Panpulmonate transcriptomes reveal candidate genes involved in the adaptation to freshwater and terrestrial habitats in gastropods

Authors' names and affiliations

Pedro E. Romero\textsuperscript{1,2}
pedro.romero@senckenberg.de
1 Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.
2 Institute for Ecology, Evolution & Diversity, Faculty of Biological Sciences, Goethe University Frankfurt, Max-von-Laue-Straße 13, 60438 Frankfurt am Main, Germany.

Barbara Feldmeyer\textsuperscript{1}
bfeldmeyer@senckenberg.de
1 Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.

Markus Pfenninger\textsuperscript{1,2}
pfenninger@bio.uni-frankfurt.de
1 Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany.
2 Institute for Ecology, Evolution & Diversity, Faculty of Biological Sciences, Goethe University Frankfurt, Max-von-Laue-Straße 13, 60438 Frankfurt am Main, Germany.
Corresponding author

Pedro E. Romero ¹,²

pedro.romero@senckenberg.de

¹ Senckenberg Biodiversity and Climate Research Centre (BiK-F), Senckenberganlage 25, 60325 Frankfurt am Main, Germany

² Institute for Ecology, Evolution & Diversity, Faculty of Biological Sciences, Goethe University Frankfurt, Max-von-Laue-Straße 13, 60438 Frankfurt am Main, Germany.
Abstract

Background

The conquest of the land from aquatic habitats is a fascinating evolutionary event that happened multiple times in different phyla. Mollusks are among the organisms that successfully invaded the non-marine realm, resulting in the radiation of terrestrial panpulmonate gastropods. We compared transcriptomes from panpulmonates to study the selective pressures that modeled the transitions from marine into freshwater and terrestrial realms in this molluscan lineage.

Results

De novo assembly of six panpulmonate transcriptomes resulted in 55,000 - 97,000 predicted open reading frames, of which 9 - 14% were functionally annotated. Adding published transcriptomes, we predicted 791 ortholog clusters shared among fifteen panpulmonate species, resulting in 702 amino acid and 736 codon-wise alignments. The branch-site test of positive selection applied to the codon-wise alignments showed twenty-eight genes under positive selection in the freshwater lineages and seven in the terrestrial lineages. Gene ontology categories of these candidate genes include actin assembly, transport of glucose, and the tyrosine metabolism in the terrestrial lineages; and, DNA repair, metabolism of xenobiotics, mitochondrial electron transport, and ribosome biogenesis in the freshwater lineages.

Conclusions

We identified candidate genes representing processes that may have played a key role during the water-to-land transition in Panpulmonata. These genes were involved in energy metabolism and gas-exchange surface development in the terrestrial lineages and in the response to the abiotic stress factors (UV radiation, osmotic pressure, xenobiotics) in the freshwater lineages. Our study expands the knowledge of possible adaptive signatures in genes and metabolic pathways related to the invasion of non-marine habitats in invertebrates.
Keywords
Realm transition, Phylogenomics, Panpulmonata, Positive selection.

Background
The invasion from marine to non-marine habitats is one of the most enthralling events in the evolution of life on Earth. The transition from sea to freshwater and land environments occurred multiple times in different branches of the tree of life. Mollusks, along arthropods and vertebrates, are among the successful phyla that invaded the non-marine realm. Several branches from the molluscan class Gastropoda (Neritimorpha, Cyclophoroidea, Littorinoidea, Rissooidea, and Panpulmonata) have colonized terrestrial habitats multiple times [1, 2]. Especially, several independent land invasions in the Panpulmonata resulted in a significant adaptive radiation and explosive diversification that likely originated up to a third of the extant molluscan diversity [3]. Therefore, panpulmonate lineages are a promising system to study evolution of adaptations to non-marine habitats.

