



MAKER
Annotate this!

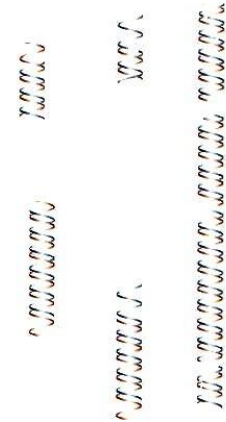
MAKER Annotation Suite

What, Why and How?

Why (is **MAKER** useful)?



Sequencing project



```
GTGAGGAAGGGGAGGGTTTTCTGGCCGGCTTGAGACCTTAAGGGCGAAGACCTGGACCCCG
CCCCGGCCCGCTGAGGGCGAGCCTTGGCCCTTGAGCCCGGAGACTATCCCGGCATCGGGATGTGGGA
GGCCCTCTACCCGGGGAGCGGAGGGATGGATCCGGSHGGCCCTACTCTGTGGGGCCCTCTAGGCC
CTCAAGGATGGGGAGACCGGAGGGGCTGACCCCTGGCCCTTAAGGGGAAAAGCGGGGGATTCCG
CCAGTGTGGAGCATCGTTTTCTGGCCCTTTGGATTGGGACCGGGACCAAGATGCTTCCGACTAAGCA
AAGCGTGTCTGTAGAGGGGGTTTTGACTCGGGGCTTTTGGCAGACTCTCGCCCTGTTTACG
CCCGACCGGTGGTCCAGGCGCTTGGCGAGGGHTTTACGGGTGAGGAGGAGGAGAGGGAGG
TGAGGCATGA
```



```
GTGAGGAAGGGGAGGGTTTTCTGGCCGGCTTGAGACCTTAAGGGCGAAGACCTGGACCCCG
CCCCGGCCCGCTGAGGGCGAGCCTTGGCCCTTGAGCCCGGAGACTATCCCGGCATCGGGATGTGGGA
GGCCCTCTACCCGGGGAGCGGAGGGATGGATCCGGSHGGCCCTACTCTGTGGGGCCCTCTAGGCC
CTCAAGGATGGGGAGACCGGAGGGGCTGACCCCTGGCCCTTAAGGGGAAAAGCGGGGGATTCCG
CCAGTGTGGAGCATCGTTTTCTGGCCCTTTGGATTGGGACCGGGACCAAGATGCTTCCGACTAAGCA
AAGCGTGTCTGTAGAGGGGGTTTTGACTCGGGGCTTTTGGCAGACTCTCGCCCTGTTTACG
CCCGACCGGTGGTCCAGGCGCTTGGCGAGGGHTTTACGGGTGAGGAGGAGGAGAGGGAGG
TGAGGCATGA
```

Sequence without annotations

```
GTGAAGGAAGGGGCAAGGGTTTTTCTGCCGCGGCTTGAGGACCTTAAAGGGCGCAAGACCTGGACCCCGG  
CCCGGGCCGCOCTGAGGCGGAGCCTGGCCTTTGAGCCCGGGACCTATCCCCGGCCATCGGGATGGTGGA  
GCCCCTTCTACCCGGGGAAGCGGAGGGATGGATCCGGGAGGCCTACTACCTGGTGGCGGCOCTCTACGCC  
CTCAAGGATGGGGAGCACCGGGAAGGGCGTACCOCTGGCCCAGGCCCTAAGGGAAAAGGCGGGGGATTCCG  
CCAGTGTGGAGCATCGTTTTCTGGCCCTTTTGATGCGGACCGGGACCAGATCGCTTTCGACTAAGGCA  
AAGCGTGGTCTGTAGAAAGGGGTTTTGACTTCGCGCGGCTTTTGGCAGACCTGCTGGCCTGGTTTAGC  
CCGGACCGGTGGGTGCAGGCCCGTTGGGCGAGGGAGTTTTACGGCGTGAAGGAAGAGGAGAAGGAGGAGG  
TGAAGGCATGA
```

It's just letters!

What (is MAKER)?



Annotation ways

Manual:

Naming features by hand
SLOWEST

With a program:

Need to install it and know about all the options for it
FASTER & EASIER

With a pipeline/suite:

A single program that runs many programs. Need very little tinkering.
FASTEST & EASIEST

Pipelines

Pipeline = A single program that runs many programs.

Many exist: BLAST2GO, ANNOVAR, MAKER, DIYA etc.

Most are hard to use, but easier than figuring out all components

Blast2Go screen example

Application Messages GO Graphs Blast/IPS Re

C04018A02

Query name (length): C04018A02 (600) Blast Version: BLASTX 2.2.28+ Database: Blast Program: References:

E-value cutoff: 0.001 Filters: L: Blast Program: Blast DB: Blast ExpectValue: Number of Blast Hits: Save results as ... Blast Desc. Annotator

