

La solución de este taller debe ser presentada en un solo archivo comprimido con nombre `NombreApellido_HW10.zip`. El archivo comprimido debe tener el script del primer literal y un cuaderno de `iPython` con las otras respuestas. Puede trabajarse en grupo de máximo tres estudiantes.

```
>gi|528476511|ref|NW_004929286.1| Homo sapiens chromosome 1 genomic sca
TAACCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
CCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
CCCAACCCCAACCCCAACCCCAACCCCAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTA
ACCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAAC
CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTCGCGGTACCTCAGCCGGCCCGCCCGCCGGTCTG
ACCTGAGGAGAACTGTCTCCGCTTCAGAGTACACCGAAATCTGTGACAGAGGACAACGCAGCTCCGCC
CTCGCGGTGCTCTCCGGGTCTGTGCTGAGGAGAACGCACCTCCGCGGCGCAGGCGCAGAGAGGCGCGCC
GCGCCGGCGCAGGCGCAGAGAGGCGCGCGCGCGCGCAGGCGCAGAGAGGCGCGCGCGCGCGCGCAG
GCGCAGAGAGGCGCGCGCGCGCGCGCGCGCAGGCGCAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCG
GCGCCGCGCGCGCGCAGGCGCAGAGAGGCGCGCGCGCGCGCGCGCGCAGGCGCAGAGAGGCGCGCGCG
CGCAGGCGCAGAGACATGCTAGCGCGTCCAGGGGGGTGGAGGCGTGGCGCAGGCGCAGAGACGCACGC
CTACGGGGGGGGGTGGGGGTGCGTGTGTTGACAGGAGCAAAGTCGACGGCGCGGGCTGGGGCGGGGG
GCCGTGCACGCGCAGAAACTCACGTACGGCGCGCGCGCGCAGAGACGGGTGGAACCTCAGTAATCCGAA
ACGCCGGGATCGACAGCCCTTGCTTGACGCGGGCACGACAGGACCGCTTGCTCACGGTGTGTGCCA
GGGCGCCCCCTGCTGGCGACTAGGGCAACTGCAGGGCTCTCTTGCTTAGAGTGGTGGCCAGCGCCCCCTG
CTGGCGCCGGGGCACTGCTGGGCCCTCTTGCTTACTGTATAGTGGTGGCACGCCGCTGCTGGCAGCTAG
GGCATTGCGAGGTCCTCTTGCTCAAGGAGTAGTGGCAGCACGCCGCTGCTGGCAGCTGGGGACACTG
CTGGGCCCTCTTGCTCAAGAGTAGTGGGGCTTATAGGGAACACGAGGAGCATATGCTGTTGGTCTG
```

1. (a) 25 pt Escriba un script de `Bash` con nombre `downbot.sh` que descargue y descomprima de forma programática los 22 autosomas del genoma humano (los archivos `.fa.gz`). Los archivos están alojados en un servidor FTP del NCBI en directorios encontrados en `ftp://ftp.ncbi.nih.gov/genomes/Homo_sapiens/`, y cuando se descomprimen el resultado son archivos de texto de tipo FASTA. Explore la estructura de los archivos. El script debe tener menos de 300 caracteres. Para descomprimir los archivos el comando `gunzip` puede ser de utilidad.
PRECAUCIÓN: tenga en cuenta que la cantidad de información a descargar es de aproximadamente 1GB.
- (b) 25 pt Escriba código en `Python` que cuente el total de A , G , T , y C en el conjunto de todos los autosomas. Tenga cuidado de no procesar el encabezado de los archivos.
- (c) 25 pt Escriba código en `Python` que cuente y guarde en cuatro listas `countA`, `countG`, `countT` y `countC` la cantidad de ocurrencias de cada base nitrogenada en las líneas de los archivos de los autosomas (cada una con 70 caracteres). Tenga cuidado de no procesar el encabezado de los archivos. Calcule el promedio, mediana y desviación estándar de cada lista.
- (d) 25 pt Haga tres histogramas cumulativos: uno que incluya en los mismos ejes los histogramas de `countG` y `countC`, los de `countA` y `countT`, y los de `countG` y `countA`. Todos los histogramas deben tener leyendas apropiadas y estar normalizados. Los resultados deben ser similares a lo mostrado en la siguiente figura y deben estar incluidos en el cuaderno de `iPython`.

Estudio de la frecuencia de las bases nitrogenadas
en los autosomas del genoma humano.

