

Research



Cite this article: Sun S, Wu Y, Ge X, Jakovlić I, Zhu J, Mahboob S, Al-Ghanim KA, Al-Misned F, Fu H. 2020 Disentangling the interplay of positive and negative selection forces that shaped mitochondrial genomes of *Gammarus pisinnus* and *Gammarus lacustris*. *R. Soc. open sci.* **7**: 190669.
<http://dx.doi.org/10.1098/rsos.190669>

Received: 12 April 2019

Accepted: 29 November 2019

Subject Category:

Genetics and genomics

Subject Areas:

evolution/molecular biology

Keywords:

Gammaridae, paraphyletic, monophyletic, high-altitude adaptation, *atp8*, non-adaptive mutational pressure

Author for correspondence:

Hongtuo Fu

e-mail: fuht@frc.cn

†These authors contributed equally to the work.

Electronic supplementary material is available online at <http://dx.doi.org/10.6084/m9.figshare.c.4777508>.

Disentangling the interplay of positive and negative selection forces that shaped mitochondrial genomes of *Gammarus pisinnus* and *Gammarus lacustris*

Shengming Sun^{1,†}, Ying Wu^{1,†}, Xianping Ge^{1,2},
Ivan Jakovlić³, Jian Zhu^{1,2}, Shahid Mahboob^{4,5},
Khalid Abdullah Al-Ghanim⁴, Fahad Al-Misned⁴
and Hongtuo Fu¹

¹Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081, People's Republic of China

²Agriculture Ministry Key Laboratory of Healthy Freshwater Aquaculture, Zhejiang Institute of Freshwater Fisheries, Huzhou 313001, People's Republic of China

³Bio-Transduction Lab, Wuhan, People's Republic of China

⁴Department of Zoology, College of Science, King Saud University, PO Box 2455, Riyadh-11451, Riyadh, Saudi Arabia

⁵Department of Zoology, GC University, Faisalabad, Pakistan

[II, 0000-0002-2461-3712](https://orcid.org/0000-0002-2461-3712); [SM, 0000-0003-4969-5387](https://orcid.org/0000-0003-4969-5387)

We hypothesized that the mitogenome of *Gammarus lacustris* (GL), native to the Qinghai-Tibet Plateau, might exhibit genetic adaptations to the extreme environmental conditions associated with high altitudes (greater than 3000 m). To test this, we also sequenced the mitogenome of *Gammarus pisinnus* (GP), whose native range is close to the Tibetan plateau, but at a much lower altitude (200–1500 m). The two mitogenomes exhibited conserved mitochondrial architecture, but low identity between genes (55% *atp8* to 76.1% *cox1*). Standard (homogeneous) phylogenetic models resolved Gammaridae as paraphyletic, but 'heterogeneous' CAT-GTR model as monophyletic. In indirect support of our working hypothesis, GL, GP and *Gammarus fossarum* exhibit evidence of episodic diversifying selection within the studied Gammaroidea dataset. The mitogenome of GL generally evolves under a strong purifying selection, whereas GP evolves under directional (especially pronounced in *atp8*) and/or relaxed

selection. This is surprising, as GP does not inhabit a unique ecological niche compared to other gammarids. We propose that this rapid evolution of the GP mitogenome may be a reflection of its relatively recent speciation and heightened non-adaptive (putatively metabolic rate-driven) mutational pressures. To test these hypotheses, we urge sequencing mitogenomes of remaining *Gammarus* species populating the same geographical range as GP.

1. Background

The amphipod crustacean genus *Gammarus* Fabricius, 1775 (Senticaudata: Gammaroidea: Gammaridae) contains more than 200 described species [1], which makes it one of the most speciose crustacean genera [2]. These species are widely distributed throughout the Northern Hemisphere in various marine, brackish, freshwater and subterranean aquatic habitats [3]. *Gammarus lacustris* (GL) is a freshwater species with a nearly circumboreal range [1], widely distributed in alpine lakes and native to the Qinghai-Tibet Plateau (QTP) [3]. The QTP is the highest and largest plateau in the world, on average about 3000–4500 m in altitude [4]. As opposed to GL, *Gammarus pisinnus* (GP) inhabits rivers and lakes in a narrow geographical range comprising the plains south of the Taihang mountains, along the Weihe River in Shanxi Province [5], several hundred kilometres eastward from the eastern edge of the QTP, and at a much lower altitude of 200–1500 m [5]. The uplifting of the QTP dramatically changed the environmental conditions from an originally humid and warm climate to the currently dry and cold one [6]. Such ecological changes are expected to result either in the extinction of a species from the habitat or in major genetic and phenotypic adaptations to the new environment. Indeed, the QTP is typically inhabited by endemic species [7,8] highly adapted to the extreme environmental conditions, such as hypoxia, low temperatures and strong ultraviolet radiation [3,4,9–14]. Although amphipod crustaceans are the predominant macroscopic aquatic invertebrate in the QTP [15], specificities of the high-altitude adaption of crustaceans remain poorly understood. Furthermore, GL has been found at Tibetan localities exceeding elevations of 5000 m, which is currently the highest known altitudinal record for gammarids [8]. This exceptional ability of GL to inhabit extremely high altitudes, matched only by some molluscs among the aquatic invertebrates [7,8], is an indication that this species may be a very good model to investigate genetic adaptations to the high-altitude environment.

Owing to the abundance of mitochondria in cells, (mostly) maternal inheritance, the absence of introns, small genomic size (in metazoans), and an increasingly large set of available orthologous sequences, mitochondrial genomes (mitogenomes) have become a popular tool in population genetics [16], phylogenetics [17–20], diagnostics [20–22] and evolution [23–26]. Although long-considered to evolve under the purifying selection, there is increasing evidence that mitogenomes can undergo episodes of directional selection in response to shifts in physiological or environmental pressures [27]. The mitochondrial genome harbours genes involved in oxidative phosphorylation, and thus plays a central role in the energy production in the organism [28]. Due to low temperatures and hypoxia, life at high altitudes is expected to generate unique evolutionary pressures on the energy management system of organisms [9,10]. However, high-altitude adaptation can occur on different metabolic levels, not all of which include the mitochondrial metabolism [29,30], so mitochondrial genome may reflect only a small fragment of these adaptive changes. Despite this limitation, a number of studies have detected signals of directional selection in the mitogenomes of animals living at high elevations [30–32], including both terrestrial [11,12,29,33,34] and aquatic [35] Tibetan fauna. Therefore, our working hypothesis was that the speciation of GL (but not that of GP) was accompanied by mitogenomic adaptations to the extreme environment of the Tibetan plateau. To test this hypothesis, we sequenced the complete mitochondrial genomes of GL and GP, conducted comparative mitogenomic and phylogenomic analyses using a set of available gammarid mitogenomes, and studied evolutionary signals imprinted upon the two mitogenomes.

2. Results

2.1. Characteristics and architecture of the two mitogenomes

The length of complete mitochondrial genomes was: 15 907 bp in GP, 15 333 bp in GL. Both possessed the standard 13 protein-coding genes (PCGs), two rRNA genes (16S and 12S), 22 tRNA genes and a control region. Both mitogenomes exhibited an identical, highly conserved, gene order and strand distribution (table 1). The genes of the two mitogenomes also exhibited very similar sizes, partially shared start/

Table 1. Organization of the mitogenomes of *Gammarus lacustris* and *Gammarus pisinnus*. Data are presented as *G. lacustris*/*G. pisinnus*. IGN is intergenic region, where a negative number represents an overlap. Sizes are given in bp (base pairs). CR is the putative control region.

