

Research



Cite this article: Stolyarova AV, Bazykin GA, Neretina TV, Kondrashov AS. 2019 Bursts of amino acid replacements in protein evolution. *R. Soc. open sci.* **6**: 181095. <http://dx.doi.org/10.1098/rsos.181095>

Received: 6 August 2018
Accepted: 25 February 2019

Subject Category:

Genetics and genomics

Subject Areas:

evolution/genomics

Keywords:

adaptive walks, punctuated equilibrium, rate of evolution, positive selection

Author for correspondence:

Anastasia V. Stolyarova
e-mail: anastasia.v.stolyarova@gmail.com

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4428893>.

Bursts of amino acid replacements in protein evolution

Anastasia V. Stolyarova¹, Georgii A. Bazykin^{1,2},
Tatyana V. Neretina^{3,4} and Alexey S. Kondrashov^{4,5}

¹Skolkovo Institute of Science and Technology, Skolkovo 143026, Russia

²Institute for Information Transmission Problems (Kharkevich Institute) of the Russian Academy of Sciences, Moscow 127994, Russia

³White Sea Biological Station, Biological Faculty, M. V. Lomonosov Moscow State University, Leninskie Gory, Moscow 119991, Russia

⁴Department of Bioengineering and Bioinformatics, M. V. Lomonosov Moscow State University, Moscow 119234, Russia

⁵Department of Ecology and Evolutionary Biology, University of Michigan, 830 North University, Ann Arbor, MI 48109-1048, USA

AVS, 0000-0002-1418-6673

Evolution can occur both gradually and through alternating episodes of stasis and rapid changes. However, the prevalence and magnitude of fluctuations of the rate of evolution remain obscure. Detecting a rapid burst of changes requires a detailed record of past evolution, so that events that occurred within a short time interval can be identified. Here, we use the phylogenies of the Baikal Lake amphipods and of Catarrhini, which contain very short internal edges which make this task feasible. We detect six salient bursts of evolution of individual proteins during such short time periods, each involving between six and 38 amino acid substitutions. These bursts were extremely unlikely to have occurred neutrally, and were apparently caused by positive selection. On average, in the course of a time interval required for one synonymous substitution per site, a protein undergoes a strong burst of rapid evolution with probability at least approximately 0.01.

1. Introduction

(Non)uniformity of the rate of evolution is one of the oldest and most contentious issues in evolutionary biology. On the one hand, at both molecular and morphological levels, evolution often occurs gradually, so that its rate is approximately constant [1–5]. In particular, this is the case for selectively neutral segments of genomes, which evolve at rates equal to the

corresponding mutation rates ('molecular clock' [6,7]). Instances of gradual adaptive evolution are also known [8,9].

On the other hand, evolution may also occur mostly through short bursts of changes alternating with long periods of stasis ('punctuated gradualism' or 'punctuated equilibrium' [10,11]). Examples of punctuated equilibrium are provided by the evolution of mammalian body weight [12], hominoid body size [13], several morphological traits of rockfish [14], intersexual signalling of cranes [15] and many other data [16–22]. Clearly, both gradual and burst-like evolution does happen, but their relative importance remains controversial [23–26]. Of course, there can be many causes for alternating episodes of stasis and evolution and of punctuation in molecular and morphological evolution.

In order to detect short bursts of changes, one needs to be able to identify evolutionary events that occurred during a short interval of time. This is easy to accomplish if a very detailed palaeontological record is available, such as those that exist for some marine invertebrates, e.g. Foraminifera [27]. However, such records are exceptions rather than the rule. Furthermore, palaeontological data usually shed light only on the morphology of organisms and, thus, cannot reveal bursts of changes at the level of genomes. Fortunately, it may be possible to identify such bursts indirectly, through comparison of genomes of extant species, as long as their phylogenetic tree contains very short internal edges. Unfortunately, despite an avalanche of genomic data, the vast majority of the currently available phylogenetic trees do not satisfy this requirement. Still, we took advantage of two phylogenetic trees that contain such edges, those of the Lake Baikal amphipods [28] and of Catarrhini [29], UCSC 100 vertebrates multiple alignment, and investigated short bursts in the evolution of their proteins.

2. Material and methods

2.1. Dataset

We used two datasets. First, we considered transcriptomes-based clusters of orthologous genes (COGs) with exactly one orthologue represented in each species for five clades of the Lake Baikal amphipods (gammarids) (64 species and 3399 COGs in total), and the phylogeny based on them. The size of a clade varies from 6 to 24 species (electronic supplementary material, figure S1A). Each clade consists of closely related species, which makes it possible to reconstruct ancestral states with high confidence. Thus, all amino acid substitutions that constitute a burst can be reliably identified. The search for bursts was performed for each clade separately.

