Received Date : 30-Jul-2016
Revised Date : 11-Oct-2016
Accepted Date : 12-Oct-2016
Article type : Original Article

Transcriptome-based phylogeny of endemic Lake Baikal amphipod species flock: fast speciation accompanied by frequent episodes of positive selection

Sergey A. Naumenko[1,2,3], Maria D. Logacheva[1,2,4], Nina V. Popova[1], Anna V. Klepikova[1,2], Aleksey A. Penin[1,2], Georgii A. Bazykin[1,2,4,5,6], Anna E. Etingova[7], Nikolai S. Mugue[8,9], Alexey S. Kondrashov[1,10] and Lev Y. Yampolsky[11]

[1]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
[2]Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, Moscow, Russia
[3]Genetics and Genome Biology Program, The Hospital For Sick Children, Toronto, ON, M5G 0A4, Canada
[4]Pirogov Russian National Research Medical University, Moscow, Russia
[5]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow, Russia
[6]Skolkovo Institute of Science and Technology, Skolkovo 143026, Russia
[7]Baikal Museum, Irkutsk Research Center, Russian Academy of Sciences, Listvyanka, Irkutsk region, Russia
[8]Laboratory of Molecular Genetics, Russian Institute for Fisheries and Oceanography (VNIRO), Moscow, Russian Federation;

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/mec.13927

This article is protected by copyright. All rights reserved

Author Manuscript

[9]Laboratory of Experimental Embryology, Koltzov Institute of Developmental Biology, Moscow, Russia
[10]Department of Ecology and Evolution, University of Michigan, Ann Arbor, MI, USA
[11]Department of Biological Sciences, East Tennessee State University, Johnson City, TN, USA

Corresponding authors:
SAN sergey.naumenko@yahoo.com and LYY yampolsk@etsu.edu

Key words not occurring in the title
adaptive radiation, ancient lakes, ecological opportunity, directional selection

Authors contributions: ASK and LYY designed the study; SAN, MDL, AEE, NSM and LYY participated in field collections and species identification; MDL, AVK and AAP did library preparation and next-gen sequencing; SAN, NVP and GAB participated in assembly, alignments orthologization and phylogenetics analysis; SAN, GAB, ASK and LYY participated in downstream analysis and manuscript preparation.

Abstract

Endemic species flocks inhabiting ancient lakes, oceanic islands and other long-lived isolated habitats are often interpreted as adaptive radiations. Yet molecular evidence for directional selection during species flocks radiation is scarce. Using partial transcriptomes of 64 species of Lake Baikal (Siberia, Russia) endemic amphipods and

This article is protected by copyright. All rights reserved

Author Manuscript

two non-endemic outgroups, we report a revised phylogeny of this species flock, and analyze evidence for positive selection within the endemic lineages. We confirm two independent invasions of amphipods into Baikal and demonstrate that several morphological features of Baikal amphipods, such as body armor and reduction of appendages and sensory organs, evolved in several lineages in parallel. Radiation of Baikal amphipods has been characterized by short phylogenetic branches and frequent episodes of positive selection which tended to be more frequent in the early phase of the second invasion of amphipods into Baikal when the most intensive diversification occurred. Notably, signatures of positive selection are frequent in genes encoding mitochondrial membrane proteins with electron transfer chain and ATP synthesis functionality. In particular, subunits of both the membrane and substrate-level ATP synthases show evidence of positive selection in the plankton species *Macrohectopus branickii*, possibly indicating adaptation to active plankton lifestyle and to survival under conditions of low temperature and high hydrostatic pressures known to affect membranes functioning. Other functional categories represented among genes likely to be under positive selection include Ca-binding muscle-related proteins, possibly indicating adaptation to Ca-deficient low mineralization Baikal waters.

Introduction

Large long lived isolated habitats – continents, islands and lakes alike – have long attracted the interest of evolutionary biologists because of the independent evolutionary diversification organisms endemic to such habitats undergo in isolation from the rest of Earth's biota (Simpson 1980; Schluter 2000). In particular, ancient lakes offer a number of spectacular examples of endemic radiations resulting in highly morphologically and ecologically diverse species flocks, in a striking contrast to much lower faunal diversity in short-lived lakes (Brooks, 1950; Martens, 1997) and often with a fascinating evidence of intrinsic nature of the endemic diversification (Cristescu et al., 2010). Frequently several independent invasions of a particular lineage into an ancient lake is followed by rapid diversification of endemic lineages (Martens 1997; Cristescu et al. 2010). The

Author Manuscript

This article is protected by copyright. All rights reserved

endemic species flocks of cichlid fishes from lakes Tanganyika, Malawi and Victoria in the African Rift valley (Meyer 1993; Allender et al. 2003; Kocher 2004; Keller et al. 2014; Fan and Meyer 2014; Brawnard et al. 2014) and of amphipod crustaceans in Lake Baikal in Siberia, Russia (Sherbakov et al. 1998; Kamaltynov 1999; Takhteev 2000ab; MacDonald et al. 2005) are among the most species rich and best studied. However, other examples are diverse and abundant (Martens 1997; Cristescu et al. 2010) and include Lake Titicaca amphipods (Pyle and Adamowicz 2015) and gastropods (Kroll et al. 2012), Lake Ohrid amphipods and gastropods (Albrecht et al. 2006; Wysocka et al. 2014; Foeller et al., 2015), Lakes Tanganiyka and Malawi gastropods (Genner et al., 2007; Van Damme and Gautier 2013), Lake Tanganyika crabs (Marijnissen et al. 2006), shrimps (von Rintelen et al. 2010) and bivalves (von Rintelen and Glaubrecht 2006) from several lakes in Sulawesi, as well as several species flocks present in Lake Baikal (Sherbakov 1999), such as sponges (Meixner et al. 2007), three independent gastropod radiations (Sitnikova 1994 2006; Hausdorf et al. 2003), planarians (Novikova et al. 2006) and sculpins (Hunt et al. 1997), among others. Several common features have been identified that are shared by ancient lakes species flocks (as well as by similar radiations in terrestrial, often insular, habitats, such as those of Galapagos Darwin's finches, Hawaiian Drosophila, Hawaiian silverswords, Hawaiian honeycreepers or Greater Antilles anoles). These common features include the following: (i) unusually high levels of ecological specialization, including feeding modes and habitat use, typically far exceeding the level of specialization in generalist non-endemic sister groups; (ii) high level of morphological diversity often uncharacteristic for non-endemic relatives; and (iii) rapid speciation allowing "explosive" radiation on the background of moderate sequence evolution (Martens 1997; Fryer 1991; Schluter 2000; Kocher 2004; Cristescu et al., 2010). Such common features shared by endemic species flocks despite their radically different taxonomy and ecology, certainly require common explanations and provide a unique opportunity to test hypotheses about evolutionary scenarios repeatedly playing out in different organisms and in different environments.

Adaptive radiations are often understood as the result of ecological opportunity (Schluter 2000; Yoder et al. 2010; Wellborn and Langerhans 2015; Berner and Salzburger 2015) which leads to accumulation of species richness by setting stage for

This article is protected by copyright. All rights reserved

diversifying selection. Ecological opportunity presents itself either due to availability of unoccupied niches in a newly colonized environment and abating as niche packing increases with the number of extant lineages (Bailey et al. 2013; Cornell 2013; Martin and Wainwright 2013; Martin and Feinstein 2014) or through morphological or ecological innovations (van Rijssel et al. 2015) that may occur relatively late in the evolutionary history of a radiation, thus possibly increasing the diversification rate with the number of lineages in a radiation. The early vs. late possibility of ecological opportunity are certainly not mutually exclusive, which makes it difficult to formulate hypotheses about rates of diversification and importance of positive selection in early vs. late phases of adaptive radiations. Furthermore, it has been argued (Losos 2010) that the idea of ecological opportunity can be elusive. Indeed, it is extremely difficult to pinpoint particular unoccupied niches or particular innovations that allow the radiation of hundreds of species of amphipods in Lake Baikal or cichlids in Great African Rift Lakes.

The gammarid species flock (Crustacea: Amphipoda) of Lake Baikal certainly displays the common features of adaptive radiations at the extreme. This assemblage inhabits the oldest (up to 30 my, with the geological history of the rift depression dating back up to 70 my, Mats et al. 2011), the deepest and the most voluminous lake on Earth, stretching over 500 km in S-N direction in East Siberia, featuring unusually cold, oligotrophic waters, with mineralization among the lowest of large lakes (<0.1 g/l), fully oxygenated depths all the way to over 1600 m and a 1-2 m thick ice cover for 5 months a year (Kozhova and Izmest'eva 1998). Baikalian amphipod species flock is the most species rich of all Baikal radiations, with over 340 described species (Bazikalova 1945; Kamaltynov 1999) and possibly many more yet undescribed cryptic species (Väinölä and Kamaltynov 1999; Danelia et al. 2011). This assemblage includes both unusually small (2-3 mm adult females in several *Micruropus* spp) and unusually large (Takhteev 2000b; Danelia et al. 2011) (up to 90 mm in several species within the genera *Acanthogammarus* and *Garjajewia*; Figure 1, see also cover image of this issue). Ecologically specialized morphs include (Takhteev 2000b) obligatory fossorial detritus-feeding species, obligate scavengers, active predators, at least two specialized abyssal clades (Sherbakov et al. 1998), body cavities parasites of other amphipods (Takhteev 2000b), specialized commensals of sponges and the world's only fully planktonic freshwater amphipod

This article is protected by copyright. All rights reserved

species, *Macrohectopus branickii* (Figure 1), complete with the entire suite of planktonic adaptations, such as vertical migrations and predation on zooplankton (Melnik et al. 1993). A previous molecular phylogeny study (McDonald et al. 2005) suggested two independent amphipods invasions into Baikal, both cladistically within the family Gammaridae. Furthermore, this phylogeny suggested that *Acanthogammaridae s. lat.*, the more ecologically and morphologically divergent and diverse of the two Baikal clades, originated within the morphologically conservative, ecologically generalistic and biogeographically palaoarctic genus *Gammarus*.

Despite strong ecological diversification among baikalian amphipods, it is difficult to explain the extant taxonomic diversity by niche availability alone, as numerous species still occupy very similar niches and are probably resource-limited, particularly considering the oligotrophic ecology of Baikal waters characterized by mean temperatures and mineralization that are among the lowest among all freshwater lakes. Indeed, stable isotope data (Yoshii 1999) indicate a broad overlap of trophic niches of amphipods from different genera within the amphipod radiation as well as with other benthic invertebrates, in particular in deep water habitats, where the complexity of food chains appears to be low. Likewise, it is difficult to specify particular morphological or ecological innovations that contributed to the observed species richness. Certainly giant forms, brood chamber parasites and the plankton species *M. branickii* have acquired very significant ecological and morphological innovations, but each such innovation led to formation of just a handful of species (the most significant ones in the case of *M. branickii* – to just one species), while the most species rich genera like *Micruropus* and *Eulimnogammarus* retain conservative morphology and ecology. It is therefore unclear whether the species diversity in Baikal is bounded (sensu Cornell 2013) and if yes, whether it has approached its limits, resulting in slowing of the diversification process (and also preventing the cosmopolitan *Gammarus lacistris* from establishing in Baikal).