gene	position		to	size bp	IGN bp	codon		strand	identity %
	from	stop				start	stop		
<i>trnY</i>	1/1	61/62	61/62	61/62				+/–	62.9
<i>trnQ</i>	58/59	113/116	113/116	56/58	–4/–4			–/–	72.41
<i>trnC</i>	369/335	426/391	426/391	58/57	255/218			–/–	62.07
<i>trnI</i>	440/583	500/642	500/642	61/60	13/191			+/+	80.33
<i>trnM</i>	502/646	562/705	562/705	61/60	1/3			+/+	78.69
<i>nad2</i>	563/706	1544/1687	1544/1687	982/982		TTG/TTG	T–T/–	+/+	62.23
<i>trnW</i>	1545/1688	1604/1748	1604/1748	60/61				+/+	86.89
<i>trnG</i>	1606/1749	1665/1808	1665/1808	60/60	1/–			+/+	83.33
<i>cox1</i>	1666/1809	3202/3345	3202/3345	1537/1537		ATC/ATA	T–T/–	+/+	76.06
<i>trnL2</i>	3203/3346	3262/3405	3262/3405	60/60				+/+	85
<i>cox2</i>	3263/3406	3941/4084	3941/4084	679/679		TTG/ATT	T–T/–	+/+	72.31
<i>trnK</i>	3942/4085	4000/4142	4000/4142	59/58				+/+	88.14
<i>trnD</i>	4001/4143	4061/4203	4061/4203	61/61				+/+	86.89
<i>atp8</i>	4062/4204	4220/4359	4220/4359	159/156		ATT/ATA	TAA/TAA	+/+	55
<i>atp6</i>	4214/4353	4885/5024	4885/5024	672/672	–7/–7	ATG/TTG	TAA/TAA	+/+	66.67
<i>cox3</i>	4885/5024	5670/5809	5670/5809	786/786	–1/–1	ATG/ATG	TAA/TAA	+/+	73.03
<i>nad3</i>	5673/5811	6026/6164	6026/6164	354/354	2/1	ATG/ATG	TAG/TAG	+/+	64.41
<i>trnA</i>	6025/6163	6084/6221	6084/6221	60/59	–2/–2			+/+	70
<i>trnS1</i>	6084/6221	6135/6271	6135/6271	52/51	–1/–1			+/+	80.77
<i>trnN</i>	6138/6274	6198/6334	6198/6334	61/61	2/2			+/+	73.77
<i>trnE</i>	6196/6332	6257/6395	6257/6395	62/64	–3/–3			+/+	85.94

(Continued.)

Table 1. (Continued.)

gene	position		to	size bp	IGN bp	codon		stop	strand	identity %
	from					start				
<i>trnR</i>	6252/6390	6311/6450	6311/6450	60/61	-6/-6				+/+	77.05
<i>trnF</i>	6310/6449	6369/6507	6369/6507	60/59	-2/-2				-/-	73.33
<i>nad5</i>	6370/6508	8082/8226	8082/8226	1713/1719		GTG/TTG	TAA/TAA		-/-	60.03
<i>trnH</i>	8083/8227	8141/8286	8141/8286	59/60					-/-	80
<i>nad4</i>	8142/8287	9456/9601	9456/9601	1315/1315		ATG/TTG	T-/T-		-/-	67.76
<i>nad4 L</i>	9450/9595	9746/9888	9746/9888	297/294	-7/-7	ATA/ATG	TAA/TAG		-/-	61.95
<i>trnT</i>	9747/9891	9807/9950	9807/9950	61/60	-2				+/+	72.13
<i>trnP</i>	9808/9950	9867/10013	9867/10013	60/64	-/-1				-/-	74.24
<i>nad6</i>	9870/10016	10373/10516	10373/10516	504/501	2/2	ATG/ATG	TAA/TAA		+/+	55.56
<i>cytB</i>	10373/10516	11507/11652	11507/11652	1135/1137	-1/-1	ATG/ATG	T-/TAA		+/+	66.14
<i>trnS2</i>	11508/11651	11568/11712	11568/11712	61/62	-/-2				+/+	64.52
<i>nad1</i>	11594/11784	12505/12716	12505/12716	912/933	25/71	TTG/TTG	TAA/TAA		-/-	67.63
<i>trnL1</i>	12506/12717	12568/12775	12568/12775	63/59					-/-	66.67
<i>trnL</i>	12569/12776	13548/13763	13548/13763	980/988					-/-	71.72
<i>trnV</i>	13549/13764	13602/13814	13602/13814	54/51					-/-	74.07
<i>trnS</i>	13603/13815	14352/14531	14352/14531	750/717					-/-	65.3
CR	14353/14532	15333/15907	15333/15907	981/1376					-/-	

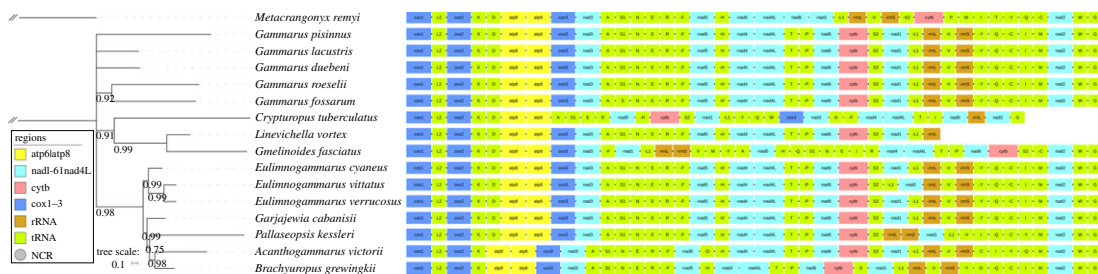


Figure 1. Mitochondrial phylogenomics and gene orders of the superfamily Gammaroidea. Phylogram was inferred using nucleotide sequences of all mitochondrial genes (PCG + rRNA + tRNA) and CAT-GTR model implemented in PhyloBayes. *Metacrangonyx remyi* is the outgroup (branch cropped). Scale bar corresponds to the estimated number of substitutions per site. Posterior probability support values less than 1.0 are shown next to nodes, and GenBank accession numbers are available in the electronic supplementary material, Additional file S1. Gene orders are shown to the right, with the legend included in the figure.

stop codons, and very similar intergenic regions (IGR) and overlaps. Size-wise, there were a few exceptions: *rrnS* gene was larger in GL than in GP (750 versus 717 bp), whereas *nad1* was much shorter in GL (912 bp) than in GP (933 bp); GP had a much larger IGR between *trnI* and *trnC* (191 versus 13 bp) and a much larger (putative) control region (1376 versus 981 bp). Several start codons were found, ATG, TTG, ATA, ATT, GTG and ATC, and most genes used the standard TAA stop codon (including an abbreviated form of it T-). An exception was *nad3* of both species and *nad4 L* of GP, which used TAG. GP had the A + T content of 70% and GL of 64.3%. None of these architectural features are novel or unique among the Gammaroidea and Senticaudata (electronic supplementary material, Additional file S1).

AT and GC skew values of complete mitogenomes were -0.068 and -0.31 (respectively) for GP; and -0.014 and -0.263 for GL. As regards individual genes, all PCGs and rRNAs exhibited predominantly negative AT skews, but tRNAs (concatenated) had positive AT skews (electronic supplementary material, Additional file S1). Several exceptions (inverted skews) were observed: the two rRNA genes in the GP mitogenome exhibited inverted AT skew values (*rrnL* = 0.004 and *rrnS* = 0.115); *rrnS* in *G. fossarum* and *Gmelinoides fasciatus* mitogenomes (0.016 and 0.017, respectively); *atp8* of *Eulimnogammarus vittatus* (0.008) and *G. roeselii* (0.036); and tRNAs of *Stygbromus indentatus* (-0.017).

However, although the architecture of the two mitogenomes was very similar, gene identity values were relatively low: 72% on average, ranging from 55% (*atp8*) to 88% (*trnK*). Among the 13 PCGs, the average identity was even lower (65.3%), with *cox1* exhibiting the highest value of 76.1%. As regards gene families, *cox* family (*cox1-3*) exhibited the highest average value (73.8%), followed by the *nad* family (62.8%) and finally the *atp* family (60.84%). Although tRNA remoulding has been reported in this group of crustaceans before [36], we did not find evidence for such evolutionary events in the two newly sequenced species.

2.2. Phylogeny and gene order

Gene order was perfectly conserved in the entire Gammaridae family, including the two newly sequenced species (figure 1). Different datasets and algorithms produced notable topological instability. Apart from the CAT-GTR analysis of the PCGRT dataset (figure 1), other analyses produced deeply paraphyletic family Gammaridae, mostly with three or even four Baikal lake families (Acanthogammaridae, Pallaseidae, Eulimnogammaridae + Micruropodidae) nested within the clade (electronic supplementary material, Additional file S2). CAT-GTR model resolved Gammaridae as monophyletic, notwithstanding several nodes in the phylogram that exhibited unresolved polyphyletic relationships. Apart from Acanthogammaridae, which were rendered paraphyletic by the nested Pallaseidae (*Pallaseopsis kessleri*), family, the remaining Gammaroidea families (Micruropodidae and Eulimnogammaridae) were also monophyletic (taxonomic details available in electronic supplementary material, Additional files S1 and S2).

