

Gramene Comparative Genome Views: How Can Gramene Leverage Rice For The Other Grasses?

Ken Youens-Clark¹, Kiran Kumar¹, Immanuel Yap², Pankaj Jaiswal², JunJian Ni², Steven Schmidt¹, William Spooner¹, Wei Zhao¹, Liya Ren¹, Susan McCouch², Edward Buckler^{3,4}, Lincoln Stein¹, Doreen Ware^{1,3}

¹Cold Spring Harbor Labs, Cold Spring Harbor, NY 11724, ²Department of Plant Breeding, Cornell University, Ithaca, NY 14853 ³USDA-ARS NAA Plant, Soil & Nutrition Laboratory Research Unit US Plant, Soil & Nutrition Laboratory, Tower Road, Ithaca, NY 14853-2901, ⁴Institute for Genomic Diversity, Cornell University, Ithaca, NY 14853

Due to the large size and complexity of many of the cereals genomes, finished genomic assemblies are unlikely to be available in the next few years. Many of these genomes will be represented by genomic sequences, ESTs, genetic and finger print contig physical maps. While these resources themselves are useful, it is often not possible to anchor many of the sequences to a physical location in their respective genomes. By leveraging the rice genome assembly, it is possible to order and orient many unanchored cereal sequenced based upon synteny in rice. These alignments will accelerate identification of genes using traditional mapped based cloning and the development of genetic and physical marker resources. Gramene (http://www.gramene.org/), with support of funds from NSF and USDA, contributes to the development and implementation of bioinformatic resources for the plant community. One such resource, CMap, is a web-based comparative genetic and physical map tool that allows a user to dynamically generate comparative map views between the cereal genomes. Gramene implements the EnsEMBL genome browser, to display the rice assembly with anchored cereal annotations and a BLAST view. Both resources provide the users with interactive displays to link within the Gramene database and to act as a web-based portal to other genome resources. Over the next year we will be improving the CMap tool (available at http://www.gmod.org/) and supporting additional plant genome views in the EnsEMBL framework.

Software

CMap and the EnsEMBL genome browser are written in Perl and rely only on open-source tools such as the Apache web server, the MySQL database, the libgd image library, and other Perl modules available on the Comprehensive Perl Archive Network (CPAN). The tools are customizable through configuration files and include tools for loading and interacting with the database

The tools are available from the following locations:

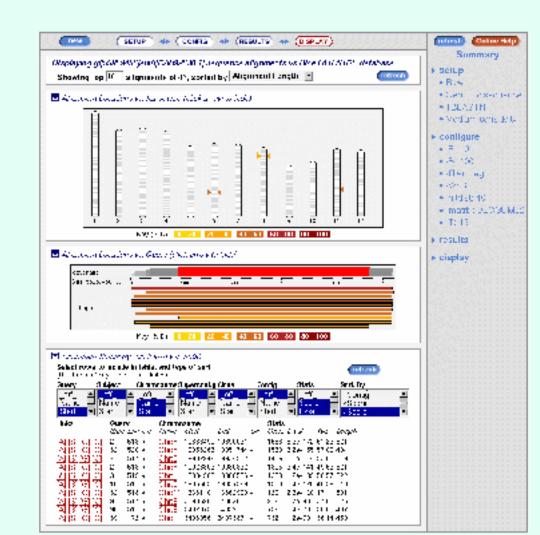
CMap Genome Browser http://www.gmod.org/cmap/http://www.ensembl.org/

How do users use the Gramene website for "Comparative Genomic" analyses?

