Excessive parallelism in protein evolution of Lake Baikal amphipod species flock

Valentina Burskaia1*, Sergey Naumenko2, Mikhail Schelkunov1,2, Daria Bedulina4,5, Tatyana Neretina2,6,7, Alexey Kondrashov7,8, Lev Yampolsky9, Georgii A. Bazykin1,2

*Author for Correspondence: Valentina Burskaia, Center of Life Sciences, Skolkovo Institute of Science and Technology, Moskow, Russia, +7 916 711 33 41, valentina.burskaia@skolkovotech.ru

1. Skolkovo Institute of Science and Technology, Ulitsa Nobelya, 3, Moscow, Moscow Oblast, 121205, Russia
2. Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevitch Institute), Bolshoy Karetny per. 19, build.1, Moscow, 127051, Russia
3. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Landmark 401 Park Drive 3 East, Boston MA 02115, USA
4. Institute of Biology, Irkutsk State University, Lenin str. 3, Irkutsk, 664003, Russia
5. Baikal Research Centre, Lenin str. 21, Irkutsk, 664003, Russia
6. N.A.Pertsov White Sea Biological Station, Lomonosov Moscow State University, Primorskiy, 184042
7. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University,

© The Author(s) 2020. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Repeated emergence of similar adaptations is often explained by parallel evolution of underlying genes. However, evidence of parallel evolution at amino acid level is limited. When the analyzed species are highly divergent, this can be due to epistatic interactions underlying the dynamic nature of the amino acid preferences: the same amino acid substitution may have different phenotypic effects on different genetic backgrounds. Distantly related species also often inhabit radically different environments, which makes the emergence of parallel adaptations less likely. Here, we hypothesize that parallel molecular adaptations are more prevalent between closely related species. We analyze the rate of parallel evolution in genome-size sets of orthologous genes in three groups of species with widely ranging levels of divergence: 46 species of the relatively recent lake Baikal amphipod radiation, a species flock of very closely related cichlids, and a set of significantly more divergent vertebrates. Strikingly, in genes of amphipods, the rate of parallel substitutions at nonsynonymous sites exceeded that at
synonymous sites, suggesting rampant selection driving parallel adaptation. At sites of parallel substitutions, the intraspecies polymorphism is low, suggesting that parallelism has been driven by positive selection and is therefore adaptive. By contrast, in cichlids, the rate of nonsynonymous parallel evolution was similar to that at synonymous sites, while in vertebrates, this rate was lower than that at synonymous sites, indicating that in these groups of species, parallel substitutions are mainly fixed by drift.

Key words: parallelism, convergence, speciation, positive selection.

Introduction

Parallel evolution of genes under similar selection pressure is widely observed in populations of the same species. Experimental bacterial evolution (Woods et al. 2006, Baym et al. 2016) and recurrent adaptation of pathogens to their hosts (Coll et al. 2018, Collins and Didelot 2018) indicate that similar substitutions can reach fixation in many populations independently. In experimental evolution, similar selection pressures emerge as the main factor determining parallel evolution (Bailey et al. 2015). There is also evidence of parallel molecular adaptation in multicellular organisms, both as a result of fixation of existing alleles (Soria-Carrasco et al. 2014, Terekhanova et al. 2014) and of recurrent origin of adaptive mutations (Karasov et al. 2010, Kreiner et al. 2019, Marcovitz et al. 2019). In genome- or transcriptome-level analyses, parallelism can be observed at the level of pathways, genes or, sometimes, specific genomic sites (Lim et al. 2019, Marcovitz et al. 2019). For example, parallel evolution of marine mammals has involved repeated loss or change in the rate of evolution of orthologous genes (Chikina et al 2016); and specific parallel amino acid substitutions have been reported in
genes including hemoglobins (Natarajan et al. 2016) and taste receptors (Baldwin et al. 2014).

Nevertheless, not all parallelism is adaptive. To claim adaptation, the level of parallelism needs to exceed that expected neutrally. However, at the whole-genome level, the scans for parallel nonsynonymous substitutions revealed no evidence for higher-than-neutral levels of molecular parallelism (Zou and Zhang 2015a; Zou and Zhang 2015b, Thomas and Hahn 2015, Bazykin et al. 2007, Foote et al. 2015), making it hard to distinguish the adaptive parallel mutations (if present) against the neutral background.

The frequency of parallel amino acid substitutions between two lineages (relative to the neutral expectation) tends to decrease with genetic distance between them (Usmanova et al. 2015, Zou and Zhang 2015b, Conte et al. 2012). This suggests that closely related lineages share amino acid propensities as a result of common constraints (negative selection) due to similarity of environmental pressure or, perhaps more importantly, genetic background at epistatically interacting sites (Kryazhimskiy et al. 2014, Storz 2016). The elevated rate of parallel evolution in closely related species can be erroneously interpreted as evidence for parallel adaptation (positive selection) if adaptive constraints are not accounted for properly (Klink et al. 2017a,b).

To compare the abundance of parallel substitutions driven by selection with neutral expectations, one can extend the conventional dN/dS-type approach, namely, estimate the relative rates of nonsynonymous and synonymous parallel evolution. Under purely neutral evolution, a site that has experienced a substitution between a pair of species can only be neutral, and if the fitness landscape is invariant, an identical substitution at this site between other two species is also expected to occur at the neutral rate. Deviation from this expectation could occur due to weak selection preventing substitutions within one of the pairs of species, or due to changing amino acid preferences between pairs (Bazykin et al. 2007); either trend will cause
parallel nonsynonymous substitutions to be less frequent than the synonymous control. Conversely, nonsynonymous parallel substitutions could be more frequent than synonymous ones if the parallel evolution is adaptive, i.e., driven by positive selection.

