

Язык программирования Python в биоинформатике

CNE 8972 Биоинформатика · 6B05105 Биотехнология

Неделя 8 · Жаниязов Ж.А., старший преподаватель

План лекции

- Почему Python
- Типы данных и работа со строками
- Условия, циклы, функции
- Чтение файлов и Biopython
- Разработка приложения с помощью ИИ

8 / 15

неделя · Python

Почему Python

- Простой и читаемый синтаксис
- Библиотеки: Biopython, pandas, matplotlib
- Работает на сервере и на личном компьютере
- Широко применяется в биоинформатике



Основные типы данных

Тип	Пример	Применение
str	"ATGC"	Последовательность
int	150	Длина, счётчик
float	0.52	Доля GC
list	["seq1", "seq2"]	Набор значений
dict	{"A": 10}	Счётчики нуклеотидов

Работа со строкой ДНК

Выражение	Результат
<code>len(seq)</code>	Длина последовательности
<code>seq.count("G")</code>	Число нуклеотидов G
<code>seq.upper()</code>	Заглавные буквы
<code>seq[0:3]</code>	Срез: первый кодон
<code>seq[::-1]</code>	Обратная строка

Условия и циклы

- if / elif / else выбирают действие по условию
- for base in seq проходит по нуклеотидам
- while повторяет действие, пока условие верно
- Блоки отделяются отступом из 4 пробелов

for

цикл по последовательности

Функции

- `def` задаёт функцию; `return` возвращает результат
- `gc_content(seq): (G + C) / длина × 100`
- Функцию можно вызывать многократно
- Хорошая функция решает одну задачу

def

своя функция

Чтение файлов

- `with open("file.fa") as f:` открывает файл и закрывает его автоматически
- `for line in f:` читает файл построчно
- `line.strip()` убирает перевод строки
- Заголовок FASTA начинается с символа «>»

open
чтение файла

Biopython

Инструмент	Назначение
SeqIO.parse	Чтение FASTA и FASTQ
Seq	Объект последовательности
reverse_complement()	Обратная комплементарная цепь
translate()	Трансляция в белок
Установка	conda install biopython

Мини-задача



Такую программу из 15–20 строк можно написать и проверить за одно занятие.

ИИ как помощник программиста

Чёткая постановка

Опишите вход, выход и формат файла

Проверка кода

Запустите на маленьком примере

Понимание

Разберите каждую строку, не копируйте вслепую

Честность

Указывайте, где использован ИИ

Хорошие практики

Понятные имена

seq, gc_content вместо a, f1

Комментарии

Коротко объясняйте шаги

Тест на малых данных

Сначала проверьте на 2–3 записях

Окружение

Фиксируйте версии в conda

Итоги и контрольные вопросы

- Какие типы данных Python вы знаете?
- Как посчитать GC-состав последовательности?
- Какие задачи решает Biopython?
- Практика: ПР №8 «Создание биоинформатического приложения с помощью ИИ». Первая промежуточная аттестация: 8 неделя

136–140

стр. · [1] Lesk