2.3. Direction and magnitude of selective pressures

The rate of non-synonymous (dN) to synonymous (dS) substitutions, $\omega = dN/dS$, is a commonly used indicator for measuring the direction and magnitude of selective pressure on genes, where $\omega < 1$

indicates purifying selection, $\omega = 1$ indicates neutral evolution and $\omega > 1$ indicates directional (or positive) selection [37]. In the dataset comprised of 13 concatenated PCGs, the number of segregating (variable) sites was 1283, total number of observed mutations was 2061, among which 1865 were synonymous, and 196 were non-synonymous.

First, we used aBSREL, which allows a global test for positive selection in the entire dataset [38]. Using the concatenated 13 PCGs dataset with all 15 species (apart from the outgroup *M. remyi*) selected as test branches, we found statistically significant evidence (p -value corrected for multiple testing ≤ 0.05) of episodic diversifying selection in two species: *Gammarus fossarum* ($p = 0.0026$), and the newly sequenced GP ($p = 0.0095$). The corrected p -value of GL was also very close to the significance threshold ($p = 0.0658$; the uncorrected p -value was 0.05). For the global test, p -values at each branch must be corrected for multiple testing, which decreases the selection detection power of this exploratory approach [38], so we then used the same tool to conduct independent tests for GL and GP (selected as the test branch), and the remaining species as reference branches. Results of these two tests were as expected, within the context of the global test: evidence of episodic diversifying selection was identified in both species (p -values: GL = 0.0052, GP = 0.0007).

Using the BUSTED tool, which 'provides a gene-wide (not site-specific) test for positive selection by asking whether a gene has experienced positive selection at at least one site on at least one branch' [39], we found evidence of gene-wide episodic diversifying selection in both species (independently selected as test branches): $p = 0.048$ for GL, and $p = 0.001$ for GP.

Relaxation of the efficiency or intensity of natural selection can result in a loss of function, but also drive evolutionary innovation [40], so we also explored the direction and magnitude of selection pressures on GL and GP using the RELAX algorithm, which asks whether the strength of natural selection has been relaxed or intensified along a specified set of test branches. While not suitable for explicitly testing for positive selection, this method is useful for identifying trends and/or shifts in the stringency of natural selection on a given locus [40]. A significant result of $K > 1$ indicates that selection strength has been intensified along the test branches, and a significant result of $K < 1$ indicates that selection strength has been relaxed along the test branches. Both analyses produced statistically significant results, but the tests indicate selection *intensification* in GL, and selection *relaxation* in GP (table 2).

To get a closer perspective and test which genes may be evolving under unique selection pressures, we divided the dataset into individual genes and conducted another round of analyses on all 13 individual genes (table 2). We did not find statistically significant evidence of directional selection (aBSREL + BUSTED) in any of the GL genes, but p -values indicate that *cox3* may be undergoing higher levels of directional selection pressures than other genes. RELAX analyses revealed that all GL genes except *nad4 L* (significant) and *nad6* (non-significant) exhibited intensified selection (six significant: *atp6*, *cox1*, *cox3*, *cytb*, *nad3* and *nad5*). In GP, *atp8* (aBSREL + BUSTED) and *nad3* (aBSREL) exhibited statistically significant evidence of directional selection. Most genes exhibited different levels of relaxed selection pressures (statistically significant in *cytb*, *nad4* and *nad6*), with only *cox1*, *nad4 L* and *nad5* exhibiting (non-significantly) intensified selection pressures.

To further explore the selection signals using a per-site-based approach in selected genes of interest (table 3), we relied on the FEL algorithm, which uses a maximum-likelihood approach to infer dN and dS substitution rates on a per-site basis for a given coding alignment and corresponding phylogeny [37]. First, we tested the entire dataset, and then genes for which other analyses (table 2) indicated that they may be evolving under unique selection pressures (significance threshold was set at $p < 0.05$ for all analyses). For the entire dataset (13PCG), the sites exhibiting predominantly negative (purifying) selection by far outnumbered the sites exhibiting predominantly positive (diversifying) selection (897 versus 17, respectively) in GL, but the situation was reversed in GP, where sites exhibiting pervasive positive/diversifying selection by far outnumbered the sites exhibiting pervasive negative/purifying selection (619 versus 0, respectively). This was reflected in individual genes as well, where all three GP genes (*atp8*, *cox3* and *nad3*) exhibited only positive selection. In GL, *nad4 L* and *nad6* also exhibited only positive selection, whereas *cox3* exhibited only negative selection.

Finally, to better examine the signal of episodic diversifying selection in *Gammarus fossarum* observed in the exploratory aBSREL test ($p = 0.0026$), we conducted a few additional analyses with this species selected as the test branch. aBSREL ($p = 0.0002$) and BUSTED found evidence ($p = 0.000$) of episodic diversifying selection. RELAX test for selection relaxation ($K = 0.84$) was also significant ($p = 0.001$, LR = 11.52), which was supported by the FEL analysis, where sites exhibiting pervasive

Table 2. Selection signals in the mitogenomes of *Gammarus lacustris* (GL) and *Gammarus pisininus* (GP) inferred using BUSTED, aBSREL and RELAX algorithms. 13PCG is the dataset comprising all 13 concatenated (non-partitioned) protein-coding genes. Sites column shows the number of sites in the alignment. *K* column: a statistically significant $K > 1$ indicates that selection strength has been intensified, and $K < 1$ that selection strength has been relaxed. LR is likelihood ratio, *p* is *p*-value and *D* indicates the direction of selection pressure change: intensified (I) or relaxed (R), where * highlights a statistically significant ($p < 0.05$) result.

data	sites	<i>Gammarus lacustris</i> (GL)				<i>Gammarus pisininus</i> (GP)							
		BUSTED		RELAX		BUSTED		RELAX					
		<i>p</i> -value	aBSREL <i>p</i> -value	<i>K</i>	<i>D</i>	<i>p</i> -value	LR	<i>p</i> -value	LR				
13PCG	3743	0.024*	0.005*	1.21	0.001	11.54	I*	0.001*	0.010*	0.63	56.39	0.000	R*
<i>atp6</i>	223	1.000	1.000	14.80	0.006	7.48	I*	0.487	1.000	0.68	1.51	0.219	R
<i>atp8</i>	58	1.000	1.000	1.02	0.897	0.02	I	0.000*	0.004*	0.13	1.14	0.286	R
<i>cox1</i>	527	1.000	1.000	1.87	0.000	12.31	I*	1.000	1.000	1.00	-0.38	1.000	I
<i>cox2</i>	231	1.000	1.000	1.60	0.097	2.76	I	1.000	1.000	1.00	0.00	1.000	R
<i>cox3</i>	262	0.345	0.087	2.24	0.005	8.05	I*	0.327	0.144	0.45	-6.96	1.000	R
<i>cytb</i>	380	1.000	0.168	2.37	0.010	6.63	I*	1.000	1.000	0.46	7.91	0.005	R*
<i>nad1</i>	313	1.000	1.000	1.18	0.343	0.90	I	0.749	1.000	0.77	3.24	0.072	R
<i>nad2</i>	330	0.718	1.000	1.01	0.543	0.37	I	0.562	0.431	0.84	0.42	0.516	R
<i>nad3</i>	119	1.000	1.000	7.91	0.008	6.98	I*	0.049*	0.454	0.64	2.28	0.131	R
<i>nad4</i>	447	1.000	1.000	1.04	0.530	0.40	I	1.000	1.000	0.58	7.75	0.005	R*
<i>nad4 L</i>	98	1.000	1.000	0.36	0.026	4.98	R*	1.000	1.000	2.05	2.61	0.106	I
<i>nad5</i>	579	0.741	0.385	1.38	0.040	4.23	I*	0.460	0.500	2.15	-45.82	1.000	I
<i>nad6</i>	176	0.766	0.500	0.95	1.000	-0.09	R	1.000	1.000	0.49	4.41	0.036	R*

Table 3. Selection signals in the selected genes of mitogenomes of *Gammarus lacustris* (GL) and *Gammarus pisinnus* (GP) inferred using a per-site-based method—FEL. 13PCG is the dataset comprising all 13 concatenated (non-partitioned) protein-coding genes. Sites column shows the number of sites in the alignment. + indicates the number of sites evolving under significant (threshold set at $p < 0.05$) pervasive positive/diversifying selection, and – under pervasive negative/purifying selection.

data	sites	GL		GP	
		+	–	+	–
13PCG	3743	17	897	619	0
<i>atp8</i>	58			18	0
<i>cox3</i>	262	0	78	37	0
<i>nad3</i>	119			17	0
<i>nad4 L</i>	98	19	0		
<i>nad6</i>	176	14	0		

positive/diversifying selection by far outnumbered the sites exhibiting pervasive negative/purifying selection (346/0 respectively; $p < 0.05$).

