

УДК: 004.75

Интеграция гетерогенных вычислительных инфраструктур для анализа данных геномного секвенирования

Аулов В.А.¹, Климентов А.А., Машинистов Р.Ю.*¹, Недолужко А.В.¹,
Новиков А.М.¹, Пойда А.А.**¹, Тертычный И.С.¹,
Теслюк А.Б.¹, Шарко Ф.С.¹

¹Национальный исследовательский центр «Курчатовский институт», Москва, Россия

Аннотация. Недавние достижения технологий в области геномного секвенирования нового поколения (NGS – Next Generation Genome Sequencing) привели к значительному увеличению объема данных, которые должны быть обработаны, проанализированы и доступны заинтересованным исследователям. Это, в свою очередь, привело к повышению требований к вычислительным платформам обработки данных: потребовалось больше оперативной памяти и более мощные процессоры. Для эффективной обработки данных необходимы принципиально новые подходы в организации вычислений. Авторами статьи было проведено исследование возможности применения методов и подходов, используемых в физике высоких энергий, для объединения гетерогенных вычислительных ресурсов в единую вычислительную платформу. Была разработана полномасштабная система управления данными и заданиями на базе вычислительных мощностей Национального исследовательского центра «Курчатовский институт». В разработанную систему был интегрирован рабочий поток для обработки данных геномного секвенирования с использованием пакета PALEOMIX. Результаты апробации разработанной системы на задаче анализа древней ДНК мамонта показали существенное уменьшение общего времени выполнения задачи.

Ключевые слова: *распределенные вычисления, суперкомпьютеры, геномное секвенирование.*

ВВЕДЕНИЕ

Недавние достижения технологий в области геномного секвенирования нового поколения (NGS – Next Generation Genome Sequencing) [1] привели к значительному увеличению объема данных, которые должны быть обработаны, проанализированы, а результаты должны быть доступны удаленным группам биоинформатиков [2, 3]. Одной из актуальных задач в данной области является анализ древней ДНК.

Древняя ДНК – генетический материал, извлеченный из древних биологических образцов. Первые исследования, с выделением и анализом древних фрагментов ДНК начались более 30 лет назад – первоначально работы проводились с небольшими участками митохондриального или ядерного генома. Спустя три десятилетия с развитием молекулярно-биологических методик, новых подходов к секвенированию

*rmashinistov@gmail.com

**poyda@wdcb.ru

(определению последовательности ДНК) и биоинформатических методов появилась реальная возможность заглянуть в далекое прошлое.

Исследователям удалось восстановить геномную информацию из самых разнообразных типов древнего биологического материала: волос [4, 5], мумифицированных тканей [6], кальцинированных зубов и костей [7–9], а также растительных останков [10, 11]. С появлением новых технических возможностей миллионы и миллиарды последовательностей ДНК могут быть получены из древних биологических образцов, благодаря массивной пропускной способности современных платформ для секвенирования следующего поколения.

К настоящему времени исследования в области древней ДНК все чаще используются для решения многих фундаментальных и прикладных вопросов. Возможность использования ДНК из археологического и палеонтологического материала позволяет решать многочисленные задачи, связанные с эволюцией экосистем в различных климатических условиях, с происхождением и эволюцией многих патогенных микроорганизмов, например чумной палочки, возбудителей туберкулеза или бруцеллеза [12]. В данной работе использовались геномные чтения шерстистого мамонта (*Mammuthus primigenius*), опубликованные ранее [13].

В настоящее время для анализа древней ДНК используются специальные программные пакеты-конвейеры, который включающие ряд программных компонент, с помощью которых осуществляется быстрая обработка данных NGS [14]. Одним из наиболее популярных программных конвейеров является пакет PALEOMIX [15]. Но его использование сопряжено с некоторыми неудобствами:

1. Требуется большой объем ручной работы, зачастую с привлечением экспертов по PALEOMIX. Весь процесс состоит из большого количества отдельных этапов, например загрузки входных данных на вычислительный кластер, конфигурации программного обеспечения, запуска вычислительной задачи, мониторинга выполнения и повторного запуска в случае возникновения ошибки, выгрузки результатов. Выполнение этих этапов требует достаточно много времени. Автоматизация данных процессов позволит значительно упростить вычисления с точки зрения конечных пользователей.

