Tomarovsky Andrey

Date of birth	22.02.1998
Phone	+7 951 145 65 65
E-mail	andrey.tomarovsky@gmail.com
Github	github.com/tomarovsky
Telegram	@atomarovsky
-	

Work experience

2022 - present — **Senior Research Assistant at the Institute of Molecular and Cellular Biology, Novosibirsk.** Research on genomics of the genus *Martes*:
- *De novo* chromosome-level assembly and following correction of mammalian genomes. Improvement of scaffold-level genome assemblies to chromosome-level using HiC-sequencing data.
- Estimation of genome size and coverage based on k-mers distribution.
- *De novo* assembly and annotation of genome repeats.
- Whole-genome phylogenetic and evolutionary analysis.
- Population structure analysis and determination of admixture levels of closely related mammalian species.
- Reconstruction of mammalian population history using sequential Markovian coalescent methods.
- Construction of haplotype networks using nuclear and mitochondrial markers.
- Creation of high-performance software pipelines based on Snakemake.

2021 - 2022 — **Research programmer at the Genomic Diversity Research Center, ITMO University.** Research on genomics of the genus *Martes*:
- Quality control and filtering of high-throughput sequencing data.
- *De novo* assembly and quality control.
- Whole genome alignments (pairwise and multiple) and coverage statistics.
- Calculating the coordinates of the pseudoautosomal region.
- Detection and annotation of repeats in genome assemblies.
- Protein-coding and non-coding genes annotation.
- Calling and filtering of genetic variants with determination of heterozygosity level.
- Work experience on a high-performance computing cluster.

2021 - present — **Technical and software support for Blastim courses.**
- Python and Linux for Bioinformatics and Biology.
- Statistics, R and data analysis.
- Analysis of NGS data.

2020 - present — **ResOps and system administration experience on computing cluster MSU FBB.**
- Assistance in writing commands and pipelines for bioinformatic analyses.
- Software support for cluster users.
- Creating a cluster documentation.

Education _____

2021 - present   **PhD student in Genetics.** Novosibirsk State University, Novosibirsk, Russia. PhD thesis: Obtaining genome assemblies and phylogenetic analysis of representatives of the genus *Martes* (fam. Mustelidae).

2019 - 2021   **MS in Bioinformatics.** Saint-Petersburg State University, St. Petersburg, Russia. MS thesis: Assembly and annotation of sable (*Martes zibellina*) and pine marten (*Martes martes*) genomes.

2015 - 2019   **BS in Biotechnologies.** Belgorod State National Research University, Belgorod, Russia.

Additional education _____

Coursera:
- Getting Started with Python (University of Michigan).
- Python Data Structures (University of Michigan).
- Capstone: Retrieving, Processing, and Visualizing Data with Python (University of Michigan).

Stepik:
- Молекулярная биология и генетика (ИБ).
- Генетика и геномика популяций (ИБ).
- Молекулярная филогенетика (ИБ).
- Введение в молекулярную биологию и биомедицину (МФТИ).
- Биотехнологии: генная инженерия (ИБ).
- Секвенирование 3-го поколения на Oxford Nanopore (ИБ).
- Введение в Linux (ИБ).
- Знакомство с R и базовая статистика (СПбГУ).
- Программирование на Python (ИБ).
- Алгоритмы: теория и практика. Структуры данных (Computer Science Center).
- Python Skills for Handling Biological Data.

Основы Git.

Skills _____

OS: Linux, Windows.
Shell: Bash. A good knowledge of the various shell tools, such as Awk, Grep, Sed.
Programming: Python.
Python libraries: Biopython, Matplotlib, Numpy, Pandas, Scikit-learn (basic level). Experience in writing various scripts and data visualizations in Jupyter Notebook and individual Python packages.
- Parsing data from files or websites to Pandas dataframes.
- Calculation of average, median, minimum and maximum values in datasets.
- Visualization of results using Matplotlib in the form of plots, histograms, Venn diagrams.
- A little experience in ML (kNN, clustering, linear regression)

Statistics: R.
R libraries: ggplot2, dplyr, readxl. Experience analyzing various datasets, such as those containing information on different types of cancer and patient survival times.
- Linear and multiple regression.
- Description and significance testing of linear models.
- Comparison of linear models.
- Testing statistical hypotheses.

Workflow managers: Snakemake. Experience in writing complex Snakemake pipelines including benchmarking, logging, task grouping and running on a compute cluster. There is experience in collaborative development.

Workload managers: Slurm, PBS. ResOps experience on computing clusters MSU FBB, ICG, IMCB and ITMO. Running large-scale computational tasks using Slurm, PBS and Snakemake. Installation and interaction with Conda environments.

Others:
- SQL (creating a database, simple and medium complexity queries)
- Circos (basic level, experience in visualization of mDNA and its coverage)
- Tcl (basic level, experience in writing module files)

On the Side _____

Snakemake pipelines:
- **BuscoClade**. Pipeline to construct species phylogenies using universal single-copy orthologs BUSCOs. Phylogenetic tree reconstruction using IQtree and Astral III. Visualization using ete3.
- **ITSpipe**. Pipeline for the analysis of ITS sequences from the ribosomal cluster. Coverage visualization using Matplotlib and variant calling using Gatk, Pisces, and Bcftools is performed.
- **varcaller**. Pipeline for calling genetic variants correctly. Includes visualization of coverage and calculation of PAR coordinates.

