

Roman Sutormin

rsutormin@gmail.com

www.bioinf.fbb.msu.ru/~roman/

Professional experience

05/2013 – present: **Lawrence Berkeley National Laboratory (Berkeley, CA, USA)**

Software developer

- Development of JSON based storage system used for keeping biological data involved in public KBase services.
- Development of KBase service for phylogenetic trees processing.
- Development of KBase service for bacterial comparative genomics.
- Development of KBase service for protein domains search.
- Generator of client/server stubs in Java based on KIDL language describing API of KBase services.
- Set of narrative widgets (visual components) allowing to start biological analysis based on public KBase services and to show results.

09/2011 – 05/2013: **Lomonosov Moscow State University, Dept. of Bioengineering and Bioinformatics (Moscow, Russia), Research Scientist**

- Enhanced the CAMPS web resource that enables multiple classification of transmembrane proteins. Optimized the speed of the HMM-based algorithm for classification of transmembrane proteins. Developed web based UI for visualization of a graph of transmembrane protein families. *This project was supervised by Prof. Dmitriy Frishman (Technical University of Munich).*
- Participated in design of HMM model for identification of transcriptional regulatory regions in eukaryotes based on statistical rules (grammar) defining the architecture of the transcription factor binding sites (TFBS). Participated in the implementation of this model in the CORECLUST algorithm. See publications list.
- Currently developing the scalable pipeline for gene annotation of eukaryotic genomes to enable high performance distributed computing on 30 nodes cluster (24 cores on each node). The cluster is operated by Torque Resource Manager. The pipeline is currently using for gene annotation in the *Fagopyrum Tataricum* and *Genlisea Margaretae*. The genome *Genlisea Margaretae* annotated by the developed pipeline is submitted to NCBI GenBank. *This project is supervised by Prof. Alex Kondrashov (University of Michigan).*
- Lecturer in Java Programming and Algorithms in Bioinformatics.

10/2009 – 08/2011: **Mail.Ru Group (Moscow, Russia), Software developer**

- Support and development of data warehouse (PostgreSQL) and administration web tool (JBoss + GWT) on server-side of MMO game “Allods Online”.

06/2008 – 09/2009: **Irbicon (Moscow/Dubna, Russia), Software developer**

- Developed the plugin providing the credit scoring functionality for SAS Data Integration Studio tool.
- Developed web portlet for SAS Information Delivery Portal providing web interface for full-text search in Subversion repositories.
- Developed web resource (GWT) for management of collection agency tasks related to a life cycle of payment debts based on SAS infrastructure.

09/2005 – 05/2008: **Lomonosov Moscow State University, Dept. of Bioengineering and Bioinformatics (Moscow, Russia), Research Scientist**

- Developed a web based UI module for identification and visualization of clusters of paralogous proteins in a bacterial genome. This module enables visualization of graph of paralogs based on protein domain similarity.
- Developed an algorithm for identification of linked (phylogenetically and by gene location) orthologous protein clusters in a group of bacterial genomes and the web module for visualization of the results.
- Developed an algorithm for the identification of syntenic regions between two bacterial genomes based on protein homology information and the web module for visualization. *These modules were developed as a result of partnership between Moscow State University and Biomax AG, and supervised by Prof. Dmitriy Frishman*
- Developed the internal web resource for student term papers management based on JSP (department order).
- Constructed the genomic data warehouse and ETL-processes for data integration from different sources like Genbank, Uniprot, PDB and others.
- Lecturer in Java Programming and Algorithms in Bioinformatics.

03/2003 – 03/2005: **Diasoft (Moscow, Russia), Software developer**

- Developed the visual tool for modeling metadata of data warehouse. Designed the unified script language to work with various dialects specific for different DBMSs such as: MSSQL, Sybase, DB2, Oracle.
- Developed the SQL-query parser and analysis tool for collecting column dependencies and SQL-query refactoring.

Bioinformatics expertise

Dynamic programming on matrices and graphs, phylogenetic tree construction, functional optimization methods (gradients, genetic algorithm, annealing, Gibbs sampling), HMM construction, pairwise and multiple alignment algorithms development.