Application of genome- and transcriptome-wide approaches to the study of adaptive radiations has started very recently, but has already shown promise in discovery of both genomic and ecological mechanisms of diversification (Berner and Salzburger 2015). Such studies have revealed complex colonization and hybridization history and led to conclusions about a possible role of gene duplications, transposable elements and

Author Manuscript

This article is protected by copyright. All rights reserved

positive selection in speciation and divergence (Kocher 2004; Keller et al. 2014; Fan and Meyer 2014; Brawnard et al. 2014; Franchini et al. 2014; Wiess et al. 2015). Yet, genomics studies of adaptive radiations have been so far largely limited to African Rift Lakes cichlids species flocks.

Here, we report the first ever non-cichlid endemic radiation transcriptomics dataset representing about 20% of described species diversity in the species flock and a significant fraction of adult-expressed open reading frames. Species were selected for sequencing, whenever possible, with the aim to increase the coverage of presumed clades at the expense of reducing taxonomic coverage within clades (see Materials and Methods).

We confirm, with a high support, the previous hypothesis about two independent radiations of gammaridean amphipods in Lake Baikal (Fig. 1), largely (with minor exceptions) corresponding to previously described families *Micruropidae* and *Acanthogammaridae* (Kamaltynov 1999). Using this robust phylogeny, we test the hypothesis that the striking feature of baikalian amphipods, namely strong body armor in the form of keels and spines, which is common in marine amphipods but not commonly observed in freshwater species, has evolved repeatedly in parallel within different baikalian clades. We also analyze other morphological adaptations: reduction of urosome and uropods and losses of sensory organs on antennae – accessory flagellum and calceoli, and show that these reductions too have occurred in parallel in several lineages.

We further utilize a detailed phylogeny with strong support for short internal branches to test the hypothesis that the diversification of Baikal amphipods occurred quickly shortly after the invasion of ancestral forms into Baikal, consistent with the idea of ecological opportunity (Wellborn and Langerhans 2015; Berner and Salzburger 2015) presented by a large number of unoccupied ecological niches and resulting in an early burst of speciation (Martin and Wainwright 2013). We also test the frequently made but infrequently tested claim that adaptive radiation by invasion into new niches and rapid accumulation of morphologically and ecologically distinct lineages was accompanied by frequent episodes of positive (diversifying) selection.

Results

Author Manuscript

This article is protected by copyright. All rights reserved

Phylogeny, parallel evolution and diversification rate

Detailed sampling, sequencing, alignment and mapping data are available in Supplementary Tables 1-6. Figure 1 shows the phylogeny of Lake Baikal amphipods based on 175 clusters of orthologous genes (COGs, with exactly one ortholog represented in each species) of all 64 Baikal species and two outgroup species obtained from whole-transcriptome datasets. Bootstrap support is shown only when it is less than 100%. Phylogenies based on a larger number of COGs (Supplementary Table 7) identified in all species within each of five color-coded clades on Fig. 1 are shown on Supplementary Figures 1 and 2. Despite very short branches in early radiation, most nodes are well supported both on the whole species flock and clade-specific phylogenies, with only a few nodes supported by <100% of bootstrap replicates. Still, requiring the presence of a gene transcript from each COG in each species results in a quick reduction in the number of COGs as the number of analyzed species increases (Supplementary Figure 3A), potentially lowering the robustness of phylogenetic inferences. Yet, the clade-specific phylogenies also had high bootstrap support and the phylogenetic positions of 64 out of the 66 analyzed species were the same as in the whole-data 66-species phylogeny, implying that our transcriptome-based phylogenetic reconstruction is robust.

The phylogeny strongly supports previously hypothesized (McDonald et al. 2005; Hou and Sket 2016) two independent invasions of amphipods into Lake Baikal. The more recent, more species rich and morphologically more diverse *Acanthogammaridae* clade is more closely related to the palaearctic *Gammarus lacustris* than to the older, morphologically less diverse *Micruropidae* clade (bootstrap support of the two sister clades 98% and 100%, respectively, Figure 1). This divergence, as well as the divergence of major clades within *Acanthogammaridae,* largely confirms the traditional morphology-based taxonomy, with several exceptions. Notably, the trademark morphological feature of Baikal amphipods – spines and keels on meso- and metasomal segments – has developed several times in parallel (black diamonds in Figure 1). The phylogeny strongly supports at least 4 independent origins (3 within *Acanthogammaridae* and 1 within *Micruropidae*) and 2 reversals to non-spiny forms, or 5 independent origins and 2

Author Manuscript

This article is protected by copyright. All rights reserved

independent losses of body armor. Note that the 2 cases of body armor loss within the *Pallasea* clade (Fig. 1, green), in *Heterogammarus sophianosii* and *Poikilogammarus spp.*,are not only reliably placed into this clade by molecular phylogeny, but also share with the genus *Pallasea* the equiramous uropod 3 as well as the coloration pattern in life animals. A possible 6th episode of spines development is observed in *Gmelinoides fasciatus (Micruropidae* clade), which possesses a row of very mild lateral bumps. It should be noted that the nature of body armor is different in these 5 clades: *Acanthogammaridae s.str*. possess spines in either or all of three rows – dorsal, median and lateral; *Odonthogammarus* is characterized by only lateral, downward pointing spines; *Eucarinogammarus* has only dorsal keels and no spines; and, finally, members of the *Micruropidae* clade have either blunt dorsal keels (*Carinogammarus)* or a single long and sharp dorsal backward pointing spine (*Macrohectopus)* (Fig. 1; Bazikalova, 1945). Similar morphologies featuring meso- and metasomal spines and keels have also repeatedly evolved outside of the Baikal within other families of amphipods in the Caspian Sea and in the oceans (Takhteev 2000a; Hou and Sket 2016).

Supplementary Figures 2A-C show other parallel changes in morphology that occurred independently in various baikalian clades. In particular, the uropod 3 (the last appendage typical for the amphipod body plan that is used for springing motion, an important taxonomic character in the family *Gammaridae* (Hou and Sket 2016) is strongly reduced in 3 independent lineages, sometimes to a tiny bump as in some *Hyalellopsis* species, and shows various degrees of reduction of its inner ramus in at least 3 other lineages (Supplementary Figure 2A). Another widespread change occurring in parallel across both major clades is the loss of sensory organs – the accessory flagellum on antenna 1 and calceoli (believed to be motion sensors) on antenna 2 (Supplementary Figures 2B,C). Accessory flagellum has been reduced to 1-2 segments (often 1 regular size and 1 rudimentary segment) in 7 independent lineages (with 2 apparent reversals in *Carinogammarus* and the genus comprised of deep-water scavenges *Ommatogammarus*). The calceoli have been lost even more frequently, with 9 apparent independent losses (possibly repeatedly occurring by these organs being first lost in females only), 1 apparent reversal (in *Eulimnogammarus verrucosus*, which possesses unusually small

Author Manuscript

This article is protected by copyright. All rights reserved

277 calceoli) and 1 gain of calceoli on antennae 1 in addition to their regular location on
278 antennae 2 (in *Micruropus wohli).*
279 Strikingly, the parasitic genus *Pachyschesis* is a strongly supported sister group to
280 one of its hosts, a much larger *Eucarinogammarus wagii*, suggesting, barring an unlikely
281 contamination artifact, a possible pathway to brood chamber parasitism through
282 intraspecific brood parasitism. Analogous to the presumed evolution of interspecific
283 brood parasitism in birds through conspecific brood parasitism as an intermediate stage
284 (Hamilton and Orians 1965), this hypothesized mechanism implies a divergence of an
285 ancestral species with conspecific brood parasitism into a large host species and a smaller,
286 possibly neotenic, parasitic species subsequently expanding to parasitize other large-
287 bodied amphipods.
288 The phylogeny also confirms the familiar patterns of fast radiation on the backdrop
289 of limited sequence evolution (Meyer et al. 1990). Morphologically strikingly different
290 species have nucleotide sequence divergence of sometimes less than 0.5%. Unlike the
291 classic example of Lake Victoria cichlids, where all the recent branches are very short,
292 Baikal amphipod radiation appears to be older, with the maximal rate of speciation
293 observed in the deeper branches of the phylogeny (Figure 2A). Both frequently used
294 models of changes in diversification rates with time (Rabosky and Lovett 2008; Höhna
295 2014) support decreasing speciation rates (Supplementary Figure 4B). Specifically,
296 Rabosky and Lovett (2008) model supports logistically decreasing speciation rate with
297 constant near-zero extinction rate (Supplementary Figure 4B, dotted line; log-likelihood
298 test vs. constant speciation model: $D = 11.4$; d.f.=1; $p<0.001$). Höhna (2014) model,
299 which takes into account incomplete lineage sampling, supports an exponentially
300 decreasing speciation rate with a non-zero extinction rate under the with the assumption
301 of "diversified" sampling strategy. Predicted extinction rate exceeds speciation rate
302 resulting in predicted negative diversification rate in the most recent 1/10 of the time
303 since divergence from the common ancestor with the Caspian species (Supplementary
304 Figure 4B, solid line; log-likelihood test vs. constant speciation model: $D = 334.9$; d.f.=2;
305 $p<0.00001$). On the other hand, in the same analysis (Höhna 2014) evaluated under the
306 assumption of "random" sampling (i.e., with sampling of lineages occurring without
307 respect to the number of species in a lineage already sampled), there was no difference in

Author Manuscript

This article is protected by copyright. All rights reserved

the likelihood of all models compared (Supplementary Figure 4B, dashed line; D = 0.01; d.f.=2; p>0.9). This difference between the results obtained under the assumptions of "diversified" vs. "random" sampling strategies may indicate that the true estimates of speciation rate change over time are probably somewhere in between of these two extremes.

Evidence of positive selection

We observed many episodes of positive selection (Table 1). Inferring positive selection is notoriously difficult, as it tends to be masked by negative selection and is radically influenced by conservative biases in the set of analyzed genes (see Discussion), so the inferred frequencies of such episodes may be underestimates.

Still, we infer approximately 120 episodes of positive selection, or 0.2 – 0.8 such episodes per gene per unit branch length (Table 1, Fig. 2B). With approximately 0.01-0.02 nucleotide substitution divergence of a typical Baikal amphipod species from the common ancestor of each of the two Baikal clades and assuming approximately 20,000 genes in the genome, this estimate corresponds to 40 - 80 genes having experienced an episode of positive selection in each lineage over the evolutionary time since the onset of amphipod radiation in Baikal. This number may be higher if the 20,000 genes is an underestimate (cf. Supplementary Table S4 that lists substantially larger number of predicted transcripts for most species studied).

We asked when these episodes had occurred. The d_N/d_S estimates and the frequency of positive selection episodes are the highest in the branches originating shortly after the second invasion of amphipod into Baikal (Supplementary Figure 4A). This corresponds to the time interval around ~9/10 of the time since divergence from the Caspian Sea outgroup (corresponds to 1 on $-\log_{10}$ scale; cf. Figure 2),. This may be an artifact of normalization by very short branch lengths, which are highly sensitive to large relative error in estimating branch lengths and numbers of positive selection episodes or number of non-synonymous changes. To ameliorate this problem, we calculated the frequency of positive selection episodes for all internal or terminal branches originating within each particular time interval (binned in such a way that each bin contains a constant fraction of cumulative branch length within each clade). This result is shown in

Author Manuscript

This article is protected by copyright. All rights reserved

Figure 2B. It demonstrates that the apparent peak of positive selection frequency around 0.1 units of relative time ago is not a normalization artifact (at least for the clades belonging to the second baikalian invasion) and that this peak corresponds to the time interval of the most intense accumulation of taxonomic diversity (Fig. 2A, B). Later in the phylogenetic time a gradual decrease of the frequency of positive selection episodes with time is observed in all four second invasion clades.