3. Discussion

To test the hypothesis that mitochondrial genome of GL may exhibit imprints of genetic adaptation to a high-altitude environment of the Tibetan plateau, we compared it to the mitogenome of a congeneric species inhabiting lower altitudes, GP, as well as the dataset comprising all currently available Gammaroidea mitogenomes. To achieve this, we sequenced the complete mitochondrial genomes of these two species, conducted comparative mitogenomic and phylogenetic analyses, and applied a number of different tests to study the evolutionary forces that shaped these mitogenomes.

3.1. General comparative mitogenomics

Both mitogenomes exhibited a fairly conserved architecture, standard for Gammaroidea and Senticaudata (figure 1; electronic supplementary material, Additional file S1). A notable exception was the truncated *nad1* gene in GL (912 bp), also shared by *Gammarus duebeni* (913 bp), whereas most other Senticaudata (including GP) exhibited a larger size (924–933 bp). As the 3' of this gene is poorly conserved in Senticaudata, this is an indication of its limited importance for the functionality of this gene, so it is unlikely that the truncation was adaptive, and more likely that it is merely a consequence of a non-adaptive random mutational event introducing a stop codon into a non-essential segment of the gene.

3.2. Strand asymmetry and skews

Strand asymmetry is a relatively common feature of organellar genomes [41–43], including those of gammarids [36], but it can obstruct phylogenetic reconstruction and produce gravely misleading phylogenetic artefacts (long-branch attraction) in cases when unrelated taxa exhibit inverted skews [44,45]. GP exhibited the lowest AT skew value among the available Gammaroidea mitogenomes (–0.068) and a low GC skew value of –0.31, whereas GL exhibited a comparatively standard AT (–0.014) and GC (–0.263) skew values. GP also exhibited the highest A + T content (70%) among the Gammaroidea mitogenomes, whereas GL had a comparatively low A + T content of 64.3%. This is an indication that the evolutionary forces (adaptive or non-adaptive) that shaped the mitogenome of GP may have differed from those of other gammarids. Whereas most PCGs had negative GC skews, *nad1*, *nad4*, *nad4 L*, *nad5* and all RNA genes had positive GC skews. Importantly, these genes are all encoded on the minus strand, which is in agreement with the hypothesis that strand asymmetry is caused by hydrolytic deamination of bases in the leading DNA strand when it is single stranded, i.e. during replication and transcription [43]. The only individual outlier in relation to this pattern was the GC skew value of tRNAs of *S. indentatus* (–0.017) (also an outlier in the AT skew). Disregarding the

minor exception of GP, gammarids exhibit relatively consistent skews, so we can tentatively conclude that there are no indications that mitogenomes are not a suitable tool for reconstructing the phylogeny of gammarids.

3.3. Phylogeny and gene order

In congruence with the topologies obtained using the standard (homogeneous) models in combination with data partitioning (electronic supplementary material, Additional file S2), paraphyly of Gammaridae has been reported in a relatively recent multi-gene-based revision of the Gammaroidea phylogeny [46], and the split of the Baikal lake Gammaroidea into two clades has also been reported by several studies [46–48]. However, using an almost complete mitogenomic dataset (PCGRT) in combination with a ‘heterogeneous’ CAT-GTR model, we managed to obtain monophyletic (polytomy notwithstanding) Gammaridae. This supports a previous observation that CAT-GTR model appears to be more efficient at dealing with compositional heterogeneity than data partitioning [45]. Although we tentatively accepted this as the correct topology, this reasoning is somewhat circular, so the monophyly of this family should be corroborated using additional datasets and data types before this issue can be declared as resolved. The close relationship of GL and *G. duebeni* in all of our topologies (polyphyly in CAT-GTR notwithstanding) is in disagreement with the findings of Hou & Sket [46], and rather unexpected, as *G. duebeni* is distributed throughout the intertidal zone of the North Atlantic region [49]. It might be merely an artefact caused by the poor sampling of gammarid mitogenomes, resulting in unresolved polyphyletic relationships, or even particular evolutionary pressures that have shaped the mitogenome of GL. A much denser sampling of closely related gammarid mitogenomes would be needed to test these hypotheses.

While the two newly sequenced mitogenomes exhibited perfectly conserved gene orders (among the Gammaridae), a representative of the Baikalian family Micrurpodidae, *Gmelinoides fasciatus*, exhibited a highly rearranged gene order (figure 1), which is very unusual among the Gammaroidea. The remaining two available Micrurpodidae mitogenomes are incomplete, so it is impossible to infer conclusions with confidence, but *Lineoichella vortex* appears to possess a conserved gammarid gene order, whereas *Crypturopus tuberculatus* also seems to possess a highly rearranged and unique gene order (figure 1). This contrast between the highly conserved architecture in the rest of the Gammaroidea dataset and highly volatile gene order in some Micrurpodidae might offer strong evidence in support of the hypothesis that the evolution of mitogenomic architecture is highly discontinuous [23]. Therefore, it would be of greatest interest to sequence further mitogenomes belonging to the Micrurpodidae and the entire Baikal group 2 of gammarids, as defined by Hou & Sket [46].

3.4. Selective pressures

Given the central role of the mitochondrial genome in the energy production in animals [31], we hypothesized that the mitogenome of GL may exhibit signs of adaptation (directional selection) to life at high altitudes. In agreement with our hypothesis, we did find evidence for directional selection in GL using two different algorithms (aBSREL and BUSTED). Although, results of RELAX and FEL algorithms indicate that signals of purifying selection outweigh the signals of directional selection in the mitogenome of this species, this is expected, as purifying selection is the predominant force driving the evolution of mitogenomes, and directional selection signals are usually constrained to specific functional sites within genes [30,50,51].

It should be noted here that GL possesses relatively high passive dispersal potential, largely due to its ability to be dispersed short distances via waterfowl, which results in a broad distribution range of this species [1,8]. This makes it unique both among other Tibetan life forms, which tend to be endemic [7], and other freshwater *Gammarus* species, which generally have weak dispersal potential [5]. Therefore, we should not exclude a possibility that Tibetan GL populations may be receiving a limited influx of migrants from other locations, which may weaken the directional selection signal. As this species is generally well-adapted to life in alpine lakes, these migrants may be able to survive in the extreme environment of the QTP. However, molecular phylogeographic patterns indicate (due to limited data, this is not strongly supported) that gammarid fauna of the Tibetan plateau is probably the result of a single recent colonization event, and that genetic admixture occurs only among different Tibetan populations [8]. Therefore, it is not a likely scenario. Furthermore, we would still expect the migrants to exhibit adaptations to life at high altitudes. Future studies may attempt to sequence mitogenomes

from lowland GL populations, and ideally even those inhabiting localities exceeding elevations of 5000 m [8], and conduct comparative analyses on the entire dataset.

Importantly, and unexpectedly, all tests indicate the existence of an even stronger signal of positive/diversifying selection, as well as overall significantly relaxed purifying selection, in the mitogenome of the other newly sequenced species, GP. This fast evolution of GP is also a likely underlying reason for the surprisingly low (for congeners) [51,52] gene identity values between GL and GP. The evidence for relaxation of functional evolutionary constraints (and/or positive selection) was reported in the transcriptome of a subterranean *Gammarus* species [53], which can be explained by adaptation to a radically different (subterranean) environment. However, this finding is surprising for GP, as this species does not inhabit a unique ecological niche compared to other gammarids.

As mentioned above, most freshwater *Gammarus* species generally have weak dispersal potential, so their phylogeny is strongly influenced by geological barriers, which results in the existence of many endemic species, often with altitude-specific distribution ranges [5]. We hypothesize that the speciation of GP was most likely a result of fragmentation of aquatic habitats and changes in the environment that accompanied the rapid uplifting of the Taihang range from a vast plain that predated it [5]. This process opened new ecological niches, where mutant phenotypes could be advantageous. For example, it has been proposed that uplifting of the Lüliang and Taihang ranges promoted several speciation events in gammarids, resulting in the existence of several (putative) gammarid species (including GP) in a relatively narrow geographical range [5]. Such scenario may explain the observed rapid evolution of the GP mitogenome, as well as the proposed heightened speciation intensity in these mountain ranges.