Software development expertise

Languages: Java, C, C++, C#, Object Pascal, Perl, Groovy, Python

Database Management Systems: MSSQL, Sybase, DB2, Oracle, MySQL, PostgreSQL, Access, JavaDB, HSQL, MongoDB

Database technologies/dialects: JDBC, Hibernate, ANSI SQL, PL/SQL, PL/pgSQL, TransactSQL

Web development: JSP, Servlets, JavaScript, GWT, ExtJS, JQuery, Dojo, FreeMarker, D3, Angular, RequireJS

Application Servers: Tomcat, JBoss, Jetty, Glassfish

Web services: SOAP, Jersey

Client-server development: sockets (client-server TCP/IP, datagrams), RMI

Other Java technologies: JavaCC (parser's compiler), multi-threading, reflection, Swing, XML (DOM, SAX, XSLT), serialization, java code profiling, byte code decompilation, Android SDK

Cluster computing: Torque Resource Manager

IDEs: Eclipse, IntelliJ IDEA, Microsoft Visual Studio, Oracle JDeveloper

Version Control Systems: CVS, Subversion, GIT

Testing Frameworks: Jenkins

Education

1997-2002: **Department of Mathematics and Mechanics, Moscow State University, Moscow, Russia. BS Mathematics and MS Computational Mathematics.**

Description of Thesis: Algorithm for simultaneous secondary structure prediction and alignment of membrane protein sequences was built based on Hidden Markov Model (HMM) approach. Resulting HMM was multiplication of two HMMs: one for pairwise alignment of two protein sequences and second for prediction of membrane regions on protein sequence. Each of these two HMMs change their parameters depending on state of another HMM.

2002-2005: **State Scientific Center GosNIIGenetika, Moscow, Russia. PhD Biophysics.**

Description of Thesis: Work contains two parts. First, series of amino acid substitution matrices were constructed for membrane regions of protein sequences using the approach similar to that applied for BLOSUM matrix construction. Second, a method was developed for better prediction of membrane regions for multiple alignment of membrane protein sequences based on Hidden Markov Model approach using substitution matrices constructed in first part of work.

2008: **Sun Certified Programmer for the Java 2 Platform, Standard Edition 5.0 certification.**

Publications

- Denisov SV, Bazykin GA, Sutormin R, Favorov AV, Mironov AA, Gelfand MS, Kondrashov AS. Weak negative and positive selection and the drift load at splice sites. *Genome Biol Evol.* 2014 May 14;6(6):1437-47
- Novichkov PS, Kazakov AE, Ravcheev DA, Leyn SA, Kovaleva GY, Sutormin RA, Kazanov MD, Riehl W, Arkin AP, Dubchak I, Rodionov DA. RegPrecise 3.0--a resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics.* 2013 Nov 1;14:745
- Leushkin EV, Sutormin RA, Nabieva ER, Penin AA, Kondrashov AS, Logacheva MD. The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics.* 2013 Jul 15;14(1):476.
- Kurmangaliyev YZ, Sutormin RA, Naumenko SA, Bazykin GA, Gelfand MS. Functional implications of splicing polymorphisms in the human genome. *Hum Mol Genet.* 2013 Sep 1;22(17):3449-59. doi: 10.1093/hmg/ddt200. Epub 2013 May 2.
- Nikulova AA, Favorov AV, Sutormin RA, Makeev VJ, Mironov AA. CORECLUST: identification of the conserved CRM grammar together with prediction of gene regulation. *Nucleic Acids Res.* 2012 Jul; 40(12):e93. Epub 2012 Mar 15.
- Sadovskaya NS, Sutormin RA, Gelfand MS. Recognition of transmembrane segments in proteins: review and consistency-based benchmarking of internet servers. *J Bioinform. Comput. Biol.* 2006 Oct; 4(5):1033-1056.
- Sutormin RA, Mironov AA. Membrane probability profile construction based on amino acids sequences multiple alignment. *Mol. Biol. (Mosk).* 2006 May-Jun;40(3):541-545.
- Sutormin RA, Rakhmaninova AB, Gelfand MS. BATMAS30: amino acid substitution matrix for alignment of bacterial transporters. *Proteins.* 2003 Apr 1;51(1):85-95.
- Kalinina OV, Makeev VJ, Sutormin RA, Gelfand MS, Rakhmaninova AB. The channel in transporters is formed by residues that are rare in transmembrane helices. *In Silico Biol.* 2003;3(1-2):197-204.