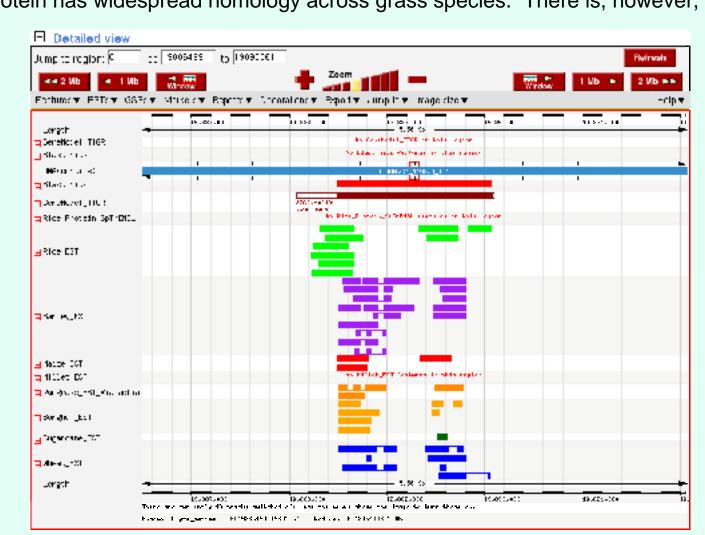
1. A TBLASTN search of a rye protein against the rice genomic sequence was carried out using BlastView. A sensitivity of "allow some local mismatch" was selected, which is a reasonable setting

for the detection of hologous coding regions in related species. NOW SETUP -O- COVER -O- RESILTS -O- CISPLAY Either Paste sequences (max 10) in ±A3. A or plain text: configure. ANTAWVESCROLGELTTE: LATUSPYLATPADVISTIOVOOMFROOOFGKLLFKKGER. 74cd Att Company ▶ results Or Upload ail le containing one or more l'ASIIA sequences : r. Not yet commercial Or Enter an existing taket ID ▶ display g. 1464 year compresse 🗅 ona quartas peptice quartes Salera the databases to courab against. species Sargrum Genomic seguance 1.▼ Feet des (Submittengere models) ■ 🚳 icha databasa . peptitie detabase. Selectine Search Look Scauch sonsi ivi y: Allow some local misso alch 💌

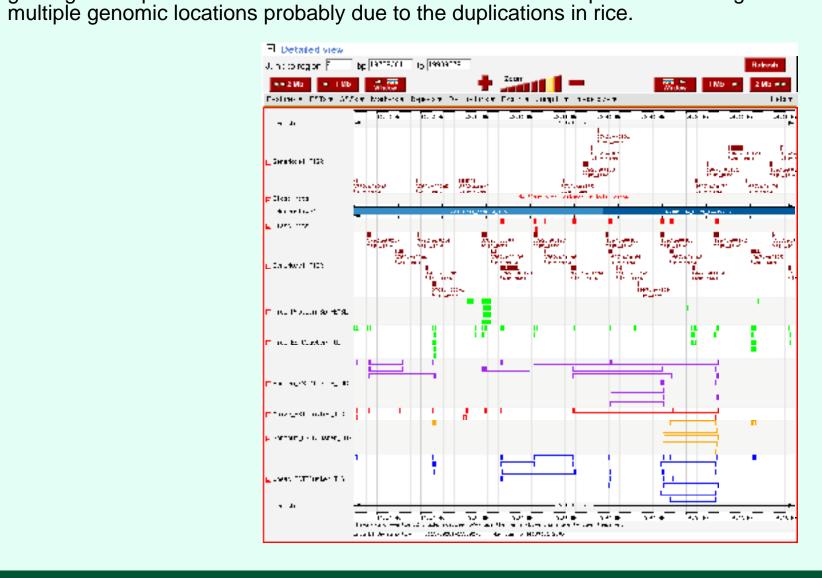
2. The top 10 alignments sorted by length are summarized in both graphical and tabular for-