A genome-wide analysis of three quartets of species of vertebrates, insects, and fungi has shown that the rate of parallelism at non-synonymous sites is lower than that at synonymous sites. This has been interpreted as evidence for weak negative selection and/or change in single-position fitness landscape between species (Bazykin et al. 2007), consistent with findings using other logic (Bazykin 2015).

Here, we apply this approach across many genome comparisons, asking how the amount of parallel amino acid evolution depends on the divergence level between the considered species. For this purpose, we use our recently published dataset of transcriptomes of the species flock of closely related baikalian amphipods (Naumenko et al. 2017), as well as two other datasets: cichlid fishes from the lake Malawi species flock, and a group of more distantly related vertebrates. In the absence of adaptation, the rate of parallel nonsynonymous to synonymous evolution should be less or equal to one. Strikingly, in the amphipod dataset, we find nonsynonymous parallel substitutions to be more frequent than synonymous ones, suggesting prevalent selection in favor of the same derived variants in different species. A similar, although weaker, effect was found in closely related cichlid fish dataset. In amphipods, within-species polymorphism at sites of past parallel nonsynonymous substitutions is low, indicating that these substitutions were driven by positive selection. By contrast, in a dataset of distantly related vertebrates, the rate of nonsynonymous parallel substitutions is lower than that of synonymous ones, consistent with prevalent negative selection.
Methods

Divergence data

The three datasets were analyzed as follows. First, we used the transcriptomic sequences of closely related amphipod species from Lake Baikal (Naumenko et al. 2017). Of the 67 species analyzed in that work, we picked the 46 species for which the sequenced sample was based on exactly one individual. Orthologous groups of genes were calculated with OrthoMCL 2.0.9 with the inflation parameter set to 1.5 (Li 2003). If a particular species carried multiple paralogous sequences of a gene, this species was excluded from the analysis of this gene. Codon-aware alignments for orthogroups were obtained with TranslatorX (Abascal et al. 2010) using the Muscle method (Edgar 2004). Poorly aligned sequences were detected and removed from the alignments using the following rule:

1) A column in an alignment was considered "good" if it carried the same nucleotide in at least 50% of species;

2) Sequences for which fewer than 50% positions were "good" were removed from the alignment.

This exclusion process was performed using TrimAl 1.4 (Capella-Gutierrez et al. 2009). It resulted in 4366 orthologous groups of genes. Alignments for all genes were concatenated, and a phylogenetic tree was reconstructed using RAxML 8.1.20 (Stamatakis 2014) with GTR+Gamma model, 20 starting maximum parsimony trees and 100 bootstrap analysis pseudoreplicates. As mutations in the third positions of codons are often synonymous, the third positions of codons accumulate substitutions quicker than the first two. Therefore, we used partitioning, with separate substitution models for the first two and for the third codon positions.
The obtained tree (Figure 1A) was similar to that obtained previously (Naumenko et al. 2017).

For the cichlid species flock from lake Malawi, exon alignments were extracted from genomic data of 62 species each mapped onto the assembly of the *Maylandia zebra* (Boulenger, 1899) (assembly ID = MetZeb1.1_prescreen) (Malinsky et al. 2018). Using this annotation (available at the Cambridge cichlid browser), we picked the longest isoform of each gene, for a total of 15318 transcripts. As the phylogenetic tree for this set of species, we used the Maximum Clade Credibility phylogenetic tree from the original paper (Malinsky et al. 2018).

For the analysis of highly divergent vertebrates, we used the exon alignments of 100 species to the hg19 human genome. All together, 21520 gene alignments were used. These data were fetched from the UCSC database (Karolchik et al. 2007) together with the corresponding tree.

**Calculation of $P$ statistic for a species quartet**

To compare the amount of amino acid-level parallelism to that expected neutrally while controlling for the heterogeneity of mutation rates between genomic sites, and to make these values comparable between groups of species, we used our approach developed previously (Bazykin et al. 2007). In brief, we rely on quartets of species with a particular phylogenetic relationship, namely, composed of two clades (‘pairs’) each involving two species (Figure 2A); and consider positions orthologous between these four species. A difference between the two species of a pair implies that a substitution has occurred in at least one of the two lineages leading to these species from their LCA, even though the direction of this substitution can be unknown. We use one pair (‘pair I’) to identify the sites that had experienced such a substitution, and ask if the same substitution has occurred in parallel in the other pair (‘pair II’). We assume parsimony, i.e., that at most one substitution has occurred between the two species within a pair;
violation of this assumption may lead to underestimation of the rate of parallelism. Pooling across different sites with substitutions in pair I, we can infer the rate of parallelism by measuring the fraction of sites at which the same substitution has also occurred between the species of pair II.

More precisely, at a site with a difference between the two species of pair I, four possible patterns can be observed in a quartet: (i) ((A,B)(A,B)), (ii) ((A,B)(B,A)), (iii) ((A,B)(A,A)) and (iv) ((A,B)(B,B)). Here, in each of the four cases, the first bracket represents the two species belonging to pair I, while the second bracket represents the two species belonging to pair II; identical letters signify identical nucleotides. The cases of mutation into a different (non-A, non-B) nucleotide in pair II were not considered. Like in Bazykin et al. 2007, we also exclude invariant sites ((A,A)(A,A)) and ((B,B)(B,B)); considering those sites would make our estimated additionally dependent on the rate of substitutions in path II, and harder to interpret.

