bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title

Comparative transcriptomics reveals candidate genes involved in the adaptation to non-marine habitats in panpulmonate mollusks

Authors' names and affiliations

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.

Barbara Feldmeyer[1]

bfeldmeyer@senckenberg.de

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

Markus Pfenninger[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.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 used panpulmonates as a model to illuminate with comparative transcriptomics the selective pressures that shaped 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 to 97,000 predicted open reading frames, of which 9 - 14% were functionally annotated. Adding published transcriptomes, we predicted 791 orthologous 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.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 land and freshwater 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 realms. 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, the largest group of amphibious teleost fishes adapted to live on mudflats, possess unique immune genes to possibly counteract novel pathogens on land, and opsin genes adapted 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 when during the water-to-land transition in tetrapods [9]. Also, vertebrate keratin genes responsible of skin rigidity underwent a

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 to from marine to non-marine habitats in invertebrates is still scarce. Only one study reported adaptation signals in gene families 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]. For example, mutations in the ATPases were suggested to provide the necessary energy to adapt to new high-energy demanding habitats [14]. Similarly, ion-ATPases and cotransporters control the cell homeostasis and can respond to new osmotic pressures during the transition [11]. DNA repair genes could help to reduce the damage produced by increased ultraviolet (UV) irradiation. One of the selected genes in Hexapoda (*Rad23*) is also present in tardigrades, invertebrates well-known for tolerating high levels of radiation [15]. Cytochrome p450s/monooxygenases, glutathione S-transferases, and ABC transporters are involved in the break-down of a multitude of endogenous and exogenous compounds were also found to be enriched in the Hexapoda lineage [11, 16]. Additionally, ribosomal proteins were targets of selection. As the ribosomal machinery is salt-sensitive, the adaptive signs in these gene family could be result of the different osmotic pressures within aquatic and terrestrial environments [17]. Furthermore, heat shock proteins Hsp70 and Hsp90, which are involved in stress response, were also under positive selection in Hexapoda. Hsp70 recognizes and targeted damaged proteins to be recycled in the proteasome [18], and Hsp90 has been shown to provide thermal resistance in terrestrial isopods [19].

In a previous paper, we explored the adaptive signals in the mitochondrial genomes of panpulmonates [20]. 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

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

production probably linked to novel demands of locomotion [21, 22], 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 [23].

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 phylogenetic 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 land and freshwater lifestyles.

Results

We generated approximately 2,100,000 - 3,400,000 reads for five ellobiids (two terrestrial species: *Carychium* sp. and *Pytha pachyodon*) and one stylommatophoran species (the land slug *Arion vulgaris*) (Table 1). The quality trimming eliminated 14 - 39% of short and low-quality fragments in our samples compared to 0.35 - 5.25% in the public data. *De novo* meta assembly in MIRA produced approximately 55,000 – 98,000 transcripts in our samples compared to 54,000 - 130,000 in the other 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 species. We obtained 702 orthologous clusters 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

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 terrestrial and freshwater 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, molecular functions, biological processes 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 that in the terrestrial lineages, the candidate genes 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).

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

Panpulmonates transitioned from marine to freshwater and terrestrial environments in several lineages and multiple times [2, 24, 25], Thus, they were a very suitable model to study the invasion of non-marine realms. However, the phylogenetic relationships within this clade are yet to be resolved [24]. Our tree topology using 382 orthologous clusters resembles the one obtained from Jörger et al. [25], 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 [26]. Still, previous phylogenies using mitochondrial and nuclear markers failed to support this clade [20, 25, 27, 28]. We also found support for Eupulmonata (*sensu* Morton [29]), a clade comprising Stylommatophora and Systellommatophora plus Ellobioidea (intertidal/terrestrial) [24], this clade was supported using a combination of mitochondrial and nuclear markers [25]. 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 definitively 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. On the one hand, 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. On the other hand, the displacement in an environment lacking the buoyancy force to float or swim, would require more energy that can be obtained increasing the glucose uptake (OG0000137) to produce energy in form of ATP in a faster rate compared to the oxidative phosphorylation pathway [30]. These hypotheses are in line with our previous finding of positive selection on the mitochondrial genomes from terrestrial panpulmonates. The adaptive signatures found in two mitochondrial genes, *cob* and *nad5*, involved in energy production in the

