Федеральная служба по надзору в сфере защиты прав потребителей и благополучия человека

**Федеральное казённое учреждение здравоохранения «Ростовский-на-Дону ордена Трудового Красного Знамени научно-исследовательский противочумный институт»** **Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека**

(ФКУЗ Ростовский-на-Дону противочумный институт Роспотребнадзора)

УТВЕРЖДАЮ

Директор ФКУЗ

Ростовский-на-Дону противочумный институт

Роспотребнадзора,

\_\_\_\_\_\_\_\_\_\_\_\_ Н.Е. Гаевская

«\_\_\_» \_\_\_\_\_\_\_\_\_\_ 2025 г.

**Методические рекомендации по анализу данных полногеномного секвенирования штаммов *Salmonella spp.* с помощью программы «Salmonella Analyzer 2»**

Ростов-на-Дону

2025 г.

**Учреждение - разработчик:**

ФКУЗ Ростовский-на-Дону противочумный институт Роспотребнадзора

**Авторы:** Горох А.М., Водопьянов А.С., Писанов Р.В., Герасименко А.А.

**Уровень внедрения —** учрежденческий

**Согласовано** с КББ ФКУЗ Ростовского-на-Дону противочумного института Роспотребнадзора ( протокол № от г.)

**Председатель КББ Цимбалистова М.В.**

**Рассмотрены** на заседании Ученого совета ФКУЗ Ростовский-на-Дону противочумный институт Роспотребнадзора протокол № от г.

**Область применения:** методические рекомендации предназначены для специалистов, занимающихся молекулярно-генетическим и биоинформационным анализом штаммов сальмонелл.

**Аннотация**

Программа **«**Salmonella Analyzer 2» предназначена для анализа данных полногеномного секвенирования штаммов *Salmonella spp.* с целью оценки качества сборки геномных последовательностей; определения антигенной структуры штаммов; выявления основных факторов патогенности; генов супероксиддисмутаз и наличия плазмид; определения INDEL-, CRISPR-типов; SNP-типов (для *Salmonella enterica* Enteritidis).

**Установка и запуск программы**

**«**Salmonella Analyzer 2» доступна для скачивания на сайте ФКУЗ Ростовского-на-Дону противочумного института Роспотребнадзора. Для запуска программы достаточно распаковать скачанный архив с программой и запустить файл Salmonella Analyzer.bat (версия для Windows) или файл SalmonellaAnalyzer2.jar (версия для Windows, Linux и MacOS).

**Входные данные**

В качестве исходных данных используется файл в fasta-формате, содержащий либо набор контигов (протяженных нуклеотидных последовательностей — результат сборки *de novo* коротких ридов, полученных в результате полногеномного секвенирования), либо в ходе гибридной сборки с последующей коррекцией длинными ридами.

**Алгоритм работы программы**

Проводится анализ нуклеотидной последовательности по заранее определенному набору последовательностей с помощью алгоритма Смита-Ватермана с аффинными штрафами за делеции-вставки, что дает возможность определения антигенного типа в соответствии с международной серологической классификации Кауффмана-Уайта, выявления генов факторов патогенности, супероксиддисмутаз, плазмидных репликонов, INDEL- и CRISPR-локусов, проведения SNP-анализа с определением принадлежности штаммов *Salmonella enterica* Enteritidis к определенной генетической линии в соответствии с внутрилабораторной классификацией ФКУЗ Ростовского-на-Дону противочумного института.

**Работа с программой в режиме графического интерфейса**

Для запуска программы с графическим интерфейсом достаточно просто запустить исполняемый файл. Внешний вид программы представлен на рисунке 1. Необходимо заполнить поле '"Result file name" для наименования файла с итоговыми результатами, выбрать интересующие опции для анализа, выбрать исследуемые последовательности по кнопке "Select files and run". Данными служат файлы с расширением .fasta или .fa. Анализ начнется автоматически. Ниже можно будет увидеть шкалу прогресса анализа как отдельной пробы, так и их совокупности (рис. 1). Ориентировочное время работы программы от 15 до 20 минут для каждого штамма (fasta-файла) со всеми включенными опциями (табл. 1) и может варьироваться в зависимости от количества анализируемых параметров и потоков. В ходе выполнения программы промежуточные результаты отображаются в окне программы.

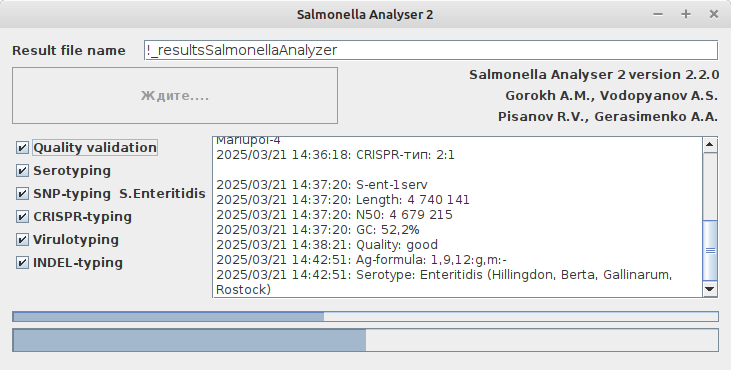


Рисунок 1 — Внешний вид программы «Salmonella Analyzer 2»

После завершения работы итоговая таблица с выбранным ранее именем будет храниться в папке, где находились выбранные для анализа последовательности. Результаты анализа каждого генома отображается в отдельной строке с указанием антигенной структуры, выявления факторов патогенности, генов супероксиддисмутаз и плазмид, определения генотипа по INDEL- и CRISPR-локусам и проведения SNP-анализа с установлением генетической линии, при этом количество и состав колонок будет зависеть от выбранных параметров при запуске (табл. 1а, б, в, г, д, е, ж).