The habitat transition must have triggered several novel adaptations in behavior, breathing, excretion, locomotion, and osmotic and temperature regulation, to overcome problems that did not exist in the oceans such as dehydration, lack of buoyancy force, extreme temperature fluctuations and radiation damage [4-6]. Studies in vertebrates showed different genomic changes involved in the adaptation to the new habitats. Mudskippers, amphibious teleost fishes adapted to live on mudflats, possess unique immune genes to possibly counteract novel pathogens on land, and opsin genes for aerial vision and for enhancement of color vision [7]. Tetrapods showed adaptation signatures in the carbamoyl phosphate synthase I (CPS1) gene involved in the efficient production of hepatic urea [8]. Primitive sarcopterygians like the coelacanth Latimeria already possess various conserved non-coding elements (CNE) that enhance the development of limbs, and an expanded repertoire of genes related to the pheromone receptor VR1 that may have facilitated the adaptation to sense airborne chemicals during the water-to-land transition in tetrapods [9]. Also, vertebrate
keratin genes responsible for skin rigidity underwent a functional diversification after the water-to-
land transition, enhancing the protection against friction imposed by the new terrestrial lifestyle [10].

Conversely, information about the molecular basis of adaptation from marine to non-marine habitats in invertebrates is still scarce. Only one study reported adaptive signals in gene families (e. g. ATPases, DNA repair, and ribosomal proteins) that may have played a key role during terrestrialization in springtails and insects (Hexapoda) [11], clades that probably had a common pancrustacean ancestor living in a shallow marine environment [12, 13]. Mutations in the ATPases were suggested to provide the necessary energy to adapt to new high-energy demanding habitats [14], DNA repair genes would have helped reducing the damage produced by increased ultraviolet (UV) irradiation, and finally, as the ribosomal machinery is salt-sensitive, adaptive signs in the ribosomal proteins could have been a result of the different osmotic pressures within aquatic and terrestrial environments [15].

In a previous paper, we explored the adaptive signals in the mitochondrial genomes of panpulmonates [16]. We found that in the branches leading to lineages with terrestrial taxa (Ellobioidea and Stylommatophora), the mitochondrial genes cob and nad5, both involved in the oxidative phosphorylation pathway that finally produces ATP, appeared under positive selection. Moreover, the amino acid positions under selection have been related to an increased energy production probably linked to novel demands of locomotion [17, 18], and to changes in the equilibrium constant physicochemical property involved in the regulation of ROS production and thus, in the ability to tolerate new abiotic stress conditions [19].

Here, we expanded our search for candidate genes related to the adaptation to non-marine habitats, using transcriptome-wide data from several panpulmonate taxa, including marine, intertidal, freshwater and terrestrial lineages. We used a phylogenomic approach to reconstruct the evolutionary relationships of Panpulmonata and then tested for positive selection in the branches
leading to freshwater and land snails. This approach aims to provide new insights into the selective pressures shaping the transition from marine to freshwater and land lifestyles.

Results

We generated approximately 2,100,000 - 3,400,000 Illumina for our six samples (five ellobiids and one stylommatophoran species, Table 1). The quality trimming eliminated 14 - 39% of short and low-quality fragments in our samples. De novo meta assembly with MIRA produced approximately 55,000 - 98,000 transcripts in our samples and 54,000 - 130,000 in the other additional samples (Table 1). For further analyses we used transcripts larger than 300 bp. This represented a reduction of less than 1% in our samples but a higher reduction in the public data (3 - 35%). The number of predicted open reading frames was very similar to the number of transcripts > 300 bp in almost all cases, the only exception was Radix balthica, where only 57% of the transcripts obtained an ORF prediction. We obtained 9,000 - 30,000 single blast hits for our data, representing 5,000 - 13,000 single annotated genes. The percentage of annotated genes from our open reading frame data was 9 - 14%.

We predicted 791 orthologous clusters shared among all species, of which 702 ortholog clusters remained after removing spurious and poorly amino acid aligned sequences in trimAL. From this dataset, MARE selected 382 informative clusters to reconstruct the phylogeny of the panpulmonate species (Additional File 1). The amount of missing data corresponds to 10.94% in the complete matrix, and 6.26% in the reduced matrix (Additional Files 2 and 3, respectively).

