

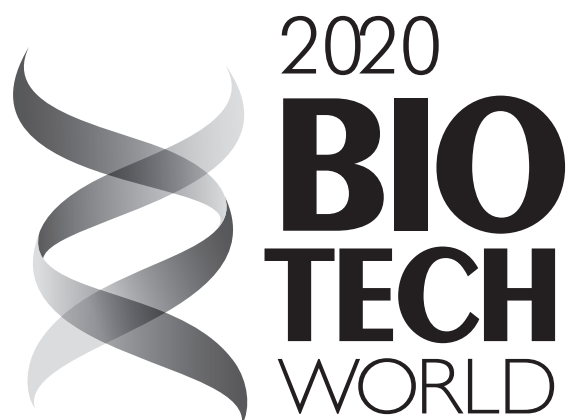
МАТЕРИАЛЫ ФОРУМА
FORUM PROCEEDINGS

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».

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.

Рассмотрены геномные структуры (взяты из базы данных RefSeq) митохондрий следующих видов: трихоплакс, пять губок, три гребневика, восемь видов типа Cnidaria (два коралловых полипа, три гидры и три из класса Мухозоа), шесть видов типа Xenacoelomorpha, четыре первичноротых и пять вторичноротых. И для них выполнена реконструкция эволюции с помощью оригинального алгоритма. В структуры отобраны белок-кодирующие гены и гены, кодирующие большие и малую субъединицы митохондриальной рибосомы. Каждая исходная структура содержала от 8 до 16 генов и состояла из одной кольцевой хромосомы, кроме трёх видов с линейными хромосомами: пять у *Clathrina clathrus*, одна у *Hydra oligactis* и две у *Hydra magnipapillata*. Направленный перебор цен показал, что следующий вариант брейкпойнтовых цен оптимален, а также соответствует биологическим представлениям об эволюции митохондрий (см. ниже): цена эволюционного события склейки краёв генов равна 2, цена их расклейки – 1, цена потери гена – 3, цена возникновения гена – 4. Алгоритм основан на следующем методе, обеспечивающем эффективность вычислений. Во внутренних вершинах дерева булевы переменные x (показатель склейки пар краёв генов) и y (показатель отсутствия гена) определяются по их известным значениям в листьях. Это делается динамическим программированием или более сложной оригинальной процедурой. Если ген отсутствует, т.е. $y=1$, склейки его краёв обнуляются, но остаётся возможность склейки одного края с двумя и более краями, что запрещено. Поэтому необходима процедура устанения такого рода «противоречий». В биологическом примере в каждой вершине возможен такой набор склеек, который приводят к одной кольцевой хромосоме. В 15 внутренних вершинах число таких виртуальных хромосом оказалось от двух до пяти. Направленным перебором всех комбинаций таких хромосом искались по кольцевой хромосоме в каждой вершине дерева, так чтобы эти хромосомы в совокупности минимизировали суммарное (по всем рёбрам дерева) расстояние между структурами на концах ребра. В качестве расстояния использовалась кратчайшая длина преобразования одной структуры в другую с ценами эволюционных событий: переклейка – 1, удаление связного участка хромосомы – 1.5, вставка такого участка – 2. Предлагаемая программа (см. <http://lab6.iitp.ru/ru/chromoggl/>) реконструировала все предковые структуры, они показаны на той же веб-странице. Отметим важность выбора цен эволюционных событий. Условие «цена склейки превышает цену расклейки» обеспечило возможность структуры из одной кольцевой хромосомы в каждой вершине дерева. Условие «цена возникновения гена превышает цену его потери» обеспечило, что удалений существенно больше, чем вставок, и большинство удалений расположено ближе к корню. В полученном сценарии эволюции 14 удалений и одна вставка. Условие «цены вставки и удаления связного участка превышают цены переклеек» позволяет соединять удаляемые гены в связный участок до их последующего удаления. Например, в сценарии при переходе от общего предка видов *Kudoa septempunctata* и *Hydra sinensis* к общему предку видов *Kudoa septempunctata* и *Kudoa iwatai* три участка генов, принадлежащих первой вершине, но не второй, сначала объединяются в связный участок, который затем удаляются. Программа обеспечивает выполнение таких общих требований, задаваемых на специальном языке спецификаций, на итоговый сценарий эволюции. Предложен алгоритм и программа вычисления сценария эволюции геномных структур с учётом биологически мотивированных цен и условий на итоговый сценарий. С помощью программы выполнена реконструкция эволюции митохондриальных геномных структур у Metazoa. Реконструкция с помощью быстрой программы позволяет уточнять филогенетическое положение видов.

Благодарность. Исследование выполнено при поддержке РФФИ (№ 18-29-13037).

