

---

# Gene Prediction - Results

Team 2 (Group 2):  
Gautham Krishna Sankar Ramalaxmi, Jyothi Guruprasad,  
Kaize H Ali, Lee Ellen Mullins, Varsha Srinivasan

---

# Gene Prediction Background

## Homology based prediction

- Finding similarity in gene sequences between the target genome and other sequences.

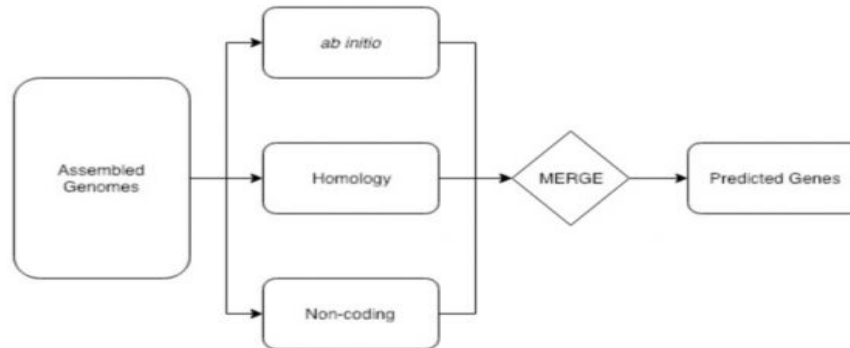
## Ab-initio prediction

- Finding protein sequences using sequence information alone
- Does not use homolog (template) structure

# Proposed Pipeline

Tools available:

- Ab-initio : GeneMarkS2, GLIMMER, Prodigal
- Homology-based : BLAST
- Merging results : Bedtools intersect, ORForise



# GeneMarkS2

- Ab-initio algorithm that finds species-specific genes to identify other genes
- Software is built using Hidden Markov Model (HMM)
- Robust method of detection
- Latest iteration was developed to increase the prediction accuracy of *where* the gene starts



[Genome Res.](#) 2018 Jul; 28(7): 1079–1089.

doi: [10.1101/gr.230615.117](https://doi.org/10.1101/gr.230615.117)

PMCID: PMC6028130

PMID: [29773659](https://pubmed.ncbi.nlm.nih.gov/29773659/)

Modeling leaderless transcription and atypical genes results in more accurate gene prediction in prokaryotes

[Alexandre Lomsadze](#),<sup>1,2,6</sup> [Karl Gemayel](#),<sup>3,6</sup> [Shiyuyun Tang](#),<sup>4</sup> and [Mark Borodovsky](#)<sup>1,2,3,4,5</sup>

# GeneMarkS2

Command line argument:

```
perl ../../GeneMarkS2/gms2.pl --genome-type bacteria --gcode 11 --output gms2_$FILE_NAME --format gff3 --gid 'gene_' --seq ../../Contigs/$FILE_NAME_FASTA.gff
```

Input parameters:

- **--genome-type** : describes type of input genome (auto/bacteria/archaea)
- **--gcode** : 11 - indicates that it is a bacteria
- **--format** : tells the output file format
- **--seq** : input file sequence (either .fna or .fasta format)
- **--output** : tells the output file name