Most branches in the panpulmonate tree received high support (Figure 1). The clade containing Stylommatophora and Systellommatophora was significantly supported (bootstrap: 94 / posterior probability: 1.0) and appeared as a sister of the monophyletic Ellobioidea (99/1.0). The Acochlidia clade was moderately supported (86/1.0). The association of the Acochlidia with the Ellobioidea, Stylommatophora, Systellommatophora clade had no significant bootstrap support but a high
posterior probability (64/1.0). The Hygrophila clade was highly supported (100/1.0). The association of Amphiboloidea and Pyramidelloidea was also highly supported (100/1.0).

We detected selection signatures on genes (codon-wise alignments) across the freshwater and terrestrial lineages in Panpulmonata. The likelihood-ratio test (LRT) comparing the branch-site model A against the null model (neutral) showed seven orthologous clusters under positive selection in the land lineages and twenty-eight clusters in the freshwater lineages (Additional File 4). There was no overlapping within positively selected genes from freshwater and terrestrial lineages. Table 2 shows examples of these candidate genes, their annotations, biological processes, molecular functions, and pathways involved. The BlastX annotations revealed candidate genes involved in the actin assembly, protein folding, transport of glucose, and vesicle transport in the terrestrial lineages. In the freshwater lineages, we found candidate genes associated to DNA repair, metabolism of xenobiotics, mitochondrial electron transport, protein folding, proteolysis, ribosome biogenesis, RNA processing and transport of lipids (Additional Files 5 and 6). We found no significant enriched GO (Gene ontology) terms neither in the freshwater nor terrestrial lineages.

Candidate genes under positive selection in the terrestrial lineages were involved in the carbohydrate digestion, endocytosis, focal adhesion, and the metabolism of lipids and tyrosine pathways. In case of the freshwater lineages, the candidate genes were involved in several metabolic pathways, for example, amino acid biosynthesis, focal adhesion, lysosome, oxidative phosphorylation, and protein signaling (Table 2, Additional Files 5 and 6).

Discussion

Panpulmonates transitioned from marine to freshwater and terrestrial environments in several lineages and multiple times [20, 2, 21]. Thus, they are a very suitable model to study the invasion of non-marine realms. However, the phylogenetic relationships within this clade are yet to be resolved [20]. Our tree topology using 382 orthologous clusters resembles the one obtained from Jörger et al. [21], based on mitochondrial and nuclear markers. In addition, we found support for the Geophila:
Stylommatophora (terrestrial) and Systellommatophora (intertidal/terrestrial) as sister groups. This clade has been proposed before based on the position of the eyes at the tip of cephalic tentacles [22]. Still, previous phylogenies using mitochondrial and nuclear markers failed to support this clade [23, 16, 24, 21]. We also found support for Eupulmonata (sensu Morton [25, 23]), a clade comprising Stylommatophora and Systellommatophora plus Ellobioidea (intertidal/terrestrial) [20], this clade was supported using a combination of mitochondrial and nuclear markers [21].

Generation of high-quality transcriptomic data for other panpulmonate clades (marine Sacoglossa and Siphonarioidea, freshwater Glacidorboidea), and additional data for terrestrial Stylommatophora and Systellommatophora, will definitely illuminate the evolutionary relationships in Panpulmonata.

Our study is the first genome-wide report on the molecular basis of adaptation to non-marine habitats in panpulmonate gastropods. In case of the terrestrial lineages, we found evidence that the different positively selected genes are involved in a general pattern of adaptation to increased energy demands. The adaptive signs found in a gene related to actin assembly (OG0001172, Table 2) can be related to the necessity to move (forage, hunt preys or escape from predators) in the terrestrial realms. Moreover, the displacement in an environment lacking the buoyancy force to float or swim requires more energy, which can be obtained by increasing the glucose uptake (OG0000137) to produce energy in form of ATP. The adaptive signatures we found previously in two mitochondrial genes, cob and nad5, involved in energy production in the mitochondrion, also suggested a response to new metabolic requirements in the terrestrial realm, such as the increase of energy demands (to move and sustain the body mass).

