

Search for convergently evolving and selected in parallel amino acids in domestic and wild animals

Romashov G.A.^{1,2*}, Yudin N.S.^{1,2}, Yurchenko A.A.¹, Larkin D.M.³

¹ *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

² *Novosibirsk State University, Novosibirsk, Russia*

³ *Royal Veterinary College, University of London, London, UK*

* romashov@bionet.nsc.ru

Key words: cattle, convergent evolution

Motivation and Aim: The study of convergent evolutionary events is a developing area of modern genetics and is of a great interest from both the theoretical and practical perspectives. The fundamental value of such a work is in obtaining a new knowledge about common genetic mechanisms of phenotype formation independently manifested in different species. Breeds of domesticated animals are a great model for studying convergent evolution of adaptation to environments because these populations have migrated with humans to new habitats and express adaptations often absent from the ancestral wild forms.

Methods and Algorithms: In the present work we looked for the nucleotide convergent changes utilizing the whole genome assembly and *de novo* annotated proteins of the cold-adapted Yakut cattle, proteins of 12 hibernating/cold-adapted/deep-diving mammals, and five species not expressing these phenotypes (outgroups). The species and phenotypic groups have been used in the analysis are: deep diving (*Odobenus divergens*, *Physeter catodon*, *Eumetopias jubatus*, *Mirounga leonine*), hibernating (*Myotis brandtii*, *Marmota marmota marmota*, *Microcebus murinus*, *Chrysochloris asiatica*), cold-resistant (*Vulpes lagopus*, *Rangifer tarandus*, *Ovis nivicola*, *Ursus maritimus*, *Odobenus divergens*) and outgroups (*Homo sapiens*, *Loxodonta Africana*, *Canis familiaris*, *Pteropus alecto*, *Bos taurus* (Hereford)). Coding sequences (CDSs) for all species have been downloaded from NCBI RefSeq or predicted with GeMoMa (v. 1.7). For each protein the longest isoform has been chosen. The protein sequences have been obtained with PRANK (<http://wasabiapp.org/software/prank/>). Orthogroups were identified with OrthoFinder (v. 2.5.2) and filtered to ensure each species had maximum one sequence in each orthogroup and each orthogroup contained ≥ 3 orthologous sequences for phenotypic group, all gaps in alignment of proteins sequences orthogroups were removed. Nucleotide sequences within each orthogroup were aligned with PRANK (v. 170703). PCOC method looking both at exact same amino acids changes and shifts in amino acid type preferences has been applied. For detection of the putative convergent changes between phenotypic groups the detection threshold has been set to 0.8. To identify convergent changes between the Yakut cattle and the ingroup species, all amino acids shared between the Yakut cattle and Hereford reference genome were filtered out.

Results: Eighteen putative convergent events were found in 18 genes (16 shared with the deep-diving and 2 with hibernating animals). This analysis confirmed the previously reported NRAP convergent mutation and identified putative events in 17 additional genes. The F_{ST} analysis for all 18 positions between the Yakut cattle and the dataset of

1000 Bull Genome Project suggests that all but one (*NRAP*) alleles were selected in parallel rather than occurred convergently in the Yakut cattle and the wild species because the Yakut cattle alleles were also present in other cattle breeds at a lower frequency and in *Bovinae* species. The functional analysis has revealed two genes previously found under selection in the Siberian human populations (*PLA2G4F* ($F_{ST}=0.98$) and *PAFAH1B2* ($F_{ST}=0.01$)), a gene involved in resistance to environmental toxicants (*FBNP4* ($F_{ST}=0.63$)), genes involved in controlling the heart function (*NRAP* ($F_{ST}=1.0$) and *POPDC2* ($F_{ST}=0.3$)), in regulation of blood pressure and insulin secretion (*SNX19* ($F_{ST}=0.53$)), and shortening of long bones (*WDR60* ($F_{ST}=0.95$)).

Conclusion: Our study demonstrates that while *de novo* convergent evolution events in breeds of domestic and wild species is a rare phenomenon (limited to *NRAP* in our study), in line with rare convergent changes between different species, a parallel selection for alternative ancestral alleles could be more widespread (17 genes). On the other hand, more ingroup and outgroup genomes should be used for a more powerful analysis. Both the convergently evolving and selected in parallel alleles should be considered when studying the biology of wild animals and livestock breeds and designing new generations of locally adapted livestock.

Acknowledgements: The study was supported by the Institute of Cytology and Genetics, SB RAS Budget Project No. FWNR-2022-0023.