleaned about taxa lacking ACY3/ASPA domain-containing homologues. However, it is clear that the functional domain exists in taxa beyond those identified with BLAST or eggNOG searches suggesting that there can be substantial sequence divergence. However, the sequence divergence and our inability to produce high quality alignments of these sequences precludes further analyses.

### Discussion

Salzburg [38] and others have stated, when referring to LGT, that extraordinary claims require extraordinary evidence, implying that LGT is an extraordinary claim. Salzberg goes on to suggest that more mundane explanations are at play, like gene loss and rate variation [38]. It seems unlikely that gene loss and rate variation alone can explain these results. On one extreme, it seems reasonable that eukaryotes may have acquired these genes from bacteria a handful of times, once in the ancestor of fungi and at least once in animals as well as an unresolved number of times in alveolates/chromophytes. On the other extreme, it also seems equally reasonable that instead, the genes have been vertically inherited in eukaryotes with dozens of gene loss events and at least one LGT to bacteria, where it could have spread further via LGT. It is not possible to be more definitive at this time given the lack of phylogenetic resolution at key position in the tree of life and the lack of sufficient genome sequencing of key taxa, like hyperartia, hyperotreti, and ctenophore as well as placozoa, porifera, cephalochordate, urochordata, hemichordate, and echniodermata. However, the most parsimonious explanation for the distribution of ACY3/ASPA homologues involves bacteria-animal LGT. This case, of a gene essential for proper brain development and function that seems to have a limited phylogenetic distribution, illustrates some of the limitations of using an h-index or BLAST-based approach, as well as how these comparisons need careful scrutiny. It highlights the need for more robust, focused analyses on the extent of LGT, gene loss, and rate variation in eukaryotes and their influence on trait acquisition. Furthermore, unbiased estimates of LGT and gene loss rates across and between different taxa are desperately needed to understand the likelihood of both events. Our understanding of the topology of the tree of life also influences these analyses, and many important branches have yet to be resolved or remain in dispute.

### Conclusions

Collectively, this analysis demonstrates our need for further high quality complete genome and transcriptome

assemblies from key phylogenetic groups in order to have the power to infer the correct relationships between both taxa and proteins of interest in order to properly evaluate claims of LGT and gene loss. Regardless, the most parsimonious explanation for the distribution of *ACY3/ASPA* genes in eukaryotes involves both gene loss and bacteria-animal LGT, albeit LGT that occurred hundreds of millions of years ago. Given the many known, well-characterized, and adaptive lateral gene transfers from bacteria to insects and nematodes in this time frame, lateral gene transfers at these time scales in the ancestors of humans is expected.

## Methods

### BLAST searches

*ACY3/ASPA* homologues were identified from a BLASTP search [45] using *ACY3* as a query (ENSG00000132744; NP\_542389; *ACY3*) and NR as a reference using the NCBI BLAST server during August and September 2017. A similar search using *ASPA* as the query produces similar results, but all subsequent analyses were conducted on the output using *ACY3* as a query. All BLASTP searches were performed with the default parameters except that 20,000 results were allowed to be returned. To identify homologues in specific clades, the BLASTP searches were restricted to these clades using the appropriate *taxon\_id* (e.g. fungi/*taxid*:4751, plants/*taxid*:3193, arthropods/*taxid*:6656, insects/*taxid*:6960, nematodes/*taxid*:6231, mollusks/*taxid*:6447, and apicomplexan/*taxid*:5794). A neighbor joining tree and a fast-minimum evolution tree were generated using the NCBI BLAST interface with maximum sequence difference of 0.85 and Grishin distance labeling sequences by taxonomic name.

### Multiple sequence alignment, model testing, and inferring/visualizing phylogenetic trees

All of the protein sequences identified from the *ACY3*-based BLASTP searches were downloaded locally and aligned with CLUSTALW v.1.4 [46] as implemented in Bioedit v.7.2.5 [47]. Poorly aligned sequences, particularly partial sequences and isoforms, were removed manually. The sequences were then re-aligned with CLUSTALW v.1.4 [46] as implemented in Bioedit v.7.2.5 [47]. The best-fit model of amino acid substitution was determined for each of the datasets with ProtTest3.2 [48]. All 15 models of protein evolution were tested in addition to the +G parameter (i.e. including models with rate variation among sites). RAXML v.8.2.10 [49] automatically removed undetermined columns and sequence duplicates and was used to infer the phylogeny with 1000 rapid bootstrap inferences, a thorough ML search, the GAMMA model of rate heterogeneity, the ML estimate of alpha-parameter, and the JTT substitution matrix using the command `raxmlHPC -f a -m`

`PROTGAMMAJTT -p 12345 -x 12,345 -N autoMRE`. Taxonomic information from the NCBI Taxonomy database was added to the RAXML output. Accessions that lack an entry in the taxonomy database were left blank. However, in some figures the genus and species designations were added in manually after confirming the lack of an entry in the taxonomy database; these are denoted with an asterisk (\*). Phylogenetic trees were visualized with Dendroscope v.3.5.7 [50].

