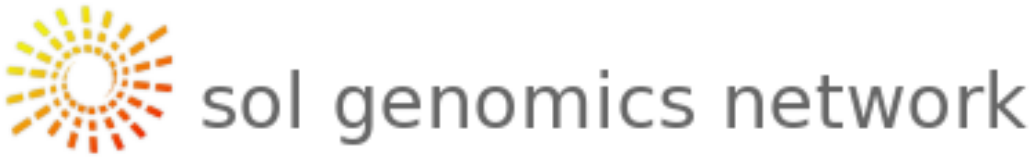
Slicing and Dicing Plant Genomes with Perl

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



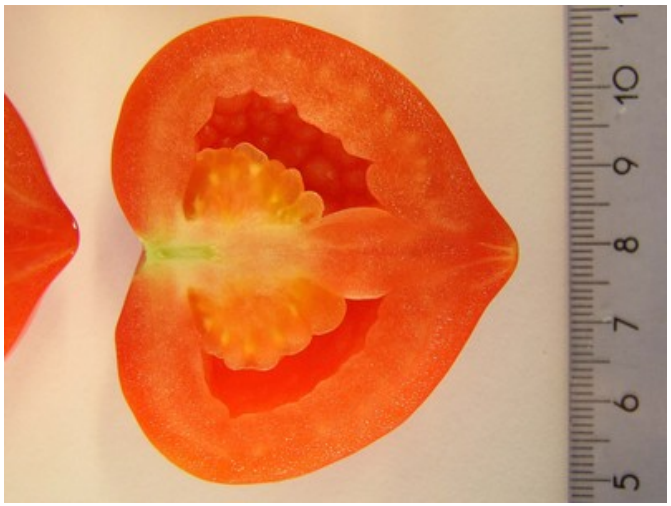
What is Sol Genomics Network? (SGN)

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
- $\sim$ 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

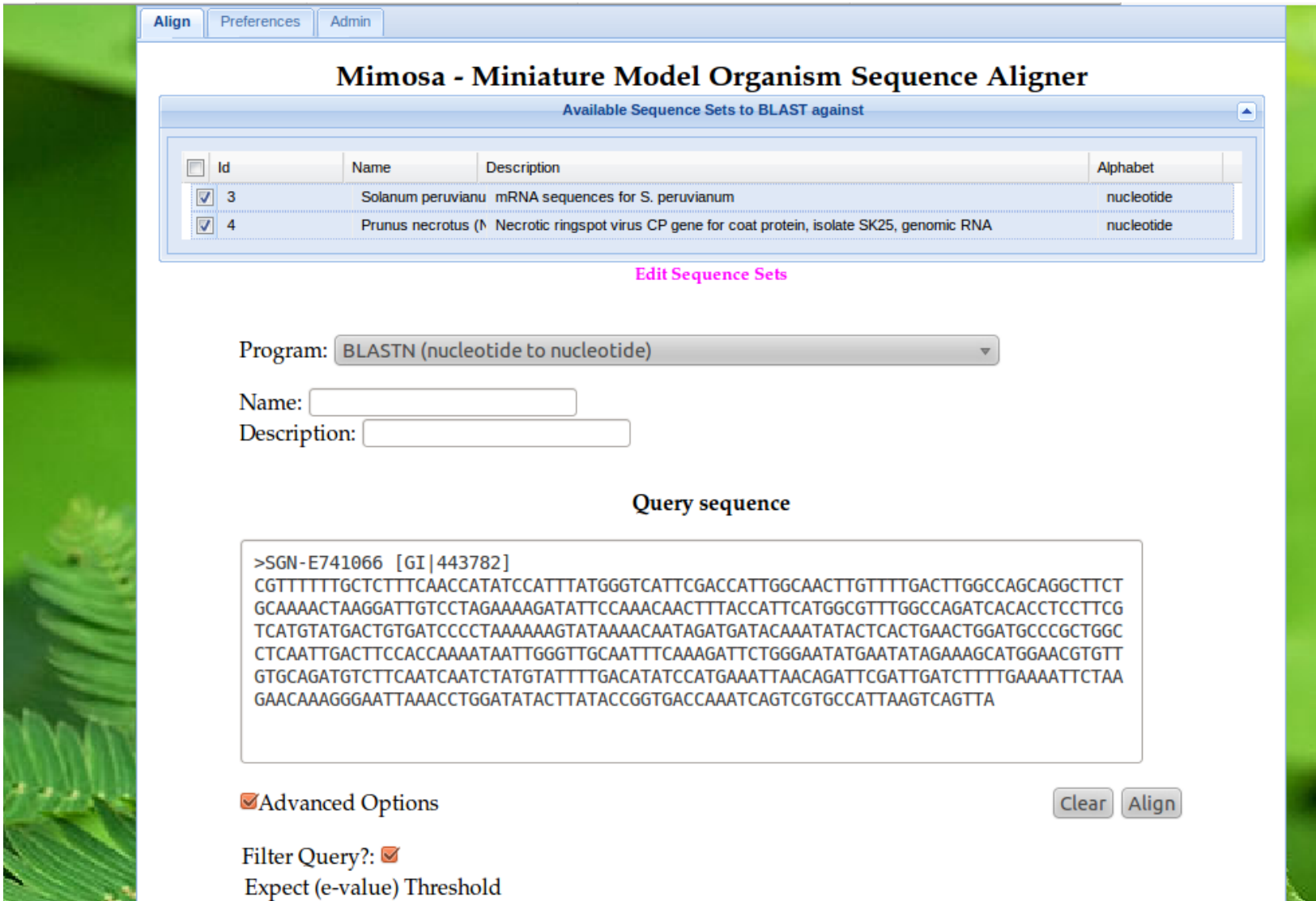
- Bio::Chado::Schema - DBIx::Class schema for Chado
- Bio::GFF3 - Manipulate GFF3 data files
- Bio::Blast - Interact with BLAST databases
- Catalyst::View::Bio::Seqio - Catalyst view for Bioperl sequences
- Yapri - Yet Another Perl R Interface

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



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