Table 2 and Supplementary Table 8 list individual COGs in which positive selection has been detected (assuming false discovery rate to correct for multiple COGs studied <0.1, and for aBSREL results, also a more stringent 0.05 Bonferroni correction for multiple tests over branches). Numerous COGs were found to contain signals of positive selection in more than one branch, either consecutively or in parallel. The two species in our dataset that are endemic to the deepest parts of the lake, *Garjajewia dershawini* and *Brachiuropus grewingkii*, are of a particular interest with respect to parallel changes. Although the number of COGs under parallel positive selection in these species (2) is not significantly higher than expected (1.4), the two cases observed are interesting functionally (see below).

Likewise, although numerous sites at numerous COGs with or without positive selection signal diversified in parallel in the two abyssal lineages studied, the frequency of such events was similar or lower than the frequency of such events occurring, along similar branch lengths, in either of the two abyssal species and a reference species *Acanthogammarus godlewskii*. Specifically, there were 45 sites in 41 COGs with parallel substitutions in the two abyssal lineages and 60 sites in 53 COGs and 46 sites in 44 COGs between each of them and *A. godlewskii*. This indicates that while parallel changes in homologous sites of homologous proteins are abundant in many baikalian lineages, their occurrence is not unusually frequent specifically in the two abyssal lineages.

Functional analysis of COGs with episodes of positive selection

Gene ontologies of the 175 COGs identified in all 66 species are summarized on Supplementary Figure 5. Proteins with nuclear, ribosomal, and membrane localization are common among these 175 COGs, as are proteins with gene expression and proteolysis functionality. Many ribosomal proteins are represented. None of the functional GO

Author Manuscript

This article is protected by copyright. All rights reserved

annotations were significantly enriched after multiple test correction in any of the COG lists with evidence of positive selection, except for the category “Respiratory chain protein” significantly enriched (FDR=0.027) in the list of proteins showing signals of positive selection in the two abyssal lineages. This is a conservative test; for example, 5 out of 5 NADH dehydrogenase subunits present in the reference list appear on some of the test lists and yet this highest possible enrichment of this GO category still does not reach significance sufficient to survive multiple test correction. Regardless of relative enrichment, a clear pattern of functional spectrum of proteins showing episodes of positive selection has emerged (Table 2). Out of the 33 top candidate genes (false discovery rate FDR<0.1), 8 code for proteins with mitochondrial membrane localization and functionality central to ATP synthesis (shown in bold in Table 2). The same results for each clade analyzed separately are listed in Supplementary Table 8. Several key components of membrane- and substrate-level ATP synthesis are represented: electron-transport chain (several NADH dehydrogenase subunits, one cytochrome b-c1 complex subunit), membrane ATPase (2 subunits), succinyl-ligase (the Krebs cycle ATP synthase) and adenine nucleotide translocase that accomplishes transport of ADP into and ATP out of mitochondria. While distributed throughout the phylogeny (Supplementary Table 8), episodes of positive selection in these proteins appear to be particularly common in abyssal species. For example, one of the flavoprotein subunits of NADH dehydrogenase shows a significant signal of positive selection when the two strictly abyssal species in the dataset, *Garjajewia dershawini* and *Brachiuropus grewingkii*, are tested against the background of other members of the same *Acanthogammaridae* clade (Supplementary Table 8). Furthermore, two out of the 4 proteins showing the strongest signals of positive selection in the plankton species *M. branickii* (which also inhabits the abyssal zone, although not limited to it) are subunits of both ATPases: the gamma- (“stalk”) subunit of membrane ATPase and one of the subunits of the succinyl-ligase (Table 2; Supplementary Table 8). Casual examination of multiple alignments of these 4 proteins (2 NADH dehydrogensase subunits and 2 subunits of ATPases) with their nearest orthologs in 2 crustaceans, 3 insects 2 other arthropods and 6 other metazoans (see Data Availability section) reveals that all substitutions that occurred in the lineages in question,

Author Manuscript

This article is protected by copyright. All rights reserved

including those for which MEME indicated a high posterior probability of positive selection, have occurred in non-conserved areas of the proteins.

Another functional group highly represented among proteins with signals of positive selection (but, again, not significantly enriched relative to the reference set) are the ribosomal proteins (italics in Table 2: 11 out of 33).

Finally, one other functional category of proteins recurrent in the list of candidate genes with signals of positive selection are proteins related to the function of muscles and calcium metabolism. This includes evidence of positive selection in the light chain on myosin in the plankton *M. branickii* and of profilin (an actin-binding protein involved in regulating the actin cytoskeleton) in several branches across the phylogeny (Supplementary Table 8; Table 2). Additionally, two other functionally related proteins have been identified as likely targets of positive selection (Supplementary Table 8): a sarcoplasmic calcium-binding protein which shows signals of positive selection in several branches of *Hyalellopsis* and *Pallasea* clades, and alpha-actinin, an actin-binding protein whose muscle isoforms help anchor the myofibrillar actin filaments. This protein shows positive selection signals in the ancestral branch of *Acanthogammaridae s. str*. as well as in the planktonic *Macrohectopus branickii.*

Discussion

High diversification rate

We observe a high diversification rate in Baikal amphipods which appears to be particularly fast in early branches of the more recent invasion of amphipods into Baikal (Figure. 1 and 2), when the majority of extant clades appeared within the timeframe corresponding to approximately 0.01-0.02 nucleotide substitutions per site. While the predictions of the diversification rates changes (Supplementary Figure 4B) may be biased (Moore et al., 2016) and they certainly suffer from incomplete taxonomic sampling in our data, there is no doubt that majority of Baikal amphipods clades originated and majority of morphological diversity was generated during this short period of time, thus being a classic case of explosive radiation.

Author Manuscript

This article is protected by copyright. All rights reserved

It is difficult to calibrate this rate in terms of astronomical time. The most comprehensive recent gammaridean amphipod phylogeny (Hou et al. 2011) is based largely on mitochondrial and rRNA genes with a single nuclear protein coding gene representing a small portion of data. Additionally, this analysis uses the upper bound estimate of Baikal age (30 My) as one of the three calibration points, so directly applying that calibration to this study would be circular; the other two calibration points are Baikal-independent, but their applicability to larger Baikal species with longer generation times is questionable. However, even if we take the estimate of Baikal age (Mats et al. 2011; Hou et al. 2011; Hou and Sket 2016) as a very conservative time point and assume that amphipod radiation occurred immediately after Baikal was formed, we estimate that the 300 extant species are the result of diversification with about 0.3 speciation events per 1 My, i.e., about 3 My between speciation events. The majority of these events occurred during the first half of the time elapsed since the onset of radiation, leading to lineages (Figure 1) representing nearly all the extant genera. Assuming that the rate of accumulation of nucleotide substitutions was constant throughout the phylogeny, this yields the estimate of 0.55 speciation events per 1 My, or approximately 2 My between speciation events in each lineage. This estimate is substantially higher than those for the peak of freshwater *Gammaridae* diversification soon after colonizing continental waters in Eocene (approx. 0.11 events per 1 My, Hou et al. 2011) and even higher than the estimates for marine ancestors of *Gammaridae* or for recent freshwater clades (0.01-0.05; Hou et al. 2011). Likewise, our estimates of diversification rates are high in comparison to those in numerous other invertebrates (typically 0.05-0.01 per 1 My), but similar to those obtained for Hawaiian Drosophila or Great African Rift lakes cichlids (Coyne and Orr 2004, Table 12.1). The analysis of genome size in Baikal amphipods (Jeffery et al., submitted) indicates that genomes are larger in quickly diversifying lineages, possibly corroborating the previously discovered elevated rate of gene duplications and transposable elements insertions in quickly diversifying Great African Rift Lakes cichlid lineages (Brawand et al. 2014).

Discordance between morphological and molecular divergence

Author Manuscript

This article is protected by copyright. All rights reserved

As repeatedly pointed out earlier for Baikal amphipods (Yampolsky et al. 1995, Sherbakov et al. 1998; McDonald et al. 2005) as well as other endemic species flocks (e.g. Meyer et al. 1990), fast diversification has led to striking morphological divergence despite shallow differentiation on the molecular level. Numerous very well defined species have sequence divergence <0.005 substitutions per site for the most conserved subset of genes (Fig. 1) and not much greater for larger subsets of genes (Supplementary Fig. 1). Furthermore, we also confirm the presence of a significant intraspecific genetic differentiation among conspecifics and among morphologically nearly identical species previously described as subspecies that is similar in magnitude to the divergence among well defined species. That is true for *Oxyacanthus spp.* (subspecies of *Oxyacanthus flavus sensu* Bazikalova 1945) and *Parapallasea spp.* (subspecies of *Parapallasea borowskii sensu* Bazikalova 1945), as well as for three pairs of conspecifics included in our study that show no apparent ecological or morphological divergence, but have been sampled in different locations (*Eulimnogammarus cruentus),* or from Baikal vs. Angara river (*Pallasea cancelloides),* or from different host species (*Pachyschesis branchialis*). While strong genetic differentiation is neither a necessary nor a sufficient condition for the recognition of cryptic species, this observation may offer further support to the hypothesis of widespread occurrence of cryptic differentiation within the Baikal radiation (Väinölä and Kamaltynov 1999; Danelia et al. 2011). In particular, lineages within the parasitic *Pachyschesis branchialis* may indeed represent incipient species (Takhteev 2000b), although both individuals included in this study were identified as the type species and not as any of host-specific taxa described by Takhteev (2000b).

Morphological homoplasies

Our robust phylogeny unequivocally confirms the previously hypothesized double invasion of amphipods into Baikal (McDonald et al. 2005) and, overall, shows a remarkable success of morphology-based systematics in resolving major lineages of Baikal amphipods despite wide-spread morphological innovations and homoplasy. These successes include both the correct placement of the Baikal radiation within the family *Gammaridae* despite enormous morphological divergence (Stebbing 1899; Sowinsky 1915; Hou and Sket 2016) and correct insights on evolutionary relationships within the

Author Manuscript

This article is protected by copyright. All rights reserved

baikalian radiation despite widespread parallel evolution in taxonomically important characters (Fig. S2). For example, the affinity of *Asprogammarus* with *Micruropidae* has been correctly recognized (Kamaltynov 1999) despite the presence of dorsal keels and 3-5 articulate accessory flagellum not observed elsewhere in this clade (Supplementary Figure 2B,C). Likewise, Bazikalova (1945) and Takhteev (2000b) pointed out similarities between *Poikilogammarus* and *Pallasea* despite secondary loss of spines in *Poikilogammarus*, an affinity strongly confirmed by our phylogeny (Figure 1, Supplementary Figure 1C). The long-recognized affinity between *Brandtia* and *Dorogostaiskia parasitica* has also been established (Yampolsky et al. 1995; Kamaltynov 1999;) despite differences in various taxonomically important traits, such as the degree of reduction of the accessory flagellum (Supplementary Figure 2B) or presence of secondary spines (a homoplasy shared by *Dorogostaiskia* and *Garjajewia*).