Patterns (i) and (ii) correspond to parallel evolution. As a scaleless estimate of the rate of parallelism for a given category of sites and substitutions, we measure the fraction of sites, among those with a substitution in pair I, that also experienced a substitution in pair II:

\[
d_p = \frac{N_{(A,B)(A,B)} + N_{(A,B)(B,A)}}{N_{(A,B)(A,B)} + N_{(A,B)(B,A)} + N_{(A,B)(B,A)} + N_{(A,B)(B,B)}}
\]

where \(N\) corresponds to the number of sites with the corresponding pattern.

\(d_p\) can be calculated for any class of sites and/or substitutions. All 4-fold degenerate sites were used to estimate the rate of synonymous parallel substitutions \(dS_p\), while all non-degenerate sites were used to estimate the rate of nonsynonymous parallel substitutions \(dN_p\). To account for the differences in mutation rates between the six single-nucleotide mutation types, namely,
A↔C, A↔T, A↔G, C↔T, C↔G and T↔G, we calculated $d_{SP}$ and $d_{NP}$ for each such mutation type separately.

Finally, we use the synonymous sites to estimate the level of parallelism expected neutrally, and calculate $P = \frac{d_{NP}}{d_{SP}}$, i.e., the ratio of the rates of parallel nonsynonymous and synonymous substitutions. For Figure 2B, the values of $P$ were averaged across the 6 mutation types. $P=1$ implies that the substitutions that occurred in pair I also occur in pair II at the neutral rate; while deviations from 1 imply selection on parallel nonsynonymous substitutions.

Between the two species of pair I in each quartet, we also calculated the numbers of nonsynonymous differences at non-degenerate sites, and of synonymous differences in four-fold degenerate sites, per such site across all codon sites of the alignment. We calculated $\frac{d_{N}}{d_{S}}$, as the ratio of these values.

Choice of quartets and filtering

For each of the three datasets, amphipods, cichlids and vertebrates, we randomly assembled 300 quartets of species (300 for amphipods, 300 for cichlids and 300 for vertebrates) with the tree topology shown in Figure 2A. For cichlids, we were concerned that the incongruence between gene trees constructed from different genes is rather high (Malinsky et al. 2018), which could make the inference of tree topology erroneous. Nevertheless, the nodes separating the six ecological groups are stable (Malinsky et al. 2018). Therefore, to ensure that unstable tree topology does not affect our results, we additionally constructed 300 quartets of cichlid species by picking two species from one ecological group and two from another, ensuring
that the last common ancestor (LCA) of the two pairs was older than the LCA of each pair (Figure S3).

The orthologous nucleotide sites of the four species within a quartet were then filtered for data quality. We required the site to be surrounded by 20 nucleotide sites (10 to the left and 10 to the right) with no gaps in any of the four species of the quartet, and by 2 nucleotide sites (1 to the left and 1 to the right) with no substitutions in any of the four species of the quartet. We also required the upstream site to be non-C, and the downstream, to be non-G, to avoid the potential biases associated with the hypermutable CpG context. Finally, to ensure local alignment quality, we required the codon carrying the analyzed site to carry no nucleotide substitutions, and to be surrounded by 20 codons (10 to the left and 10 to the right) carrying in total no more than 3 nonsynonymous substitutions, between any of the species of the quartet.

Some quartets yielded very few parallel substitutions for some mutation types, making the estimates of $P$ unreliable. To account for this, we only retained a quartet if the number of both synonymous and nonsynonymous parallel substitutions exceeded five for each of the 6 mutation types. This filtering retained 195 out of 300 quartets of amphipods, 147 of 300 quartets of vertebrates, and all 300 quartets of cichlids. Filtering and exclusion of quartets did not radically affect the results of the $P$ test (Figure S4).

Comparing the numbers of parallel and divergent substitutions

To visualize the relationships between parallel and divergent evolution in amphipods (Figure 1B,C), we compared, for each quartet, the number of sites with the identical (((A,B) (A,B)) or ((A,B)(B,A))) and different (((A,B)(A,C)), ((A,B)(C,A)), ((A,B)(B,C)), ((A,B)(C,B)), or ((A,B)(C,D)), where C and D are distinct non-A, non-B nucleotides) nucleotide substitutions between species in Path I and in Path II.
Validation by Sanger sequencing

The primers used are listed in Table S9. Purified PCR products were bidirectionally sequenced on an ABI 3500 Genetic Analyzer (Applied Biosystems) using the BigDye Terminator v 3.1 Cycle Sequencing Kit (Applied Biosystems) and the same primers as for PCR.

Polymorphism data

We used two sources of data on within-species polymorphism in amphipods at sites that had experienced a parallel pair of substitutions.

First, using TopHat (Trapnell 2009), we remapped the Illumina sequencing reads (Naumenko et al. 2017) corresponding to each sample back onto the assembly of that sample, and filtered out positions with quality<10 or with coverage<20. To detect polymorphic sites, i.e., sites heterozygous within the analyzed individual, we performed SNP calling by samtools-1.3.1 (Li, 2011). This analysis is further referred to as “individual-based polymorphism test”.