# GeneMarkS2

Output obtained:

```
##gff-version 3
# GeneMark.hmm-2 prokaryotic version: 1.25.lic
# File with sequence: ../../Contigs/CGT2006.fasta
# File with native parameters: GMS2.mod
# Native species name and build: unspecified GeneMarkS-2-1.14_1.25.lic
# File with MetaGeneMark parameters: /home/team2/group2/GeneMarkS2/mgm_11.mod
# translation table: 11
# output date start: Mon Mar 13 15:28:07 2023

##sequence-region k141_6 1 262
# k141_6      total_logodd    0      average_length  0      average_density 0.00

##sequence-region k141_12 1 319
k141_12 GeneMark.hmm2  gene   3      317      .      +      .      ID=gene_k141_12_1;
k141_12 GeneMark.hmm2  CDS    3      317      38.95  +      0      ID=k141_12_1; Parent=gene_k141_12_1; gene_type=native; partial=11; gc=56; le
ngth=315;
# k141_12      total_logodd    38.9526 average_length  315      average_density 3.13

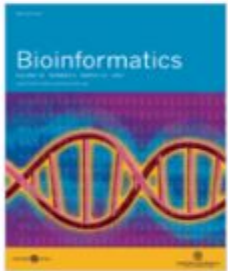
##sequence-region k141_10 1 300
k141_10 GeneMark.hmm2  gene   66      299      .      -      .      ID=gene_k141_10_1;
k141_10 GeneMark.hmm2  CDS    66      299      3.51   -      0      ID=k141_10_1; Parent=gene_k141_10_1; gene_type=atypical; partial=01; gc=32;
length=234;
# k141_10      total_logodd    3.51192 average_length  234      average_density 3.33

##sequence-region k141_2 1 441
k141_2  GeneMark.hmm2  gene   206      439      .      -      .      ID=gene_k141_2_1;
k141_2  GeneMark.hmm2  CDS    206      439      31.10  -      0      ID=k141_2_1; Parent=gene_k141_2_1; gene_type=native; partial=01; gc=50; leng
th=234;
# k141_2      total_logodd    31.0979 average_length  234      average_density 2.27

##sequence-region k141_13 1 347
k141_13 GeneMark.hmm2  gene   3      338      .      -      .      ID=gene_k141_13_1;
k141_13 GeneMark.hmm2  CDS    3      338      27.49  -      0      ID=k141_13_1; Parent=gene_k141_13_1; gene_type=native; partial=10; gc=54; le
ngth=336;
# k141_13      total_logodd    27.4856 average_length  336      average_density 2.88
```

# GLIMMER

- Find genes in prokaryotic DNA
- 1st system to use interpolated Markov model (IMM) to identify coding regions
- Has 2 built-in programs:
  - Build IMM - takes an input of many sequences, output is an IMM
  - Glimmer - using the above IMM, identifies the presumes gene



Volume 23, Issue 6  
march 2007

## JOURNAL ARTICLE

### Identifying bacterial genes and endosymbiont DNA with Glimmer

Arthur L. Delcher , Kirsten A. Bratke, Edwin C. Powers, Steven L. Salzberg

[Author Notes](#)

*Bioinformatics*, Volume 23, Issue 6, march 2007, Pages 673–679,

<https://doi.org/10.1093/bioinformatics/btm009>

# GLIMMER

Command line argument:

```
$glimmerpath/glimmer3 $glimmeropts -b $tag.motif -P $startuse $genome $tag.icm $tag
```

Input parameters:

- **--options** : can be gc\_percent/codon list/gene length etc
- **--sequence-file** : input file given for analysis
- **Other parameters/flags exist, but is outside the scope of our intended goal**



# GLIMMER

- For our situation, we used g3-iterated.sh, a script that uses g3-from-scratch.sh and g3-from-training.sh
- These pre-configured scripts exists in glimmer3.02/scripts
- Flags used were -o50: minimum nucleotide length set to 50. -g110: minimum distance between adjacent genes. -t30: minimum threshold score is set to 30
- Ran on Macbook Pro M2 with a core of 8 CPU

```
kc — kali45@biogenome2023:/home/team2/group2/ALL_RESULTS/glimmer — ssh kali45@biogeno
[kali45@biogenome2023 glimmer]$ ls
convert2gff.sh gff3 iterated iterated.sh run_iterated_on_all.sh runtime.log
[kali45@biogenome2023 glimmer]$ cat run_iterated_on_all.sh
#!/bin/bash

outdir="/Users/kc/downloads/iterated"

for fasta_file in *.fasta; do
    output_name=$(basename "$fasta_file" .fasta)
    output_file="$outdir/$output_name.glimmer"
    iterated.sh "$fasta_file" "$output_file"
done
[kali45@biogenome2023 glimmer]$
```

