

УДК 004.75

BioInfoWF – СИСТЕМА АВТОМАТИЧЕСКОЙ ГЕНЕРАЦИИ WEB-ИНТЕРФЕЙСОВ И WEB-СЕРВИСОВ ДЛЯ БИОИНФОРМАЦИОННЫХ ИССЛЕДОВАНИЙ

© 2012 г. М.А. Генаев¹, Е.Г. Комышев², К.В. Гунбин¹, Д.А. Афонников^{1,2}

¹ Учреждение Российской академии наук Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия, e-mail: mag@bionet.nsc.ru;

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия

Поступила в редакцию 5 июля 2012 г. Принята к публикации 25 июля 2012 г.

В настоящей работе представлена система BioInfoWF для автоматической генерации Web-интерфейсов и Web-сервисов для вычислительных модулей в области биоинформатики. Для каждого вычислительного модуля, используемого в этой системе, вводится метаописание на языке XML. На основе метаописаний BioInfoWF автоматически генерируют Web-интерфейсы и Web-сервисы, которые в дальнейшем могут использоваться как в различных биоинформационных системах, так и непосредственно в самой системе BioInfoWF. Вычислительные модули в нашей системе могут объединяться в конвейеры, для которых автоматически генерируется пользовательский Web-интерфейс. Разработанный нами инструмент существенно упрощает разработку и публикацию модулей анализа биоинформатических данных в сети, что обеспечивает их доступность для сообществ биологов и биоинформатиков. Система BioInfoWF распространяется под свободной лицензией GNU GPL. Дистрибутив и пользовательская документация системы BioInfoWF доступны на сайте <http://bioinfowf.bionet.nsc.ru>.

Ключевые слова: биоинформатика, интеграция данных, конвейерная обработка данных, метаописание данных, Web-интерфейс, Web-сервис.

ВВЕДЕНИЕ

Методы биоинформатики все шире применяются в самых различных областях биологии: при анализе данных высокопроизводительного секвенирования (Pop, Salzberg, 2008), в биомедицине (Fernald *et al.*, 2011), генетике (Moore *et al.*, 2010), изучении молекулярной эволюции (Sánchez *et al.*, 2011), при анализе фенотипических признаков организмов (Hartmann *et al.*, 2011). Для решения конкретной биологической задачи рутинными процедурами являются обращение к базам данных, использование большого числа программ, объединенных в цепочки, визуализация полученных результатов на разных этапах обработки данных. Успех решения зависит от эффективной обработки огромного количества информации, представленной в разных форматах; применения большого числа

математических моделей, которые реализованы разными вычислительными программами и алгоритмами. Все это обуславливает важность использования в биоинформатике конвейерных систем организации обработки данных (Deelman *et al.*, 2009).

При разработке подобных систем актуальной задачей является обеспечение возможности повторного использования разработанных конвейеров, их доступности другим пользователям, в том числе и в виде отдельных модулей. Такие возможности предоставляет Интернет, в частности, организация работы конвейеров на основе Web-сервисов. Web-сервис – это программная система, идентифицируемая строкой универсального идентификатора ресурса (universal resource identifier, URI), его общедоступные интерфейсы определены на языке XML и основаны на базе открытых стандартов

и протоколов. Web-сервис является единицей модульности при использовании сервис-ориентированной архитектуры приложения и обеспечивает взаимодействие программных систем независимо от платформы.

При большом разнообразии имеющихся Web-сервисов встает задача их структуризации. В настоящее время разработано несколько популярных систем каталогизации и конструирования конвейеров в области биоинформатики, которые используют Web-сервисы. Проект BioCatalogue (Bhagat *et al.*, 2010) является реестром биологических Web-сервисов. BioCatalogue предоставляет общий интерфейс для регистрации, просмотра и аннотирования Web-служб в сообществе наук о жизни. При аннотации сервисов учитываются их технический тип, биоинформатическая категория, поставщик этого сервиса, различные пользовательские теги, форматы входных/выходных данных. Сервисы BioCatalogue также являются предметом постоянного мониторинга, позволяющего идентифицировать проблемы с обслуживанием и изменением сервиса. Это позволяет осуществлять фильтрацию недоступных или ненадежных ресурсов. Сервисы в BioCatalogue доступны не только в виде программного интерфейса, но и с помощью Web-интерфейса на основе технологий «Web 2.0» (Bhagat *et al.*, 2010).