Finally, Bedulina et al. (2014) expressed doubt in monophyly of the genus *Eulimnogammarus* and its subgenera (Basikalova 1945) elevated to genera by Barnard and Barnard (1983). Indeed, the genus *Eulimnogammarus* appears to be polyphyletic. A representative of the subgenus *Heterogammarus* (Bazikalova 1945) with typical *Pallasea*-like equiramous uropods 3 is branching reliably within *Pallasea* (Figure 1). Two species of the subgenus *Eurybiogammarus* sensu Bazikalova (1945), *E. violaceus* and *E. ussolzewii*, characterized by mildly reduced inner branch of uropod 3, are paraphyletic with the rest of *Eulimnogammarus* with respect to the clade that contains parasitic genus *Pachyschesis* and one of its hosts *Eucarinogammaris wagii* (Figure 1, Supplementary Figure 1E). The representatives of subgenera *Philolimnogammarus* and *Eulimnogammarus s.str* forming a monophyletic clade in our reconstruction share strong reduction of the inner branch of uropod 3 and the loss of calceoli, but the two subgenera are paraphyletic with respect to each other's members (Figure 1, Supplementary Figure 1E, Supplementary Figure 2A).

One feature of the phylogenetic topology that morphology-based taxonomy failed to recognize is the most basal group of the acanthogammarid invasion into Baikal. In our reconstruction a representative of the morphologically primitive and ecologically, presumably, generalist genus *Echiuropus*, traditionally included into the micruropid clade (Kamaltynov 1999), branches as the sister clade to all other acanthogammarids.

This article is protected by copyright. All rights reserved

Sequencing of additional species of this genus is needed to determine if it indeed represents the ancestral morphology and ecology for the more recent of the two invasions.

To conclude, several morphological changes, in particular the development and losses of spines and body parts reduction, occurring in parallel in different lineages of Baikal amphipods corroborates previously known examples of frequent parallel morphological evolution within endemic species flocks (Kocher et al., 1993; Martens, 1997; Rüber et al., 1999; Allender et al. 2003; Salzburger 2009).

Frequency of positive selection episodes

We report frequent occurrence of episodes of positive selection during adaptive radiation within Baikal amphipod species flock. We almost certainly underestimate the occurrence of positive selection due to several reasons. First, positively selected genes tend to be evolutionarily transient (Vishnoi et al. 2010) and expressed at a lower levels and in tissue-specific manner (Kosiol et al. 2008). All these factors bias our sample of conserved multispecies transcriptomics-based COGs towards negatively-selected genes, and reduce the likelihood that a positively selected gene would be represented in it. Indeed, the d_N/d_S ratio decreases with the number of species in which a gene is represented, indicative of negative selection (Supplementary Figure 3B). Second, many episodes of positive selection are likely to be missed due to low power of HyPhy models to detect positive selection on very short branches (Smith et al. 2015). For example, the majority of amino acid changes identified by MEME (Murrell et al. 2012) as likely to have occurred by selection are double-nucleotide changes or transversions, indicating that any changes that required a single transition would be overlooked even when they were indeed accomplished by selection. Even with these biases, we estimate that dozens of genes have been under positive selection in each lineage of Baikal endemic amphipods.

We also show that the frequency of episodes of positive selection per unit of branch length detected by the aBSREL model (Smith et al 2015) has peaked during the time bracket of the most intensive lineage diversification. There are three caveats that need to be addressed. First, and obviously, this conclusion relies on correct phylogenetic resolution of very short branches, which, even with high bootstrap support, may be subject to errors. Second, we may have underestimated the frequency of positive

Author Manuscript

This article is protected by copyright. All rights reserved

554 selection episodes in early branches, because many early internal branches are very short,
555 resulting in lower power of positive selection detection. However, binning positive
556 selection episodes and branch length over time intervals in which each branch initiated, in
557 addition to ameliorating to the normalization bias discussed above, also take care of this
558 problem, because earlier time intervals include both the typically short internal branches
559 and long terminal branches, while late time intervals include some of the shortest
560 terminal branches. Thus, there was no difference in the distribution of branch length
561 between the early and the late time intervals and thus probably no large difference in the
562 power to detect positive selection (Smith et al. 2015).
563 Finally, and on the other hand, we should keep in mind that alignments and
564 ancestral reconstructions are likely to become less reliable for deeper nodes than for most
565 recent ones. Potentially, this could inflate the frequency of apparent selection episodes in
566 early branches, meaning that we overestimate the decrease of their frequency with time.
567 Yet, due to very low divergence among sequences in our dataset, homoplastic changes
568 are rare and saturation is not an issue. Thus we believe that the observed pattern of the
569 increased occurrence of positive selection during the most intense diversification of
570 Baikal amphipod species flock is real. It may indicate that the ecological opportunity
571 indeed appears to elicit selective response, as predicted for adaptive radiations (Yoder et
572 al. 2010; Wellborn and Langerhans 2015; Berner and Salzburger 2015) and that this
573 opportunity may have been the highest when early amphipod invaders into Baikal
574 encountered a large number of unoccupied niches.
575 What may have been the exact nature of these ecological opportunities, given the
576 likely paucity of trophic niches (Yoshii 1999) in Baikal's oligotrophic communities? We
577 believe that this question should be approach in conjunction with the enigma of how
578 Baikal biodiversity, including amphipods, may have survived repeated glaciations
579 (Karabanov et al. 2004). While it is difficult to calibrate the time of the second invasion
580 in absolute time due to the lack of fossil evidence and doubtful applicability of external
581 calibrations, it is clear that, despite short branch lengths, this radiation is much older than
582 postglacial. During the last 5 my Baikal experienced a significant cooling, occurring in
583 two separate episodes (Williams et al., 1997) that followed by repeated, much more
584 severe cooling episodes during Quaternary glaciation cycles. Initial diversification thus

Author Manuscript

This article is protected by copyright. All rights reserved

almost certainly occurred in much warmer, much more oxygen poor Baikal, while the most recent terminal branches are likely to have originated during glacial cycles or even since the last glacial maximum. Although Baikal has never been completely frozen or filled with glacial ice (BDP Members 2000), ice cover was extremely thick and primary production plummeted during the last glaciation maximum (and possibly previous maxima; Karabanov et al. 2004). Karabanov et al (2004) suggested that shallow euphotic areas may have remained inhabitable due to river inflow and could have served as refugia from which lineages recolonized the abyssal zone. Alternatively, one may hypothesize that Baikal biodiversity survived (and possibly evolved) in deep-water refugia supported by hydrothermal vents. Today Baikal contains several types of such deep-water vent habitats (both hot and cold) in which a diverse fauna that includes amphipods is supported by a community of chemoautotrophic primary producers (Zemskaya et al. 2012). Ecological opportunity (which perhaps should be called in this context ecological necessity) may have been associated with repeated colonizations of such deep-water, oxygen-rich refugia.

On the other hand, long periods of "near-dead" state of Baikal ecosystem proposed by Karabanov et al. (2004) are not supported by several molecular phylogenetic studies of lake Baikal endemic groups (Hunt et al. 1997; Hausdorf et al. 2003; McDonald et al. 2005), all of which, including the one presented here, support long-lived lineages clearly predating the last glaciation maximum. For example it is difficult to imagine that specialized endemics of Baikal abyssal zone (and the planktonic *Macrohectopus branickii)* may have accomplished the observed radical morphological changes in the last 10-15K years. In particular, the sharp and prolonged environmental deterioration proposed by Karabanov et al. (2004) would have radically affected Baikal endemic fish species. Yet, the radiation of Baikal sculpins is dated at 3.3-8.0 Mya (Hunt et al. 1997; Goto et al. 2015) and shows no evidence for mass extinction. Similarly, phylogeography of mitochondrial and microsatellite markers in Siberian sturgeon *Acipencer baerii* presently distributed in all major Siberian rivers strongly indicates that Baikal was a single refugium for this species during (Barmintseva and Mugue, in press). All these evidence supports that glaciations did not affect the Baikal biota as severely as proposed by Karabanov et al. (2004). Severe depletion of diatoms in the sediment core, reported by

Author Manuscript

This article is protected by copyright. All rights reserved

Karabanov et al. (2004) may be due to other reasons than drop of lake ecosystem productivity. Pollen record of terrestrial plants, collected from the same core, indicate that surrounding coastal landscapes were occupied by cold steppe floral assemblages, not snow deserts (Tarasov et al. 2007). Thus, at least deep-water components of Baikal ecosystem may have retained active life even during glaciation maxima. Either way, it appears that adaptation to abyssal habitats may have been an important feature of the evolutionary past of all Baikal amphipods, not just those that are abyssal endemics today. The position of *Garjajewia*, an obligatory abyssal taxon, basal to the *Acanthogammaridae* + *Pallaseidae* clade, is consistent with this hypothesis (Fig.1).

Functional genomics and parallel evolution

We observe numerous episodes of positive selection in members of the same COG in different lineages in parallel (Table 2, Suppl. Table 8). This observation, as well as a number of observed morphological homoplasies (Suppl. Fig. 2), is consistent with high occurrence of parallel evolution in species flocks (Berner and Salzburger 2015). However, tests for enrichment of parallel substitutions relative to random expectations failed to detect any deviation from randomness in the two abyssal clades that we expected to show the highest frequency of parallelisms. It is entirely possible that such parallelisms could have been detected if we had an opportunity to analyze more than two independent abyssal lineages. However, this finding is consistent with the conclusion that many detectable episodes of parallel evolution may in fact occur entirely through neutral processes (Bazykin et al. 2007) and not represent any excess of frequency over random expectations (Thomas and Hahn 2015).

