

МЕЖДУНАРОДНЫЙ КОНГРЕСС

**БИОТЕХНОЛОГИЯ:
СОСТОЯНИЕ
И ПЕРСПЕКТИВЫ
РАЗВИТИЯ**

ВЫПУСК 19

26 - 29 ОКТЯБРЯ 2021
МОСКВА

INTERNATIONAL CONGRESS

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

ISSUE 19

26 - 29 OCTOBER, 2021
MOSCOW

WWW.BIOMOS.RU

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

Материалы международного конгресса
«Биотехнология: состояние и перспективы
развития»
26 - 29 ОКТЯБРЯ 2021 г.

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

Материалы тезисов публикуются в авторской
версии. Организаторы не несут ответственности
за неточности и упущения 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 978-5-6045396-1-3
ISSN: 2312-640X
DOI: 10.37747/2312-640X-2021-19

© 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 конгресса
«Биотехнология: состояние и перспективы
развития»

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

The proceedings of International congress
«Biotechnology: state of the art and perspectives»
26 - 29 OCTOBER, 2021

DISCLAIMER

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

Opinions expressed may not necessarily reflect the
position of the Scientific Council of congress.
Information in the «Biotechnology: state of the art
and perspectives» 2021 Congress 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 congress.

ISBN 978-5-6045396-1-3
ISSN: 2312-640X
DOI: 10.37747/2312-640X-2021-19

Copyright © LTD «EXPO-BIOHIM-TECHNOLOGIES»
apt. 154, bldg 10/2, Olimpiyskiy Pr-kt, Moscow,
info@biomos.ru www.biomos.ru

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

УДК 575.83 DOI:10.37747/2312-640X-2021-19-207-209

РЕКОНСТРУКЦИЯ ЭВОЛЮЦИИ ГЕНОМНЫХ СТРУКТУР С ПАРАЛОГАМИ

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

Для любых нагруженных ориентированных chain-cycle graphs a и b (называемых структурами) и любых равных цен операций (переклеек и дубликации), получен алгоритм, который, последовательно применяя эти операции к a , выдаёт b , если первая структура не содержит паралогов (рёбер с повторяющимся именем), а во второй не более двух паралогов для каждого ребра. Алгоритм имеет мультипликативную ошибку не более $13/9 + \epsilon$, где ϵ – любое строго положительное число, и время его работы порядка $n^{O(\epsilon^{-2.6})}$, где n – размер исходной пары графов. Также получены алгоритмы реконструкции эволюции геномных структур с условием на предковые структуры и вдоль филогенетического дерева.

Ключевые слова: геномная структура; геномные перестройки; преобразование генома; DCJ-операции; дубликация гена; мультипликативно точный алгоритм.

Структурой называется любой ориентированный граф, компоненты которого – цепи и циклы, а рёбрам приписаны имена (в этом состоит нагруженность графа). Ребро всегда рассматривается в паре с его именем (но не с ориентацией). Структура описывает геном, где каждая компонента играет роль хромосомы, а ребро графа – роль гена. Повторение ребра состоит в присутствии в структуре второго ребра с тем же именем; число разрешённых повторений существенно влияет на сложность задачи. Рассматриваются задачи преобразования и реконструкции структур, которые состоят в поиске: (1) кратчайшего преобразования одной структуры a в другую b посредством заданных операций над структурами; и (2) кратчайшей реконструкции структур, заданных в листьях дерева, на его внутренние вершины. Задачу (2) иногда описывают как поиск кратчайшей расстановки структур, заданных в листьях дерева, по его внутренним вершинам. В этих задачах «кратчайшее» означает минимум соответствующего функционала. В первом случае функционал это – суммарная цена преобразования a в b , а во втором случае – суммарная цена расстановки структур по внутренним вершинам дерева. В точке минимума говорят, соответственно, о «кратчайшей цене» преобразования/расстановки и «кратчайшем преобразовании»/«кратчайшей расстановке». Таким образом, обе задачи описывают эволюцию геномных структур a в b , и реконструкцию эволюции вдоль филогенетического дерева видов. В структуре ребро с повторяющимся именем называется паралогичным, а каждое его повторение – паралогом (соответствующего ребра). Звездой называется дерево, состоящее из корня и рёбер, которые ведут из корня в листья. Локально кратчайшая реконструкция (расстановка) – такая, что в любой одной вершине дерева замена структуры в ней на любую другую не уменьшает значение функционала. Если две или более структур не имеют паралогов и в них множества имён рёбер равны, то говорят, что это – структуры с равным составом; если не имеют паралогов и множества имён рёбер не обязательно равны – с неравным составом. Структуры, в которых паралоги могут присутствовать, но множества имён равны, назовём с квазиравным составом. DCJ-Операции, которыми разрешено преобразовывать структуру a в b , в последовательности промежуточных структур от a до b могут повторяться неограниченное число раз. Каждая операция имеет свою цену – строго положительное рациональное число. Здесь снова различают два случая: цены всех операций равны (тогда их можно считать равными 1) или это не так (разные цены). Различают два случая задачи преобразования: все структуры без паралогов или паралоги разрешаются; соответственно говорят о задаче преобразования без паралогов и с паралогами. Задача с паралогами алгоритмом линейной сложности сводится к задаче целочисленного линейного программирования. В присутствии паралогов эта задача NP-трудная. Для NP-трудных задач важна разработка эффективных приближённых, но с хорошей точностью алгоритмов. Эффективность предполагает, что время работы алгоритма описывается полиномом низкой степени; обычно 1-ой или 2-ой степени от размера исходных данных. Кроме алгоритма, указанного в абстракте, получены, в частности, следующие результаты. При равном множестве имён в структурах и равных ценах операций рассмотрены следующие условия на них: все структуры из одного цикла, все из одной цепи, все из цепей. Для каждого условия получен точный (т.е. с нулевой ошибкой) алгоритм преобразования данных a в b , квадратичный по времени работы. Для другого списка операций (склейка и расклейка вершины, удаление и добавление ребра) над структурами и

