

## Task 2: Identifying Open Reading Frames of Inserts using both web-based and command-line based GenMarks

# Web-based GenMarks

- Web Site
  - <http://exon.gatech.edu/Genemark/genemarks.cgi>
- Copy data from
  - <https://github.com/EvdH0/ASMworkshop/blob/master/data/sample.fasta>

# GeneMarkS

**John Besemer, Alexandre Lomsadze and Mark Borodovsky**

[GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions.](#)

*Nucleic Acids Research* (2001) 29, pp 2607-2618

This webpage provides accesses to version 4.28 of gene prediction program GeneMarkS. This version combines the original 2001 prokaryotic GeneMarkS with later development, which extended the unsupervised gene prediction to intronless eukaryotes, eukaryotic viruses, phages and EST/cDNA sequences.

[Browse GeneMarkS manual](#)

## Input sequence

**Enter sequence (FASTA or multi FASTA format)**

```
>insert_1
GAATT CATT AAAGAGGAGAAAGGT ACCGGGCCCCCCT CGAGGT CCCT CGGGGAT CGAT CCAGGT GGCAC
TTTT CGGGGAAAT GT GCGCGGAACCCCT ATTT GTTT ATTTTT CT AAAT ACATT CAAAT AT GT AT CCGCT C
AT GAGACAAT AACCTT GAT AAAT GCTT CAAT AAT ATT GAAAAAGGAAGT GT AT GAGT ATT CAACATTT CC
GTGT CGCCCTT ATT CCCTTTT TGCGGCATTTT GCCTT CCT GTTTT TCT CACCCAGAAACGCT GGT GAA
```

**or, upload file:**

Choose File

No file chosen

Action

## Options

Sequence type	Output format for gene prediction	Output options	Optional: results by E-mail
<input checked="" type="radio"/> Prokaryotic <input type="radio"/> Intronless eukaryotic <input type="radio"/> Virus <input type="radio"/> Phage <input type="radio"/> EST/cDNA	<input checked="" type="radio"/> LST <input type="radio"/> GFF	<input checked="" type="checkbox"/> Protein sequence <input checked="" type="checkbox"/> Gene nucleotide sequence  Coding potential graph (not for multi FASTA) <input type="checkbox"/> PDF <input type="checkbox"/> PostScript	E-mail <input type="text"/> Subject GeneMarkS <input type="checkbox"/> Compress files

## Advanced options

- ☒ genetic code 11
- ☐ "TGA" codon as a Tryptophan (not as a stop codon), genetic code 4
- ☐ "TGA" codon as a Glycine (not as a stop codon), genetic code 25
- ☐ Switch off search for gene start related motif(s)

[Contact Us](#) | [Home](#)

# GeneMarkS

**John Besemer, Alexandre Lomsadze and Mark Borodovsky**

[GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions.](#)

*Nucleic Acids Research* (2001) 29, pp 2607-2618

This webpage provides accesses to version 4.28 of gene prediction program GeneMarkS. This version combines the original 2001 prokaryotic GeneMarkS with later development, which extended the unsupervised gene prediction to intron-less eukaryotes, eukaryotic viruses, phages and EST/cDNA sequences.

[Browse GeneMarkS manual](#)

GeneMarkS is running

job ID = genemarks.20170530.150504.24841 ; Estimated run time: 1 second(s)

## Gene Prediction Results

output	link
Coordinates of predicted genes	<a href="#">gms.out</a>
Protein sequences	<a href="#">gms.out.faa</a>
Nucleotide sequence	<a href="#">gms.out.fnn</a>

[Contact Us](#) | [Home](#)

# Command-line GenMarks

- Go to this website
  - <http://exon.gatech.edu/GeneMark/>
- Go Downloads
  - [http://exon.gatech.edu/GeneMark/license\\_download.cgi](http://exon.gatech.edu/GeneMark/license_download.cgi)
  - Choose:
    - GeneMarkS v.4.30
      - LINUX 64
    - Fill in your information
    - Press button “I agree to the terms of this license agreement”

# Download instructions

## GeneMark\* software

If you are an academic, non-profit institution or U.S. Governmental agency, you may use these Products royalty free. All other interested parties should use [this link](#).