Moreover, this may also explain the much more pronounced directional selection pressure signal in the mitogenome of GP, compared to that of GL. The geological ages of the two habitats are very different; the uplift of the Tibetan plateau began during the Upper Cretaceous, some 70 Ma (million-years-ago), whereas the uplift of the Taihang mountain range occurred during the Miocene (23 to 5 Ma) [3,5]. Therefore, it is possible that GL underwent the adaptation to the new environment much earlier than GP. Such scenario would allow the purifying selection to swamp the signal of episodic positive evolution in GL, but not in GP. Note that this appears incongruent with the proposal mentioned above, that Tibetan GL populations are the result of a single recent colonization event [8], but the authors did not propose the exact time frame for 'recent', so we cannot reject it with certainty.

Furthermore, as accelerated mitochondrial evolution can be driven by directional adaptive selection or by non-adaptive mutational pressures [14,27,45], we should not exclude a possibility that the comparatively strong signal in GP could be a combination of both. The exceptionally high GC skew value of the GP mitogenome may be an indirect evidence for this, as base composition skews are believed to be mostly caused by non-adaptive mutational pressures, usually driven by architectural rearrangements or heightened metabolic rates [43–45]. Given the highly conserved architecture of the gammarid mitogenomes, we can probably exclude architectural rearrangements as the underlying cause, and instead propose that it may be associated with a (hypothetical) comparatively high metabolic rate in the GP. Although it is not excessive to speculate that species living in moderate climate environments would exhibit higher metabolic rates than species inhabiting extremely cold environment of the Tibetan plateau, this hypothesis would first have to be tested. Specific questions that remain open include: do other species described in the same geographical range [5] exhibit heightened evolutionary rates?; and do they exhibit higher evolutionary rates than *Gammarus* species inhabiting other moderate environments? Within the studied dataset, three (out of five) *Gammarus* species exhibited significant signals of episodic diversifying/relaxed selection: GL, GP and *G. fossarum*. This is an indication that some *Gammarus* species indeed do exhibit comparatively accelerated mitochondrial evolution rates in comparison to other Senticaudata. As these questions warrant further studies, it would be very interesting to sequence the mitogenomes of the remaining species described in that geographical range [5] and test the hypothesis that these mitogenomes should also exhibit comparatively high evolutionary rates within the entire *Gammarus* dataset. These molecular data would also allow us to further test the extent of isolation and gene flow between these proposed species.

In agreement with results obtained using the concatenated dataset, individual genes in the mitogenome of GL predominantly exhibited varying levels of purifying selection, whereas that of GP mostly exhibited relaxed selection. However, there were some outliers: *nad4 L* and *nad6* of GL exhibited relaxed selection pressure (only *nad4 L* significantly), whereas *cox1*, *nad4 L* and *nad5* of GP exhibited intensified selection pressure (all non-significant). Some of these outliers are expected, as *cox1* is generally the most conserved gene in animal mitogenomes [51,52,54], which includes both

other gammarids (Baikal lake) [36] and other Tibetan animals [31]. Our comparative analyses also indicate relatively high levels of purifying selection in the *cox* gene family of the two species (73–76% interspecific identity), and somewhat relaxed (or positive) selection in the *nad* family (55–68% identity). *nad4 L* is a very small gene that often overlaps with other genes (*nad4*) in many crustacean mitogenomes [17], which is an indication of reduced mutational constraints [55]; and *nad6* is often among the fastest-evolving mitochondrial genes [56,57], and sometimes even altogether missing from them [58]. Although this may explain the outliers in the GL mitogenome, it remains unclear why *nad4 L* and *nad5* of GP exhibited intensified selection pressure. As both results were non-significant, it may be just a statistical fluke.

A surprising finding was that, among the identified genes that may be especially strongly affected by selective pressures, GP *atp8* exhibited exceptionally strong evidence of directional selection. This is intriguing, as *atp8* is a very small, ‘dispensable’ gene, often missing from metazoan mitogenomes, and usually evolving under highly relaxed selection pressures [58]. Combined with the absence of evidence for a relaxation of selection constraints and very low similarity to the GL orthologue, we can conclude that our data indicate that *atp8* is indeed evolving under exceptionally strong directional selection in GP, which warrants further functional studies.

4. Conclusion

We identified the existence of directional evolution signal in the mitogenome of GL, which indirectly supports our working hypothesis, but a surprising finding is that the signal of relaxed/directional selective pressures is much stronger in the mitogenome of GP. This is surprising, as GP does not inhabit a unique ecological niche compared to other gammarids. We hypothesize that the speciation of GP was most likely a result of fragmentation of aquatic habitats and changes in the environment that accompanied the rapid uplifting of the Taihang range from a vast plain that pre-dated it, which opened new ecological niches where mutant phenotypes could be advantageous. Such scenario may explain the observed rapid evolution of the GP mitogenome, as well as the proposed heightened speciation intensity in these mountain ranges. We hypothesize that much older age of the Tibetan plateau may have allowed the purifying selection to swamp the signal of episodic positive evolution in GL, whereas the relative recency of its speciation may explain the stronger signal in GP. We propose that this hypothesis can be tested by sequencing the mitogenomes of the remaining six (proposed) species populating the same narrow geographical range. As all of these species are believed to be a result of these same geological events, their mitogenomes should exhibit similar evolutionary footprints. Furthermore, as we found indications that some *Gammarus* species exhibit comparatively accelerated mitochondrial evolution rates in comparison to other Senticaudata, we propose that future studies should not exclude the existence of unusually strong non-adaptive mutational pressures in their mitogenomes.

5. Methods

5.1. Sampling

GL samples were collected at the altitude of ca3400 m, in the Cuomujiri lake (92°09′–98°47′ N, 26°52′–30°40′ E) near the Nyingchi city in Tibet (figure 2). GP samples were collected at the altitude of ca510 m, from the Yellow River near the Xian City, Shanxi province, China (34°16′ N, 108°54′ E). Both samplings were conducted in August 2017 using fine-meshed hand nets. Approximately 50 specimens were collected at each site, immediately preserved in 95% ethanol in the field, and long-term stored at –20°C for later use. Samples were morphologically identified under the light microscope as described before [3,5]. As gammarids are non-protected invertebrates, no permits were needed for sampling and sample processing.

5.2. DNA extraction, amplification and sequencing

These steps were conducted as described before [23,57,59], so only an overview is provided. DNA was isolated from a complete single specimen (AidLab DNA extraction kit, AidLab Biotechnologies, Beijing, China) of each species after rinsing them in distilled water. Conserved fragments of several genes were amplified and sequenced using degenerate primer pairs (electronic supplementary

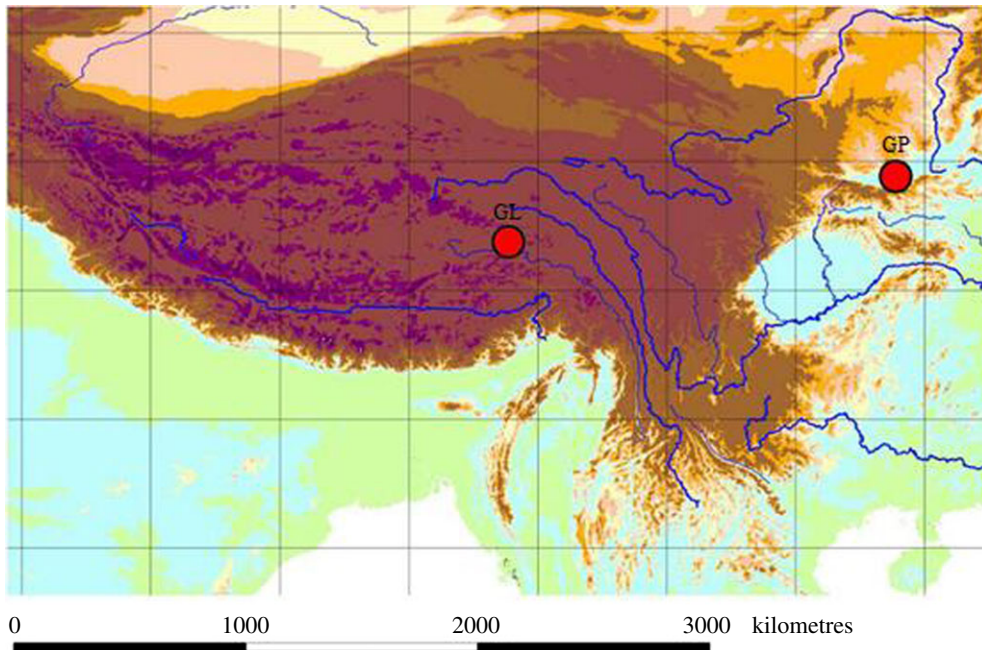


Figure 2. Approximate sampling localities for *Gammarus pisinnus* (GP) and *Gammarus lacustris* (GL).

material, Additional file S3) designed using available gammarid mitogenomes as templates. These gene fragments were then used to design specific primers (overlapping by approx. 100 bp) for amplification and sequencing of the whole mitogenome. PCR reaction mixture (20 μ l): 7.4 μ l ddH₂O, 10 μ l 2 \times PCR buffer (Takara, Dalian, China), 0.6 μ l of each primer, 0.4 μ l rTaq polymerase (250 U, Takara) and 1 μ l DNA template. Amplification: 98°C/2 min, 40 cycles of 98°C/10 s, 50°C/15 s, 68°C 1 min kb⁻¹ and finally 68°C/10 min. PCR products were sequenced bi-directionally using the Sanger method and the same set of primers.