При решении конкретной задачи в области биоинформатики, как правило, используется набор методов, каждый из которых реализован в виде отдельного приложения или Web-сервиса. В настоящее время разработано несколько популярных систем в области биоинформатики, позволяющих организовывать такие методы в конвейеры. Схема конвейера в этом случае является планом эксперимента для решения конкретной биологической задачи. Одной из популярных систем, предоставляющих конвейерную обработку данных, является Galaxy (Goecks *et al.*, 2010). Пользователи, не имеющие опыта программирования, могут легко задать параметры, запуская вычислительные модули и конвейеры. При проведении вычислений Galaxy фиксирует промежуточную информацию, для того чтобы любой пользователь мог повторить и подробно проанализировать результаты численного анализа. В системе Galaxy пользователи могут совместно использовать и публиковать конвейеры через

Web и создавать Web-документы, описывающие протоколы и смысл исследований.

Другой популярной системой подобного типа является Taverna (Oinn *et al.*, 2004). Это графическая среда для управления и запуска конвейеров, реализованная на языке Java в виде приложения с графическим интерфейсом (GUI). Taverna обеспечивает удобство интеграции программ, баз данных и Web-сервисов, доступных в Интернет. Это позволяет биоинформатикам конструировать конвейеры для решения задач в различных областях, таких, как секвенирование и аннотация геномов. Распределенная Web-сервисная архитектура системы позволяет использовать много различных провайдеров ресурсов для проведения анализа. Схему готового конвейера можно сохранить, отредактировать и запустить повторно, причем не только через GUI интерфейс, но и как консольное приложение.

Популярность конвейерной обработки данных и накопление огромного числа различных готовых конвейеров, каждый из которых нацелен на решение узкой биологической задачи, привели к появлению таких ресурсов, как myExperiment (Goble *et al.*, 2010). myExperiment – среда для совместной работы, где ученые могут публиковать свои планы эксперимента, конвейеры, делиться ими с другими пользователями. Конвейеры, объекты (программы и данные из баз), их связи (так называемые пакеты) можно искать, сортировать и классифицировать как фотографии и видео в Интернете. Однако в отличие от Facebook или MySpace myExperiment полностью ориентирован на потребности исследователей. myExperiment обеспечивает удобный доступ к знаниям (информации, научным данным) для следующего поколения ученых, предоставляя пул научных методов, формируя сообщества и определяя протокол общения внутри этого сообщества. Все это приводит к уменьшению времени выхода на биологический эксперимент.

Широта решаемых задач и популярность работы в системах, подобных Taverna и Galaxy, напрямую зависят от доступности Web-сервисов и разнообразия задач, которые они способны выполнять. В настоящей работе мы представляем систему BioInfoWF для автоматической генерации Web-сервисов и Web-интерфейсов. Предложено формальное метаописание вычислительных модулей на XML. На основе таких

метаописаний система BioInfoWF автоматически генерирует Web-интерфейсы и Web-сервисы, которые в дальнейшем могут использоваться как в различных биоинформационных системах (Biocatalogue, Taverna, myExperiment), так и непосредственно в самой системе BioInfoWF. Вычислительные модули в нашей системе могут объединяться в конвейеры, для которых автоматически генерируется пользовательский Web-интерфейс. Разработанный нами инструмент существенно упрощает разработку и публикацию модулей анализа биоинформатических данных в сети и их доступность в сообществах биоинформатиков и биологов, таких, как Biocatalogue, myExperiment и др.

