

Архивация и установка программ: tar, gzip, apt, conda

CNE 8972 Биоинформатика · 6B05105 Биотехнология
Неделя 5 · Жаниязов Ж.А., старший преподаватель

План лекции

- Архивация и сжатие файлов
- gzip и tar
- Менеджеры пакетов apt и conda
- Окружения и bioconda
- Установка FastQC и SPAdes

5 / 15

неделя • установка программ

Зачем архивация

- Данные NGS велики, сжатие экономит место
- Каталог удобнее передавать одним файлом-архивом
- Расширения: .gz, .tar, .tar.gz (.tgz), .zip
- Сжатые FASTQ программы читают напрямую

.gz

сжатие данных

Сжатие gzip

Команда	Что делает
gzip файл	Сжимает файл (получится файл.gz)
gzip -k файл	Сжимает, сохраняя исходный файл
gunzip файл.gz	Распаковывает
zcat файл.gz	Выводит содержимое без распаковки

Архивы tar

Задача	Команда
Создать архив	<code>tar -czvf архив.tar.gz каталог</code>
Распаковать	<code>tar -xzvf архив.tar.gz</code>
Посмотреть содержимое	<code>tar -tf архив.tar.gz</code>

Ключи tar

c

create: создать архив

x

extract: извлечь

z

сжатие gzip

f

имя файла архива

Запомните: **czf** — упаковать, **xzf** — распаковать.

Менеджеры пакетов

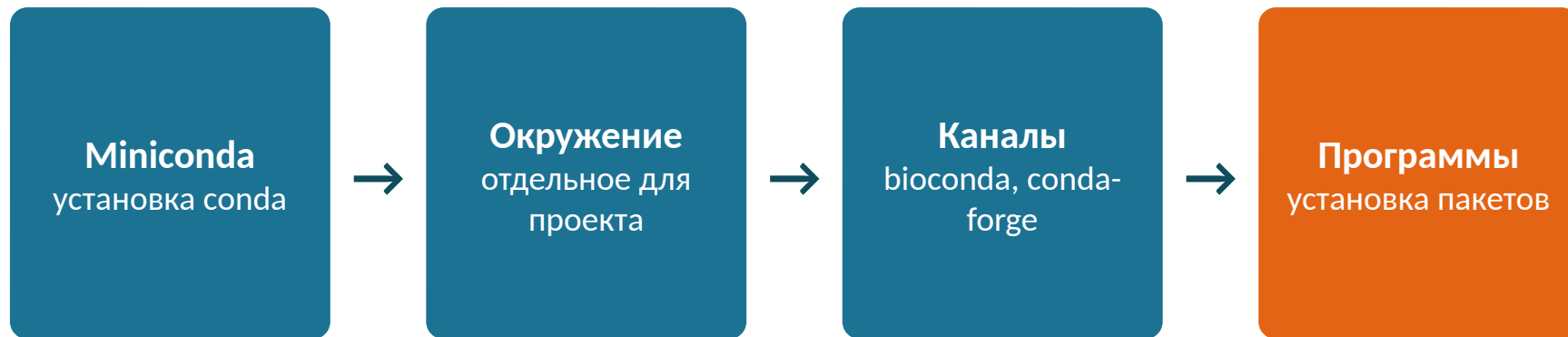
- Пакет — программа вместе с зависимостями
- apt: системные пакеты Debian и Ubuntu, нужен sudo
- conda: окружения и биоинформатические программы (канал bioconda)
- pip: пакеты Python

пакеты
установка без сборки

Команды apt

Команда	Что делает
<code>sudo apt update</code>	Обновляет список пакетов
<code>sudo apt install имя</code>	Устанавливает пакет
<code>apt search имя</code>	Ищет пакет
<code>sudo apt remove имя</code>	Удаляет пакет

Окружения conda



Окружение исключает конфликты версий между проектами.

Команды conda

Команда	Что делает
<code>conda create -n qc</code>	Создать окружение qc
<code>conda activate qc</code>	Включить окружение
<code>conda install -c bioconda fastqc</code>	Установить FastQC
<code>conda list</code>	Показать установленные пакеты
<code>conda deactivate</code>	Выйти из окружения

Программы курса

FastQC

Оценка качества прочтений

SPAdes

Сборка геномов микроорганизмов

Trimmomatic

Очистка прочтений

Prokka

Аннотация геномов

Проверка установки

- `fastqc --version` и `spades.py --version`
- `which fastqc` показывает путь к программе
- Ошибка «command not found»: окружение не активировано
- Проверьте переменную `PATH`



which
путь к программе

Итоги и контрольные вопросы

- Чем tar отличается от gzip?
- Когда использовать apt, а когда conda?
- Зачем нужны окружения?
- Практика: ПР №5 «Установка FastQC и SPAdes». СРО №1: групповые презентации по основам Linux

75–120

стр. · [1] Lesk