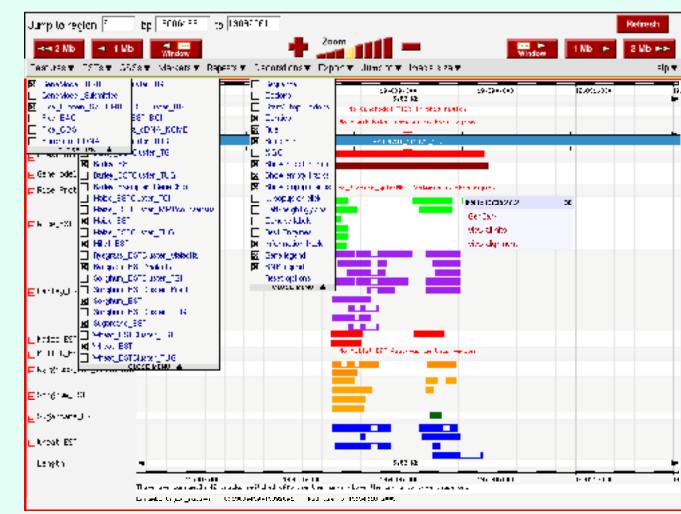


3. The link to ContigView was followed for the highest scoring alignment. This displays the region of the alignment extended by 2000 bp flanks. The ContigView display was configured to show rice gene models, proteins, and ESTs for several species. The rye protein corresponds very closely with an annotated rice gene model, and there are corresponding ESTs for all species except millet. This suggests that the protein has widespread homology across grass species. There is, however, no SpTrEMBL protein mapped to this region.

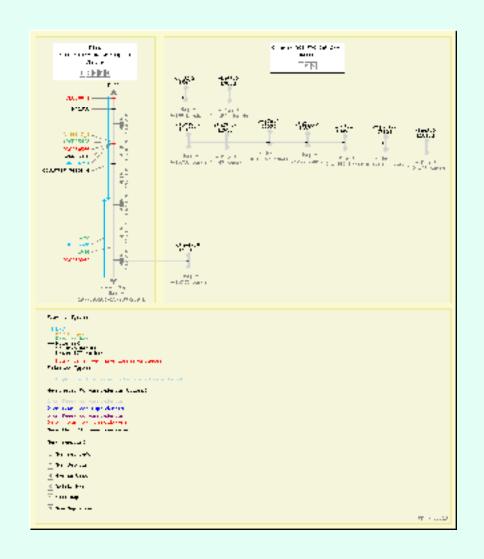


4. By using ContigView's zoom controls, a larger genomic area can be viewed - in this case 200,000bp. As the depth of overlapping EST features can be large at such resolutions, the EST tracks have been replaced by EST cluster (TUG) tracks. The rye protein sequence has seven significant alignments in this region, each spanning the greater part of the length of an annotated Rice Gene Model. This pattern suggests gene duplication in rice. The EST clusters for non-rice species are elongated as they match against





5. The CMAP TIGR assembly was displayed by following the link from ContigView's 'jump-to' menu. The Maize Curated AGI FPC map has been opened for comparison.



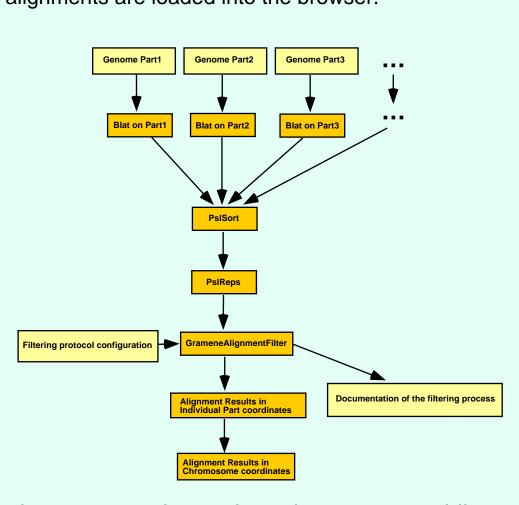
How does 'Comparative Genomic' data get into Gramene?