Таким образом, BioInfoWF позволяет решать задачи конвейерной обработки данных, обеспечивает генерацию удобного пользовательского интерфейса «Web 2.0» для запуска вычислительных модулей и конвейеров, а также дает возможность автоматически создавать Web-сервисы.

МАТЕРИАЛЫ И МЕТОДЫ

Архитектура системы

Предлагаемая система включает (рис. 1):

1) менеджер запуска вычислительных модулей и готовых конвейеров BioInfoWF;

2) репозитории, в которых хранятся: а) описания вычислительных модулей и схем конвейеров, которые являются наборами XML файлов; б) вычислительные модули, базы данных и программы визуализации данных; в) пользовательские данные;

3) системы генерации Web-интерфейсов и Web-сервисов.

Описание конвейера и его исполнение

В системе BioInfoWF конвейер представляет собой набор вычислительных модулей, связанных топологией выполнения задачи. Модули представляют собой программы, запускаемые в консольном режиме в среде Linux. Управляющие параметры (названия входных и выходных файлов, параметры алгоритмов) передаются в расчетные модули через командную строку или переменные окружения. В ходе выполнения задачи выходные данные одного модуля могут подаваться на вход другому модулю.

Разработанная нами система позволяет интегрировать любые вычислительные модули, организованные подобным образом. При этом данные могут находиться как на локальной машине, так и на удаленной. Схема интеграции (порядок выполнения процедур) и метаописания вычислительных модулей определяются на языке XML.



Рис. 1. Основные структурные элементы системы BioInfoWF.

Серверная часть, реализованная на языке Perl, выполняет запуск конвейера и отслеживает статус выполнения каждого вычислительного модуля. На вход приложению подается описание схемы конвейера и вычислительных модулей. Приложение запускает конвейер, создавая файл с отчетом. В отчете указывается статус выполнения каждого узла в конвейере. Серверная часть поддерживает параллельный запуск узлов конвейера, при этом максимальное количество потоков определяется в конфигурационном файле приложения.

Язык описания вычислительных модулей схем конвейеров

Для формального описания схемы конвейера и вычислительных модулей, из которых он состоит, нами был разработан язык на основе XML. Описание конвейера представлено двумя файлами. Первый описывает вычислительные модули, второй задает топологию конвейера (рис. 2).

Описания вычислительных модулей содержатся в хранилище и состоят из следующих разделов:

– *Входные файлы.* Описание того, какие входные файлы подаются на вход модулю; для каждого файла указывается его идентификатор, даются описание, формат файла.

– *Выходные файлы.* Описание выходных файлов вычислительного модуля.

– *Параметры и опции.* Описание параметров и опций для вычислительного модуля. Для каж-

дого параметра задаются идентификатор, описание, тип параметра (например, строка, число или бинарное значение), значение по умолчанию, внешний вид поля запроса значения для параметра на странице WWW-браузера.

– *Правила генерации командной строки.* Описание набора входных параметров вычислительного модуля и списка входных и выходных файлов. На основе этой информации менеджер запуска генерирует командную строку для запуска вычислительного модуля.

– *Правила поведения пользовательского интерфейса.* Опциональный раздел, который указывает, каким образом Web-интерфейс будет динамически реагировать на действия пользователей.

Пример описания в формате XML команды kill, которая прекращает выполнение какого-либо вычислительного процесса, приведен на рис. 3. Программа на вход принимает идентификационный номер процесса, который надо завершить. Вторая строчка файла (рис. 3) описывает название вычислительного модуля и путь, где располагается исполняемый файл модуля. В нашем случае kill – это команда окружения bash, поэтому указание полного пути не требуется. Секция output (строки 7–14) описывает выходные файлы, в этом примере описываются два файла с идентификаторами *stdout* и *stderr*, которые мы в дальнейшем проассоциируем со стандартными потоками вывода 1 и 2 соответственно. Аналогичным образом описывается секция input для входных файлов. В нашем примере входных файлов нет,

