A peer-reviewed version of this preprint was published in PeerJ on 15 June 2018.

View the peer-reviewed version (peerj.com/articles/5067), which is the preferred citable publication unless you specifically need to cite this preprint.

Hernandez AM, Ryan JF. 2018. Horizontally transferred genes in the ctenophore *Mnemiopsis leidyi*. PeerJ 6:e5067
https://doi.org/10.7717/peerj.5067

Horizontally transferred genes in the ctenophore *Mnemiopsis leidyi*

Alexandra M Hernandez [1, 2] , **Joseph F Ryan** Corresp. [1, 2]

[1] Whitney Laboratory for Marine Bioscience, St. Augustine, Florida, United States
[2] Department of Biology, University of Florida, Gainesville, FL, United States

Corresponding Author: Joseph F Ryan
Email address: joseph.ryan@whitney.ufl.edu

Horizontal gene transfer has had major impacts on the biology of a wide range of organisms from antibiotic resistance in bacteria to adaptations to herbivory in arthropods. A growing body of literature shows that horizontal gene transfer (HGT) between non-animals and animals is more commonplace than previously thought. In this study, we present a thorough investigation of HGT in the ctenophore *Mnemiopsis leidyi*. We applied tests of phylogenetic incongruence to identify nine genes that were likely transferred horizontally early in ctenophore evolution from bacteria and non-metazoan eukaryotes. All but one of these HGTs (an uncharacterized protein) are homologous to characterized enzymes, supporting previous observations that genes encoding enzymes are more likely to be retained after HGT events. We found that the majority of these nine horizontally transferred genes were expressed during development, suggesting that they are active and play a role in the biology of *M. leidyi*. This is the first report of HGT in ctenophores, and contributes to an ever-growing literature on the prevalence of genetic information flowing between non-animals and animals.

Horizontally transferred genes in the ctenophore *Mnemiopsis leidyi*

Alexandra M. Hernandez[1,2] and Joseph F. Ryan[1,2]

[1] Whitney Laboratory for Marine Bioscience, University of Florida, St. Augustine, FL, USA

[2] Department of Biology, University of Florida, Gainesville, FL, USA

Corresponding Author:

Joseph Ryan[1]

joseph.ryan@whitney.ufl.edu

Abstract

Horizontal gene transfer has had major impacts on the biology of a wide range of organisms from antibiotic resistance in bacteria to adaptations to herbivory in arthropods. A growing body of literature shows that horizontal gene transfer (HGT) between non-animals and animals is more commonplace than previously thought. In this study, we present a thorough investigation of HGT in the ctenophore *Mnemiopsis leidyi*. We applied tests of phylogenetic incongruence to identify nine genes that were likely transferred horizontally early in ctenophore evolution from bacteria and non-metazoan eukaryotes. All but one of these HGTs (an uncharacterized protein) are homologous to characterized enzymes, supporting previous observations that genes encoding enzymes are more likely to be retained after HGT events. We found that the majority of these nine horizontally transferred genes were expressed during development, suggesting that they are active and play a role in the biology of *M. leidyi*. This is the first report of HGT in ctenophores, and contributes to an ever-growing literature on the prevalence of genetic information flowing between non-animals and animals.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3462v3 | CC BY 4.0 Open Access | rec: 6 Jun 2018, publ: 6 Jun 2018

Introduction

Evolution is commonly thought to occur by descent with modification from a single lineage. However, evidence has shown that genomes from bacteria, archaea, and eukaryotes are typically chimeric, resulting from horizontal (or lateral) gene transfers (Garcia-Vallvé et al. 2000; Katz 2002). As such, horizontal gene transfer (HGT) has likely impacted evolution more than originally thought by creating opportunities for rapid genetic diversification and contributing to speciation events. Moreover, HGT is a potential catalyst for organisms to acquire novel traits (Soucy et al. 2015) and creates opportunities for HGT receivers to exploit new ecological niches (Boto 2010). For example, HGTs have played an important role in herbivory in arthropods (Wybouw et al. 2016), venom recruitment in parasitoid wasps (Martinson et al. 2016), cellulose production in urochordates (Dehal et al. 2002) and plant parasitism in nematodes (Haegeman et al. 2011).

Although HGT is generally accepted as an important evolutionary mechanism in prokaryotes (Boto 2014), it remains controversial whether it occurs in animals, despite many convincing studies (Madhusoodanan 2015). Much of the skepticism has been fueled by high-profile reports of HGT (e.g., Lander et al. 2001; Boothby et al. 2015) that were later shown to be largely incorrect due to contamination or taxon sampling (Stanhope et al. 2001; Koutsovoulos et al. 2016). In addition, HGT in animals is hypothesized to be rare due to the origin of a sequestered germ line, which provides fewer opportunities for germ cells to be exposed to foreign DNA (Doolittle 1999; Andersson et al. 2001; Jensen et al. 2016). However, the presence and absence of germline sequestration is not well described across the animal tree of life, and

there are inconsistencies between studies regarding which animal groups have sequestered germlines (Buss, 1983; Radzvilavicius et al. 2016; Jensen et al. 2016).

The major challenges for HGT detection efforts have been taxon sampling and contamination. Many early reports of HGT in animals were overturned due to limited representation of taxa in public genomic databases (e.g., Salzberg et al. 2001). For example, a gene present in bacteria and humans, but absent from nematodes and drosophilids (the most highly represented taxa at the time) may have been considered the result of HGT, until discovering that the gene is present in many other animal genomes that were not available at the time of the initial claim. In these cases, the limited representation of taxa made it difficult to distinguish HGTs from differential gene loss (Andersson et al. 2006; Keeling & Palmer 2008). More recently, contamination has led to both overestimation and likely underestimation of HGT events. In several recent cases, contamination in newly generated datasets has been interpreted as HGT but later shown to be cross-contaminants present in genome sequences (Bhattacharya et al. 2013; Delmont & Eren 2016; Koutsovoulos et al. 2016). On the other hand, the presence of contaminants in public databases (e.g., a bacteria sequence labeled as an animal sequence) makes it difficult to identify *bona fide* HGTs, as "animal" sequences will appear among the top BLAST hits for a particular HGT, leading to false negatives (Kryukov & Imanishi 2016). As such, contamination remains a major hurdle to contemporary studies of HGT.

Pairwise BLAST-based similarity scores (e.g., alien index (Gladyshev et al. 2008) and the HGT index (Boschetti et al. 2012)) are the most common criteria used to detect HGT in animals. However, these measures largely ignore phylogenetic information associated with

sequence data. While a positive BLAST-based result may be due to HGT, it may also result from gene loss, selective evolutionary rates, convergent evolution, sequence contamination, and species misassignment (Hall et al. 2005). Previous HGT studies have demonstrated that HGT predictions need to be carefully considered and a combination of methods are required to rule out false positives (Schönknecht et al. 2013). Hypothesis tests incorporating phylogenetic incongruence are one such method that has been used to test HGT. While some studies in animals have incorporated these techniques (e.g., Eliáš et al. 2016), they are more commonly deployed in studies involving non-animals (e.g., Bapteste et al. 2003; Richards et al. 2006).

HGT has yet to be thoroughly explored in Ctenophora. Ctenophores (comb jellies) are marine invertebrates that are morphologically characterized by eight rows of cilia used for movement. They typically live in the water column, but the group includes benthic species as well (Song & Hwang 2010; Alamaru et al. 2015; Glynn et al. 2017). Phylogenomic evidence from studies including ctenophores has suggested that ctenophores are the sister group to all other animals (Dunn et al. 2008; Hejnol et al. 2009; Ryan et al. 2013; Moroz et al. 2014; Borowiec et al. 2015; Chang et al. 2015; Torruella et al. 2015; Whelan et al. 2015; Arcila et al. 2017; Shen et al. 2017; Whelan et al. 2017), but the position remains controversial with some evidence supporting sponges as the sister group to the rest of animals (Philippe et al. 2009; Pick et al. 2010; Pisani et al. 2015; Telford et al. 2015; Simion et al. 2017; Feuda et al. 2017). Thus, investigating HGT in ctenophores is essential to understanding its implications on early animal evolution.

Here, we apply a rigorous framework to identify and confirm HGTs in the ctenophore *Mnemiopsis leidyi*. Our process includes identification of HGT candidates by alien index and confirmation by phylogenetic hypothesis testing, providing statistical support in an evolutionary framework. Furthermore, we analyze gene expression profiles during development to obtain clues as to the function of these HGTs in *M. leidyi*.

Material and Methods

All command lines, parameters, and version numbers of programs are in the supplementary text.

Identification of HGT candidates by alien_index

As part of this project, we developed the program alien_index and complimentary metazoan/non-metazoan sequence databases to automate the generation of alien index (Gladyshev et al. 2008) and HGT index scores (Boschetti et al. 2012). We BLASTed the entire set of *M. leidyi* gene models (ML2.2) (Ryan et al. 2013) against a database of animal and non-animal sequences (alien_index_db version 0.01) and then calculated alien index values as the logarithmic difference between the best BLASTP E-values for animal and non-animal hits (as outlined in Gladyshev et al. (2008)) (Fig. 1A). In more simple terminology, the alien index reflects the difference between the E-value of the best non-animal BLAST hit and that of the best animal hit. The database used includes translated gene models from curated genomes that include bacteria (5), archaea (2), non-animal eukaryotes (5), and animals (12). See Table S1 or http://ryanlab.whitney.ufl.edu/downloads/alien_index/ for the entire list of taxa. HGT index

values were computed by the difference in the highest non-animal and animal bit scores generated from the alien_index database. The alien_index program is available at: https://github.com/josephryan/alien_index

Confirmation of HGTs

We applied a phylogenetic approach to confirm putative HGTs. HGT candidates from alien_index were used as queries for BLASTP against NCBI's RefSeq database (O'Leary et al. 2016) using the NCBI BLAST interface. We collected the top ten sequences each from bacteria, eukaryotes, fungi, and animals with an E-value cutoff of 0.1. We included only the first sequence if there were hits to sequences from species in the same genus (Fig. 1B). We also added the top BLAST hit (E-value ≤ 0.1) from each of the following fully sequenced animals from version 0.01 of the alien_index database: *Amphimedon queenslandica*, *Trichoplax adhaerens*, *Nematostella vectensis*, *Capitella teleta*, *Drosophila melanogaster*, and *Homo sapiens*. Sequences were aligned against the corresponding putative HGT using MAFFT (Katoh et al. 2002; Katoh & Standley 2013) and trimmed with Gblockswrapper (Castresana 2000) (Fig. 1C). There were six genes (ML012034a, ML06718a, ML03277a, ML02232a, ML18354a, ML219316a) with BLASTP hits to non-animals but not to animals (E-value ≤ 0.1), preventing us from performing additional phylogenetic analyses on these sequences. We considered the lack of animal BLASTP hits below our cutoff as sufficient evidence that these six were clearly HGTs. ML018031a and ML00882a only had two BLASTP hits to animal sequences. Since it was unclear if this resulted from contamination, we were unable to test these genes using phylogenetic approaches, so they were removed from contention as HGTs.