### Gene loss

Taxa with > 5000, > 6000, > 7000, > 8000, and > 10,000 known proteins in NR were determined by using the NCBI protein server (<https://www.ncbi.nlm.nih.gov/protein>) to search PDB, RefSeq, UniProtKB/Swiss-Prot, DDBJ, EMBL, GenBank, and PIR with the appropriate *taxon\_id* in November 2017. These results were overlaid on reference phylogenetic trees for the eukaryotic lineages that were concatenated from trees retrieved from the Tree of Life website ([tolweb.org](http://tolweb.org)) [51–62].

### Rate variation

In order to identify *ACY3/ASPA* homologues that may be subject to rate variation, pre-computed orthologous clusters in eggNOG were examined ([http://eggnogdb.embl.de/#/app/results#COG2988\\_datamenu](http://eggnogdb.embl.de/#/app/results#COG2988_datamenu)) as well as hidden markov model (HMM) search results generated by PFAM were overlaid on a species tree using the PFAM server ([http://pfam.xfam.org/family/AstE\\_AspA#tabview=tab7](http://pfam.xfam.org/family/AstE_AspA#tabview=tab7)).

## Additional files

**Additional file 1:** EggNOG tree for COG2988 as of March 2, 2018. (PDF 229 kb)

**Additional file 2:** TBLASTN search results against NR of *A. pisum* sequence from PFAM (J9KWH7) with homology to *ASPA/ACY3* homologues. (PDF 264 kb)

**Additional file 3:** BLASTN search of *A. pisum* strain LSR1 unplaced genomic scaffold, *Acyr\_2.0 Scaffold2139* (NW\_003385628.1) against NT allowing for 20,000 matches with an e-value below 0.00001. (PDF 284 kb)

**Additional file 4:** BLASTX search of *A. pisum* strain LSR1 unplaced genomic scaffold, *Acyr\_2.0 Scaffold2139* (NW\_003385628.1) against NR allowing for 20,000 matches with an e-value below 0.00001. (PDF 269 kb)

### Abbreviations

HGT: Horizontal gene transfer; HMM: Hidden Markov model; LGT: Lateral gene transfer; ML: Maximum likelihood

### Acknowledgements

We would like to thank John Werren at the University of Rochester for helpful discussions, James Munro at the Institute for Genome Science for his help with running PROTTEST, and John Mattick for his help adding taxonomy information to the phylogenetic trees from RAXML at the Institute for Genome Science.

### Funding

This work was funded by the National Science Foundation Advances in Biological Informatics (ABI-1457957) and an NIH Director's Transformative Research Award (1-R01-CA206188) to JCDH.

**Availability of data and materials**

The RAXML outputs generated are available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.4jv6312>.

**Authors' contributions**

JCDH conducted all analyses and wrote the manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD 21201, USA. <sup>2</sup>Department of Microbiology and Immunology, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

<sup>3</sup>Greenebaum Cancer Center, University of Maryland School of Medicine, Baltimore, MD 21201, USA.

