BeanCAP: Objective 3: PhaseolusGenes database

Paul Gepts
Dept. Plant Sciences / MS1
University of California, Davis

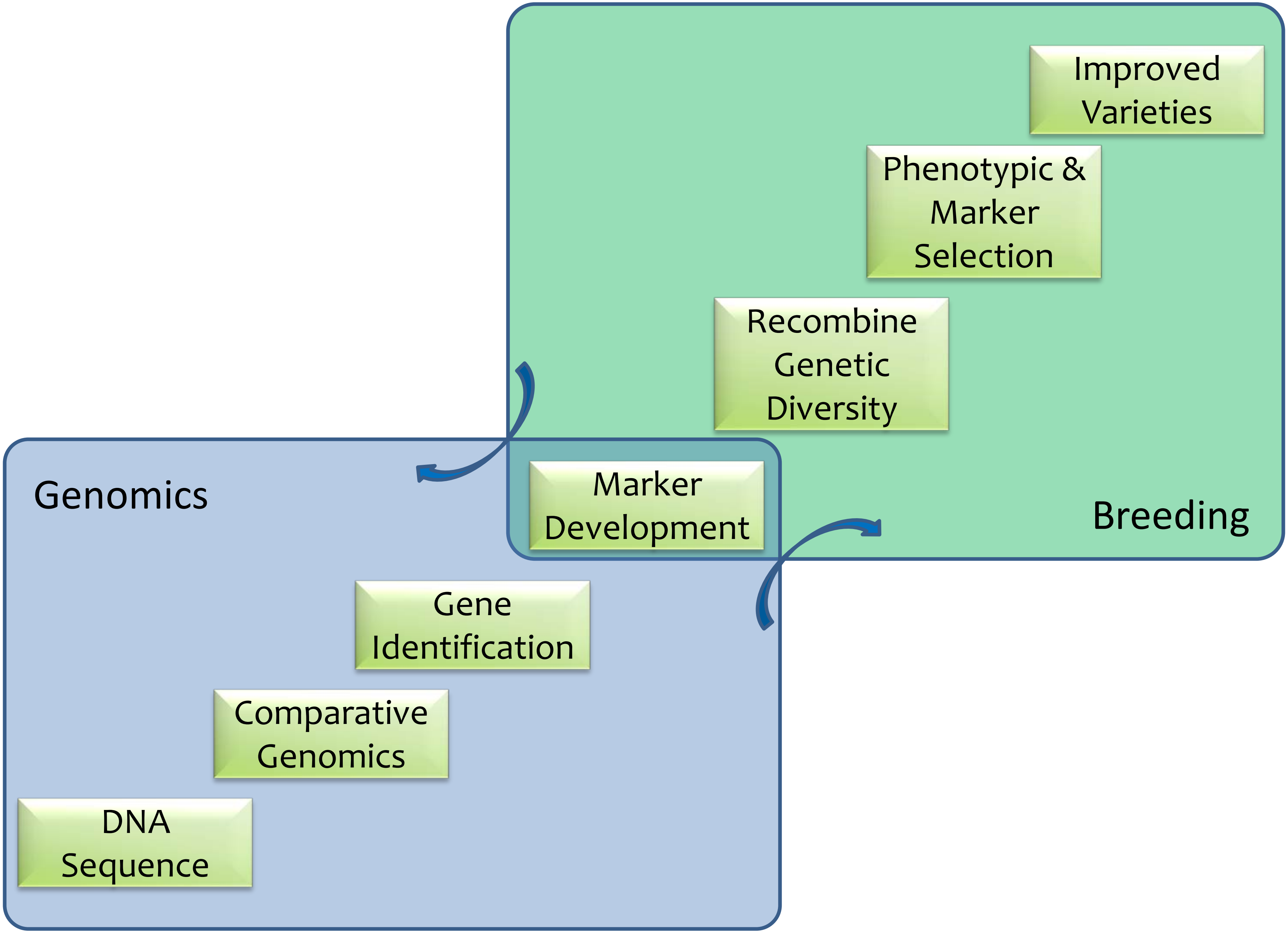
Annual Meeting, San Diego, CA
Jan. 10, 2013

BeanCAP Objective 3

- Challenge from breeding perspective
  - Long-term: Gene discovery, basis of phenotypes
  - Short-term: Marker discovery
- Prerequisite
  - Mapped sequence information
- Goal of PhaseolusGenes
  - Collate and present information in *Phaseolus* sp. for marker development and utilization

Database Development Participants

- Dawei Lin, Director Bioinformatics Core, UC Davis Genome Center
- Jose Boveda, Web Interface
- Monica Britton, Analysis 1x DNA sequence
- Joe Fass, Senior programmer
- Nikhil Joshi, Programmer, Synteny & SSR analyses
- Zhi-Wei Lu, Hardware/Software; External drive
- James Kami, Marker analysis and development
- José Vicente Gomes dos Santos: visiting undergraduate (ESALQ): data curation
- Shelby Repinski: graduate student: data curation
- Adriana Navarro Gómez: data curation: CBB
- Erin Wilkus, data curation: updates
- Paul Gepts, Co-PI



Genomics

DNA Sequence

Comparative Genomics

Gene Identification

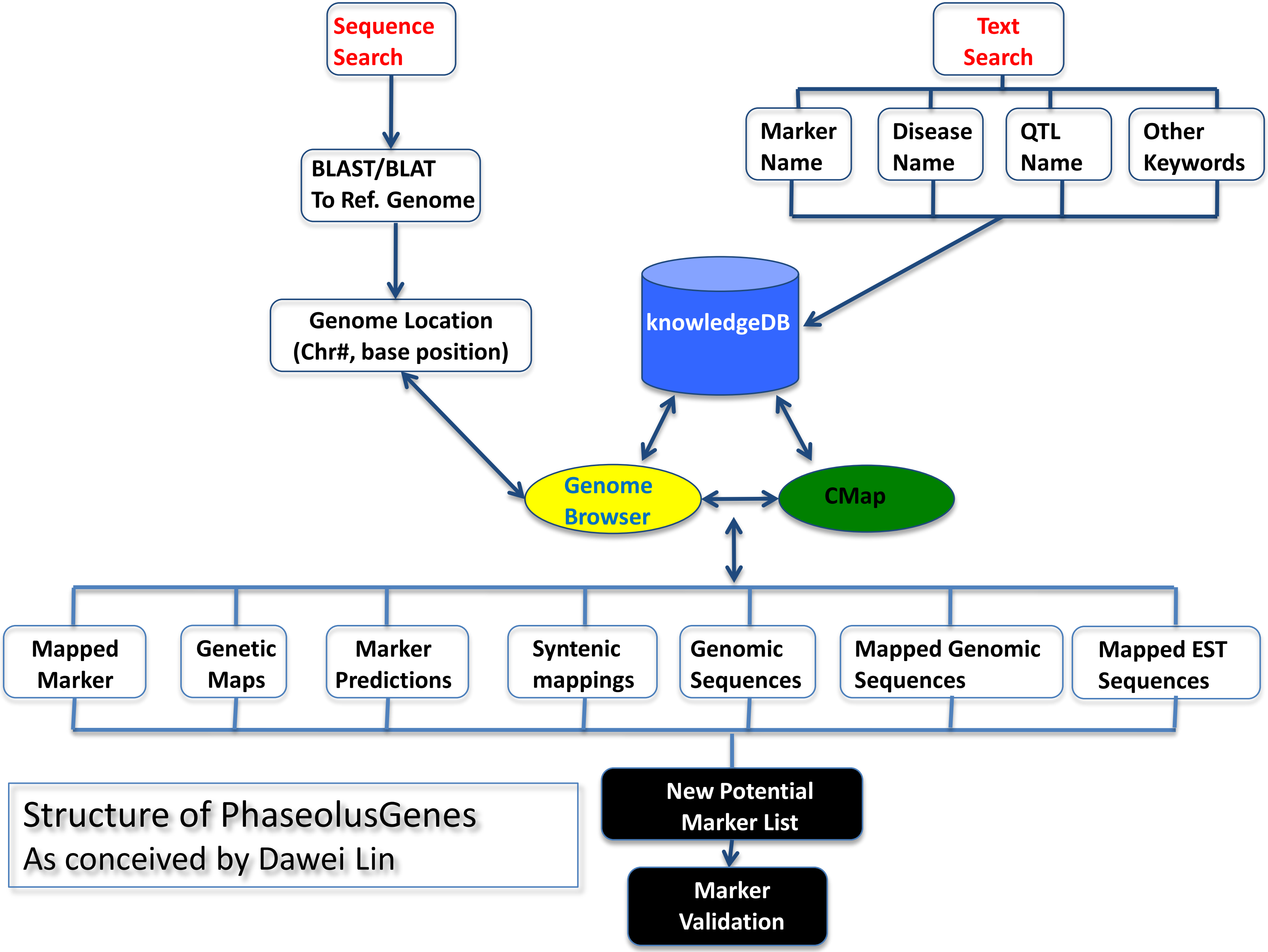
Marker Development

Recombine Genetic Diversity

Phenotypic & Marker Selection

Improved Varieties

Breeding



Sequence Search

BLAST/BLAT To Ref. Genome

Genome Location (Chr#, base position)