We performed maximum-likelihood analyses on the remaining 29 alignments using RAxML (Stamatakis 2014) (Fig. 1C). Since the RefSeq database has many instances of contamination (Pible et al. 2014), we allowed a maximum of two non-ctenophore animal sequences to fall outside of the main animal clade. To implement this, we pruned putative contaminants if the removal of two taxa resulted in a monophyletic animal clade (Fig. S1). We discarded any HGT candidates with more than two taxa disrupting animal monophyly.

We explicitly tested topologies in opposition to HGT (i.e., animal monophyly) with the SOWH test using SOWHAT (Church et al. 2015) and the AU test using CONSEL (Shimodaira and Hasegawa 2001) (Fig. 1C). The SOWH and AU test evaluate statistical support for phylogenetic incongruence by comparing the likelihood values between trees to a distribution of trees generated by parametric sampling in the SOWH test and non-parametric sampling in the AU test. We required that these two different approaches to hypothesis testing agreed to ensure that our criteria confirming *bona fide* HGTs was stringent. To address any potential problems of selection bias in the AU test (causing the likelihood value to bias upwards for the maximum likelihood best tree when included in the dataset), we performed multiple AU analyses using bootstrap trees as suboptimal trees (similar to Eliáš et al. 2016). We generated 100 bootstrap trees using RAxML rapid bootstrap analyses, and verified there were no duplicate trees in our 100 bootstrap set using the ape package in R (Paradis et al. 2004). RAxML was used to generate per-site log likelihoods for the best maximum-likelihood tree, the tree constraining the putative HGT to metazoans (i.e., metazoan-constraint tree), and suboptimal trees, to perform the AU test implemented through CONSEL. To test the effectiveness of comparing to bootstrap trees, we

manually created a set of suboptimal trees for each HGT candidate by shuffling clades of three (Fig. S2) and running the same analyses. We evaluated the tree space covered by suboptimal trees in the AU test (i.e., bootstrap and manually generated trees) by visualizing the data using violin plots. We calculated likelihood proportions for each tree by dividing individual likelihood scores by the average likelihood score of suboptimal trees. This was done to make the data comparable for visualization since the likelihood scores differ between sets of gene trees. The trees and scripts used to automate these phylogenetic analyses are available in the accompanying GitHub site (https://github.com/josephryan/2018-Hernandez_and_Ryan_HGT).

We verified that HGT candidates were not the result of bacterial contaminants by using the *M. leidyi* genome browser (Moreland et al. 2014) to examine the intron/exon structure of each HGT candidate, as well as the origin of their neighboring genes. We examined each intron to determine whether they are actively handled by spliceosomes (which are only found in eukaryotes), since bacteria, archaea, and viruses contain Group I and II introns. U2 spliceosomal introns were identified by conserved GT dinucleotides at the 5' end and conserved AG dinucleotides at the 3' end of introns. We also conducted reciprocal best BLASTP searches for each HGT candidate (identified from the genome and gene models from an *M. leidyi* individual collected in Woods Hole, MA) against the transcriptome of an *M. leidyi* individual collected from St. Augustine, Florida, as well as seven other ctenophore transcriptomes reported in Moroz et al. (2014): *Bolinopsis infundibulum, Beroe abyssicola, Dryodora glandiformis, Pleurobrachia bachei, Vallicula multiformis, Coeloplana astericola, Euplokamis dunlapae*. For these searches we used default parameters and an E-value cutoff of 0.1.

HGT developmental expression profiles

An extensive transcriptomic developmental timecourse of *M. leidyi* was recently generated from single-embryos over the first 20 hours of embryogenesis (Levin et al. 2016). To examine whether HGTs might play a role in development we used these data (GSE70185), as well as additional time points for *M. leidyi* generated after this publication (GSE111748).

The Levin et al. (2016) data was produced by using three replicate timecourses that each consisted of 20 isolated embryos from fertilization to 20 hours. Embryos were flash frozen and RNA was extracted with TRIzol and sequenced using Illumina sequencing according to the CEL-Seq protocol (Hashimshony et al. 2012). For each replicate, reads were mapped to *M. leidyi* gene models (ML2.2) using bowtie2 version 2.2.3 (Langmead & Salzberg 2012) with default settings and reads per transcript were counted using htseq-count (Anders et al. 2015). Normalization of read counts was performed by dividing by the total number of counted reads and multiplying by 10^6. Since the CEL-Seq protocol involves sequencing only from the 3' end of transcripts, results are not normalized by length of transcript.

Since the publication of Levin et al. (2016), six additional time points (four replicates each) for hours 14-19 (not included in the original study) have been sequenced and submitted to the Gene Expression Omnibus (GSE111748). These additional data were produced by the same researchers (i.e., Itai Yanai and Mark Martindale) from the original study using the same methods and facilities. To create a baseline for what is considered adequate expression during development, we summed median transcripts per million (tpm) values for all replicates along the

25 time points for each of our 9 confirmed HGTs. HGTs that had summed median read counts of 100 or greater were classified as being expressed sufficiently to have roles in development. We chose a value that was 10 times stricter than the minimum criteria in Levin et al. (2016) (i.e., 10 transcripts) to err on the side of caution.

HGT origins and functions

To uncover the functional roles of HGTs, we used the BLAST interface provided by UniProt and the UniProtKB database (Pundir et al. 2017) to identify homologous sequences used to characterize genes. Annotations of the top hits (E-value ≤0.1) were assigned to HGT candidates. We also associated HGTs with Pfam-A domains using the MGP Portal under the *Mnemiopsis* Gene Wiki (Moreland et al. 2014). In all cases, the annotations based on BLAST and Pfam-A analyses were consistent with the results from our phylogenetic analyses. To identify the origin of the HGTs lacking animal hits (ML012034a, ML18354a, ML219316a), we performed phylogenetic analyses on the sequences collected at the start of the study from RefSeq using RAxML.

Results

Mnemiopsis leidyi HGTs

Figure 1 shows our pipeline and results for each method during this analysis. We calculated an alien index for every *M. leidyi* gene model using a database of 12 animals and 12

non-animals (Table S1). We identified 37 genes with alien indices greater than 45 and designated these as HGT candidates (Fig. 1A; Table S2). In addition to the alien_index database, we BLASTed the RefSeq database at NCBI restricting hits to bacteria, then to animals, and then to non-animal eukaryotes (Fig. 1B). All but six HGT candidates had BLAST hits to animals with E-values ≤ 0.1. We classified these six (ML012034a, ML06718a, ML03277a, ML02232a, ML18354a, ML219316a) as absent from all other animals. We analyze these six further below.

For the remaining 29 candidates, we conducted detailed phylogenetic analyses using the top 10 hits of unique non-animal and animal taxa from each of the RefSeq searches along with sequences from *Amphimedon queenslandica, Trichoplax adhaerens*, *Nematostella vectensis*, *Capitella teleta*, *Drosophila melanogaster*, and *Homo sapiens* that were top hits from our initial BLASTs of the alien_index database (Fig. 1C). HGT candidates that formed a clade with all other animals were ruled out as potential HGTs, while candidates that disrupted animal monophyly were tested further. We discarded 14 candidates with more than 2 non-ctenophore animal sequences disrupting animal monophyly; in cases of 2 or less sequences, the disrupting sequences were considered potential contaminants and pruned (e.g., Fig. S1). We discarded three more candidates after pruning because the trees continued to result in a non-monophyletic clade of animals. We then applied the SOWH and AU tests to the remaining 12 candidates to compare the maximum-likelihood topology to the alternative hypothesis that HGT candidates were more closely related to animals (Fig. 2). This involved comparing likelihood values of optimal trees to those that were constrained to produce a monophyletic Animalia. Our results showed that the AU test was more conservative in confirming HGTs than the SOWH test (Table 1). For perspective on how optimal trees compared to constrained trees, we performed AU tests comparing optimal

and constrained trees to bootstrap trees (Fig. 3). The likelihood scores of the constrained trees from HGTs supported by the AU test tend to fall outside or on the tails of the distribution of likelihood scores of suboptimal trees, whereas the likelihood scores of constrained trees for unsupported HGTs were all closer to the most likely tree than the bootstrap trees (Fig. 3). We confirmed seven HGTs in which gene trees significantly differed ($p < 0.05$) from the metazoan constraint trees in both the SOWH and AU analyses (Table 1).

We then analyzed the BLAST results of the seven HGT candidates confirmed by phylogenetic analyses and the six genes which were absent in other animals (ML012034a, ML06718a, ML03277a, ML02232a, ML18354a, ML219316a). We removed four of these genes from contention (ML092610a, ML06718a, ML03277a, ML02232a) because the top BLAST hits against RefSeq were either Choanoflagellatea or Filasterea (two of the closest protistan lineages to animals (Hehenberger et al. 2017; Torruella et al. 2015)) (E-value ≤ 0.1). If ctenophores are the sister group to the rest of animals, vertical inheritance remains a possibility for these cases. As such, these tests support a total of nine HGTs.

We used the *M. leidyi* genome browser to examine the intron/exon structure of each of these nine HGTs, as well as the origin of their neighboring genes for evidence of bacterial contamination (lack of introns would indicate bacterial contamination). Eight HGTs were found on scaffolds with intron-containing genes and eight HGTs contained introns (Table 2). ML49231a (itself containing 6 introns) is the only gene on its scaffold. All intron-containing genes had U2 spliceosomal introns (except ML219315a, a gene on the same scaffold as an

277 HGT). These data suggest that the majority of the HGT candidates did not appear to be bacterial
278 contaminants.
279
280 To further test these nine genes for contamination, we confirmed using BLAST that the
281 genes were also present in a transcriptome from an *M. leidyi* individual collected from St.
282 Augustine, FL (*M. leidyi* genome and gene models were from individuals collected in Woods
283 Hole, MA). We also performed reciprocal best BLASTP searches for each of the nine HGTs
284 against seven of the ctenophore transcriptomes published in Moroz et al. (2014): *Bolinopsis*
285 *infundibulum, Beroe abyssicola, Dryodora glandiformis, Pleurobrachia bachei, Vallicula*
286 *multiformis, Coeloplana astericola, Euplokamis dunlapae*. Each HGT was present in the
287 transcriptome of at least one other ctenophore species and in the Florida *M. leidyi* transcriptome
288 (Fig. 4). Furthermore, the gene lacking introns (ML012034a) is expressed in all examined
289 ctenophore transcriptomes. Because it is unlikely that the same species contaminated each of
290 these datasets, these comparisons provide additional evidence against HGT sequences resulting
291 from contamination. Here, we have included as much evidence as possible to carefully confirm
292 nine HGTs in *M. leidyi*.
293
294 **HGTs are expressed in development**
295
296 We summed tpm values (medians for each set of expression values at 25 time points)
297 from single-embryo RNA-Seq analyses over 20 hours for each of the nine confirmed HGTs to
298 identify those HGTs that were expressed during development. Six of the nine HGTs had sums
299 greater than 100 (Fig. 5), suggesting that these had some role in development. ML00955a was

expressed maternally (0 hours post fertilization (hpf)) and throughout early cleavage stages (1-3 hpf) with reduced expression later in development. Three genes (ML005129a, ML18354a, ML012034a) were expressed later in development with spikes during tentacle morphogenesis (9-12 hpf). ML02771a and ML219316a displayed cyclic expression throughout development suggesting a potential role in cell cycle.