Furthermore, it is difficult to make any conclusions about the non-randomness of functional parallelisms, i.e., occurrences of episodes of positive selection in non-orthologous, but functionally related proteins, as observed in this study. Because the entire (reference) dataset is highly enriched in certain categories of genes relative to the whole transcriptome (see above), it impossible to distinguish between two reasons why a particular functional category occurs among discovered positive selection episodes. On the one hand, positive selection may be equally common throughout the genome regardless of functionality, while a particular category of proteins is more likely to occur

Author Manuscript

This article is protected by copyright. All rights reserved

647 in the full-occupancy set of COGs (for example, due to their high conservation and/or
648 high expression, making them likely targets during library construction and easy
649 transcripts to align). On the other hand, the subset of COGs in our datasets may be more
650 or less random with respect to functionality, but selection is indeed much more frequent
651 in particular functional categories. In the absence of a possible ecological context for
652 selection in a particular functional group, such as, for example, ribosomal proteins (Table
653 2) it is difficult to hypothesize on which of the two possibilities is more likely.
654 On the other hand, episodes of positive selection observed in other functional
655 categories, such as membrane proteins with respiratory chain and ATP synthesis
656 functionality, may be easier to explain, as they are concordant with the above hypothesis
657 of abyssal refugia that allowed Baikal amphipods to survive glaciation periods. Effects
658 of abyssal (>400 m) depths on animals, virtually non-existent in freshwaters outside of
659 Baikal, have been well characterized in marine animals (Somero 1992; Sébert et al. 2004).
660 One of the most profound effects of high hydrostatic pressure (exacerbated by low
661 temperature) is the reduction of membrane fluidity that in particular affects functions of
662 inner mitochondrial membranes (Sébert et al. 2004; Scaion et al. 2008). At this point it is
663 only possible to hypothesize whether the observed amino acid substitutions in NADH
664 dehydrogenase, ATPases and ATP translocase subunits have any adaptive value in terms
665 of ameliorating the adverse effect of increased rigidity of the inner mitochondrial
666 membrane. Structural locations of these substitutions (Yampolsky et al., in preparation)
667 offer at least some insight on this. While the two NADH dehydrogenase subunits
668 (flavoproteins 1 and 2) identified in our analysis as likely subjects of positive selection
669 are not located in the proximity of the membrane and probably do not interact with it,
670 sites that appear to be under selection in ATPase gamma subunit in the *M. branickii* are
671 indeed located in the immediate proximity to the membrane. The shaft-forming gamma
672 subunit is believed to be essential in transmitting the energy from the rotating c-ring to
673 the conformational changes in the catalytic sites of F1 subunits (von Ballmoos et al.
674 2009) and its interactions with the membrane-imbedded subunits is critical for the
675 mechanism of ATP synthesis (Zimmermann et al. 2006; Okazaki and Hummer 2015).
676 One may hypothesize that selection may have introduced amino acid changes that allow
677 the planktonic *M. branickii* that occurs from surface to as deep as 1400+ m and

Author Manuscript

This article is protected by copyright. All rights reserved

accomplishes diurnal vertical migrations of up to 250 m to maintain high level of ATP synthesis necessary for the active life of a plankton predator regardless of hydrostatic pressure it experiences. The fact that the other ATP synthase, the Krebs cycle succinate lyase, is also among the top genes identified as likely targets for positive selection further corroborates this hypothesis.

Materials and Methods

Field collections: Baikal amphipods we collected during the field seasons of 2012 and 2013 at various locations (Supplementary Tables 1 and 2) within Lake Baikal and neighboring water bodies. Caspian sea species *Pandorites podoceroides* and cosmopolitan *Gammarus lacustris* (sampled in a small lake adjacent, but not connected to to Baikal) were used as outgroups. Collection of specimens at depths 0 to 20 m were accomplished by hand collection and dipnet sweeps (from shore or by scuba diving), at 15 – 100 m and below 400 m by baited traps and at 20 - 60 m on soft substrate by dredging. Collected individuals were kept without food at 6°C for 24 hours to reduce the occurrence of food RNA in guts and preserved individually in either liquid nitrogen (whole body) or RNAlater (after fragmenting into fragments not thicker than 3 mm). When unequivocal species identification was not possible at the time of collection, it was conducted in the laboratory on ethanol-preserved voucher specimens whose conspecific identity with specimens preserved for sequencing was confirmed in the field. Whenever possible conspecific pairs captured in pre-copula were used for this purpose, with female preserved for sequencing and male preserved for identification as the voucher specimen. Generally, when available and when sex identification was possible, females were used for sequencing. Large specimens obtained in small numbers were cut longitudinally and left side was preserved for sequencing, while the right side preserved in ethanol as the voucher for identification. Species were selected for sequencing with the goal to increase the number of lineages covered (i.e., additional available species redundant with respect to large clades such as *Micruropus, Eulimnogammarus* or *Pallasea* were not included

Author Manuscript

This article is protected by copyright. All rights reserved

into sequencing); thus the "diversified" model of phylogenetic incompleteness was assumed for diversification rate analysis (Höhna 2014), see below.

Note on taxonomy: Taxonomy is given following Bazikalova (1945) and Kamaltynov, (1999). It disagrees with the observed phylogeny (Fig.1, Supplementary Fig.1) in the following: *Micruropidae* clade includes *Macrohectopus*, but not *Echiuropus*; *Acanthogammaridae s.str* clade (*Acanthogammarinae sensu* Kamaltynov 1999) does not include *Eucarinogammarus* (which shows affinity to *Eulimnogammarus*, see results) or the genera *Pallasea, Brandtia* and *Dorogostaikia* (which form a separate clade together with *Parapallasea*, see results); *Acanthogammaridae s.lat.* includes the genus *Pachyschesis*. Additionally, several Bazikalova's (1945) subgenera within the genera *Eulimnogammarus, Pallasea* and *Acanthogammarus* are either elevated to genus level (*Heterogammarus*, *Corophiomorphus* (*Sluginella*), *Homalogammarus, Brachiuropus*) or eliminated as suggested by Bedulina et al. 2014 (see Discussion). We use "*Acanthogammaridae s. lat"* to describe the second amphipod invasion into Baikal, i.e. all Baikal clades except *Micruropidae.*

RNA next-generation sequencing: RNAs were extracted from whole bodies of liquid nitrogen- or RNALater-preserved specimens <3 mm long and from a cross-section of a metasome segment in larger individuals using Qiagen RNeasy mini kit (Qiagen, Germany). RNA quality was checked using capillary electrophoresis on Bioanalyzer 2100 (Agilent, USA). Library preparation was performed using TruSeq RNA sample preparation kit (Illumina, USA) according to manufacturers instructions. Resulting libraries were quantified using Qubit fluorometer (Thermofisher, USA) and qPCR and sequenced on Illumina Hiseq instrument with 101 + 101 read length.

Filtering, de novo transcriptome assembly, alignment and orthologization

After removing duplicate and low complexity reads (Schmieder and Edwards 2011) we filtered out mtRNA reads using BLASTN search (Altschul et al 1990) with e-value cutoff of 10^{-5} against the mitochondrial genome of *Eulimnogammarus verrucosus* (Rivarola-Duarte et al 2014). We blasted the left and the right read of each pair separately and removed all those pairs having at least one hit. Overlapping reads in a pair were merged

Author Manuscript

This article is protected by copyright. All rights reserved

739 using fastqjoin (Aronesty 2011; min overlap = 20 bases). All reads were trimmed from
740 adapters and low quality ends using fastqc
741 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) to detect adapters and
742 trimmomatic (Bolger et al. 2014; with the minimum base retention quality 20 and the
743 minimum length of reads 50) for trimming. Filtration statistics is shown in Table S3.
744 We then assembled reads de novo using Trinity (Grabherr et al 2011, versions
745 20140413p1 20140717), retaining all isoforms of a given locus for the further analysis.
746 Assembly statistics is shown in Table S4. In order to check the assembled contigs for
747 contamination, all contigs were blasted against NT database (released 2014.05.24
748 2014.10.11) with e-value cutoff 10^{-5} and identity cutoff (ID $\geq$ 95%). Loci having
749 isoforms with top hits outside of Arthropoda were excluded. Loci without hits at all were
750 retained. Contamination statistics is shown in Table S5. We also filtered assembled
751 transcripts versus UniVec database. Protein coding sequences were extracted from
752 assembled transcripts with TransDecoder (Grabherr et al 2011) using PFAM database
753 (Finn et al 2016) and with minimal length parameter set to 70. In cases when
754 TransDecoder found two CDs per transcript the longer CDS was retained. CDS statistics
755 is shown in Table S6.
756 For orthology analysis we used the longest protein coding sequence for the locus.
757 We inferred the orthology relationship using standard OrthoMCL pipeline v.2.0 (Li et al
758 2003). We filtered out clusters having paralogues by using only those OrthoMCL
759 pipeline groups in which each species was represented by the single transcript. This
760 resulted in 175 clusters of unique orthologs for 66 species. We then aligned protein-
761 coding sequences with mafft (Katoh and Standley 2013) and reverse translated protein
762 alignments to cDNA alignments with RevTrans (Wernersson and Pedersen 2003). The
763 best fit substitution model was determined by jmodeltest2 (Darriba et al. 2012; Guindon
764 and Gascuel 2003). For all clades except the *Acanthogammaridae* clade, GTR+Γ+I was
765 the best scoring model, while for *Acanthogammaridae* the best model was TIM1+Γ+I
766 followed by GTR+Γ+I with delta BIC=18. We used GTR+Γ+I model for all phylogeny
767 inferences. We built the phylogenetic tree using RAxML 8.2.4 (Stamatakis 2014) with
768 1000 bootstrap replicates. For comparison we also reconstructed phylogenies using

Author Manuscript

This article is protected by copyright. All rights reserved

MrBayes 3.2 (Ronquist et al., 2012), which resulted in identical topologies; only RAxML results reported.

We repeated the steps starting from mafft (Katoh and Standley 2013) and RevTrans (Wernersson and Pedersen 2003) alignment separately for the five clades marked with color triangles in Fig.1, and for all 66 samples, resulting in sets of COGs of various size represented in each species of these 5 clades and in a set of COGs represented in variable number of lineages (Supplementary Table 7).

Additionally we generated a set of 523 COGs with >2/3 occupancy (i.e., present in at least 45 species out of 66) with iterative mapping-and-variant calling approach (Naumenko 2015). These clusters contain more species than generated for particular clades; however, they have inferior alignment quality.

Diversification rates

Branching times were estimated from the chronotree obtained by chronos function (Kim and Sanderson 2008) of ape R package. Diversification rates along the reconstructed phylogeny were estimated using the models of Rabosky and Lovett (2008) and the "diversified" version of Höhna (2014); the "random" version of this model was also evaluated for comparison. Both these models suffer from a number of methodological problems and their predictions may be biased (Moore et al., 2016), so these results must be interpreted with caution.

Detection of positive selection and parallel molecular evolution

We estimated the dn/ds ratio and detected genes with sites having dn/ds > 1 in some branches using likelihood ratio test for branch-site model in PAML (Jeffares et al 2015; Yang 2007). For the comparison of dn/ds ratios between branches with different branching times, the two estimates obtained for two lineages originating from each node were averaged and this average juxtaposed to the node age estimated as described above.

Alignment-wide tests for episodes of positive selection, lineage-specific branch-site tests and site-specific test for episodes of diversification were conducted using, respectively, BUSTED (Murrell et al. 2015), aBSREL (Smith et al. 2015) and MEME (Murrell et al. 2012) modules of the HyPhy package. The following filters were applied

Author Manuscript

This article is protected by copyright. All rights reserved

to HyPhy analysis to reduce the rate of false positives caused by unreliable alignment: (i) five codons surrounding indels of size >=3 bp were removed from the analysis and (ii) any amino acid sites within 16 codons from either end of the alignment if there were > 2 different amino acids per site at any two sites were removed from the analysis. Both PAML and HyPhy analyses resulted in similar lists of COGs and branches with evidence of positive selection (with HyPhy showing somewhat higher statistical power); only HyPhy COG-specific results will presented.

To detect parallel evolution of COG sequences during independent invasion of 2 lineages into the abyssal zone (Sherbakov et al. 1998) we focused on *Garjejewia dershawini* and *Brachiuropus grewingkii*, each of which does not occur above 100 m and has the maximal occurrence below 400 m (Bazikalova 1945). For each of 976 COGs represented in these two species and two shallow-water reference representatives of the *Acanthogammarus* clade we applied aBSREL model (Smith et al. 2015) specifying these two species as the test group and all other lineages in the dataset as the background reference.