Second, we considered a multiple alignment of protein-coding genes from 11 primates species obtained from the 100 vertebrates genomes alignment of the UCSC Genome Browser together with the corresponding reconstructed phylogenetic tree (electronic supplementary material, figure S1B). In total, there were 17 755 alignments of protein-coding genes of primates containing columns without gaps.

Only internal edges, i.e. segments of the phylogenetic tree ancestral to more than one species, were used in our analysis. For both datasets, we only considered internal edges of length less than 0.005 dS units. We used *codeml* program of the PAML package [30] to reconstruct substitution histories of sequences and to estimate dN/dS values. Only gapless alignment columns were considered.

Presumptive functions of amphipod genes were inferred from blast2GO predictions [28] and from the genome annotation of a related species *Hyaletella azteca* [31]; functions of primate genes were inferred the human genome annotation (hg38).

2.2. Analysis

We assumed the neutral null model ($dN = dS$), which makes it possible to detect only the longest and the fastest bursts (see Discussion). We related dN of the gene on a particular edge of the phylogenetic tree to the length of this edge. Length edge was measured in the units of dS on the basis of all the available genes. We used this approach, instead of considering the dS value of only the gene that underwent a burst, because it is impossible to estimate the dS for an individual gene on a short edge with precision. For example, the expected number of synonymous substitutions in a gene encoding a 200 amino acid long protein on the edge of length 0.005 dS is only approximately 1. For a particular edge, the *p*-value associated with a non-synonymous burst was calculated for each gene as the probability of observing this many or more non-synonymous substitutions in a Poisson distribution with the

parameter equal to the edge length (dS) (or, for bursts spanning multiple edges in a row, the sum of their lengths) multiplied by the number of non-synonymous sites in the gene (estimated according to [32]). The obtained *p*-values were adjusted for multiple comparisons using the Benjamini–Hochberg method.

As a control, we also searched for statistically significant bursts of synonymous substitutions on the same set of short edges using an analogous approach.

2.3. Filtering

The primary set of putative bursts with Benjamini–Hochberg adjusted *p*-values < 0.05 was filtered as follows.

First, alignments containing more than 50% of columns with gaps were excluded. Second, to ensure the precise phylogenetic positioning of all substitutions that constituted a burst, we excluded sites with *codeml* posterior probabilities for the reconstructed ancestral variant less than 0.8 (which will be the case for discordant genes, e.g. caused by incomplete lineage sorting) and recalculated the statistics for this gene.

Third, to safeguard against contribution of anciently divergent paralogues or pseudogenes rather than orthologues to our findings, we applied additional filtering. We required that the dS value that characterized the edge of the putative burst obtained using the considered gene was lower than the dS value for this edge obtained using all genes (adjusted *p*-value > 0.001). Genes with substitutions in multiple repeated domains were excluded. Next, we required the absence of evidence for paralogues or duplications of the considered gene as follows. For each gene, we determined the pre-burst sequence as the reconstructed sequence of the phylogenetic node immediately ancestral to the burst-carrying edge(s), and the post-burst sequence as the reconstructed sequence of the phylogenetic node immediately descendant to it. For gammarids, we mapped raw transcriptomic reads of the considered gene from all species onto the pre-burst and post-burst sequences. If any reads from any of the species descendant to the edge of the provisional burst supported the pre-burst variant, or if any reads from any of the species not descendant to the edge of the provisional burst supported the post-burst variant, this gene was discarded. For primates, we aligned the pre-burst and post-burst sequences onto the assembled genomes of this gene from all species, and proceeded analogously. For primates, we also required that the genomic position of each burst-carrying gene was conserved in all primate genomes.

Finally, the burst-containing alignments that survived these filters were curated manually for any evidence for alignment errors, low complexity and unexpected patterns in substitutions. If most substitutions constituting a burst fell into regions of poor alignment or were located in the very beginning or the very end of the gene, the corresponding putative burst was discarded.

3. Results

We searched for bursts of amino acid substitutions (bursts) within internal edges of phylogenetic trees that are shorter than 0.005 dS. Suitable edges are present in five clades of the phylogenetic tree of gammarids from Lake Baikal [28]: *Eulimnogammarus* and related genera (18 edges), *Pallasea* and related genera (10), *Hyalleopsis* (3), Acanthogammaridae s. str. (7) and Micruropidae (4); as well as within the Catarrhini clade (3 edges) of the tree of vertebrates (electronic supplementary material, figure S1) [29]. A burst consists of several amino acid substitutions which occurred in a protein within such an edge or, perhaps, within several successive edges of combined length below 0.005 dS.