5.3. Sequence assembly and analysis

These steps, assembly, annotation and comparative analyses, were also conducted following the methodology described before in more detail [23,57,59]. Briefly, sequenced fragments were quality-proofed by visually inspecting the electropherograms and queried against the GenBank using BLAST to confirm that the amplicon is the target sequence. During the manual assembly, conducted with the help of DNASTar v. 7.1 [60], we made sure that the overlaps were identical, mitogenome circular, and that no mitochondrial DNA copies (numts) [61] were incorporated. ORFs for PCGs were located using DNASTar, and manually fine-tuned via a comparison with available gammarid orthologues using BLAST and BLASTx. tRNAscan [62] and ARWEN [63] were used to identify tRNAs. PhyloSuite [64] was used to parse and extract mitogenomes annotated in Microsoft Word documents, as well as to create the files for submission to GenBank and generate comparative mitogenomic architecture tables. Following the arguments put forward before [65], instead of referring to strands as heavy and light, we used plus (+) and minus (–) to denominate the strands encoding a majority and minority of genes, respectively. The mitogenomes are available from the GenBank repository under the accession numbers MK354235 (GL) and MK354236 (GP).

5.4. Comparative mitogenomic and phylogenetic analyses

To minimize the impact of compositional heterogeneity, for phylogenetic and selection pressure analyses we used only all 13 mitogenomes belonging to the superfamily Gammaroidea available in the GenBank, plus *Metacrangonyx remyi* (Senticaudata: Hadzioidea) [66] as the outgroup. As some Gammaroidea mitogenomes in that dataset were incomplete, for comparative analyses, we retrieved all 15 available mitogenomes belonging to the infraorder Senticaudata (class Malacostraca: order Amphipoda) available in the RefSeq database (electronic supplementary material, Additional file S1; alignments available in the electronic supplementary material, Additional file S4). PhyloSuite was used to

batch-download all selected mitogenomes from the GenBank, extract genomic features, translate genes into amino acid sequences, and semi-automatically re-annotate ambiguously annotated tRNA genes with the help of the ARWEN output. The GenBank taxonomy was automatically replaced with the WoRMS database taxonomy using PhyloSuite, as the latter tends to be more up to date [67].

To test the stability of topology, we conducted phylogenetic analyses using different methods and datasets. Datasets used were: amino acid sequences of all 13 concatenated PCGs (PCG_AAs), nucleotides of concatenated 13 PCGs (PCG_NUC), and nucleotides of all mitochondrial genes (PCGRT: PCGs + rRNAs + tRNAs). Genes were aligned in batches with MAFFT [68], using ‘-auto’ strategy, and ‘codon’ alignment mode for nucleotides and ‘normal’ mode for amino acid sequences of PCGs; whereas all RNA genes were aligned using Q-INS-i algorithm, which takes secondary structure information into account. PartitionFinder [69] was used to infer the best partitioning strategy and select best-fit evolutionary models for each partition using the greedy algorithm and AICC criterion (electronic supplementary material, Additional file S4). To infer the phylogeny, we used two standard methods, both of which assume compositional homogeneity of data (herein referred to as ‘homogeneous’ models), maximum-likelihood (ML) conducted using IQ-TREE [70] and Bayesian inference (BI) conducted using MrBayes 3.2.6 [71]. ML analysis was conducted with 100 standard bootstraps, and BI analysis was conducted with default settings, 2×10^6 metropolis-coupled MCMC generations (analysis was stopped as the average standard deviation of split frequencies was sufficiently low: 0.0024). All these analyses were conducted with the help of PhyloSuite. As compositional heterogeneity can compromise phylogenetic analyses, phylogeny was also inferred using the CAT-GTR site mixture model implemented in PhyloBayes-MPI 1.7a, which allows for site-specific rates of mutation (herein referred to as ‘heterogeneous’ model) [72]. This analysis was conducted on beta version of the Cipres server (<https://cushion3.sdsc.edu/portal2/tools.action>) [73], with default parameters (burnin = 500, invariable sites automatically removed from the alignment, two MCMC chains), -dgam set to 8, and the analysis was automatically stopped when the conditions considered to indicate a good run were reached: maxdiff less than 0.1 and minimum effective size greater than 300 (PhyloBayes manual). Phylograms and gene orders were visualized in iTOL [74] using files generated by PhyloSuite.

5.5. Selection pressure analyses

We used nucleotides of 13 PCGs (aligned using the codon mode; electronic supplementary material, Additional file S4) and PhyloBayes topology (figure 1) as the guidance tree. The numbers of segregating sites and mutations were estimated using DnaSP [75]. We tested a number of different evolutionary hypotheses using tools available from the Datamonkey webserver [76]: aBSREL to test if positive selection has occurred on a proportion of branches [38], BUSTED to test for a gene-wide (not site-specific) positive selection [39], RELAX to detect relaxed selection [40], and FEL to infer non-synonymous (dN) and synonymous (dS) substitution rates on a per-site basis [37].

Animal ethics. This study was approved by the Animal Care and Use Committee of the Centre for Applied Aquatic Genomics at Chinese Academy of Fishery Sciences.

Permission to carry out fieldwork. As sampling was conducted on unprotected invertebrates and on public land, fieldwork permits were not required.

Data accessibility. The two new mitogenomes are available from the GenBank repository under accession numbers MK354235 (*Gammarus lacustris*) and MK354236 (*Gammarus pisinus*), and all other data generated or analysed during this study are included in this published article and its electronic supplementary material, information files. **Authors’ contributions.** S.S. and X.G. participated in the conceptualization and design of the study, collection of samples, laboratory work, data analysis and drafting the manuscript; Y.W. participated in the collection of samples, laboratory work, data analysis and critically revised the manuscript for important intellectual content; I.J. participated in the design of the study, data analysis, interpretation and visualization and drafting the manuscript; J.Z., S.M., K.A.A.G. and F.A.M. participated in the molecular laboratory work, data analysis and visualization, and critically revised the manuscript for important intellectual content; H.F. conceived and coordinated the study, and critically revised the manuscript for important intellectual content. All authors approved the final version and agree to be accountable for all aspects of the work.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by National Natural Science Foundation of China (no. 31402280); the New Varieties Creation Major Project in Jiangsu Province (PZCZ201745); the Science & Technology Supporting Program of Jiangsu Province (BE2016308); the China Agriculture Research System-48 (CARS-48); and the Deanship of Scientific Research at King Saud University (1440-0138).