# Prodigal

- PROkaryotic DYnamic programming Genefinding ALgorithm
- Constructs its training set by examining the GC frame plot
- Uses dynamic programming to identify heavily overlapped ORFs
- Lowers false positives (<5%)

Software | [Open Access](#) | [Published: 08 March 2010](#)

## Prodigal: prokaryotic gene recognition and translation initiation site identification

[Doug Hyatt](#) , [Gwo-Liang Chen](#), [Philip F LoCascio](#), [Miriam L Land](#), [Frank W Larimer](#) & [Loren J Hauser](#)

# Prodigal

Command line argument:

```
prodigal -i "../../../Contigs/$FILE_NAME_FASTA" -c -m -f gff -o cds_$FILE_NAME.gff 2>&1 | tee log_$FILE_NAME #
```

Input parameters:

- **-i** : input file parameter (.fasta format)
- **-o** : name of output file generated
- **-f** : format of output file (.gff format in this case)
- **-c** : closed ends - does not allow partial genes at end of sequence
- **-m** : gap-handling behavior - genes cannot run into gaps

# Prodigal

## Output obtained:

```
#gff-version 3
# Sequence Data: seqnum=1;seqlen=262;seqhdr="k141_6 flag=0 multi=15.6860 len=262"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=2;seqlen=319;seqhdr="k141_12 flag=1 multi=1.0000 len=319"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=3;seqlen=300;seqhdr="k141_10 flag=1 multi=1.0000 len=300"
# Files 1 Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=4;seqlen=441;seqhdr="k141_2 flag=1 multi=1.0000 len=441"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
k141_2 Prodigal.v2.6.3 CDS 206 421 19.7 - 0 ID=4_1;partial=00;start_type=GTG;rbf_motif=GGA/GAG/AGG;rbf_spacer=5-10bp;gc_cont=0.495;conf=98.94;score=19.72;cscore=25.28;sscore=-5.50;rscore=2.50;uscore=1.60;tscore=6.46;
# Sequence Data: seqnum=5;seqlen=347;seqhdr="k141_13 flag=1 multi=1.0000 len=347"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=6;seqlen=711;seqhdr="k141_8 flag=0 multi=38.0000 len=711"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
k141_8 Prodigal.v2.6.3 CDS 63 680 62.5 - 0 ID=6_1;partial=00;start_type=ATG;rbf_motif=AGGAG;rbf_spacer=5-10bp;gc_cont=0.570;conf=100.00;score=62.51;cscore=46.36;sscore=16.15;rscore=14.94;uscore=2.36;tscore=3.56;
# Sequence Data: seqnum=7;seqlen=473;seqhdr="k141_3 flag=1 multi=1.0000 len=473"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=8;seqlen=1214;seqhdr="k141_7 flag=0 multi=34.0000 len=1214"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=9;seqlen=367;seqhdr="k141_11 flag=1 multi=1.0000 len=367"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
# Sequence Data: seqnum=10;seqlen=7471;seqhdr="k141_4 flag=0 multi=17.2149 len=7471"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
k141_4 Prodigal.v2.6.3 CDS 281 502 7.3 + 0 ID=10_1;partial=00;start_type=GTG;rbf_motif=None;rbf_spacer=None;gc_cont=0.509;conf=84.30;score=7.31;cscore=25.26;sscore=-17.95;rscore=-10.43;uscore=1.23;tscore=6.28;