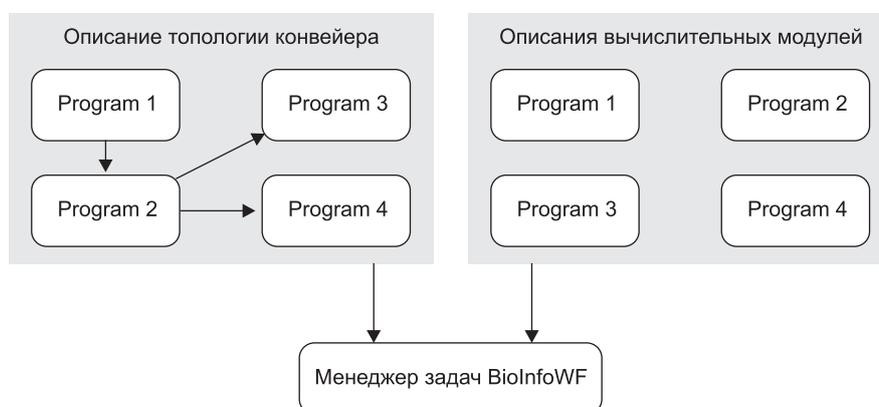


Рис. 2. Описание любого конвейера включает описание вычислительных модулей, вовлеченных в конвейер, и связей между этими модулями.

поэтому секция отсутствует. В секции `options` описывается единственная опция, которая будет передавать идентификатор `id` процесса команде `kill`. Опция имеет тип `int` и представление `text`, которое будет соответствовать `<input type=«text» />` при генерации Web-интерфейса. Значение по умолчанию для опции не задано. Секция `cmdline` описывает правила генерации командной строки. На входе мы имеем две хеш-таблицы, `$options` и `$files`. Ключами в этих хеш-таблицах служат `id` из секций `input`, `output` и `cmdline`. На выходе необходимо сформировать переменную `$cmd`, которая содержала бы готовую командную строку для вычислительного модуля.

Генерация пользовательского Web-интерфейса

Пользовательский интерфейс генерируется автоматически на основе XML описаний вычислительных модулей и схемы конвейера. Пользователю предлагается работать с уже готовыми схемами. В текущей версии клиентской части возможна работа только с последовательными конвейерами.

Схема генерации Web-страниц представлена на рис. 4. Менеджер запуска задач BioInfoWF получает на вход описание схемы конвейера и

вычислительных модулей в формате XML и с использованием библиотеки `Perl HTML::Template` генерирует файл в формате `html`. Автоматизация достигается за счет того, что в описании каждого входного параметра указывается тип элемента HTML для его визуального представления (выпадающее меню, радиокнопка, текстовое поле и т. п.). Пользователь может управлять поведением конвейера, изменяя его схему и входные данные для вычислительных модулей. Для визуализации и редактирования входных и выходных данных имеются возможности подключения внешних программ, реализованных, как правило, в виде `Java Applet` приложений.

Реакция интерфейса на действие пользователя при его работе с HTML-страницами достигается за счет внедрения в описание каждого модуля правил поведения, реализованных с помощью библиотеки `jQuery`. Динамическое изменение интерфейса удобно использовать, когда существуют зависимости между параметрами исполняемых модулей. Например, в модуле, описывающем задачу множественного выравнивания белковых последовательностей программой `Mafft`, выбор матрицы сравнения аминокислот обуславливает ряд дополнительных опций, которые зависят от ее типа. При выборе матрицы на Web-странице отображаются только опции, связанные типом выбранной матрицы.

```

1 <programs>
2 <program name="Kill" exe="kill" cluster="off">
3 <description>
4   kill - terminate a process
5 </description>
6
7 <output>
8 <file id="stdout" type="text" name="STDOUT" description="Standard output" />
9 <file id="stderr" type="text" name="STDERR" description="Standard error" />
10 </output>
11
12 <options>
13 <option id="PID" name="PID" description="PID" view="text" type="int" default="" />
14 </options>
15
16 <cmdline>
17   $cmd = " $options{PID} "
18   "1>\"$files{stdout}" ".
19   "2>\"$files{stderr}" ";
20 </cmdline>
21
22 </program>
23 </programs>

```

Рис. 3. Пример XML описания вычислительного модуля.