Text Search

Marker Name

Disease Name

QTL Name

Other Keywords

knowledgeDB

Genome Browser

CMap

Mapped Marker

Genetic Maps

Marker Predictions

Syntenic mappings

Genomic Sequences

Mapped Genomic Sequences

Mapped EST Sequences

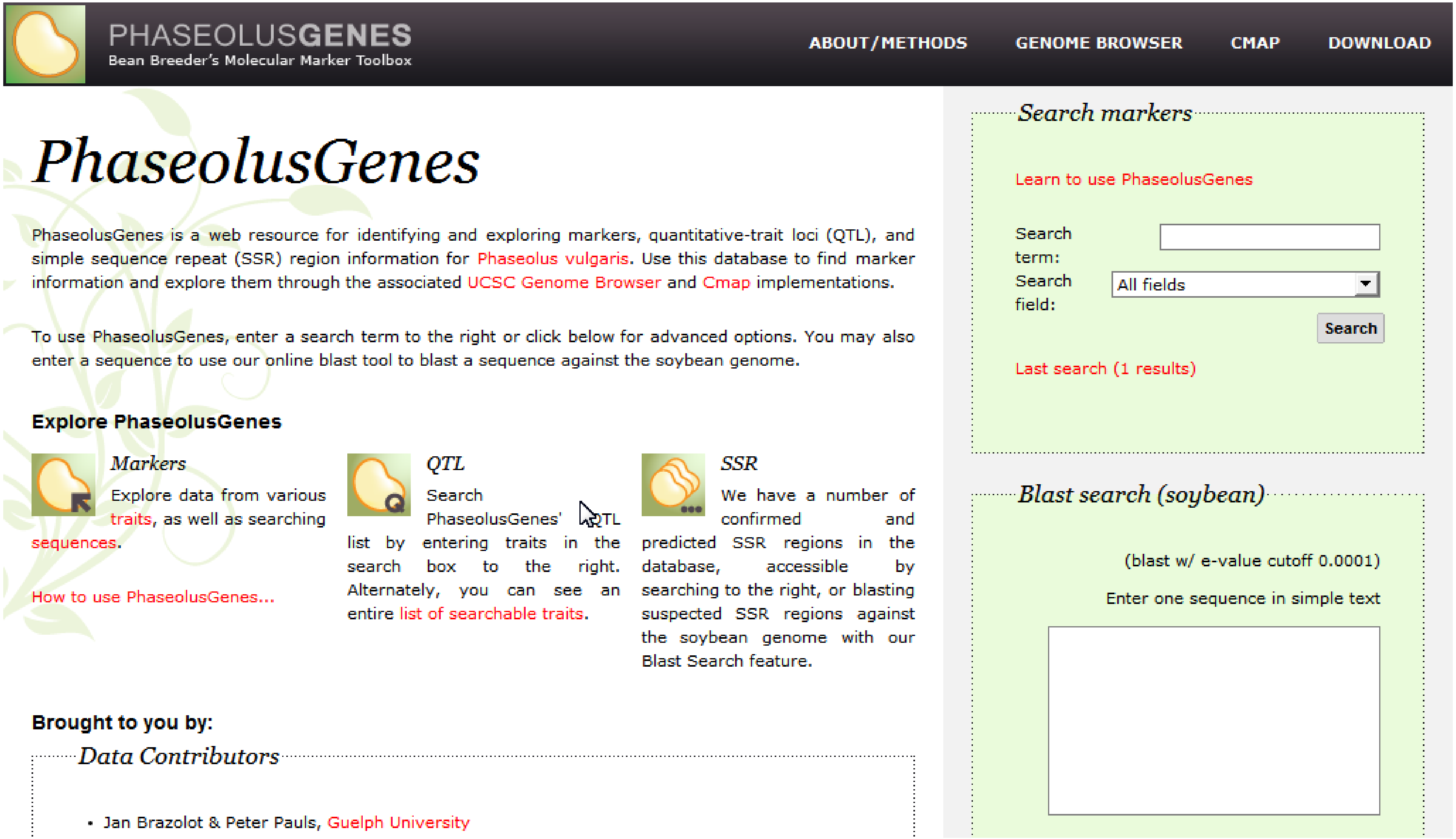
New Potential Marker List

Marker Validation

Structure of PhaseolusGenes
As conceived by Dawei Lin

PhaseolusGenes: Front page

bit.ly/pgenes or phaseolusgenes.bioinformatics.ucdavis.edu



PhaseolusGenes

PhaseolusGenes is a web resource for identifying and exploring markers, quantitative-trait loci (QTL), and simple sequence repeat (SSR) region information for Phaseolus vulgaris. Use this database to find marker information and explore them through the associated UCSC Genome Browser and Cmap implementations.

To use PhaseolusGenes, enter a search term to the right or click below for advanced options. You may also enter a sequence to use our online blast tool to blast a sequence against the soybean genome.

Explore PhaseolusGenes

Markers

Explore data from various traits, as well as searching sequences.

How to use PhaseolusGenes...

QTL

Search PhaseolusGenes' QTL list by entering traits in the search box to the right. Alternately, you can see an entire list of searchable traits.

SSR

We have a number of confirmed and predicted SSR regions in the database, accessible by searching to the right, or blasting suspected SSR regions against the soybean genome with our Blast Search feature.

Brought to you by:

Data Contributors

- Jan Brazolot & Peter Pauls, Guelph University

Search markers

Learn to use PhaseolusGenes

Search term:

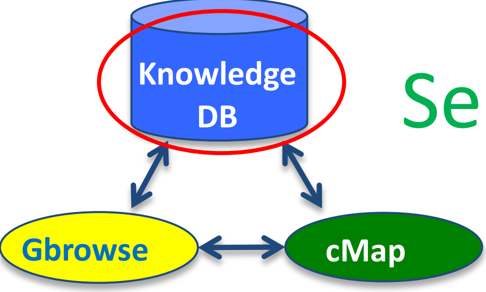
Search field: All fields

Last search (1 results)

Blast search (soybean)

(blast w/ e-value cutoff 0.0001)

Enter one sequence in simple text

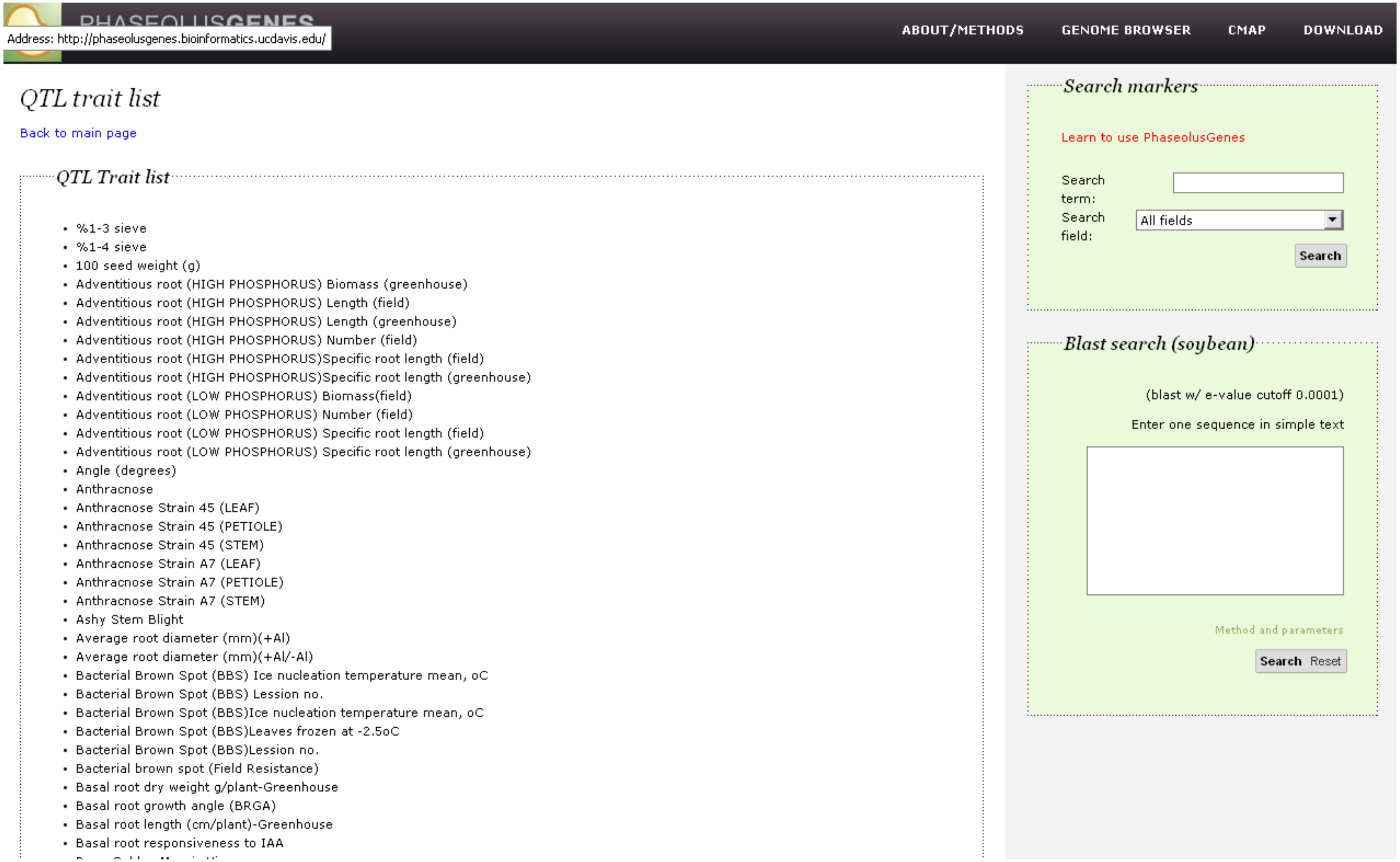


Searchable Marker & Trait Database

Type	Source	Name	Total	Mapped
STS	Hougaard et al. 2008	Leg	195	105
	Various		12	11
	Murray et al. 2002	Bng	130	89
	Kami et al. , unpubl.	D	113	9
	Kim et al. 2006	COS	120	0
	McConnell et al. 2010	g	302	295
	David et al. 2008	B4 BACs	114	114
		Total STS	**986**	**623**
Genes			25	15
SCARs			78	69
SSRs	Blair et al. , various	BM	692	210
	Yu et al. 2000	Pv	37	19
	Buso et al. 2006, Grisi et al. 2007	PVBR	81	55
	Hanai et al. 2010	PVM	140	45
	Hanai et al. 2007	FJ	40	0
	Benchimol et al. 2007, Oblessuc et al. 2009	IAC, IAC-SSR	173	0
	Cardoso et al. 2008	SSR-IAC	99	0
		Total SSR	**1262**	**329**
		General total	**2351**	**1036**