One gene found under positive selection in the terrestrial genus Pythia, was involved in the metabolism of tyrosine (OG0000060). Tyrosine is the principal component of the thyroid hormones (TH). Despite invertebrates lack the thyroid gland responsible of the production of TH’s; the synthesis of TH’s has been demonstrated in mollusks and echinoderms. In these organisms, iodine is ligated to the tyrosine in the peroxisomes, producing thyroid hormones [26]. Notably, it has been suggested that iodinated tyrosine may have been essential in vertebrates during the transition to
terrestrial habitats for TH's are required in the expression of transcription factors involved in the embryonic development and differentiation of the lungs [27]. Land snails adapted to breath air by losing their gills and transforming the inner surface of their mantle into a lung [5]. Therefore, we propose that the tyrosine pathway was also a key component in invertebrates probably promoting the development of novel gas exchange tissues in land snails.

In case of the freshwater lineages, one of the positively selected genes was similar to the subunit 4 of the cytochrome c oxidase (cob) respiratory complex (OG0004174, Table 2). As mentioned above, cob is part of the energy production pathway in the cell. This enzyme complex contains many subunits encoded both in the mitochondrial and nuclear genome. The subunit 4 belongs to the nuclear genome and has an essential role in the assembly and function of the cob complex [28]. In agreement with our previous results that found the mitochondrial cob subunit under positive selection [16], we suggest that this gene was also involved in enhancing the metabolic performance of the enzyme and aided to cope with the new energy demands the realm transition.

A gene similar to cytochrome P450 was also found under positive selection (OG000120). Cytochrome P450s are proteins involved in the metabolism of xenobiotics. They were also under positive selection in the terrestrial Hexapoda lineages in comparison to other water-dwelling arthropods [11]. This result suggests that adaptations in these genes probably improve the response to new organic pollutants and toxins absent in the marine realm.

Another gene that showed adaptive signatures was the 40S ribosomal protein S3a (OG0002708). Likewise, ribosomal genes were also identified in a previous study on land-to-water transitions in hexapods [11] and plants [15]. In the latter study, it was suggested that the difference in the osmotic pressure from aquatic and terrestrial realms could affect the salt-sensitive ribosomal machinery, triggering adaptations to tolerate new salt conditions. This could also be the case for the freshwater animals (hypertonic) in comparison to the marine ones (hypotonic).
Finally, we found adaptive signatures in a DNA methyltransferase gene (OG0004116). This enzyme is part of the DNA repair system in the cell. Specifically, it removes methyl groups from O6-methylguanine produced by carcinogenic agents and it has been showed that its expression is regulated by the presence of ultraviolet B (UVB) radiation [29]. Positive selection on DNA repair genes has been found in hexapods [11], and in vertebrates living in high altitude environments (Tibetan antelopes) [14] or in mudflats (mudskippers) [7], suggesting an important role in the maintenance of the genomic integrity in response to the rise of temperature gradients or UV radiation in the terrestrial realms. In case of the aquatic environments, an extensive review has found an overall negative UVB effect on marine and freshwater animals [30]. However, the authors did not find a significant difference of the survival among taxonomic groups or levels of exposure in marine and freshwater realms, and suggested that the negative effects are highly variable among organisms and depends on several factors including cloudiness, ozone concentration, seasonality, topography, and behavior. Interestingly, it has been reported that survival in the freshwater snail *Physella acuta* (Hygrophila) depends of the combination of a photoenzymatic repair system plus photoprotection provided by the shell thickness and active selection of locations below the water surface avoiding the sunlight [31]

**Conclusions**

We found that the positively selected genes in the terrestrial lineages were related to motility and to the development of novel gas-exchange tissues; while most of the genes in freshwater lineages were related to the response to abiotic stress such osmotic pressure, UV radiation and xenobiotics. These adaptations at the genomic level combined with novel responses in development and behavior probably facilitated the success during the transitions to the non-marine realm. Our results are very promising to understand the genomic basis of the adaptation during the sea-to-land transitions, and also highlight the necessity of more genome-wide studies especially in invertebrates, comparing marine, freshwater and terrestrial taxa, to unravel the evolution of the molecular pathways involved in the invasion of new realms.
Methods