We normalized the observed count of positive selection episodes by the number of genes in a dataset and by branch length (measures in nucleotide substitutions per site). In order to increase robustness of normalized frequency of positive selection episodes, in addition to estimating it for each branch of the phylogeny (many of which are very short and are likely to have been measured with a large relative error) we normalized the number of episodes observed in all branches originating in time intervals each containing 20% of the combined length of all branches in each clade. To ameliorate problems arising from potentially non-normal nature of such binned time variable, we evaluated the correlation between normalized frequency of positive selection episodes and branching time by Spearman's rank correlation, in addition to a parametric regression model.

To analyze functional patterns of genes with evidence of positive selection and to test whether functionally similar but not necessarily homologous proteins showed parallel evolution, we annotated all COGs in all datasets using blast2go (Conesa 2005; Supplementary Figure 5) and performed functional enrichment tests (Fisher's Exact Test) using the considered list of COGs as the test and the remainder of the annotation dataset as the reference.

Author Manuscript

This article is protected by copyright. All rights reserved

Data accessibility Sequenced reads have been deposited to NCBI Sequence read archive (SRP07418; SRR3467040-3467102, Supplementary Table 9); Transcriptome Shotgun Assembly project (PRJNA321360) has been deposited at DDBJ/EMBL/GenBank under accessions listed in the Supplementary Table S9. Versions of individual assemblies described in this paper are the first versions, i.e. GERM01000000. Alignments listed in Supplementary Table 7, and phylogenetic trees in newick format are available at http://faculty.etsu.edu/yampolsk/data/BaikalAmphipods. Data for Fig. 2 are available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.4ck36.

Acknowledgements

We are grateful to colleagues at Irkutsk State University Baikal Biological station and Baikal Museum and to the crew of RV Professor Treskov for hospitality and assistance with field work, to Valery Chernykh for providing diving support, to Fyodor Kondrashov, Yuliya Kulikova, Yulia Bibilo and Natalia Sashina for field work volunteering, to Sergey Kirilchik for kindly provided specimens and to Boris Sket, Gauthier Chapelle, Kara Woo and Elizaveta Kondrateva for images of amphipods. This study was partially supported by the Ministry of Education and Science of the Russian Federation (grant 11.G34.31.0008) to ASK, the Russian Foundation for Basic Research grant 15-34-21135-mol_a_ved to GAB, and US National Science Foundation grant 1136710 to LYY.

References Cited

Albrecht C, Wolff C, Glöer P, Wilke T (2006) Concurrent evolution of ancient sister lakes and sister species: the freshwater gastropod genus *Radix* in lakes Ohrid and Prespa. In: Wilke T et al., eds., Patterns and Processes of Speciation in Ancient Lakes. Proceedings of the Fourth Symposium on Speciation in Ancient Lakes, Berlin, Germany, pp. 157–167.

Allender CJ, Seehausen O, Knight ME, Turner GF, MacLean N (2003) Divergent selection during speciation of Lake Malawi cichlid fishes inferred from parallel radiations

Author Manuscript

This article is protected by copyright. All rights reserved

in nuptial coloration. Proceedings of the National Academy of Sciences of the United States of America, 100, 14074–14079.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol., **215**, 403-410.

Aronesty E (2011) ea-utils: Command-line tools for processing biological sequencing data. Available: http://code.google.com/p/ea-utils

Bailey SF, Dettman JR, Rainey PB, Kassen R (2013) Competition both drives and impedes diversification in a model adaptive radiation. Proc. Royal Soc. B-Biological Sciences, **280**, 20131253. DOI: 10.1098/rspb.2013.1253

Barmintseva AE, Mugue NS. Natural genetic polymorphism and phylogeography of Siberian sturgeon Acipenser baerii Brandt, 1869. Rus J Genetics, in press.

Barnard JL, Barnard CM (1983). Freshwater Amphipoda of the World. Mt. Hayfield Associates, Vernon, VA.

Bazikalova AY (1945) Amphipods of Lake Baikal. Trudy Baikalskoj Limnologicheskoj Stantsii, **11**. (In Russian.)

Bazykin GA, Kondrashov FA, Brudno M, Poliakov A, Dubchak I, Kondrashov AS (2007) Extensive parallelism in protein evolution. Biol Direct, **2**, 20.

BDP Members (The Baikal Drilling Project Group) (2000) The Late Cenozoic paleoclimate record from Lake Baikal (results of 600 m deep drilling core investigations). Russian Geology and Geophysics, **41**, 3–32. (In Russian).

Bedulina DS, Takhteev VV, Pogrebnyak SG, Govorukhina EB, Madyarova EV, Lubyaga YuA et al. (2014) On *Eulimnogammarus messerschmidtii*, sp. n. (Amphipoda: Gammaridea) from Lake Baikal, Siberia, with redescription of *E. cyanoides* (Sowinsky) and remarks on taxonomy of the genus *Eulimnogammarus*. Zootaxa, **3838**, 518–544. dx.doi.org/10.11646/zootaxa.3838.5.2

Berner D, Salzburger W (2015) The genomics of organismal diversification illuminated by adaptive radiations. Trends in Genetics, **31**, 491-499. DOI: 10.1016/j.tig.2015.07.002

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics btu170.

Author Manuscript

This article is protected by copyright. All rights reserved

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S et al. (2014) The genomic substrate for adaptive radiation in African cichlid fish. Nature, **513**, 375-381.

Brooks JL (1950) Speciation in ancient lakes. The QuarterlyReview of Biology, 25, 30–60.

Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, **21**, 3674-3676.

Cornell HV (2013) Is regional species diversity bounded or unbounded? Biological Reviews, **88**, 140-165. DOI: 10.1111/j.1469-185X.2012.00245.x

Coyne JA, Orr HA (2004) Speciation. Sinauer, Sutherland, MA.

Cristescu ME, Adamowicz SJ, Vaillant JJ, Haffner, DG (2010) Ancient lakes revisited: from the ecology to the genetics of speciation. Molecular Ecology, 19: 4837-4851.

Daneliya ME, Kamaltynov RM, Väinolä R (2011) Phylogeography and systematics of *Acanthogammarus s. str*, giant amphipod crustaceans from Lake Baikal. Zoologica Scripta, **40**, 623-637.

Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nature Methods, **9**, 772.

Fan S, Meyer A (2014) Evolution of genomic structural variation and genomic architecture in the adaptive radiations of African cichlid fishes. Frontiers in Genetics, **5**, 163.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL et al. (2016) The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. Database Issue **44**, D279-D285.

Foeller K, Stelbrink B, Hauffe T, Wilke, T (2015) Constant diversification rates of endemic gastropods in ancient Lake Ohrid: ecosystem resilience likely buffers environmental fluctuations. Biogeosciences, 12: 7209-7222.

Franchini P, Fruciano C, Spreitzer ML, Jones JC, Elmer KR, Henning F, Meyer A (2014) Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. Molecular Ecology, **23**, 1828-1845.

Author Manuscript

This article is protected by copyright. All rights reserved

Fryer G (1991) Comparative aspects of adaptive radiation and speciation in Lake Baikal and the Great Rift lakes in Africa. Hydrobiologia, 21, 137–146.

Genner MJ, Todd JA, Michel E, Erpenbeck D, Jimoh A, Joyce DA Piechocki A, Pointier JP (2007) Amassing diversity in an ancient lake: evolution of a morphologically diverse parthenogenetic gastropod assemblage in Lake Malawi. Molecular Ecology, 16: 517-530.

Goto A, Yokoyama R, Sideleva VG (2015) Evolutionary diversification in freshwater sculpins (Cottoidea): a review of two major adaptive radiations. Environ Biol Fish, **98**, 307-335. DOI 10.1007/s10641-014-0262-7

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol, **29**, 644-652. doi: 10.1038/nbt.1883.

Guindon S, Gascuel O (2003) A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Systematic Biology, **52**, 696-704.

Hamilton WJ, III, Orians GH (1965) Evolution of brood parasitism in altricial birds. The Condor, 67: 361-382.

Hausdorf B, Ropstorf P, Riedel F (2003) Relationships and origin of endemic Lake Baikal gastropods (Caenogastropoda: Rissoidea) based on mitochondrial DNA sequences. Molecular Phylogenetics and Evolution, **26**, 435-443.

Höhna S (2015) The time-dependent reconstructed evolutionary process with a key-role for mass-extinction events. J. Theor. Biology, **380**, 321-331.

Hou Z, Sket B, Fišer C, Li S (2011) Eocene habitat shift from saline to freshwater promoted Tethyan amphipod diversification. Proc. Natl. Acad. Sci. USA, **108**, 14533–14538.

Hou Z, Sket B (2016) A review of Gammaridae (Crustacea: Amphipoda): the family extent, its evolutionary history, and taxonomic redefinition of genera. Zool. J. Linnean Soc., **176**, 323-348.

Hunt DM, Fitzgibbon J, Slobodyanyuk SJ, Bowmaker JK, Dulai KS (1997) Molecular evolution of the cottoid fish endemic to Lake Baikal deduced from nuclear DNA evidence. Molecular Phylogenetics and Evolution, **8**, 415-422.

Author Manuscript

This article is protected by copyright. All rights reserved

Jeffares DC, Tomiczek B, Sojo V, dos Reis M (2015) A beginners guide to estimating the non-synonymous to synonymous rate ratio of all protein-coding genes in a genome. Methods Mol Biol., **1201**, 65-90. doi: 10.1007/978-1-4939-1438-8_4.

Jeffery NW, Yampolsky LY, Gregory TR. (2016) Nuclear DNA content correlates with depth, body size, and diversification rate in amphipod crustaceans from ancient Lake Baikal, Russia. Genome, in press.

Kamaltynov RM (1999) On the higher classification of Lake Baikal amphipods. Crustaceana (Leiden), **72**, 933-944.

Karabanov E, Williams D, Kuzmin M, Sideleva V, Khursevich G, Prokopenko A et al. (2004) Ecological collapse of Lake Baikal and Lake Hovsgol ecosystems during the Last Glacial and consequences for aquatic species diversity. Palaeogeography Palaeoclimatology Palaeoecology, **209**, 227-243.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol., **30**, 772-780.

Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, Seehausen O (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular Ecology, **22**, 2848–2863.

Kim J, Sanderson MJ (2008) Penalized likelihood phylogenetic inference: bridging the parsimony-likelihood gap. Systematic Biology, **57**, 665-674.

Kocher T D (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nature Rev. Genetics, **5**, 288-298.

Kocher TD, Conroy JA, McKaye KR, Stauffer JR (1993) Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to convergence. Molecular Phylogenetics and Evolution, 2, 158–165.

Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A (2008) Patterns of positive selection in six Mammalian genomes. PLoS Genetiscs, **4**, e1000144. doi: 10.1371/journal.pgen.1000144.

Kozhova OM, Izmest'eva LR (1998) Lake Baikal: Evolution and biodiversity. Backhuys Publ., Leiden.

Author Manuscript

This article is protected by copyright. All rights reserved

Kroll O, Hershler R, Albrecht C, Terrazas EM, Apaza R, Fuentealba C, Wolff C, Wilke, T (2012) The endemic gastropod fauna of Lake Titicaca: correlation between molecular evolution and hydrographic history. Ecology and Evolution, 2: 1517-1530.

Li L, Stoeckert CJJr, Roos DS (2003) OrthoMCL: Identification of ortholog groups for eukaryotic genomes. Genome Research, **13**, 2178-2189.

Losos JB (2010) Adaptive radiation, ecological opportunity, and evolutionary determinism. American Naturalist, **175**, 623-639.