In gammarids, five statistically significant bursts occurred in five proteins within two clades. Three of them occurred over the time period corresponding to a short individual edge of the phylogeny, while the remaining two spanned two very short adjacent edges. In Catarrhini, there was one significant burst (table 1). Each burst consisted of between 6 and 38 amino acid substitutions, or between 6 and 39 non-synonymous substitutions (as some amino acid sites underwent multiple non-synonymous substitutions), scattered throughout the protein (electronic supplementary material, figure S2). All edges that harbour bursts have 100% bootstrap support. Unfortunately, there is no accepted dating for amphipods diversification, so we cannot estimate the time of the corresponding bursts with precision; all that we know is that they have occurred after Baikal has originated in the Miocene, i.e. less than 30 Ma. Genes that harboured bursts are enriched in proteins located in mitochondria: they constitute three of the five such genes, although only 14% of the initial set of COGs are annotated as

Table 1. Bursts of evolution in proteins of the Baikal gammarids and Catarrhini.

clade	edge number	edge length (dS)	gene name (predicted)	description of the protein	overall dN/dS for the gene (excluding the edge with the burst)	substitutions during burst		adjusted p -value
						non-syn.	syn.	
<i>Pallasea</i> and related genera	1	0.0040	<i>DNAJC11</i>	DnaJ-like protein subfamily c member 11	0.41	39	1	1.01×10^{-25}
	1	0.0040	<i>MRPL22</i>	mitochondrial ribosomal protein L22	0.55	10	1	0.012
	2 + 3	0.0007 + 0.0008	<i>NOP16</i>	nucleolar protein 16-like	0.35	6	2	0.046
<i>Eulimnogammarus</i> and related genera	4	0.0011	<i>MRPS25</i>	mitochondrial ribosomal protein S25	0.31	6.5	0.5	0.0011
	5 + 6	0.0024 + 0.0010	<i>AKR1</i>	aldo-keto reductase	0.57	10	1	0.032
primates	7	0.0043	<i>PKR</i>	interferon-induced protein kinase R	1.27	18	2	0.0010

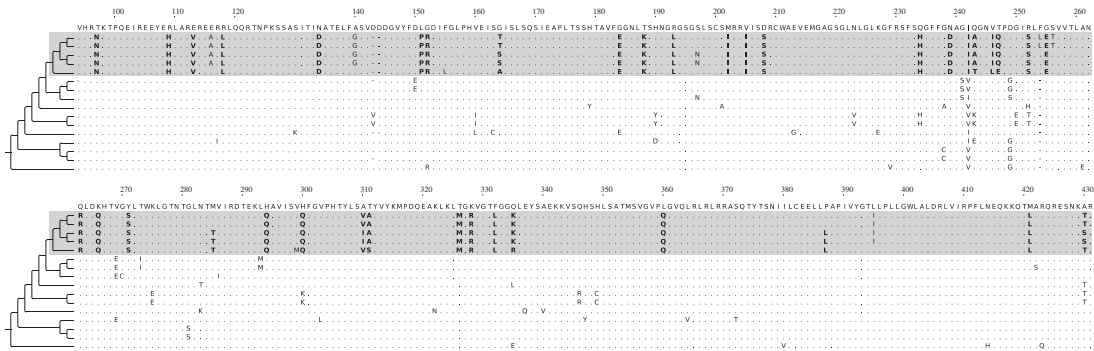


Figure 1. A fragment of the alignment of orthologous *DNAJC11* genes of *Pallasea* gammarids, which contains 39 non-synonymous substitutions on the internal edge 0.004 dS in length. Alleles derived in the adaptive burst are shown in bold. Total length of the alignment is 1677 nucleotides, or 1515 nucleotides without gaps.