Dataset collection

The dataset from Zapata et al. [32] was used as a starting point for our study. We added to this dataset the transcriptome from *Radix balthica* [33] and retrieved additional freshwater specimens from the NCBI Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra). We complemented the dataset with five intertidal and terrestrial specimens from Ellobioidea (*Carychium* sp., *Cassidula plecotremata, Melampus flavus, Pythia pachyodon, Trimusculus sp.*) and one terrestrial Stylommatophora (*Arion vulgaris*), collected in Japan (2013) and Germany (2014), respectively (Additional File 7). RNA was isolated following the RNeasy kit (QIAGEN) following the manufacturer's protocol. cDNA production and sequencing on the Illumina NextSeq500 platform (150 bp paired-end reads) was performed by StarSEQ GmbH (Mainz, Germany), according to their Illumina standard protocol. The final dataset comprised fifteen transcriptomes of panpulmonate species occurring in marine, intertidal, freshwater and terrestrial habitats (Table 1).

Read processing and quality checking

FastQC [34] was used for initial assessment of reads quality. Then, Trimomatic v0.33 [35] was used to remove and trim Illumina adaptor sequences and other reads with an average quality below 15 within a 4-base wide sliding window. In addition, we repeated the trimming analysis specifying a minimum length of 25 nt for further assembly comparisons. The same procedure was applied to all samples, except for *Radix* (454 reads). In this latter case, we got the transcriptome assembly directly from the author [33].

Transcriptome assembly

*De novo* assembly was performed for all samples, except *Radix* (see last section), using Trinity v2.0.6 [36] with a minimum contig length of 100 amino acids, and Bridger v2014-12-01 [37] with default options. Bridger required the trimmed set with the minimum length of 25 nt. We combined the results from Trinity and Bridger in a meta-assembly using MIRA [38] with default settings. Only sequences with longer than 100 aa were retained for further analyses. This step was done to
improve the accuracy in ortholog determination and facilitate phylogenomic analyses [39]. Furthermore, we used the ORFpredictor server [40] to predict open reading frames (ORF) within the transcripts.

**Construction of ortholog clusters**

Ortholog clusters shared among protein sequences of the fifteen panpulmonate species were predicted using OrthoFinder [41] with default parameters. In case clusters contained more than one sequence per species, only a single sequence per species with the highest average similarity was selected using a homemade script. The predicted amino acid sequences from each ortholog cluster were aligned using MAFFT [42] with standard parameters. Nucleotide sequences in each orthogroup were aligned codon-wise using TranslatorX [43] taking into account the information from the amino acid alignments. Ambiguous aligned regions from the amino acid or codon alignments were removed using Gblocks [44] with standard settings. We used TrimAL [45] to remove poorly aligned or incomplete sequences in each ortholog cluster, using a minimum residue overlap score of 0.75.

**Phylogenomic analyses**

Phylogenetic relationships among the Panpulmonata were reconstructed based on a subset of 382 ortholog clusters. The subset selection was done using MARE [46], a tool designed to find informative subsets of genes and taxa within a large phylogenetic dataset of amino acid sequences. The concatenated amino acid alignment length resulted in 88622 positions. Data were partitioned by gene using the partition scheme suggested in PartitionFinder [47] using the -rcluster option (relaxed hierarchical clustering algorithm), suitable for phylogenomic data [48]. We reconstructed an unrooted tree to be used as an input for the selection analyses. Maximum likelihood analyses were conducted in RAxML-HPC2 (8.0.9) [49]. We followed the “hard and slow way” suggestions indicated in the manual and selected the best-likelihood tree after 1000 independent runs. Then, branch support was evaluated using bootstrapping with 100 replicates, and confidence values were drawn in the best-scoring tree. Bayesian inference was conducted in MrBayes v3.2.2 [50].
simultaneous Monte Carlo Markov Chains (MCMC) were run, with the following parameters: eight chains of 20 million generations each, sampling every 20000 generations and a burn-in of 25%. Tracer 1.6 [51] was used to evaluate effective sample sizes (ESS > 200). We assume that a bootstrap value of >70% and a posterior probability of > 0.95 are evidence of significant nodal support.