уже произвольных цен операций получен алгоритм продолжения структур, заданных в листьях дерева, на его внутренние вершины. Алгоритм точен, если дерево – звезда; при этом в листьях структуры могут быть даже с неравным множеством имён или с паралогами. Время работы алгоритма порядка $nX + n^2 \log(n)$, где n – число имён в листьях и X – легко вычисляемая характеристика структур в листьях. В общем случае алгоритм находит локально минимальное решение и время его работы кубическое.

Благодарность. Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 18–29–13037.

UDC 575.83 DOI:10.37747/2312-640X-2021-19-207-209

RECONSTRUCTION OF EVOLUTION OF GENOME STRUCTURES WITH PARALOGS

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: lyubetsky@iitp.ru*

For any weighted directed chain-cycle graphs a and b (referred to as *structures*) and any equal costs of operations (intermergings and duplication), we obtain an algorithm which, by successively applying these operations to a , outputs b if the first structure contains no paralogs (edges with a repeated name) and the second has no more than two paralogs for each edge. The algorithm has a multiplicative error of at most $13/9 + \varepsilon$, where ε is any strictly positive number, and its runtime is of the order of $n^{O(\varepsilon^{-2.6})}$, where n is the size of the initial pair of graphs. We also obtain algorithms for reconstruction of the evolution of genome structures with a condition on ancestor structures and along the phylogenetic tree.

Keywords: genome structure; genome rearrangements; genome transformation; DCJ operations; gene duplication; multiplicatively exact algorithm.

A *structure* is any directed graph in which the components are chains and cycles, and edges are assigned with names (in this respect, the graph is weighted). An edge is always considered in a pair with its name (but not with its direction). A structure describes a genome, where every component plays the role of a chromosome and every edge of the graph plays the role of a gene. Repetition of an edge consists in the presence of another edge with the same name in the structure; the number of allowed repetitions strongly influences the complexity of the problem. We consider the problems of transformation and reconstruction of structures, which consist in (1) finding *the shortest transformation* of one structure, a , into another, b , using pre-fixed operations over structures; and (2) finding *the shortest reconstruction* of structures specified at leaves of a tree at its interior vertices. Problem (2) is sometimes described as finding *the shortest arrangement* of structures specified at leaves of a tree over its interior vertices. In these problems, 'the shortest' means the minimum value of the corresponding functional. In the first case, the functional is the *total cost of the transformation of a into b* ; in the second, *the total cost of arrangement* of structures over interior vertices of the tree. At the minimum point, the terms 'shortest cost' of a transformation/arrangement and, respectively, 'shortest transformation'/'shortest arrangement' are used. Thus, the two problems describe the evolution of genome structures, a into b , and reconstruction of the evolution along the phylogenetic tree of species. An edge with a repeated name in a structure is said to be *paralogic*, and each of its repetitions is called a *paralog* (of the corresponding edge). A star is a tree consisting of a root and edges leading from the root to leaves. A reconstruction (arrangement) is said to be *locally shortest* if at any single vertex of the tree, changing the structure in it to any other does not reduce the value of the functional. If two or more structures have no paralogs and sets of edge names in them are the same, these structures are said to have *equal content*; if there are no paralogs but the sets of edge names are not necessarily the same, they have *unequal content*. Structures where paralogs may occur but sets of names are the same will be called structures with *quasi-equal content*. DCJ operations allowed to be used to transform a structure a into b can be repeated unlimitedly many times in a sequence of intermediate structures from a to b . Each operation has its own cost, a strictly positive rational number. Here, again two cases are distinguished: costs of all operations are *equal* (in this case we may assume them to be equal to 1), or this is not so (*unequal costs*). Two cases of the transformation problem are distinguished: all structures have no paralogs, or paralogs are allowed; accordingly, we speak about the transformation problem *without* or *with paralogs*. The problem