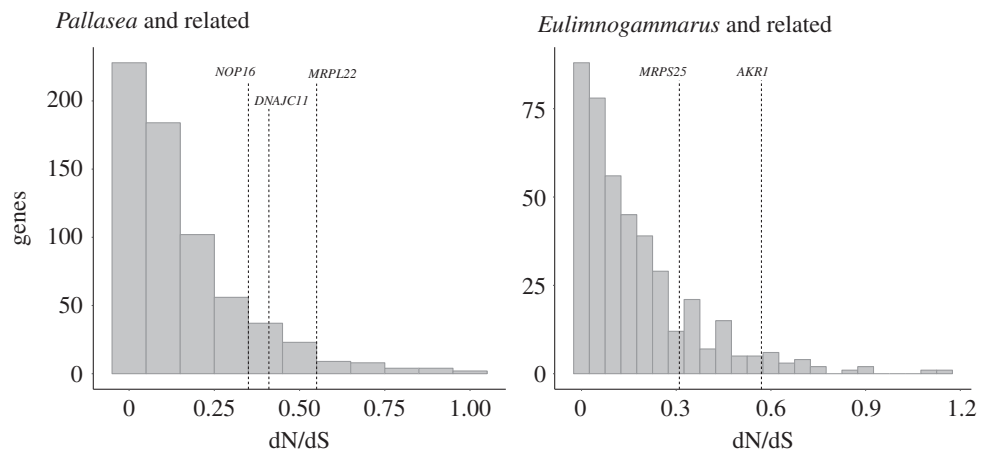


Figure 2. Distribution of dN/dS in genes of Baikla Amphipods in the clades carrying bursts. Genes with confirmed bursts are shown with dashed lines.

components of mitochondria (binomial test, p -value = 0.02; electronic supplementary material, figure S3) [28]. No significant synonymous bursts were observed.

The most remarkable burst involving 39 non-synonymous substitutions occurred in the mitochondrial chaperone gene (*DNAJC11*) on an edge of length 0.004 dS in the *Pallasea* clade (figure 1). This edge also harboured another burst in a protein located in mitochondria (L22 ribosome protein) (table 1). Bursts that are confined to one edge of the phylogenetic tree do not extend to preceding and/or successive edges (p -values for such edges greater than 0.27) (electronic supplementary material, figure S4). Hence, the characteristic duration of a burst is short, approximately 10^{-3} dS. The overall rate of evolution of some burst-carrying genes was somewhat higher than the average (table 1 and figure 2). Still, after multiple testing correction, there remains no genes with more than one statistically significant burst (p -values > 0.018 , adjusted p -values = 1).

Phylogenetic tree of 11 species of Catarrhini has only three internal edges shorter than 0.005 dS (electronic supplementary material, figure S1B), and only one statistically significant burst has been detected. This small number may be due to several reasons: Catarrhini species are more distant from each other, which results in longer phylogenetic edges and less confident ancestral state reconstruction; moreover, a larger initial dataset leads to a more substantial multiple testing correction.

The detected burst occurred on the internal edge ancestral to two macaques species. Based on the divergence time estimates, we assume that the burst occurred at approximately 8–3 Ma [33]. The gene with the burst encodes the PKR protein (also known as EIF2AK), which is the eukaryotic translation initiation factor 2 kinase activated during viral infection. As in gammarids, the substitutions are scattered along the sequence (figure 3). Primate PKR contains two dsRNA binding motifs (DRBMs 1 and 2) and C-terminal catalytic kinase domain. The kinase domain carries 14 amino acid substitutions on the selected edge, and DRBM1 the remaining 4. Most substitutions lie in α D, α G and α H helices