Second, we used additional pooled transcriptomics data obtained from multiple individuals for one of the studied species, *Eulimnogammarus verrucosus* (Gerstfeldt, 1858). The dataset included 19 samples, each pooled across 4 individuals (Drozdova et al. 2019). The transcriptomics Illumina reads for each sample were mapped onto the reference assemblies of these species, filtering out positions with quality<10 or with coverage<20. To detect polymorphic sites, we performed SNP calling by samtools-1.3.1 (Li, 2011) for each sample individually. We considered a site polymorphic if 8 or more samples were preserved by filtering, and one or more sample was either heterozygous or homozygous with respect to a non-reference allele. This analysis is further referred to as “population-based polymorphism test”.
Polymorphism at sites of a parallel substitution

Using the obtained amphipod SNP data, we estimated, among the sites that had underwent the same parallel A↔B substitution according to the reference genomes (Figure 1A), the fraction of those that also carry both alleles A and B within a single species.

The individual-based polymorphism test was performed on the same 300 quartets of amphipod species that were previously used for the P test. For each quartet, we counted the sites corresponding to the ((A,B)(A,B)) or ((A,B)(B,A)) pattern, and, among those sites, the polymorphic sites were variants A and B were also both present in at least one of the four species. These values were summed over all 300 quartets, and the ratio of these sums is shown in Figure 3B. We compare the thus assessed SNP fractions for the nonsynonymous and synonymous sites (parallel pN/pS statistic).

For the population-based polymorphism test, we repeated the procedure for generating 300 species quartets, but this time required each quartet to include *E. verrucosus*. The fraction of polymorphic sites among the parallel sites was estimated in the same way as for the individual-based polymorphism test (Figure 3C).

To obtain the baseline polymorphism level at sites of a (non-parallel) substitution, we additionally calculated the number of polymorphic and monomorphic sites at positions that underwent a substitution between species of Path I (independently of whether a substitution has occurred in Path II), pooled these numbers over the 300 quartets, and showed the ratio of these sums in Figure 3A. The comparison of these values for the nonsynonymous and synonymous sites yielded the nonparallel pN/pS statistic.
Search for possible phenotypic parallelism

To test for possible dependencies between parallel phenotypic and genotypic changes in amphipods, we randomly selected 40 quartets such that each path included one deepwater and one shallow water species (Table S10). Additionally, we considered detailed phenotypes of species that formed the quartets with the highest and the lowest values of the $P$ statistic (Table S11). All phenotype descriptions are based on (Bazikalova et al, 1945).

Results

Phylogeny of Lake Baikal amphipods

The phylogenetic tree of Lake Baikal amphipods that we obtain based on 4366 orthologs (many of which were found only in a fraction of species) is similar to that obtained previously on the basis of ~175 groups of universal orthologs (Naumenko et al. 2017), indicating that phylogenetic reconstruction is robust. In particular, it confirms that amphipods populated Lake Baikal at least twice (Figure 1A and Figure 5S), corresponding to the two invasion events. Most nodes have bootstrap support above 80.

High rate of parallel nonsynonymous evolution

To illustrate the amount of parallel and divergent evolution in amphipods, we calculate the number of substitutions of these two types in each of the considered quartets; this analysis is similar to that of Castoe et al (2009). We observe a linear dependence between the numbers of divergent and parallel substitutions (Figure 1B,C) reflective of the differences in evolutionary distances between species quartets. We find that the rate of parallel substitutions, relative to that
Fig. 1: Phylogenetics of Lake Baikal amphipods and analysis of parallelism. A: Phylogenetic tree of Lake Baikal amphipods. Dots indicate bootstrap values above 80%. B,C: Parallel vs divergent substitutions in amphipods. Each dot corresponds to one species quartet. B: Quartets were selected only from the closely related group originating from the second invasion. C: Quartets consist of two species from the first invasion and two species from the second invasion; these pairs are genetically more distant. Slopes of linear regression models differ significantly.
(ANOVA test, p-value <2.2e-16) for synonymous and nonsynonymous substitutions at both pictures.

of divergent substitutions, is higher for the nonsynonymous substitutions than for synonymous ones, as indicated by a steeper slope of the regression line for the former.

Although this finding implies an increased rate of parallel evolution among functional positions, this can arise both from parallel adaptation and differences in strength of negative selection among sites (Bazykin et al. 2007, Rokas and Carroll 2008, Povolotskaya and Kondrashov 2010). To distinguish between these alternatives, we next compare the rate of nonsynonymous parallel evolution to that expected neutrally.

Parallel amino acid differences between amphipod species are more frequent than expected neutrally

In genes of compared amphipod species, the $dN/dS$ ratio averaged over all sites and all quartets of compared species (n=195) equaled 0.11, consistent with previous findings (Naumenko et al. 2017) and in line with the universal prevalence of negative selection (Figure 2C). In stark contrast, the ratio of nonsynonymous and synonymous parallel substitutions $P$ equaled 2.17, which implies that the rate of parallel nonsynonymous substitutions exceeded that of parallel synonymous substitutions by 117%. $P$ exceeded 1 for 193 out of the 195 considered quartets of amphipod species (Figure 2B, Table S1). The excess of nonsynonymous parallel, compared to synonymous parallel, substitutions was observed for all types of mutations, and therefore is not due to heterogeneity of mutation rates (Figure S1, Table S1). The value of $P$ was nearly independent of phylogenetic distance between the two species pairs (Table S4).
To confirm that our results are not an artifact of the NGS sequencing technology used, we confirmed a subset of observed parallel sites using Sanger technology. We randomly picked 3 sites of parallel nonsynonymous substitutions for each of the 2 species of amphipod: *Ommatogammarus albinus* (Dybowski, 1874) and *Eulimnogammarus marituji* Bazikalova, 1945 (Table S8). The individuals used for resequencing were different from the ones used for the transcriptomic analysis. We obtained sequences for all 6 samples; in 5 of these 6 cases, the only allele present was the one called for this species, and there was no evidence for the presence of an alternative allele. (In the remaining case, the alternative allele observed in *Eulimnogammarus marituji* coincided with a SNP observed in this species; see below).

