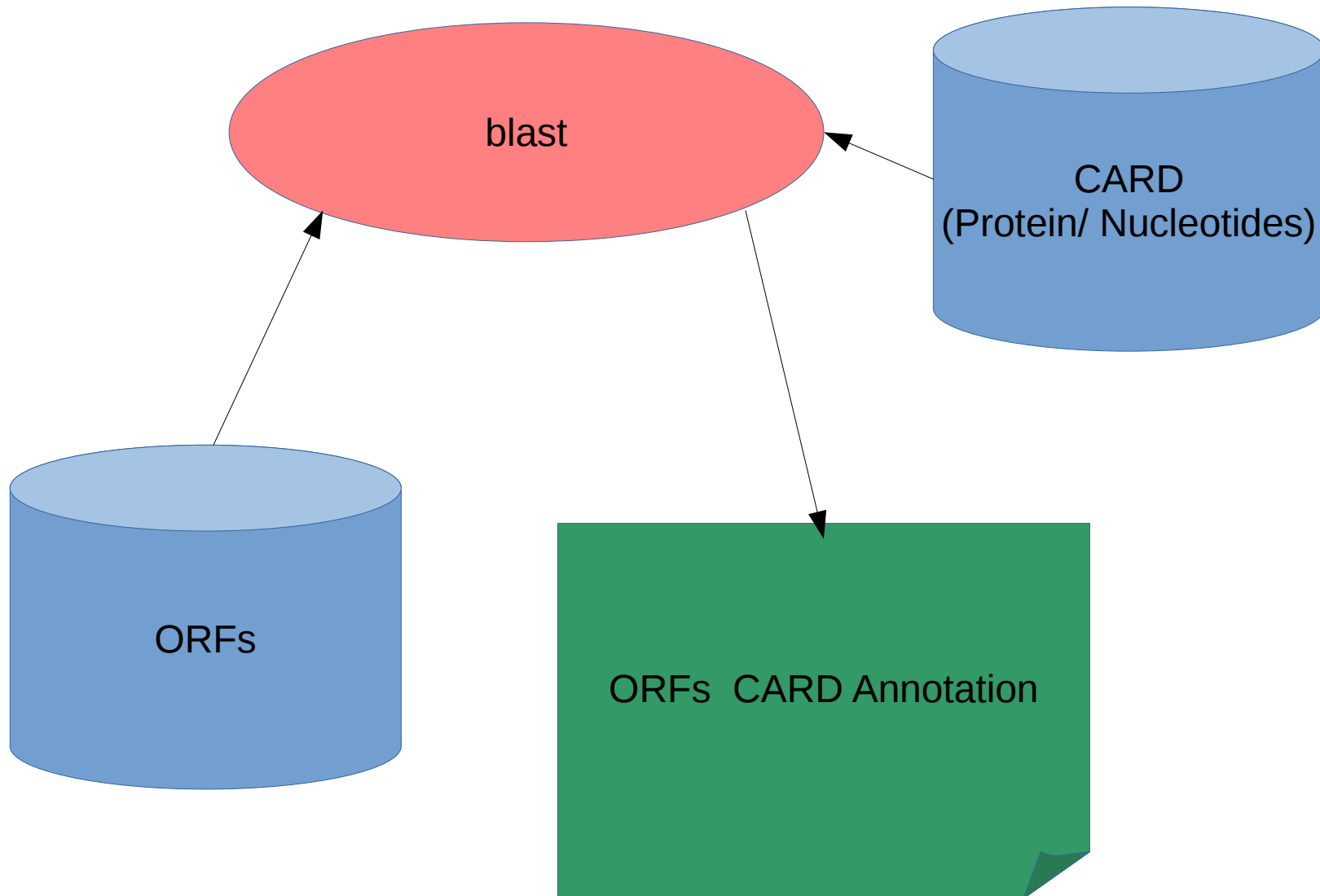


Task 5: CARD Annotation of Metagenomic ORFs

Task 6: Summary



Task 5: Download CARD

```
# getting and installing
```

```
cd ~/Databases
```

```
mkdir CARD
```

```
cd CARD
```

```
mkdir RawData
```

```
cd RawData
```

```
wget https://card.mcmaster.ca/download/0/broadstreet-  
v1.1.8.tar.gz
```