**Параметр «Quality validation»** нужен для оценки качества генома исследуемой пробы. Результат работы программы с данным включенным флагом представляет собой нахождение длины последовательности (Length), N50, GC % нуклеотидный состав и качество сборки (Quality), содержит 10 генов жизнеобеспечения, необходимые для поддержания базовых клеточных процессов. В колонке Reason quality указывается причина, по которой геном признан невалидным (при наличии) (табл. 1а).

Таблица 1а – Результат работы программы с параметром Quality validation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| File | Strain | Length | N50 | GC | Quality | Reason quality |
| S-ent-203.fasta | S-ent-203 | 4752558 | 730548 | 52,3 | good |  |
| S-ent-107.fasta | S-ent-107 | 4738997 | 4679367 | 52,2 | good |  |
| S-ent-9296.fasta | S-ent-9296 | 4867750 | 200337 | 52,1 | good |  |

**Параметр «Serotyping»** позволяет произвести серотипирование штамма *in silico*. При этом выявляется тип по О-антигену и точность определения в % (O-Ag); по жгутиковому антигену фазы I (H1); фазы II (H2); дается полная антигенная формула (Ag-formula) и название серотипа (Serotype) (табл. 1б).

Таблица 1б – Результат работы программы с параметром «Serotyping»

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| File | Strain | O-Ag | H1: | H2: | Ag-formula | Serotype |
| S-ent-203.fasta | S-ent-203 | 28 (99.47%) | z10 (100,00%) | e,n,x (99,93%) | 28:z10:e,n,x | Umbilo |
| S-ent-107.fasta | S-ent-107 | 1,9,12 (99.70%) | g,m (100,00%) | - (-) | 1,9,12:g,m:- | Enteritidis (Hillingdon, Berta, Gallinarum, Rostock) |
| S-ent-9296.fasta | S-ent-9296 | 8 (100%) | d (100,00%) | 1,2 (99,87%) | 8:d:1,2 | Muenchen (Virginia, II) |

**Параметр «SNP-typing»** представляет филогенетический анализ с определением генетической линии исследуемого изолята, если он принадлежит к серотипу Enteritidis. В колонке Lineage можно увидеть конкретную генетическую линию, а в Lineage Path - родительские ветви, от которых произошел исследуемый изолят (табл. 1в).

Таблица 1в – Результат работы программы с параметром «SNP-typing»

|  |  |  |  |
| --- | --- | --- | --- |
| File | Strain | Lineage | Lineages path |
| S-ent-203.fasta | S-ent-203 | not Enteritidis | not Enteritidis |
| S-ent-107.fasta | S-ent-107 | Mariupol-3 | XVII, XVIII, Glob-3, DPR, Mariupol-3 |
| S-ent-9296.fasta | S-ent-9296 | not Enteritidis | not Enteritidis |

**Параметр «CRISPR-typing»** позволяет получить CRISPR-тип с указанием точности найденной аллели в % (CRISPR %) изучаемого штамма и последовательности аллелей (CRISPR1\_seq, CRISPR2\_seq) (табл. 1г).

Таблица 1г – Результат работы программы с параметром «CRISPR-typing»

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| File | Strain | CRISPR-тип | CRISPR% | CRISPR1\_seq | CRISPR2\_seq |
| S-ent-203.fasta | S-ent-203 | CRISPR: 31:new | 100.00 / - | ACGTTCT… | CACAGAT… |
| S-ent-107.fasta | S-ent-107 | CRISPR: 1:1 | 100.00 / 100.00 | ACGTCCT… | CACAGAT… |
| S-ent-9296.fasta | S-ent-9296 | CRISPR: 41:75 | 100.00 / 100.00 | ACGTCCT.. | CACAGAT… |

**Параметр «Virulotyping»** позволяет проводить поиск генов и островков вирулентности. При этом в табличном файле будут присутствовать все найденные гены и точность найденного совпадения с референсным геном в % (табл. 1д).

Таблица 1д – Результат работы программы с параметром «Virulotyping»

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| File | Strain - Strain | SPI-1 - fur regulated Salmonella iron transporter | SPI-2 - tetrathionate reductase complex, subunit A | SPI-3 - ATP binding protein | SPI4-23 |
| S-ent-203.fasta | S-ent-203 | 99,50% | 99,0% | not found | … |
| S-ent-107.fasta | S-ent-107 | 98,5% | 99,0% | 92,3% | … |
| S-ent-9296.fasta | S-ent-9296 | 97,5% | 99,0% | not found | … |

Также с данной опцией находятся гены супероксиддисмутаз и определяется наличие плазмид в исследуемом изоляте (табл. 1е).