HGTs are enzymes originating from non-animal eukaryotes and bacteria

We used phylogenetic evidence to determine the origin of these nine HGTs. Four HGTs originated from bacteria and five from non-animal eukaryotes (Table 3). We found no evidence of HGTs that were transferred from Archaea. Specific lineage origins of three HGTs appear to be from Proteobacteria, Firmicutes, and Rhodophyta. We were unable to identify the lineage origins of the remaining HGTs. To characterize gene function, we BLASTed the nine confirmed HGTs against the UniProt database. All HGTs except one uncharacterized protein (ML219316a) were homologous to known characterized enzymes (Table 3).

Discussion

HGTs in ctenophores and their implications

It had been previously speculated that ctenophores had HGTs since initial profiling revealed that many ‘bacteria-like’ genes in ctenophores contained introns and were on chromosomes with vertically inherited (i.e., non-HGT) genes (Artamonova et al. 2015). We

confirmed that all HGTs except ML012034a had spliceosomal introns and were on scaffolds with other spliceosomal intron-containing genes (Table 2). This provided evidence that these candidates were not the result of extrinsic contamination. We provided additional evidence that candidates were not contaminants by showing that all HGTs were found in both Massachusetts and Florida *M. leidyi* individuals, as well as in many other ctenophore species (Fig. 4). Six HGTs were present in the *E. dunlapae* transcriptome suggesting that the majority of these HGT events occurred very early in ctenophore evolution (Fig. 4). This deep evolutionary history suggests that these HGTs may have had important impacts on the biology of ctenophores.

Mechanisms driving HGT in ctenophores

While we are uncertain about the mechanisms driving HGT, we speculate that some of these may have resulted from symbiotic relationships with bacteria and non-animal eukaryotes. *Proteobacteria* is the most abundant group of bacteria associated with ctenophores (Daniels & Breitbart 2012) and have been identified as donors of the gene ML00955a in the *M. leidyi* genome (Table 3) and confirmed in almost all other ctenophore transcriptomes (Fig. 4). Other possible donors could have been gymnamoebae symbionts that have been described living on the surface of comb plates and on the ectoderm of ctenophores (Moss et al. 2001). However, studies investigating symbiotic relationships with ctenophores are limited. Future studies are needed to improve our understanding of how symbiotic relationships impact HGT, as well as to understand the mechanisms that drive HGT between organisms.

***Mnemiopsis leidyi* HGTs are expressed during development and encode enzymes**

346

347 Many HGTs are likely to be deleterious and lost, but some HGTs will be neutral or
348 provide a selective advantage and spread throughout a population (Thomas & Nielsen 2005).
349 HGT integration is thought to mainly occur in neutral genes with low levels of expression (Park
350 & Zhang 2012). Once integrated, neutral HGTs may become a source of novel genetic variation
351 upon which selection can act (Soucy et al. 2015). HGTs may then become more highly expressed
352 after recruitment of transcription factors and regulators from the host genome (Lercher & Pál
353 2008). Six of the nine HGTs we identified showed high expression during the first 20 hours of
354 development, suggesting potentially important developmental roles (Fig. 5). ML02771a is
355 expressed during development and encodes penicillin acylase or amidase, which catalyzes the
356 hydrolysis of benzylpenicillin. This reaction creates key intermediates for penicillin synthesis
357 and may be important to defend against microbial infection or colonization.

358

359 Observations of HGT patterns in prokaryotes have also suggested that there is a
360 preference to retain operational genes (e.g., metabolic enzymes) rather than informational genes
361 (Lawrence & Roth 1996; Jain et al. 1999; Garcia-Vallvé et al. 2000). Informational genes, such
362 as those involved in DNA replication, transcription, and translation are seldom found in sets of
363 HGTs (Thomas & Nielsen 2005). This propensity for operational genes is thought to occur
364 because informational genes are involved in larger and complex systems (Jain et al. 1999).
365 Recently, this pattern has also been observed in animal HGTs (Boto 2014) (e.g., Zhu et al. 2011;
366 Boschetti et al. 2012; Sun et al. 2013; Eyres et al. 2015; Conaco et al. 2016). These reports
367 suggest that operational genes are preferentially transferred and/or retained in both prokaryotes

368 and eukaryotes. Our data support this idea since all of the characterizable genes in our HGT set
369 encode enzymes.
370
371 **Commonly used BLAST-based methods for identifying HGTs in animals are insufficient**
372
373 Identifying HGTs can be challenging due to bacterial associations with hosts
374 (Artamonova & Mushegian 2013; Chapman et al. 2010; Fraune & Bosch 2007), DNA extraction
375 kits and reagents that have led to contamination (Naccache et al. 2013; Salter et. al 2014), and/or
376 laboratory conditions that can contaminate preparations during DNA extraction (Laurence et al.
377 2014; Strong et al. 2014). These challenges associated with sequencing and assembly have
378 resulted in contamination in public databases (Longo et al. 2011; Merchant et al. 2014) and make
379 HGT predictions difficult. Moreover, while BLAST-based approaches (i.e., alien index and the
380 HGT index) are useful for identification of HGT candidates, they are difficult to implement, lack
381 an evolutionary perspective, and do not address problems associated with contamination.
382
383 To overcome some of these challenges, we developed alien_index to automate the
384 generation of alien index and HGT index scores for rapid identification of HGT candidates. We
385 confirmed HGTs by using rigorous phylogenetic approaches to address the problems associated
386 with the lack of evolutionary perspective from BLAST methods. Our phylogenetic tests of
387 incongruence provided clear metrics from which to judge the level of certainty applied to each
388 HGT candidate. Our study showed that many of the predictions based on BLAST did not stand
389 up to hypothesis testing and suggest that the similarity between sequences that cause high alien
390 indices do not necessarily provide true phylogenetic signal. Consequently, incorporation of

phylogenetic likelihood-based methods are necessary when performing HGT analyses in animals.

Conclusion

The importance of HGT as an evolutionary mechanism in prokaryotes and eukaryotes has been underestimated. While studies of HGT in animals are gradually becoming more accepted, many challenges remain to quantify the extent of HGT and its impacts. To mitigate some of these challenges, rigorous approaches that employ both BLAST- and phylogenetic likelihood-based methods should be applied to future HGT studies in animals. Here we provided evidence of nine cases of HGT in ctenophores by applying these rigorous methods (among others), and found similar patterns of transfer between prokaryotes and eukaryotes with preference for operational genes. It should be noted that we implemented an extremely conservative approach and there are likely to be more HGTs in *M. leidyi.* However, many more studies will be necessary to gain a comprehensive overview of HGT and the mechanisms by which HGT occurs in animals.

Acknowledgements

We would like to thank Melissa DeBiasse for constructive comments on earlier versions of this manuscript.

References

Alamaru A., Brokovich E., Loya Y. 2015. Four new species and three new records of benthic ctenophores (Family: Coeloplanidae) from the Red Sea. *Marine Biodiversity* 46:261-279. DOI: 10.1007/s12526-015-0362-4

Anavy L., Levin M., Khair S., Nakanishi N., Fernandez-Valverde SL., Degnan BM., Yanai Y. 2014. BLIND ordering of large-scale transcriptomic developmental timecourses. *Development* 141:1161–1166. DOI: 10.1242/DEV.105288

Anders S., Pyl PT., Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169. DOI: 10.1093/bioinformatics/btu638

Andersson JO., Doolittle WF., Nesbø CL. 2001. Are There Bugs in Our Genome? *Science* 292:1848-1850.

Andersson JO., Hirt RP., Foster PG., Roger AJ. 2006. Evolution of four gene families with patchy phylogenetic distributions: influx of genes into protist genomes. *BMC Evolutionary Biology* 6:27. DOI: 10.1186/1471-2148-6-27

Arcila D., Ortí G., Vari R., Armbruster JW., Stiassny MLJ., Ko KD., Sabaj MH., Lundberg J., Revell LJ., Betancur-R. R. 2017. Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. *Nature Ecology & Evolution* 1:20. DOI: 10.1038/s41559-016-0020

Artamonova II., Mushegian AR. 2013. Genome sequence analysis indicates that the model eukaryote Nematostella vectensis harbors bacterial consorts. *Applied and Environmental Microbiology* 79:6868–6873. DOI: 10.1128/AEM.01635-13

Artamonova II., Lappi T., Zudina L., Mushegian AR. 2015. Prokaryotic genes in eukaryotic genome sequences: when to infer horizontal gene transfer and when to suspect an actual microbe. *Environmental Microbiology* 17:2203–2208. DOI: 10.1111/1462-2920.12854

Bapteste E., Moreira D., Philippe H. 2003. Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase. *Gene* 318:185-191.

Bhattacharya D., Pelletreau KN., Price DC., Sarver KE. 2013. Genome Analysis of Elysia chlorotica Egg DNA Provides No Evidence for Horizontal Gene Transfer into the Germ Line of This Kleptoplastic Mollusc. *Molecular Biology and Evolution* 30:1843–1852. DOI: 10.1093/molbev/mst084

Boothby TC., Tenlen JR., Smith FW., Wang JR., Patanella KA., Nishimura EO., Tintori SC., Li Q., Jones CD., Yandell M., Messina DN., Glasscock J., Goldstein B. 2015. Evidence for extensive horizontal gene transfer from the draft genome of a tardigrade. *Proceedings of the*

National Academy of Sciences of the United States of America 112:15976–15981. DOI: 10.1073/pnas.1510461112

Borowiec ML., Lee EK., Chiu JC., Plachetzki DC. 2015. Extracting phylogenetic signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* 16:987. DOI: 10.1186/s12864-015-2146-4

Boschetti C., Carr A., Crisp A., Eyres I., Wang-Koh Y., Lubzens E., Barraclough TG., Micklem G., Tunnacliffe A. 2012. Biochemical Diversification through Foreign Gene Expression in Bdelloid Rotifers. *PLoS Genetics* 8:e1003035. DOI: 10.1371/journal.pgen.1003035

Boto L. 2010. Horizontal gene transfer in evolution: facts and challenges. *Proceedings. Biological Sciences* 277:819–827. DOI: 10.1098/rspb.2009.1679

Boto L. 2014. Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proceedings of the Royal Society of London B: Biological Sciences* 281:20132450. DOI: 10.1098/rspb.2013.2450

Buss LW. 1983. Evolution, development, and the units of selection. *Proceedings of the National Academy of Sciences* 80:1387-1391.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17:540-552. DOI: 10.1093/oxfordjournals.molbev.a026334

Chang ES., Neuhof M., Rubinstein ND., Diamant A., Philippe H., Huchon D., Cartwright P. 2015. Genomic insights into the evolutionary origin of Myxozoa within Cnidaria. *Proceedings of the National Academy of Sciences of the United States of America* 112:14912–14917. DOI:10.1073/pnas.1511468112

Chapman JA., Kirkness EF., Simakov O., Hampson SE., Mitros T., Weinmaier T., Rattei T., Balasubramanian PG., Borman J., Busam D., Disbennett K., Pfannkoch C., Sumin N., Sutton GG., Viswanathan LD., Walenz B., Goodstein DM., Hellsten U., Kawashima T., Prochnik SE., Putnam NH., Shu S., Blumberg B., Dana CE., Gee L., Kibler DF., Law L., Lindgens D., Martinez DE., Peng J., Wigge PA., Bertulat B., Guder C., Nakamura Y., Ozbek S., Watanabe H., Khalturin K., Hemmrich G., Franke A., Augustin R., Fraune S., Hayakawa E., Hayakawa S., Hirose M., Hwang JS., Ikeo K., Nishimiya-Fujisawa C., Ogura A., Takahashi T., Steinmetz PRH., Zhang X., Aufschnaiter R., Eder M-K., Gorny A-K., Salvenmoser W., Heimberg AM., Wheeler BM., Peterson KJ., Böttger A., Tischler P., Wolf A., Gojobori T., Remington KA., Strausberg RL., Venter JC., Technau U., Hobmayer B., Bosch TCG., Holstein TW., Fujisawa T., Bode HR., David CN., Rokhsar DS., Steele RE. 2010. The dynamic genome of Hydra. *Nature* 464:592–596. DOI: 10.1038/nature08830.

