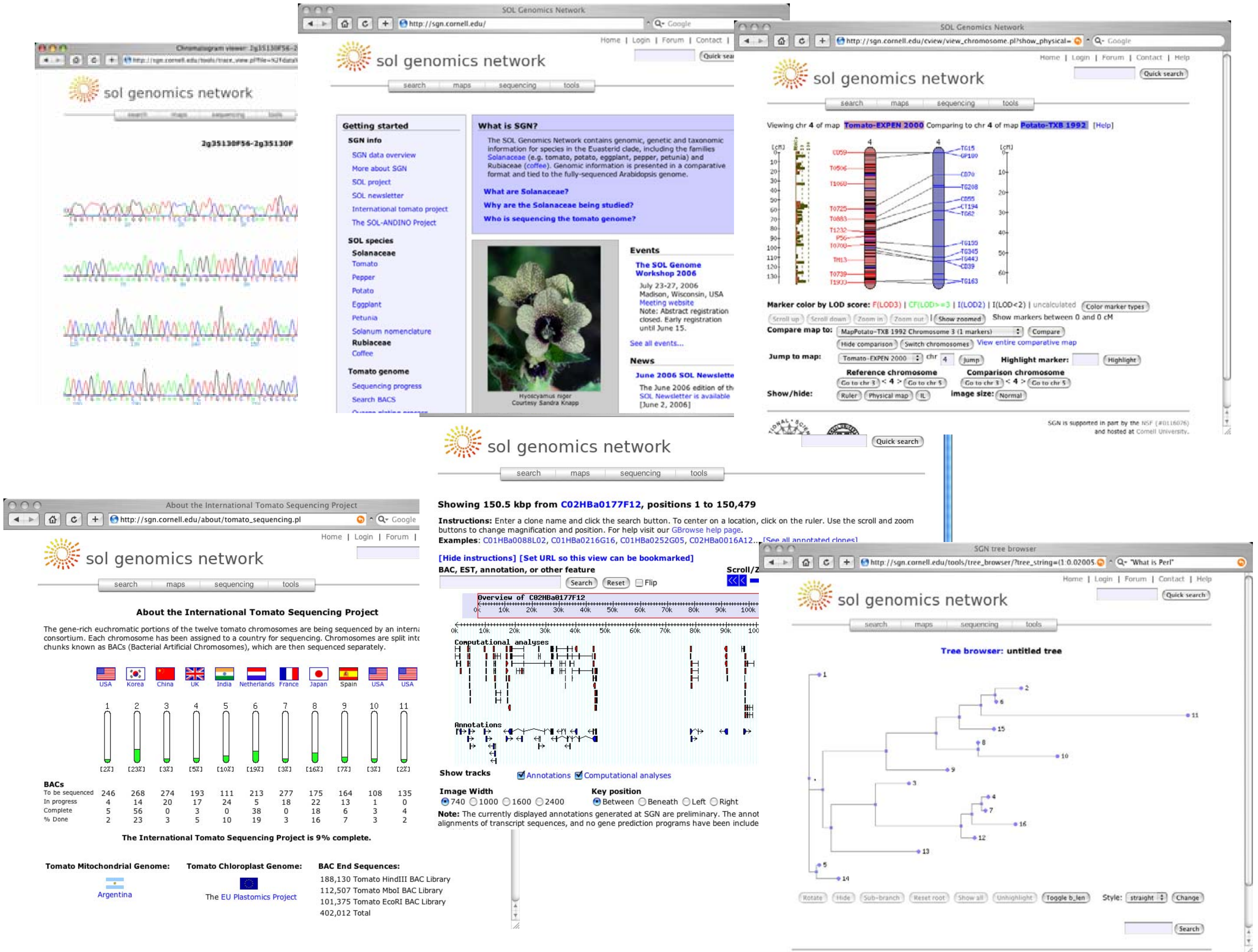
PerlCyc

Pathway Tools Workshop
SRI, Menlo Park
June 12-16, 2006

Lukas Mueller
Cornell University
http://sgn.cornell.edu/

sol genomics network

- Database of Solanaceae plants and related organisms
- tomato, potato, eggplant, pepper, petunia
- tobacco, Atropa belladonna
- Coffee
- Dealing with many organisms: Clade Oriented Database (COD)
- Data types: Genetic maps, EST sequences, unigene builds, gene families, pathway data, tomato genome data



Chromatogram viewer: 2g35130F56-2

sol genomics network

search maps sequencing tools

2g35130F56-2g35130F

SOL Genomics Network

http://sgn.cornell.edu/

Home | Login | Forum | Contact |

sol genomics network

search maps sequencing tools

Getting started

SGN info

SGN data overview
More about SGN
SOL project
SOL newsletter
International tomato project
The SOL-ANDINO Project

SOL species

Solanaceae

Tomato
Pepper
Potato
Eggplant
Petunia
Solanum nomenclature

Rubiaceae

Coffee

Tomato genome

Sequencing progress
Search BACS

What is SGN?

