

**PhaseolusGenes :**  
**A Genomics Information Management and**  
**Marker Discovery Toolkit**  
**For A Non-Model Organism**

**Dawei Lin, Ph.D. et al.**

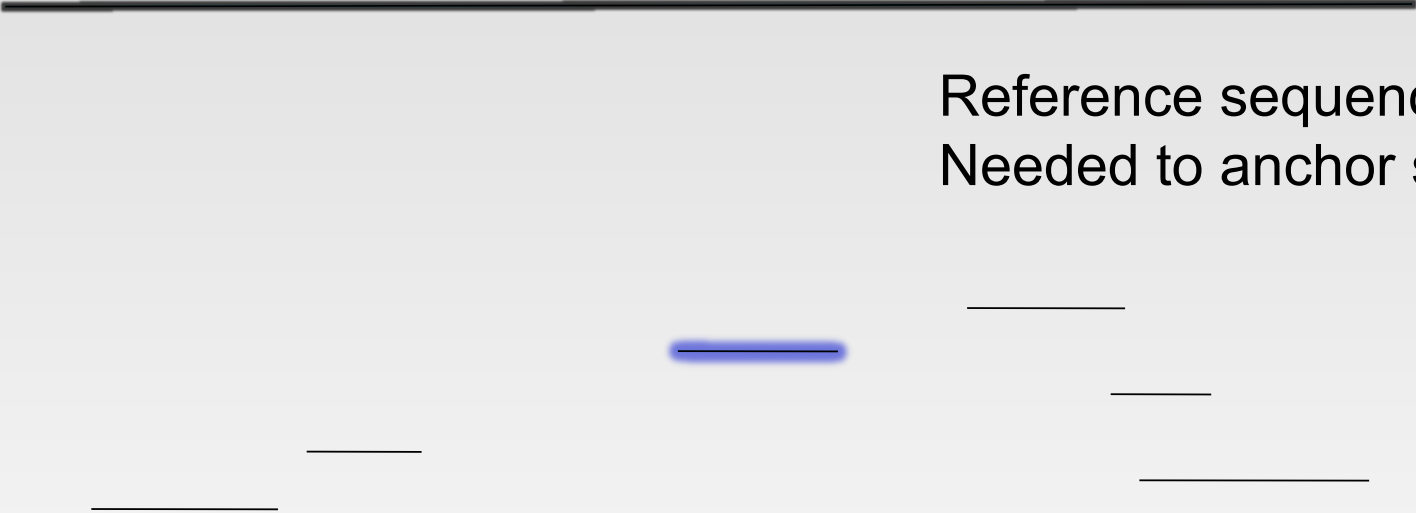
**Director of Bioinformatics Core**  
**Genome Center, UC Davis**

Jan 15, 2010, GMOD 2010

# Marker Discovery

---

Reference sequence is  
Needed to anchor sequences



# Problems



Sequences can not be anchored for non-model organisms that do not have a reference genome sequence

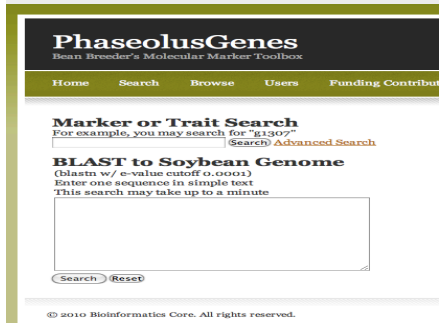
# Data Challenges

- Genomics Sequences (~400,000)
- BAC End Sequences (~90,000)
- ESTs (~70,000)
- Marker Information (~1400)
  - Description, Type, PCR primers, DB links, literatures
- Genetic Maps
- + Non Model Organism