2. Требование достаточно специализированной вычислительной инфраструктуры. PALEOMIX не предназначен для распараллеливания между вычислительными машинами. В то же время для корректной работы он требует большой объем оперативной памяти. Таким образом, для его запуска требуется мощный сервер с большим объемом оперативной памяти. В то же время использование обычных кластеров и вычислительных машин позволит отказаться от закупки дорогостоящих серверов.

3. Большие временные затраты. Так, например, задача анализа древней ДНК, может считаться до двух месяцев даже с использованием мощных вычислительных ресурсов. Однако в крупных научных центрах доступны вычислительные ресурсы коллективного пользования (университетские кластеры, суперкомпьютеры), загрузка которых сильно варьируется и может достигать 80–90%. Обеспечение возможности использования свободного времени ЦПУ этих ресурсов может позволить значительно увеличить эффективность научных исследований в случае обеспечения автоматизированных механизмов мониторинга и захвата доступных ресурсов. Эта задача также может быть возложена на систему управления данными и заданиями, которая ко всему прочему будет настроена на работу с конкретными вычислительными ресурсами, обеспечивая аутентификацию пользователей, соблюдение выделенных квот и политик безопасности.

Для решения описанных выше проблем могут быть использованы методы организации распределенных вычислений и автоматизации управления данными и задачами, применяемые в других областях наук, например, в физике высоких энергий.

Например, для обработки данных эксперимента ATLAS [16] во время первого рабочего запуска БАК [17] были использованы десятки вычислительных центров и хранилищ мультипетабайтного масштаба по всему миру. Использование специальной системы управления потоком заданий (Workload Management System) PanDA [18, 19] позволило объединить гетерогенные вычислительные мощности (в основном входящие в инфраструктуру Грид – WLCG [20]) и предоставить ученым унифицированный доступ к вычислительным ресурсам.

Целью данной работы стало исследование возможности применения опыта организации распределенных вычислений на основе системы PanDA в физике высоких энергий для организации распределенных вычислений в других научных областях, требующих привлечения больших вычислительных ресурсов.

В статье описана разработанная авторами полномасштабная система управления данными и заданиями в среде федеративных гетерогенных ресурсов на базе вычислительных мощностей Национального исследовательского центра «Курчатовский институт» (<http://computing.kiae.ru/>), а также ее применение для задачи обработки данных геномного секвенирования с использованием пакета PALEOMIX. Разработанная система позволяет:

- запускать задачи PALEOMIX по анализу данных геномного секвенирования на распределенной гетерогенной вычислительной инфраструктуре;
- использовать для вычисления одной задачи вычислительные ресурсы с различной инфраструктурой: от суперкомпьютеров до обычных вычислительных машин;
- контролировать и управлять данными и задачами в автоматическом режиме, сведя участие пользователя к минимуму.

Мы разработали информационный портал, объединяющий различные вычислительные инфраструктуры, в том числе суперкомпьютер с пиковой производительностью 122 TFLOPS и платформу облачных вычислений. В рамках данной работы мы развернули систему PanDA, адаптировав ее к уникальной архитектуре вычислительного центра института и реализовав дополнительные программные модули, необходимые для работы портала.

ОПИСАНИЕ ИСПОЛЬЗУЕМОГО ПОДХОДА И АРХИТЕКТУРЫ РЕАЛИЗУЮЩЕЙ ЕГО ПРОГРАММНОЙ СИСТЕМЫ

Разработанный портал можно условно разделить на три части. Основным компонентом портала является система PanDA. Наиболее важными составляющими системы являются сервер и подсистема пилотных заданий. Сервер распределяет поступающие задачи на доступные вычислительные ресурсы, отслеживает состояние и жизненный цикл задач. Пилотные задания осуществляют механизм поздней привязки задачи к среде вычислений. Они запускаются на ресурсах, предназначенных для выполнения пользовательских задач, проверяют их состояние, резервируют, собирают информацию и передают на сервер, после чего инициируют запуск полученной от сервера задачи непосредственно на ресурсах и контролируют процесс ее выполнения.

Вторым важным компонентом является вычислительная инфраструктура, включающая суперкомпьютер и облачную платформу. Суперкомпьютер НИЦ «Курчатовский институт» – высокопроизводительный вычислительный кластер второго поколения с пиковой производительностью 122,9 TFLOPS, был сдан в эксплуатацию с сентября 2011 года. Кластер состоит из 1280 счётных двухпроцессорных узлов, объединенных высокопроизводительной сетью передачи данных и сообщений InfiniBand DDR, имеет суммарную оперативную память 20,5 Тбайт и систему хранения данных на 144 Тбайт. На счётных узлах кластера установлена операционная система Linux (CentOS). Система хранения данных построена на параллельной файловой системе Lustre 2.0. Для управления распределением ресурсов и выполнением счетных

заданий используется менеджер ресурсов SLURM. Характеристики узлов приведены в таблице 1.