Sequences producing significant alignments

Gene Name	AC
qii2213425[emb]CAB09799.1[hypothetical protein [Citrus x paradisi]]	CAB09799
qii557552355[gb]ESR62984.1[hypothetical protein CICLE_v10016136mg [Citrus clementina]]	ESR62984
qii557552356[gb]ESR62985.1[hypothetical protein CICLE_v10016136mg [Citrus clementina]]	ESR62985
qii557552357[gb]ESR62986.1[hypothetical protein CICLE_v10016136mg [Citrus clementina]]	ESR62986
qii462407619[gb]EMJ12953.1[hypothetical protein PRUPE_ppa009462mg [Prunus persica]]	EMJ12953
qii557552354[gb]ESR62983.1[hypothetical protein CICLE_v10016134mg [Citrus clementina]]	ESR62983
qii224078584[ref]XP_002305564.1[unpredicted protein [Populus trichocarpa]]	XP_002305564
qii22344529[gb]EE66075.1[cytochrome b, homology family protein [Populus trichocarpa]]	EE66075
qii359483362[ref]XP_002273346.2[PREDICTED: putative lipoxygenase [Vitis vinifera]]	XP_002273346
qii302144130[emb]CBI23235.3[unannotated protein product [Vitis vinifera]]	CBI23235
qii388493080[gb]AFK34606.1[unknown [Medicago truncatula]]	AFK34606
qii470114187[ref]XP_004293298.1[PREDICTED: putative lipoxygenase [Vitis vinifera]]	XP_004293298
qii35655674[ref]XP_003546155.1[PREDICTED: putative lipoxygenase [Glycine max]]	XP_003546155
qii356531939[ref]XP_00354533.1[PREDICTED: putative lipoxygenase [Glycine max]]	XP_00354533
qii460373807[ref]XP_004232705.1[PREDICTED: putative lipoxygenase [Solanum lycopersicum]]	XP_004232705
qii211906514[gb]ACJ11750.1[lipoxygenase [Gossypium hirsutum]]	ACJ11750
qii502078142[ref]XP_004486167.1[PREDICTED: putative lipoxygenase isoform X2 [Cicer arietinum]]	XP_004486167
qii502078139[ref]XP_004486166.1[PREDICTED: putative lipoxygenase isoform X1 [Cicer arietinum]]	XP_004486166
qii388520197[gb]AFK48160.1[unknown [Lotus japonicus]]	AFK48160
qii357480325[ref]XP_003610448.1[lipoxygenase [Medicago truncatula]]	XP_003610448
qii355511603[gb]AFS92645.1[lipoxygenase [Medicago truncatula]]	AFS92645

Status: 31446, Connected to public database: b2g_sep13 Memory usage: 179MB of 966MB Blasting 83%

Blast Configuration

BLAST already running, please wait...

Basic Advanced

Important note:
Via this function you communicate directly with the NCBI BLAST service. Please use this service in a responsible fashion, identify yourself providing your email address and do not run BLAST searches in parallel. Any issues regarding the performance or obtained results depend on the NCBI BLAST. Questions regarding the NCBI BLAST service can be sent to: blast-help@ncbi.nlm.nih.gov

Your e-mail (for NCBI Blast):

Blast Program:

Blast DB:

Blast ExpectValue:

Number of Blast Hits:

Save results as ...
 xml text html

Blast Desc. Annotator:

Waiting 49 seconds for 98RJKEM015
 Waited 49 seconds for 98RJKEM015
 Waited 49 seconds for 98RJKEM015
 Waited 49 seconds for 98RJKEM015
 Waited 49 seconds for 98RJKEM015
 7-----request removed
 7-----request removed
 7-----request removed
 7-----request removed
 Waited 79 seconds for 98RJKEM015
 7-----request removed

MAKER Description

Ab initio gene prediction	EST alignment	Protein alignment	Repeats	Small RNA
SNAP	Exonerate	Exonerate	Repeatmasker	tRNAscan-SE
Augustus	BlastN	BlastX	Repeatrunner	snoscan
FGENESH*		tBlastX		
GeneMark-ES				
GeneMark-S				

*Commercial product

How (do we use **MAKER**)?



Overview

- 1) *Log onto mobaXterminal
- 2) *Load pre-installed modules
- 3) Create configuration file
- 4) Edit configuration file
- 5) Run a short MAKER test
- 6) Check the results
- 7) Run MAKER for real
- 8) Wait....
- 9) Wait....
- 10) Check the results
- 11) Ponder the results!**

Running MAKER is easy

```
module load maker
```

```
maker -CTL
```

```
maker maker_bopts.ct1 maker_exe.ct1 maker_opts.ct1
```

About the files

Maker_exe.ctl

Installation paths for all software; only relevant if installing

Maker_bopts.ctl

Settings for all prediction programs like BLAST; feel free to tinker around

Maker_opts.ctl

Input files (like genome) and analysis to run. **MUST** be tinkered with.

What we won't cover

Downloading and installing MAKER:

http://yandell.topaz.genetics.utah.edu/cgi-bin/maker_license.cgi

Linux experience necessary!

Extra resources:

http://gmod.org/wiki/MAKER_Tutorial

Don't just click and go

Problems are NOT obvious:

Did we compare it with enough data?

Was the criteria for matches too strict or too loose?

Was there something related to the type of species?

Prokaryotic/eukaryotic/archaea?

GC-content?

Repetitive segment content?

Etc.

Take-away message

Individual programs

Harder to start, error checking is fast
Gives very specific predictions

Pipeline

Easy to start, takes long to check for errors
Solution: Test on a portion of the data!

Gives many types of predictions
Solution: Crop the output, or make sure you know what everything is.

HPC Maker

Very few prediction programs

Data is just as an example

Runs 5-10 minutes

If you finish early

1. Play with the .ctl files (bopts, opts)
2. See if you can run MAKER with very small files (<20 MB) from NCBI uploaded to HPC
3. If you sit on a LINUX pc and have time, try installing MAKER – It's relatively easy