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mitochondrion, 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 [31]. Notably, it has been suggested that iodinated tyrosines may have been essential in vertebrates during the transition to non-marine habitats for TH's are required in the expression of transcription factors involved in the embryonic development and differentiation of the lungs [32]. 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 [33]. In agreement with our previous results that found the mitochondrial *cob* subunit under positive selection [20], 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

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [17]. 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 [34]. 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 [35]. 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 [36]

Conclusions

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 non-marine realms. 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. [37] was used as a starting point for our study. We added to this dataset the transcriptome from *Radix balthica*. [38] and retrieved additional freshwater specimens from the NCBI Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra). All species were sequenced using the Illumina platform. Also, 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 HiSeq2000 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 [39] was used for initial assessment of reads quality. Then, Trimmomatic v0.33 [40] was used to remove and trim Illumina adaptor sequences and other reads with an average quality below

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 [38].

Transcriptome assembly

De novo assembly was performed for all samples, except *Radix* (see last section), using Trinity v2.0.6 [41] with a minimum contig length of 100 amino acids, and Bridger v2014-12-01 [42] 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 [43] 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 [44]. Furthermore, we used the ORFpredictor server [45] to predict open reading frames (ORF) within the transcripts.

Construction of orthologous clusters

Orthologous clusters shared among protein sequences of the fifteen panpulmonate species were predicted using OrthoFinder [46] 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 orthologous cluster were aligned using MAFFT [47] with standard parameters. Nucleotide sequences in each orthogroup were aligned codon-wise using TranslatorX [48] taking into account the information from the amino acid alignments. Ambiguous aligned regions from the amino acid or codon alignments were removed using Gblocks [49] with standard settings. We used TrimAL [50] to remove poorly aligned or incomplete sequences in each orthologous cluster, using a minimum residue overlap score of 0.75. These trimmed alignments were also checked by eye.

Phylogenetic analyses

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Phylogenetic relationships among the Panpulmonata were reconstructed based on a subset of 382 orthologous clusters. The subset selection was done using MARE [51], 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 [52] using the *-rcluster* option (relaxed hierarchical clustering algorithm), suitable for phylogenomic data [53]. We reconstructed an unrooted tree to be used as an input in the selection analyses. Maximum likelihood analyses were conducted in RAxML-HPC2 (8.0.9) [54]. 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 [55]. Four 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 [56] 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 orthologous clusters (codon-wise alignments) in CODEML implemented in the software PAML v4.8 [57]. PAML estimated the omega ratio (ω = dN: non-synonymous sites / dS: synonymous sites); $\omega = 1$ indicates neutral evolution, $\omega < 1$ purifying selection, and $\omega > 1$ indicates positive selection [58]. 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 [59] in CODEML was applied (model = 2, NSsites = 2) for each orthologous cluster. The unrooted tree obtained using maximum likelihhod 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

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

compared in a likelihood ratio test (LRT= 2*(lnL model A – lnL 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 [60] 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*, or using a randomization procedure with 1000 replicates implemented in a custom Python script.

Functional annotation

The transcripts were annotated using BlastX [61]. 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 (OG) terms for each BLASTx search were obtained in the Blast2GO suite [62]. Functional annotation information was obtained from InterPro database [63] using the InterProScan [64]. 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 [65]. This server assigns orthology identifiers from the KEGG database (Kyoto Encyclopedia of Genes and Genomes).