k141_4 Prodigal.v2.6.3 CDS 699 1316 44.3 - 0 ID=10_2;partial=00;start_type=ATG;rbf_motif=AGGA/GGAG/GAGG;rbf_spacer=11-12bp;gc_cont=0.492;conf=100.00;score=44.27;cscore=37.54;sscore=0.73;rscore=3.79;uscore=0.83;tscore=3.56;
k141_4 Prodigal.v2.6.3 CDS 1979 2392 58.6 + 0 ID=10_3;partial=00;start_type=ATG;rbf_motif=GGA/GAG/AGG;rbf_spacer=5-10bp;gc_cont=0.457;conf=100.00;score=58.63;cscore=49.49;sscore=9.14;rscore=2.93;uscore=2.64;tscore=3.56;
k141_4 Prodigal.v2.6.3 CDS 2524 3432 139.4 - 0 ID=10_4;partial=00;start_type=ATG;rbf_motif=AGGAG;rbf_spacer=5-10bp;gc_cont=0.494;conf=100.00;score=139.38;cscore=119.43;sscore=19.94;rscore=14.94;uscore=2.09;tscore=3.56;
k141_4 Prodigal.v2.6.3 CDS 3635 4648 103.4 - 0 ID=10_5;partial=00;start_type=ATG;rbf_motif=GGAG/GAGG;rbf_spacer=5-10bp;gc_cont=0.510;conf=100.00;score=103.40;cscore=87.24;sscore=16.16;rscore=10.70;uscore=1.91;tscore=3.56;
k141_4 Prodigal.v2.6.3 CDS 4739 5655 57.5 - 0 ID=10_6;partial=00;start_type=ATG;rbf_motif=GGA/GAG/AGG;rbf_spacer=3-4bp;gc_cont=0.508;conf=100.00;score=57.46;cscore=58.25;sscore=-0.78;rscore=4.20;uscore=0.51;tscore=3.56;
k141_4 Prodigal.v2.6.3 CDS 5755 6213 20.0 + 0 ID=10_7;partial=00;start_type=GTG;rbf_motif=GGAG/GAGG;rbf_spacer=5-10bp;gc_cont=0.495;conf=99.00;score=19.98;cscore=12.74;sscore=7.23;rscore=10.70;uscore=2.04;tscore=5.50;
k141_4 Prodigal.v2.6.3 CDS 6264 7106 124.5 + 0 ID=10_8;partial=00;start_type=ATG;rbf_motif=AGGA/GGAG/GAGG;rbf_spacer=11-12bp;gc_cont=0.509;conf=100.00;score=124.49;cscore=114.71;rscore=9.78;rscore=3.79;uscore=1.70;tscore=3.56;
# Sequence Data: seqnum=11;seqlen=31814;seqhdr="k141_9 flag=0 multi=16.9778 len=31814"
# Model Data: version=Prodigal.v2.6.3;run_type=Single;model="Ab initio";gc_cont=51.44;transl_table=11;uses_sd=1
k141_9 Prodigal.v2.6.3 CDS 225 680 27.2 - 0 ID=11_1;partial=00;start_type=ATG;rbf_motif=AGGA;rbf_spacer=5-10bp;gc_cont=0.318;conf=99.81;score=27.21;cscore=8.07;sscore=19.14;rscore=9.90;uscore=4.81;tscore=3.56;
k141_9 Prodigal.v2.6.3 CDS 719 1000 4.7 - 0 ID=11_2;partial=00;start_type=ATG;rbf_motif=GGA/GAG/AGG;rbf_spacer=5-10bp;gc_cont=0.309;conf=74.61;score=4.69;cscore=-2.26;sscore=6.95;rscore=2.93;uscore=1.01;tscore=3.56;
k141_9 Prodigal.v2.6.3 CDS 1181 1306 9.2 + 0 ID=11_3;partial=00;start_type=ATG;rbf_motif=AGGAG;rbf_spacer=5-10bp;gc_cont=0.286;conf=89.31;score=9.23;cscore=0.68;sscore=9.91;rscore=7.35;uscore=1.31;tscore=1.75;
k141_9 Prodigal.v2.6.3 CDS 2089 3180 177.4 - 0 ID=11_4;partial=00;start_type=ATG;rbf_motif=GGA/GAG/AGG;rbf_spacer=5-10bp;gc_cont=0.528;conf=100.00;score=177.40;cscore=165.33;sscore=12.07;rscore=2.93;uscore=5.57;tscore=3.56;
k141_9 Prodigal.v2.6.3 CDS 3297 3801 68.6 - 0 ID=11_5;partial=00;start_type=GTG;rbf_motif=AGGAG;rbf_spacer=5-10bp;gc_cont=0.525;conf=100.00;score=68.58;cscore=64.72;sscore=3.86;rscore=9.90;uscore=0.11;tscore=5.50;
k141_9 Prodigal.v2.6.3 CDS 4157 4435 35.8 + 0 ID=11_6;partial=00;start_type=ATG;rbf_motif=GGAGG;rbf_spacer=5-10bp;gc_cont=0.505;conf=99.97;score=35.79;cscore=18.36;sscore=17.43;
```