MacDonald KS, Yampolsky L, Duffy JE (2005) Molecular and morphological evolution of the amphipod radiation of Lake Baikal. Molecular Phylogenetics and Evolution, **35**, 323–343.

Marijnissen SAE, Michel E, Daniels SR, Erpenbeck D, Menken SBJ, Schram FR (2006) Molecular evidence for recent divergence of Lake Tanganyika endemic crabs (Decapoda : Platythelphusidae). Molecular Phylogenetics and Evolution, 40: 628-634.

Martens, K (1997) Speciation in ancient lakes. Trends in Ecology & Evolution 12: 177-182.

Martin CH, Feinstein LC (2014) Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. Molecular Ecology, **23**, 1846-1862.

Martin CH, Wainwright PC (2013) Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. Science, **339**, 208-211. doi:10.1126/science.1227710

Mats VD, Shcherbakov DYu, Efimova IM (2011) Late Cretaceous–Cenozoic history of the Lake Baikal depression and formation of its unique biodiversity. Stratigraphy and Geological Correlation, **19**, 404–423.

Meixner MJ, Lueter C, Eckert C, Itskovich V, Janussen D, von Rintelen T et al. (2007) Phylogenetic analysis of freshwater sponges provide evidence for endemism and radiation in ancient lakes. Molecular Phylogenetics and Evolution, **45**, 875-886.

Melnik NG, Timoshkin OA, Sideleva VG, Pushkin SV, Mamylov VS (1993) Hydroacoustic measurement of the density of the Baikal macrozooplankter Macrohectopus branickii. Limnology And Oceanography, 38: 425-434.

Author Manuscript

This article is protected by copyright. All rights reserved

Meyer A (1993) Phylogenetic relationships and evolutionary processes in east African cichlid fishes. Trends in Ecology and Evolution, 8, 279–284.

Meyer A, Kocher TD, Basasibwaki P, Wilson AC (1990) Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature, **347**, 550-553.

Moore BR, Hohna S, May MR, Rannala B, Huelsenbeck JP. (2016) Critically evaluating the theory and performance of Bayesian analysis of macroevolutionary mixtures. Proc. Natl. Acad. Sci. U.S.A. 113:9569-9574.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward et al. (2015) Gene-wide identification of episodic selection. Mol Biol Evol., **32**, 1365-1371. doi: 10.1093/molbev/msv035.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond S. L (2012) Detecting individual sites subject to episodic diversifying selection. PLoS Genetics, **8**, e1002764.

Naumenko S (2015) Iterative Target Contig Assembler (iTCA). Available: http://itas2015.iitp.ru/pdf/1570161515.pdf.

Novikova OA, Naumova TV, Timoshkin OA (2006) Karyotypes and current approaches to the systematics of endemic Baikal representatives of *Bdellocephala* genus (Turbellaria Dendrocoelidae). Hydrobiologia, **568**, Suppl. 1, 183-191.

Okazaki K, Hummer G (2015) Elasticity, friction, and pathway of γ-subunit rotation in FoF1-ATP synthase. Proc Natl Acad Sci U S A, **112**, 10720-10725.

Pyle MN, Adamowicz SJ (2015) Mode and tempo of diversification of Hyalella (Crustacea: Amphipoda) in ancient Lake Titicaca (2015) Genome, 58: 268-268.

Rabosky DL, Lovette IJ (2008) Explosive evolutionary radiations: Decreasing speciation or increasing extinction through time? Evolution, **62**, 1866-1875.

Rivarola-Duarte L, Otto C, Juehling F, Schreiber S, Bedulina D, Jakob L et al. (2014) A first glimpse at the genome of the Baikalian amphipod *Eulimnogammarus verrucosus*. J. Exper. Zool. B: Molecular and Developmental Evolution, **322**, 177-189.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) MrBayes 3.2: efficient Bayesian

Author Manuscript

This article is protected by copyright. All rights reserved

phylogenetic inference and model choice across a large model space. Syst Biol. 61: 539-542.

Rüber L, Verheyen E, Meyer A (1999) Replicated evolution of trophic specializations in an endemic cichlid fish lineage from Lake Tanganyika. Proceedings of the National Academy of Sciences of the United States of America, 96, 10230–10235.

Salzburger W (2009) The interaction of sexually and naturally selected traits in the adaptive radiation of cichlid fishes. Molecular Ecology, 18, 169–185.

Scaion D, Belhomme M, Sébert P (2008) Pressure and temperature interactions on aerobic metabolism of migrating European silver eel. Respiratory Physiology & Neurobiology, **164**, 319–322.

Schluter D (2000) The Ecology of Adaptive Radiation. Oxford University press, Oxford.

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics, **27**, 863-864.

Sébert P, Theron M, Vettier A (2004) Pressure and temperature interactions on cellular respiration: a review. Cellular and Molecular Biol., **50**, 491-500.

Sherbakov DY, Kamaltynov RM, Ogarkov OB, Verheyen E (1998) Patterns of evolutionary ehange in Baikalian gammarids inferred from DNA sequences (Crustacea, Amphipoda). Molecular phylogenetics and evolution, **10**, 160–167.

Sherbakov DY (1999) Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. Trends in Ecology & Evolution, 14: 92-95.

Simpson GG (1990) Splendid Isolation: The Curious History of South American Mammals. Yale University Press, New Haven CT.

Sitnikova TY (1994) Recent views on the history and diversity of the Baikalian malacofauna. Ergebnisse der Limnologie, **44**, 319-326.

Sitnikova TYa (2006) Endemic gastropod distribution in Baikal. Hydrobiologia, **568**, Suppl. 1, 207-211.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol., **32**, 1342-1353. doi: 10.1093/molbev/msv022.

Author Manuscript

This article is protected by copyright. All rights reserved

Somero GN (1992) Adaptations to high hydrostatic pressure. Annu. Rev. Physiol., 54, 557-577

Sowinsky VK (1915) Amphipoda ozera Baikala (Sem. Gammaridae). Zoologicheskiye issledovaniya ozera Baikala. IX. St Vladimir Univ Press, Kiev.

Stamatakis A (2014) RAxML Version 8: A tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics, **30**, 1312-1313. doi: 10.1093/bioinformatics/btu033

Stebbing TRR (1906) Amphipoda I. Gammaridea. In: Dar Tierreich. Verlag von R. Friedländer und Sohn, Berlin.

Stock JH (1969) Members of Baikal Amphipod genera in European waters, with description of a new species, Eulimnogammarus macrocarpus, from Spain. Proceedings Koninklijke Nederlandse Akademie van Wetenschappen. Serie C: Biologica, **72**, 66-75.

Takhteev VV (2000a) Trends in evolution of Baikal amphipods and evolutionary parallels with some marine malacostracan faunas. Adv. Ecol. Res., **31**, 197–220.

Takhteev VV (2000b) Essays on the amphipods of Lake Baikal (systematics, comparative ecology, evolution). Irkutsk State Univ Press, Irkutsk, Russia. (In Russian.)

Tarasov P, Bezrukova E, Karabanov E, Nakagawa T, Wagner M, Kulagina N et al. (2007) Vegetation and climate dynamics during the Holocene and Eemian interglacials derived from Lake Baikal pollen records. Palaeogeography, Palaeoclimatology, Palaeoecology, **252**, 440–457.

Thomas GW, Hahn MW (2015) Determining the null model for detecting adaptive convergence from genomic data: a case study using echolocating mammals. Mol Biol Evol., **32**, 1232-1236. doi: 10.1093/molbev/msv013.

Väinölä R, Kamaltynov RM (1999) Species diversity and speciation in the endemic amphipods of Lake Baikal: molecular evidence. Crustaceana (Leiden), **72**, 945–956.

Van Damme D, Gautier A (2013) Lacustrine mollusc radiations in the Lake Malawi Basin: experiments in a natural laboratory for evolution. Biogeosciences, 10: 5767-5778.

Author Manuscript

This article is protected by copyright. All rights reserved

van Rijssel JC, Hoogwater ES, Kishe-Machumu MA, van Reenen E, Spits KV, van der Stelt RC et al. (2015) Fast adaptive responses in the oral jaw of Lake Victoria cichlids. Evolution, **69**, 179-189.

Vishnoi A, Kryazhimskiy S, Bazykin GA, Hannenhalli S, Plotkin JB (2010) Young proteins experience more variable selection pressures than old proteins. Genome Res., **20**, 1574-1581.

von Ballmoos C, Wiedenmann A, Dimroth, P (2009) Essentials for ATP synthesis by F1F0 ATP synthases. Annu Rev Biochem., **78**, 649-672. doi: 10.1146/annurev.biochem.78.081307.104803.

von Rintelen T, Glaubrecht M (2006) Rapid evolution of sessility in an endemic species flock of the freshwater bivalve Corbicula from ancient lakes on Sulawesi, Indonesia. Biology Letters, 2: 73-77.

von Rintelen K, Glaubrecht M, Schubart CD, Wessel A, von Rintelen T (2010) Adaptive radiation and ecological diversification of Sulawesi's ancient lake shrimps. Evolution, 64: 3287-3299.

Weiss JD, Cotterill FPD, Schliewen UK (2015) Lake Tanganyika - a 'melting pot' of ancient and young cichlid lineages (Teleostei: Cichlidae). PLoS ONE, **10**, e0125043.

Wellborn GA, Langerhans RB (2015) Ecological opportunity and the adaptive diversification of lineages. Ecology and Evolution, **5**, 176-195.

Wernersson R, Pedersen AG (2003) RevTrans: Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res., **31**, 3537-3539.

Williams DF, Peck J, Karabanov EB, Prokopenko AA, Kravchinsky V, King J, Kuzmin MI (1997) Lake Baikal record of continental climate response to orbital insolation during the past 5 million years. Science, 278: 1114-1117.

Wysocka A, Grabowski M, Sworobowicz L, Mamos T, Burzynski A, Sell J. (2014) Origin of the Lake Ohrid gammarid species flock: ancient local phylogenetic lineage diversification. Journal of Biogeography, 41: 1758-1768.

Yampolsky LYu, Kamaltynov RM, Filatov DA, Ebert D, Chernykh VI (1994) Variation of allozyme loci in endemic gammarids of Lake Baikal. Biol. J. Linn. Soc., **53**, 309-329.

Author Manuscript

This article is protected by copyright. All rights reserved

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol., **24**, 1586-1591.

Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe W et al. (2010) Ecological opportunity and the origin of adaptive radiations. J. Evol. Biol., **23**, 1581-1596. DOI: 10.1111/j.1420-9101.2010.02029.x

Yoshii K (1999) Stable isotope analyses of benthic organisms in Lake Baikal. Hydrobiologia, 411: 145-159

Zemskaya TI, Sitnikova TY, Kiyashko SI, Kalmychkov GV, Pogodaeva TV, Mekhanikova IV et al. (2012) Faunal communities at sites of gas- and oil-bearing fluids in Lake Baikal. Geo-Marine Letters, **32**, 437-451.

Zimmermann B, Diez M, Börsch M, Gräber P (2006) Subunit movements in membrane-integrated EF0F1 during ATP synthesis detected by single-molecule spectroscopy. Biochim Biophys Acta, **1757**, 311-319.

