

# Gene Prediction & Filtering

Qiaoshan Lin

## Abstract

**Citation:** Qiaoshan Lin Gene Prediction & Filtering. **protocols.io**

dx.doi.org/10.17504/protocols.io.k7sczne

**Published:** 10 Dec 2017

## Protocol

### Predict genes by BRAKER1

#### Step 1.

```
/home/CAM/qlin/LF10_Genome_Annotation/BRAKER/t0
```

```
/home/CAM/qlin/LF10_Genome_Annotation/BRAKER/t1
```

```
/home/CAM/qlin/LF10_Genome_Annotation/BRAKER/t2
```

### Separate complete genes from BRAKER1

#### Step 2.

```
/home/CAM/qlin/scripts/geneFilter4gtf.pl
```

```
/home/CAM/qlin/LF10_Genome_Annotation/BRAKER/t1/braker/lewisii/t2*
```

### Align complete genes to M.guttatus proteins by DIAMOND

#### Step 3.

```
/home/CAM/qlin/LF10_Genome_Annotation/diamond/t2
```

### Get protein sequences from the protein alignment file

#### Step 4.

```
/home/CAM/qlin/scripts/fetchSeq4fa2.pl
```

```
/home/CAM/qlin/LF10_Genome_Annotation/diamond/t2/t2.match.txt
```

```
/home/CAM/qlin/LF10_Genome_Annotation/diamond/t2/t2.match.aa
```

```
/home/CAM/qlin/LF10_Genome_Annotation/diamond/t2/t2.unmatch.aa
```

### Align the guttatus-protein-matched genes and unmatched genes to Pfam-A.hmm

#### Step 5.

```
/home/CAM/qlin/LF10_Genome_Annotation/hmm/t2*
```

Get t2.unmatch.aa from t2.unmatch.domtblout

**Step 6.**

/home/CAM/qlin/scripts/domtbloutFetch.pl

/home/CAM/qlin/LF10\_Genome\_Annotation/hmm/t2.unmatch.aa

Separate t2.unmatch.aa into mono- or multi-exonic genes

**Step 7.**

/home/CAM/qlin/LF10\_Genome\_Annotation/hmm/t2.unmatch.mono.aa

/home/CAM/qlin/LF10\_Genome\_Annotation/hmm/t2.unmatch.multi.aa

Align guttatus-unmatched proteins with Pfam domains to

/isg/shared/databases/Diamond/RefSeq/plant.protein.faa.85.dmnd database by diamond

**Step 8.**

/home/CAM/qlin/LF10\_Genome\_Annotation/diamond/t2\_plants

Separate into mono-exonic or multi-exonic

**Step 9.**

Check if the corresponding guttatus proteins of the 959 matched but no Pfam domain proteins have Pfam domains or not

**Step 10.**