Таблица 1е – Результат работы программы с параметром «Virulotyping», продолжение

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Гены супероксиддисмутаз - sodA | sodB | sodC | sodC2 | Плазмидные репликоны - Col156 | Col440I | ColRNAI | Plasmid N |
| 98,5% | 99,50% | 99,06% | 87,9% | not found | not found | not found | … |
| 98,3% | 99,83% | 99,06% | 87,9% | not found | not found | not found | … |
| 98,7% | 99,16% | 99,06% | 87,9% | not found | 91,2% | 82,6% | … |

**Параметр «INDEL-typing»** позволяет проводить поиск INDEL-локусов изучаемых изолятов с указанием названия локуса и его длины в парах оснований (табл. 1ж).

Таблица 1ж – Результат работы программы с параметром «INDEL-typing»

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| File | Strain - Strain | INDEL-маркеры - STY4288 | STY3837 | STY1159 | STY…. |
| S-ent-203.fasta | S-ent-203 | 85 b.p. | 100 b.p. | 74 b.p. | … |
| S-ent-107.fasta | S-ent-107 | 97 b.p. | 89 b.p. | 92 b.p. | … |
| S-ent-9296.fasta | S-ent-9296 | 85 b.p. | 100 b.p. | 74 b.p. | … |

**Запуск программы в консольном режиме**

Запуск в консольном режиме проводится с помощью команды

java -jar SalmonellaAnalyzer2.jar -i <genome.fasta> -r <result file name>

где:

<genome.fasta> - путь к исходному fasta-файлу генома

<result file name> - путь к файлу результата в формате json

По умолчанию при запуске в консольном режиме проводится анализ с использованием всех возможных параметров. При необходимости сократить анализ возможно использование дополнительных параметров:

-noQuality – не проводить оценку качества

-noSero – не проводить серотипирование

-noCRISPR – не проводить анализ CRISPR

-noVirulome – не проводить поиск генов вирулентности

-noINDEL – не проводить поиск INDEL-маркеров

-noSNP – не проводить SNP-анализ

При запуске в консольном режиме результат выдается в виде json-файла (рис. 2).

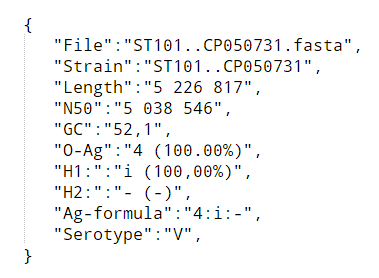


Рисунок 2 — Пример результата в json-формате

**Запуск программы в консольном многопоточном режиме**

Дополнительно реализована возможность запуска в консольном многопоточном режиме с помощью скрипта на языке multiThreads.py (требуется установленный Python не ниже версии 3.6). Для запуска программы в многопоточном режиме и получения результата в json формате нужно запустить из рабочей директории в терминале или командной строке скрипт multiThreads.py

Необходимые аргументы:

**-j** — путь до исполняемой программы (SalmonellaAnalyzer2.jar);

**-i** — путь до папки с пробами;

**-o** — путь до папки, где будут храниться результаты;

**-t** — число выделяемых потоков для работы.

После запуска вышеприведенного скрипта необходимо ответить на предложенные вопросы, которые позволят выбрать, какие характеристики геномов необходимо исследовать. Если нужно включить или отключить какую-либо из предложенных опций, следует ответить «да» (y) или «нет» (n) (рис. 3). Итоговые файлы в формате json можно увидеть в папке, которая была выбрана через аргумент -o.

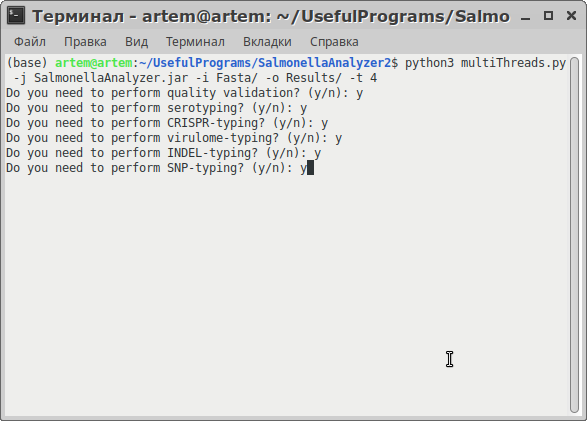


Рисунок 3 — Работа скрипта multiThreads.py для работы в многопоточном режиме

**Обратная связь**

Все пожелания, замечания, выявленные ошибки можно направлять по электронной почте (gorokh\_am@antiplague.ru; [vodopyanov\_as@antiplague.ru](https://mail.antiplague.ru/SOGo/so/gorokh_am@antiplague.ru/Mail/view)), телефону 89081705309; 8(863)240-22-66 или на почтовый адрес 344002, г. Ростов-на-Дону, ул. М. Горького, 117/40, Горох Алевтине Михайловне или Водопьянову Алексею Сергеевичу.