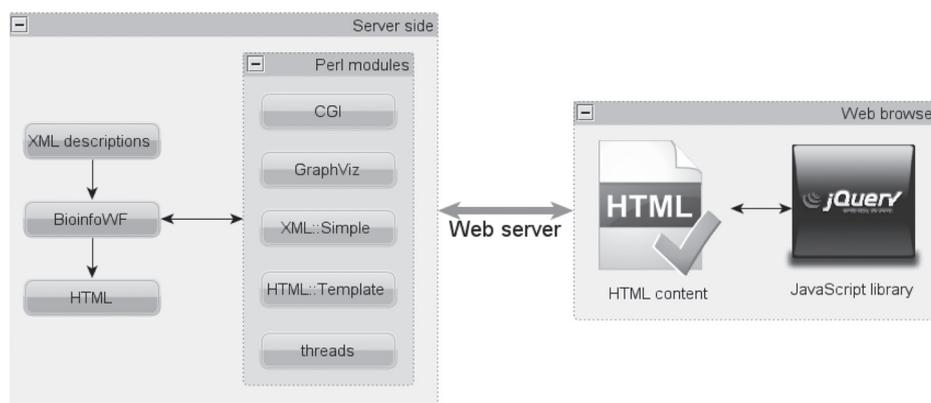


Рис. 4. Схема процесса генерации Web-интерфейса в системе BioInfoWF.

Предложенный подход позволяет автоматически создавать Web-интерфейсы для любых вычислительных модулей, описанных в системе.

В текущей версии интерфейса доступны следующие базовые опции управления конвейером.

- Установка входных файлов, параметров и опций для каждого вычислительного модуля в конвейере.

- Старт с произвольного узла и остановка на произвольном узле в конвейере.

- Отслеживание статуса выполнения каждого вычислительного модуля в конвейере.

- Просмотр входных/выходных файлов для каждого этапа расчета в конвейере.

Привязка форматов входных/выходных файлов к различным приложениям для их визуализации.

Генерация Web-сервисов

Генерация Web-сервисов производится автоматически приложением Java, модулем генерации Web-сервисов (рис. 5). Интерфейс генерируется автоматически на основе метаописания вычислительных модулей, содержащего всю необходимую для этого информацию. Для генерации необходимо запустить модуль WebServicesGen в консольном режиме, передав в параметрах файл с метаописанием вычислительных модулей и название вычислительного модуля, после чего Web-сервис будет сгенерирован и размещен на сервере. Таким образом, в системе BioInfoWF реализована быстрая автоматическая генерация Web-сервисов для новых вычислительных модулей.

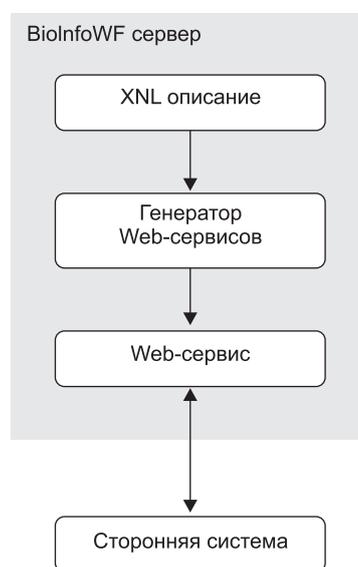


Рис. 5. Схема генерации Web-сервисов и взаимодействия их со сторонней системой.

Задача сгенерированных Web-сервисов – предоставление доступа к вычислительным модулям системы BioInfoWF. При этом модуль генерации Web-сервисов производит Java-сервлет, аналог клиентского апплета, исполняющегося на сервере. На каждый вычислительный модуль генерируется отдельный сервлет, который в свою очередь реализует как REST, так и WSDL/SOAP Web-сервис как синхронного, так и асинхронного типа на основе протокола HTTP.