**Selection analyses**

The test of positive selection was performed for 736 ortholog clusters (codon-wise alignments) in CODEML implemented in the software PAML v4.8 [52]. PAML estimated the omega ratio ($\omega = \frac{dN}{dS}$: non-synonymous sites / $dS$: synonymous sites); $\omega = 1$ indicates neutral evolution, $\omega < 1$ purifying selection, and $\omega > 1$ indicates positive selection [53]. To detect positive selection affecting sites along the terrestrial or freshwater branches (foreground) in comparison to the intertidal or marine lineages (background), the branch-site model A [54] in CODEML was applied (model = 2, NSsites = 2) for each orthologous cluster. The unrooted tree obtained using maximum likelihood was set as the guide tree. In order to avoid problems in convergence in the log-likelihood calculations, we ran three replicates of model A with different initial omega values ($\omega = 0.5$, $\omega = 1.0$, $\omega = 5.0$). We also calculated the likelihood of the null model (model = 2, NSsites = 2, fixed $\omega = 1.0$). Both models were compared in a likelihood ratio test ($LRT = 2^* (\ln L \text{ model A} - \ln L \text{ null model})$). The Bayes Empirical Bayes (BEB) algorithm implemented in CODEML was used to calculate posterior probabilities of positive selected sites. We corrected p-values with a false discovery rate (FDR) cut-off value of 0.05 using the Benjamini and Hochberg method [55] implemented in R. The statistical significance of the overlap between positively selected genes from freshwater and terrestrial lineages was calculated using the R function `phyper`.

**Functional annotation**

The transcripts were annotated using BlastX [56]. We blasted the nucleotide sequences against the invertebrate protein sequence RefSeq database (release 73, November 2015), with an e-value cut-off of $10^{-6}$. We selected top hits with the best alignment and the lowest e-value. Gene ontology (GO)
terms for each BLASTx search were obtained in the Blast2GO suite [57]. Functional annotation
information was obtained from InterPro database [58] using the InterProScan [59]. GO terms were
then assigned to each orthologous group that was found under positive selection. In addition, we
added to this clusters the metabolic pathway information retrieved from the KAAS server [60]. This
server assigns orthology identifiers from the KEGG database (Kyoto Encyclopedia of Genes and
Genomes). Functional enrichment analysis using the Fisher exact test was also performed in
Blast2GO comparing the genes under positive selection against all ortholog clusters.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent to publish
Not applicable.

Availability of data and materials
Raw sequence data is deposited in the Sequence Read Archive as BioProject (PRJNA339817), in
the NCBI database. All other data sets (including trees, alignments, orthologous clusters, and
scripts) supporting the results are available in the FigShare database: https://dx.doi.org/.

Competing interests
The authors declare that they have no competing interests.

Funding
The project was supported by the German funding program “LOEWE – Landes-Offensive zur
Entwicklung Wissenschaftlich-ökonomischer Exzellenz” of the Hessen State Ministry of Higher
Education, Research and the Arts. PER also received a scholarship from
CONCYTEC/CIENCIACTIVA: Programa de becas de doctorado en el extranjero del Gobierno del
Perú (291-2014-FONDECYT).

Author’s contributions
PER carried out the fieldwork, transcriptome assembly, phylogenetic and molecular evolution
analyses, conceived the study and wrote the manuscript. BF participated in the transcriptome
assemblies and molecular evolution analyses, and helped to draft the manuscript. MP participated in the design and coordination of the study and revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Yasunori Kano and Annette Klussmann-Kolb for their assistance in the fieldwork and Tilman Schell for his assistance in the bioinformatics analyses.

References


### Table 1. Descriptive statistics of the assembled transcriptomes. *De novo* assemblies of new panpulmonate transcriptomes are highlighted in bold.

