Genome Organization

Download the file CryptoDB-50_CparvumIOWA-ATCC.gff to your working directory

```
cryptoFile=CryptoDB-50_CparvumIOWA-ATCC.gff

if [ ! -f $cryptoFile ]; then
  wget https://cryptodb.org/common/downloads/Current_Release/CparvumIOWA-
ATCC/gff/data/CryptoDB-50_CparvumIOWA-ATCC.gff
fi
```

1. How many genes do you have in the gff file?

```
cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f3 | grep -v "#" | grep -c
"gene"
```

2. How many CDS?

```
cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f3 | grep -v "#" | grep -c
"CDS"
```

3. How many exons?

```
cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f3 | grep -v "#" | grep -c
"exon"
```

4. How many ncRNA (ribosomal RNAs, tRNAs, etc.)?

5. How many of each?

```
ncDnaNames=("tRNA" "rRNA" "five_prime_UTR" "three_prime_UTR" "snRNA")
nNcDna=0

for i in ${!ncDnaNames[@]}
do
  echo "Number of ${ncDnaNames[$i]}: "
  nTmp=$(cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f3 | grep -v "#" |
grep -c ${ncDnaNames[$i]})
  echo $nTmp
  nNcDna=$((nNcDna+nTmp))
```

```
done

echo "Total number of ncDNA: "
echo $nNcDna
```

6. How many genes by Chromosome?

```
chroms=($(cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f1 | grep -v "#" |
sort | uniq))
for chrom in "${chroms[@]}"
do
  nChrom=$(cat CryptoDB-50_CparvumIOWA-ATCC.gff | cut -f1-3 | grep -v "#" |
grep "gene" | grep -c "$chrom")
  echo "${chrom}: $nChrom"
done
```

7. How many genes have a product name?

8. How many are Uncharacterized or Hypothetical?

9. What is the ID/Name of the last gene in CpIOWA-ATCC_Chr1?

   - CPATCC_0039550