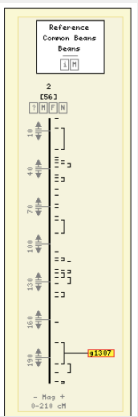
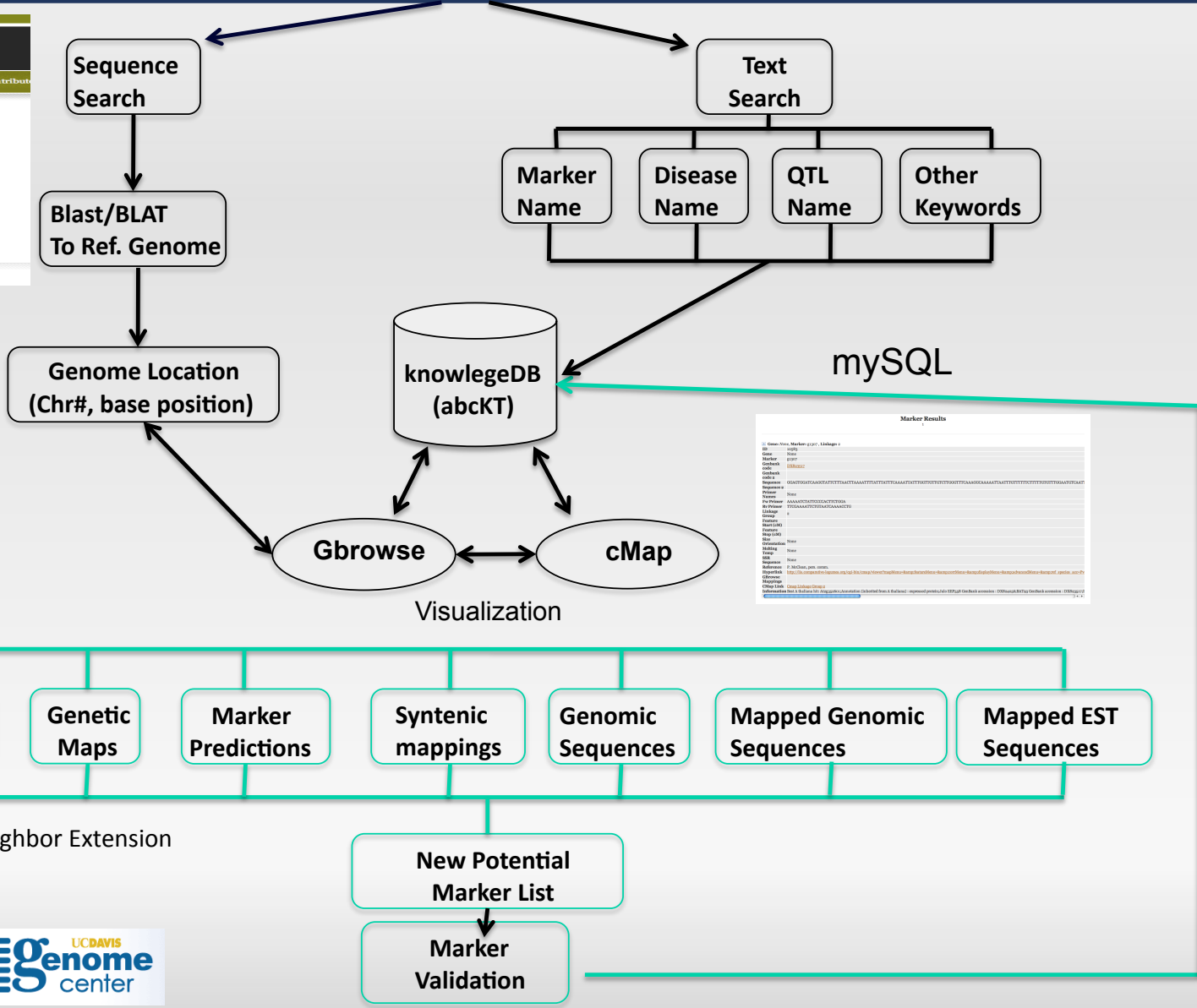
# Computational Challenges