<table>
<thead>
<tr>
<th>Clade</th>
<th>Species</th>
<th>Habitat</th>
<th>Accession Number (SRA database)</th>
<th>Number of raw reads</th>
<th>For Trinity</th>
<th>For Bridger</th>
<th>All transcripts</th>
<th>Transcripts &gt; 300 (bp)</th>
<th>N50 (bp)</th>
<th>Predicted ORF's</th>
<th>Single hits</th>
<th>Single genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acoclidia</td>
<td>Microhedyle glandulifera</td>
<td>Marine</td>
<td>SRR1505118</td>
<td>6194970</td>
<td>6103287</td>
<td>6028350</td>
<td>129453</td>
<td>84468</td>
<td>742</td>
<td>83983</td>
<td>28750</td>
<td>12772</td>
</tr>
<tr>
<td>Acoclidia</td>
<td>Strubellia wawrai</td>
<td>Freshwater</td>
<td>SRR1505137</td>
<td>24132673</td>
<td>24049385</td>
<td>23436737</td>
<td>82681</td>
<td>79947</td>
<td>1582</td>
<td>79735</td>
<td>27911</td>
<td>13735</td>
</tr>
<tr>
<td>Amphiboloidea</td>
<td>Phallomedusa solida</td>
<td>Intertidal</td>
<td>SRR1505127</td>
<td>25685273</td>
<td>25496722</td>
<td>24822510</td>
<td>68633</td>
<td>65500</td>
<td>1424</td>
<td>64978</td>
<td>24560</td>
<td>12722</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Carychium sp.</td>
<td>Terrestrial</td>
<td>XXXXXXXX</td>
<td>33608344</td>
<td>23461211</td>
<td>23461211</td>
<td>87994</td>
<td>87719</td>
<td>2035</td>
<td>87242</td>
<td>23586</td>
<td>11672</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Cassidula plecotremata</td>
<td>Intertidal</td>
<td>XXXXXXXX</td>
<td>26316221</td>
<td>18928297</td>
<td>17989294</td>
<td>62318</td>
<td>62031</td>
<td>1533</td>
<td>61569</td>
<td>9563</td>
<td>5717</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Melampus flavus</td>
<td>Intertidal</td>
<td>SRR4103303</td>
<td>21068142</td>
<td>16919851</td>
<td>16052613</td>
<td>97629</td>
<td>97334</td>
<td>1640</td>
<td>96816</td>
<td>22187</td>
<td>10778</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Ophicardelus sulcatus</td>
<td>Intertidal</td>
<td>SRR1505124</td>
<td>16026272</td>
<td>15737826</td>
<td>15425623</td>
<td>74467</td>
<td>71073</td>
<td>1499</td>
<td>70573</td>
<td>20851</td>
<td>10937</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Pythia pachyodon</td>
<td>Terrestrial</td>
<td>XXXXXXXX</td>
<td>24016251</td>
<td>20555474</td>
<td>19399330</td>
<td>92548</td>
<td>92264</td>
<td>2035</td>
<td>91759</td>
<td>24077</td>
<td>11776</td>
</tr>
<tr>
<td>Ellobioidea</td>
<td>Trimusculus sp.</td>
<td>Intertidal</td>
<td>SRR4102394</td>
<td>25613160</td>
<td>16343735</td>
<td>15696090</td>
<td>55728</td>
<td>55391</td>
<td>1240</td>
<td>55213</td>
<td>9310</td>
<td>5547</td>
</tr>
<tr>
<td>Hygrophila</td>
<td>Planorbarius comeus</td>
<td>Freshwater</td>
<td>SRR1185333</td>
<td>28040804</td>
<td>27417761</td>
<td>26569686</td>
<td>86067</td>
<td>82815</td>
<td>2587</td>
<td>78934</td>
<td>39051</td>
<td>16959</td>
</tr>
<tr>
<td>Hygrophila</td>
<td>Biomphalaria glabrata</td>
<td>Freshwater</td>
<td>SRR942795</td>
<td>172317158</td>
<td>170810180</td>
<td>164139487</td>
<td>70439</td>
<td>67493</td>
<td>2070</td>
<td>66689</td>
<td>33757</td>
<td>17015</td>
</tr>
<tr>
<td>Hygrophila</td>
<td>Radix bathica</td>
<td>Freshwater</td>
<td>SRR097739</td>
<td>16923850</td>
<td>NA</td>
<td>NA</td>
<td>54450</td>
<td>54450</td>
<td>679</td>
<td>30883</td>
<td>14614</td>
<td>8140</td>
</tr>
<tr>
<td>Pyramidelloidea</td>
<td>Turbonilla sp.</td>
<td>Marine</td>
<td>SRR1505139</td>
<td>26619896</td>
<td>26219791</td>
<td>25806515</td>
<td>132978</td>
<td>127301</td>
<td>1023</td>
<td>126409</td>
<td>34180</td>
<td>14316</td>
</tr>
<tr>
<td>Stylommatophora</td>
<td>Arion vulgaris</td>
<td>Terrestrial</td>
<td>XXXXXXXX</td>
<td>24874185</td>
<td>19179005</td>
<td>17977774</td>
<td>92316</td>
<td>91984</td>
<td>1454</td>
<td>91417</td>
<td>29208</td>
<td>13134</td>
</tr>
<tr>
<td>Systellommatophora</td>
<td>Onchidella floridana</td>
<td>Intertidal</td>
<td>SRR1505123</td>
<td>14872953</td>
<td>14797399</td>
<td>14528352</td>
<td>79502</td>
<td>74540</td>
<td>1057</td>
<td>74146</td>
<td>28257</td>
<td>13679</td>
</tr>
</tbody>
</table>
Table 2. Examples of ortholog clusters under positive selection in the terrestrial and freshwater lineages. The complete information can be found in the Additional Files 4 and 5.