478 Church SH., Ryan JF., Dunn CW. 2015. Automation and Evaluation of the SOWH Test with
479 SOWHAT. *Systematic Biology* 64:1048–1058. DOI: 10.1093/sysbio/syv055

480 Conaco C., Tsoulfas P., Sakarya O., Dolan A., Werren J., Kosik K. S. 2016. Detection of
481 Prokaryotic Genes in the Amphimedon queenslandica Genome. *PLOS ONE* 11:e0151092.
482 DOI: 10.1371/journal.pone.0151092

483 Daniels C., Breitbart M. 2012. Bacterial communities associated with the ctenophores
484 Mnemiopsis leidyi and Beroe ovata. *FEMS Microbiology Ecology* 82:90–101. DOI:
485 10.1111/j.1574-6941.2012.01409.x

486 Dehal P., Satou Y., Campbell RK., Chapman J., Degnan B., De Tomaso A., Davidson B., Di
487 Gregorio A., Gelpke M., Goodstein DM., Harafuji N., Hastings KEM., Ho I., Hotta K.,
488 Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen IA., Necula S., Nonaka
489 M., Putnam N., Rash S., Saiga H., Satake M., Terry A., Yamada L., Wang H-G., Awazu S.,
490 Azumi K., Boore J., Branno M., Chin-bow S., DeSantis R., Doyle S., Francino P., Keys
491 DN., Haga S., Hayashi H., Hino K., Imai KS., Inaba K., Kano S., Kobayashi K., Kobayashi
492 K., Lee B-I., Makabe KW., Manohar C., Matassi G., Medina M., Mochizuki Y., Mount S.,
493 Morishita T., Miura S., Nakayama A., Nishizaka S., Nomoto H., Ohta F., Oishi K.,
494 Rigoutsos I., Sano M., Sasaki A., Sasakura Y., Shoguchi E., Shin-i T., Spagnuolo A.,
495 Stainier D., Suzuki MM., Tassy O., Takatori N., Tokuoka M., Yagi K., Yoshizaki F., Wada
496 S., Zhang C., Hyatt PD., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T.,
497 Richardson P., Lucas S., Kohara Y., Levine M., Satoh N., Rokhsar DS. 2002. The Draft
498 Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. *Science*
499 298:2157-2167. DOI: 10.1126/science.1080049

500 Delmont TO., Eren AM. 2016. Identifying contamination with advanced visualization and
501 analysis practices: metagenomic approaches for eukaryotic genome assemblies. *PeerJ*
502 4:e1839. DOI: 10.7717/peerj.1839.

503 Doolittle WF. 1999. Phylogenetic Classification and the Universal Tree. *Science* 284:2124-2128.

504 Dunn CW., Hejnol A., Matus DQ., Pang K., Browne WE., Smith SA., Seaver E., Rouse GW.,
505 Obst M., Edgecombe GD., Sørensen M V., Haddock SHD., Schmidt-Rhaesa A., Okusu A.,
506 Kristensen RM., Wheeler WC., Martindale MQ., Giribet G. 2008. Broad phylogenomic
507 sampling improves resolution of the animal tree of life. *Nature* 452:745–749. DOI:
508 10.1038/nature06614.

509 Eliáš M., Klimeš V., Derelle R., Petrželková R., Tachezy J. 2016. A paneukaryotic genomic
510 analysis of the small GTPase RABL2 underscores the significance of recurrent gene loss in
511 eukaryote evolution. *Biology direct* 11:5. DOI: 10.1186/s13062-016-0107-8

512 Eyres I., Boschetti C., Crisp A., Smith TP., Fontaneto D., Tunnacliffe A., Barraclough TG. 2015.
513 Horizontal gene transfer in bdelloid rotifers is ancient, ongoing and more frequent in species
514 from desiccating habitats. *BMC Biology* 13. DOI: 10.1186/s12915-015-0202-9.

515 Fraune S., Bosch TCG. 2007. Long-term maintenance of species-specific bacterial microbiota in
516 the basal metazoan Hydra. *Proceedings of the National Academy of Sciences of the United*
517 *States of America* 104:3146–13151. DOI: 10.1073/PNAS.0703375104

518 Feuda R., Dohrmann M., Pett W., Philippe H., Rota-Stabelli O., Lartillot N., Wörheide G., Pisani
519 D. 2017. Improved Modeling of Compositional Heterogeneity Supports Sponges as Sister to
520 All Other Animals. *Current Biology* 27:3864-3870. DOI: 10.1016/j.cub.2017.11.008.

521 Garcia-Vallvé S., Romeu A., Palau J. 2000. Horizontal gene transfer in bacterial and archaeal
522 complete genomes. *Genome Research* 10:1719–1725. DOI: 10.1101/GR.130000

523 Gladyshev EA., Meselson M., Arkhipova IR. 2008. Massive horizontal gene transfer in bdelloid
524 rotifers. *Science* 320:1210–1213. DOI: 10.1126/science.1156407

525 Glynn PW., Coffman B., Fuller MPC., Moorhead SG., Williams MK., Primov KD., Fortson TN.,
526 Barrales RN., Glynn PJ. 2017. Benthic ctenophores (Platyctenida: Coeloplanidae) in south
527 Florida: environmental conditions, habitats, abundances, and behaviors. *Invertebrate*
528 *Biology* 136:379-393. DOI: 10.1111/ivb.12189.

529 Haegeman A., Jones JT., Danchin EG. 2011. Horizontal gene transfer in nematodes: a catalyst
530 for plant parasitism?. *Molecular Plant-Microbe Interactions* 24:879-887.

531 Hall C., Brachat S., Dietrich, FS. 2005. Contribution of horizontal gene transfer to the evolution
532 of Saccharomyces cerevisiae. *Eukaryotic Cell* 4:1102–1115. DOI: 10.1128/EC.4.6.1102-
533 1115.2005

534 Hashimshony T., Wagner F., Sher N., Yanai I. 2012. CEL-Seq: Single-Cell RNA-Seq by
535 Multiplexed Linear Amplification. *Cell Reports* 2:666–673. DOI:
536 10.1016/J.CELREP.2012.08.003.

537 Hehenberger E., Tikhonenkov D V., Kolisko M., del Campo J., Esaulov AS., Mylnikov AP.,
538 Keeling PJ. 2017. Novel Predators Reshape Holozoan Phylogeny and Reveal the Presence
539 of a Two-Component Signaling System in the Ancestor of Animals. *Current Biology*
540 27:2043–2050.e6. DOI: 10.1016/J.CUB.2017.06.006.

541 Hejnol A., Obst M., Stamatakis A., Ott M., Rouse GW., Edgecombe GD., Martinez P., Baguñà
542 J., Bailly X., Jondelius U., Wiens M., Müller WEG., Seaver E., Wheeler WC., Martindale
543 MQ., Giribet G., Dunn CW. 2009. Assessing the root of bilaterian animals with scalable
544 phylogenomic methods. *Proceedings of the Royal Society of London B: Biological Sciences*
545 276:4261–4270. DOI: 10.1098/rspb.2009.0896.

546 Jain R., Rivera MC., Lake JA. 1999. Horizontal Gene Transfer among Genomes: The
547 Complexity Hypothesis. *Proceedings of the National Academy of Sciences of the United*
548 *States of America Evolution* 96:3801–3806. DOI: 10.1073/pnas.96.7.3801

549 Jensen L., Grant JR., Laughinghouse HD., Katz LA. 2016. Assessing the effects of a sequestered
550 germline on interdomain lateral gene transfer in Metazoa. *Evolution* 70:1322–1333. DOI:
551 10.1111/evo.12935.

552 Katoh K., Misawa K., Kuma K., Miyata T. 2002. MAFFT: a novel method for rapid multiple
553 sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066.
554 DOI: 10.1093/nar/gkf436.

555 Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7:
556 Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780.
557 DOI: 10.1093/molbev/mst010.

558 Katz LA. 2002. Lateral gene transfers and the evolution of eukaryotes: theories and data.
559 *International Journal of Systematic and Evolutionary Microbiology* 52:1893–1900. DOI:
560 10.1099/00207713-52-5-1893

561 Keeling PJ., Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. *Nature Reviews*
562 *Genetics* 9:605–618. DOI: 10.1038/nrg2386

563 Koutsovoulos G., Kumar S., Laetsch DR., Stevens L., Daub J., Conlon C., Maroon H., Thomas
564 F., Aboobaker AA., Blaxter M. 2016. No evidence for extensive horizontal gene transfer in
565 the genome of the tardigrade *Hypsibius dujardini*. *Proceedings of the National Academy of*
566 *Sciences* 113:5053-5058. DOI: 10.1073/pnas.1600338113.

567 Kryukov K., Imanishi T. 2016. Human Contamination in Public Genome Assemblies. *PLOS*
568 *ONE* 11:e0162424. DOI: 10.1371/journal.pone.0162424.