- Last year (2011): ~1550 markers; now (2012): ~2,350
- Additional markers: putative SSRs from 1x BAT93 sequencing: ~ 160K sequences
- Added SSRs of Garcia et al. 2011; SNPs of Hyten et al. 2010
- Added SSRs from CIAT (Bodo Raatz)

PhaseolusGenes: QTLs



Address: http://phaseolusgenes.bioinformatics.ucdavis.edu/

QTL trait list

Back to main page

QTL Trait list

- %1-3 sieve
- %1-4 sieve
- 100 seed weight (g)
- Adventitious root (HIGH PHOSPHORUS) Biomass (greenhouse)
- Adventitious root (HIGH PHOSPHORUS) Length (field)
- Adventitious root (HIGH PHOSPHORUS) Length (greenhouse)
- Adventitious root (HIGH PHOSPHORUS) Number (field)
- Adventitious root (HIGH PHOSPHORUS)Specific root length (field)
- Adventitious root (HIGH PHOSPHORUS)Specific root length (greenhouse)
- Adventitious root (LOW PHOSPHORUS) Biomass(field)
- Adventitious root (LOW PHOSPHORUS) Number (field)
- Adventitious root (LOW PHOSPHORUS) Specific root length (field)
- Adventitious root (LOW PHOSPHORUS) Specific root length (greenhouse)
- Angle (degrees)
- Anthracnose
- Anthracnose Strain 45 (LEAF)
- Anthracnose Strain 45 (PETIOLE)
- Anthracnose Strain 45 (STEM)
- Anthracnose Strain A7 (LEAF)
- Anthracnose Strain A7 (PETIOLE)
- Anthracnose Strain A7 (STEM)
- Ashy Stem Blight
- Average root diameter (mm)(+Al)
- Average root diameter (mm)(+Al/-Al)
- Bacterial Brown Spot (BBS) Ice nucleation temperature mean, oC
- Bacterial Brown Spot (BBS) Lession no.
- Bacterial Brown Spot (BBS)Ice nucleation temperature mean, oC
- Bacterial Brown Spot (BBS)Leaves frozen at -2.5oC
- Bacterial Brown Spot (BBS)Lession no.
- Bacterial brown spot (Field Resistance)
- Basal root dry weight g/plant-Greenhouse
- Basal root growth angle (BRGA)
- Basal root length (cm/plant)-Greenhouse
- Basal root responsiveness to IAA

Search markers

Learn to use PhaseolusGenes

Search term:

Search field: All fields

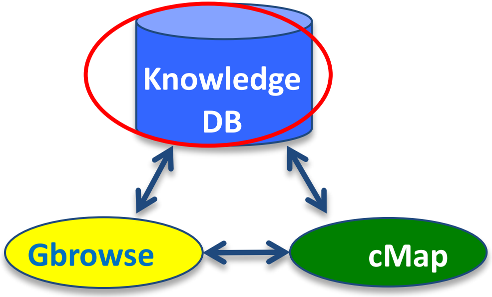
Blast search (soybean)

(blast w/ e-value cutoff 0.0001)

Enter one sequence in simple text

Method and parameters

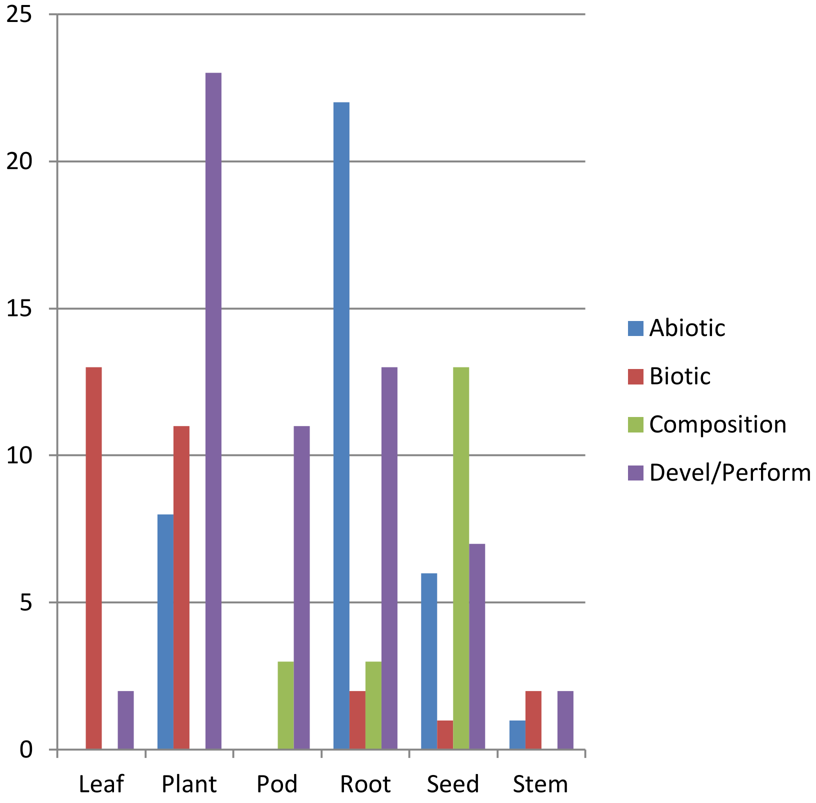
QTLs



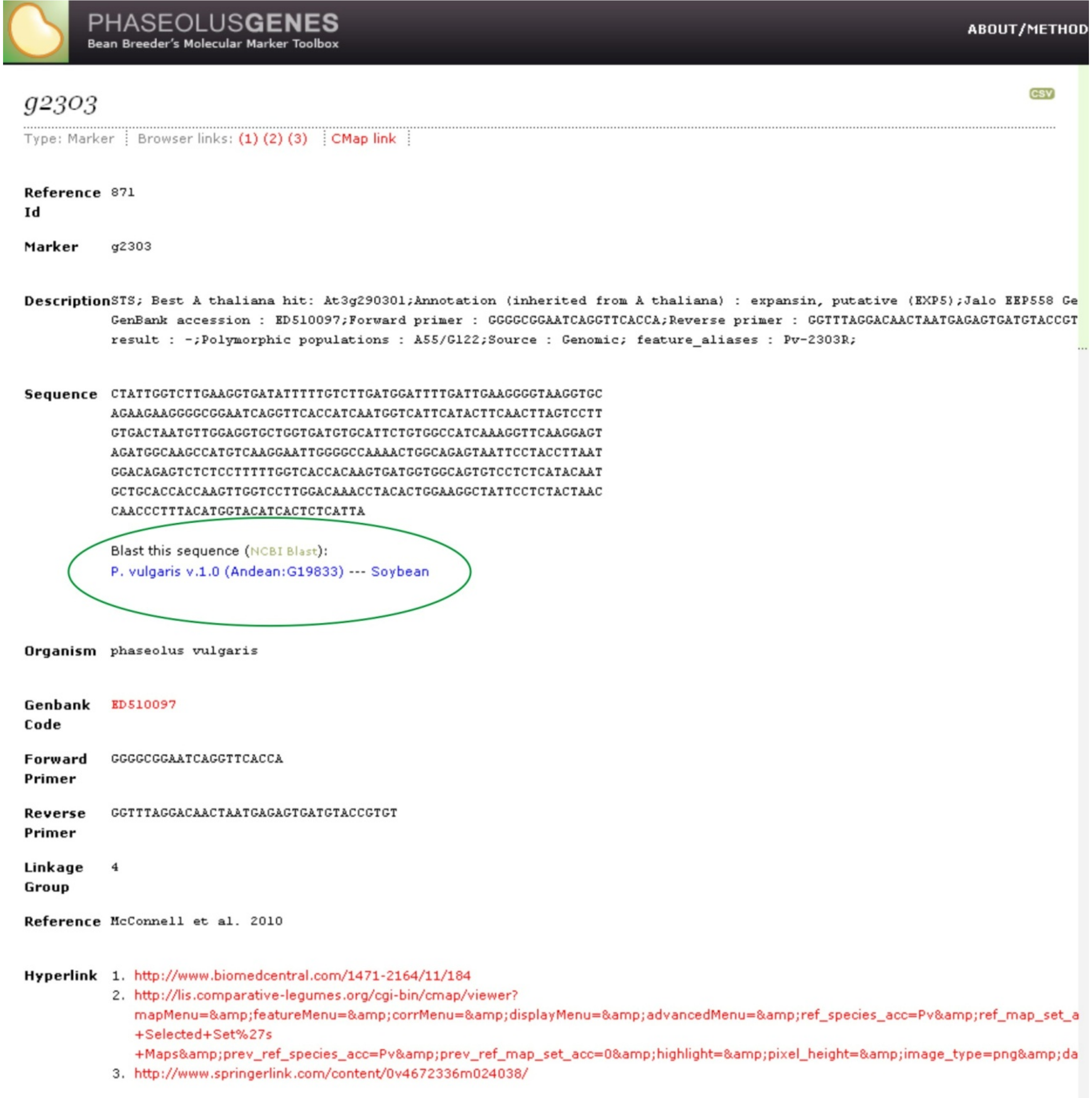
Knowledge DB

Gbrowse

cMap



- QTL = trait x location x year
- Last year: 925; this year: 1148
- 143 categories: organ x trait class
- Included in searchable table of PhaseolusGenes, but need to be added to Cmap
- Difficulty: many markers not-sequence-based and not included in database → need for inter- or extrapolation