Таблица 1. Характеристики узлов суперкомпьютера НИЦ “Курчатовский институт”

Счётные узлы на процессорах Intel Xeon E5450 (3,00 ГГц, 4 ядра)	
Количество процессоров на узел	2
Количество ядер на узел	8
Оперативная память на узел (Гбайт)	16
Оперативная память на ядро (Гбайт)	2
Локальная дисковая память на узел (Гбайт)	120
Общее количество узлов	1280
Общее количество процессоров	2560
Общее количество ядер	10240
Общая пиковая производительность (TFLOPS)	122,9

Инфраструктура облачных вычислений, развернутая в НИЦ “Курчатовский институт”, состоит из 16 узлов и дополнительного хранилища на 60 ТБ.

Третьей составляющей являются разработанные программные компоненты портала, включающие веб-интерфейс формирования и запуска задач, файловый каталог и систему транспортировки файлов.

Для более эффективного взаимодействия пользователей с системой нами был разработан интерфейс, позволяющий загружать входные файлы из различных хранилищ, определять и запускать пользовательские задачи, контролировать процесс и отслеживать статус их выполнения (мониторинг). Данный интерфейс состоит из нескольких программных модулей.

NEW JOB

SOFTWARE:

INPUT FILES: drag & drop 2 files
 Drag and Drop Files Here

INPUT FILES: ftp dir

INPUT FILES: http url

INPUT FILES: guid

INPUT FILES: container
 One file one job

CORES:

RUN SCRIPT:

```
mkdir tmp; mkdir out; bam_pipeline run --max-threads=2 --jar-root=$JAR_ROOT --temp-root=tmp --destination=out ./Mammoth.yam
```

TAGS:

Рис. 1. Унифицированная веб-форма для определения новых пользовательских задач.

Непосредственное взаимодействие с пользователем берет на себя веб-интерфейс представленный на рисунке 1, который предоставляет унифицированную веб-форму для определения новых пользовательских задач. С помощью формы пользователь

генерирует описание задачи и отправляет ее на сервер, после чего он может наблюдать за состоянием своей задачи с помощью встроенной в веб-интерфейс системы мониторинга. Портал выполняет аутентификацию с помощью логина и пароля пользователей – сотрудников Курчатовского Института, использующих портал для обработки своих данных. При этом также обеспечивается прозрачность при запуске вычислительных задач. От пользователя скрыты определенные параметры запуска задач и некоторые технические характеристики сервера.

В ряде случаев необходима генерация большого числа вычислительных задач. При этом использование веб-интерфейса может быть неудобным. В таких случаях научные коллективы могут использовать собственные приложения или программные сценарии, позволяющие создавать описания вычислительных задач. Для интеграции с такими средствами был создан API (интерфейс прикладного программирования), позволяющий производить действия, доступные пользователям веб-интерфейса, с помощью HTTP запросов.

Перемещением пользовательских и системных файлов между распределенными вычислительными узлами занимается система транспортировки файлов. Также система поддерживает механизм репликации и обеспечение согласованности реплик. Все файлы, задействованные в работе, их реплики и правила доступа к ним описываются в специальном файловом каталоге. Совместное использование системы транспортировки файлов и файлового каталога обеспечивает возможность быстрого подключения сторонних систем хранения (например, Dropbox, Google Drive и т.д.).

