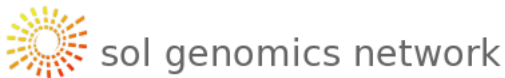


# Slicing and Dicing Plant Genomes with Perl

Jonathan "Duke" Leto  
Sol Genomics Network  
<http://solgenomics.net>



# What is Sol Genomics Network? (SGN)

SGN is an NSF-funded Model Organism Database on the web.

- A model organism is a species that is extensively studied to understand particular biological phenomena, with the expectation that discoveries made in the organism model will provide insight into the workings of other organisms.

# What is Sol Genomics Network? (Continued)

SGN is part of the Lukas Mueller research lab at Boyce Thompson Institute for Plant Research (BTI), located on the Cornell University campus.



We study the Solanaceae (Nightshade family): Tomato, Potato, Petunia, Eggplant, and many more

## Who uses SGN?

- Evolutionary plant biologists
- Plant biochemists
- Plant disease researchers
- Plant breeders



# Code By Numbers

Every line of code goes on Github, anything reusable goes to CPAN.

- 29 public Github repos
- 10 committers
- ~ 10 CPAN modules released (more on the way)
- Over 200K lines of Perl
  - Lots of legacy code, slowly modernizing
  - Converted to Catalyst last year

## Some CPAN Modules

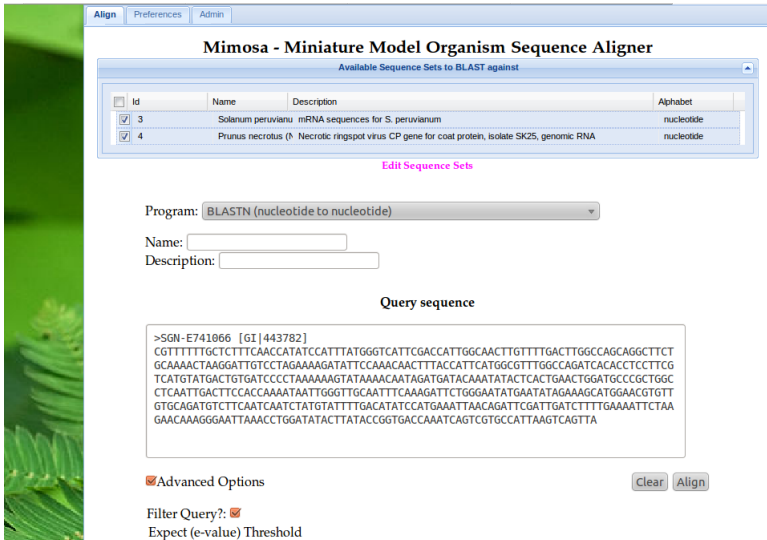
- Bio::Chado::Schema - DBIx::Class schema for Chado
- Bio::GFF3 - Manipulate GFF3 data files
- Bio::Blast - Interact with BLAST databases
- Catalyst::View::Bio::Seqio - Catalyst view for Bioperl sequences
- Yapri - Yet Another Perl R Interface

# What we depend on most

- Moose
- Catalyst
- DBIx::Class
- Plack
- Starman
- BioPerl
- Mason
- WWW::Mechanize
- autodie
- Test::\*

# My current project

## Mimosa: Miniature Model Organism Sequence Aligner

The image shows a screenshot of the Mimosa web interface. The interface has a green background on the left side with a close-up image of a plant. The main content area is white. At the top, there are three tabs: 'Align', 'Preferences', and 'Admin'. Below the tabs is the title 'Mimosa - Miniature Model Organism Sequence Aligner'. Underneath the title is a section titled 'Available Sequence Sets to BLAST against'. This section contains a table with columns: 'Id', 'Name', 'Description', and 'Alphabet'. There are two rows in the table, both of which are selected with checkboxes. The first row has Id '3', Name 'Solanum peruvianum', Description 'mRNA sequences for S. peruvianum', and Alphabet 'nucleotide'. The second row has Id '4', Name 'Prunus necrotus (N Necrotic ringspot virus CP gene for coat protein, isolate SK25, genomic RNA', and Alphabet 'nucleotide'. Below the table is a link 'Edit Sequence Sets'. Further down, there is a 'Program:' dropdown menu set to 'BLASTN (nucleotide to nucleotide)'. Below that are input fields for 'Name:' and 'Description:'. Then, there is a section titled 'Query sequence' which contains a text box with a long DNA sequence. At the bottom, there is a checkbox for 'Advanced Options' which is checked, and two buttons: 'Clear' and 'Align'.

Align Preferences Admin

### Mimosa - Miniature Model Organism Sequence Aligner

Available Sequence Sets to BLAST against

<input type="checkbox"/>	Id	Name	Description	Alphabet
<input checked="" type="checkbox"/>	3	Solanum peruvianum	mRNA sequences for S. peruvianum	nucleotide
<input checked="" type="checkbox"/>	4	Prunus necrotus (N Necrotic ringspot virus CP gene for coat protein, isolate SK25, genomic RNA		nucleotide

[Edit Sequence Sets](#)

Program: BLASTN (nucleotide to nucleotide)

Name:

Description:

#### Query sequence

```
>SGN-E741066 [GI|443782]
CGTTTTTTGCTCTTTCAACCATATCCATTTATGGGTCATTGACCATTTGGCAACTGTTTTGACTTGGCCAGCAGGCTTCT
GCAAACTAAGGATTGCTCTAGAAAAGATATTCAAACAACCTTTACCATTATGCGGCTTTGGCCAGATCACACCTCCTTCG
TCATGTATGACTGTGATCCCTAAAAAAGTATAAAACAATAGATGATACAAATATACTCACTGAACCTGGATGCCCGCTGGC
CTCAATTGACTTCCACCAAAATAATTGGGTGCAATTTCAAGATTCTGGGAATATGAATATAGAAAGCATGGAACGTGTT
GTGCAGATGCTTCAATCAATCTATGATTTTGACATATCCATGAAATTAACAGATTTCGATTGATCTTTGAAAAATCTAA
GAACAAAGGGAATTAACCTGGATATACTTATACCGGTGACCAATCAGTCGTGCCATTAAGTCAGTTA
```

☒ Advanced Options

Filter Query?: ☒

Expect (e-value) Threshold

Clear Align



# Resources

- <http://solgenomics.net>
- <http://github.com/solgenomics>
- #cxgn on irc.perl.org

# Thanks!

- Boyce Thompson Institute for Plant Research
- National Science Foundation
- All the Perl + CPAN hackers that make SGN possible