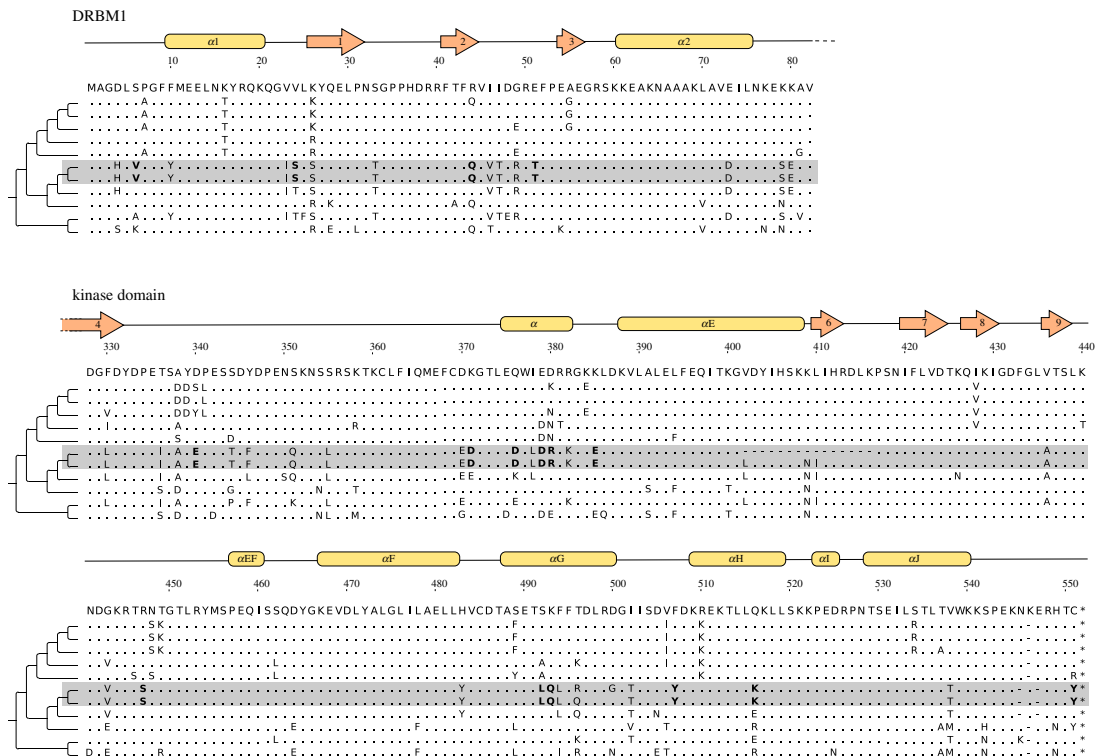


Figure 3. Alignment of *PKR* genes of Catarrhini (fragments) containing 18 non-synonymous substitutions on the internal edge ancestral to *Macaca mulatta* and *Macaca fascicularis*. The majority of substitutions (14) occurred in the kinase domain of the protein, the others fall in the dSRNA binding motif (DRBM1). Alleles derived in the adaptive burst are shown in bold.

or nearby, which have been shown to be enriched in positively selected sites [34]. αG helix and specifically positions with amino acid substitutions on the selected edge are involved in PKR interaction with *eIF2 α* [35]; however, there is no evidence of bursts of evolution in *eIF2 α* gene on the same edge.

4. Discussion

Allele replacements driven by positive selection are the fundamental genetic mechanism of adaptive evolution. These replacements can occur independently of each other or be correlated [36]. *A priori*, there is a continuum of possibilities, from fully independent individual substitutions to bursts of adaptive evolution, each consisting of multiple substitutions that occurred over a short period of time. We searched for such bursts within individual proteins, taking advantage of two phylogenetic trees, of the Lake Baikal gammarids [28] and of Catarrhini [29].

Using only internal edges is essential because in this case the derived sequence is observed in more than one species, so that rare sequencing and alignment errors would not lead to false discovery of bursts. Our criteria for detection of bursts were rather stringent: conservative filtering of alignments, a neutral null model ($dN = dS$) and multiple testing correction. Unfortunately, it is hard to define the correct null model which takes into account all possible features of protein evolution, for example, those resulting from non-uniformity of the mutation rate along the genomes. Thus, our *p*-values should be viewed with caution. Even neutrally evolving sequences can carry increased number of substitutions because of variation in the evolution rate over large time scales (overdispersed molecular clock [37–39]). However, the aim of our work was not to identify consistent deviations from the Poisson expectation, but to find the most radical outliers. The six bursts that we have found are likely to be ‘real’, in the sense of being caused by simultaneous or near-simultaneous action of positive selection at multiple sites within a protein.

Berglund *et al.* [40] and Galtier *et al.* [41] have used methods similar to ours to find episodes of accelerated evolution which could have non-adaptive explanations, such as gene conversion. By contrast, we used stringent criteria for detection of bursts, and those that we found are likely to occur

due to positive selection. In particular, the most radical bursts in Galtier *et al.* involve both synonymous and non-synonymous substitutions; while the bursts described in our paper, in particular in the novel amphipod dataset, are limited to non-synonymous sites. Unlike Galtier and Berglund, we see no GC bias in the bursts-composing substitutions: e.g. the number of AT \rightarrow GC and GC \rightarrow AT substitutions in the strongest burst are 14 and 21, correspondingly. This is inconsistent with gene conversion, and instead further supports the adaptive explanation.

Mutations that initiated amino acid substitutions that together constitute a burst are extremely unlikely to appear simultaneously as parts of one complex mutational event. Thus, a burst is likely to involve substitutions that were not precisely synchronous. In other words, a burst lasts longer than a substitution. Still, the bursts that we detected are quite short, at the evolutionary time scale. Indeed, four bursts were confined to just one internal edge shorter than 0.005 dS. The remaining two bursts, involving six and nine amino acid substitutions, each occurred on two successive internal edges, of lengths 0.0007 and 0.0008, and 0.0024 and 0.0010 dS, suggesting that approximately three to four non-synonymous substitutions occurred per 0.001 dS of evolutionary time. Among the substitutions involved in such composite bursts, neither occurred on both edges after a cladogenesis (multiple non-synonymous substitutions did not occur on edges leading to *Eulimnogammarus testaceus* or *Pallasea cancellus*).