References

- Vainola R, Witt JDS, Grabowski M, Bradbury JH, Jazdzewski K, Sket B. 2008 Global diversity of amphipods (Amphipoda; Crustacea) in freshwater. *Hydrobiologia* **595**, 241–255. (doi:10.1007/s10750-007-9020-6)
- Costa FO, Henzler CM, Lunt DH, Whiteley NM, Rock J. 2009 Probing marine *Gammarus* (Amphipoda) taxonomy with DNA barcodes. *Syst. Biodivers.* **7**, 365–379. (doi:10.1017/S147720009990120)
- Hou Z, Li S. 2018 Four new *Gammarus* species from Tibetan Plateau with a key to Tibetan freshwater gammarids (Crustacea, Amphipoda, Gammaridae). *Zookeys* **747**, 1–40. (doi:10.3897/zookeys.747.21999)
- Qi D, Chao Y, Guo S, Zhao L, Li T, Wei F, Zhao X. 2012 Convergent, parallel and correlated evolution of trophic morphologies in the subfamily Schizothoracinae from the Qinghai-Tibetan Plateau. *PLoS ONE* **7**, e34070. (doi:10.1371/journal.pone.0034070)
- Hou Z, Li J, Li S. 2014 Diversification of low dispersal crustaceans through mountain uplift: a case study of *Gammarus* (Amphipoda: Gammaridae) with descriptions of four novel species. *Zool. J. Linnean Soc.* **170**, 591–633. (doi:10.1111/zoj.12119)
- Wu Z, Barosh PJ, Wu Z, Hu D, Zhao X, Ye P. 2008 Vast early Miocene lakes of the central Tibetan Plateau. *Geol. Soc. Am. Bull.* **120**, 1326–1337. (doi:10.1130/B26043.1)
- Clewing C, Albrecht C, Wilke T. 2016 A complex system of glacial sub-refugia drives endemic freshwater biodiversity on the Tibetan Plateau. *PLoS ONE* **11**, e0160286. (doi:10.1371/journal.pone.0160286)
- Clewing C, Wilke T, Ilge A, Albrecht C. 2016 Phylogenetic patterns of freshwater amphipods inhabiting the Tibetan Plateau. *Crustaceana* **89**, 239–249. (doi:10.1163/15685403-00003518)
- Qiu Q *et al.* 2012 The yak genome and adaptation to life at high altitude. *Nat. Genet.* **44**, 946–949. (doi:10.1038/ng.2343)
- Qu Y *et al.* 2013 Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat. Commun.* **4**, 2071. (doi:10.1038/ncomms3071)
- Gou X *et al.* 2014 Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**, 1308–1315. (doi:10.1101/gr.171876.113)
- Bigham A *et al.* 2010 Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genet.* **6**, e1001116. (doi:10.1371/journal.pgen.1001116)
- Zhang D *et al.* 2017 Genetic adaptation of Schizothoracine fish to the phased uplifting of the Qinghai–Tibetan Plateau. *G3: Genes, Genomes, Genetics* **7**, 1267–1276. (doi:10.1534/g3.116.038406)
- Guan L, Chi W, Xiao W, Chen L, He S. 2014 Analysis of hypoxia-inducible factor alpha polyploidization reveals adaptation to Tibetan plateau in the evolution of schizothoracine fish. *BMC Evol. Biol.* **14**, 192. (doi:10.1186/s12862-014-0192-1)
- Scheinfeldt LB, Tishkoff SA. 2010 Living the high life: high-altitude adaptation. *Genome Biol.* **11**, 133. (doi:10.1186/gb-2010-11-9-133)
- Bourguignon T *et al.* 2018 Transoceanic dispersal and plate tectonics shaped global cockroach distributions: evidence from mitochondrial phylogenomics. *Mol. Biol. Evol.* **35**, 970–983. (doi:10.1093/molbev/msy013)
- Zou H, Jakovlic I, Zhang D, Chen R, Mahboob S, Al-Ghanim KA, Al-Misned F, Li WX, Wang GT. 2018 The complete mitochondrial genome of *Cymothoa indica* has a highly rearranged gene order and clusters at the very base of the Isopoda clade. *PLoS ONE* **13**, e0203089. (doi:10.1371/journal.pone.0203089)
- Perkins EM, Donnellan SC, Bertozzi T, Whittington ID. 2010 Closing the mitochondrial circle on paraphyly of the Monogenea (Platyhelminthes) infers evolution in the diet of parasitic flatworms. *Int. J. Parasitol.* **40**, 1237–1245. (doi:10.1016/j.ijpara.2010.02.017)
- Li H, Leavengood Jr, JM, Chapman EG, Burkhardt D, Song F, Jiang P, Liu J, Zhou X, Cai W. 2017 Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs. *Proc. Biol. Sci.* **284**, 20171223. (doi:10.1098/rspb.2017.1223)
- Liu ZQ *et al.* 2018 Sequencing of complete mitochondrial genomes confirms synonymization of *Hyalomma asiaticum asiaticum* and *kazlovi*, and advances phylogenetic hypotheses for the Ixodidae. *PLoS ONE* **13**, e0197524. (doi:10.1371/journal.pone.0197524)
- Huysse T, Buchmann K, Littlewood DT. 2008 The mitochondrial genome of *Gyrodactylus derjavinoi* (Platyhelminthes: Monogenea)—a mitogenomic approach for *Gyrodactylus* species and strain identification. *Genes* **417**, 27–34. (doi:10.1016/j.gene.2008.03.008)
- Lan T, Gill S, Bellemain E, Bischof R, Nawaz MA, Lindqvist C. 2017 Evolutionary history of enigmatic bears in the Tibetan Plateau-Himalaya region and the identity of the yeti. *Proc. Biol. Sci.* **284**, 20171804. (doi:10.1098/rspb.2017.1804)
- Zou H, Jakovlic I, Chen R, Zhang D, Zhang J, Li WX, Wang GT. 2017 The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class. *BMC Genomics* **18**, 840. (doi:10.1186/s12864-017-4237-x)
- Ballard JWO, Pichaud N, Fox C. 2014 Mitochondrial DNA: more than an evolutionary bystander. *Funct. Ecol.* **28**, 218–231. (doi:10.1111/1365-2435.12177)
- Wolff JN, Ladoukakis ED, Enríquez JA, Dowling DK. 2014 Mitonuclear interactions: evolutionary consequences over multiple biological scales. *Phil. Trans. R. Soc. B* **369**, 20130443. (doi:10.1098/rstb.2013.0443)
- Boore JL, Fuerstenberg SI. 2008 Beyond linear sequence comparisons: the use of genome-level characters for phylogenetic reconstruction. *Phil. Trans. R. Soc. B* **363**, 1445–1451. (doi:10.1098/rstb.2007.2234)
- Botero-Castro F, Tilak MK, Justy F, Catzeflis F, Delsuc F, Douzery EJP. 2018 In cold blood: compositional bias and positive selection drive the high evolutionary rate of vampire bats mitochondrial genomes. *Genome Biol. Evol.* **10**, 2218–2239. (doi:10.1093/gbe/evy120)
- Sluis VD *et al.* 2015 Parallel structural evolution of mitochondrial ribosomes and OXPHOS complexes. *Genome Biol. Evol.* **7**, 1235–1251. (doi:10.1093/gbe/evv061)
- Beall CM. 2007 Two routes to functional adaptation: Tibetan and Andean high-altitude natives. *Proc. Natl. Acad. Sci. USA* **104**, 8655–8660. (doi:10.1073/pnas.0701985104)
- Scott GR, Schulte PM, Egginton S, Scott ALM, Richards JG, Milsom WK. 2011 Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Mol. Biol. Evol.* **28**, 351–363. (doi:10.1093/molbev/msq205)
- Hassanin A, Ropiquet A, Couloux A, Cruaud C. 2009 Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). *J. Mol. Evol.* **68**, 293–310. (doi:10.1007/s00239-009-9208-7)
- Ma X, Kang J, Chen W, Zhou C, He S. 2015 Biogeographic history and high-elevation adaptations inferred from the mitochondrial genome of Glyptosternoid fishes (Sisoridae, Siluriformes) from the southeastern Tibetan Plateau. *BMC Evol. Biol.* **15**, 233. (doi:10.1186/s12862-015-0516-9)
- Peng Y *et al.* 2011 Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Mol. Biol. Evol.* **28**, 1075–1081. (doi:10.1093/molbev/msq290)
- Li Y, Wu D-D, Boyko AR, Wang G-D, Wu S-F, Irwin DM, Zhang Y-P. 2014 Population variation revealed high-altitude adaptation of Tibetan mastiffs. *Mol. Biol. Evol.* **31**, 1200–1205. (doi:10.1093/molbev/msu070)
- Yang L, Wang Y, Zhang Z, He S. 2015 Comprehensive transcriptome analysis reveals accelerated gene evolution in a Tibet fish, *Gymnodiptychus pachycheilus*. *Genome Biol. Evol.* **7**, 251–261. (doi:10.1093/gbe/evu279)
- Romanova EV, Aleoshin VV, Kamal'tynov RM, Mikhailov KV, Logacheva MD, Sirotnina EA,