- Multiple tools
- Computational Capacity
- Geographical locations

# Bean breeder's toolbox



Django



# Demo

## Live Demo



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

# Home Page

## PhaseolusGenes

Bean Breeder's Molecular Marker Toolbox



[Home](#) [Search](#) [Browse](#) [Users](#) [Funding Contributors](#) [Data Contributors](#) [About](#) [Log Out](#)

### Marker or Trait Search

For example, you may search for "g1307"

[Search](#) [Advanced Search](#)

### BLAST to Soybean Genome

(blastn w/ e-value cutoff 0.0001)

Enter one sequence in simple text

This search may take up to a minute

[Search](#) [Reset](#)

© 2010 Bioinformatics Core. All rights reserved.



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>



# Database Record For Each Marker

## Marker Results

1

+ Gene: None, Marker: g1307, Linkage: 2

ID	10583
Gene	None
Marker	g1307
Genbank code	<a href="#">DX823517</a>
Genbank code 2	
Sequence	GGAGTGGATCAAGGTATTCTTTAACTTAAAAATTTTATTTATTTCAAAATTATTTGGTTGTGTCTTGGGTTTCAAAGGCCAAAAATTAATTTGTTTTCTTTTTGTGTTTGGAAATGTCAAT
Sequence 2	
Primer Names	None
Fw Primer	AAAAATCTATTCCCCACITCTGGA
Rv Primer	TTCGAAAATTCTGTAATCAAAACCTG
Linkage Group	2
Feature Start (cM)	
Feature Stop (cM)	
Size	
Orientation	None
Melting Temp	None
SSR Sequence	None
Reference	P. McClean, pers. comm.
Hyperlink	<a href="http://lis.comparative-legumes.org/cgi-bin/cmap/viewer?mapMenu=&amp;featureMenu=&amp;corrMenu=&amp;displayMenu=&amp;advancedMenu=&amp;ref_species_acc=Pv">http://lis.comparative-legumes.org/cgi-bin/cmap/viewer?mapMenu=&amp;featureMenu=&amp;corrMenu=&amp;displayMenu=&amp;advancedMenu=&amp;ref_species_acc=Pv</a>
GBrowse Mappings	
CMap Link	<a href="#">Cmap Linkage Group 2</a>
Information	Best A thaliana hit: At2g352601;Annotation (inherited from A thaliana) : expressed protein;Jalo EEP558 GenBank accession : DX824056;BAT93 GenBank accession : DX823517;I

# Blast Results

## Soy Search Results

Default cutoff e-value: 0.0001

- sequence: lcl|Gm01
- score: 236.4 bits (119.0)
- e value: 2.5238e-60
- frame: +1/+1
- positives: 240
- identity: 240
- [GBrowse Link](#)

```
Score 119 (236 bits), expectation 2.5e-60, alignment length 282
Query:      123 TGACAGACTGATAACATTCTTGCTCCCATTATCACACATGGCATA...CAG 393
             |||...|||
Sbjct:52342487 TGACAGACTGATAATATTCTTGCTCCCATTATCACACATGGCATA...CAG 52342766
```

- sequence: lcl|Gm11
- score: 234.4 bits (118.0)
- e value: 9.97248e-60
- frame: +1/-1
- positives: 255
- identity: 255
- [GBrowse Link](#)

```
Score 118 (234 bits), expectation 1.0e-59, alignment length 300
Query:      396 AACCTGCATATTCTGTATGTACGAGACAAGAAGTATAATGAGTAT...ACA 106
             |||...|||
Sbjct: 3187509 AACCTGCATATTCTGTATGTATGAGACAAAACATGAGTAT...ACA 3187806
```

# Markers on GBrowse

**Search**  
Landmark or Region:

**Data Source**

**Reports & Analysis:**

**Scroll/Zoom:**       Flip

**Overview**

**Details**

**Gene**

**mRNA**

**Phaseolus\_vulgaris\_McClean**

**Methyl-Filtrated BAT93**

[Clear highlighting](#)