# Prodigal Stats

```

PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
3954486 gtg875q   20   0   200072   54292   2276 R   93.8   0.1   0:05.06 prodigal
   1 root      20   0   257112   8464    5164 S    0.0   0.0   5:48.46 systemd
   2 root      20   0         0         0        0 S    0.0   0.0   0:05.10 kthreadd
   3 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 rcu_gp
   4 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 rcu_par_gp
   6 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 kworker/70:0H-events_highpri
   9 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 mm_percpu_wq
  10 root     20   0         0         0        0 S    0.0   0.0   0:00.00 rcu_tasks_rude_
  11 root     20   0         0         0        0 S    0.0   0.0   0:00.00 rcu_tasks_trace
  12 root     20   0         0         0        0 S    0.0   0.0   2:26.37 ksoftirqd/0
  13 root     20   0         0         0        0 I    0.0   0.0  14:19.56 rcu_sched
  14 root      rt   0         0         0        0 S    0.0   0.0   0:00.87 migration/0
  15 root      rt   0         0         0        0 S    0.0   0.0   0:00.82 watchdog/0

top - 06:53:15 up 25 days, 1:22,  8 users,  load average: 0.92, 0.63, 0.39
Tasks: 258 total,  2 running, 255 sleeping,   0 stopped,   1 zombie
%Cpu(s):  9.9 us,  0.6 sy,  0.0 ni, 89.5 id,  0.0 wa,  0.0 hi,  0.0 si,  0.0 st
MiB Mem : 64146.0 total, 54089.9 free,  2332.1 used,  7724.1 buff/cache
MiB Swap: 16383.0 total, 15928.9 free,   454.1 used. 60854.1 avail Mem

PID USER      PR  NI    VIRT    RES    SHR S  %CPU  %MEM     TIME+ COMMAND
3954534 gtg875q   20   0   198468   53520   1940 R   93.8   0.1   0:01.01 prodigal
   1 root      20   0   257112   8464    5164 S    0.0   0.0   5:48.46 systemd
   2 root      20   0         0         0        0 S    0.0   0.0   0:05.10 kthreadd
   3 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 rcu_gp
   4 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 rcu_par_gp
   6 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 kworker/70:0H-events_highpri
   9 root      0 -20         0         0        0 I    0.0   0.0   0:00.00 mm_percpu_wq
  10 root     20   0         0         0        0 S    0.0   0.0   0:00.00 rcu_tasks_rude_
  11 root     20   0         0         0        0 S    0.0   0.0   0:00.00 rcu_tasks_trace
  12 root     20   0         0         0        0 S    0.0   0.0   2:26.37 ksoftirqd/0
  13 root     20   0         0         0        0 I    0.0   0.0  14:19.58 rcu_sched
  14 root      rt   0         0         0        0 S    0.0   0.0   0:00.87 migration/0
  15 root      rt   0         0         0        0 S    0.0   0.0   0:00.82 watchdog/0