- Gornov AV, Anikin AS, Sherbakov DY. 2016 Evolution of mitochondrial genomes in Baikalian amphipods. *BMC Genomics* **17**, 1016. (doi:10.1186/s12864-016-3357-z)
37. Kosakovsky Pond SL, Frost SDW. 2005 Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* **22**, 1208–1222. (doi:10.1093/molbev/msi105)
38. 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)
39. Murrell B *et al.* 2015 Gene-wide identification of episodic selection. *Mol. Biol. Evol.* **32**, 1365–1371. (doi:10.1093/molbev/msv035)
40. Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K. 2015 RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* **32**, 820–832. (doi:10.1093/molbev/msu400)
41. Wei S-J, Shi M, Chen X-X, Sharkey MJ, Achterberg CV, Ye G-Y, He J-H. 2010 New views on strand asymmetry in insect mitochondrial genomes. *PLoS ONE* **5**, e12708. (doi:10.1371/journal.pone.0012708)
42. Hassanin A. 2006 Phylogeny of Arthropoda inferred from mitochondrial sequences: strategies for limiting the misleading effects of multiple changes in pattern and rates of substitution. *Mol. Phylogenet. Evol.* **38**, 100–116. (doi:10.1016/j.ympev.2005.09.012)
43. Reyes A, Gissi C, Pesole G, Saccone C. 1998 Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. *Mol. Biol. Evol.* **15**, 957–966. (doi:10.1093/oxfordjournals.molbev.a026011)
44. Hassanin A, Léger N, Deutsch J. 2005 Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 277–298. (doi:10.1080/10635150590947843)
45. Zhang D, Zou H, Hua C-J, Li W-X, Mahboob S, Al-Ghanim KA, Al-Misned F, Jakovlić I, Wang G-T. 2019 Mitochondrial architecture rearrangements produce asymmetrical nonadaptive mutational pressures that subvert the phylogenetic reconstruction in Isopoda. *Genome Biol. Evol.* **11**, 1797–1812. (doi:10.1093/gbe/evz121)
46. Hou Z, Sket B. 2016 A review of Gammaridae (Crustacea: Amphipoda): the family extent, its evolutionary history, and taxonomic redefinition of genera. *Zool. J. Linn. Soc.* **176**, 323–348. (doi:10.1111/zoo.12318)
47. MacDonald III KS, Yampolsky L, Duffy JE. 2005 Molecular and morphological evolution of the amphipod radiation of Lake Baikal. *Mol. Phylogenet. Evol.* **35**, 323–343. (doi:10.1016/j.ympev.2005.01.013)
48. Lee C-W, Nakano T, Tomikawa K, Min G-S. 2018 The complete mitochondrial genome of *Pseudocrangonyx daejeonensis* (Crustacea: Amphipoda: Pseudocrangonyctidae). *Mitochondrial DNA Part B* **3**, 823–824. (doi:10.1080/23802359.2018.1495116)
49. Krebs L, Bastrop R. 2012 The mitogenome of *Gammarus duebeni* (Crustacea Amphipoda): a new gene order and non-neutral sequence evolution of tandem repeats in the control region. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **7**, 201–211. (doi:10.1016/j.cbd.2012.02.004)
50. Sun SE, Hui M, Wang M, Sha Z. 2018 The complete mitochondrial genome of the alvinocaridid shrimp *Shinkaicaris leurokolos* (Decapoda, Caridea): insight into the mitochondrial genetic basis of deep-sea hydrothermal vent adaptation in the shrimp. *Comp. Biochem. Physiol. Part D Genomics Proteomics* **25**, 42–52. (doi:10.1016/j.cbd.2017.11.002)
51. Oliveira DCSG, Raychoudhury R, Lavrov DV, Werren JH. 2008 Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp *Nasonia* (Hymenoptera: Pteromalidae). *Mol. Biol. Evol.* **25**, 2167–2180. (doi:10.1093/molbev/msn159)
52. Xiao J-H, Jia J-G, Murphy RW, Huang D-W. 2011 Rapid evolution of the mitochondrial genome in Chalcidoid wasps (Hymenoptera: Chalcidoidea) driven by parasitic lifestyles. *PLoS ONE* **6**, e26645. (doi:10.1371/journal.pone.0026645)
53. Carlini DB, Fong DW. 2017 The transcriptomes of cave and surface populations of *Gammarus minus* (Crustacea: Amphipoda) provide evidence for positive selection on cave downregulated transcripts. *PLoS ONE* **12**, e0186173. (doi:10.1371/journal.pone.0186173)
54. Pons J, Bauzá-Ribot MM, Jaume D, Juan C. 2014 Next-generation sequencing, phylogenetic signal and comparative mitogenomic analyses in Metacrangonyctidae (Amphipoda: Crustacea). *BMC Genomics* **15**, 566. (doi:10.1186/1471-2164-15-566)
55. Doublet V, Ubbrig E, Alioua A, Bouchon D, Marcade I, Marechal-Drouard L. 2015 Large gene overlaps and tRNA processing in the compact mitochondrial genome of the crustacean *Armadillidium vulgare*. *RNA Biol.* **12**, 1159–1168. (doi:10.1080/15476286.2015.1090078)
56. Zhang D, Li WX, Zou H, Wu SG, Li M, Jakovlić I, Zhang J, Chen R, Wang GT. 2018 Mitochondrial genomes of two diplectanids (Platyhelminthes: Monogenea) expose parphyly of the order Dactylogryidea and extensive tRNA gene rearrangements. *Parasites Vectors* **11**, 601. (doi:10.1186/s13071-018-3144-6)
57. Zhang D, Zou H, Wu SG, Li M, Jakovlić I, Zhang J, Chen R, Li WX, Wang GT. 2018 Three new Diplozoidae mitogenomes expose unusual compositional biases within the Monogenea class: implications for phylogenetic studies. *BMC Evol. Biol.* **18**, 133. (doi:10.1186/s12862-018-1249-3)
58. Gissi C, Iannelli F, Pesole G. 2008 Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**, 301–320. (doi:10.1038/hdy.2008.62)
59. Zhang D, Zou H, Wu SG, Li M, Jakovlić I, Zhang J, Chen R, Wang GT, Li WX. 2017 Sequencing of the complete mitochondrial genome of a fish-parasitic flatworm *Paratetraochoides inermis* (Platyhelminthes: Monogenea): tRNA gene arrangement reshuffling and implications for phylogeny. *Parasites Vectors* **10**, 462. (doi:10.1186/s13071-017-2404-1)
60. Burland TG. 2000 DNASTAR's Lasergene sequence analysis software. In *Methods in molecular biology*TM (eds S Misener, SA Krawetz), pp. 71–91. Totowa, NJ: Humana Press.
61. Hazkani-Covo E, Zeller RM, Martin W. 2010 Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* **6**, e1000834. (doi:10.1371/journal.pgen.1000834)
62. Schattner P, Brooks AN, Lowe TM. 2005 The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**, W686–W689. (doi:10.1093/nar/gki366)
63. Laslett D, Canback B. 2008 ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics* **24**, 172–175. (doi:10.1093/bioinformatics/btm573)
64. Zhang D, Gao F, Li WX, Jakovlić I, Zou H, Zhang J, Wang GT. 2018 PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetic studies. *bioRxiv.* 489088. (doi:10.1101/489088)
65. Lima NCB, Prosdociimi F. 2017 The heavy strand dilemma of vertebrate mitochondria on genome sequencing age: number of encoded genes or G + T content? *Mitochondrial DNA Part A* **29**, 300–302. (doi:10.1080/24701394.2016.1275603)
66. Bauzá-Ribot MM, Juan C, Nardi F, Oromí P, Pons J, Jaume D. 2012 Mitogenomic phylogenetic analysis supports continental-scale vicariance in subterranean Thalassoid crustaceans. *Curr. Biol.* **22**, 2069–2074. (doi:10.1016/j.cub.2012.09.012)
67. Costello MJ *et al.* 2013 Global coordination and standardisation in marine biodiversity through the World Register of Marine Species (WoRMS) and related databases. *PLoS ONE* **8**, e51629. (doi:10.1371/journal.pone.0051629)
68. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)
69. Lanfear R, Calcott B, Ho SYW, Guindon S. 2012 PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol. Biol. Evol.* **29**, 1695–1701. (doi:10.1093/molbev/mss020)
70. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
71. Ronquist F *et al.* 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
72. Lartillot N, Brinkmann H, Philippe H. 2007 Suppression of long-branch attraction artefacts in the animal phylogeny using a site-

- heterogeneous model. *BMC Evol. Biol.* **7**, S4. (doi:10.1186/1471-2148-7-S1-S4)
73. Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Grid Computing Environments Workshop, GCE*. IEEE. (doi:10.1109/GCE.2010.5676129).
74. Letunic I, Bork P. 2007 Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128. (doi:10.1093/bioinformatics/btl529)
75. Librado P, Rozas J. 2009 DnaSP v5. *Bioinformatics* **25**, 1451–1452. (doi:10.1093/bioinformatics/btp187)
76. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018 Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* **35**, 773–777. (doi:10.1093/molbev/msx335)