569 Lander ES., Linton LM., Birren B., Nusbaum C., Zody MC., Baldwin J., Devon K., Dewar K.,
570 Doyle M., FitzHugh W., Funke R., Gage D., Harris K., Heaford A., Howland J., Kann L.,
571 Lehoczky J., LeVine R., McEwan P., McKernan K., Meldrim J., Mesirov JP., Miranda C.,
572 Morris W., Naylor J., Raymond C., Rosetti M., Santos R., Sheridan A., Sougnez C., Stange-
573 Thomann N., Stojanovic N., Subramanian A., Wyman D., Rogers J., Sulston J., Ainscough
574 R., Beck S., Bentley D., Burton J., Clee C., Carter N., Coulson A., Deadman R., Deloukas
575 P., Dunham A., Dunham I., Durbin R., French L., Grafham D., Gregory S., Hubbard T.,
576 Humphray S., Hunt A., Jones M., Lloyd C., McMurray A., Matthews L., Mercer S., Milne
577 S., Mullikin JC., Mungall A., Plumb R., Ross M., Shownkeen R., Sims S., Waterston RH.,
578 Wilson RK., Hillier LW., McPherson JD., Marra MA., Mardis ER., Fulton LA., Chinwalla
579 AT., Pepin KH., Gish WR., Chissoe SL., Wendl MC., Delehaunty KD., Miner TL.,
580 Delehaunty A., Kramer JB., Cook LL., Fulton RS., Johnson DL., Minx PJ., Clifton SW.,
581 Hawkins T., Branscomb E., Predki P., Richardson P., Wenning S., Slezak T., Doggett N.,
582 Cheng J-F., Olsen A., Lucas S., Elkin C., Uberbacher E., Frazier M., Gibbs RA., Muzny
583 DM., Scherer SE., Bouck JB., Sodergren EJ., Worley KC., Rives CM., Gorrell JH., Metzker
584 ML., Naylor SL., Kucherlapati RS., Nelson DL., Weinstock GM., Sakaki Y., Fujiyama A.,
585 Hattori M., Yada T., Toyoda A., Itoh T., Kawagoe C., Watanabe H., Totoki Y., Taylor T.,
586 Weissenbach J., Heilig R., Saurin W., Artiguenave F., Brottier P., Bruls T., Pelletier E.,
587 Robert C., Wincker P., Rosenthal A., Platzer M., Nyakatura G., Taudien S., Rump A.,

588 Smith DR., Doucette-Stamm L., Rubenfield M., Weinstock K., Lee HM., Dubois J., Yang
589 H., Yu J., Wang J., Huang G., Gu J., Hood L., Rowen L., Madan A., Qin S., Davis RW.,
590 Federspiel NA., Abola AP., Proctor MJ., Roe BA., Chen F., Pan H., Ramser J., Lehrach H.,
591 Reinhardt R., McCombie WR., de la Bastide M., Dedhia N., Blöcker H., Hornischer K.,
592 Nordsiek G., Agarwala R., Aravind L., Bailey JA., Bateman A., Batzoglou S., Birney E.,
593 Bork P., Brown DG., Burge CB., Cerutti L., Chen H-C., Church D., Clamp M., Copley RR.,
594 Doerks T., Eddy SR., Eichler EE., Furey TS., Galagan J., Gilbert JGR., Harmon C.,
595 Hayashizaki Y., Haussler D., Hermjakob H., Hokamp K., Jang W., Johnson LS., Jones TA.,
596 Kasif S., Kaspryzk A., Kennedy S., Kent WJ., Kitts P., Koonin E V., Korf I., Kulp D.,
597 Lancet D., Lowe TM., McLysaght A., Mikkelsen T., Moran J V., Mulder N., Pollara VJ.,
598 Ponting CP., Schuler G., Schultz J., Slater G., Smit AFA., Stupka E., Szustakowki J.,
599 Thierry-Mieg D., Thierry-Mieg J., Wagner L., Wallis J., Wheeler R., Williams A., Wolf
600 YI., Wolfe KH., Yang S-P., Yeh R-F., Collins F., Guyer MS., Peterson J., Felsenfeld A.,
601 Wetterstrand KA., Myers RM., Schmutz J., Dickson M., Grimwood J., Cox DR., Olson M
602 V., Kaul R., Raymond C., Shimizu N., Kawasaki K., Minoshima S., Evans GA., Athanasiou
603 M., Schultz R., Patrinos A., Morgan MJ. 2001. Initial sequencing and analysis of the human
604 genome. *Nature* 409:860–921. DOI: 10.1038/35057062.

605 Langmead B., Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*
606 9:357–359. DOI: 10.1038/nmeth.1923

607 Laurence M., Hatzis C., Brash DE. 2014. Common Contaminants in Next-Generation
608 Sequencing That Hinder Discovery of Low-Abundance Microbes. *PLoS ONE* 9:e97876.
609 DOI: 10.1371/journal.pone.0097876

610 Lawrence JG., Roth JR. 1996. Selfish Operons: Horizontal Transfer May Drive the Evolution of
611 Gene Clusters. *Genetics* 143:1843-1860.

612 Lercher MJ., Pal C. 2008. Integration of Horizontally Transferred Genes into Regulatory
613 Interaction Networks Takes Many Million Years. *Molecular Biology and Evolution* 25:559–
614 567. DOI: 10.1093/molbev/msm283

615 Levin M., Anavy L., Cole AG., Winter E., Mostov N., Khair S., Senderovich N., Kovalev E.,
616 Silver DH., Feder M., Fernandez-Valverde SL., Nakanishi N., Simmons D., Simakov O.,
617 Larsson T., Liu S-Y., Jerafi-Vider A., Yaniv K., Ryan JF., Martindale MQ., Rink JC.,
618 Arendt D., Degnan SM., Degnan BM., Hashimshony T., Yanai I. 2016. The mid-
619 developmental transition and the evolution of animal body plans. *Nature* 531:637–641.
620 DOI: 10.1038/nature16994.

621 Longo MS., O'Neill MJ., O'Neill RJ., Clamp M., Barton G. 2011. Abundant Human DNA
622 Contamination Identified in Non-Primate Genome Databases. *PLoS ONE* 6:e16410. DOI:
623 10.1371/journal.pone.0016410.

624 Madhusoodanan J. 2015. Horizontal Gene Transfer a Hallmark of Animal Genomes? Available
625 at http://www.the-scientist.com/?articles.view/articleNo/42420/title/Horizontal-Gene-
626 Transfer-a-Hallmark-of-Animal-Genomes-/

627 Martinson EO., Martinson VG., Edwards R., Mrinalini., Werren JH. 2016. Laterally Transferred
628 Gene Recruited as a Venom in Parasitoid Wasps. *Molecular Biology and Evolution*
629 33:1042–1052. DOI: 10.1093/molbev/msv348.

630 Merchant S., Wood DE., Salzberg SL. 2014. Unexpected cross-species contamination in genome
631 sequencing projects. *PeerJ*, 2:e675. DOI: 10.7717/peerj.675

632 Moreland RT., Nguyen A-D., Ryan JF., Schnitzler CE., Koch BJ., Siewert K., Wolfsberg TG.,
633 Baxevanis AD. 2014. A customized Web portal for the genome of the ctenophore
634 Mnemiopsis leidyi. *BMC genomics* 15:316. DOI: 10.1186/1471-2164-15-316.

635 Moroz LL., Kocot KM., Citarella MR., Dosung S., Norekian TP., Povolotskaya IS., Grigorenko
636 AP., Dailey C., Berezikov E., Buckley KM., Ptitsyn A., Reshetov D., Mukherjee K., Moroz
637 TP., Bobkova Y., Yu F., Kapitonov V V., Jurka J., Bobkov Y V., Swore JJ., Girardo DO.,
638 Fodor A., Gusev F., Sanford R., Bruders R., Kittler E., Mills CE., Rast JP., Derelle R.,
639 Solovyev V V., Kondrashov FA., Swalla BJ., Sweedler J V., Rogaev EI., Halanych KM.,
640 Kohn AB. 2014. The ctenophore genome and the evolutionary origins of neural systems.
641 *Nature* 510:109–114. DOI: 10.1038/nature13400.

642 Moss AG., Estes AM., Muellner LA., Morgan DD. 2001. Protistan epibionts of the ctenophore
643 *Mnemiopsis mccradyi* Mayer. *Hydrobiologia* 451:295–304.

644 Naccache SN., Greninger AL., Lee D., Coffey LL., Phan T., Rein-Weston A., Aronsohn A.,
645 Hackett J., Delwart EL., Chiu CY. 2013. The perils of pathogen discovery: origin of a novel
646 parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *Journal of*
647 *virology* 87:11966–11977. DOI: 10.1128/JVI.02323-13.

648 O'Leary NA., Wright MW., Brister JR., Ciufo S., Haddad D., McVeigh R., Rajput B., Robbertse
649 B., Smith-White B., Ako-Adjei D., Astashyn A., Badretdin A., Bao Y., Blinkova O., Brover
650 V., Chetvernin V., Choi J., Cox E., Ermolaeva O., Farrell CM., Goldfarb T., Gupta T., Haft
651 D., Hatcher E., Hlavina W., Joardar VS., Kodali VK., Li W., Maglott D., Masterson P.,
652 McGarvey KM., Murphy MR., O'Neill K., Pujar S., Rangwala SH., Rausch D., Riddick
653 LD., Schoch C., Shkeda A., Storz SS., Sun H., Thibaud-Nissen F., Tolstoy I., Tully RE.,
654 Vatsan AR., Wallin C., Webb D., Wu W., Landrum MJ., Kimchi A., Tatusova T., DiCuccio
655 M., Kitts P., Murphy TD., Pruitt KD. 2016. Reference sequence (RefSeq) database at
656 NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids*
657 *Research* 44:D733–D745. DOI: 10.1093/nar/gkv1189.

658 Park C., Zhang J. 2012. High Expression Hampers Horizontal Gene Transfer. *Genome Biology*
659 *and Evolution* 4:523–532. DOI: 10.1093/gbe/evs030

660 Paradis E., Claude J. Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R
661 language. *Bioinformatics* 20:289-290.

662 Philippe H., Derelle R., Lopez P., Pick K., Borchiellini C., Boury-Esnault N., Vacelet J., Renard
663 E., Houliston E., Quéinnec E., Da Silva C., Wincker P., Le Guyader H., Leys S., Jackson

664 DJ., Schreiber F., Erpenbeck D., Morgenstern B., Wörheide G., Manuel M. 2009.
665 Phylogenomics revives traditional views on deep animal relationships. *Current biology*
666 19:706–12. DOI: 10.1016/j.cub.2009.02.052.

667 Pible O., Hartmann EM., Imbert G., Armengaud J. 2014. The importance of recognizing and
668 reporting sequence database contamination for proteomics. *EuPA Open Proteomics* 3:246–
669 249. DOI: 10.1016/J.EUPROT.2014.04.001.

670 Pick KS., Philippe H., Schreiber F., Erpenbeck D., Jackson DJ., Wrede P., Wiens M., Alié A.,
671 Morgenstern B., Manuel M., Wörheide G. 2010. Improved phylogenomic taxon sampling
672 noticeably affects nonbilaterian relationships. *Molecular biology and evolution* 27:1983–7.
673 DOI: 10.1093/molbev/msq089.

674 Pisani D., Pett W., Dohrmann M., Feuda R., Rota-Stabelli O., Philippe H., Lartillot N., Wörheide
675 G. 2015. Genomic data do not support comb jellies as the sister group to all other animals.
676 *Proceedings of the National Academy of Sciences of the United States of America*
677 112:15402–7. DOI: 10.1073/pnas.1518127112.

678 Pundir S., Martin MJ., O'Donovan C. 2017. Uniprot protein knowledgebase. *Protein*
679 *Bioinformatics: From Protein Modifications and Networks to Proteomics*, 41-55.

680 Radzvilavicius AL., Hadjivasiliou Z., Pomiankowski A., Lane N. 2016. Selection for
681 Mitochondrial Quality Drives Evolution of the Germline. *PLoS biology* 14:e2000410. DOI:
682 10.1371/journal.pbio.2000410.