**Tracks**

**Analysis**  All on  All off

Restriction Sites

**General**  All on  All off

<input type="checkbox"/> CDS	<input type="checkbox"/> Exon	<input checked="" type="checkbox"/> mRNA
<input type="checkbox"/> Clone from soybase	<input type="checkbox"/> Five Prime UTR	<input checked="" type="checkbox"/> Phaseolus_vulgaris_McClean
<input type="checkbox"/> DNA/GC Content	<input checked="" type="checkbox"/> Methyl-Filtrated BAT93	<input type="checkbox"/> Three Prime UTR

**Genes**  All on  All off

Gene

**Display Settings**

<input checked="" type="checkbox"/> Show grid	<b>Image Width</b> <input type="radio"/> 450 <input type="radio"/> 640 <input checked="" type="radio"/> 800 <input type="radio"/> 1024	<b>Highlight feature(s) (feature1 feature2...)</b> <input type="text" value="g1307@yellow"/>
<input checked="" type="checkbox"/> Cache tracks	<b>Track Name Table</b> <input checked="" type="radio"/> Alphabetic <input type="radio"/> Varying	<b>Highlight regions (region1:start.end region2:start.end)</b> <input type="text"/>

Search

Landmark or Region:

Gm01:52337567..52347566 Search

Data Source

Soybean (Glyma1, Dec2008, ftp.jgi-psf.org)

Overview

Reports & Analysis:

Annotate Restriction Sites Configure... Go

Scroll/Zoom: <<< Show 10 kbp >>> Flip

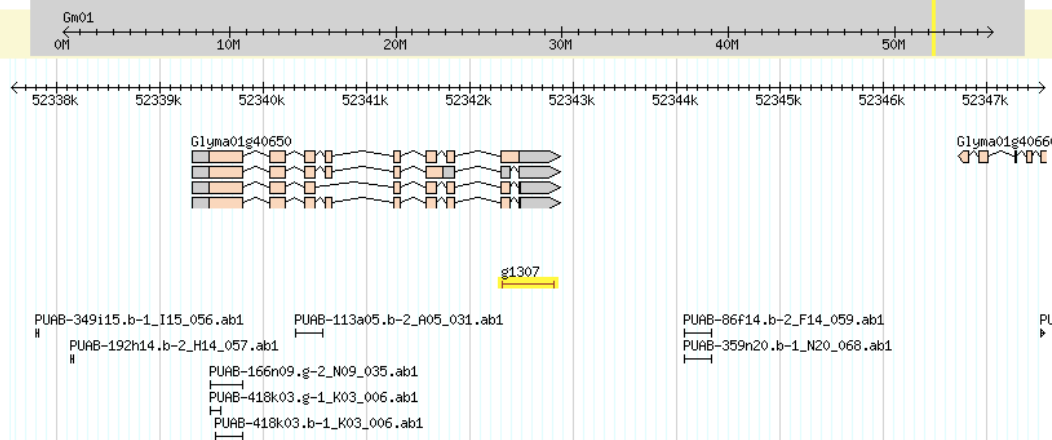
Details

Gene

mRNA

Phaseolus vulgaris\_McClean

Methyl-Filtrated BAT93



Clear highlighting

Update Image

Tracks

Analysis All on All off

Restriction Sites

General All on All off

CDS

Clone from soybase

DNA/GC Content

Genes All on All off

Exon

Five Prime UTR

Methyl-Filtrated BAT93

mRNA

Phaseolus\_vulgaris\_McClean

Three Prime UTR

Search: |cl|Gm01:52342487..52342766 Search

Data Source

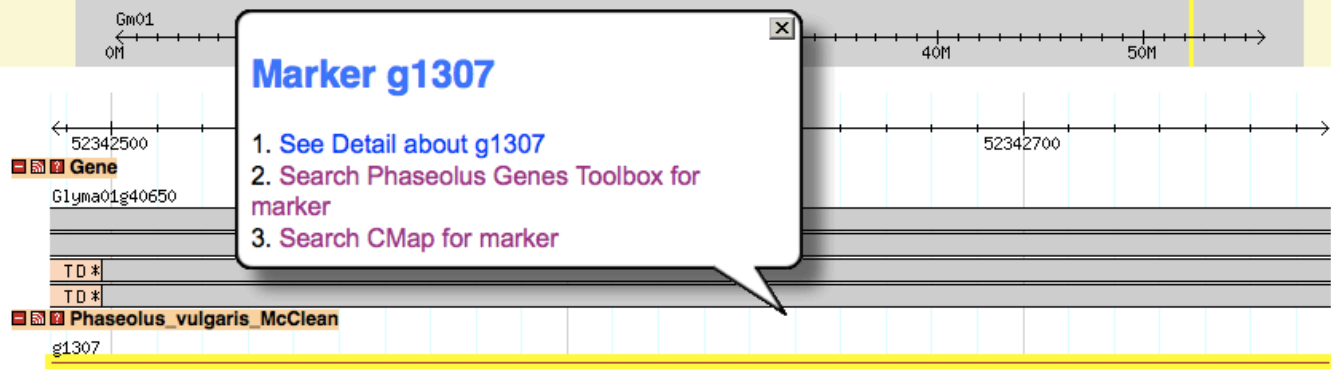
Soybean (Glyma1, Dec2008, ftp.jgi-psf.org)

Annotate Restriction Sites Configure... Go

Scroll/Zoom: <<< << < Show 280 bp + >> >>> Flip

Overview

Details



Clear highlighting

Tracks

Analysis  All on  All off

Restriction Sites

General  All on  All off

CDS

Exon

mRNA

Clone from soybase

Five Prime UTR

Phaseolus\_vulgaris\_McClean

DNA/GC Content

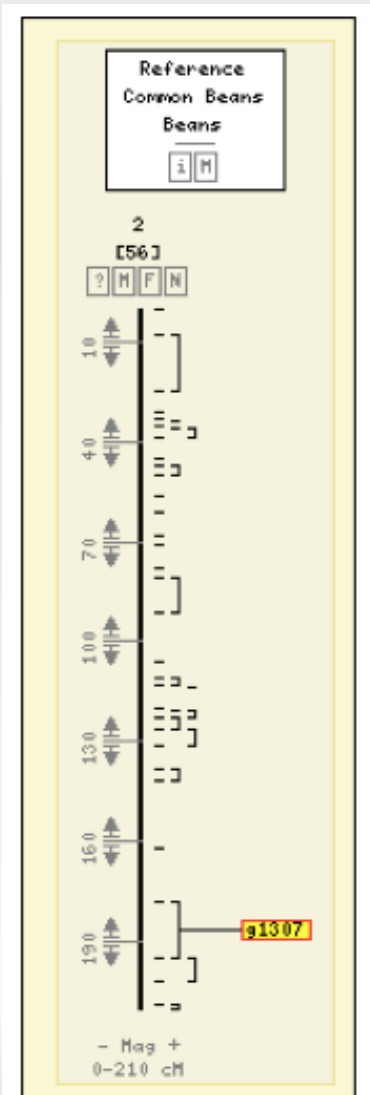
Methyl-Filtrated BAT93

Three Prime UTR

Genes  All on  All off

Gene





# Bioinformatics Core Members



Dawei Lin, Vince Buffalo, Joe Fass, Monica Britton, Jose Boveda, Nikhil Joshi, Zhiwei Lu



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>

# Acknowledgement

- Prof. Paul Gepts and groups
- UC Davis Genome Center

Kirkhouse  
Trust



United States Department of Agriculture

National Institute of Food and Agriculture



Bioinformatics Core, <http://bioinformatics.ucdavis.edu>