**Результаты**:

1. Выполнен сравнительный обзор наиболее популярных современных платформ конвейерной обработки данных, применяемые в научных исследованиях с интенсивным использованием данных, в том числе, RAPIDMINER, KNIME, PIPLINE PILOT, TAVERNA. Сравнение произведено по таким характеристикам, как модульность, требования к наличию навыков программирования у исследователя, открытость исходного кода, интеграция с языком программирования R, случаи применения в геномике и метагеномике (исп. Паскал).
2. Выполнен аналитический обзор программного обеспечения для анализа метагеномных данных. Составлен список программных решений аналогичных по функциональности применяемому в настоящее время в ЛИН СО РАН комплексу MOTHUR с описанием форм применяемого лицензирования, интерфейсов прикладного программирования для реализации собственных модулей, платформ исполнения (исп. Красовицкая).
3. Реализован прототип расширения “NEW GENERATION SEQUENCING” для платформы конвейерной обработки данных RAPIDMINER, включающее в себя интеграцию с программным обеспечением MOTHUR. В расширение включены следующие модули: TRIM, SUMMARY, UNIQUE, ALIGN, SCREEN и др. Реализована семантическая обработка файлов журналирования программы MOTHUR для синтеза модулей, расширяющих платформу RAPIDMINER. Расширение RAPIDMINER позволяет построить наглядную информационную модель обработки метагеномных данных с использованием реализованных модулей (исп. Паскал, Красовицкая, Черкашин).
4. Спроектирована схема хранения данных, получаемых при исследовании микробиома озера Байкал. Схема охватывает информацию о сборе проб, анализе физико-химических и биологических параметров этих проб, результатов секвенирования, применяемом оборудовании, таксономических базах, методиках анализа собранных материалов, публикациях результатов исследований, а также исследователях, принимающих участие в получении результатов. Также она позволяет хранить используемые вычислительные цепочки обработки и анализа метагеномных данных, включая информацию о программных инструментах, командах и конфигурационных файлов (исп. Малков, Шигаров).
5. Разработан прототип модели веб-ориентированной информационной системы (ИС) управления данными результатов исследования микробиома озера Байкал на протяжение их жизненного цикла. Реализована опытная версия ИС на основе инструментальной платформы DJANGO и системы управления данными POSTGRESQL. Реализован пользовательский веб-интерфейс ИС доступа к данным. В настоящее время ИС развернута в облачной вычислительной инфраструктуре ИДСТУ СО РАН. Начат опытный ввод имеющихся данных ЛИН СО РАН по исследованию микробиома озера Байкал (исп. Малков, Шигаров).

**Публикации**

1. Малков Ф.С., Галачьянц Ю.П., Шигаров А.О., Морозов А.А., Михайлов И.С., Ломакина А.В., Захаренко А.С. Управление данными в исследовании микробиома оз. Байкал // Тезисы докладов 18-й Всеросс. конф. молодых ученых по математическому моделированию и информационным технологиям. Иркутск, 2017. c. 82.
2. Черкашин Е.А.. Шигаров А.О., Орлова И.В., Михайлов И.С. Использование технологий Linked Open Data при публикации текстовых документов // Материалы Всеросс. конф. «Знания-Онтологии-Теории». Новосибирск, 2017. Т. 2, с. 138-147.
3. Cherkashin E., Shigarov A., Orlova I., Mikhailov I. Authoring and Publishing Text Documents by means of Linked Open Data Technologies // In Procs. Int. Conf. on Applied Internet and Information Technologies. Zrenjanin, Serbia, 2017. pp. 98-109.
4. Cherkashin E., Shigarov A., Malkov F., Pascal K., Morozov A. An Environment for Metagenomic Analysis // In Proc. Int. Conf. on Applied Internet and Information Technologies. Zrenjanin, Serbia, 2017. pp. 110-117.
5. Ожидается статья в Известиях ИГУ (Морозов А. и др.)

**Рисунки**

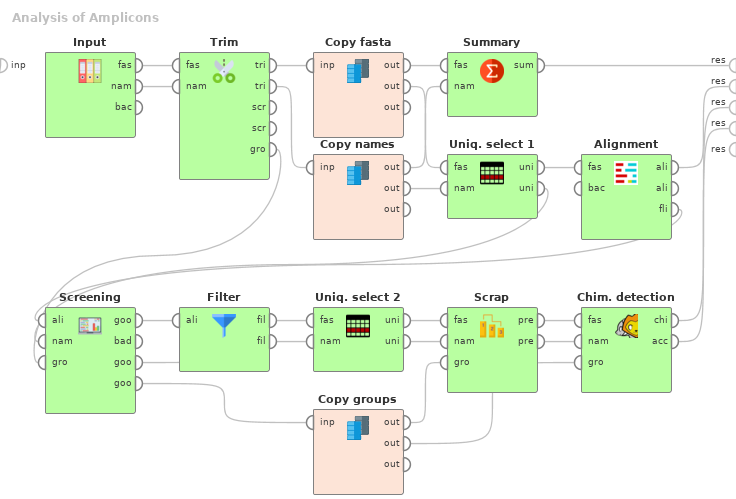


Рис. 1. Построение вычислительных цепочек обработки и анализа метагеномных данных с помощью расширения “NEW GENERATION SEQUENCING” для платформы конвейерной обработки данных RAPIDMINER.



Рис. 2. Упрощенное представление схемы данных результатов исследования микробиома, ребра, подписанные (1..n), соответствуют отношениям типа один-ко-многим, подпись (n..m) показывает отношения типа многие-ко-многим.

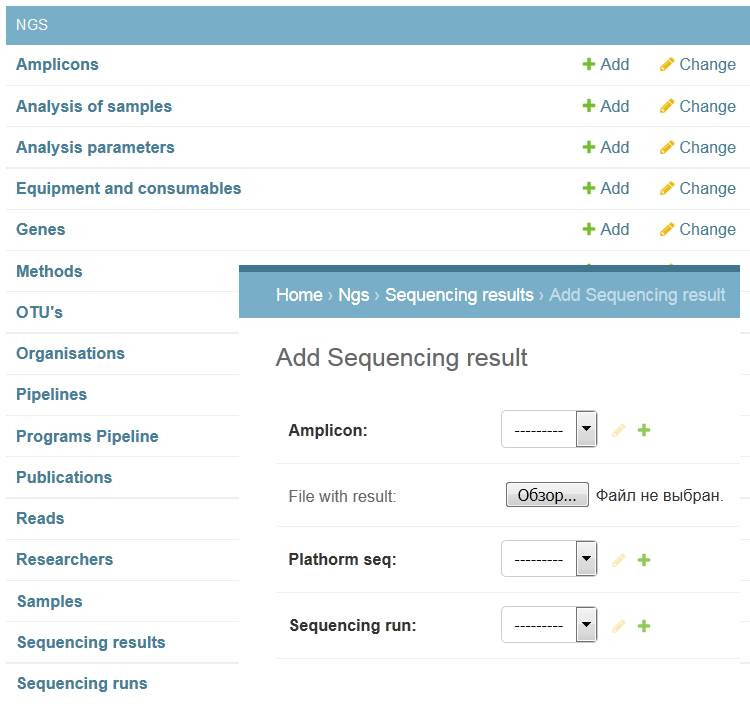


Рис. 3. Снимок экрана пользовательского веб-интерфейса доступа к данным в опытной версии информационной системы управления данными результатов исследования микробиома озера Байкал.