

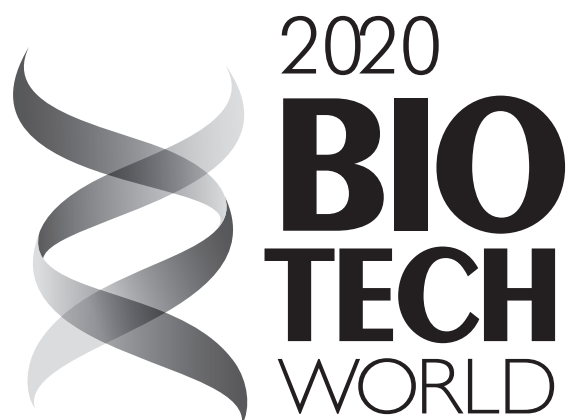
МАТЕРИАЛЫ ФОРУМА
FORUM PROCEEDINGS

БИОТЕХНОЛОГИЯ:
СОСТОЯНИЕ
И ПЕРСПЕКТИВЫ
РАЗВИТИЯ
МЕЖДУНАРОДНЫЙ ФОРУМ

BIOTECHNOLOGY:
STATE OF THE ART
& PERSPECTIVES
INTERNATIONAL FORUM

2022

MOSCOW
OCTOBER
28-30
ОКТАБРЯ
МОСКВА



WWW.BIOMOS.RU

МЕЖДУНАРОДНЫЙ ФОРУМ

INTERNATIONAL FORUM

**БИОТЕХНОЛОГИЯ:
СОСТОЯНИЕ
И ПЕРСПЕКТИВЫ
РАЗВИТИЯ**

**BIOTECHNOLOGY:
STATE OF THE ART
AND PERSPECTIVES**

ВЫПУСК 18

ISSUE 18

28-30 ОКТЯБРЯ 2020
МОСКВА

28-30 OCTOBER, 2020
MOSCOW

**МЕЖДУНАРОДНЫЙ ФОРУМ
«БИОТЕХНОЛОГИЯ: СОСТОЯНИЕ И
ПЕРСПЕКТИВЫ РАЗВИТИЯ»**

Материалы международного форума
«Биотехнология: состояние и перспективы
развития»
28-30 ОКТЯБРЯ 2020 г.

Настоящие материалы форума созданы на
основании информации, предоставленной
участниками форума и одобренные
руководителями секций.

Материалы тезисов публикуются в авторской
версии. Организаторы не несут ответственности
за неточности и опущения в названиях и адресах,
представленных в данном сборнике.
Любое копирование и использование
материалов без письменного разрешения
Программного комитета не разрешено.

УДК 575.1/2::612.017.1 ББК 28.072
ISBN 978-5-6045396-0-6
ISSN: 2312-640X
DOI 10.37747/2312-640X-2020-18

© ООО «ЭКСПО-БИОХИМ-ТЕХНОЛОГИИ»
129090, Москва г, Олимпийский пр-кт,
дом № 10, к.2, кв.154
info@biomos.ru, www.biomos.ru

Все права на издание принадлежат
ООО «ЭКСПО-БИОХИМ-ТЕХНОЛОГИИ» -
организатор международного форума
«Биотехнология: состояние и перспективы
развития»

**INTERNATIONAL FORUM «BIOTECHNOLOGY:
STATE OF THE ART AND PERSPECTIVES»**

The proceedings of International forum
«Biotechnology: state of the art and perspectives»
28 - 30 OCTOBER, 2020.

DISCLAIMER

This book contains abstracts and complete papers
approved by the Forum Review Committee. Authors
are responsible for the content and accuracy.

Opinions expressed may not necessarily reflect the
position of the Scientific Council of forum.
Information in the Biotechnology: state of the art and
perspectives» 2020 Forum Proceedings is subject
to change without notice. No part of this book may
be reproduced or transmitted in any form or by any
means, electronic or mechanical, for any purpose,
without the express written permission of the
International Scientific Council of forum.