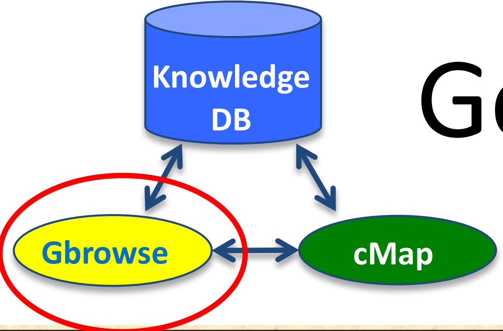


g2303

Type: Marker | Browser links: (1) (2) (3) | CMap link

Reference Id	871
Marker	g2303
Description	STS; Best A thaliana hit: At3g290301;Annotation (inherited from A thaliana) : expansin, putative (EXP5);Jalo EEP558 Ge GenBank accession : ED510097;Forward primer : GGGGCGGAATCAGGTTCACCA;Reverse primer : GGTTTAGGACAACTAATGAGAGTGATGTACCGT result : -;Polymorphic populations : A55/G122;Source : Genomic; feature_aliases : Pv-2303R;
Sequence	CTATTGGTCTTGAAGGTGATATTTTTGTCTTGATGGATTTTGATTGAAGGGGTAAGGTGC AGAAGAAGGGGCGGAATCAGGTTCACCATCAATGGTCATTCATACTTCAACTTAGTCCTT GTGACTAATGTTGGAGGTGCTGGTGATGTGCATTCTGTGGCCATCAAAGGTTCAAGGAGT AGATGGCAAGCCATGTCAAGGAATTGGGGCCAAAACTGGCAGAGTAATTCCTACCTTAAT GGACAGAGTCTCTCCTTTTTGGTCACCACAAGTGATGGTGGCAGTGTCCTCTCATACAAT GCTGCACCACCAAGTTGGTCCTTGGACAAACCTACACTGGAAGGCTATTCCTCTACTAAC CAACCCTTTACATGGTACATCACTCTCATTA Blast this sequence (NCBI Blast): P. vulgaris v.1.0 (Andean:G19833) --- Soybean
Organism	phaseolus vulgaris
Genbank Code	ED510097
Forward Primer	GGGGCGGAATCAGGTTCACCA
Reverse Primer	GGTTTAGGACAACTAATGAGAGTGATGTACCGTGT
Linkage Group	4
Reference	McConnell et al. 2010
Hyperlink	1. http://www.biomedcentral.com/1471-2164/11/184 2. http://lis.comparative-legumes.org/cgi-bin/cmap/viewer? mapMenu=&featureMenu=&corrMenu=&displayMenu=&advancedMenu=&ref_species_acc=Pv&ref_map_set_a +Selected+Set%27s +Maps&prev_ref_species_acc=Pv&prev_ref_map_set_acc=0&highlight=&pixel_height=&image_type=png&da 3. http://www.springerlink.com/content/0v4672336m024038/

Genome Browser

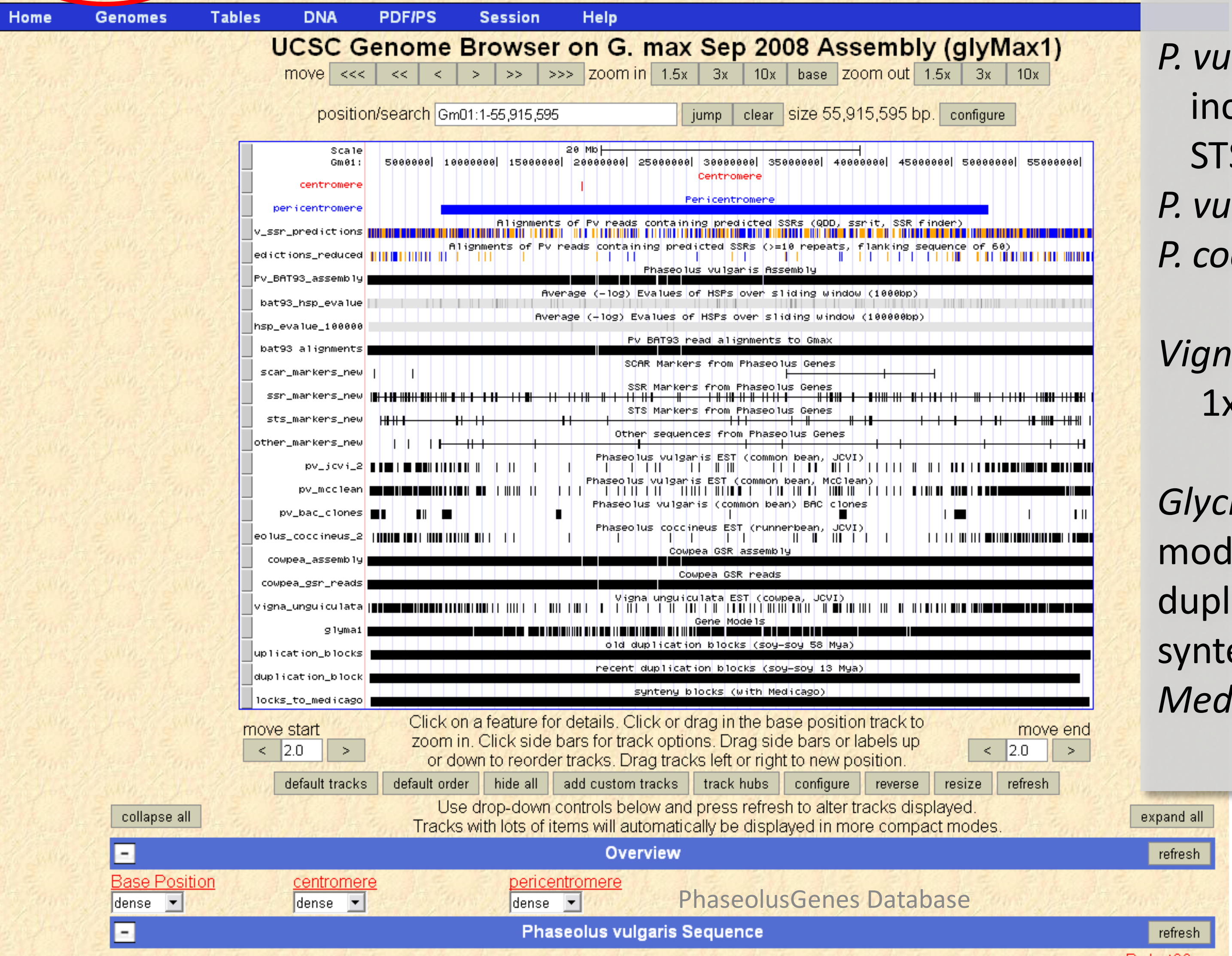


P. vulgaris BAT93 1x MF genomic DNA;
P. vulgaris SSR-containing sequences;

P. vulgaris markers, including SCARs, SSRs, and STSs;
P. vulgaris ESTs;
P. coccineus ESTs;

Vigna unguiculata (cowpea) 1x MF seq. & ESTs

Glycine max (soybean) gene models, old and recent duplication blocks, and synteny blocks with *Medicago*



UCSC Genome Browser on G. max Sep 2008 Assembly (glyMax1)