683 Richards TA., Dacks JB., Campbell SA., Blanchard JL., Foster PG., McLeod R., Roberts CW.
684 2006. Evolutionary origins of the eukaryotic shikimate pathway: gene fusions, horizontal
685 gene transfer, and endosymbiotic replacements. *Eukaryotic cell* 5:1517–31. DOI:
686 10.1128/EC.00106-06.

687 Ryan JF., Pang K., Schnitzler CE., Nguyen A-D., Moreland RT., Simmons DK., Koch BJ.,
688 Francis WR., Havlak P., Smith SA., Putnam NH., Haddock SHD., Dunn CW., Wolfsberg
689 TG., Mullikin JC., Martindale MQ., Baxevanis AD. 2013. The Genome of the Ctenophore
690 *Mnemiopsis leidyi* and Its Implications for Cell Type Evolution. *Science* 342:1242592. DOI:
691 10.1126/science.1242592

692 Salter SJ., Cox MJ., Turek EM., Calus ST., Cookson WO., Moffatt MF., Turner P., Parkhill J.,
693 Loman NJ., Walker AW. 2014. Reagent and laboratory contamination can critically impact
694 sequence-based microbiome analyses. *BMC Biology* 12:87. DOI: 10.1186/s12915-014-
695 0087-z.

696 Salzberg SL., White O., Peterson J., Eisen JA. 2001. Microbial Genes in the Human Genome:
697 Lateral Transfer or Gene Loss? *Science* 292:1903-1906. DOI: 10.1126/science.1061036

698 Schönknecht G., Weber APM., Lercher MJ. 2014. Horizontal gene acquisitions by eukaryotes as
699 drivers of adaptive evolution. *BioEssays*, 36:9–20. DOI: 10.1002/bies.201300095

700 Shen X-X., Hittinger CT., Rokas A. 2017. Contentious relationships in phylogenomic studies can
701 be driven by a handful of genes. *Nature ecology & evolution* 1:0126. DOI: 10.1038/s41559-
702 017-0126

703 Shimodaira H., Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree
704 selection. *BIOINFORMATICS APPLICATIONS NOTE* 17:1246–1247.

705 Simion P., Philippe H., Baurain D., Jager M., Richter DJ., Di Franco A., Roure B., Satoh N.,
706 Quéinnec É., Ereskovsky A., Lapébie P., Corre E., Delsuc F., King N., Wörheide G.,
707 Manuel M. 2017. A Large and Consistent Phylogenomic Dataset Supports Sponges as the
708 Sister Group to All Other Animals. *Current Biology* 27:958–967. DOI:
709 10.1016/j.cub.2017.02.031.

710 Song J-I., Hwang S-J. 2010. New Species of Genus *Coeloplana* (Ctenophora: Tentaculata:
711 Platyctenida) from Korea. *Animal Systematics, Evolution and Diversity* 26:217–221. DOI:
712 10.5635/KJSZ.2010.26.3.217

713 Soucy SM., Huang J., Gogarten JP. 2015. Horizontal gene transfer: building the web of life.
714 *Nature Reviews Genetics* 16:472–482. DOI: 10.1038/nrg3962

715 Stamatakis A. 2014 RAxML Version 8: A tool for phylogenetic analysis and post-analysis of
716 large phylogenies. *Bioinformatics* 30:1312–1313.

717 Stanhope MJ., Lupas A., Italia MJ., Koretke KK., Volker C., Brown JR. 2001. Phylogenetic
718 analyses do not support horizontal gene transfers from bacteria to vertebrates. *Nature*
719 411:940–944. DOI: 10.1038/35082058.

720 Strong MJ., Xu G., Morici L., Bon-Durant SS., Baddoo M., Lin Z., Fewell C., Taylor CM.,
721 Flemington EK. 2014. Microbial Contamination in Next Generation Sequencing:
722 Implications for Sequence-Based Analysis of Clinical Samples. *PLoS Pathogens*
723 10:e1004437. DOI: 10.1371/journal.ppat.1004437.

724 Sun BF., Xiao JH., He SM., Liu L., Murphy RW., Huang DW. 2013. Multiple ancient horizontal
725 gene transfers and duplications in lepidopteran species. *Insect Molecular Biology* 22:72–87.
726 DOI: 10.1111/imb.12004.

727 Telford MJ., Budd GE., Philippe H. 2015. Phylogenomic Insights into Animal Evolution.
728 *Current Biology* 25:R876–R887. DOI: 10.1016/j.cub.2015.07.060

729 Thomas CM., Nielsen KM. 2005. Mechanisms of, and Barriers to, Horizontal Gene Transfer
730 between Bacteria. *Nature Reviews Microbiology* 3:711–721. DOI: 10.1038/nrmicro1234

731 Torruella G., de Mendoza A., Grau-Bové X., Antó M., Chaplin MA., del Campo J., Eme L.,
732 Pérez-Cordón G., Whipps CM., Nichols KM., Paley R., Roger AJ., Sitjà-Bobadilla A.,
733 Donachie S., Ruiz-Trillo I. 2015. Phylogenomics Reveals Convergent Evolution of

734 Lifestyles in Close Relatives of Animals and Fungi. *Current Biology* 25:2404–2410. DOI:
735 10.1016/j.cub.2015.07.053.

736 Whelan NV., Kocot KM., Moroz LL., Halanych KM. 2015. Error, signal, and the placement of
737 Ctenophora sister to all other animals. *Proceedings of the National Academy of Sciences of*
738 *the United States of America* 112:5773–8. DOI: 10.1073/pnas.1503453112.

739 Whelan NV., Kocot KM., Moroz TP., Mukherjee K., Williams P., Paulay G., Moroz LL.,
740 Halanych KM. 2017. Ctenophore relationships and their placement as the sister group to all
741 other animals. *Nature Ecology & Evolution* 1:1737–1746. DOI: 10.1038/s41559-017-0331-
742 3.

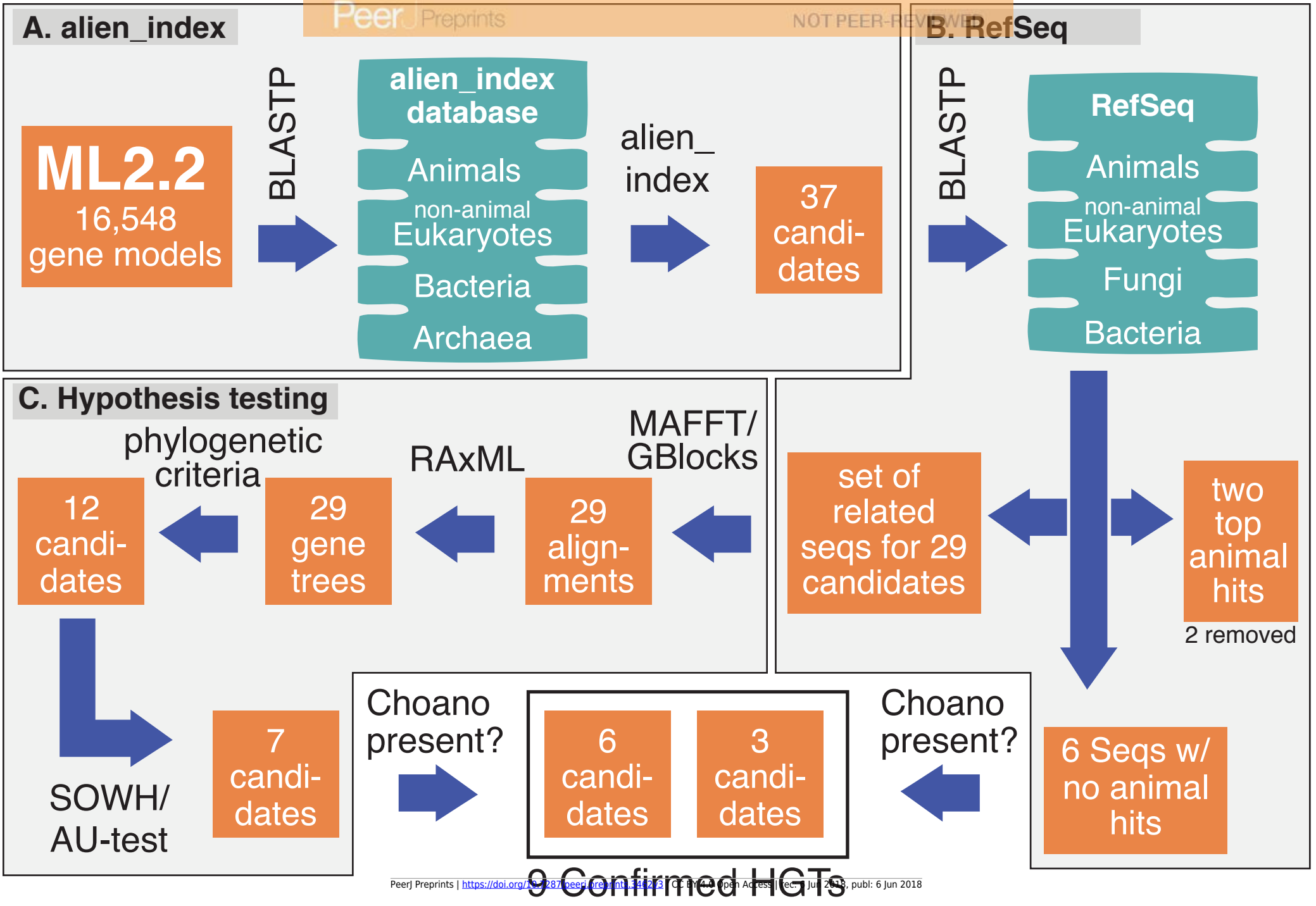
743 Wybouw N., Pauchet Y., Heckel DG., Van Leeuwen T. 2016. Horizontal Gene Transfer
744 Contributes to the Evolution of Arthropod Herbivory. *Genome biology and evolution*
745 8:1785–801. DOI: 10.1093/gbe/evw119.

746 Zhu B., Lou M-M., Xie G-L., Zhang G-Q., Zhou X-P., Li B., Jin G-L. 2011. Horizontal gene
747 transfer in silkworm, *Bombyx mori*. *BMC Genomics* 12:248. DOI: 10.1186/1471-2164-12-
748 248.

Figure 1(on next page)

Pipeline and outputs to identify and confirm HGTs.

(A) alien_index was used to identify 37 HGT candidates. (B) These candidates were then BLASTed against RefSeq; two candidates were removed because they only had two significant animals hits and six were set aside for future testing because they lacked animal hits. (C) The remaining 29 candidates were tested by phylogenetic analyses and hypothesis testing (SOWH and AU test). The 6 candidates that lacked animal hits and those that passed hypothesis testing (7 candidates) were screened for significant hits to Choanoflagellates. More details on genes passing through the pipeline are described in Table S2.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3462v3 | CC BY 4.0 Open Access | rec: Jun 2018, publ: 6 Jun 2018

Figure 2(on next page)

Maximum-likelihood best tree and metazoan-constraint tree compared in the SOWH and AU tests.