REST – подход к архитектуре построения Web-приложений – позволяет использовать простые HTTP запросы, содержащие параметры в заголовках, а передаваемые данные – в теле HTTP запроса. Таким образом, простой REST

Web-сервис возможно использовать в системах, не поддерживающих более сложный SOAP протокол. Для использования REST Web-сервиса вычислительного модуля необходимо послать HTTP запрос PUT на адрес *http://BioInfoWF.server/ServiceName/* (где *BioInfoWF.server* – адрес сервера BioInfoWF в сети, а *ServiceName* – название соответствующего вычислительного модуля системы BioInfoWF), в заголовке которого содержатся параметры вычислительного модуля, а в теле – данные. Ответ на этот запрос будет содержать посчитанные вычислительным модулем данные. Для получения краткой справочной информации о вычислительном модуле необходимо послать GET запрос на тот же адрес.

Язык описания Web-сервисов WSDL вместе с протоколом обмена структурированными сообщениями в распределенной вычислительной среде SOAP позволяют выполнять более гибкие запросы и предоставлять формальную информацию о Web-сервисе. Системе, в которой реализована поддержка SOAP/WSDL Web-сервисов, достаточно предоставить WSDL описание, указав URI, по которому оно находится, после чего эта система может взаимодействовать с Web-сервисом.

Асинхронный тип Web-сервиса полезен при вычислениях, требующих временных затрат, когда поддержание постоянного HTTP соединения на запрос–ответ проблематично. В этом случае для загрузки входных данных при инициализации вычисления, их выгрузке и проверке состояния вычисления используются отдельные HTTP запросы и ответы.

Java-сервлет, реализующий эти Web-сервисы совместно, упаковывается в war-файл (jar архив для сервлетов), который может быть развернут на любом контейнере сервлетов. Для контейнера сервлетов Catalina пакета Tomcat достаточно поместить этот файл в папку Webapps/.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Система BioInfoWF распространяется под свободной лицензией GNU GPL. Дистрибутив и пользовательская документация системы доступны на официальном сайте <http://bioinfowf.bionet.nsc.ru>. Приведем пример использования

системы BioInfoWF для решения задач молекулярной эволюции генов и белков.

Решение этой задачи обычно заключается в последовательном выполнении набора операций с нуклеотидными или аминокислотными последовательностями (извлечение гомологов из банка данных, выравнивание, реконструкция филогении, оценка режима эволюции). Ранее были созданы несколько вариантов специализированных конвейерных систем для решения задач молекулярной эволюции и филогении (Dereeper *et al.*, 2008; Sбnchez *et al.*, 2011). При помощи системы BioInfoWF нами был создан пакет анализа молекулярной эволюции SAMEM (Гунбин и др., 2011; Gunbin *et al.*, 2012). SAMEM состоит из двух основных конвейеров, анализа эволюции генов (I) и анализа эволюции белков (II) и двух дополнительных конвейеров, собирающих выборки генов и белков (III) и производящих их первичный анализ (IV). Уникальной особенностью конвейера I является возможность исследования данных ранее предложенными методами оценки K_R/K_C (Zhang, 2000; Smith, 2003), а также оригинальным, впервые предложенным, методом оценки K_R/K_C (Gunbin *et al.*, 2012). При анализе данных конвейером II существенной особенностью является оригинальный метод анализа скоростей фиксации различных типов аминокислотных замен на ветвях филогенетического дерева (Гунбин и др., 2011; Gunbin *et al.*, 2012). Метод основан на Марковском моделировании эволюции (Pupko *et al.*, 2002) и непараметрическом перестановочном тесте (Gunbin *et al.*, 2011), осуществляющем сравнение числа ожидаемых и наблюдаемых аминокислотных замен, и позволяет анализировать режимы молекулярной эволюции на глубоких ветвях филогенетического дерева. Подход, реализованный в системе SAMEM, также позволяет 1) использовать все известные свойства аминокислот и 2) проводить статистическое соотнесение изменения этих свойств с признаками фенотипа.