Received: 22 December 2017 Accepted: 29 May 2018

Published online: 18 June 2018

**References**

- Domanska J. If all the seas were one sea. New York: Aladdin Children's Books; 1971.
- Darwin C. On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life. London: J. Murray; 1859.
- Griffith F. The significance of pneumococcal types. *J Hyg (Lond)*. 1928;27:113–59.
- Avery OT, Macleod CM, McCarty M. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a Desoxyribonucleic acid fraction isolated from pneumococcus type iii. *J Exp Med*. 1944;79:137–58.
- Dunning Hotopp JC, Clark ME, Oliveira DC, Foster JM, Fischer P, Torres MC, Giebel JD, Kumar N, Ishmael N, Wang S, et al. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science*. 2007;317:1753–6.
- Li ZW, Shen YH, Xiang ZH, Zhang Z. Pathogen-origin horizontally transferred genes contribute to the evolution of lepidopteran insects. *BMC Evol Biol*. 2011;11:356.
- Wheeler D, Redding AJ, Werren JH. Characterization of an ancient lepidopteran lateral gene transfer. *PLoS One*. 2013;8:e59262.
- Zhu B, Lou MM, Xie GL, Zhang GQ, Zhou XP, Li B, Jin GL. Horizontal gene transfer in silkworm, *Bombyx mori*. *BMC Genomics*. 2011;12:248.
- Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJ, et al. Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. *Science*. 2010;327:343–8.
- Moran NA, Jarvik T. Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science*. 2010;328:624–7.
- Nikoh N, McCutcheon JP, Kudo T, Miyagishima SY, Moran NA, Nakabachi A. Bacterial genes in the aphid genome: absence of functional gene transfer from *Buchnera* to its host. *PLoS Genet*. 2010;6:e1000827.
- Nikoh N, Nakabachi A. Aphids acquired symbiotic genes via lateral gene transfer. *BMC Biol*. 2009;7:12.
- Husnik F, Nikoh N, Koga R, Ross L, Duncan RP, Fujie M, Tanaka M, Satoh N, Bachtrog D, Wilson AC, et al. Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell*. 2013;153:1567–78.
- Ioannidis P, Lu Y, Kumar N, Creasy T, Daugherty S, Chibucos MC, Orvis J, Shetty A, Ott S, Flowers M, et al. Rapid transcriptome sequencing of an invasive pest, the brown marmorated stink bug *Halyomorpha halys*. *BMC Genomics*. 2014;15:738.
- Novakova E, Moran NA. Diversification of genes for carotenoid biosynthesis in aphids following an ancient transfer from a fungus. *Mol Biol Evol*. 2012;29:313–23.
- Sloan DB, Nakabachi A, Richards S, Qu J, Murali SC, Gibbs RA, Moran NA. Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol Biol Evol*. 2014;31:857–71.
- Nakabachi A, Ishida K, Hongoh Y, Ohkuma M, Miyagishima SY. Aphid gene of bacterial origin encodes a protein transported to an obligate endosymbiont. *Curr Biol*. 2014;24:R640–1.
- Woolfit M, Iturbe-Ormaetxe I, McGraw EA, O'Neill SL. An ancient horizontal gene transfer between mosquito and the endosymbiotic bacterium *Wolbachia pipiensis*. *Mol Biol Evol*. 2009;26:367–74.
- Klasson L, Kambris Z, Cook PE, Walker T, Sinkins SP. Horizontal gene transfer between *Wolbachia* and the mosquito *Aedes aegypti*. *BMC Genomics*. 2009;10:33.
- Kondo N, Nikoh N, Ijichi N, Shimada M, Fukatsu T. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc Natl Acad Sci U S A*. 2002;99:14280–5.
- Nikoh N, Tanaka K, Shibata F, Kondo N, Hizume M, Shimada M, Fukatsu T. *Wolbachia* genome integrated in an insect chromosome: evolution and fate of laterally transferred endosymbiont genes. *Genome Res*. 2008;18:272–80.
- Aikawa T, Anbutsu H, Nikoh N, Kikuchi T, Shibata F, Fukatsu T. Longicorn beetle that vectors pinewood nematode carries many *Wolbachia* genes on an autosome. *Proc Biol Sci*. 2009;276:3791–8.
- Acuna R, Padilla BE, Florez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee SJ, Yeats TH, Egan AN, Doyle JJ, Rose JK. Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A*. 2012;109:4197–202.
- Altincicek B, Kovacs JL, Gerardo NM. Horizontally transferred fungal carotenoid genes in the two-spotted spider mite *Tetranychus urticae*. *Biol Lett*. 2012;8:253–7.
- Craig JP, Bekal S, Hudson M, Domier L, Niblack T, Lambert KN. Analysis of a horizontally transferred pathway involved in vitamin B-6 biosynthesis from the soybean cyst nematode *Heterodera glycines*. *Mol Biol Evol*. 2008;25:2085–98.
- Danchin EGJ, Rossoo M-N, Vieiraa P, Jd A-E, Coutinhob PM, Henriessatb B, Abada P. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A*. 2010;107:17651–6.
- McNulty SN, Foster JM, Mitreva M, Dunning Hotopp JC, Martin J, Fischer K, Wu B, Davis PJ, Kumar S, Brattig NW, et al. Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer. *PLoS One*. 2010;5:e11029.
- Mayer WE, Schuster LN, Bartelmes G, Dieterich C, Sommer RJ. Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. *BMC Evol Biol*. 2011;11:13.
- Ioannidis P, Johnston KL, Riley DR, Kumar N, White JR, Olarte KT, Ott S, Tallon LJ, Foster JM, Taylor MJ, Dunning Hotopp JC. Extensively duplicated and transcriptionally active recent lateral gene transfer from a bacterial *Wolbachia* endosymbiont to its host filarial nematode *Brugia malayi*. *BMC Genomics*. 2013;14:639.
- Wu B, Novelli J, Jiang D, Dailey HA, Landmann F, Ford L, Taylor MJ, Carlow CK, Kumar S, Foster JM, Slatko BE. Interdomain lateral gene transfer of an essential ferrochelatase gene in human parasitic nematodes. *Proc Natl Acad Sci U S A*. 2013;110:7748–53.
- Koutsovoulos G, Makepeace B, Tanya VN, Blaxter M. Palaeosymbiosis revealed by genomic fossils of *Wolbachia* in a stronglyloidean nematode. *PLoS Genet*. 2014;10:e1004397.
- Metcalf JA, Funkhouser-Jones LJ, Brileya K, Reysenbach A-L, Bordenstein SR. Antibacterial gene transfer across the tree of life. *eLife*. 2014;25:3.
- Chou S, Daugherty MD, Peterson SB, Biboy J, Yang Y, Jutras BL, Fritz-Laylin LK, Ferrin MA, Harding BN, Jacobs-Wagner C, et al. Transferred interbacterial antagonism genes augment eukaryotic innate immune function. *Nature*. 2014; Gladyshev EA, Meselson M, Arkhipova IR. Massive horizontal gene transfer in bdelloid rotifers. *Science*. 2008;320:1210–3.
- Dunning Hotopp JC, Estes AM. Biology wars: the eukaryotes strike back. *Cell Host Microbe*. 2014;16:701–3.
- Crisp A, Boschetti C, Perry M, Tunnacliffe A, Micklem G. Expression of multiple horizontally acquired genes is a hallmark of both vertebrate and invertebrate genomes. *Genome Biol*. 2015;16:50.
- Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al. Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis. *Nature*. 2000;403:785–9.
- Salzberg SL. Horizontal gene transfer is not a hallmark of the human genome. *Genome Biol*. 2017;18:85.
- UniProtKB - Q96HD9 (ACY3\_HUMAN) [<http://www.uniprot.org/uniprot/Q96HD9>].