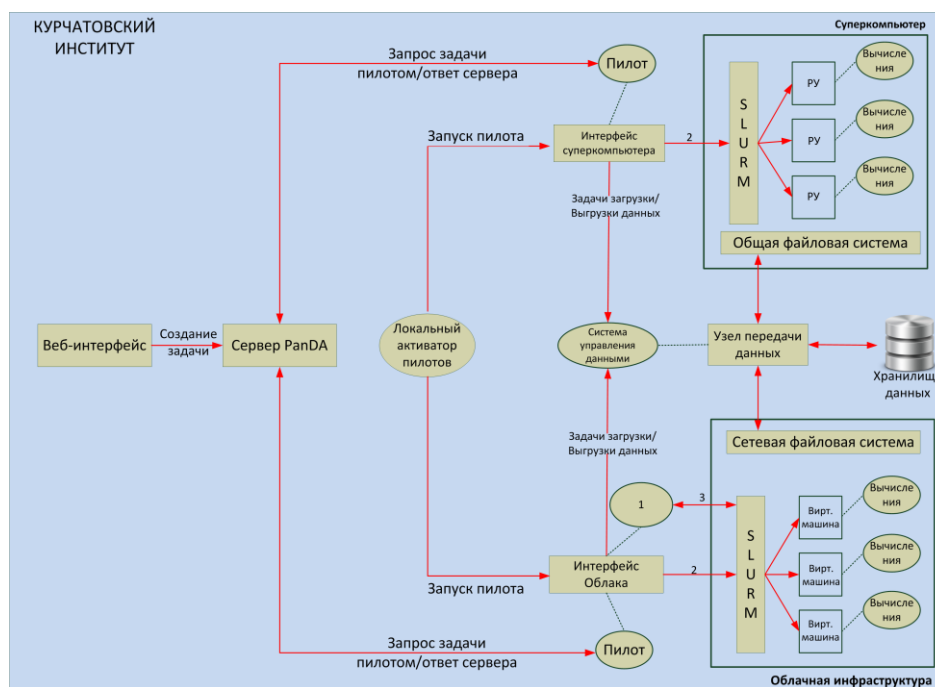


Рис. 2. Схема разработанного портала: 1 – балансировщик ресурсов, 2 – запуск вычислений, 3 – балансировка ресурсов.

На рисунке 2 представлена схема портала. Вычислительные задачи, определенные пользователями с помощью веб-интерфейса, передаются на сервер PanDA. Программный сценарий локального активатора пилотов инициирует запуск пилотов на виртуальных машинах, являющихся шлюзами для суперкомпьютера (Интерфейс суперкомпьютера) и облачной инфраструктуры (Интерфейс Облака). Пилот выполняет https запрос к серверу PanDA и получает на обработку вычислительную задачу. Далее пилот ставит вычислительную задачу на выполнение в очередь при помощи менеджера ресурсов SLURM. При этом система управления данными выполняет перемещение

входных файлов, необходимых задаче, из хранилища данных в файловую систему вычислительного ресурса (Общая файловая система, Сетевая файловая система). Балансировщик ресурсов управляет динамическим изменением состава выделенных виртуальных машин на платформе облачных вычислений в зависимости от загрузки системы.

ЗАДАЧИ АНАЛИЗА ДАННЫХ ГЕНОМНОГО СЕКВЕНИРОВАНИЯ

Одним из важных направлений в современной биологии являются исследования с использованием геномного секвенирования нового поколения (Next Generation Sequencing, NGS). Последние достижения технологий в этой области привели к увеличению объемов данных, которые необходимо обработать, проанализировать, а затем предоставить доступ к результатам биологам по всему миру.

Одной из актуальных задач в данной области является анализ древней ДНК. Для решения таких задач часто применяется популярный программный конвейер PALEOMIX, который включает в себя ряд программных компонент с открытым исходным кодом, с помощью которых осуществляется быстрая обработка данных NGS.

Предварительно с помощью конвейера PALEOMIX был проведен анализ данных геномного секвенирования шерстистого мамонта. Было обработано более 350 Гигабайт входных данных, содержащих более 900 миллионов парных чтений, что потребовало около двух месяцев счета на одном 80-ядерном сервере с 512 ГБ оперативной памяти.

Некоторые из компонент PALEOMIX могут обрабатывать отдельные фрагменты входных данных независимо от остальных фрагментов. Авторы статьи воспользовались этим обстоятельством и распараллелили работу PALEOMIX путем разбиения входных данных на большее число файлов меньшего размера и параллельной обработки этих файлов с последующей сборкой выходных данных из результатов обработки отдельных файлов. На рисунке 3 представлена схема модификации классического алгоритма работы PALEOMIX в параллельный алгоритм PanDA PALEOMIX.