```

```

#!/bin/bash

# Set the filename for the log file
LOGFILE=tops/prodigal_tops.log

# Clear the log file before we start
> $LOGFILE

# Run top and capture the top 50 lines every 5 seconds
while true; do
    top -b -n 1 | head -n 20 >> $LOGFILE
    sleep 60
done

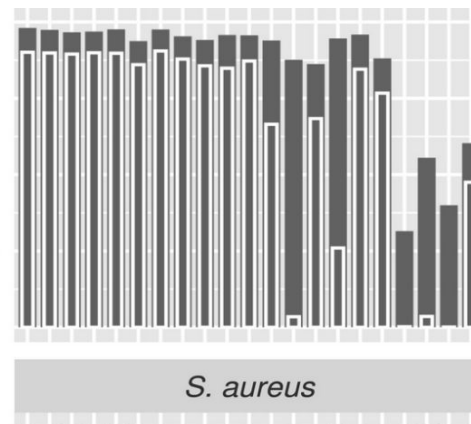
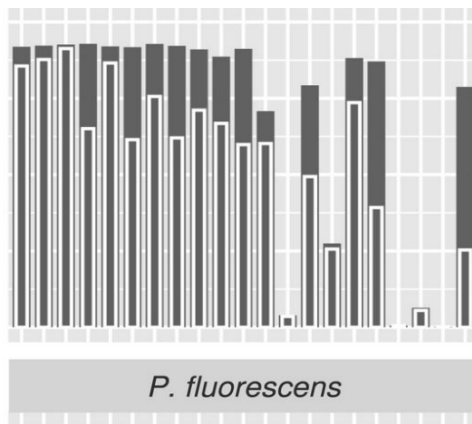
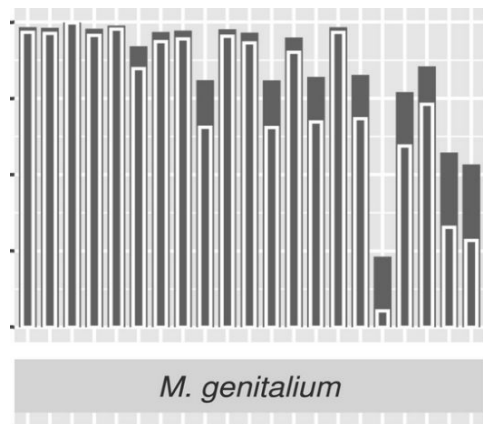
```

# Performance Comparison

Parameters/Tools	GeneMarkS2	GLIMMER	Prodigal
Time	47m 2.262s in real time 46m 14.530s in user-mode 0m 24.795s in system mode	3m 39s and in user-mode 1.63 seconds in system (Kernel) mode	4 min 53.985s in real time 4 min 51.031s inn user-mode 0 min 2.551s in system-mode
CPU Capacity	97%	98%	93.8%

# Tools Comparison

## Prodigal vs GeneMarkS2 vs GLIMMER



- A : Prodigal
- B : GeneMarkS -2
- C : MetaGeneAnnotator
- D : MetaGeneMark
- E : GeneMarkS
- F : MetaGene
- G : GeneMark.hmm/ *E.coli* Model
- H : GeneMark -HA
- I : Augustus/ *E.coli* Model
- J : FragGeneScan
- K : GLIMMER -3