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 (PRJNAXXX), both 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.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Kameda Y, Kato M: **Terrestrial invasion of pomatiopsid gastropods in the heavy-snow region of the Japanese Archipelago**. *BMC evolutionary biology* 2011, **11**:118.
2. Romero PE, Pfenninger M, Kano Y, Klussmann-Kolb A: **Molecular phylogeny of the Ellobiidae (Gastropoda: Panpulmonata) supports independent terrestrial invasions**. *Molecular phylogenetics and evolution* 2016, **97**:43-54.
3. Kocot KM, Halanych KM, Krug PJ: **Phylogenomics supports Panpulmonata: opisthobranch paraphyly and key evolutionary steps in a major radiation of gastropod molluscs**. *Molecular phylogenetics and evolution* 2013, **69**(3):764-771.
4. Little C: **The terrestrial invasion: An ecophysiological approach to the origins of land animals**. Cambridge: Cambridge University Press; 1990.
5. Laurin M: **How vertebrates left water**. Berkeley: University of California Press; 2010.
6. McGhee G: **When the invasion of land failed: The legacy of the Devonian extinctions**. New York: Columbia University Press; 2013.
7. You X, Bian C, Zan Q, Xu X, Liu X, Chen J, Wang J, Qiu Y, Li W, Zhang X *et al*: **Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes**. *Nature communications* 2014, **5**:5594.
8. Amemiya CT, Alfoldi J, Lee AP, Fan S, Philippe H, Maccallum I, Braasch I, Manousaki T, Schneider I, Rohner N *et al*: **The African coelacanth genome provides insights into tetrapod evolution**. *Nature* 2013, **496**(7445):311-316.
9. Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, Suzuki H, Okuno M, Aibara M, Ngatunga BP *et al*: **Coelacanth genomes reveal signatures for evolutionary transition from water to land**. *Genome research* 2013, **23**(10):1740-1748.
10. Vandebergh W, Bossuyt F: **Radiation and functional diversification of alpha keratins during early vertebrate evolution**. *Molecular biology and evolution* 2012, **29**(3):995-1004.
11. Faddeeva A, Studer RA, Kraaijeveld K, Sie D, Ylstra B, Marien J, op den Camp HJ, Datema E, den Dunnen JT, van Straalen NM *et al*: **Collembolan transcriptomes highlight molecular evolution of hexapods and provide clues on the adaptation to terrestrial life**. *PloS one* 2015, **10**(6):e0130600.
12. Misof B, Liu S, Meusemann K, Peters RS, Donath A, Mayer C, Frandsen PB, Ware J, Flouri T, Beutel RG *et al*: **Phylogenomics resolves the timing and pattern of insect evolution**. *Science* 2014, **346**(6210):763-767.
13. von Reumont BM, Jenner RA, Wills MA, Dell'ampio E, Pass G, Ebersberger I, Meyer B, Koenemann S, Iliffe TM, Stamatakis A *et al*: **Pancrustacean phylogeny in the light of new phylogenomic data: support for Remipedia as the possible sister group of Hexapoda**. *Molecular biology and evolution* 2012, **29**(3):1031-1045.
14. Ge RL, Cai Q, Shen YY, San A, Ma L, Zhang Y, Yi X, Chen Y, Yang L, Huang Y *et al*: **Draft genome sequence of the Tibetan antelope**. *Nature communications* 2013, **4**:1858.
15. Forster F, Beisser D, Grohme MA, Liang C, Mali B, Siegl AM, Engelmann JC, Shkumatov AV, Schokraie E, Muller T *et al*: **Transcriptome analysis in tardigrade species reveals specific molecular pathways for stress adaptations**. *Bioinformatics and biology insights* 2012, **6**:69-96.
16. Ranson H, Claudianos C, Ortelli F, Abgrall C, Hemingway J, Sharakhova MV, Unger MF, Collins FH, Feyereisen R: **Evolution of supergene families associated with insecticide resistance**. *Science* 2002, **298**(5591):179-181.
17. Wissler L, Codoner FM, Gu J, Reusch TB, Olsen JL, Procaccini G, Bornberg-Bauer E: **Back to the sea twice: identifying candidate plant genes for molecular evolution to marine life**. *BMC evolutionary biology* 2011, **11**:8.
18. Feder ME, Hofmann GE: **Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology**. *Annual review of physiology* 1999, **61**:243-282.
19. Gusev O, Ziegler TA, Saigusa M: **Expression and structure of stress chaperon hsp90 in terrestrial decapods, Coenobita (Anomura: Coenobitidae) and Chiromantes (Brachyura: Sesarmidae)**. *Crustacean Research* 2006, **6**:103-113.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Romero PE, Weigand AM, Pfenninger M: **Positive selection on panpulmonate mitogenomes provide new clues on adaptations to terrestrial life**. *BMC Evol Bio* 2016.