Please select software and operating system and fill in other fields below (\* required).

If you are not sure which program fits best to you needs please follow [this link](#) for additional information. GeneMark-ES and GeneMark-ET algorithms are distributed as a single package GeneMark-ES/ET.

Download **Braker-1** from [here](#).

[AUGUSTUS](#) and GeneMark-ET (below) are required for Braker-1. LINUX 32/64 and OS X are supported.

<input type="button" value="Reset"/>	<b>Software*</b>	<b>Operating system*</b>
	<input type="radio"/> GeneMark-ES / ET v.4.33	<input type="radio"/> LINUX 32 <input type="radio"/> LINUX 64 <input type="radio"/> Mac OS X
	<input checked="" type="radio"/> GeneMarkS v.4.30	<input type="radio"/> LINUX 32 <input checked="" type="radio"/> LINUX 64 <input type="radio"/> Mac OS X
	<input type="radio"/> GeneMark.hmm eukaryotic	<input type="radio"/> LINUX 32 <input type="radio"/> LINUX 64 <input type="radio"/> Mac OS X
	<input type="radio"/> MetaGeneMark v.3.38	<input type="radio"/> LINUX 32 <input type="radio"/> LINUX 64 <input type="radio"/> Mac OS X
	<input type="radio"/> ParseRNA seq	<input type="radio"/> LINUX 64
	<input type="radio"/> GeneTack	<input type="radio"/> LINUX 64
	<input type="radio"/> MetaGeneTack	<input type="radio"/> UNIX
	<input type="radio"/> GeneMarkS-T	<input type="radio"/> LINUX 64

<b>Name*:</b>	Mostafa Mostafa Hashim Ellabaan
<b>Institution*:</b>	Copenhagen
<b>Address:</b>	3, Lygten
<b>City:</b>	Copenhagen NV
<b>State:</b>	
<b>Country*:</b>	Denmark
<b>E-mail*:</b>	mostafa.mhashim@gmail.com

If you agree to the terms of the license given below, please confirm by pressing the button and you will receive the download link.

## Academic License Agreement

---

Georgia Institute of Technology  
Atlanta, Georgia, USA

## Download instructions

### **GeneMark\* software**

Thank you for interest to GeneMark family software.

Please download program [here](#)

Please download key [32 bit](#) or [64 bit](#)

---

[Contact Us](#) | [Home](#)

# GenMarks Installation

```
cd ~/Programs
```

```
## we need to install wget to get the data from the web.
```

```
##sudo yum install wget
```

```
## copy link address of the download program here
```

```
## and wget it
```

```
wget <the link from the website after accepting the license >
```

```
##http://topaz.gatech.edu/GeneMark/tmp/GMtool_guX8y/genemark_suite_linux_64.tar.gz
```

```
## copy link address of the key using 64_bit version
```

```
wget <the link from the website after accepting the license > ##http://topaz.gatech.edu/GeneMark/tmp/GMtool_guX8y/gm_key_64.gz
```

```
## extract both the program and the key.
```

```
tar -xvf genemark_suite_linux_64.tar.gz
```

```
gzip -d gm_key_64.gz
```

```
## copy genemarks key to the home directory
```

```
cp gm_key_64 ~/.gm_key
```

```
cd genemark_suite_linux_64/gmsuite
```

```
ls
```



# Running genemark

```
## to see options available just write this
```

```
cd ~/Data
```

```
## we need git to get the data from the workshop
```

```
sudo yum install git
```

```
## get the data from the workshop
```

```
git clone https://github.com/EvdH0/ASMworkshop
```

```
cp ASMworkshop/data/sample.fasta .
```

```
gmsuite=~/Programs/genemark_suite_linux_64/gmsuite
```

```
$gmsuite/gmhmp -m $gmsuite/MetaGeneMark_v1.mod \
```

```
-A orfs.protein.fa -D orfs.nucleotide.fa -o all.orfs.result sample.fasta
```

```
cat orfs.nucleotide.fa
```

```
cat orfs.protein.fa
```

```
## to know how many orfs are there
```

```
grep ">" orfs.protein.fa | wc -l
```

```
## or
```

```
grep ">" orfs.nucleotide.fa | wc -l
```