# Bedtools



```
>k141_19:106802-108947
TTAACGCCGCTCGGCGGTGTCAGGTTTTGCTTTTGGTTTTACTATAACGCTGGCAGCGGAGGCTGCTCTATCGGAAAAACAGGTAAGGCATTAGTAAACGCTGCTCGGTAGTTTTTGGTCAACAAACGCTTATCATAAATGACCACCT
CTCCCCGACAGGGCGTGCTACGGATCACTGCCTACCTGCTGAATTAATTAAGAGACGCGCTTGGTAAGCTTTGTACCTCAAAACGGATAACGATTAGGCTTTTTAGCCATTTCGCTTTCGGTGATCAGACCGGGCTGTCGATCGGC
GGGAAGGCAATTTATGGATATGACCTGCGTGCAGCAGCTCGCCTTTAAGATCCAGCCCTTCGGCAAAAGACTGAAGCCCAACAGCAGCTGCGTTCGCCGCTCTCCACCCGTTTACGGTGTAGCTCACCAGCGGGTAACGCGGCTG
GTCGCCCTGGACCAAGGAGGAGTAAGCGCACATCAGCGACATGCTCCAGAAAACGCTGCTCGCCGCCGCTGGCGAACAGTACCAGCATACCATGATGTTTTTGTCTCCAGTTGCTCAGAAAAATAGGCCGCAATTCGGCAATAT
GCTGCTCTTCTGTTATCAATGGTGGTTCGTAGCGCATTGAGGAAATCACCAGTTTCCCTGTTCCACATGATTAAACGGCGAATCCAGCGCCCAAAAGCGGTGCGCGGCTTTTCTTTAAAGCGCTCATTTCTGCGAGACGCGAAAAG
CTATTGAGCGAGGCGAGTGTCCGCGAGGTGACAAATAATGTGCGGTACGCTGCGCCATAAACAATCTTTCCAGCTGATCGCTGACCGGGATACCGACACAATGGAACACAGCGTGGAGTTGCCCTTCGCGTATTCACGCTGCGCCATTT
GGACACCGGCGGCTGAAGACTGCGCCATCGAAGCCAGCGCCATAAATTTATGCGCTTCGAACATTCAGCGCGCGGTTCTATCTGAAAAATACCCGATGCAAAACGCAATAATCTGTCGCGACCCGTTTTTTCGCTGAGATCGT
TGAGAAAGGATTCCGCGACGCCACGAGCGTCTCGGTGAGCTTCGCGAGTCGCTGACAGATTTCCATGACCTCTATCAGGCAACTCGCCCAATTGCAAAACGGTGCTCCGCTCTGCGCTGCGGGCATATACAGATTAAAGATGGCATTG
AGCGAGGCGATCAGTTCTGATACCTCCTCGCAATGGCGTTAAGTCGCTCAGGTTTCGTAACGCGGCGTGGTTTTAGGGCGAAATGCTCCATGCAAGTCCGACCAAGTTGCTGAACAGATCCAGTTGCAAGCGATACCGAGGGG
GGTAATTTCCGCGCTCATCTCCAGCGCATCGCGAGCGACATCGGGCAGATGATGACCTTCATCCAGTACCAACAGCAGATGCTTTGGTTCGGCAGTACGGCTTCACCTTTCCATCGCGGCCATTACCAGCGCGTGATTCGCCACCACGA
CTTCGCGTTCTCTGATTTTCAGCTCGGGCGACAAAGAATGGACATTCGCGGTAATAGTGACAGTTGCGGTTTCAGAGCAACTGGCTTTATCGGATCTTAATCGCCGCCACAGATCGTCGTCAATAGCAATGTCGGTATGATCCGCGACAGCA
TCCCATTTGTAACCGTCAAGATCCCTTTGAGCTGCGCAACGCTTTTGTCTTCCTGGTTGTTAGGCGTCAGTTGCTGATCAAGAAAGGCGAGCAGATCTGCTGCTGAGGCTGCTACTGCGGAGCGCGCCAGGTTTCGCGGACA
CAGGTAACGCCCGCGGCAAGCGGCGGTAAGCGCAGATCGGGGATGATTTTTGCTAGCAGCGGCAAAATCTTTGCTAAAGATCTGATCCTGTAACGCCAGGTTAGCGGTAAGTACGACCAAGTGTGTTTTGCTCTCCGCGGCAATGG
CGATACCGGGAATAAGATAGGAGAGGGTTTTCCGACGCCGCTGCGGCTTAATAGCCAGATGCTCCCTCTTCCCGCGGCGAGCTTGGCGACGTCAGCAATCATCTGCCGCTGCGGCGCACGGGGAATAAGTCGGGATCTGG
TCCTGAAGCGCTTATACAGGCGGCAATTTGCGCTTCAGCGCAGCGGTCAATGCCAT
>k141_19:109154-110516