with paralogs can be *reduced* by a linear complexity algorithm to an integer linear programming problem. When paralogs are allowed, this problem is NP-hard. For NP-hard problems, designing efficient approximate algorithms but with good accuracy is of importance. The efficiency assumes that the algorithm runtime is described by a low-degree polynomial; usually, of degree 1 or 2 in the size of initial data. Besides the algorithm mentioned in the abstract, we have obtained, in particular, the following results. For equal sets of names in the structures and equal operation costs, we have the following conditions on them: all structures are from one cycle, all are from one chain, all are from chains. For each of the conditions, we have obtained an exact (i.e., zero-error) algorithm for the transformation of a given a into a given b with quadratic runtime. For another list of operations (join and cut of a vertex, deletion and insertion of an edge) over structures and now arbitrary operation costs, we have obtained an algorithm for the extension of structures specified at leaves of a tree onto its interior vertices. The algorithm is exact if the tree is a star; in this case, structures in the leaves may even have unequal sets of names or paralogs. The runtime of the algorithm is of the order of $nX + n^2 \log(n)$, where n is the number of names in the leaves and X is an easily computable characteristic of the structures at the leaves. In the general case, the algorithm finds a locally minimal solution, and its runtime is cubic.

Acknowledgments. This research was funded by the Russian Foundation for Basic Research under research project No. 18-29-13037.

УДК 544.165 DOI:10.37747/2312-640X-2021-19-209-211

РЕПОЗИЦИОНИРОВАНИЕ ЛЕКАРСТВ ДЛЯ ТЕРАПИИ COVID-19

Поройков В.В., Дружиловский Д.С., Савосина П.И., Гомазков О.А., Соболев Б.Н., Веселовский А.В., Тарасова О.А., Бизюкова Н.Ю., Ионов Н.С., Иванов С.М., Рудик А.В., Столбов Л.А., Погодин П.В., Карасев Д.А., Дмитриев Д.А., Лагунин А.А., Филимонов Д.А.

Институт биомедицинской химии им. В.Н.Ореховича, Москва, Россия
119121, Москва, ул. Погодинская, д. 10, стр. 8
e-mail: vvp1951@yandex.ru

Репозиционирование лекарств – выявление новых показаний у разрешенных к медицинскому применению лекарственных препаратов – является естественным оперативным ответом на вирусную инфекцию SARS-CoV-2/COVID-19. В докладе рассмотрено современное состояние поиска и разработки лекарственных препаратов для терапии COVID-19 с применением *in silico* и *in vitro* методов.

Ключевые слова: SARS-CoV-2; COVID-19; фармакологические мишени; репозиционирование лекарств; виртуальный скрининг; молекулярное моделирование; машинное обучение; исследования *in vitro*

Под репозиционированием лекарств понимают выявление новых показаний для разрешенных к медицинскому применению препаратов. Наличие информации о фармакологических и токсикологических характеристиках известного препарата обеспечивает условия для его быстрого применения по новой нозологии (Poroikov and Druzhilovskiy, 2019). Необходимость оперативного реагирования на пандемию COVID-19 стимулировала масштабные исследования в этом направлении. Поиск в Google по запросу «COVID-19 AND drug repurposing» дает около 5 миллионов документов. Лекарственные препараты Ремдесивир, Фавипиравир и Умифеновир (Арбидол) исходно разрабатывались для других показаний и были репозиционированы для терапии инфекции SARS-CoV-2. Сюда же можно отнести препараты Триазавирин, Нобазит, Нафамостат и ряд других, для которых в настоящее время проводятся клинические испытания на пациентах с COVID-19. Для поиска новых фармакологических эффектов у известных лекарств проводят исследования *in silico* и *in vitro*. Компьютерные оценки получают путем моделирования взаимодействия анализируемых соединений с молекулярными мишенями, идентификацией аналогов на основе структурного сходства, анализом зависимостей «структура-активность» методами машинного обучения, установлением ассоциаций методами сетевой фармакологии (Muratov et al., 2021). Первоначальный отбор «хитов» проводят с помощью методов *in silico*, которые могут применяться к виртуальным (еще не синтезированным) молекулам, с последующей экспериментальной валидацией компьютерных предсказаний *in vitro*. Определение антикоронавирусной активности *in vitro* осуществляют с использованием биохимических и клеточных модельных систем (Savosina et al., 2021; Mslati et al., 2021). Предварительный отбор потенциально активных соединений на основе оценок *in silico* существенно повышает шансы на успех (Gentile et al., 2020).