Gene IDs (in black) denote the putative HGTs. (A), (C), and (E) are examples of RAxML best trees for HGT candidates validated by phylogenetic analyses and hypothesis testing. (B), (D), and (F) are examples of trees where putative HGTs have been constrained to produce monophyletic Animalia and have been optimized in RAxML. Taxa that are prefixed "META_" are from our alien_index database version 0.01 (i.e., META_NVEC (*Nematostella vectensis*), META_TADH (*Trichoplax adhaerens*), META_HSAP (*Homo sapiens*), META_CTEL (*Capitella teleta*), META_DMEL (*Drosophila melanogaster*), META_AQUE (*Amphimedon queenslandica*). MET=Metazoa; BAC=Bacteria; EUK=Eukaryota; FUN=Fungi; More details for each taxon are specified in Table S3. The asterisk indicates a gene that is later removed from contention.

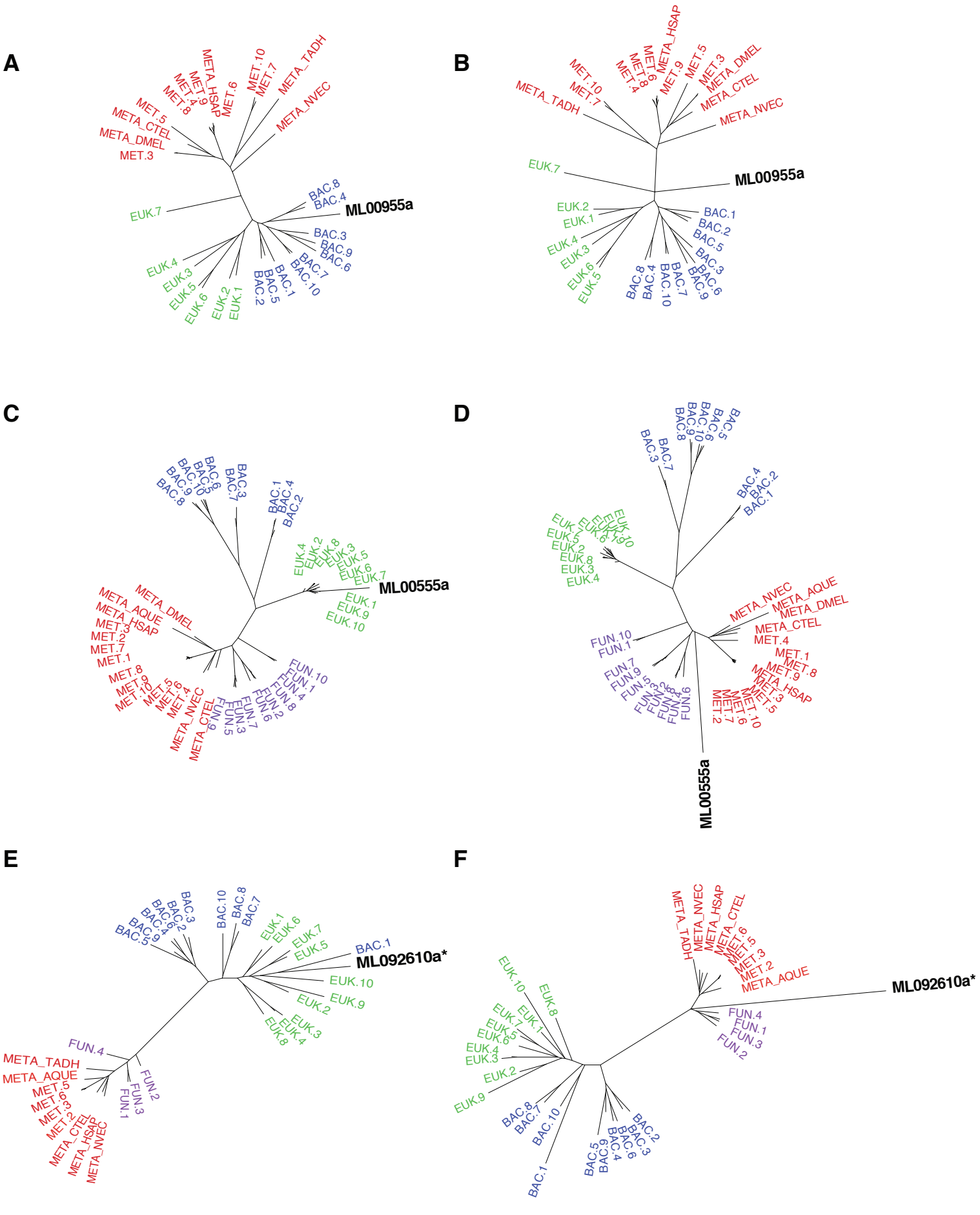
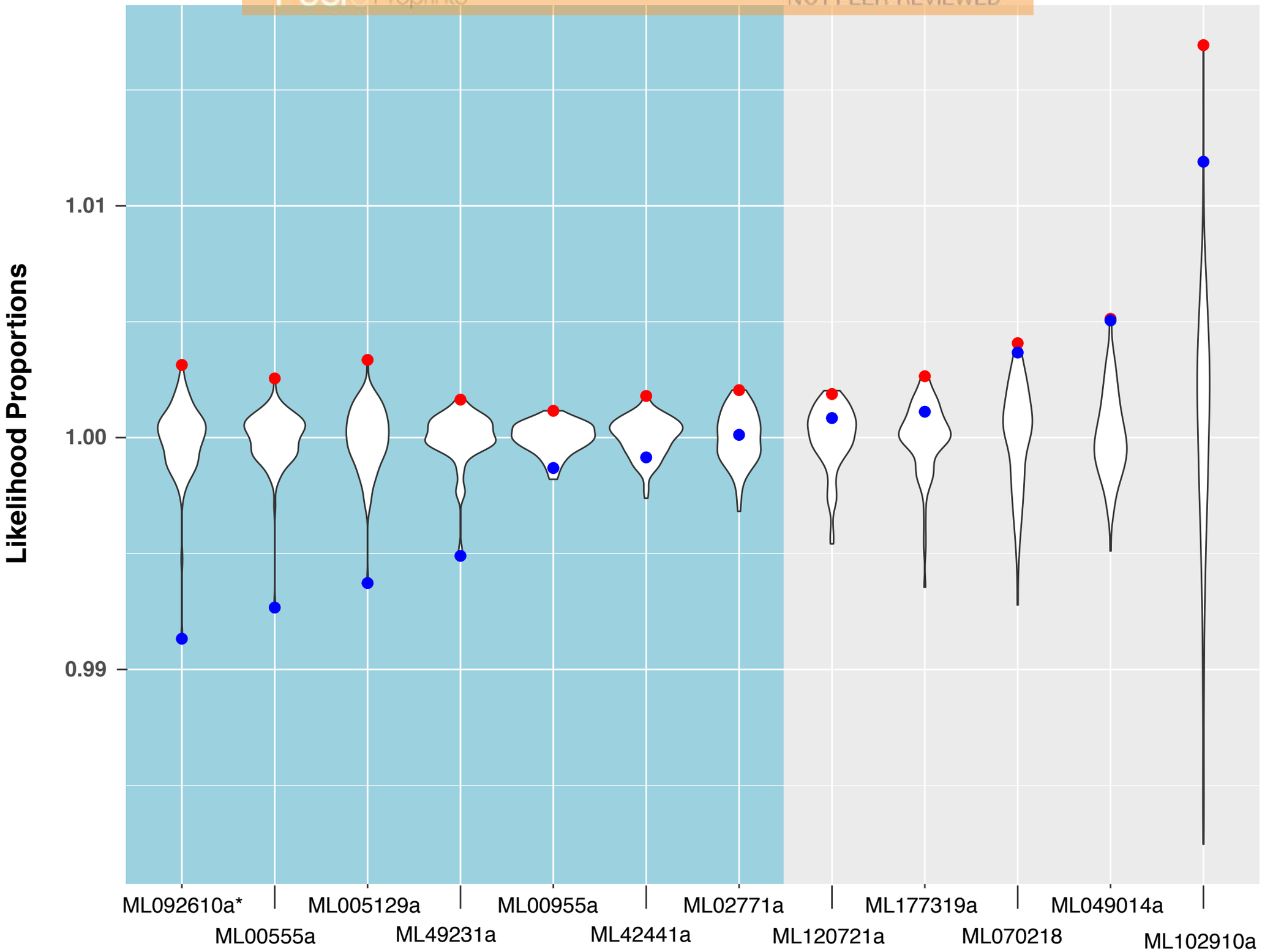


Figure 3(on next page)

A comparison of likelihood proportions between the best tree, metazoan-constrained tree, and bootstrap trees for HGT candidates with BLAST hits to Metazoa.

Likelihood proportions are individual likelihood values divided by the average likelihood value for suboptimal trees (i.e., bootstrap trees). Red points indicate likelihood proportions of the best tree (i.e., tree indicating HGT). Blue points indicate likelihood proportions of the metazoan constrained tree (i.e., tree contradicting HGT). The violin plot shows the distribution of likelihood proportions of 100 bootstrap trees for each HGT candidate. The side in teal shows HGT candidates validated by hypothesis testing and the side in gray shows HGT candidates unsupported by hypothesis testing. The asterisk indicates a gene that is later removed from contention.