21. Shen YY, Liang L, Zhu ZH, Zhou WP, Irwin DM, Zhang YP: **Adaptive evolution of energy metabolism genes and the origin of flight in bats**. *Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**(19):8666-8671.
22. McClellan DA, Palfreyman EJ, Smith MJ, Moss JL, Christensen RG, Sailsbery JK: **Physicochemical evolution and molecular adaptation of the cetacean and artiodactyl cytochrome b proteins**. *Molecular biology and evolution* 2005, **22**(3):437-455.
23. Mittler R: **Oxidative stress, antioxidants and stress tolerance**. *Trends Plant Sci* 2002, **7**(9):405-410.
24. Schrödl M: **Time to say "Bye-bye Pulmonata"?** *Spixiana* 2014, **37**(2):161-164.
25. Jorger KM, Stoger I, Kano Y, Fukuda H, Knebelsberger T, Schrodl M: **On the origin of Acochlidia and other enigmatic euthyneuran gastropods, with implications for the systematics of Heterobranchia**. *BMC evolutionary biology* 2010, **10**:323.
26. Dayrat B, Tillier S: **Evolutionary relationships of euthyneuran gastropods (Mollusca): a cladistic re-evaluation of morphological characters**. *Zoological Journal of the Linnean Society* 2002, **135**(4):403-470.
27. White TR, Conrad MM, Tseng R, Balayan S, Golding R, de Frias Martins AM, Dayrat BA: **Ten new complete mitochondrial genomes of pulmonates (Mollusca: Gastropoda) and their impact on phylogenetic relationships**. *BMC evolutionary biology* 2011, **11**:295.
28. Dayrat B, Conrad M, Balayan S, White TR, Albrecht C, Golding R, Gomes SR, Harasewych MG, de Frias Martins AM: **Phylogenetic relationships and evolution of pulmonate gastropods (Mollusca): New insights from increased taxon sampling**. *Molecular phylogenetics and evolution* 2011, **59**(2):425-437.
29. Morton JE: **The evolution of the Ellobiidae with a discussion on the origin of the Pulmonata**. *Proceedings of the Zoological Society of London* 1955, **125**(1):127-168.
30. Lunt SY, Vander Heiden MG: **Aerobic glycolysis: meeting the metabolic requirements of cell proliferation**. *Annual review of cell and developmental biology* 2011, **27**:441-464.
31. Heyland A, Price DA, Bodnarova-Buganova M, Moroz LL: **Thyroid hormone metabolism and peroxidase function in two non-chordate animals**. *Journal of experimental zoology Part B, Molecular and developmental evolution* 2006, **306**(6):551-566.
32. Crockford SJ: **Evolutionary roots of iodine and thyroid hormones in cell-cell signaling**. *Integrative and comparative biology* 2009, **49**(2):155-166.
33. Li Y, Park JS, Deng JH, Bai Y: **Cytochrome c oxidase subunit IV is essential for assembly and respiratory function of the enzyme complex**. *Journal of bioenergetics and biomembranes* 2006, **38**(5-6):283-291.
34. Lee J, Rhee K-J, Kim SH, Kang YW, Cho Y, Yang S-J, Pan C-H, Kim YS: **O6-Methylguanine-DNA methyltransferase (MGMT) gene expression is associated with ultraviolet B (UVB)-induced cell growth inhibition and recovery**. *Genes & Genomics* 2015, **37**(9):789-796.
35. Bancroft BA, Baker NJ, Blaustein AR: **Effects of UVB radiation on marine and freshwater organisms: a synthesis through meta-analysis**. *Ecology letters* 2007, **10**(4):332-345.
36. Olson MH, Barbieri NE: **Mechanisms of ultraviolet radiation tolerance in the freshwater snailPhysa acuta**. *Freshwater Science* 2014, **33**(1):66-72.
37. Zapata F, Wilson NG, Howison M, Andrade SC, Jorger KM, Schrodl M, Goetz FE, Giribet G, Dunn CW: **Phylogenomic analyses of deep gastropod relationships reject Orthogastropoda**. *Proceedings Biological sciences / The Royal Society* 2014, **281**(1794):20141739.
38. Feldmeyer B, Wheat CW, Krezdorn N, Rotter B, Pfenninger M: **Short read Illumina data for the de novo assembly of a non-model snail species transcriptome (Radix balthica, Basommatophora, Pulmonata), and a comparison of assembler performance**. *BMC genomics* 2011, **12**:317.
39. **FastQC** [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]
40. Bolger AM, Lohse M, Usadel B: **Trimmomatic: a flexible trimmer for Illumina sequence data**. *Bioinformatics* 2014, **30**(15):2114-2120.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

41. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q *et al*: **Full-length transcriptome assembly from RNA-Seq data without a reference genome**. *Nature biotechnology* 2011, **29**(7):644-652.
42. Chang Z, Li G, Liu J, Zhang Y, Ashby C, Liu D, Cramer CL, Huang X: **Bridger: a new framework for de novo transcriptome assembly using RNA-seq data**. *Genome biology* 2015, **16**(1):30.
43. Chevreux B, Wetter T, Suhai S: **Genome Sequence Assembly Using Trace Signals and Additional Sequence Information** In: *Computer Science and Biology: Proceedings of the German Conference on Bioinformatics (GCB).* 1999: 45-56.
44. Zhao L, Zhang N, Ma PF, Liu Q, Li DZ, Guo ZH: **Phylogenomic analyses of nuclear genes reveal the evolutionary relationships within the BEP clade and the evidence of positive selection in Poaceae**. *PloS one* 2013, **8**(5):e64642.
45. Min XJ, Butler G, Storms R, Tsang A: **OrfPredictor: predicting protein-coding regions in EST-derived sequences**. *Nucleic acids research* 2005, **33**:W677-680.
46. Emms DM, Kelly S: **OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy**. *Genome biology* 2015, **16**:157.
47. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability**. *Molecular biology and evolution* 2013, **30**(4):772-780.
48. Abascal F, Zardoya R, Telford MJ: **TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations**. *Nucleic acids research* 2010, **38**(Web Server issue):W7-13.
49. Talavera G, Castresana J: **Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments**. *Systematic biology* 2007, **56**(4):564-577.
50. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T: **trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses**. *Bioinformatics* 2009, **25**(15):1972-1973.
51. Misof B, Meyer B, von Reumont BM, Kuck P, Misof K, Meusemann K: **Selecting informative subsets of sparse supermatrices increases the chance to find correct trees**. *BMC Bioinformatics* 2013, **14**:348.
52. Lanfear R, Calcott B, Ho SY, Guindon S: **Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses**. *Molecular biology and evolution* 2012, **29**(6):1695-1701.
53. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A: **Selecting optimal partitioning schemes for phylogenomic datasets**. *BMC evolutionary biology* 2014, **14**:82.
54. Stamatakis A: **RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models**. *Bioinformatics* 2006, **22**(21):2688-2690.
55. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Hohna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP: **MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space**. *Systematic biology* 2012, **61**(3):539-542.
56. Rambaut A, Suchard MA, Xie D, Drummond AJ: **Tracer v1.6**; 2014.
57. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood**. *Molecular biology and evolution* 2007, **24**(8):1586-1591.
58. Yang ZH, Nielsen R, Goldman N, Pedersen AMK: **Codon-substitution models for heterogeneous selection pressure at amino acid sites**. *Genetics* 2000, **155**(1):431-449.
59. Zhang J, Nielsen R, Yang Z: **Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level**. *Molecular biology and evolution* 2005, **22**(12):2472-2479.
60. Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing**. *J Roy Stat Soc B Met* 1995, **57**(1):289-300.
61. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs**. *Nucleic acids research* 1997, **25**(17):3389-3402.
62. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A: **High-throughput functional annotation and data mining with the Blast2GO suite**. *Nucleic acids research* 2008, **36**(10):3420-3435.