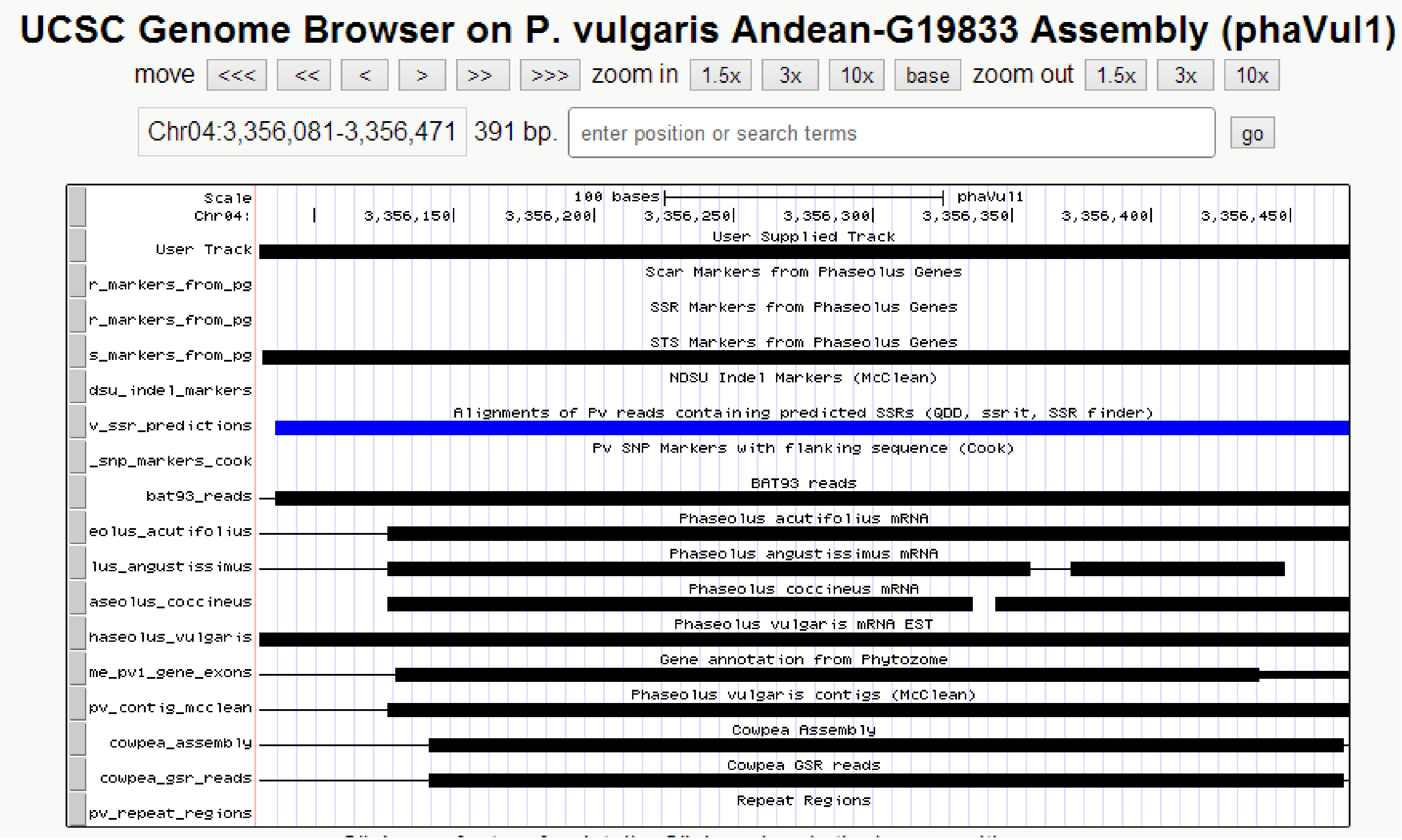
position/search Gm01:1-55,915,595 size 55,915,595 bp.

Click on a feature for details. Click or drag in the base position track to zoom in. Click side bars for track options. Drag side bars or labels up or down to reorder tracks. Drag tracks left or right to new position.

Use drop-down controls below and press refresh to alter tracks displayed. Tracks with lots of items will automatically be displayed in more compact modes.

Base Position: dense
centromere: dense
pericentromere: dense

Genome Browser Anchored to *P. vulgaris*



UCSC Genome Browser on P. vulgaris Andean-G19833 Assembly (phaVul1)

Chr04:3,356,081-3,356,471 391 bp.

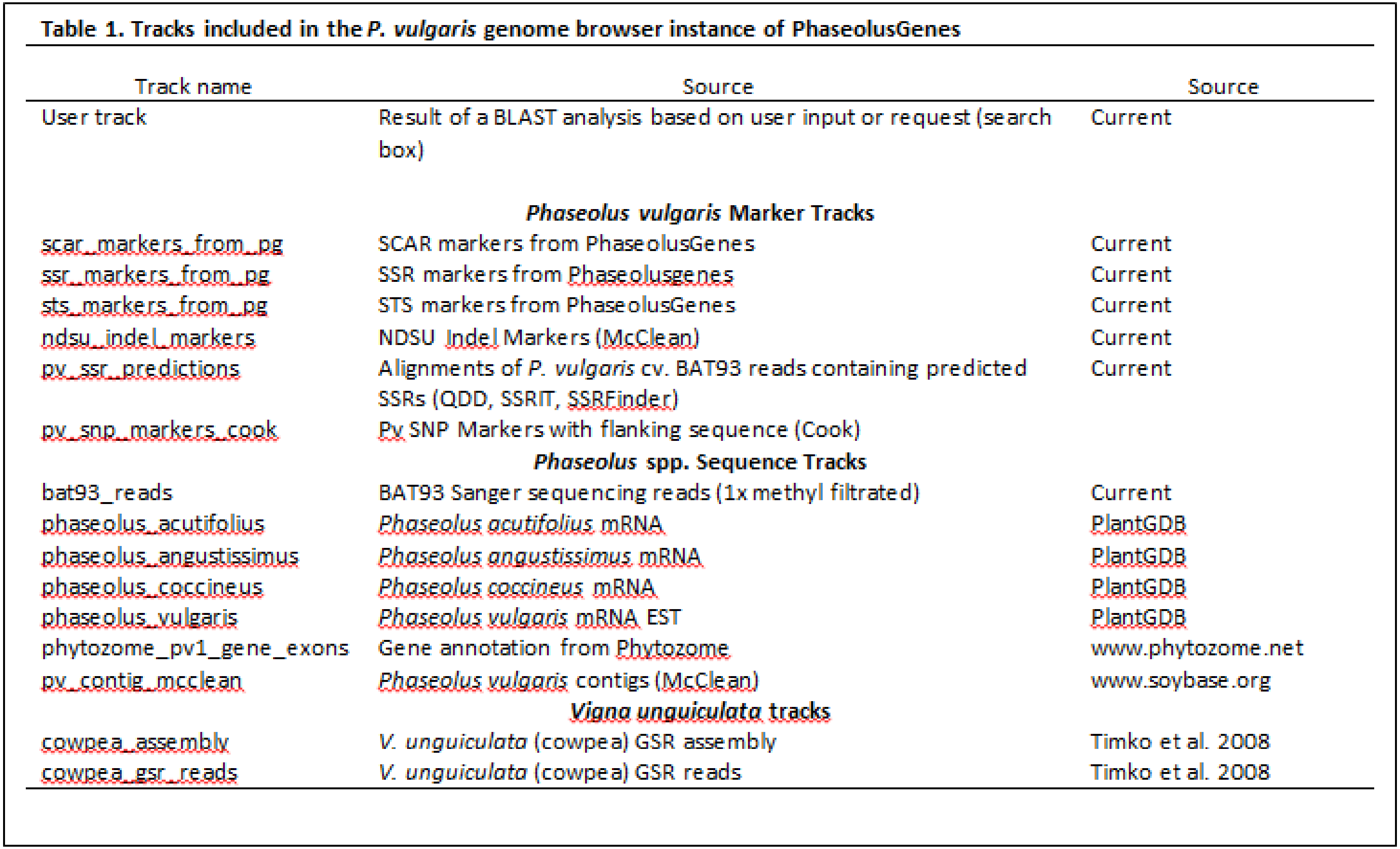


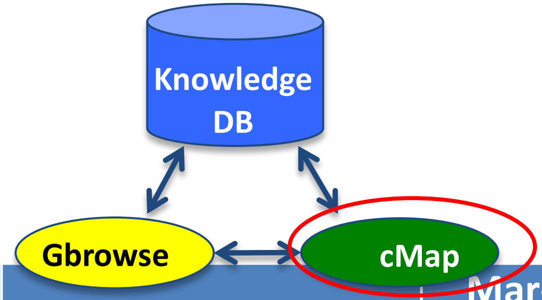
Table 1. Tracks included in the *P. vulgaris* genome browser instance of PhaseolusGenes

Track name	Source	Source
User track	Result of a BLAST analysis based on user input or request (search box)	Current
	***Phaseolus vulgaris* Marker Tracks**	
scar_markers_from_pg	SCAR markers from PhaseolusGenes	Current
ssr_markers_from_pg	SSR markers from Phaseolusgenes	Current
sts_markers_from_pg	STS markers from PhaseolusGenes	Current
ndsu_indel_markers	NDSU Indel Markers (McClean)	Current
pv_ssr_predictions	Alignments of *P. vulgaris* cv. BAT93 reads containing predicted SSRs (QDD, SSRIT, SSRFinder)	Current
pv_snp_markers_cook	Pv SNP Markers with flanking sequence (Cook)	
	***Phaseolus* spp. Sequence Tracks**	
bat93_reads	BAT93 Sanger sequencing reads (1x methyl filtrated)	Current
phaseolus_acutifolius	*Phaseolus acutifolius* mRNA	PlantGDB
phaseolus_angustissimus	*Phaseolus angustissimus* mRNA	PlantGDB
phaseolus_coccineus	*Phaseolus coccineus* mRNA	PlantGDB
phaseolus_vulgaris	*Phaseolus vulgaris* mRNA EST	PlantGDB
phytozome_pv1_gene_exons	Gene annotation from Phytozome	www.phytozome.net
pv_contig_mcclean	*Phaseolus vulgaris* contigs (McClean)	www.soybase.org
	***Vigna unguiculata* tracks**	
cowpea_assembly	*V. unguiculata* (cowpea) GSR assembly	Timko et al. 2008
cowpea_gsr_reads	*V. unguiculata* (cowpea) GSR reads	Timko et al. 2008

