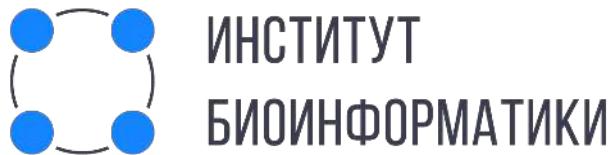


Автоматизация тестирования инструментов множественного выравнивания последовательностей в среде KNIME



Руководитель:
Яснев Олег (KNIME.com)

Студенты:
Захарова Александра
Пантелеева Александра

MSA

последовательное выравнивание:

- рафинирующее(время)
- консистентное(качество)

Крупномасштабное выравнивание:

- глубокое выравнивание
- длинное выравнивание

Как оценить MSA?

SP score - sum of pairs score (матрицы BLOSUM и PAM)

$$SP = \sum_i^C \sum_{j,k,j \neq k}^N \text{score}(a_{j,i}, b_{k,i})$$

C – число столбцов, N – число последовательностей,
score(a, b) – оценка замены пары(a, b)

Программы для MSA и бенчмарки для их тестирования

Clustal O
MUSCLE
T-Coffee
PRANK
MAFFT
Clustal W

BaliBASE
BaliFam
OxBench
HOMSTRAD

Бенчмарки для тестирования

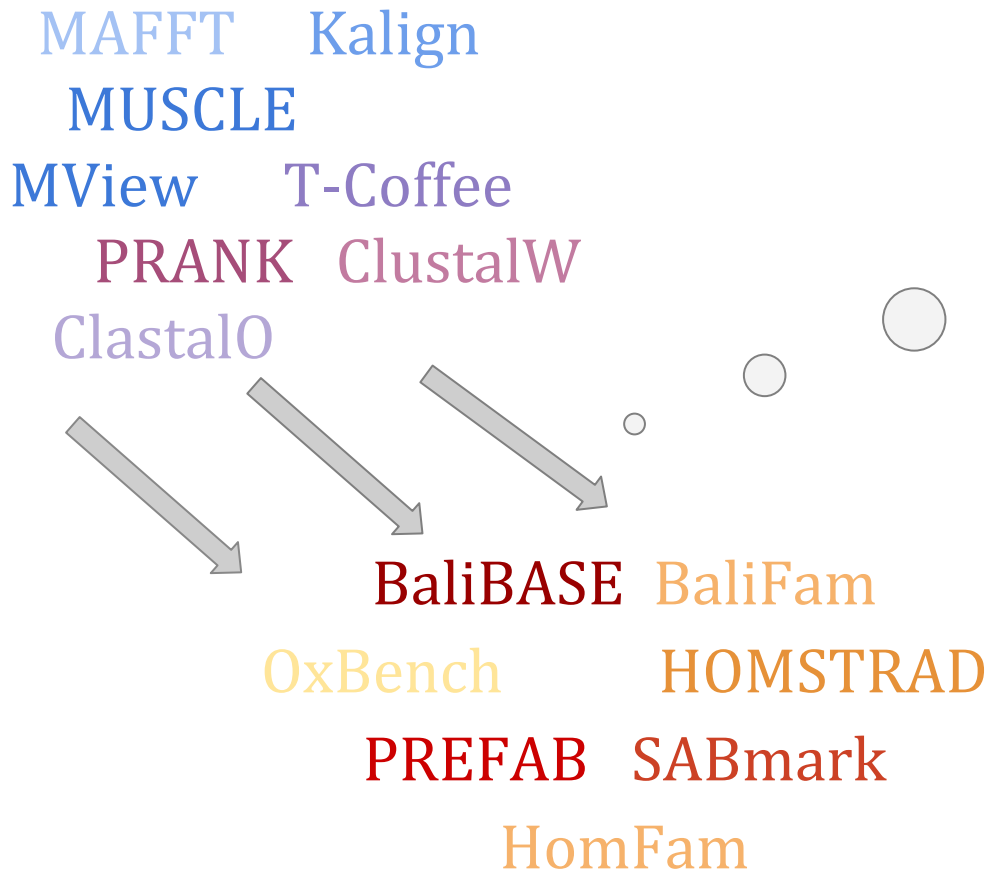
BAlIiBASE

- RV11, RV12 - равноудаленные последовательности примерно одинаковой длины.
- RV20 - RV20 - семейства белков + до трех сиротских последовательностей
- RV30 - до четырех белковых подсемейств на каждый тест, которые должны быть объединены в единое выравнивание.
- RV40 - длинные расширения с левого и правого конца отдельных последовательностей.
- RV50 - тесты выравниваний с последовательностями с большими внутренними вставкам

SP score – это доля корректно выровненных пар аминокислот.

TC score – (total column score) это оценка всего столбца.