To put these results into perspective, we considered other groups of species. Among the 100 species of vertebrates, the value of the $P$ statistic is below 1 for all quartets (mean = 0.40, n=147, Table S3), consistent with previous results (Bazykin et al. 2007; Zou and Zhang 2015b). In this group of species, $P$ declines with phylogenetic distance between species (Figure 2B, Figure S1, Table S6). Among the 100 species of cichlids, $P$ is close to 1 (mean = 1.13, n=300; Figure 2B, Table S2). Similarly to the amphipod sample, it shows almost no dependence on phylogenetic distance (Table S5). The overall $dN/dS$ ratio in vertebrates was below that in amphipods (0.02), and the $dN/dS$ ratio in cichlids, above that in amphipods (0.34) (Figure 2C); the differences in these values may reflect data filtering, differences in effective population sizes leading to differences in efficiency of negative selection against slightly deleterious mutations (Nikolaev et al. 2007), and, perhaps more importantly, an excess of slightly deleterious unfixed variants in comparisons of closely related amphipods.

**High amino acid parallelism shows no clear link with phenotypic parallelism**

Species of Lake Baikal amphipods inhabit depths between 0 and >1500m. Reasoning that
Fig. 2: The \( P \) test for excess parallelism. A: A schematic representation of the calculation of the level of parallelism \( d_P \). In a quartet of species with the depicted topology, green and red colors identify evolutionary paths I and II respectively. \( d_P \) measures the probability that a substitution that has occurred along path I has also occurred along path II. \( P \) is the ratio of \( d_P \) values at nonsynonymous and synonymous sites. B: \( P \) for quartets of cichlids, amphipods and vertebrates. Each point represents the mean value of \( P \) for a species quartet across the 6 mutation types; only those 642 out of the 900 quartets for which sufficient data are available for each mutation type are shown (see Methods). The horizontal axis shows the phylogenetic distance between the last common ancestors of pairs I and II, measured in number of substitutions per nucleotide site (note the logarithmic axis). C: \( dN/dS \) values for two species in Path I, calculated for three groups of species. D: Distribution of values of the \( P \) statistic in all the 300 analyzed quartets (cyan) and in...
the 40 quartets with evidence of parallel phenotypic adaptations to abyssal environment (blue). The mean values of the two distributions are similar (Wilcoxon test, p=0.2709).

Excess parallel evolution is more likely in those species which undergo similar adaptations to the same environment, we hypothesised that those quartets which include both deepwater and shallow water species in each path should have elevated values of the $P$ statistic. However, we found that the data did not support this hypothesis: the values of the $P$ statistic were similar to those of randomly selected 300 quartets (Figure 2D).

In addition, we analyzed the quartets with the highest and the lowest values of the $P$ statistic, among the 300 quartets used in the main test. We hypothesised that the quartets with the high values of the $P$ statistic will possess some phenotypic changes parallel between the two pairs. However, these quartets demonstrated no clear excess of similarity between species of different pairs in any of the considered phenotypic traits (Table S11). Since the excess of the $P$ statistic is rather uniform among amphipod quartets (Figure 2B) and is driven by a large number of genes, it is perhaps unsurprising that most of these parallel events have no obvious phenotypic manifestation.

Among the sites with parallel changes between species, many are polymorphic within species

The observed differences between genomes or transcriptomes of different species could correspond to fixed differences between these species or to SNPs segregating within them. We asked whether the SNPs at sites of nonsynonymous parallel changes between species occur less or more often, compared to the analogous synonymous sites. A deficit of nonsynonymous SNPs
is indicative of negative selection against one of the variants, while an excess of nonsynonymous SNPs can indicate balancing selection maintaining polymorphism at these sites.

To address this, we used within-species polymorphism data from two sources: heterozygous sites within the reference transcriptomes (individual-based polymorphism test) or SNPs detected in population samples of two amphipod species (population-based polymorphism test). For all six possible mutation types, both tests showed that among nonsynonymous parallel sites, a lower fraction carried SNPs, compared to synonymous parallel sites (Figure 3B,D and Table S7).

To put this observation into perspective, we compared this deficit of SNPs at nonsynonymous parallel sites to that observed at nonparallel ones, i.e., those where a single substitution had occurred between the path I species, independent of the similarity between the path II species (Figure 3A,C). The values of the pN/pS statistic were lower at parallel than at nonparallel sites, suggesting stronger negative selection in the former. This was true both for individual-based (Figure 3E) and population-based (Figure 3F) tests.

To validate SNP calling, we included in our sample for Sanger resequencing (see above) one position that was polymorphic in the resequenced species (Eulimnogammarus marituji) according to the NGS data. Resequencing of that position also gave us a polymorphic site, with the same pair of nucleotides at the SNP (A/C; Table S8).

Parallel and non-parallel amino acid substitutions are similar in their properties

Among the 4366 analyzed genes of amphipods, 2514 (57.6%) carried a parallel nonsynonymous substitution in at least one of the species quartets. We were unable to detect any
specific properties of these genes. According to the GO analysis, the set of genes with parallel

Fig. 3: Polymorphism at sites of parallel substitutions. A-D: Proportion of sites carrying a SNP among those with a nucleotide substitution between species in path I (A, C) or among sites with parallel substitutions (B, D). A,B: Individual-based polymorphism test. C,D: Population-based polymorphism test. E, F: Ratio of proportions of sites carrying a SNP between nonsynonymous and synonymous sites (pN/pS). E: Individual-based polymorphism test (paired t-test, p=0.0031; unpaired t-test, p=0.0033). F: Population-based polymorphism test (paired t-test, p=0.0091;
nonsynonymous substitutions was indistinguishable from the remaining genes.