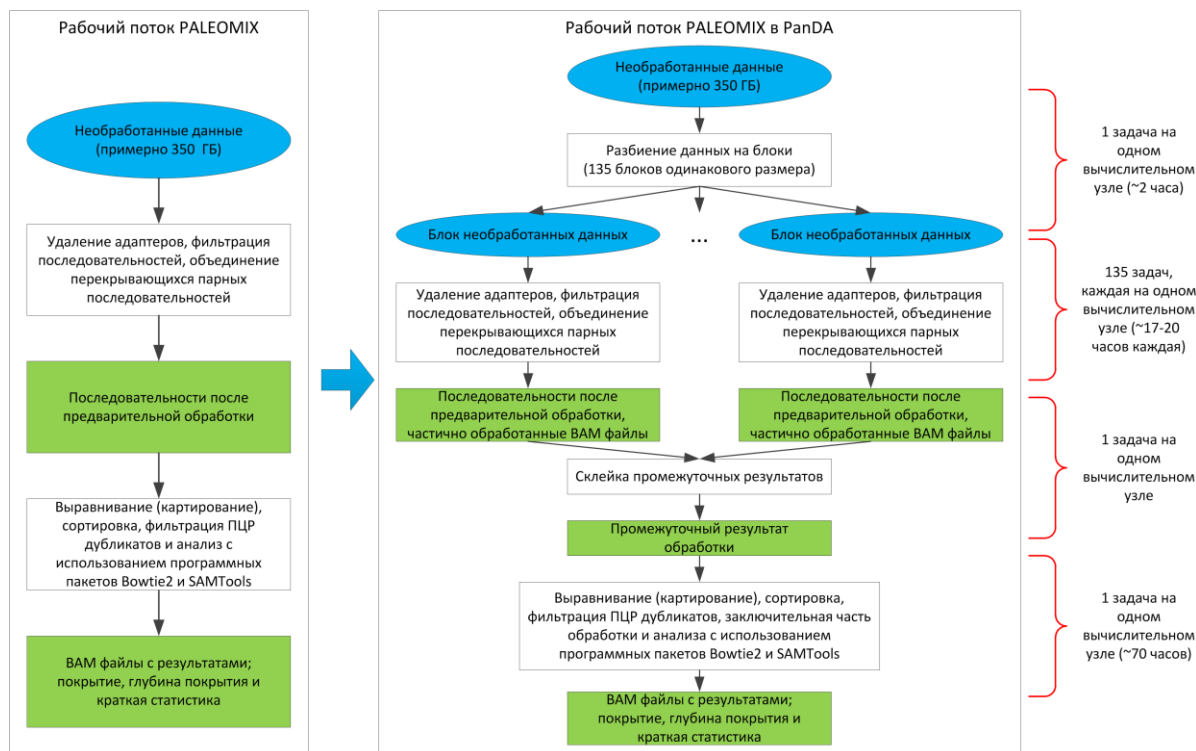


Рис. 3. Схема модификации классического алгоритма работы PALEOMIX в параллельный алгоритм PanDA PALEOMIX.

На первом этапе производится разбиение всего пространства данных на отдельные файлы. Этот процесс оформлен в виде одной PanDA-задачи, выполняемой на одном узле. На втором этапе для каждого полученного файла запускается конвейер PALEOMIX в виде отдельной PanDA-задачи. Эти задачи могут выполняться в параллельном режиме на распределенной инфраструктуре. На третьем этапе при помощи конвейера PALEOMIX объединяются все результаты с предыдущего этапа, что реализуется в одной PanDA-задаче.

Таким образом, для запуска приложений по секвенированию генома в разработанном портале был использован принятый в физике высоких энергий подход к обработке больших файлов: исходный файл разрезается на большое число фрагментов меньшего размера; каждый из полученных файлов обрабатывается как независимое вычислительное задание на разных узлах суперкомпьютера; результаты выполнения всех заданий объединяются в один выходной файл. Такой подход позволил существенно уменьшить общее время на обработку данных полногеномного секвенирования с нескольких недель до нескольких дней.

ЗАКЛЮЧЕНИЕ

В ходе выполнения проекта была разработана полномасштабная система управления данными и заданиями в среде федеративных гетерогенных ресурсов на базе вычислительных мощностей Национального исследовательского центра «Курчатовский институт». В разработанную систему был интегрирован рабочий поток для обработки данных геномного секвенирования с использованием пакета PALEOMIX, что позволило:

- Запускать задачи PALEOMIX по анализу данных геномного секвенирования на распределенной гетерогенной вычислительной инфраструктуре. Размер и количество файлов – фрагментов исходных данных – могут быть подобраны таким образом, что бы с максимальной эффективностью использовать конкретные вычислительные ресурсы. Размер файла напрямую связан с объемом требуемой оперативной памяти и временем обработки.