ISBN 978-5-6045396-0-6
ISSN: 2312-640X
DOI 10.37747/2312-640X-2020-18

Copyright © LTD «EXPO-BIOHIM-TECHNOLOGIES»
Pr-kt Olimpiiskii, d.10, c. 2, kv Moscow,
info@biomos.ru, www.biomos.ru

All Rights Reserved by LTD «EXPO-BIOHIM-
TECHNOLOGIES» - organizer of the International forum
«Biotechnology: state of the art and perspectives».

UDC 615.015.11:004.852: 616.379-008.64 DOI 10.37747/2312-640X-2020-18-257-258

ANTIGLYCATION ACTIVITY MODELING BY THE METHODS OF MACHINE LEARNING

Litvinov R.A., Vasiliev P.M.

Volgograd State Medical University, Volgograd, Russian Federation
 400131, 1, Pavshikh Bortsov Sq., Volgograd,
 e-mail: litvinov.volggu@mail.ru

A multilevel consensus model of the antiglycation activity has been built by the machine learning methods using IT Microcosm.

Key words: antiglycation activity; in silico; IT Microcosm; prediction strategies; multilevel consensus model.

On the basis of ChEMBL open source [1], the structure-activity database of the reference antiglycation compounds had been created. Structural formulas, activity and research method have been verified by qualified experts, chemists and pharmacologists. Data on antiglycation activity had been clustered into 4 classes depends on the activity level: "high", "moderate", "low", "inactive". By the usage of IT Microcosm [2] a multilevel model of the chemical compounds' antiglycation activity had been built (based on the consensus of prediction strategies). The leave-one-out cross validation method was used for accuracy verification of the final model. The results of the accuracy assessment: conservative strategy gives 77%, 84%, 83%; normal strategy gives 75%, 88%, 89%; risk strategy gives 84%, 87%, 87% of accuracy – for the activity levels "high", "high/moderate", "high/moderate/low", respectively. High activity level has been better predicted by the risk strategy, manifested activity (if any activity level is presented) has been better predicted by the normal strategy. The results obtained, indicate that the data on the antiglycation activity of the reference compounds have been adequately processed and are validly correlated with their chemical structure. The multilevel consensus model of the antiglycation activity of the chemical compounds obtained by the usage of IT Microcosm will be used for the target-oriented searching for new antiglycation compounds, which was started early [3, 4].

Investigation was carried out under financial support of the grant of the President of the Russian Federation for young scientists – candidates of sciences, MK-1887.2020.7.

References

1. ChEMBL: a manually curated database of bioactive molecules with drug-like properties. URL: <https://www.ebi.ac.uk/chembl/> (date of the application: 19.05.2020).
2. Vassiliev P.M., Spasov A.A., Kosolapov V.A., Kucheryavenko A.F., Gurova N.A., Anisimova V.A. Consensus Drug Design Using IT Microcosm. In Application of Computational Techniques in Pharmacy and Medicine / Eds. L. Gorb, V. Kuz'min, E. Muratov – Vol. 17. – Dordrecht (Netherlands): Springer Science + Business Media, 2014. – 550 p. – P. 369-431.
3. Savateev K., Fedotov V., Butorin I., Eltsov O., Slepukhin P., Ulomsky E., Rusinov V., Litvinov R., Babkov D., Khokhlacheva E., Radaev P., Vassiliev P., Spasov A. Nitrothiadiazolo[3,2-a]pyrimidines as promising antiglycating agent // Eur J Med Chem. 202. Vol 185. Art. No. 111808.
4. Spasov A.A., Brel A.K., Litvinov R.A., Lisina S.V., Kucheryavenko A.F., Budaeva Yu.N., Salaznikova O.A., Rashchenko A.I., Shamshina D.D., Batrakov V.V., Ivanov A.V. Evaluation of N-hydroxy-, N-methoxy-, and N-acetoxybenzoyl-substituted derivatives of thymine and uracil as new substances for prevention and treatment of long-term complications of diabetes mellitus // Russian Journal of Bioorganic Chemistry. 2018. Vol. 44. № 6. P. 769-777.

УДК 577.29 DOI 10.37747/2312-640X-2020-18-258-260

ПРЕДСКАЗАНИЕ ПОТЕРЬ ГЕНОВ НА ОСНОВЕ ГЕНОМНЫХ СТРУКТУР

Рубанов Л.И., Шиловский Г.А., Селиверстов А.В., Зверков О.А., Любецкий В.А.

