

Аннотация генома и анализ генов

CNE 8972 Биоинформатика · 6B05105 Биотехнология
Неделя 11 · Жаниязов Ж.А., старший преподаватель

План лекции

- Что такое аннотация
- Особенности генов бактерий
- Prokka и его результаты
- Функциональный анализ генов
- Ограничения и групповые презентации

11 / 15

неделя · аннотация генома

Что такое аннотация

- Определение положения и функции элементов генома
- Структурная: поиск генов, рРНК, тРНК
- Функциональная: назначение продукта и функции
- Следует за сборкой генома

Gene

где и что делает

Что ищем в геноме

CDS

Белок-кодирующие гены

рРНК и тРНК

Гены рибосомальных и транспортных РНК

ncRNA

Некодирующие РНК

Другие элементы

CRISPR-массивы, регуляторные участки

Особенности генов бактерий

- Интронов нет, гены расположены плотно
- ORF идёт от старт-кодона до стоп-кодона
- Старты: ATG, реже GTG и TTG; стопы: TAA, TAG, TGA
- Перед геном находится участок связывания рибосомы

ORF

открытая рамка считывания

Поиск ORF



Prokka использует Prodigal для предсказания генов.

Prokka

- Быстрая аннотация геномов бактерий
- Prodigal ищет гены, Aragorn — тРНК, Barrnap — рРНК
- Функции назначаются по базам белков
- Вывод: GFF, GenBank, FASTA белков



Команда Prokka

Параметр	Значение
--outdir out --prefix sample	Каталог и префикс файлов
--genus --species	Название организма
--kingdom Bacteria	Группа организмов
--cpus 8	Число потоков
contigs.fasta	Входная сборка

Файлы результатов

Расширение	Содержание
.gff	Аннотация (положение признаков)
.gbk	Формат GenBank
.faa	Белки в FASTA
.ffn	Нуклеотидные последовательности генов
.tsv	Таблица признаков
.txt	Итоговая статистика

Статистика аннотации

contigs

число контигов

bases

общая длина

CDS

белок-кодирующие
гены

**tRNA,
rRNA**

РНК-гены

Сравните значения с ожидаемыми для вашего вида.

Функциональный анализ

Hypothetical protein

Функция пока неизвестна

Гомологи

BLAST, UniProt

Домены

Pfam, InterPro

Специальные базы

CARD: устойчивость к антибиотикам, VFDB: вирулентность

Анализ конкретного гена

- Найдите ген в файле .gff командой grep
- Извлеките белок из файла .faa
- Проверьте гомологи в BLAST и UniProt
- Оцените, где расположен ген и какие рядом

grep

поиск гена в .gff

Ограничения

Границы генов

Автоматические ошибки в старт-кодонах

Гипотетические белки

Много генов без известной функции

Зависимость от баз

Качество аннотации зависит от базы

Ручная проверка

Важные гены нужно проверять

Итоги и контрольные вопросы

- Что такое аннотация генома?
- Какие программы использует Prokka?
- Какие файлы создаёт Prokka?
- Практика: ПР №11 «Prokka». СРО №3: групповые презентации

180–190

стр. · [1] Lesk