Of course, the two phylogenetic trees which we studied almost certainly contained other bursts of positive selection-driven amino acid substitutions which we could not detect with certainty. This would be the case for any burst that occurred within an external edge of a tree, or within an internal edge that is not short enough, or even within a short internal edge as long as the burst itself involved only a small number of substitutions. Unfortunately, we cannot estimate the number of such real but not confidently detectable bursts.

Obviously, our ability to detect a burst depends on the length of the internal edge. Roughly speaking, all bursts that involve at least approximately eight amino acid substitutions within a protein of less than 300 amino acids can be detected within internal edges of length below 0.005 dS. Because we investigated 3411 proteins and the total length of all such edges in the gammarid tree was 0.15 dS, five bursts that we found in them imply that during an interval of time required for one synonymous substitution to occur per site, a protein undergoes a strong burst of adaptive evolution with probability approximately 0.01 (the estimate derived from primates is similar). If so, such bursts are not uncommon.

What can we say about the genes that underwent bursts? Not much: they evolve faster than an average gene, but only marginally so. Unexpectedly, three out of five genes encode proteins that are located in mitochondria: a mitochondrial chaperone and two mitochondrial ribosome proteins. This observation is hard to explain. Mitochondrial genomes of Baikal Lake gammarids have been shown to undergo intensive rearrangement, which in combination with mito-nuclear discordance and epistatic interactions between mitochondrial proteins coded in nuclear and mitochondrial genomes might lead to this phenomenon [42].

Because multiple nucleotide substitutions that constitute a burst occur very close to each other, making recombination between them negligible, Hill–Robertson interference [43] can be expected to impede their fixations. Let us consider the most extreme burst comprising 39 non-synonymous substitutions on the internal edge of length 0.004 dS. Assuming the per nucleotide per generation mutation rate is approximately 10^{-8} , as in a number of animals [44], this edge corresponds to approximately 400 000 generations, leaving approximately 10 000 generations per each substitution, if they occurred without overlaps. Is this feasible? Every generation, $2N\mu$ mutations occur at a site, where N is the census population size and μ is the mutation rate. An advantageous mutation will eventually reach fixation with probability $2sN_e/N$, where N_e is the effective population size (see [45, p. 158]). Thus, the per generation probability of fixation of a particular advantageous mutation is $4N_e\mu s$. Under assumptions of $N_e = 10^5$ (limited data indicate nucleotide diversity approx. 0.01 in several Lake Baikal amphipods), $\mu = 10^{-8}$ and $s = 10^{-2}$, this probability becomes 4×10^{-5} , which is not very different from 10^{-4} . Thus, successive accumulation of substitutions that constitute a burst, which makes them immune to the Hill–Robertson interference, cannot be ruled out. This would be especially the case if during the whole course of a burst selection favours mutations at all the 39 sites that constitute it, or, in other words, the order in which the substitutions occur is not prescribed [46]. If so, the target for advantageous mutation at a particular moment of time consists of all sites where substitutions did not yet occur, and their order depends on the order in which mutations appear.

Correlated positive selection at multiple sites that leads to a burst may emerge due to a variety of mechanisms. One possibility, of course, is a sudden, drastic change of the adaptive landscape of a protein. However, a burst can also occur as a result of only a small change of the landscape, if it caused a fold bifurcation which eliminated a fitness peak that was occupied by a protein [47, fig. 4].

This mechanism is compatible with the fact that all genes with bursts in gammarids show a low overall dN/dS ratio on the entire phylogenetic tree (less than 0.57), implying that these bursts of evolution affected genes that usually evolve slowly. By contrast, the *PKR* gene of primates possessed a high dN/dS ratio (greater than 1), implying that the burst in this gene involved an episode of additional acceleration of evolution which was generally fast. Hopefully, the number of available dense phylogenetic trees will soon become much larger, which will make it possible to study bursts of rapid evolution in more detail.

Data accessibility. Data available from the Dryad Digital Repository: <https://doi.org/10.5061/dryad.40vp30d> [48].

Authors' contributions. A.V.S. carried out the data analyses, participated in the design of the study and drafted the manuscript; G.A.B. participated in the design of the study and helped draft the manuscript; T.V.N. provided experimental validation of the results; A.S.K. conceived of the study and helped draft the manuscript. All authors gave final approval for publication. Competing interests. The authors declare no competing interests.

Funding. This study was supported by Russian Science Foundation (grant no. 16-14-10173) to A.S.K.