Вычислительные модули пакета SAMEM могут быть доступны как в качестве пользовательского Web-интерфейса, так и в качестве Web-сервисов. Примеры Web-интерфейса SAMEM и использование модулей SAMEM в виде Web-сервисов в системе Taverna приведены на рис. 6.

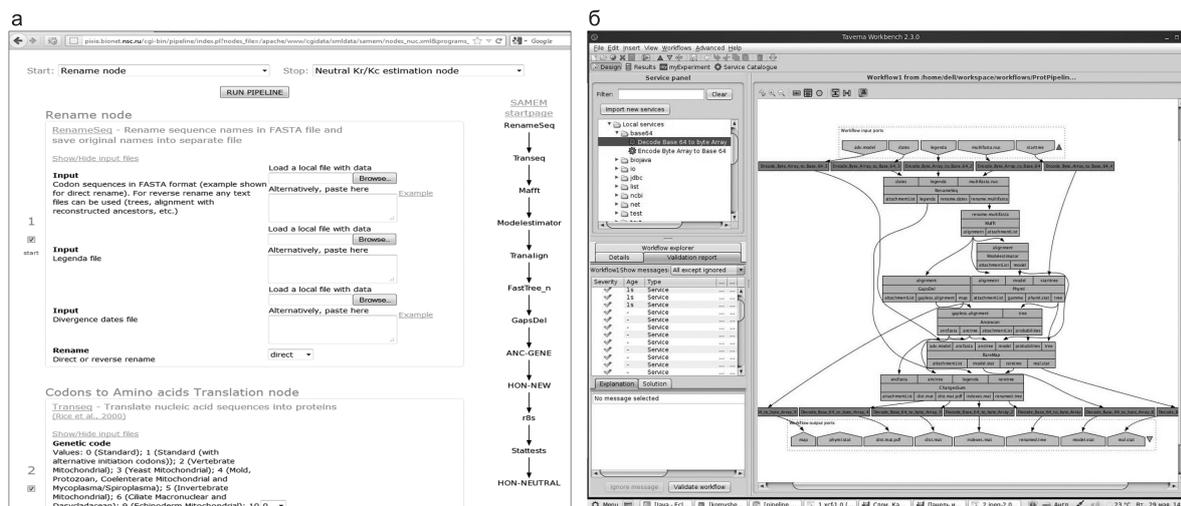


Рис. 6. Интерфейс пользователя системы SAMEM, реализованной на платформе BioInfoWF.

а – реализация конвейера в виде Web-интерфейса; б – реализация этого конвейера в системе Taverna с использованием Web-сервисов.

ЗАКЛЮЧЕНИЕ

В настоящей работе предложена система BioInfoWF, которая позволяет генерировать Web-сервисы для решения биоинформатических задач, организовывать их работу в виде конвейера, автоматически генерировать пользовательский Web-интерфейс. Удобство системы заключается в быстром подключении вычислительных модулей на основе исполняемых приложений и Web-сервисов. Использование системы было продемонстрировано на примере создания конвейеров решения задач по анализу молекулярной эволюции генов и белков – SAMEM.

Работа поддержана грантом Министерства образования и науки РФ в рамках ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007–2013 гг.» (07.514.11.4023).

ЛИТЕРАТУРА

Гунбин К.В., Генаев М.А., Турнаев И.И., Афонников Д.А. Компьютерная система анализа режимов молекулярной эволюции генов и белков: анализ эволюции циклинов В // Вестн. Томского гос. ун-та. Биология. 2011. 4. С. 175–189.