We also asked whether the parallel amino acid substitutions in amphipods differ in their properties from the substitutions that occurred in just one of the species. For this, we compared the distribution of Miyata distances between those substitutions that occurred in paths I and II (parallel substitutions) and those that only occurred in path I (non-parallel substitutions). The two distributions were indistinguishable (X-squared goodness of fit test, $X^2 = 2019.8$, df = 55, p-value < 2.2e-16; Figure S2).

**Discussion**

If the substitution rate is uniform between sites and invariable in time, the probability that a substitution occurs is independent of whether the same substitution has occurred at the same position in a related genome. However, multiple processes can make the substitution rates heterogenic between genomic sites, and such heterogeneity should inflate the observed rate of parallel substitutions at orthologous sites. Specifically, the rate of a parallel substitution, compared to a non-parallel one, can be elevated even in the absence of selection due to differences between genomic sites in point mutation rates (Hodkinson and Eyre-Walker 2011, Seplyarskiy et al. 2012). Besides, differences in strength of negative selection between genomic regions should result in an excess of substitutions at those regions were this selection is relaxed, and if this difference in selection pressures is conserved between species, this should lead to an excess of cases whereby the same (beneficial, neutral or even deleterious) substitution has occurred in parallel, compared to the expectation formulated without regard to selection heterogeneity.
Our approach of comparing the rates of nonsynonymous and synonymous parallel substitutions, while accounting for the mutation type, aims to single out the effect of parallel adaptation, mostly controlling for these confounding effects. Indeed, mutational heterogeneity should equally affect synonymous and nonsynonymous sites, which is the logic underlying the conventional dN/dS test as well as its derivatives such as the McDonald-Kreitman test (MacDonald and Kreitman 1991, Smith et al. 2002). Both in the conventional dN/dS test and in the \( P \) test, a deficit of nonsynonymous substitutions compared to the synonymous control implies negative selection preventing their fixation. By contrast, in the conventional dN/dS test, an excess of nonsynonymous substitutions implies positive selection in their favor; similarly, in the \( P \) test, \( P>1 \) suggests selection in favor of the substitutions parallel to those that also occurred at another lineage (Bazykin et al. 2007).

Consistent with previous results using few species, we find \( P\sim0.8 \) in vertebrates. By contrast, and strikingly, we observe a genome-wide \( P>1 \) in amphipods, over all substitution types. The cichlids demonstrate an intermediate pattern (\( P\sim1 \)).

What is the cause of high nonsynonymous parallelism in amphipods? While high parallelism could potentially arise artifactually, most sources of artifacts would equally affect synonymous and nonsynonymous sites, and it is difficult to think of a mechanism that could lead to \( P>1 \). Conceivably, some of the observed parallel differences between species could arise from artifactual misassembly of paralogous segments of DNA if different copies of the paralogs are included in analyses for different species. Since we use transcriptomics data, similar problems can also result from alternative splicing involving the segment covering the parallel site. Such artifacts, however, should affect all categories of sites equally, and are not expected to lead to the observed excess of nonsynonymous parallel substitutions. The presence of just a single variant in
the species DNA, coinciding with the variant determined by transcriptomics sequencing, was also confirmed by Sanger resequencing. Additionally, cross-sample contamination is made unlikely by the fact that different individuals were used for DNA and transcriptome sequencing. While some of the observed parallelisms could still be caused by artifactual assembly, it is not clear how it could lead to \(P>1\) observed in our data.

Phylogenetic patterns consistent with parallel evolution can arise artifactualy due to survival of ancestral polymorphism throughout the time between subsequent divergence events (incomplete lineage sorting), hybridisation or errors in phylogenetic reconstruction. This leads to hemiplasy, i.e., the situation when two alleles that seem to have originated in parallel are in fact identical by descent (Hahn and Nakhleh 2016). Yet these patterns are generally not expected to lead to \(P>1\), unless they are accompanied by selection in favor of new variants. The only conceivable scenario involving a hemiplasy that could result in \(P>1\) is adaptive hybridization, i.e., hybridization followed by preferential fixation of introgressed loci. Adaptive hybridization has been proposed as a mechanism for species flock emergence, as it increases within-species diversity facilitating further adaptation (Seehausen 2004); and reticular speciation has been observed in other crustaceans (daphnia, Giessler et al. 1999). However, the maximum distance between our amphipod species exceeds 10%, and we observe \(P>1\) even for the species that are that remote, while hybridisation is very unlikely between species at these distances (Jančúchová-Lásková et al 2015, Fitzpatrick 2004).

By exclusion, our observation of \(P>1\) requires parallel selection favoring the novel variant. There are two potential mechanisms for it: long-term balancing selection and recurrent episodes of positive selection. Balancing selection may increase the lifetime of two alleles cosegregating at a site, and their allele frequencies, over that at neutral (synonymous) sites; at a
fraction of sites, the alleles observed in the reference genomes of the four compared species will differ between species according to the pattern in Figure 2A, and this will increase $P$ compared to that at synonymous sites. Alternatively, $P>1$ can arise from positive selection favoring the novel adaptive variant at least at two of the four compared species.