<table>
<thead>
<tr>
<th>Orthologous cluster</th>
<th>BlastX annotation</th>
<th>Molecular function</th>
<th>Biological process</th>
<th>KEGG pathway</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Terrestrial</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0000060</td>
<td>tyramine beta-</td>
<td>Copper ion binding,</td>
<td>Oxidation-reduction</td>
<td>Tyrosine metabolism</td>
</tr>
<tr>
<td></td>
<td>hydroxylase-like</td>
<td>oxidoreductase</td>
<td>process</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>activity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0000137</td>
<td>sodium glucose</td>
<td>Transmembrane</td>
<td>Transporter activity</td>
<td>Carbohydrate digestion and</td>
</tr>
<tr>
<td></td>
<td>cotransporter 4-like</td>
<td>transport</td>
<td></td>
<td>absorption</td>
</tr>
<tr>
<td>OG0001172</td>
<td>alpha- sarcomeric-</td>
<td>Actin filament</td>
<td>Actin crosslink</td>
<td>Focal adhesion</td>
</tr>
<tr>
<td></td>
<td>like isoform X2</td>
<td>binding, calcium ion</td>
<td>formation, actin</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>binding</td>
<td>filament bundle</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>assembly</td>
<td></td>
</tr>
<tr>
<td><strong>Freshwater</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0000120</td>
<td>cytochrome P450</td>
<td>Monooxygenase</td>
<td>Xenobiotic metabolic</td>
<td>Aminobenzoate degradation,...</td>
</tr>
<tr>
<td></td>
<td>3A7-like</td>
<td>activity, iron ion</td>
<td>process</td>
<td>steroid hormone biosynthesis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>binding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0004116</td>
<td>methylated-DNA--</td>
<td>methylated-DNA-[protein]-cysteine S methyltransferase activity</td>
<td>DNA repair</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>-cysteine</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>methyltransferase-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>like isoform X2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0004174</td>
<td>cytochrome c</td>
<td>Cytochrome c</td>
<td>Proton transport,</td>
<td>Oxidative phosphorylation</td>
</tr>
<tr>
<td></td>
<td>oxidase subunit 4</td>
<td>oxidase activity</td>
<td>mitochondrial electron transport, cytochrome c to oxygen</td>
<td></td>
</tr>
<tr>
<td></td>
<td>isoform</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>mitochondrial-like</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG0002708</td>
<td>40S ribosomal</td>
<td>-</td>
<td>RNA binding, protein</td>
<td>rRNA processing, translation</td>
</tr>
<tr>
<td></td>
<td>protein S3a</td>
<td></td>
<td>binding</td>
<td></td>
</tr>
</tbody>
</table>