Тестирование



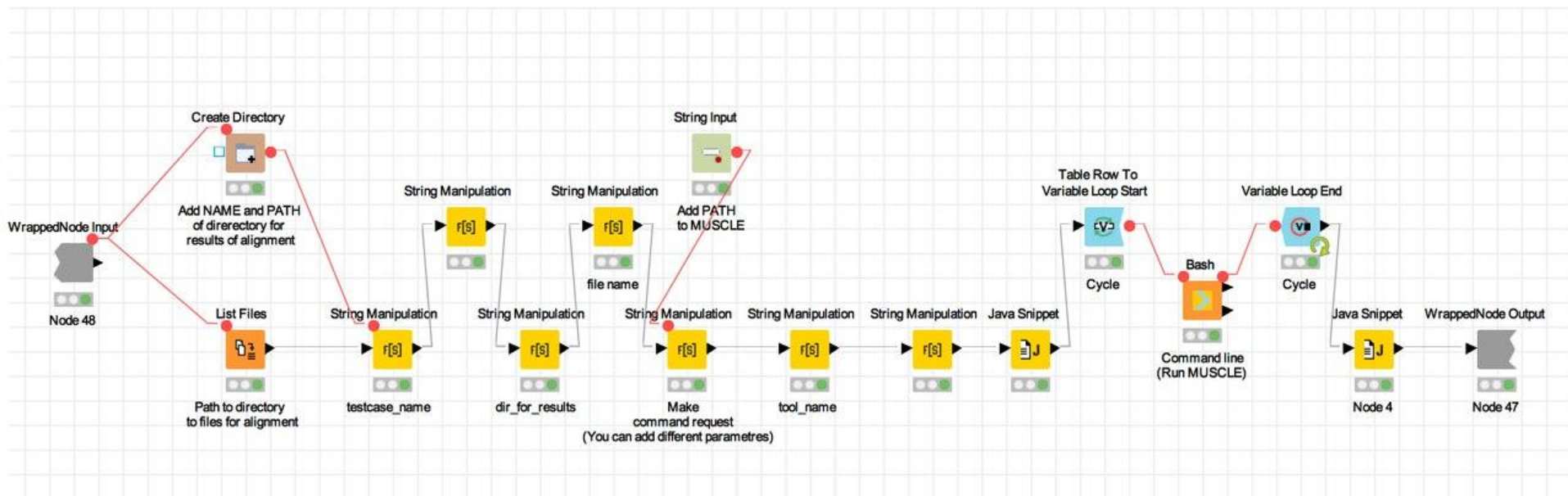
Цель

Автоматизировать работу тестирования программ для множественного выравнивания в среде KNIME

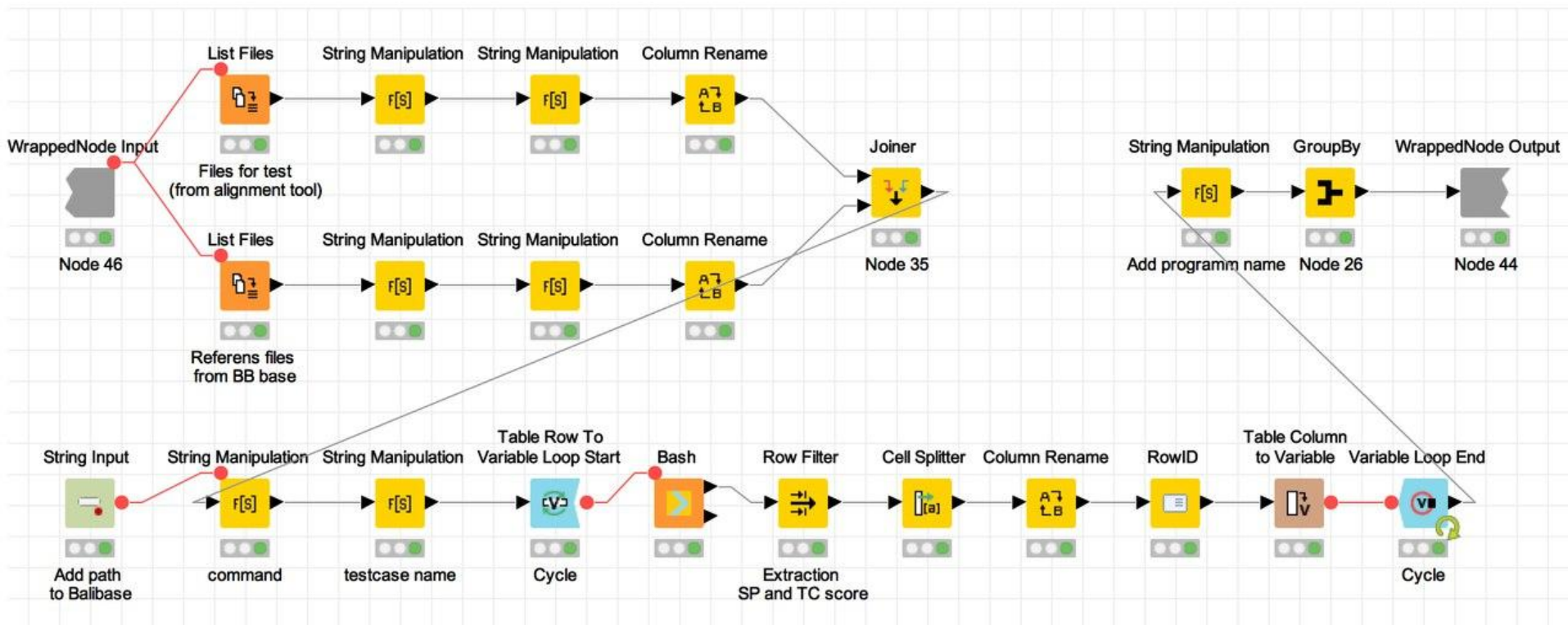
Задачи:

1. Познакомиться с существующими тулами для множественного выравнивания и бенчмарками для их тестирования.
2. Интегрировать выбранные тулы в KNIME.
3. Развернуть решение на сервере с оформлением в виде веб-интерфейса.

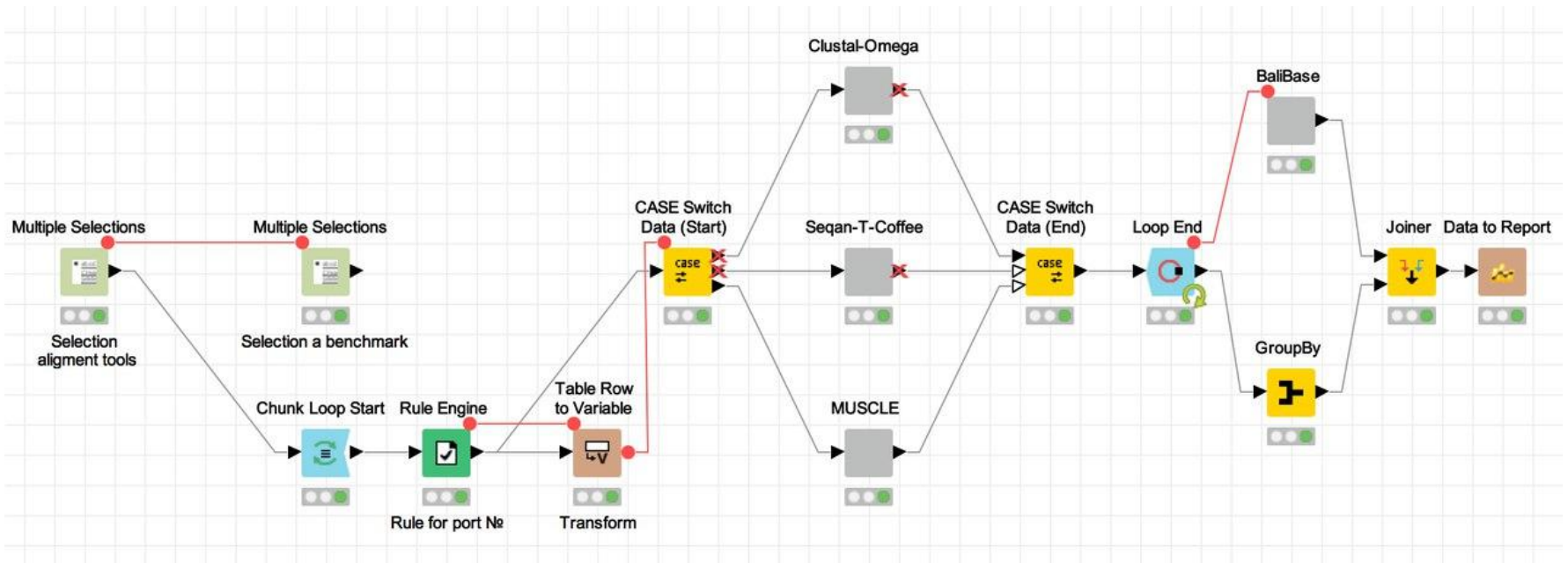
Добавление в KNIME программ для множественного выравнивания



Добавление в KNIME бенчмарка BALiBASE



Объединение смысловых блоков



Выходные данные

KNIME Report

Knime report powered by *Birt*

"tool_name"	"runtime"
Clustal Omega	569
MUSCLE	1982
SeqAn::T-Coffee	1913

Tool name	Mean(SP score)	Mean(TC score)
Clustal Omega		
RV11	0,481	0,247
RV12	0,847	0,683
RV20	0,823	0,306
RV30	0,76	0,373
RV40	0,799	0,428
RV50	0,735	0,358
MUSCLE		
RV11	0,465	0,229
RV12	0,846	0,679
RV20	0,806	0,249
RV30	0,718	0,246
RV40	0,759	0,338
RV50	0,702	0,304
SeqAn::T-Coffee		
RV11	0,513	0,261
RV12	0,855	0,679
RV20	0,823	0,252
RV30	0,716	0,241
RV40	0,779	0,366
RV50	0,74	0,351

Итоги

1. Познакомиться с существующими тулами для множественного выравнивания и бенчмарками для их тестирования.
2. Интегрировать выбранные тулы в KNIME.
3. Развернуть решение на сервере с оформлением в виде веб-интерфейса(в процессе)

Спасибо за внимание