Others:
- **Biocrutch**. A custom python package for bioinformatics research. My project contains bioinformatics scripts for genome and coverage statistics, repeat masking, determining coordinates of pseudoautosomal region, filtering 10XGenomics linked reads, PSMC data formatting and others.
- **Bashare**. The repository contains custom Bash scripts and pipelines for data processing.

Grants _____

Russian Foundation for Basic Research, grant № 20-04-00808 A, "Genomes and genetic diversity of mustelids (fam. Mustelidae) of Russia and South-Eastern Asia".

Articles _____

Google scholar:

1. (submitted) Kliver, S.; Houck, M.L.; Perelman, P.L.; Totikov, A.A.; Tomarovsky, A.A.; DNA Zoo Consortium; Chan, S.; Hastie, A.; Komissarov, A.; Ryder, O.A.; Graphodatsky, A.S.; Johnson, W.E.; Maldonado, J.E.; Pukazhenthi, B.S.; Marinari, P.E.; Wildt, D.E.; Koepfli, K.-P. Chromosome-length genome assembly and karyotype of the endangered black-footed ferret (Mustela nigripes). Journal of Heredity.
2. Тотиков А.А., Томаровский А.А., Якупова А.Р., Графодатский А.С., Кливер С.Ф. Обзор методов реконструкции демографической истории популяций в природоохранной биологии // Экологическая генетика. 2023. Т. 21. № 1. С. 85–102. DOI: 10.17816/ecogen120078
3. Yakupova, A.; Tomarovsky, A.; Totikov, A.; Beklemisheva, V.; Logacheva, M.; Perelman, P.; Komissarov, A.; Dobrynin, P.; Krasheninnikova, K.; Tamazian, G.; Serdyukova, N.; Rayko, M.; Bulyonkova, T.; Grachev, M.; Cherkasov, N.; Pylev, V.; Varnavsky, A.; Peterfeld, V.; Penin, A.; Balanovska, E.; Lapidus, A.; O'Brien, S.; Graphodatsky, A.; Kepfli, K.-P.; Kliver, S. Chromosome length genome assembly of the Baikal seal (Pusa sibirica) reveals fewer answers than new mysteries. Genes. 2023; 14(3):619. DOI: 10.3390/genes14030619
4. Derežanin, L.; Blažytė, A.; Dobrynin, P.; Duchêne, D.A.; Grau, J.H.; Hofreiter, M.; Jeon, S.; Kliver, S.; Koepfli, K.P.; Meneghini, D.; Preick, M.; Tomarovsky, A.; Totikov, A.; Fickel, J.; Förster, D.W. Multiple types of genomic variation contribute to adaptive traits in the mustelid subfamily Guloninae. Molecular Ecology 2022. DOI: 10.1111/mec.16443

5. Totikov, A.; Tomarovsky, A.; Prokopov, D.; Yakupova, A.; Bulyonkova, T.; Derezanin, L.; Rasskazov, D.; Wolfsberger, W.; Koepfli, K.P.; Oleksyk, T.K. and Kliver, S. Chromosome-Level Genome Assemblies Expand Capabilities for Conservation Biology. Genes 2021, 12, x. DOI: 10.3390/genes12091336

Conferences

1. Totikov, A.; **Tomarovsky, A.**; Perelman, P.; Serdyokova, N.; Beklemisheva, V.; Bulyonkova, T.; Zub, K.; Panov, V.; Mukhacheva, A.; Abramov, A.; Koepfli, K.; Graphodatsky, A.; Melo-Ferreira, J.; Kliver, S. Reconstruction of the demographic history for three populations of the least weasel Mustela nivalis. Moscow Conference on Computational Molecular Biology (MCCMB-2021). ISBN: 978-5-901158-32-6
2. **Tomarovsky, A.**; Totikov, A.; Beklemisheva, V.; Perelman, P.; Serdyokova, N.; Bulyonkova, T.; Koniaeva, K.; Abramov, A.; Graphodatsky, A.; Koepfli, K.; Powell, R.; Kliver, S. Assembly and annotation of the sable (*Martes zibellina*) and pine marten (*Martes martes*) genomes. Moscow Conference on Computational Molecular Biology (MCCMB-2021). ISBN: 978-5-901158-32-6
3. Тотиков А.А., **Томаровский А.А.**, Анализ кластеров генов биосинтеза метаболитов в бактериальных геномах. 24-я Международная Пущинская конференция молодых ученых. 2020, Пущино. Сборник тезисов, 2020. – 446 с. ISBN 978-5-91874-901-2
4. **Томаровский А.А.**, Лихидченко В.И., Поливанова Е.Ю., Серикова Н.В., Анализ и сравнение результатов после применения фермента Alphalase AFP на клетки *Corynebacterium glutamicum*. CXIX Международная научнопрактическая конференция «Молодой исследователь: вызовы и перспективы», 05/2019