Task 5: Compile Fasta CARD

```
ls
```

```
##sudo yum install bzip2
```

```
tar -xvf broadstreet-v1.1.8.tar.gz
```

```
ls
```

```
## to show the card nucleotide fasta files
```

```
ls nucleotide_fasta_*
```

```
cat nucleotide_fasta_* > ../All.CARD.NA.fa
```

```
ls protein_fasta_*
```

```
cat protein_fasta_* > ../All.CARD.AA.fa
```

Task 5: Compile blastdb CARD

```
# to insure that you are at the right place.
```

```
cd ~/Databases/CARD/
```

```
mkdir blastNA blastAA
```

```
## compiling blast protein database for CARD
```

```
cd blastAA
```

```
~/Programs/ncbi-blast-2.6.0+/bin/makeblastdb -in ../All.CARD.AA.fa -title CARDProt \  
-out CARDProt -input_type fasta -hash_index -dbtype prot
```

```
## to confirm that you are at ~/Databases/CARD/
```

```
cd ..
```

```
## compiling blast Nucleotides database for CARD
```

```
cd blastNA
```

```
~/Programs/ncbi-blast-2.6.0+/bin/makeblastdb -in ../All.CARD.NA.fa -title CARDNucl \  
-out CARDNucl -input_type fasta -hash_index -dbtype nucl
```

Task 5: Annotate NA ORFs using NA CARD

```
## Run CARD databses
```

```
cd ~/Data
```

```
mkdir CARD
```

```
cd CARD
```

```
## nucleotide versus card nucleotide
```

```
~/Programs/ncbi-blast-2.6.0+/bin/blastn -query ../orfs.nucleotide.fa -db ~/Databases/CARD/blastNA/CARDNucl -outfmt 6 \
-max_target_seqs 10 -evalue 1E-50 -word_size 6 -num_threads 28 -out orf.card.NA.versus.NA.tab
```

```
## to get the top hit based on bitscore.
```

```
sort -k1,1 -k12,12nr orf.card.NA.versus.NA.tab | awk -F"\t" '{if(FNR==1) {geneID=$1; print $0} else
{if(geneID!=$1) {geneID=$1; print $0}}}' > orf.card.NA.versus.NA.tab.top1.txt
```

```
wc -l orf.card.NA.versus.NA.tab.top1.txt
```

```
## to show the frequency of Antibiotic resistance genes observed.
```

```
cut -f1,2 orf.card.NA.versus.NA.tab.top1.txt | sort -u | cut -f2 | sort |
uniq -c | awk 'BEGIN {print "Gene\tNumberOfORFsObserved"; sum=0;}
{print $2"\t"$1; sum=sum+$1} END {print "Total\t"sum}'
```

Task 5: ORF Proteins using AA CARD

protein versus card nucleotide

```
~/Programs/ncbi-blast-2.6.0+/bin/tblastn -query ../orfs.protein.fa -db
```

```
~/Databases/CARD/blastNA/CARDNucl -outfmt 6 \
```

```
-max_target_seqs 10 -evaluate 1E-50 -word_size 4 -num_threads 28 -out orf.card.AA.versus.NA.tab
```

```
sort -k1,1 -k12,12nr orf.card.AA.versus.NA.tab | awk -F"\t" '{if(FNR==1) {geneID=$1; print $0}  
else {if(geneID!=$1) {geneID=$1; print $0}}}' > orf.card.AA.versus.NA.tab.txt.top1.txt
```

to show the frequency of Antibiotic resistance genes observed.

```
wc -l orf.card.AA.versus.NA.tab.txt.top1.txt
```

to show the frequency of Antibiotic resistance genes observed.

```
cut -f1,2 orf.card.AA.versus.NA.tab.txt.top1.txt | sort -u | cut -f2 | sort |
```

```
uniq -c | awk 'BEGIN {print "Gene\tNumberOfORFsObserved"; sum=0;}
```

```
{print $2"\t"$1; sum=sum+$1} END {print "Total\t"sum}'
```