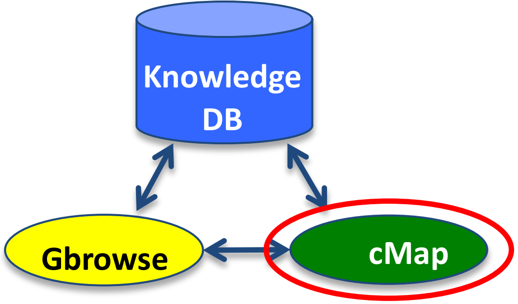
Table 2. Tracks included in the soybean genome browser instance of PhaseolusGenes

Track name	Description	Source
centromere	Soybean centromere	www.soybase.org
pericentromere	Soybean pericentromere	www.soybase.org
	***Phaseolus vulgaris* sequences**	
Pv_ssr_predictions	Alignments of *P. vulgaris* cv. BAT93 reads containing predicted SSRs (QDD, SSRIT, SSRFinder)	Current
Pv_BAT93_assembly	All *P. vulgaris* sequences resulting from the 1X methyl-filtrated dideoxy sequencing	Current
Bat93_hsp_evalue	Average (-log) Evalues of HSPs over sliding window (1000bp)	Current
Bat93_hsp_evalue	Average (-log) Evalues of HSPs over sliding window (100000bp)	Current
Pv BAT93 alignments	*P. vulgaris* cv. BAT93 read alignments to *G. max*	Current
	***Phaseolus vulgaris* tracks**	
scar_markers	SCAR markers from PhaseolusGenes	Current
ssr_markers	SSR markers from Phaseolusgenes	Current
sts_markers	STS markers from PhaseolusGenes	Current
pv_jcvi_2	*P. vulgaris* EST (common bean, JCVI)	http://plantta.jcvi.org/cgi-bin/plantta_release.pl
pv_mcclean	*P. vulgaris* EST (common bean, McClean)	www.soybase.org
pv_bac_clones	*P.vulgaris* BAC clones	www.soybase.org
	***Phaseolus coccineus* tracks**	
phaseolus_coccineus_2	*P. coccineus* EST (runner bean, JCVI)	http://plantta.jcvi.org/cgi-bin/plantta_release.pl
cowpea_assembly	*V. unguiculata* (cowpea) GSR assembly	Timko et al. 2008
cowpea_gsr_reads	*V. unguiculata* (cowpea) GSR reads	Timko et al. 2008
vigna_unguiculata	*V. unguiculata* EST (cowpea, JCVI)	http://plantta.jcvi.org/

CMap



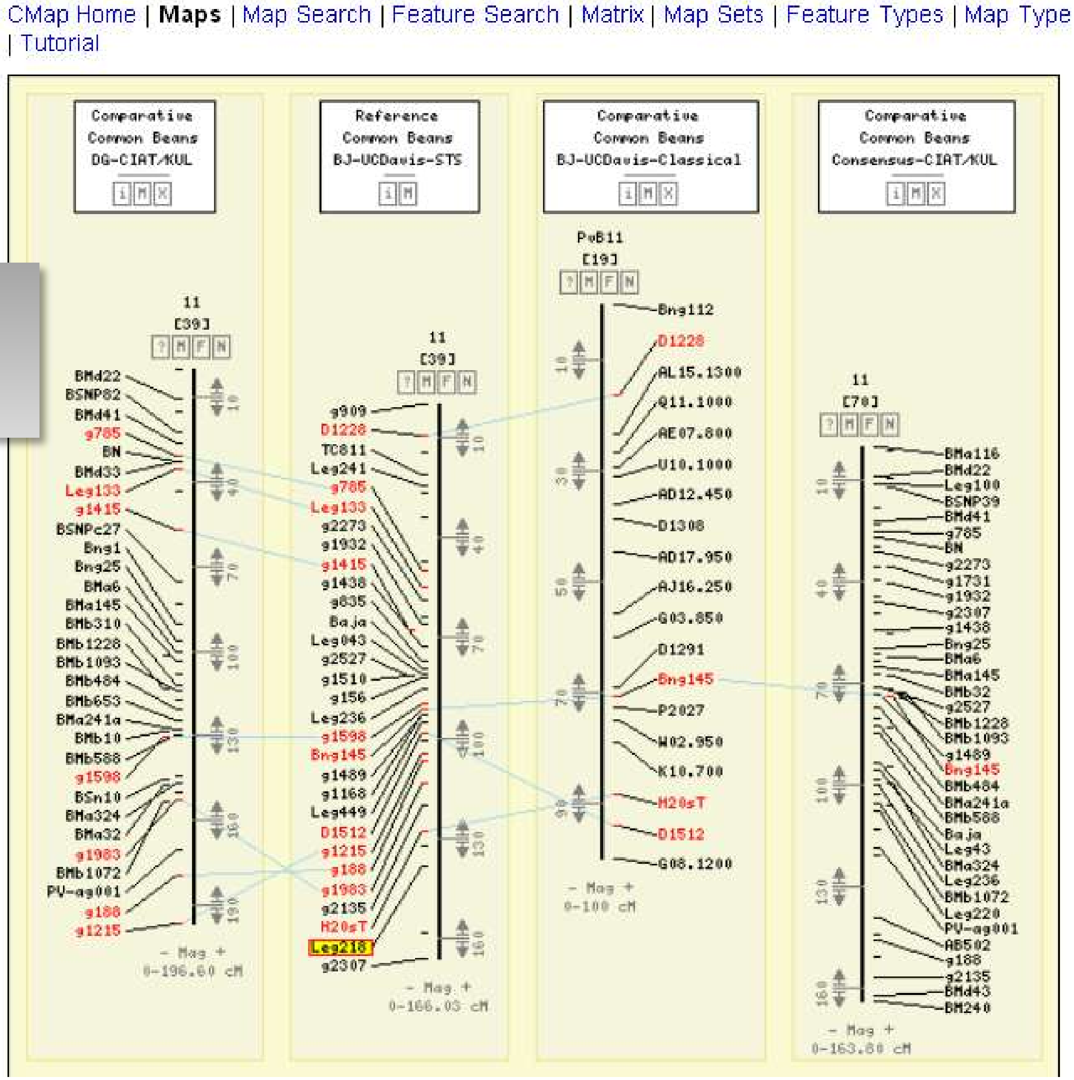
Name	Parents	Generation	Type	Source	Type of markers	Markers
BJ-UCDavis-Classical	BAT93 x Jalo EEP558	RI	Meso x Andean	Freyre et al. 1998	Phenotypic, RFLP, RAPD, AFLP; SCAR; Allozyme	234
BJ-NDSU	BAT93 x Jalo EEP558	RI	Meso x Andean	McClean 2007; McConnell et al. 2010	STS: g, CAPS, dCAPS, indel; SCAR; RAPD, AFLP	450
BJ-UCDavis-STS	BAT93 x Jalo EEP558	RI	Meso x Andean	Navarro Gomes & Gepts, unpubl.	STS: g, Leg, Bng, D, RFLP;	485
BJ-CIAT/KUL	BAT93 x Jalo EEP558	RI	Meso x Andean	Galeano et al. 2011	STS: Leg; SNP: g	424
DB-CIAT/KUL	DOR364 x BAT477	RI	Meso x Meso	Galeano et al. 2011	SSR, SSR-BES, RAPD, AFLP	291
DG-CIAT/KUL	DOR364 x G19833	RI	Meso x Andean	Galeano et al. 2011	SSR, SNP, STS: RFLP, Leg	499
Consensus CIAT-KUL	Consensus	NA	NA	Galeano et al. 2011	SSR; SNP; STS: Leg, g, Bng	1010



CMap

Multiple map comparison

- within species



Availability of potential alternative markers around the SN02 marker in their physical order based on synteny with soybean (PhaseolusGenes) and available whole-genome sequence scaffolds (Phytozome.net)

Trait	Gene (LG)	Markers	Type	Soybean synteny[a]	Common bean scaffold: length[b]	Position on scaffold
		Leg722	Yes	Gm14+Gm12	s00485: 232 kbp	194 kbp
		g2316	STS	Gm14+Gm02	s00012: 2.4 Mbp	749 kbp
		SSR[c]	SSR	Gm02+Gm14		563 kbp
		g796	STS	Gm02+Gm14		434 kbp
		SSR[d]	SSR	Gm14+Gm02		282 kbp
		D1472	STS	Gm14+Gm02		135 kbp
		PvM012	SSR	Gm14+Gm08		101 kbp
		Bng73 (R)	STS	Gm14+Gm02		042 kbp
		SM02	SCAR	Gm14+Gm16	s00126: 703 kbp	011 kbp
		Bmb352	SSR	Gm18+Gm14		032 kbp
ALS	*Phg-2* (08)	**SN02**	SCAR	Gm14		069 kbp
		BARC-PV-0003175	SNP	Gm14+Gm02		087 kbp
		Leg196	Yes	Gm02+ Gm14		127 kbp
		SSR[e]	SSR	Gm08+Gm14		177 kbp
		g2061	STS	Gm14+Gm02		360 kbp
		SAA19	SCAR	Gm14		617 kbp
		IAC102	SSR	Gm14+Gm02		633 kbp
		SSR-IAC183	SSR	Gm14+Gm02		633 kbp
		D1234	STS	Gm12+Gm14		701 kbp
		PvM119	SSR	Gm02+Gm14	s00322: 352 kbp	304 kbp
		SSR[f]	SSR	Gm02+Gm14		235 kbp
		BMc331	SSR	Gm14		175 kbp
		BMc251	SSR	Gm14		175 kbp
		CP-025	SNP	Gm02+Gm14		017 kbp
		SSR[g]	SSR	Gm02-Gm14	s01390: 62 kbp	28 kbp