TTACTGGCAGGCGTGAATTCGCGGGGACGCGCGGACGTGCTTTGTCGCCACGCGGTTTTCCCTCGCTGTGCGCGGTTTTGTTTTCGCTTCGCCGCGCGGTATCTGACGTGCGGCTGCTGACTGCGCGCGCTCCCGCT
GATTTTGTGCGCGTCCAGCAGCGCGCGCGCGCTGCTGACGACGTTCTGAATCGGCTCGCTTTAATCAGCGGTCGCGCTCATAGCCCGGCGTGAATGCGTGGGATCTCTTTTTAGCAACTTTTCGATATCAGCAACAAT
TTATGTTGCTCAACACAGCAGCAGCAGCGCTCGCCAGTTGCCGCGCGCGACCGGTACGGCAATACGTTGATCATAAATCTTCGCGCAGGTTTCGCGAGCTCATAGTTGACCAGTGCAGCGAGCTCTTCAATATCCAGACACGCGC
CGCAATATTCGCTCGCCACAGCAGCAATATCGCGGATTTAAATCGGCCAGCGCGCGGTACGTGCGCTCGCGATTATACGTTGATGCGCGGCTGCGGAATACCGTCTTTATTAAAGCTGTTCCGCGAGATGGTTCGCGCAT
GCTTGGTGGCGGTGAAGACCAACACCTGCTGCCAGTTGCCCTGGCCAAATCAATTTGACAGCAGTTTCGCGCTTACGTTTTTATCGACAAAGTGACGCTGCTGCGTCACCTGCTCTGAGGCGGTGTTGCGGCGGCCACTTCGATTTTC
AGCGGGTTGCGCAACAGCTTTTCGCGAGCGCTTTGATGTCGTCGGAGAAGCTGGCGGAGAACAGCAGATTCTGACGCTTCGCGGCGAGTTTCGCGACACGCGACGGATGTCGTAATAAGCCCATATCCAGCATACGGTTCGGCTTC
ATCCAGTACAGAAATCTCAATCTGATCCAGCTTCACGGCATTTGATGTTCCAGGTGAGCAGAGCTCGCGGCTGCGCACCAGCAGCATCCAGCGCGCGCGGAGTTTATCATCTCGCGATTGATGCTTACGCGCGCAACACCCACA
GCGAGCGAATGTTGAGGTAATTTACTGTAATCTCTGACGTTTTTCGCAATTTGCGCGCGCAGCTGTCGCGGTTGGCGTGAGGATCAACGCGCGCACCGGCGGACGACCTTACGCTGCGGCTGATGGGTAATCAGATGCTGCAACAATGG
AGGGTAAACCTCGCGTTTTACCGGTTCCGGTCTGGGCTGCGCATCAGATCGCGGCTTCAGCAGCGCGGGAATCGCTGTTGCTGAATAGGGGTAGGTTACGGTAACCTGCTCGGCAATGCGCGCAGGATATCAGGTTTAA
TCCAGGGATCAAAAGACAT
>k141_17:32-209
TTACGCTTTTCAACCTGAACGAGGTTCTGATGTGACGGGTTCCCTTCGCCAGCGGAGAAGGACGTTGGGTGCTGAGTACGTTAATACATCCGCCCTGATCCACGCGCTTTGATCCGGGTATACACGCGCTTCTCCAGGGCGA
CCACGCCCGCATCATGCGCGGCTCAC
(group2) [gtg875q@biogenome2023 lem]$ bedtools getfasta -fi CGT2006.fasta -bed ./prodigal_results/cds_CGT2006.gff -fo 2006.fasta
(group2) [gtg875q@biogenome2023 lem]$
```



# Task Delegation

- Jyothi Guruprasad - Background & PPT
- Gautham Krishna Sankar Ramalaxmi - GeneMarkS2
- Kaize H Ali - GLIMMER
- Varsha Srinivasan - Prodigal
- Lee Ellen Mullins - Bedtools & Comparison

# References

- [https://github.com/nigyta/dfast\\_core/blob/master/docs/FAQ.md](https://github.com/nigyta/dfast_core/blob/master/docs/FAQ.md)
- <http://ccb.jhu.edu/software.shtml>
- <https://wiki.gacrc.uga.edu/wiki/GLIMMER-Teaching>
- <https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-11-119>
- <https://github.com/hyattpd/Prodigal>