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук
 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.
 e-mail: lyubetsk@iitp.ru

Разработан эффективный алгоритм, реализованный в программе для многопроцессорных систем, позволяющий обнаруживать гены, потерянные или приобретённые в ходе эволюции большинством видов из

заданного множества. Предложенный подход учитывает взаимное расположение генов на хромосоме и позволяет одновременно рассматривать сотни видов.

Ключевые слова: ортологичные гены; синтения; потеря генов; эффективная компьютерная программа; оценка времени работы программы; суперкомпьютер.

Часто используемые методы выявления ортологов и паралогов включают филогенетический анализ, основанный на сравнении гомологичных последовательностей и согласовании деревьев, анализ локального расположения генов (синтении) и реконструкцию эволюционных сценариев на основе геномных структур (глобального расположения генов на хромосомах). Известно, что анализ последовательностей и синтении может привести к несовместимым результатам. Разработано большое число алгоритмов и программ, посвящённых анализу ортологичности и сопутствующим задачам. Существенной особенностью нашего подхода к поиску потерянных и приобретённых генов является совместное сравнение больших наборов видов. Программа *lossgainRSL* разработана для поиска генов, потерянных, приобретённых или сохранившихся в большинстве видов из заданных наборов. Алгоритм реализован в программе, позволяющей производить параллельные вычисления в среде MPI под управлением операционных систем Windows и Linux. Программа свободно доступна по адресу <http://lab6.iitp.ru/ru/lossgainrsl/>. Программа *lossgainRSL* имеет следующую схему. Для выбранного вида рассматриваются все его гены (с окрестностью фиксированного размера), присутствующие или отсутствующие в нескольких заданных на входе наборах видов. Условие отбора генов задаётся пользователем. Программа допускает работу с неполными геномами. Упомянутая окрестность гена (определяемая параметром программы) используется для проверки синтении в её пределах. Важно, что размер окрестности можно обоснованно выбирать. А именно, трёхмерная структура ДНК включает топологически ассоциированные домены, размеры которых определяют эту окрестность и представляются биологически обоснованными [1].

Время поиска ортолога заданного гена в заданной окрестности сравнимо со временем чтения таблиц ортологов и паралогов и ограничено линейной функцией от длины входа программы. Поэтому полное время работы программы ограничено полиномом второй степени от длины входа даже на однопроцессорной системе, что позволяет на многопроцессорной системе рассматривать одновременно сотни видов.

С помощью программы мы определили гены мыши, потерянные или ставшие псевдогенами у человека, но сохранившиеся (и кодирующие белки) у не менее чем четырёх из пяти человекообразных обезьян: *Gorilla gorilla*, *Nomascus leucogenys*, *Pan paniscus*, *Pan troglodytes* и *Pongo abelii*. Среди отобранных генов присутствует важный ген мыши *Smah* (ENSMUSG00000016756), кодирующий гидроксиллазу цитидин монофосфо-N-ацетилнейраминной кислоты. Действительно, у людей наблюдается видоспецифический дефицит N-гликолилнейраминной кислоты, обусловленный псевдогенизацией *СМАН*, произошедшей у предков гоминин 2–3 млн лет назад.

Предложенные метод и программа обеспечивают эффективное предсказание ортологов с учётом геномных структур при одновременном рассмотрении сотен видов.

Благодарности. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18-29-13037. Вычисления выполнялись с использованием суперкомпьютеров Межведомственного суперкомпьютерного центра Российской академии наук (МСЦ РАН).

Литература

1. Razin S.V., Gavrilov A.A. *Structural-functional domains of the eukaryotic genome // Biochemistry (Moscow)*. 2018. Vol. 83, no. 4. P. 302–312.

UDC 577.29 DOI 10.37747/2312-640X-2020-18-258-260

GENE LOSS PREDICTION BASED ON GENOMIC STRUCTURE

Rubanov L.I., Shilovsky G.A., Seliverstov A.V., Zverkov O.A., Lyubetsky V.A.

*Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia
Bolshoy Karetny per. 19, build.1, Moscow 127051 Russia
e-mail: lyubetsky@iitp.ru*

We have developed an efficient algorithm implemented in a program for a multiprocessor computing system, which makes it possible to discover genes lost or acquired during the evolution of most species from a given set. The new approach takes into account the mutual arrangement of genes on a chromosome and allows us to simultaneously consider hundreds of species.

Key words: orthologous gene; synteny; gene loss; effective computer program; evaluation of the program's time; supercomputer.

Among the commonly used methods of inferring orthology and paralogy relations are: phylogenetic analysis based on the homologous sequences comparison and tree reconciliation, analysis of local gene arrangement (synteny) and reconstruction of evolutionary scenarios based on genome structures (the global arrangement of genes on chromosomes). It is known that sequence alignment and synteny can lead to inconsistent results. There is an ever growing set of algorithms and tools devoted to orthology analysis and related tasks. The essential feature of our approach to the search for lost or acquired genes is the simultaneous comparison of large sets of species.

The lossgainRSL program was developed to search for genes lost, acquired or preserved in most species from several species sets. The algorithm was implemented as a CLI program providing parallel computing in the MPI environment under Windows and Linux operating systems. The program is freely available at <http://lab6.iitp.ru/en/lossgainrsl/>. The lossgainRSL program has the following scheme. For a selected species, all its genes (each with a fixed-size neighborhood) are considered, which are present in or missing from several pre-defined parts of input sets of species. The condition for the gene selection can be arbitrary and is user defined. It is possible to work with incomplete genomes. The mentioned neighborhood of a gene (defined by a program's parameter) is used to check for synteny within its limits. It is important that the size of the neighborhood can be chosen reasonably. Namely, the 3D structure of DNA includes topologically associated domains (TADs). The TAD size defines the neighborhood and appears biologically sound [1].

The search time for a given gene ortholog in a given neighborhood is comparable to the time it takes to read the input tables of orthologs and paralogs and is bounded by a linear function of the program input length. Consequently, the total running time of the program is bounded by a second-degree polynomial of the input length, even for a single-processor machine. It allows us to simultaneously consider hundreds of species on a multiprocessor system.

Using the program, we have found the mouse genes, which were lost or became pseudogenes in human but are preserved (and have protein products) in no less than four out of five examined apes: *Gorilla gorilla*, *Nomascus leucogenys*, *Pan paniscus*, *Pan troglodytes*, and *Pongo abelii*. Among these genes there is an important mouse gene Cmah (ENSMUSG00000016756) which encodes cytidine monophospho-N-acetylneuraminic acid hydroxylase. Indeed, humans exhibit a species-specific deficiency of the N-glycolylneuraminic acid, due to pseudogenization of CMAH which occurred in hominin ancestors 2 to 3 MYA.

The proposed method and program provide an effective prediction of orthologs taking into account genomic structures while examining hundreds of species.

Acknowledgments. The reported study was funded by RFBR according to the research project № 18-29-13037. The research was carried out using supercomputers at the Joint Supercomputer Center of the Russian Academy of Sciences (JSCC RAS).

References

1. Razin S.V., Gavrilov A.A. Structural-functional domains of the eukaryotic genome // *Biochemistry (Moscow)*. 2018. Vol. 83, no. 4. P. 302–312.

УДК 575.83 DOI 10.37747/2312-640X-2020-18-260-262

ЭВОЛЮЦИЯ МИТОХОНДРИАЛЬНЫХ ГЕНОМНЫХ СТРУКТУР У МЕТАЗОА: АЛГОРИТМ И ПРОГРАММА

Горбунов К.Ю., Любецкий В.А.

Федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук. 127051, г. Москва, Большой Каретный переулок, д.19 стр. 1.
e-mail: lyubetsk@iitp.ru

Получен быстрый алгоритм и соответствующая программа для реконструкции геномных структур вдоль филогенетического дерева с произвольно назначаемыми ценами эволюционных событий. В зависимости от размера исходных данных, используются локальные, суперкомпьютерные или облачные вычисления. Программа тестировалась на искусственных данных и затем была применена для реконструкции предковых геномных структур митохондрий животных (Metazoa).

Ключевые слова: реконструкция, геномная структура, митохондрия, эффективный суперкомпьютерный алгоритм.