Table 1. Summary of episodes of positive selection detected using BUSTED (positive selection anywhere in the phylogeny; Murrell et al. 2014) and aBSREL (positive selection in a specific branch; Smith et al. 2015) models applied to either whole-phylogeny datasets with 100% or >75% occupancy or to each clade separately. Values in cells represent the total number of inferred episodes (FDR<0.1) and the same number per gene per unit branch length (nucleotide substitution per site). SD among individual clades in parentheses.

Dataset	Episodes of positive selection			
	BUSTED		aBSREL	
	Total	Per gene per branch length	Total	Per gene per branch length
All species, 175 COGs, 100% occupancy	33	0.101	20	0.061
All species, 523 COGs, >66% occupancy	101	0.077	56	0.043

Author Manuscript

This article is protected by copyright. All rights reserved

Each clade (Fig.1) separately, variable number of COGs per clade (Suppl.Table 7), 100% occupancy	115	0.112 (0.055)	124	0.117 (0.031)

Table 2. Annotations of 33 proteins with evidence for positive selection reported by either BUSTED or aBSREL (FDR <0.1), out of the 175 proteins in the analyzed data set. For aBSREL, each protein also survived the 0.05 Bonferroni correction for multiple tests over all branches in the phylogeny. If both methods detect positive selection the higher of the two FDRs is shown. To one exception (ancestral node of *Macrohectopus branickii* clade), all episodes identified to a branch were found in terminal branches. Mitochondrial respiratory chain and ATP synthesis proteins are shown in bold.

#	COG ID	Description	FDR	Detected by BUSTED	Detected by aBSREL	Detected by aBSREL in:
1	cog006728	**NADH dehydrogenase**	0	+		
2	cog006711	profilin	0	+	+	Ancestor of *Macrohectopus branickii* clade 2 parallel ancestral branches of *Eulimnogammarus* clade, *Heterogammarus sophianosii*, *Eucarinogammarus wagii*
3	cog006775	**NADH dehydrogenase**	0	+		
4	cog006681	**mitochondrial ATP synthase f chain**	1.5e-14	+	+	*Baikalogammarus pullus*, *Macrohectopus branickii*, *Micruropus vortex*
5	cog006814	**ATP synthase gamma subunit mitochondrial**	1.2e-11	+	+	*Macrohectopus branickii*
6	cog006708	**NADH dehydrogenase**	1.5e-10	+	+	*Poekilogammarus pictoides*
7	cog006909	**adenine nucleotide translocase**	8.7e-9	+		
8	cog006784	*mitochondrial*	6.6e-7	+		

Author Manuscript

This article is protected by copyright. All rights reserved

		ribosomal protein s7				
9	cog006777	high mobility group protein dsp1-like	9.2e-7	+		
10	cog006925	peptidylprolyl isomerase	0.001	+		
11	cog006781	*probable 39s ribosomal protein mitochondrial*	0.001	+		
12	cog006769	**cytochrome b-c1 complex subunit mitochondrial**	0.001	+		
13	cog006705	*ribosomal protein l13*	0.001	+		
14	cog006773	agap004964-pa-like protein	0.002	+		
15	cog006908	*60s ribosomal protein l27a*	0.003	+		
16	cog006720	*60s ribosomal protein l21*	0.008	+		
17	cog006825	DNA-directed RNA polymerase III subunit rpabc2	0.014		+	*Pachyschesis branchialis*
18	cog006876	*40s ribosomal protein s24*	0.014	+	+	*Poekilogammarus pictoides*
19	cog006836	myosin light chain	0.014	+	+	*Echiuropus macronychus*
20	cog006702	cre-rpl-17 protein	0.014	+	+	*Echiuropus macronychus, Micruropus parvulus*
21	cog006786	translocating chain-associated membrane protein 1-like	0.014	+	+	*Axelboeckia carpenteri*
22	cog006767	aldo-keto	0.019	+		

Author Manuscript

This article is protected by copyright. All rights reserved

		reductase				
23	cog006691	*28s ribosomal protein mitochondrial*	0.019	+	+	*Echiuropus macronychus*
24	cog006797	*s10e ribosomal protein*	0.022	+	+	*Eulimnogammarus similis*
25	cog006717	viral iap-associated factor	0.043	+		
26	cog006698	---NA---	0.044	+	+	*Micruropus parvulus*
27	cog006778	proliferating cell nuclear antigen	0.047	+		
28	cog006893	*60s ribosomal protein l35a*	0.048	+		
29	cog006844	**succinyl-ligase**	0.076	+	+	*Macrohectopus branickii*
30	cog006715	*40s ribosomal protein s3-like*	0.077	+		
31	cog006757	*60s ribosomal protein l10*	0.085	+		
32	cog006779	small nuclear ribonucleoprotein sm d3	0.086	+		
33	cog006846	heat shock protein	0.089	+		

Figure Legends

Fig. 1 Partial transcriptome-based phylogeny (175 COGs; ~50K sites) of two endemic radiations of gammarid amphipods in Lake Baikal, *Micruropudae* (bracket 1) and *Acanthogammaridae* (bracket 2). Numbers correspond to bootstrap values, with only values <100 shown. Outgroups (bold underlined): Caspian sea (*Pandorites podoceroides*) and palaearctic freshwater (*Gammarus lacustris*) species. Clades analyzed individually are marked by different colors. Species with body armor are marked with black diamonds, and species with presumed secondary loss of body armor, with open diamonds. A, abyssal species; Pl, plankton species; Pr, (semi)parasitic species. Select Baikal species are

Author Manuscript

This article is protected by copyright. All rights reserved

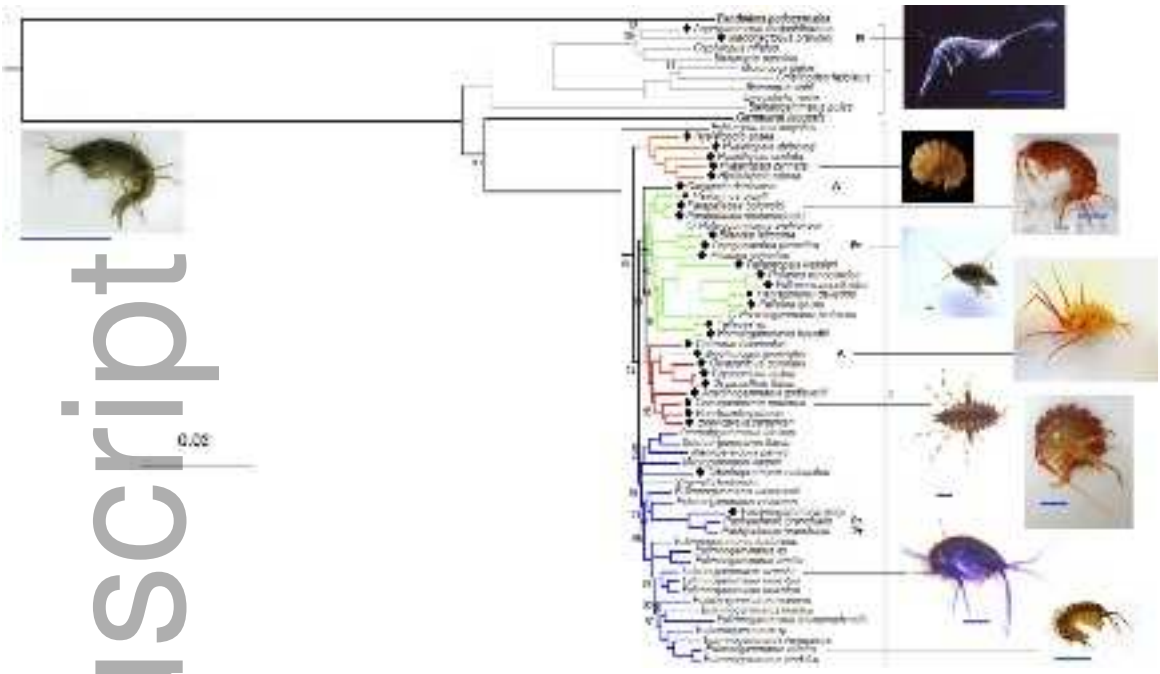
shown on the right; *Gammarus lacustris* shown as an inset. Blue bars: 1 cm. Red bars: 1 mm. Scale: nucleotide substitutions per nucleotide site. Image credits: *Macrohectopus branickii*: Dr. B.Sket/University of Ljubljana; *Acanthogammarus maximus*: Dr. G. Chapelle / Royal Belgian Institute of Natural Sciences; *Gammarus lacustris* and *Eulimnogammarus vittatus*: E. Kondrateva; *Parapallasea borowskii*: K. Woo.

Fig. 2. Number of lineages and positive selection signals in the studied subset of Baikal amphipods species flock lineages plotted vs. relative node age, $-\log_{10}$ scale (i.e., time flows to the right; 0 corresponds the time of divergence from the Caspian sea outgroup). A: number of Baikal lineages in the whole dataset (black) and in individual clades (colors on A and B matching color-coded clades on Fig.1). Dotted line: extrapolation based on diversification rate estimated at the root of the 2nd amphipod invasion into Baikal. B: occurrence of episodes of positive selection vs. branching time span binned into bins with equal total branches length (number of episodes per gene per 1 nucleotide substitution). Open circles: non-Baikal branches. Color symbols: baikalian clades analyzed separately. Gray and red vertical dashed lines show branching times of the first (*Micruropidae*) and the second (*Acanthogammaridae s.l.*) amphipod invasions into Baikal, respectively. Linear regression lines draw through all points (dotted line; regression coefficient R=-0.34 ±0.15; t=-2.26; $p<0.035$) and through points corresponding to the second Baikal invasion (solid line; regression coefficient R=-0.54 ±0.13; t=-4.22; $p<0.0005$.). Spearman's rank correlations were $\rho = -0.44$; $P<0.032$ and $\rho = -0.78$; $P<0.0001$ for both invasions and for the second invasion alone, respectively. Spearman's rank correlation was also significant for the *Acanthogammarus* clade (red dots) analyzed separately ($\rho = -1$; $P<0.001$).

Author Manuscript

This article is protected by copyright. All rights reserved

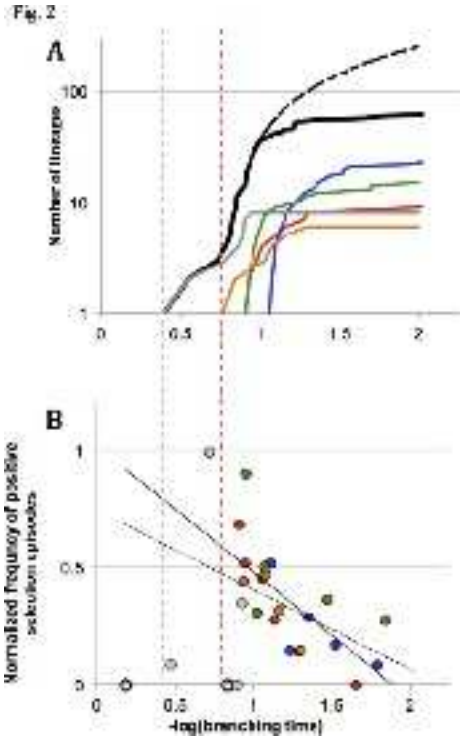
Author Manuscript



mec_13927_f1.tiff

This article is protected by copyright. All rights reserved

Author Manuscript



mec_13927_f2.tiff

This article is protected by copyright. All rights reserved