Andrey Tomarovsky

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

Education

September, 2021 - present	**PhD student in Genetics.** Novosibirsk State University, Novosibirsk, Russia. PhD thesis: Obtaining genomic assemblies and phylogenetic analysis of members of the genus *Martes* (fam. Mustelidae).
September, 2019 - July 2021	**MS in Bioinformatics.** Saint-Petersburg State University, St. Petersburg, Russia. MS thesis: Assembly and annotation of the sable (*Martes zibellina*) and pine marten (*Martes martes*) genomes.
September, 2015 - July 2019	**BS in Biotechnologies.** Belgorod State National Research University, Belgorod, Russia.

Additional education

Online courses:

1. Молекулярная биология и генетика (ИБ).
2. Генетика и геномика популяций (ИБ).
3. Молекулярная филогенетика (ИБ).
4. Введение в молекулярную биологию и биомедицину (МФТИ).
5. Биотехнологии: генная инженерия (ИБ).
6. Секвенирование 3-го поколения на Oxford Nanopore (ИБ).
7. Введение в Linux (ИБ).
8. Знакомство с R и базовая статистика (СПбГУ).
9. Программирование на Python (ИБ).
10. Алгоритмы: теория и практика. Структуры данных (Computer Science Center).
11. Getting Started with Python (University of Michigan).
12. Python Data Structures (University of Michigan).
13. Capstone: Retrieving, Processing, and Visualizing Data with Python (University of Michigan).
14. Python Skills for Handling Biological Data.
15. Основы Git.

Work experience

1) <u>March, 2021 - present</u> **Research programmer at the Genomic Diversity Research Center, ITMO University**. Conducts research on genomics of the genus Martes:
   - Quality control and filtering of sequencing data.
   - *De novo* assembly and quality control.
   - Genome annotation:
     - Whole genome alignments (pairwise and multiple) and coverage statistics.
     - Calculating the coordinates of the pseudoautosomal region (custom algorithm was developed).
     - *De novo* assembly and annotation of genome repeats.
     - Protein-coding and non-coding genes annotation.

- Calling and filtering of genetic variants.
- Visualization and analysis of heterozygosity.
- Phylogenetic and evolutionary analysis.
- Comparison and analysis of the results obtained.
- Pipeline development based on Bash and Snakemake.

2) <u>July, 2021 - present</u> **Technical and software support for Blastim courses**:
   a) Python and Linux for Bioinformatics and Biology (July 2021).
   b) Analysis of NGS data (September 2021, February 2022).

3) <u>December, 2020 - present</u> **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](#).

Skills

English: Pre-Intermediate.
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: readxl, dplyr, car, cowplot, ggplot2. 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:

1) **BuscoClade** ([link to github](#))
   Pipeline to construct species phylogenies using universal single-copy orthologs BUSCOs.

2) **ITSpipe** ([link to github](#))
   Pipeline for the analysis of ITS sequences from the ribosomal cluster. Coverage visualization using Matplotlib and variant calling using Gatk, Pisces, and Bcftool is performed.

3) **varcaller** ([link to github](#))
   Pipeline for calling genetic variants correctly. Includes visualization of coverage and calculation of PAR coordinates.

Others:

1) **Biocrutch** ([link to github](#))
   A custom python package for bioinformatics research. My project contains bioinformatics scripts for genome and coverage statistics, repeats masking, determining coordinates of pseudoautosomal region, filtering 10XGenomics linked reads, PSMC date combine and others.

2) **Bashare** ([link to github](#))
   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".

Conferences

1. Tomarovsky, A.; Azamat A. Totikov; Violetta R. Beklemisheva; Polina L. Perelman; Natalia A. Serdyokova; Tatiana Bulyonkova; Ksenia A. Koniaeva; Alexei V. Abramov; Alexander S. Graphodatsky; Klaus-Peter Koepfli; Roger A. Powell; Sergei F. Kliver, **Assembly and annotation of sable (Martes zibellina) and pine marten (Martes martes) genomes**, Moscow Conference on Computational Molecular Biology (MCCMB), July 30th — August 2nd, 2021

2. Azamat A. Totikov; Andrey A. Tomarovsky; Polina Perelman; Natalia Serdyokova; Violetta R. Beklemisheva; Tatiana M. Bulyonkova; Inês Miranda; Liliana Farelo; Karol Zub; Viktor V. Panov; Anna S. Mukhacheva; Alexei V. Abramov; Klaus-Peter Koepfli; Alexander S. Graphodatsky; Jose Melo-Ferreira; Sergei F. Kliver, **Reconstruction of the demographic history for three populations of the least weasel Mustela nivalis**, Moscow Conference on Computational Molecular Biology (MCCMB), July 30th — August 2nd, 2021

3. Totikov, A., Tomarovsky, A., Derezanin, L., Dudchenko, O., Lieberman-Aiden, E., Koepfli, K., & Kliver, S. (2020, November). **Chromosome-Level Genome Assemblies Expanded Capabilities of Conservation Biology**. In Presented at the 1st International Electronic Conference on Genes: Theoretical and Applied Genomics (Vol. 2, p. 30)

4. 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

5. 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

Articles

Google scholar page:

1. 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. ISBN: 978-5-901158-32-6.

2. 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*. ISBN: 978-5-901158-32-6.

3. 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*. ISBN: 978-5-901158-32-6.

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.