Alignment Pipeline of the Comparative Datasets to Rice Genome

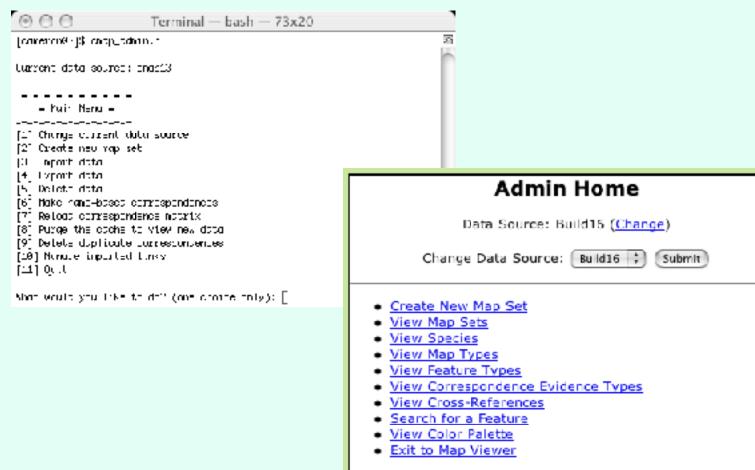
BLAT is used to generate alignments.

These alignments are passed through various filtering steps which are configurable
 Documentation for the whole alignment process is generated automatically based

on the configuration of alignment parameters and filtering parameters
Different alignment and filtering protocols are used for different types of datasets.
The filtered alignments are loaded into the browser.



CMap has two curation main tools: a command-line tool for batch uploads and a web-based administration interface.



Other Comparative Views

Here are additional comparative views from CMap showing synetry among

many species (rice, maize, sorghum

and wheat)

± bi − top Vina

File of the second state o

Fig.	K IAF Fair ABF Sai							
The state of the s				activate feature and all the		AC 15th home shall be call.	721 X.F.	
1	position atte		,		1			
## 14 15 15 15 15 15 15 15	**************************************							
### Company of the Co	EX. 443.000 TA -	<u>_2</u> -	rushtas.	48	. H 4885 = -	=4-1 44		7 104
STATE OF THE PROPERTY OF THE P		41			1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			
Section Sect					1 NO P 11	50 50 50 50 50 50 50 50 50 50 50 50 50 5	A95-1.57-1. F.	
### To be a second of the control of	Petr Nings	404	metatatieni	10 No.	W (7 a) 1924	REPORT OF THE PARTY OF THE PARTY OF THE PARTY.	(Sept. 1)	The second second
### 15 19 19 19 19 19 19 19	# SH I/	-4:	6 108055 = TERM	*** **	★ 付款を表表する。	No. 1986 (1985) 100 -		
Service of the control of the contro	zent Attyat (A/XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	37	Sign			A - SE ON SHARLA		THE RESERVE OF LAND CONTRACTOR
Service of the control of the contro	20 JST41047 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	45 -		Manager of the Control of the Contro	2000年中央1997年	The state of the s	A 1 1 2 3 2	
Service of the servic	and the same of the same		Sandaria (Sandaria)	杨杨小素篇47 翻译等于	大田東 共成で作品で む。	7 E E E	, II. I	A PART OF THE
wheat the state of	acc 19 h	#:	**************************************	你們我你Symm **[[<u>##</u>] レガ	Terretor.	Service California Company Control	1907.10	William Street
wheat the state of	20 J. 18 C. 18 T.	STREET-LINE	- ******* =	第四位的 二二二十二十二	(2) 本書 「「「「「「「「「」」」」	我的人,有 实验上的工工工作的 1967年中央主义	717 2 287 3 24	Conversed
The second secon	17 1000 1 1 1 1 2 2 2 2 2 2 2				ジェング 田中原をあったのださん	Control of the Contro		
Service Control of the control of th			######################################	A2 m201 201 4 - 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	10 (1994) H. (1994) Same of the con-	The state of the s	- 1	· · · · · · · · · · · · · · · · · · ·
The second secon	a-1000 = 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	J. The Royal P	**************************************	A 1900—— \$11 E 2	一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一 一		Z*4 T	※到車を対かくは4 10 V m
The second secon				CST-PERFORMANCE III	- H ****	CSU098720490-1		THE RESERVE OF THE PERSON OF T
Set to instance of the control of th			er filologica 🖽	400 mm	III Characa	- THE - PARKET	-\$-1	S T T T T T T T T T T T T T T T T T T T
** Table 1 - And 1 - A			- 4a 1 -51,51-5,31,51 - 1		- 15 - 10 - 2			- 10g (1 1 10 1070,000 kg
Analysis of analysis Analysis of analysis Analysis of analysis Analysis of analysis The bounds Analysis of analysis The bounds Analysis of analysis Analysis of analysis Analysis of analysis The bounds of analysis The bounds of analysis								
- Analysis of a second								
date the second state of the second s								
- To the second	delector WT							
Signature of the second	— taxa							
The design of the second control of the seco	11/1 / T							
Fallow Spani of the a law the as +1 metal constant for Spania To real to the Spania To law the spania To law the spania	- des Silvata							
Service Law through the serviced For the Law through the Control of the Control		rea.						
Transport to the state of the s		constant.						
· Par Maralla Traffic Annua								
$\overline{\underline{}}$ to h_{ab}	day of Application							
-	_							
	Transaction							