The SOL Genomics Network contains genomic, genetic and taxonomic information for species in the Euasterid clade, including the families Solanaceae (e.g. tomato, potato, eggplant, pepper, petunia) and Rubiaceae (coffee). Genomic information is presented in a comparative format and tied to the fully-sequenced Arabidopsis genome.

What are Solanaceae?
Why are the Solanaceae being studied?
Who is sequencing the tomato genome?

Hyoscyamus niger
Courtesy Sandra Knapp

Events

The SOL Genome Workshop 2006
July 23-27, 2006
Madison, Wisconsin, USA
Meeting website
Note: Abstract registration closed. Early registration until June 15.

See all events...

News

June 2006 SOL Newsletter
The June 2006 edition of th
SOL Newsletter is available
[June 2, 2006]

SOL Genomics Network

http://sgn.cornell.edu/cview/view_chromosome.pl?show_physical=

Home | Login | Forum | Contact | Help

sol genomics network

Quick search

search maps sequencing tools

Viewing chr 4 of map Tomato-EXPEN 2000 Comparing to chr 4 of map Potato-TXB 1992 [Help]

Marker color by LOD score: F(LOD3) | CF(LOD>=3 | I(LOD2) | I(LOD<2) | uncalculated [Color marker types]

[Scroll up] [Scroll down] [Zoom in] [Zoom out] [Show zoomed] Show markers between 0 and 0 cM

Compare map to: MapPotato-TXB 1992 Chromosome 3 (1 markers) [Compare]

[Hide comparison] [Switch chromosomes] View entire comparative map

Jump to map: Tomato-EXPEN 2000 chr 4 [Jump] Highlight marker: [Highlight]

Reference chromosome: [Go to chr 3] < 4 > [Go to chr 5] Comparison chromosome: [Go to chr 3] < 4 > [Go to chr 5]

Show/hide: [Ruler] [Physical map] [IL] image size: [Normal]

SGN is supported in part by the NSF (#0116076) and hosted at Cornell University.

About the International Tomato Sequencing Project

http://sgn.cornell.edu/about/tomato_sequencing.pl

Home | Login | Forum |

sol genomics network

search maps sequencing tools

About the International Tomato Sequencing Project

The gene-rich euchromatic portions of the twelve tomato chromosomes are being sequenced by an interna consortium. Each chromosome has been assigned to a country for sequencing. Chromosomes are split into chunks known as BACs (Bacterial Artificial Chromosomes), which are then sequenced separately.

	USA	Korea	China	UK	India	Netherlands	France	Japan	Spain	USA	USA
	1	2	3	4	5	6	7	8	9	10	11
	[2%]	[23%]	[3%]	[5%]	[10%]	[19%]	[3%]	[16%]	[7%]	[3%]	[2%]
BACs											
To be sequenced	246	268	274	193	111	213	277	175	164	108	135
In progress	4	14	20	17	24	5	18	22	13	1	0
Complete	5	56	0	3	0	38	0	18	6	3	4
% Done	2	23	3	5	10	19	3	16	7	3	2

The International Tomato Sequencing Project is 9% complete.

Tomato Mitochondrial Genome: Argentina

Tomato Chloroplast Genome: The EU Plastomics Project

BAC End Sequences:
188,130 Tomato HindIII BAC Library
112,507 Tomato MboI BAC Library
101,375 Tomato EcoRI BAC Library
402,012 Total

sol genomics network

Quick search

search maps sequencing tools

Showing 150.5 kbp from C02HBa0177F12, positions 1 to 150,479

Instructions: Enter a clone name and click the search button. To center on a location, click on the ruler. Use the scroll and zoom buttons to change magnification and position. For help visit our GBrowse help page.