There are aspects of data that argue against both these options. On the one hand, while many of the parallel nonsynonymous sites that we observed are polymorphic within a species, the level of polymorphism is lower than that at parallel synonymous sites or non-parallel nonsynonymous sites, while balancing selection is expected to increase the within-site diversity, compared to neutral. The deficit of polymorphism suggests ongoing negative selection in one species pair in favor of the variant also acquired in the other; together with the $P>1$, this implies that selection that had favored this acquisition was positive. On the other hand, the presence of some polymorphism is inconsistent with the action of positive selection, whereby positive selection should rid sites of polymorphism. The available data is not sufficient to distinguish between these alternatives. Moreover, we see no systematic directionality in the amino acid substitutions (Figure S2) or preference for particular gene categories, arguing against a genome-wide trend in adaptation, e.g. towards a change in protein thermostability; although this doesn’t exclude the possibility that the parallel adaptation in this system is promiscuous with respect to the substitutions it favors or genes that it affects.

Detection of adaptation from molecular data is complicated by the fact that it is hard to distinguish from the background of neutral and deleterious mutations. Analyzing substitutions between closely related species alleviates this problem, because at low phylogenetic distances, neutral (or slightly deleterious) substitutions have not yet had sufficient time to accumulate (Wolf et al. 2009, Stolyarova et al. 2019). The observation of $P>1$ in the flock of closely related
amphipod species, but not in more distantly related vertebrates, as well as the decrease in $P$ with phylogenetic distance in vertebrates (Figure 2B), is consistent with this explanation. Still, in a group of even more closely related cichlid fish, $P$ is lower than that in amphipods; and we see no dependence of $P$ on phylogenetic distance in amphipods or cichlids (Figure 2B). Moreover, vertebrate and amphipod species at close phylogenetic distances demonstrate very different values of $P$ (which are much higher in amphipods). In total, while dependent on phylogenetic distance in one of the groups (vertebrates), $P$ is more strongly determined by the identity of the analyzed group (Figure 2B).

Therefore, the enormous observed amount of nonsynonymous parallel evolution appears to be a specific feature of the baikalian amphipod species flock. More extensive, preferably whole-genome, data on within- and between-species variation is needed to clarify its cause. Such data, for example, could provide evidence of selective sweeps from reduced polymorphism in the vicinity of parallel sites, which would constitute an independent signature of positive selection. More generally, parallel adaptation may be not as rare as it seems, and worth a systematic survey in many groups, particularly those of closely related species.

**Literature**

Abascal F, Zardoya R, and Telford MJ. 2010. TranslatorX: Multiple Alignment of Nucleotide Sequences Guided by Amino Acid Translations. Nucleic Acids Res 38 (Web Server issue): W7–13.

Bailey SF., Rodrigue N, and Kassen R. 2015. The Effect of Selection Environment on the Probability of Parallel Evolution. Mol Biol Evol. 32(6): 1436-1448.

Baldwin MW et al. 2014. Evolution of Sweet Taste Perception in Hummingbirds by
Transformation of the Ancestral Umami Receptor. Science. 345(6199): 929-33.

Baym M et al. 2016. Spatiotemporal Microbial Evolution on Antibiotic Landscapes. Science 353 (6304): 1147–51.

Bazikalova AY et al. 1945. Amphipods of Lake Baikal. Trudy Baikals-koj Limnologicheskoj Stantsii, 11, 1–439. (In Russian.)

Bazykin GA et al. 2007. Extensive Parallelism in Protein Evolution. Biol Direct 2:20.

Bazykin GA. 2015. Changing preferences: deformation of single position amino acid fitness landscapes and evolution of proteins. Biol Lett. 11(10): 20150315.

Capella-Gutierrez S, Silla-Martinez JM, and Gabaldon T. 2009. TrimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. Bioinformatics 25(15):1972-3.

Castoe TA et al. 2009. Evidence for an ancient adaptive episode of convergent molecular evolution. Proc Natl Acad Sci USA. 106(22): 8986-91

Chikina M, Robinson JD and Clark NL. 2016. Hundreds of Genes Experienced Convergent Shifts in Selective Pressure in Marine Mammals. Mol Biol Evol 33 (9): 2182–92.

Coll F et al. 2018. Author Correction: Genome-Wide Analysis of Multi- and Extensively Drug-Resistant Mycobacterium Tuberculosis. Nature Genetics 50 (5): 764.

Collins C, Didelot X. 2018. A Phylogenetic Method to Perform Genome-Wide Association Studies in Microbes That Accounts for Population Structure and Recombination. PLOS Comput Biol 14(2): e1005958.

Conte GL, Arnegard ME, Peichel CL, Schluter D. 2012. The Probability of Genetic Parallelism and Convergence in Natural Populations. Proc Biol Sci. 279 (1749): 5039–47.

Drozdova P et al. 2019. Comparison between transcriptomic responses to short-term
stress exposures of a common Holarctic and endemic Lake Baikal amphipods. BMC Genomics, 20(1): 712.

Edgar RC. 2004. MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput. Nucleic Acids Res. 32(5): 1792-7.

Fitzpatrick BM. 2004. Rates of evolution of hybrid inviability in birds and mammals. Evolution, 58(8): 1865–1870.

Foote AD et al. 2015. Convergent Evolution of the Genomes of Marine Mammals. Nature Genetics. 47: 272–275.

Giessler S, Mader E, Schwenk K. 1999. Morphological evolution and genetic differentiation in Daphnia species complexes. J. Evol. Biol. 12(4), 710–723.

Hahn MW, and Nakhleh L. 2016. Irrational Exuberance for Resolved Species Trees. Evolution 70 (1): 7–17.

Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. Nat. Rev. Genet. 12(11): 756-66.