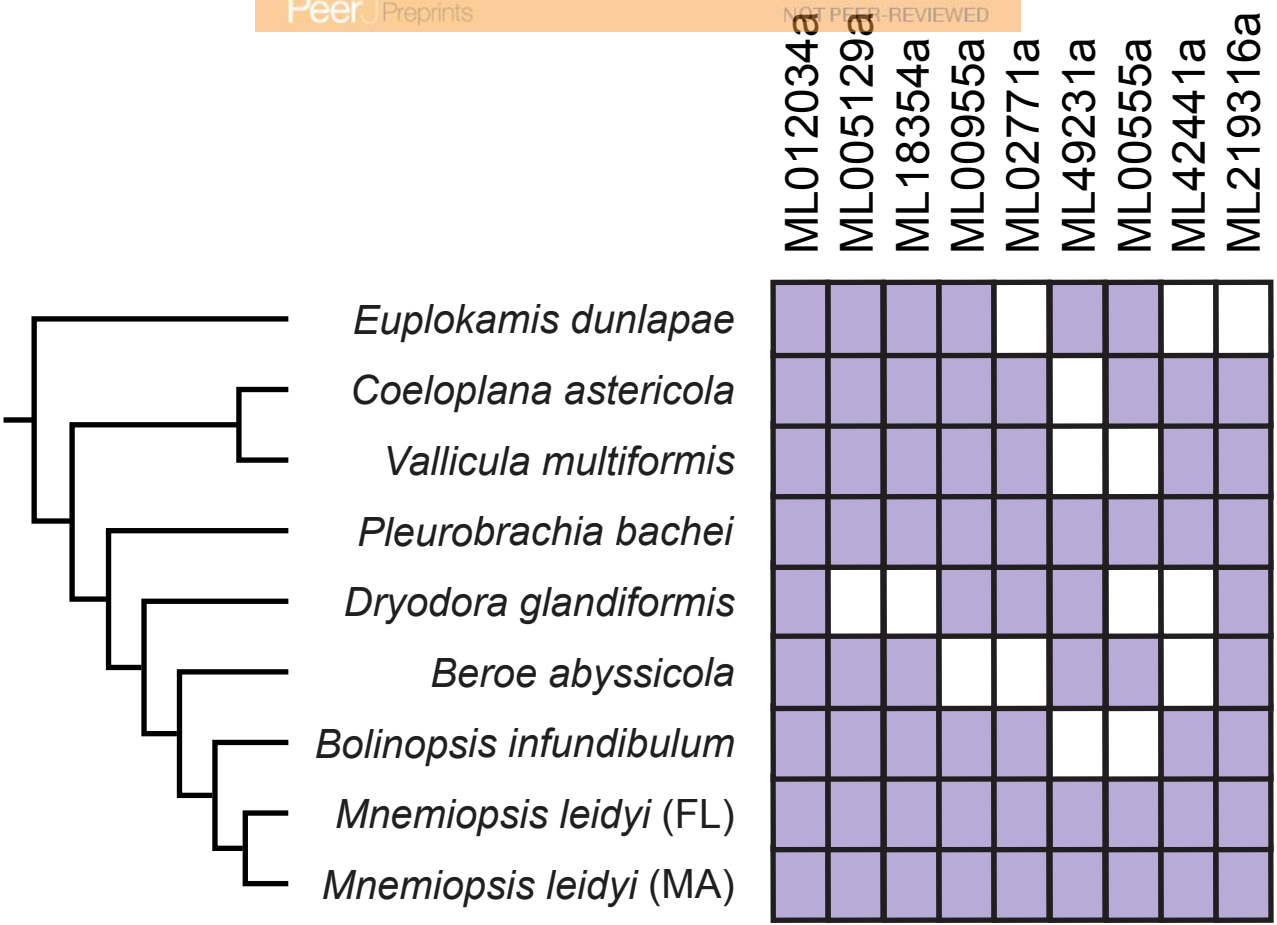


PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3462v3 | CC BY 4.0 Open Access | rec: 6 Jun 2018, publ: 6 Jun 2018

Figure 4(on next page)

Expression of confirmed HGTs from the *M. leidyi* genome in ctenophore transcriptomes.

Purple boxes indicate the specified HGT is present in the species' transcriptome confirmed by reciprocal best BLAST hits; white boxes indicate the gene is absent in the species' transcriptome. Tree was inferred by Moroz et al. (2014). Percent identity among genes are described in Table S4.

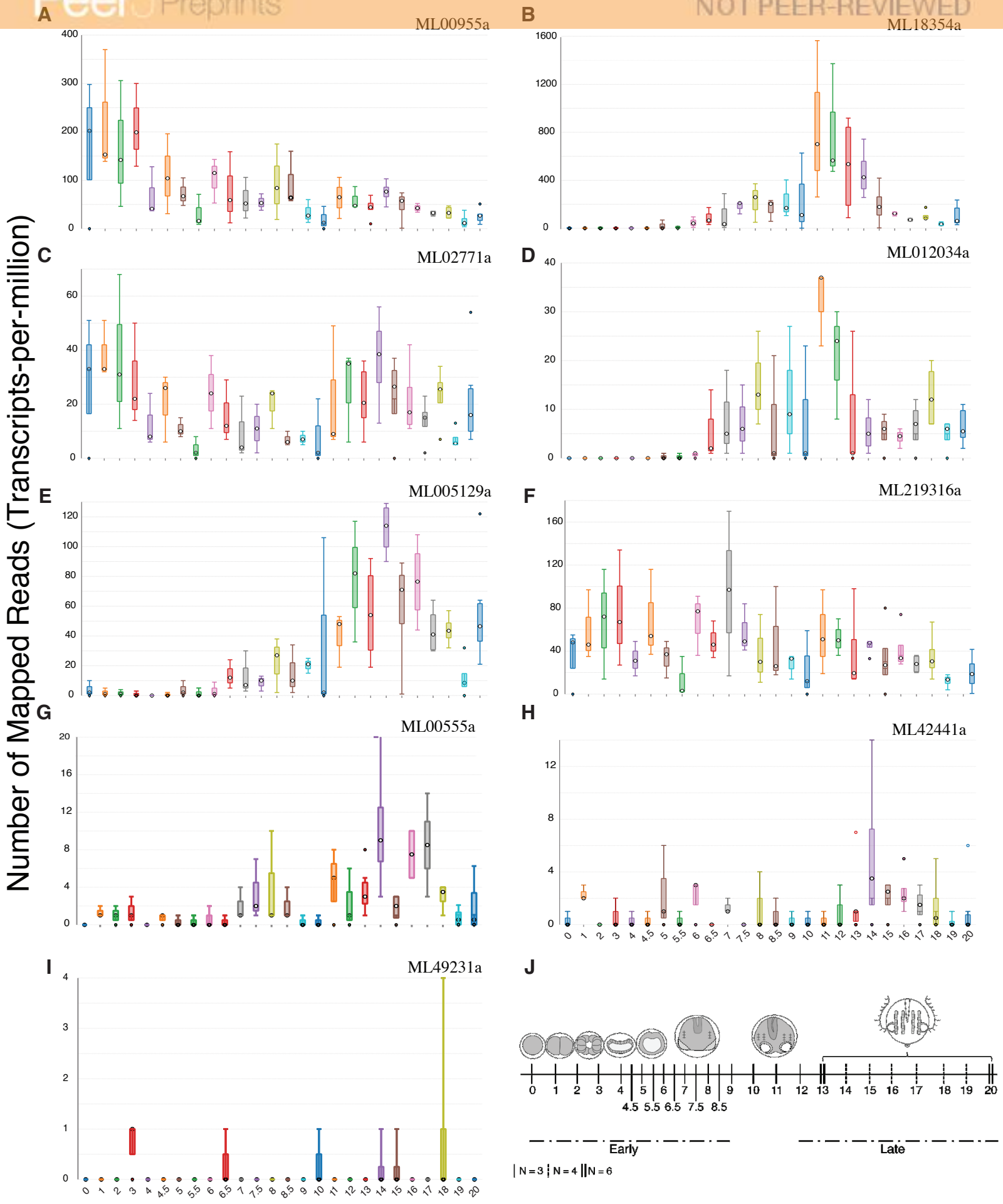


Peerj Preprints | https://doi.org/10.7287/peerj.preprints.3462v3 | CC BY 4.0 Open Access | rec: 6 Jun 2018, publ: 6 Jun 2018

Figure 5(on next page)

Expression profiles of the nine HGTs identified in this study.

Single-embryo RNA-Seq analyses were performed over 20 hours. (A-F) Confirmed HGTs with tpm values (medians for each set of time point replicates) greater than or equal to 100 over 20 hours (25 time points) are shown. (G-I) Confirmed HGTs with tpm values less than 100 over 20 hours. (J) Ctenophore stages of development over the timecourse. Early cleavage stages occur at 1-3 hpf. Gastrulation occurs at 4-6 hpf . Tentacle morphogenesis occurs at 9-12 hpf. N refers to the number of replicates.



PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3462v3 | CC BY 4.0 Open Access | rec: 6 Jun 2018, publ: 6 Jun 2018

Table 1(on next page)

Hypothesis testing on HGT candidates that were confirmed by phylogenetic analyses.

P-values indicate the level of support for HGTs in comparison to the metazoan constraint tree for the SOWH test and suboptimal trees (bootstrap and manually generated) in the AU test. Candidates in blue have significant values in all three tests ($p \leq 0.05$) likely HGTs. The asterisk indicates a gene that is later removed from contention.

Genes	SOWH p-value	AU Bootstrap p-value	AU Manual p-value
ML00555a	<0.001	4.00E-45	7.00E-06
ML49231a	<0.001	2.00E-44	7.00E-103
ML092610a*	<0.001	2.00E-31	4.00E-68
ML005129a	<0.001	1.00E-04	6.00E-06
ML00955a	<0.001	0.021	0.002
ML02771a	<0.001	0.023	0.029
ML42441a	<0.001	0.047	0.022
ML177319a	<0.001	0.226	0.042
ML120721a	<0.001	0.48	0.245
ML049014a	0.985	0.862	0.604
ML070218a	0.262	0.849	0.361
ML102910a	0.229	0.719	0.255

1

Table 2(on next page)

Intron structure of nine HGTs and surrounding genes.

The genes highlighted in red are the HGT candidates. The gene with an asterisk indicates one non-spliceosomal intron.

Candidate HGT	Number of introns in candidate HGTs and surrounding genes						
ML00955a	ML00952a 5	ML00953a 0	ML00954a 4	ML00955a 1	ML00956a 0	ML00957a 7	ML00958a 2
ML18354a	ML18351a 6	ML18352a 14	ML18353a 7	ML18354a 5	ML18355a 1	ML18356a 0	ML18357a 0
ML02771a				ML02771a 7	ML02772a 8	ML02773a 16	ML02774a 3
ML012034a	ML012031a 6	ML012032a 5	ML012033a 6	ML012034a 0	ML012035a 3	ML012036a 6	
ML005129a	ML005126a 0	ML005127a 7	ML005128a 2	ML005129a 1	ML005130a 0	ML005131a 4	ML005132a 8
ML219316a	ML219313a 6	ML219314a 3	ML219315a 7*	ML219316a 4	ML219317a 6		
ML00555a	ML00552a 0	ML00553a 12	ML00554a 0	ML00555a 14	ML00556a 3	ML00557a 0	ML00558a 5
ML42441a				ML42441a 1	ML42442a 14	ML42443a 3	ML42444a 10
ML49231a				ML49231a 6			

1

Table 3(on next page)

Sumary of confirmed HGT origins and functions.

HGT functions were determined by BLAST against the UniProt database and associated Pfam-A domains were searched on the *Mnemiopsis* Genome Portal. The origin column shows the domains of life from which these genes are predicted to have been transferred (Bac = Bacteria; Euk = Eukaryota). The RefSeq column shows a more detailed classification for the origin of HGTs. All rows highlighted in orange indicate genes that show developmental expression.

Genes	Function	Pfam Domains	Origin	Lineage
ML00955a	Putative metalloendopeptidase	Peptidase family M13	Bac	Proteobacteria
ML005129a	2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	2OG-Fe(II) oxygenase superfamily	Euk	Unknown
ML02771a	Penicillin acylase	Penicillin amidase	Bac	Unknown
ML012034a	Uncharacterized protein	2OG-Fe(II) oxygenase superfamily	Euk	Unknown
ML18354a	Putative chalcone and stilbene synthase	Chalcone and stilbene synthases, 3-Oxoacyl- synthase III, FAE1/Type III polyketide synthase	Bac	Unknown
ML219316a	Uncharacterized protein		Bac	Firmicutes
ML00555a	Phospholipase D alpha 1	C2, Phospholipase D	Euk	Unknown
ML49231a	Phospholipase D gamma 1	Phospholipase D	Euk	Rhodophyta
ML42441a	NADH dehydrogenase, putative	Pyridine nucleotide-disulphide oxidoreductase	Euk	Unknown

1