Examples: C01HBa0088L02, C01HBa0216G16, C01HBa0252G05, C02HBa0016A12... [See all annotated clones]

[Hide instructions] [Set URL so this view can be bookmarked]

BAC, EST, annotation, or other feature [Search] [Reset] Flip

Overview of C02HBa0177F12

Computational analyses

Annotations

Show tracks ☑ Annotations ☑ Computational analyses

Image Width [● 740] 1000 1600 2400

Key position [● Between] Beneath Left Right

Note: The currently displayed annotations generated at SGN are preliminary. The annot alignments of transcript sequences, and no gene prediction programs have been include

SGN tree browser

http://sgn.cornell.edu/tools/tree_browser/?tree_string=(1:0.02005

Home | Login | Forum | Contact | Help

sol genomics network

Quick search

search maps sequencing tools

Tree browser: untitled tree

[Rotate] [Hide] [Sub-branch] [Reset root] [Show all] [Unhighlight] [Toggle b_len] Style: straight [Change]

[Search]

PerlCyc

- **What is it?**
  - **A Perl interface for Pathway Tools**
- **What is Perl?**
  - **Practical Extraction and Reporting Language**
  - **Language of the “C” family, developed from UNIX shell languages (yes, it's ugly!)**
  - **Excellent text handling capabilities**
  - **Object oriented models available**
  - **Popular in web programming and bioinformatics**