bioRxiv preprint doi: https://doi.org/10.1101/072389; this version posted August 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

63. Mitchell A, Chang HY, Daugherty L, Fraser M, Hunter S, Lopez R, McAnulla C, McMenamin C, Nuka G, Pesseat S *et al*: **The InterPro protein families database: the classification resource after 15 years**. *Nucleic acids research* 2015, **43**(Database issue):D213-221.

64. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G *et al*: **InterProScan 5: genome-scale protein function classification**. *Bioinformatics* 2014, **30**(9):1236-1240.

65. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M: **KAAS: an automatic genome annotation and pathway reconstruction server**. *Nucleic acids research* 2007, **35**(Web Server issue):W182-185.

Tables titles and legends

Table 1. Descriptive statistics of the assembled transcriptomes. *De novo* assemblies of new transcriptomes from terrestrial panpulmonates are highlighted in bold.

				Trimmed reads		**MIRA meta-assembly**				**BlastX**	
Clade	**Species**	**Accession Number**	**Number of raw reads**	**For Trinity**	**For Bridger**	**All transcripts**	**Transcripts > 300 (bp)**	**N50 (bp)**	**Predicted ORF's**	**Single hits**	**Single genes**
Acochlidia	*Microhedyle glandulifera*	SRR1505118	6194970	6103287	6028350	129453	84468	742	83983	28750	12772
Acochlidia	*Strubellia wawrai*	SRR1505137	24132673	24049385	23436737	82681	79947	1582	79735	27911	13735
Amphiboiloidea	*Phallomedusa solida*	SRR1505127	25685273	25496722	24822510	68633	65500	1424	64978	24560	12722
Ellobioidea	***Carychium* sp.**	**XXXXXXXXX**	**33608344**	**23461211**	**23461211**	**87994**	**87719**	**2035**	**87242**	**23586**	**11672**
Ellobioidea	***Cassidula plecotremata***	**XXXXXXXXX**	**26316221**	**18928297**	**17989294**	**62318**	**62031**	**1533**	**61569**	**9583**	**5717**
Ellobioidea	***Melampus flavus***	**XXXXXXXXX**	**21068142**	**16919851**	**16052613**	**97629**	**97334**	**1640**	**96816**	**22187**	**10778**
Ellobioidea	*Ophicardelus sulcatus*	SRR1505124	16026272	15737826	15425623	74467	71073	1499	70573	20851	10937

Ellobioidea	***Pythia pachyodon***	**XXXXXXXXX**	**24016251**	**20555474**	**19399330**	**92548**	**92264**	**2035**	**91759**	**24077**	**11776**
Ellobioidea	***Trimusculus sp.***	**XXXXXXXXX**	**25613160**	**16343735**	**15696090**	**55728**	**55391**	**1240**	**55213**	**9310**	**5547**
Hygrophila	*Planorbarius corneus*	SRR1185333	28040804	27417761	26569686	86067	82815	2587	78934	39051	16959
Hygrophila	*Biomphalaria glabrata*	SRR942795	172317158	170810180	164139487	70439	67493	2070	66689	33757	17015
Hygrophila	*Radix balthica*	SRP005151	16923850	NA	NA	54450	54450	679	30883	14614	8140
Pyramidelloidea	*Turbonilla* sp.	SRR1505139	26619896	26219791	25806515	132978	127301	1023	126409	34180	14316
Stylommatophora	***Arion vulgaris***	**XXXXXXXXX**	**24874185**	**19179005**	**17977774**	**92316**	**91984**	**1454**	**91417**	**29208**	**13134**
Systellommatophora	*Onchidella floridana*	SRR1505123	14872953	14797399	14528352	79502	74540	1057	74146	28257	13679

Table 2. Examples of orthologous clusters under positive selection in the terrestrial and freshwater lineages. The complete information can be found in the Additional Files 4 and 5.

Orthologous cluster	**BlastX annotation**	**Molecular function**	**Biological process**	**KEGG pathway**
Terrestrial				
OG0000060	tyramine beta-hydroxylase-like	Copper ion binding, oxidoreductase activity	Oxidation-reduction process	Tyrosine metabolism
OG0000137	sodium glucose cotransporter 4-like	Transmembrane transport	Transporter activity	Carbohydrate digestion and absorption
OG0001172	alpha- sarcomeric-like isoform X2	Actin filament binding, calcium ion binding	Actin crosslink formation, actin filament bundle assembly	Focal adhesion
Freshwater				
OG0000120	cytochrome P450 3A7-like	Monooxygenase activity, iron ion binding	Xenobiotic metabolic process	Aminobenzoate degradation, steroid hormone biosynthesis

OG0004116	methylated-DNA-- -cysteine methyltransferase-like isoform X2	methylated-DNA-[protein]-cysteine S methyltransferase activity	DNA repair	-
OG0004174	cytochrome c oxidase subunit 4 isoform mitochondrial-like	Cytochrome c oxidase activity	Proton transport, mitochondrial electron transport, cytochrome c to oxygen	Oxidative phosphorylation
OG0002708	40S ribosomal protein S3a	-	RNA binding, protein binding	rRNA processing, translation

615