

Inferring phylogeny of Baikal gammarids using transcriptome sequencing

Nina Popova

Laboratory of Evolutionary Genomics, Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Leninskiye Gory 1-73, Moscow 119992, Russia, nina.tolmacheva@gmail.com

Sergey Naumenko

Laboratory of Evolutionary Genomics, Department of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Leninskiye Gory 1-73, Moscow 119992, Russia; Institute for Information Transmission Problems (Kharkevich Institute), Russian Academy of Sciences, Bolshoi Karetny pereulok 19, Moscow 127994, Russia, sergey.naumenko@yahoo.com

Gammaridae, a family of *Amphipods* (*Crustaceans* subphylum), is a very diverse phylum. Lake Baikal is populated by more than 270 species of *Gammaridae*. The goal of our project is to create a genomic resource for the study of speciation.

To our knowledge, current genomic resources of *Crustaceans* are few. *Daphnia pulex* genome (300 Mbp) and transcriptome [1] could not be used in this research because of large evolutionary distance to *Gammaridae*. Two transcriptome studies of *Parhyale hawaiiensis* [2,3] provide assembled transcripts with 19,067 and 12,271 protein blast hits respectively.

Transcriptomes of 34 samples of Baikal Gammarids collected by L. Yampolsky and N. Mugue were sequenced at the Laboratory of evolutionary genomics by M. Logacheva and A. Penin.

We assembled transcriptomes of these samples using Agalma [4], Trinity[5], Oases[6] software packages and “Makarich” [7] computer cluster. On average we got 8,000 protein blast hits per sample. We extracted protein coding regions from the assembled transcripts and translated nucleotides to amino acid sequences. Then we found 1:1 orthologies with OrthoMCL [8] and build the phylogenetic tree (figure 1) with RaxML [9]. This tree is in agreement with previous published *Gammaridae* phylogenetic studies [10].

This project established a wide collaboration. We thank A. Kondrashov, L. Yampolsky, G.Bazykin, M.Gelfand for invaluable discussions and M. Logacheva, A. Penin and M. Schelkunov for sequencing and data processing. Authors were supported by the

Ministry of Education and Science of the Russian Federation (grants № 11.G34.0008 and 8814) and by the Russian Foundation for Basic Research (grant № 12-07-31261).

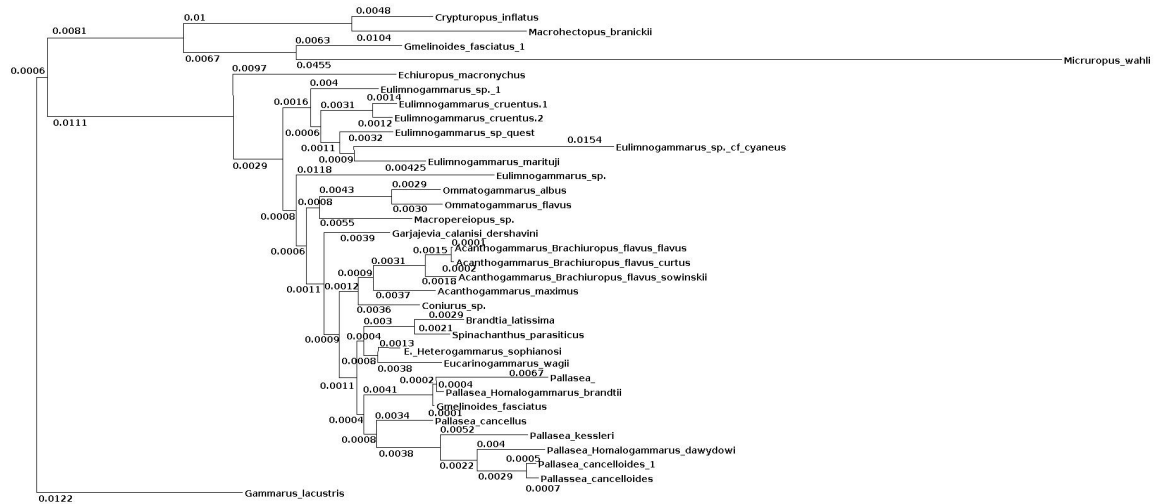


Figure 1. Phylogeny of 34 *Gammaridae*.

1. J.K. Colbourne et al. (2011) *Science*, **331**(6017):555-61.
2. V. Zeng et al. (2011) *BMC Genomics*, **12**:581.
3. M.J. Blythe et al. (2012) *PLoS ONE*, **7**(3): e33784.
4. M. Howison et al. (2012) Proceedings of TaPP '12.
5. M.G. Grabherr et al. (2011) *Nature Biotechnology*, **29**:644-652.
6. M.H. Schulz et al. (2012) *Bioinformatics*, **28**(8), 1086-92.
7. R.Arifulov, S.Naumenko. (2012) Proceedings of ITAS2012. (in Russian).
8. L. Li et al. (2003) *Genome Research* **13**(9): 2178-2189.
9. A. Stamatakis (2006) *Bioinformatics*, **22**(21):2688–2690.
10. K.S. Macdonald et al. (2005) *Molecular Phylogenetics and Evolution*, **35**, 323–343.