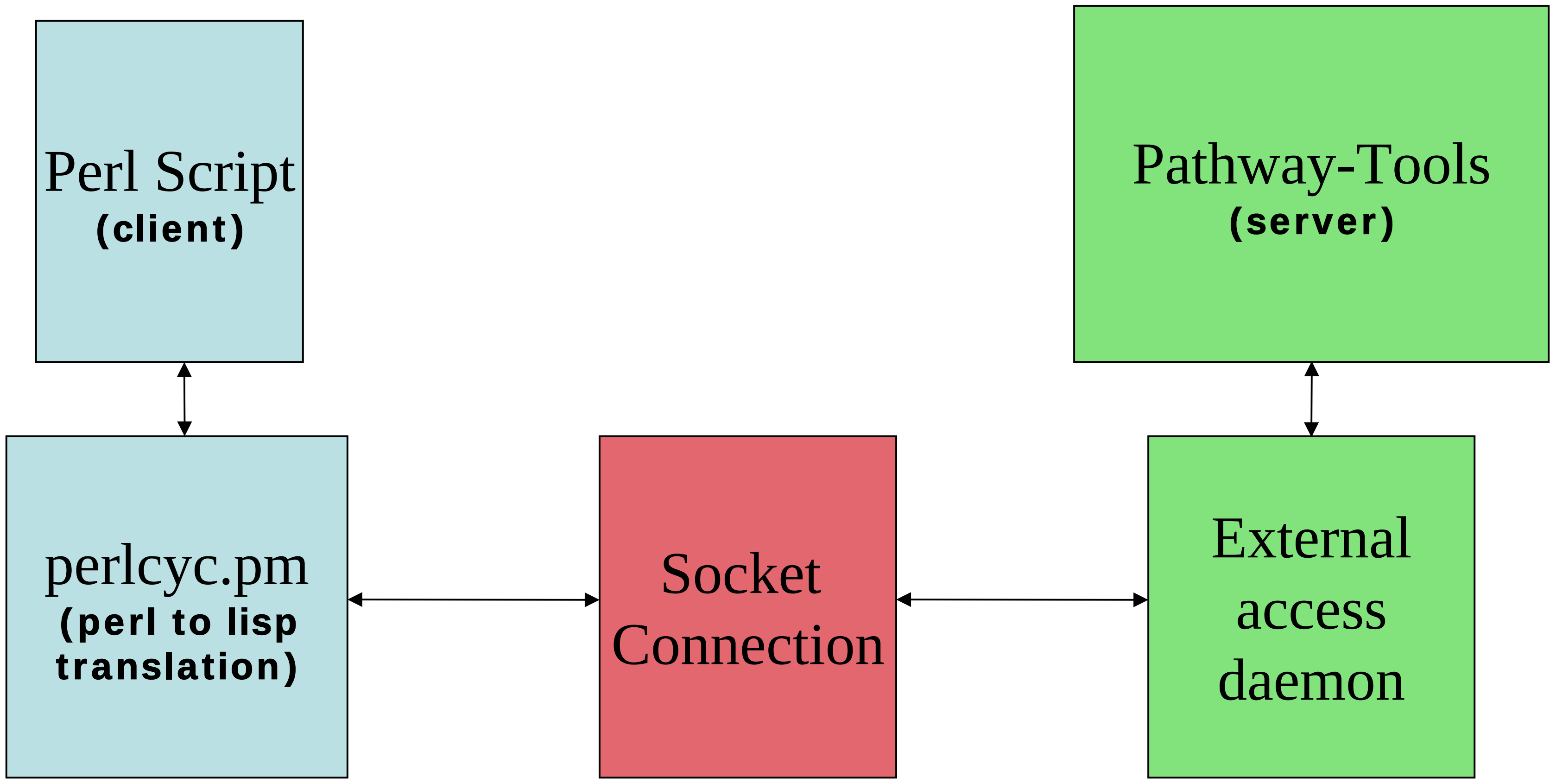
Motivation

- Only few bioinformaticians know Lisp
- Make Pathway-Tools more accessible to the bioinformatics community
- Most bioinformaticians know Perl and/or Java.

Perlcyc features

- Allows to make calls to GFP and Pathway Tool functions from Perl
- Simple implementation
  - Written as a easy-to-use Perl module using general Object Perl conventions
- Secure
  - User needs access to file system

PerlCyc implementation



PerlCyc API

- Implements both
  - Generic Frame Protocol (GFP) functions
  - Pathway Tools functions

- Function name conventions:
  - Replace dashes with underlines
  - Replace question marks with ‘_p’

Data type equivalents

- **Note: Perl is not strongly typed!**
  - **Lisp list = Perl array**
  - **Lisp string = Perl scalar**
  - **Lisp symbol = Perl scalar**
  - **Lisp boolean (“NIL” | “t”) = Perl scalar (NIL|t)**
  - **Lisp integer, etc = Perl scalar**

PerlCyc limitations

- Only one PerlCyc script can be running at a time (socket connection)
- Works only on UNIX
- Optional parameters to functions are not supported
- Certain functions may not be available
- It's just a thin layer - frames are not implemented in Perl

Installing and running PerlCyc

- Requirements:
  - UNIX installation of Pathway Tools
  - Perl 5.6 or later
- Installation
  - Download PerlCyc from TAIR
    http://arabidopsis.org/biocyc/perlcyc/
  - Follow installation instructions
  - Start pathway-tools using the -api option
  - Write and run PerlCyc perl scripts

Usage examples

```
#!/usr/bin/perl
use strict;
use perlcyc;
my $cyc = perlcyc -> new(“ARA”);
my @result = $cyc -> most_pathway_tools_functions();
# do something with results...
# ....
# at the end of script, clean up
$cyc -> close();
```

Sample Scripts

```
#!/usr/bin/perl
use strict;
use perlcyc;

my $cyc = perlcyc -> new(“ARA”);
my @reactions = $cyc -> all_rxns();

foreach my $reaction (@reactions) {
     print "$reaction\n";
     my $rname = $cyc -> get_slot_value($reaction, “in-pathwa
     print "$rname\n";
}
$cyc->close();
```

```
#!/usr/bin/perl
use strict;
use perlcyc;
my $cyc = perlcyc -> new(“ARA”);
foreach my $g ($cyc->get_class_all_instances("|Genes|"))
  {
   my $dblink = $cyc -> get_slot_value($g, "DBLINKS");
   print "DBLINKS: $dblink\n";
}
```

```
use strict;
my $added =0;  my $genesprocessed=0;
my $cyc = perlcyc -> new ("ARA");
my @genes = $cyc -> get_class_all_instances ("|Genes|");
print "Adding TAIR links...\n";
foreach my $g (@genes) {
     $genesprocessed++;
     my $common_name = $cyc -> get_slot_value($g,
        "common-name");
     if ($common_name && ($common_name ne "NIL")) {
               $cyc -> put_slot_value ($g, "dblinks", "(TAIR:
                                             \"$common_name\")");
           $added++;
     }
      if ($genesprocessed % 1000 == 0) { print
    "$genesprocessed\r";}
}
print "Processed $genesprocessed genes and added
```

JavaCyc

- **A similar interface for Java**
- **Written by Thomas Yan (TAIR)**
- **Available from TAIR**

Future directions

- **Improvements to address limitations**
- **Support for more languages**
- **Library of PerlCyc scripts**

Acknowledgments

- **Peter Karp, SRI**
- **Suzanne Paley, SRI**
- **Sue Rhee, TAIR**
- **Danny Yoo, TAIR**

Perl scripting as a spectator sport