40. GTEx Portal [<https://www.gtexportal.org/home/>].
41. Consortium GT. Human genomics. The genotype-tissue expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–60.
42. UniProtKB - P45381 (ACY2\_HUMAN) [<http://www.uniprot.org/uniprot/P45381>].
43. Simakov O, Kawashima T, Marletaz F, Jenkins J, Koyanagi R, Mitros T, Hisata K, Bredeson J, Shoguchi E, Gyoja F, et al. Hemichordate genomes and deuterostome origins. *Nature*. 2015;527:459–65.
44. Flajnik MF, Kasahara M. Origin and evolution of the adaptive immune system: genetic events and selective pressures. *Nat Rev Genet*. 2010;11:47–59.
45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215:403–10.
46. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*. 1994;22:4673–80.
47. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999;41:95–8.
48. Darriba D, Taboada GL, Doallo R, Posada D. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*. 2011;27:1164–5.
49. Stamatakis A, Ludwig T, Meier H. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*. 2005;21:456–63.
50. Huson DH, Scornavacca C, Huson DH, Scornavacca C. Dendroscope 3- an interactive viewer for rooted phylogenetic trees and networks. *Syst Biol*. 2012; <https://doi.org/10.1093/sysbio/sys062>. *Systematic Biology* 2012
51. The Tree of Life Web Project. [<http://tolweb.org>].
52. Animals. Metazoa. Version 01 [<http://tolweb.org/Animals/2374/2002.01.01> i].
53. Bilateria. Triploblasts, Bilaterally symmetrical animals with three germ layers. Version 01 [<http://tolweb.org/Bilateria/2459/2002.01.01>].
54. Deuterostomia. Version 01 January 2002 [<http://tolweb.org/Deuterostomia/2466/2002.01.01>].
55. Chordata. Version 01 January 1995 [<http://tolweb.org/Chordata/2499/1995.01.01>].
56. Craniata. Animals with skulls. Version 01 January 1997 [<http://tolweb.org/Craniata/14826/1997.01.01>].
57. Vertebrata. Animals with backbones. Version 01 January 1997 (under construction). [<http://tolweb.org/Vertebrata/14829/1997.01.01>].
58. Gnathostomata. Jawed Vertebrates. Version 01 January 1997 [<http://tolweb.org/Gnathostomata/14843/1997.01.01>].
59. Sarcopterygii. The lobe-finned fishes & terrestrial vertebrates. Version 01 January 1995 [<http://tolweb.org/Sarcopterygii/14922/1995.01.01>].
60. Terrestrial Vertebrates. Stegocephalians: Tetrapods and other digit-bearing vertebrates. Version 21 April 2011. [[http://tolweb.org/Terrestrial\\_Vertebrates/14952/2011.04.21](http://tolweb.org/Terrestrial_Vertebrates/14952/2011.04.21)].
61. Amniota. Mammals, reptiles (turtles, lizards, Sphenodon, crocodiles, birds) and their extinct relatives. Version 30 January 2012. [<http://tolweb.org/Amniota/14990/2012.01.30>].
62. Mammals and their extinct relatives. Version 14 August 2011. [<http://tolweb.org/Synapsida/14845/2011.08.14>].

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](http://biomedcentral.com/submissions)