Acknowledgements. We gratefully acknowledge Dr E. Braun and the anonymous reviewer whose comments helped to improve and to clarify the manuscript.

References

- Martin AP, Palumbi SR. 1993 Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl Acad. Sci. USA* **90**, 4087–4091. (doi:10.1073/pnas.90.9.4087)
- Kumar S, Blair Hedges S. 1998 A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920. (doi:10.1038/31927)
- dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z. 2012 Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc. R. Soc. B* **279**, 3491–3500. (doi:10.1098/rspb.2012.0683)
- Tamura K *et al.* 2012 Estimating divergence times in large molecular phylogenies. *Proc. Natl Acad. Sci. USA* **109**, 19 333–19 338. (doi:10.1073/pnas.1213199109)
- O'Meara BC, Ané C, Sanderson MJ, Wainwright PC. 2006 Testing for different rates of continuous trait evolution using likelihood. *Evolution Int. J. Org. Evolution* **60**, 922–933. (doi:10.1111/j.0014-3820.2006.tb01171.x)
- Zuckerandl E, Pauling L. 1965 Evolutionary divergence and convergence in proteins. In *Evolving genes and proteins* (eds V Bryson, HJ Vogel), pp. 97–166. New York, NY: Academic Press.
- Kimura M, Ohta T. 1974 Probability of gene fixation in an expanding finite population. *Proc. Natl Acad. Sci. USA* **71**, 3377–3379. (doi:10.1073/pnas.71.9.3377)
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, Kim JF. 2009 Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* **461**, 1243–1247. (doi:10.1038/nature08480)
- Mahler DL, Revell LJ, Glor RE, Losos JB. 2010 Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean anoles. *Evolution Int. J. Org. Evolution* **64**, 2731–2745. (doi:10.1111/j.1558-5646.2010.01026.x)
- Gould SJ, Eldredge N. 1993 Punctuated equilibrium comes of age. *Nature* **366**, 223–227. (doi:10.1038/366223a0)
- Stanley SM. 1998 *Macroevolution: pattern and process*. Baltimore, MD: Johns Hopkins University Press.
- Mattila TM, Bokma F. 2008 Extant mammal body masses suggest punctuated equilibrium. *Proc. R. Soc. B* **275**, 2195–2199. (doi:10.1098/rspb.2008.0354)
- Bokma F. 2002 Detection of punctuated equilibrium from molecular phylogenies. *J. Evol. Biol.* **15**, 1048–1056. (doi:10.1046/j.1420-9101.2002.00458.x)
- Ingram T. 2011 Speciation along a depth gradient in a marine adaptive radiation. *Proc. R. Soc. B* **278**, 613–618. (doi:10.1098/rspb.2010.1127)
- Mooers A, Vamossi SM, Schluter D. 1999 Using phylogenies to test macroevolutionary hypotheses of trait evolution in cranes (Gruinae). *Am. Nat.* **154**, 249. (doi:10.1086/303226)
- Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. 2006 Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biol. Direct* **1**, 34. (doi:10.1186/1745-6150-1-34)
- Hunt G. 2007 The relative importance of directional change, random walks, and stasis in the evolution of fossil lineages. *Proc. Natl Acad. Sci. USA* **104**, 18 404–18 408. (doi:10.1073/pnas.0704088104)
- Hunt G. 2008 Gradual or pulsed evolution: when should punctuational explanations be preferred? *Paleobiology* **34**, 360–377. (doi:10.1666/07073.1)
- Strotz LC, Allen AP. 2013 Assessing the role of cladogenesis in macroevolution by integrating fossil and molecular evidence. *Proc. Natl Acad. Sci. USA* **110**, 2904–2909. (doi:10.1073/pnas.1208302110)
- Bedford T *et al.* 2014 Integrating influenza antigenic dynamics with molecular evolution. *eLife* **3**, e01914. (doi:10.7554/eLife.01914)
- Hunt G, Hopkins MJ, Lidgard S. 2015 Simple versus complex models of trait evolution and stasis as a response to environmental change. *Proc. Natl Acad. Sci. USA* **112**, 4885–4890. (doi:10.1073/pnas.1403662111)
- Voje KL. 2016 Tempo does not correlate with mode in the fossil record. *Evolution Int. J. Org. Evolution* **70**, 2678–2689. (doi:10.1111/evo.13090)
- Gillespie JH. 1991 *The causes of molecular evolution*. Oxford, UK: Oxford University Press.
- Pagel M, Venditti C, Meade A. 2006 Large punctuational contribution of speciation to evolutionary divergence at the molecular level. *Science* **314**, 119–121. (doi:10.1126/science.1129647)
- Venditti C, Pagel M. 2008 Speciation and bursts of evolution. *Evolution: Educ. Outreach* **1**, 274–280. (doi:10.1007/s12052-008-0049-4)
- Pennell MW, Harmon LJ, Uyeda JC. 2014 Is there room for punctuated equilibrium in macroevolution? *Trends Ecol. Evol.* **29**, 23–32. (doi:10.1016/j.tree.2013.07.004)
- Malmgren BA, Berggren WA, Lohmann GP. 1983 Evidence for punctuated gradualism in the Late Neogene *Globorotalia tumida* lineage of planktonic foraminifera. *Paleobiology* **9**, 377–389. (doi:10.1017/S0094837300007843)
- Naumenko SA *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**, 536–553. (doi:10.1111/mec.13927)
- Rosenbloom KR *et al.* 2015 The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**(Database issue), D670–D681. (doi:10.1093/nar/gku1177)
- Yang Z. 2007 PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591. (doi:10.1093/molbev/msm088)
- Poynton HC *et al.* 2018 The toxicogenome of *Hyalella azteca*: a model for sediment ecotoxicology and evolutionary toxicology. *Environ. Sci. Technol.* **52**, 6009–6022. (doi:10.1021/acs.est.8b00837)
- Nei M, Gojobori T. 1986 Simple methods for estimating the numbers of synonymous and

- nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426.
33. Perelman P *et al.* 2011 A molecular phylogeny of living primates. *PLoS Genet.* **7**, e1001342. (doi:10.1371/journal.pgen.1001342)
 34. Rothenburg S, Seo EJ, Gibbs JS, Dever TE, Dittmar K. 2009 Rapid evolution of protein kinase PKR alters sensitivity to viral inhibitors. *Nat. Struct. Mol. Biol.* **16**, 63–70. (doi:10.1038/nsmb.1529)
 35. Krishna KH, Vadlamudi Y, Kumar MS. 2016 Viral evolved inhibition mechanism of the RNA dependent protein kinase PKR's kinase domain, a structural perspective. *PLoS ONE* **11**, e0153680. (doi:10.1371/journal.pone.0153680)
 36. Bazykin GA, Kondrashov FA, Ogurtsov AY, Sunyaev S, Kondrashov AS. 2004 Positive selection at sites of multiple amino acid replacements since rat-mouse divergence. *Nature* **429**, 558–562. (doi:10.1038/nature02601)
 37. Ohta T, Kimura M. 1971 On the constancy of the evolutionary rate of cistrons. *J. Mol. Evol.* **1**, 18–25. (doi:10.1007/BF01659391)
 38. Gillespie JH. 1984 The molecular clock may be an episodic clock. *Proc. Natl Acad. Sci. USA* **81**, 8009–8013. (doi:10.1073/pnas.81.24.8009)
 39. Cutler DJ. 2000 Understanding the overdispersed molecular clock. *Genetics* **154**, 1403–1417.
 40. Berglund J, Pollard KS, Webster MT. 2009 Hotspots of biased nucleotide substitutions in human genes. *PLoS Biol.* **7**, e1000026. (doi:10.1371/journal.pbio.1000026)
 41. Galtier N *et al.* 2009 GC-biased gene conversion promotes the fixation of deleterious amino acid changes in primates. *Trends Genet.* **25**, 1–5. (doi:10.1016/j.tig.2008.10.011)
 42. Romanova EV *et al.* 2016 Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genomics* **17**(Suppl. 14), 1016. (doi:10.1186/s12864-016-3357-z)
 43. Hill WG, Robertson A. 1966 The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269–294. (doi:10.1017/S0016672300010156)
 44. Lynch M. 2010 Evolution of the mutation rate. *Trends Genet.* **26**, 345–352. (doi:10.1016/j.tig.2010.05.003)
 45. Ferrière R, Dieckmann U, Couvet D. 2004 *Evolutionary conservation biology*. Cambridge, UK: Cambridge University Press.
 46. Kondrashov FA, Kondrashov AS. 2001 Multidimensional epistasis and the disadvantage of sex. *Proc. Natl Acad. Sci. USA* **98**, 12 089–12 092. (doi:10.1073/pnas.211214298)
 47. Dodson MM, Hallam A. 1977 Allopatric speciation and the fold catastrophe. *Am. Nat.* **111**, 415–433. (doi:10.1086/283176)
 48. Stolyarova AV, Bazykin GA, Neretina TV, Kondrashov AS. 2019 Data from: Bursts of amino acid replacements in protein evolution. Dryad Digital Repository. (doi:10.5061/dryad.40vp30d)