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

EVOLUTION OF MITOCHONDRIAL GENOMIC STRUCTURES IN METAZOANS: ALGORITHM AND SOFTWARE

Gorbunov K.Yu., 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: lyubetsk@iitp.ru

A fast algorithm and software for the reconstruction of genomic structures along a phylogenetic tree with specified cost of evolutionary events have been generated. Depending on the data size, local, supercomputer, or cloud computing can be used. The software was tested on synthetic data and then applied to reconstruct the ancestral genomic structures of metazoan mitochondria.

Key words: reconstruction, genomic structure, mitochondria, efficient supercomputer algorithm.

Mitochondrial genomic structures (from RefSeq) of the Trichoplax, five sponges, three ctenophores, eight cnidarians (two anthozoans, three hydras, and three myxozoans), six xenacoelomorphs, four protostomes, and five deuterostomes were considered and the original algorithm was used to reconstruct their evolution. The structures included protein-coding genes and genes coding for the small and large subunits of the mitochondrial ribosome. Each initial structure contained 8 to 16 genes in a single circular chromosome except three species with linear chromosomes; five chromosomes in *Clathrina clathrus*; one in *Hydra oligactis*; and two in *Hydra magnipapillata*. A directional search of the costs has identified the following optimal set corresponding to the biological considerations (see below): the breakpoint gluing gene extremities cost equals 2; the ungluing gene extremities cost, 1; the gene loss cost, 3; and the gene emergence cost, 4. The algorithm rely on the following computation-efficient method. In the internal nodes of the tree, the Boolean variables x (the index of gluing gene extremities) and y (the index of gene absence) are defined from their available values in the leaves. This is performed by dynamic programming or a more complex original procedure. If a gene is missing, i.e., $y=1$, its gluing indices are set to zero but multiple gluings of the same extremity are yet possible, which necessitates a procedure eliminating such problems. In the above biological example, gluings are possible in any node, which results in a single circular chromosome. The number of such possible chromosomes ranged from two to 5 in 15 internal nodes. A directional search of all combinations of all such chromosomes was used to identify single circular chromosomes with the minimum total (for all tree edges) distance between the structures at the edge termini in all nodes. This distance is the shortest sequence of evolutionary events transforming one structure into another using the following costs: cut-and-paste, 1; deletion of the connected region, 1.5; insertion of the connected region, 2. The reconstruction was done using the above mentioned original software. Both the program and reconstruction are available at <http://lab6.iitp.ru/ru/chromogg/>. Notice the significance of evolutionary event costs. The condition that the gluing cost is greater than the ungluing cost made possible the formation of a single circular chromosome in each node. The condition that the gene emergence cost is greater than the gene loss cost provided that there were much more deletions than insertions and that most of deletions were close to the root. The resulting scenario included 14 deletions and 1 insertion. The condition that the insertion and deletion costs are greater than the cut-and-paste cost allowed the deleted genes to be pooled before their deletion. For instance, in the scenario the transition from the common ancestor of *Kudoa septempunctata* and *Hydra sinensis* to the common ancestor of *Kudoa septempunctata* and *Kudoa iwatai*, three gene regions that belong to the first but not the second node are first pooled and then deleted. The software allows such general evolutionary conditions to be satisfied in the generated scenario. An algorithm and software computing the evolutionary scenario for the genomic structures with an account of biologically motivated costs and conditions were proposed. The software was used to reconstruct the evolution of mitochondrial genomic structures in Metazoans. Such fast computer-aided reconstruction can be used to refine the phylogenetic position of species.

Acknowledgments. This study was funded by Russian Foundation for Basic Research (project no. 18-29-13037).

УДК 573.22 DOI 10.37747/2312-640X-2020-18-262-265

БИОИНФОРМАТИЧЕСКИЙ АНАЛИЗ КЛАСТЕРОВ ГЕНОВ БИОСИНТЕЗА НЕРИБОСОМНЫХ ПЕПТИДОВ БАКТЕРИЙ НА ОСНОВЕ ПРЕДСКАЗАНИЯ ЭФФЕКТИВНОСТИ ЭЛОНГАЦИИ ТРАНСЛЯЦИИ

Клименко А.И., Лашин С.А., Афонников Д.А., Матушкин Ю.Г.

Институт цитологии и генетики СО РАН, Новосибирск,
 Россия, 630090, Новосибирск, пр. Лаврентьева 10
 e-mail: mat@bionet.nsc.ru

Обнаружено, что бактерии, имеющие и не имеющие кластеры генов биосинтеза для нерибосомных пептид-синтетаз, существенно различаются по предпочтительным эволюционным стратегиям оптимизации первичной структуры кодирующих частей генов.

Ключевые слова: бактерии, эффективность элонгации трансляции, нерибосомные пептид синтетазы.

Эффективность элонгации трансляции – это характеристика «оптимальности» первичной структуры генов в организме: чем активнее происходит экспрессия указанных генов, тем выше индекс эффективности элонгации (ИЭЭ) [1]. Оптимизация может проходить по частотам используемых кодонов, минимизации