What 'Comparative Genome' data is there in the database?

Mapping of 3rd party sequence databases to the TIGR2 Oryza sativa (ssp. japonica) genome assembly. Results for each mapping contribute a track to the Gramene Continuity display

Database/Track name (SequenceType_Project)	Sequences Total Aligned Alignments
Rice GeneModel_TIGR: Protein_SpTrEMBL: BAC: CDS: EST: ESTCluster_TGI: ESTCluster_TUG: Marker_RFLP: Marker_SSR: Marker_TOS17:	58561 58561 23881 36620 2172 12617 1954 7864 193240 455341 51863 182863 48554 118585 6753 6753 2625 3083 16328 16793
RiceBrachyantha BACend_OMAP:	34930 57825
RiceIndica EST_BGI: ESTCluster_BGI:	67893 152320 19775 52443
RiceJaponica BACend_IRGSP: cDNA_KOME:	71673 76867 23485 103433
RiceNivara BACend_OMAP:	76621 86263
RiceRufipogan BACend_OMAP:	53049 59178
Barley EST: ESTCluster_TGI: ESTCluster_TUG: Exemplar_GeneChip:	226577 616030 23221 80407 25867 80421 10987 46316
Maize BACend: EST: ESTCluster_MMPcornsensus: ESTCluster_TGI: ESTCluster_TUG: HiCot_Bennetzen: HiCotCluster_TIGR: HiCotMethylFilterCluster_TIGR: MethylFilter_CSHL: MethylFilter_Orion: Mu_insert: Marker:	73397 150596 175532 467840 6778 28126 23760 83180 24175 73938 73686 145231 34572 83099 52421 143021 12559 22750 127294 255765 29544 49040 2750 11529
Millet EST:	694 1836
Ryegrass EST_Vialactia: ESTCluster_Vialactia: MethylFilter_Orion: MethylFilterCluster_Orion:	10210 27825 4893 14790 36216 68577 21859 50487
Sorghum CDNA: EST: ESTCluster_Pratt: ESTCluster_TGI: ESTCluster_TUG: GSS_Klein: Marker: MethylFilter_Orion:	1855 5078 121745 315686 11683 30012 19949 64899 12522 36274 117 235 743 1835 35346 83526
Sugarcane EST:	141702 407248
Wheat EST: ESTCluster_TGI: ESTCluster_TUG: Marker:	296094 814457 47206 146388 33160 94361 3442 9217

CMap Maps by Type					
QTL	93				
Genetic	35				
Physical	5				
Sequence	2				
Deletion	1				

Future Plans

- Adding a GUI front end to configure and run the analysis pipeline
 CMap 0.14 will offer more controls over how maps are selected, aligned,
- oriented, ordered, and ornamented

 The new 'comparative' functionality for the Genome Browser will be the
- The new 'comparative' functionality for the Genome Browser will be the addition of the Arabidopsis assembly and completed Maize contigs into Ensembl Compara

Funded By

Gramene is funded by a Plant Genome Initiative grant from the National Science Foundation, an IFAFS grant from the USDA Cooperative State Research and Education Service (CSREES), and was previously supported by a Specific Cooperative Agreement through the USDA Agricultural Research Service.