[a] PhaseolusGenes

[b] Phytozome.net: G19833 (Andean) sequence; scaffolds from Mesoamerican sequence from BAT93 to be added later

[c] SSR_(TA)10_2ndhit_552_PUAB-416l10.b-1_L10_037.ab1

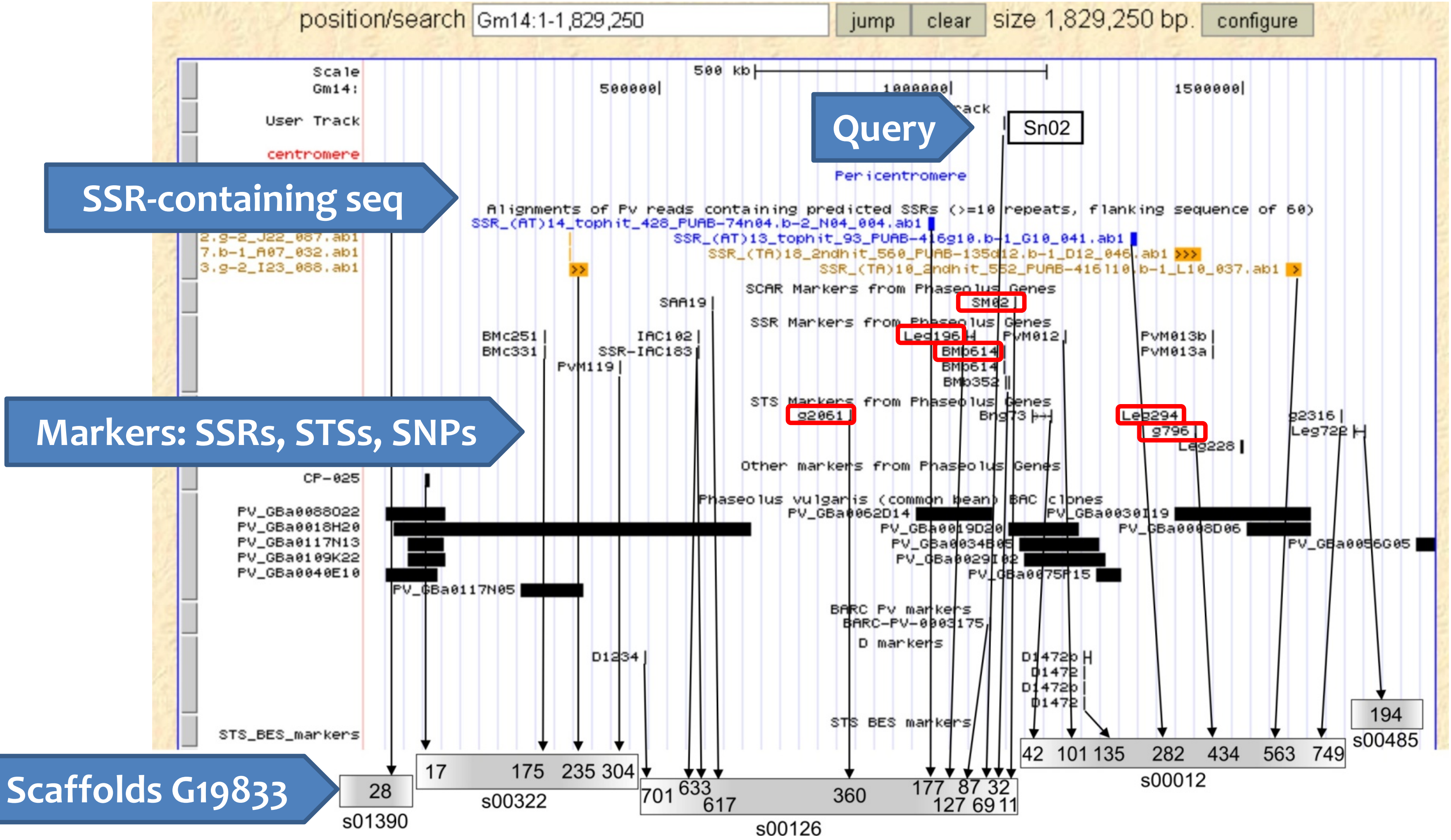
[d] SSR_(AT)13_tophit_93_PUAB-416g10.b-1_G10_041.ab1

[e] SSR_(AT)14_tophit_428_PUAB-74n04.b-2_N04_004.ab1

[f] SSR_(AT)16_2ndhit_428_PUAB-88f07.g-2_F07_028.ab1

[g] SSR_(CT)12_2ndhit_353_PUAB-140i23.g-2_I23_088.ab1

PhaseolusGenes: Scaffolds with markers from Genome Browser


position/search: Gm14:1-1,829,250

size 1,829,250 bp.

Query

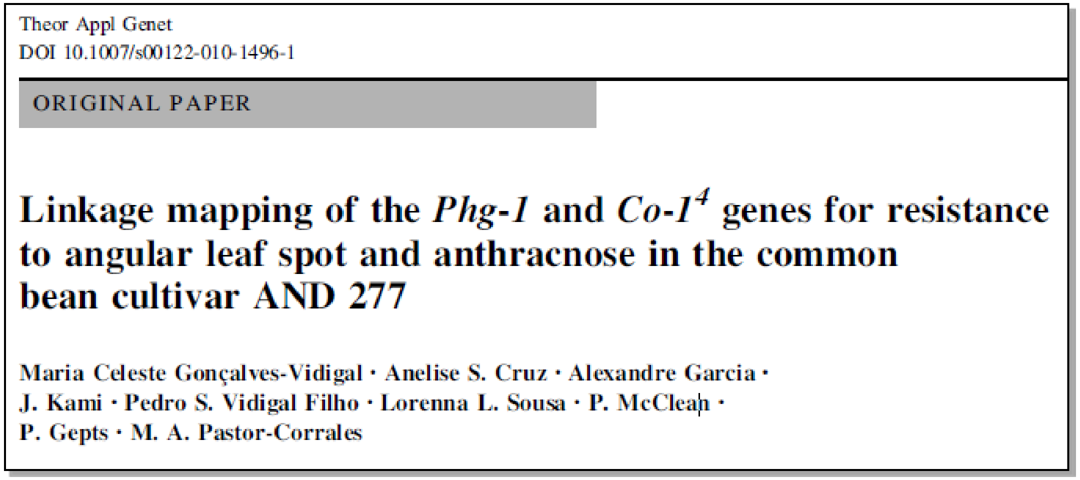
SSR-containing seq

Markers: SSRs, STSs, SNPs

Scaffolds G19833

~25 markers over 2 Mbp

Examples of Synteny Utilization



Theor Appl Genet
DOI 10.1007/s00122-010-1496-1

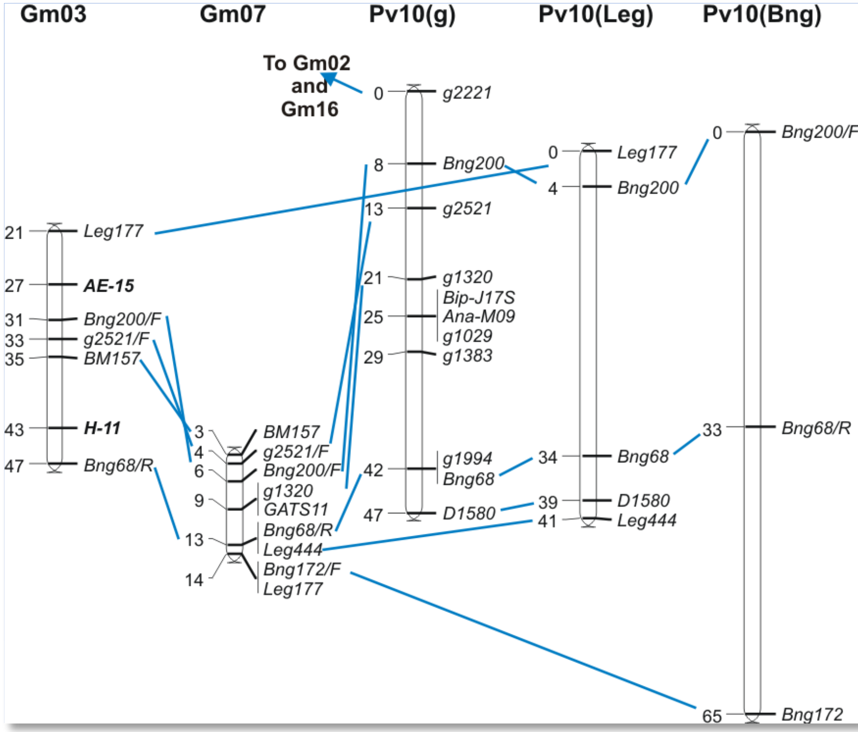
ORIGINAL PAPER

Linkage mapping of the *Phg-1* and *Co-1*^{4} genes for resistance to angular leaf spot and anthracnose in the common bean cultivar AND 277

Maria Celeste Gonçalves-Vidigal · Anelise S. Cruz · Alexandre Garcia · J. Kami · Pedro S. Vidigal Filho · Lorenna L. Sousa · P. McClean · P. Gepts · M. A. Pastor-Corrales