Jančúchová-Lásková J, Landová E, Frynta D. 2015. Are genetically distinct lizard species able to hybridize? A review. Current Zoology 61(1): 155-180.

Karasov T, Messer PW, Petrov DA. 2010. Evidence that adaptation in Drosophila is not limited by mutation at single sites. PLoS Genet, 6(6): e1000924

Karolchik DR et al. 2007. The UCSC Genome Browser Database: 2008 Update. Nucleic Acids Res. 36(Database issue): D773-9.

Klink GV, Bazykin GA. 2017. Parallel Evolution of Metazoan Mitochondrial Proteins. Genome Biol Evol 9(5): 1341-50.

Klink GV, Golovin AV, Bazykin GA. 2017. Substitutions into amino acids that are
pathogenic in human mitochondrial proteins are more frequent in lineages closely related to human than in distant lineages. PeerJ 5:e4143.

Kreiner JM et al. 2019. Multiple modes of convergent adaptation in the spread of glyphosate-resistant Amaranthus tuberculatus. Proc Natl Acad Sci USA

Kryazhimskiy S, Rice DP, Jerison ER, Desai MM. 2014. Microbial Evolution. Global Epistasis Makes Adaptation Predictable despite Sequence-Level Stochasticity. Science 344 (6191): 1519–22.

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21): 2987-93.

Li L, Stoeckert CJ Jr, Roos DS. 2003. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. Genome Res. 13: 2178-2189.

Lim MCW, Witt CC, Graham CH, Dávalos LM. 2019. Parallel Molecular Evolution in Pathways, Genes, and Sites in High-Elevation Hummingbirds Revealed by Comparative Transcriptomics. Genome Biol Evol 11(6): 1573–1585.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351(6328): 652-4.

Malinsky M et al. 2018. Whole Genome Sequences of Malawi Cichlids Reveal Multiple Radiations Interconnected by Gene Flow. Nat Ecol Evol 2(12): 1940-1950.

Marcovitz A et al. 2019. A functional enrichment test for molecular convergent evolution finds a clear protein-coding signal in echolocating bats and whales. Proc Natl Acad Sci U S A. 116(42): 21094-21103.

Chandrasekhar N et al. 2016. Predictable Convergence in Hemoglobin Function Has
Unpredictable Molecular Underpinnings. Science 354 (6310): 336–39.

Naumenko S et al 2017. Transcriptome-Based Phylogeny of Endemic Lake Baikal Amphipod Species Flock: Fast Speciation Accompanied by Frequent Episodes of Positive Selection. Mol Ecol. 26(2): 536-553.

Nikolaev SI, Montoya-Burgos JI, Popadin K, Parand L, Margulies EH. 2007. Life-history traits drive the evolutionary rates of mammalian coding and noncoding genomic elements. Proc Natl Acad Sci U S A. 104(51): 20443-8.

Povolotskaya IS, Kondrashov FA. 2010. Sequence space and the ongoing expansion of the protein universe. Nature. 465: 922–926.

Rokas A, Carroll SB. 2008. Frequent and Widespread Parallel Evolution of Protein Sequences. Mol Biol Evol. 25(9): 1943–1953.

Seehausen O. 2004. Hybridization and adaptive radiation. Trends Ecol Evol 19(4): 198-207.

Seplyarskiy VB, Kharchenko P, Kondrashov AS, Bazykin GA. 2012. Heterogeneity of the Transition/Transversion Ratio in Drosophila and Hominidae Genomes. Mol Biol Evol. 29(8): 1943-55.

Smith NG, Eyre-Walker A. 2002. Adaptive protein evolution in Drosophila. Nature, 415(6875): 1022-4.

Soria-Carrasco V et al. 2014. Stick Insect Genomes Reveal Natural Selection’s Role in Parallel Speciation. Science. 344: 738–742.

Stamatakis A. 2014. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. Bioinformatics 30(9):1312-3.

Stolyarova AV et al. 2019. Senescence and entrenchment in evolution of amino acid
Storz JF. 2016. Causes of Molecular Convergence and Parallelism in Protein Evolution. Nat Rev Genet. 17(4): 239–50.

Terekhanova NV et al. 2014. Fast Evolution from Precast Bricks: Genomics of Young Freshwater Populations of Threespine Stickleback Gasterosteus Aculeatus. PLoS Genet. 10(10): e1004696.

Thomas GWC, Hahn MW. 2015. Determining the Null Model for Detecting Adaptive Convergence from Genomic Data: A Case Study Using Echolocating Mammals. Mol Biol Evol. 32(5): 1232-6.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering Splice Junctions with RNA-Seq. Bioinformatics. 25(9): 1105-1111.

Usmanova DR, Ferretti L, Povolotskaya IS, Vlasov PK, Kondrashov FA. 2015. A Model of Substitution Trajectories in Sequence Space and Long-Term Protein Evolution. Mol Biol Evol. 32(2): 542–54.

Woods R, Schneider D, Winkworth CL, Riley MA, and Lenski RE. 2006. Tests of Parallel Molecular Evolution in a Long-Term Experiment with Escherichia Coli. Proc Natl Acad Sci USA. 103(24): 9107-9112.

Zou Z, Zhang J. 2015a. No Genome-Wide Protein Sequence Convergence for Echolocation. Mol Biol Evol 32(5): 1237–41.

Zou Z, Zhang J. 2015b. Are Convergent and Parallel Amino Acid Substitutions in Protein Evolution More Prevalent Than Neutral Expectations? Mol Biol Evol. 32(8): 2085–96.