

Задания

1. Найдите в интернете текст песни Элтона Джона Rocket Man на английском. С помощью nano создайте файл, добавьте в него этот текст. Затем переведите все строчные буквы в заглавные, а пробелы в нижнее подчеркивание. Сохраните первые 5 строк как task1.txt
2. Замените дуб на куб, цепь на степь (со всеми склонениями) и любое другое слово на слово, подходящее по рифме в Pushkin.txt, сохраните первые 10 строк в task2.txt
3. Отсортируйте Pushkin.txt по второму слову, выберите 8 последних строк и допишите 2 последние строки от себя (но в рифму!). Сохраните в task3.txt
4. Посчитайте, сколько раз в man-е по программе grep встречается слово grep, сохраните в task4.txt;
5. Напишите скрипт, который будет приветствовать пользователя "Здравствуйте, %username%. Вы находитесь в папке %текущая папка%". Сохраните скрипт как task5.txt
6. Скачайте размеры хромосом S.cerevisiae и назовите, как вам удобно:
<http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/bigZips/sacCer3.chrom.sizes>
Сохраните длину четырнадцатой хромосомы **с помощью кода** в файл task6.txt
7. Установите пакет tree. Прочитайте мануал, выполните команду в рабочей папке, результат сохраните в файл task7.txt
8. Скачайте Гамлета Шекспира по ссылке
https://shakespeare.folger.edu/downloads/txt/hamlet_TXT_FolgerShakespeare.txt
Найдите все строки, начинающиеся с букв от v до x (не важно, строчных или заглавных) и не содержащие строчные буквы от a до e. Запишите результат в task8.txt
9. Скачайте файл GFF3 для Saccharomyces cerevisiae по ссылке
http://ftp.ensembl.org/pub/release-105/gff3/saccharomyces_cerevisiae/Saccharomyces_cerevisiae.R64-1-1.105.gff3.gz
Разархивируйте. Какую он имеет структуру? Если необходимо, найдите описание такой структуры файлов в интернете. Найдите число элементов каждого типа (например, гены, CDS, экзоны, интроны и т.д.), сохраните в task9.txt
10. Сохраните длину первых 5 генов в файле
Saccharomyces_cerevisiae.R64-1-1.105.gff3 в task10.txt
11. Сохраните историю в файл history.txt

Последние задания взяты из файла Department of Computational Biology - Cornell University. Если у вас получилось их сделать, вы молодец!

11*. Перейдите по ссылке

https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/

1) Найдите clinvar_20211212.vcf.gz. Как получить на него прямую ссылку?
Загрузите clinvar_20211212.vcf.gz в новый каталог.

- 2) Каков размер полученного архива? Распакуйте. Каков размер несжатого vcf-файла?
- 3) Какова структура файла?
- 4) Подсчитайте строки в vcf-файле, которые начинаются с # (заголовок и информация).
- 5) Подсчитайте строки с самим содержимым файла.
Сохраните информативные строки в новый файл и удалите их из clinvar_20211212.vcf.
- 6) Вырежьте столбцы CHROM, POS, REF и ALT из clinvar_20211212.vcf и сохраните как .
- 7) Найдите самый мутабельный нуклеотид REF на 15 хромосоме.
- 8) Найдите наиболее распространенную пару нуклеотидов REF-ALT в 15 хромосоме.