TGA1.1 & CV542014 on Pv01

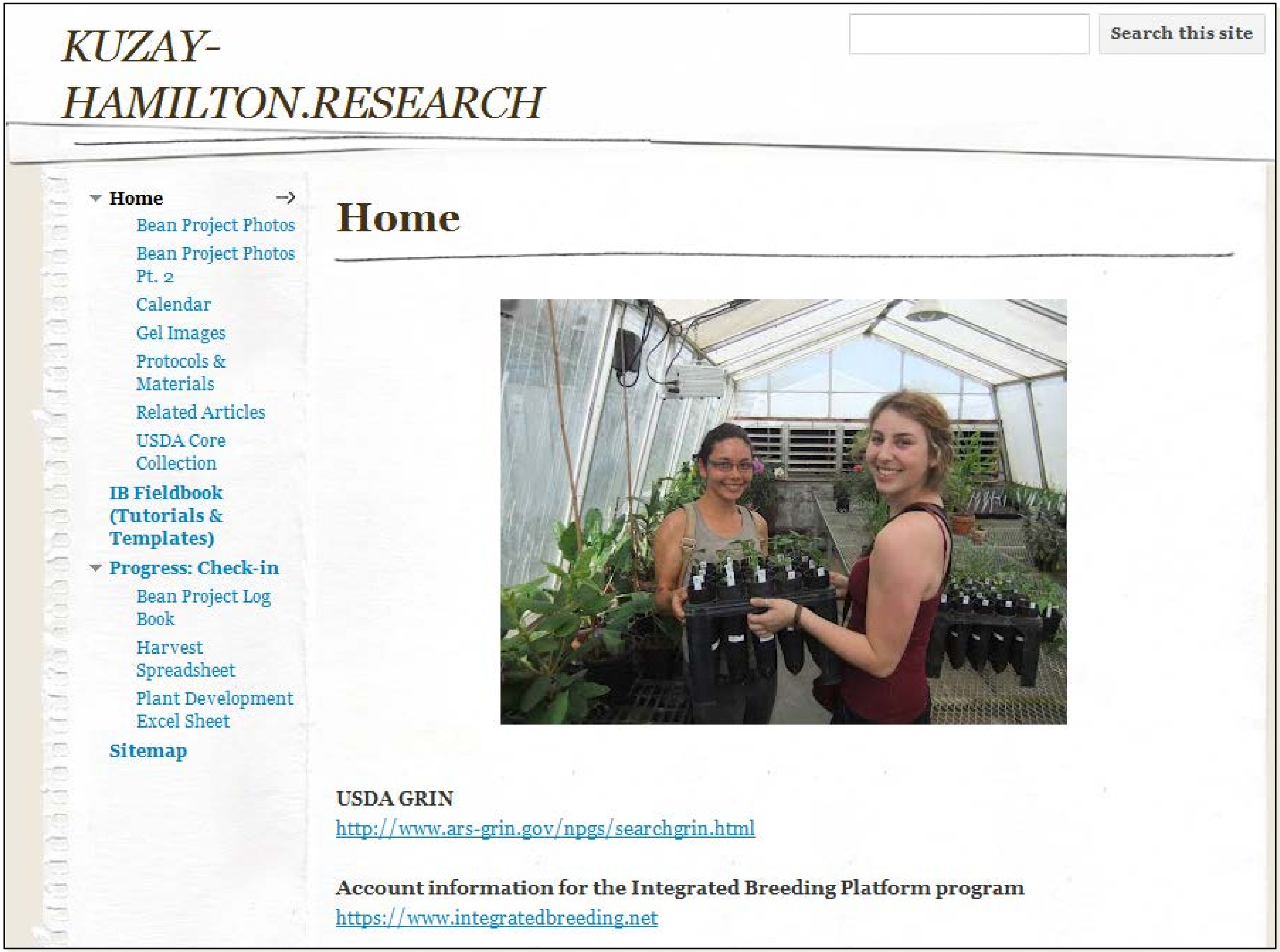
- Shatterproof gene of Arabidopsis on Pv06: Nanni et al. 2011
- *Pse-1* and *Pse-2* on Pv10: Miklas et al. 2011
- Markers surrounding BM154: CIAT study
- Ouro Negro ANT and ALS R genes on Pv04, submitted



Future Plans

1. 1x methyl-filtrated BAT93 (Mesoamerican) Sanger sequencing
2. Two other WGS
3. Addition of newer SNPs
4. Reduced representation or re-sequencing efforts
5. Lima bean sequences, developed at UC Davis, and any other sequences that will be published
6. QTLs on the CMap
7. Speed up the initial search.
8. Other wishes? E.g., phenotypic data of BeanCAP and elsewhere?

DNA Extraction Core Collection



KUZAY-HAMILTON.RESEARCH

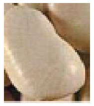
Home

USDA GRIN
http://www.ars-grin.gov/npgs/searchgrin.html

Account information for the Integrated Breeding Platform program
https://www.integratedbreeding.net

Application of genomics to species other than P. vulgaris: Population-specific Genomics in Plant Breeding & Genotyping by Sequencing

Large Lima
(Andean)



P1

UC 92
Bush

x

Baby Lima
(Mesoamerican)

P2

UC Haskell
Viny

F1

Sequencing of P1, P2
BLAST onto bean and soybean genomes
--> physical map
Use RI population for genetic map and QTL analysis
Identify polymorphisms
--> potential markers

Recombinant Inbred population

Genetic Analysis

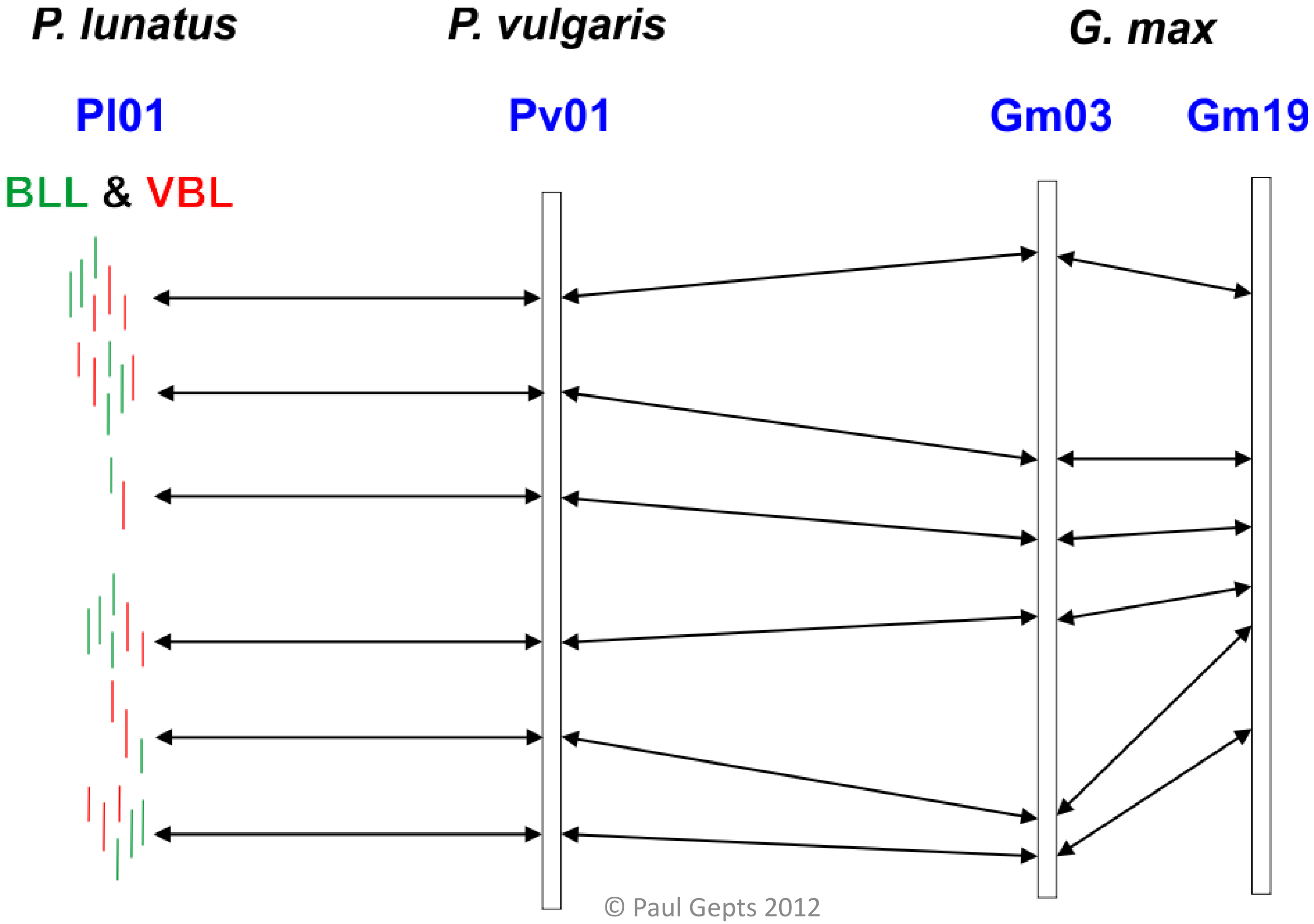
Inbred Backcross population

Other breeding populations

Field Breeding

© Paul Gepts 2012

“Instant” Mapping in Lima Bean through Synteny With Common Bean and Soybean



© Paul Gepts 2012