- Использовать для вычисления одной задачи вычислительные ресурсы с различной инфраструктурой: от суперкомпьютеров до обычных вычислительных машин. В частности для решения задачи анализа ДНК мамонта был использован суперкомпьютер НИЦ КИ. В исходном виде задача не могла быть решена ни на Курчатовском суперкомпьютере, ни на обычном кластере из-за высоких требований к объему оперативной памяти на ядро.

- Контролировать и управлять данными и задачами в автоматическом режиме, сведя участие пользователя к минимуму. Использованный подход обеспечивает гибкость и надежность вычислений. Авторизованный конечный пользователь взаимодействует с системой посредством интуитивно-понятного интерфейса.

Разработанная система была успешно применена для вычислительной задачи анализа ДНК мамонта. Результаты показали существенное уменьшение общего времени обработки данных (с двух месяцев до нескольких дней).

Представленный проект показал, что программные средства и методы обработки больших объемов экспериментальных данных, которые были разработаны в области физики высоких энергий для экспериментов на ускорителе БАК, могут быть успешно применены в других областях науки.

Данный метод наиболее применим в случае, когда задача может быть представлена в следующем виде. Набор типовых задач для выбранной предметной области представляется в виде программных сценариев, на вход которым подаются файлы исходных данных. Входными данными являются как непосредственно файлы, содержащие данные для обработки, так и файлы конфигураций программного обеспечения. На выходе сценария получаются файлы обработанных данных. При

выполнении этого условия задача может быть успешно решена за разумный промежуток времени с помощью современных высокопроизводительных вычислительных ресурсов.

Реализация представленной системы стала возможной благодаря наработкам, осуществленным в системах PanDA, PALEOMIX и опытом использования GRID-систем.

Данная работа выполнена в рамках мегагранта правительства РФ, контракт № 14.Z50.31.0024, и при частичной финансовой поддержке РФФИ в рамках научного проекта № 16-37-00249. Данная работа была выполнена с использованием высокопроизводительных вычислительных ресурсов федерального центра коллективного пользования в НИЦ «Курчатовский институт». Мы благодарны нашим коллегам из НИЦ «Курчатовский институт» и ЦЕРН за обсуждение результатов работы, особая благодарность группе разработчиков WMS PanDA и группе биоинформатики НИЦ «Курчатовский институт». Также мы выражаем благодарность Миккелю Шуберту (Университет Копенгагена) за руководства по использованию пакета PALEOMIX.

СПИСОК ЛИТЕРАТУРЫ

1. Skryabin K.G., Prokhortchouk E.B., Mazur A.M., Boulygina E.S., Tsygankova S.V., Nedoluzhko A.V., Rastorguev S.M., Matveev V.B., Chekanov N.N., Goranskaya D.A., Teslyuk A.B., Gruzdeva N.M., Velikhov V.E., Zaridze D.G., Kovalchuk M.V. Combining two technologies for full genome sequencing of human. *Acta Nat.* 2009. V. 1. № 3. P. 102–107.
2. Kawalia A., Motameny S., Wonzczak S., Thiele H., Nieroda L., Jabbari K., Borowski S., Sinha V., Gunia W., Lang U., Achter V., Nurnberg P. Leveraging the Power of High Performance Computing for Next Generation Sequencing Data Analysis: Tricks and Twists from a High Throughput Exome Workflow. *PLoS One.* 2015. V. 10. № 5. Article No e0126321.
3. Bao R., Huang L., Andrade J., Tan W., Kibbe W.A., Jiang H., Feng G. Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform.* 2014. V. 13. Suppl. 2. P. 67–82.
4. Miller W., Drautz D.I., Ratan A., Pusey B., Qi J., Lesk A.M., Tomsho L.P., Packard M.D., Zhao F., Sher A., Tikhonov A., Raney B., Patterson N., Linblad-Toh K., Lander E.S., Knight J.R., Irzyk G.P., Fredrikson K.M., Harkins T.T., Sheridan S., Pringle T., Schuster S.C. Sequencing the nuclear genome of the extinct woolly mammoth. *Nature.* 2008. V. 456. P. 387–390.
5. Rasmussen M., Li Y., Lindgreen S., Pedersen J.S., Albrechtsen A., Moltke I., Metspalu M., Metspalu E., Kivisild T., Gupta R., et al. Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature.* 2009. V. 463. P. 757–762.
6. Keller A., Graefen A., Ball M., Matzas M., Boisguerin V., Maixner F., Leidinger P., Backes C., Khairat R., Forster M., et al. New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-genome sequencing. *Nature Communications.* 2011. V. 3.
7. Allentoft M.E., Collins M., Harker D., Haile J., Oskam C.L., Hale M.L., Campos P.F., Samaniego J.A., Gilbert M.T., Willerslev E., et al. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. *Proc Biol Sci.* 2012. V. 279. P. 4724–4733.
8. Nedoluzhko A.V., Boulygina E.S., Sokolov A.S., Tsygankova S.V., Gruzdeva N.M., Rezepkin A.D., Prokhortchouk E.B. Analysis of the Mitochondrial Genome of a Novosvobodnaya Culture Representative using Next-Generation Sequencing and Its Relation to the Funnel Beaker Culture. *Acta Naturae.* 2014. V. 6. P. 31–35.
9. Sokolov A.S., Nedoluzhko A.V., Boulygina E.S., Tsygankova S.V., Gruzdeva N.M., Shishlov A.V., Kolpakova A., Rezepkin A.D., Skryabin K.G., Prokhortchouk E.B. Six