- Bhagat J., Tanoh F., Nzuobontane E. *et al.* BioCatalogue: a universal catalogue of Web services for the life sciences // *Nucl. Acids Res.* 2010. V. 38. P. 689–694.
- Deelman E., Gannon D., Shields M., Taylor I. Workflows and e-Science: An overview of workflow system features and capabilities // *Future Generation Computer Systems.* 2009. V. 25. P. 528–540.
- Dereeper A., Guignon V., Blanc G. *et al.* Phylogeny.fr: robust phylogenetic analysis for the non-specialist // *Nucl. Acids Res.* 2008. V. 36. P. 465–469.
- Fernald G.H., Capriotti E., Daneshjou R. *et al.* Bioinformatics challenges for personalized medicine // *Bioinformatics.* 2011. V. 27. P. 1741–1748.
- Goble C.A., Bhagat J., Alekseyevs S. *et al.* myExperiment: a repository and social network for the sharing of bioinformatics workflows // *Nucl. Acids Res.* 2010. V. 38. P. 677–682.
- Goecks J., Nekrutenko A., Taylor J., Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences // *Genome Biol.* 2010. 11:R86.
- Gunbin K.V., Suslov V.V., Genaev M.A., Afonnikov D.A. Computer system for analysis of molecular evolution modes (SAMEM): analysis of molecular evolution modes at deep inner branches of the phylogenetic tree // *In Silico Biol.* 2012. In press.
- Gunbin K.V., Suslov V.V., Turnaev I.I. *et al.* Molecular evolution of cyclin proteins in animals and fungi // *BMC Evol. Biol.* 2011. V. 11. P. 224.
- Hartmann A., Czauderna T., Hoffmann R. *et al.* HTPPheno: an image analysis pipeline for high-throughput plant phenotyping // *BMC Bioinformatics.* 2011. V. 12. P. 148.
- Moore J.H., Asselbergs F.W., Williams S.M. Bioinformatics challenges for genome-wide association studies //

- Bioinformatics. 2010. V. 26. P. 445–455.
- Oinn T., Addis M., Ferris J. *et al.* Taverna: a tool for the composition and enactment of bioinformatics workflows // Bioinformatics. 2004. V. 20. P. 3045–3054.
- Pop M., Salzberg S.L. Bioinformatics challenges of new sequencing technology // Trends Genet. 2008. V. 24. P. 142–149.
- Pupko T., Pe'er I., Hasegawa M. *et al.* A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families // Bioinformatics. 2002. V. 18. P. 1116–1123.
- Sánchez R., Serra F., Tárraga J. *et al.* Phylemon 2.0: a suite of Web-tools for molecular evolution, phylogenetics, phylogenomics and hypotheses testing // Nucl. Acids Res. 2011. V. 39. P. 470–474.
- Smith N.G. Are radical and conservative substitution rates useful statistics in molecular evolution? // J. Mol. Evol. 2003. V. 57. P. 467–478.
- Zhang J. Rates of conservative and radical nonsynonymous nucleotide substitutions in mammalian nuclear genes // J. Mol. Evol. 2000. V. 50. No. 1. P. 56–68.

BioInfoWF – WEB SERVICES AND WEB INTERFACES GENERATOR FOR BIOINFORMATICS ANALYSIS

M.A. Genaev¹, E.G. Komyshev², K.V. Gunbin¹, D.A. Afonnikov^{1,2}

¹ Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia,
e-mail: mag@bionet.nsc.ru;

² Novosibirsk National Research State University, Novosibirsk, Russia

Summary

The BioInfoWF (Bioinformatics WorkFlow) system for automated generation of Web interface and Web services for bioinformatics programs. Each program module used in the system has metadescription in XML. The metadescriptions are used for automated generation of Web interface and Web services that can be used further in bioinformatics workflows. Computational modules can be organized in workflows. The tool we have developed significantly simplify the design and publication of modules for bioinformatics data analysis via the internet and their availability for scientific communities. The developed system makes is distributed under GNU GPL. The Source codes and documentation for BioInfoWF are available at <http://bioinfowf.bionet.nsc.ru>.

Key words: bioinformatics, data integration, workflow data processing, data metadescription, Web interface, Web service.