- complete mitochondrial genomes from Early Bronze Age humans in the North Caucasus. *Journal of Archaeological Sciences*. 2016. V. 73. P. 138–144.
10. Martin M.D., Cappellini E., Samaniego J.A., Zepeda M.L., Campos P.F., Seguin-Orlando A., Wales N., Orlando L., Ho S.Y., Dietrich F.S., et al. Reconstructing genome evolution in historic samples of the Irish potato famine pathogen. *Nature Communications*. 2013. V. 4.
 11. Yoshida K., Schuenemann V.J., Cano L.M., Pais M., Mishra B., Sharma R., Lanz C., Martin F.N., Kamoun S., Krause J., et al. The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. *eLife*. 2013. V. 2.
 12. Lorenzen E.D., Nogues-Bravo D., Orlando L., Weinstock J., Binladen J., Marske K.A., Ugan A., Borregaard M.K., Gilbert M.T., Nielsen R., et al. Species-specific responses of Late Quaternary megafauna to climate and humans. *Nature*. 2011. V. 479. P. 359–364.
 13. Lynch V.J., Bedoya-Reina O.C., Ratan A., Sulak M., Drautz-Moses D.I., Perry G.H., Miller W., Schuster S.C. Elephantid Genomes Reveal the Molecular Bases of Woolly Mammoth Adaptations to the Arctic. *Cell Rep*. 2015. V. 12. № 2. P. 217–228.
 14. Massie M., Nothaft F., Hartl C., Kozanitis C., Schumacher A., Joseph A.D., Patterson D.A. ADAM: genomics formats and processing patterns for cloud scale computing. *Report No.: UCB/EECS-2013-207*. Berkeley: EECS Department, University of California, Berkeley. 2013.
 15. Schubert M., Ermini L., Sarkissian C.D., Jonsson H., Ginolhac A., Schaefer R., Martin M.D., Fernandez R., Kircher M., McCue M., Willerslev E., Orlando L. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 2014. V. 9. P. 1056–1082.
 16. The ATLAS Collaboration, G Aad et al., The ATLAS Experiment at the CERN Large Hadron Collider. *Journal of Instrumentation*. 2008. V. 3. <http://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003>
 17. *The Large Hadron Collider*. URL: <http://home.cern/topics/large-hadron-collider> (дата обращения: 28.09.2016).
 18. Maeno T. on behalf of PANDA team and ATLAS collaboration. PanDA: distributed production and distributed analysis system for ATLAS. *Journal of Physics: Conference Series*. 2008. V. 119. № 6.
 19. Maeno T., De K., Klimentov A., Nilsson P., Oleynik D., Panitkin S., Petrosyan A., Schovancova J., Vaniachine A., Wenaus T. Evolution of the ATLAS PanDA workload management system for exascale computational science. *J. Phys.: Conf. Ser.* 2014. V. 513.
 20. *Worldwide LHC Computing Grid*. URL: <http://wlcg.web.cern.ch/> (дата обращения: 28.09.2016).

Рукопись поступила в редакцию 15.07.2016.

Дата опубликования 21.10.2016.