

Gramene Exercises Cereal Genomics Course Cold Spring Harbor, New York, USA October 19th, 2016

These exercises will illustrate the power of comparative plant genomics in research using the resources in Gramene.

Exercise 1. View a phylogenetic tree for a family of transcription factors

In this exercise, we will navigate a phylogenetic tree for plant genes in the TCP family of transcription factors (named after the first characterized protein members: maize <u>TB1</u>, snapdragon <u>C</u>YC, and rice <u>P</u>CF), highlight species-specific orthologs/paralogs with particular GO annotations in the tree. We will then proceed to generate lists of orthologs/paralogs and download both, images and tables with our results.

a. How many orthologs can you identify for maize TB1?

Hint: You may find the answer for this through different approaches. Gramene's new search will give you the quickest answer through a snapshot of the *tb1* (Zm00001d033673) gene tree. Other approaches are described in subsequent exercises.

- 1. Go to <u>www.gramene.org</u>. This is Gramene's homepage.
- 2. Enter TB1 in the search box. This will redirect you to search.gramene.org.
- 3. Find the maize *tb1*(Zm00001d033673) gene. Click on the "Homology" tab.

Answer: There are 54 orthologs of maize TB1 in the current Gramene build #51.



b. What is the most prominent TCP domain among members of the gene tree? How many maize genes have a TCP domain?

Note: By looking at the maize TB1 gene tree in Gramene's genome browser, 3 InterPro domains with TCP features appear to be shared among family members. IPR005333 is considered a "family" of protein domains as it encompasses TCP domains: IPR017887 and IPR017888.

Answer: Again, there are multiple ways to answer a question.

- 1) Via search.gramene.org:
 - a) Go to the "Homology" tab of the Search results for maize *tb1* (see above).
 - b) Click on the most prominent blue colored domain (IPR017887).
 - c) Simple answer: IPR017887 Transcription factor TCP subgroup. Note: From a closer look, wheat genes in the tree appear to be lacking protein domain annotation even though their protein sequence lines up with TCP domains.
- 2) Via the genome browser:
 - a) From the "Homology" tab in your search results (see above), click on "Ensembl Gene Tree view" OR go to ensembl.gramene.org, search for maize TB1 and click on (Plant Compara) Gene Tree (EPIGT00820000103607).

- b) Select InterPro domains in the annotations table. By selecting an individual domain, all members that share it will be highlighted in the tree.
- c) Detailed answer:
 - i) 479 members have <u>IPR005333</u> Transcription factor, TCP.
 - ii) 477 members <u>IPR017887</u> Transcription factor TCP subgroup
 - iii) 141 members IPR017888 CYC/TB1, R domain

Zea mays (AGPv4) V Location	n: 1:270,553,676-270,554,776 Gen	e: Zm00001d033673 Tr	rans: Zm00001d033673_T001 Jobs V	
Gene-based displays				
Summary Splice variants	Gene: Zm00001d033	673		
- Transcript comparison	Description	Zm00001d033673	1	
	Location	Chromosome 1:2	70.553.676-270.554.776 forward strand.	
Secondary Structure	About this gene	This gone has 1 tr	anecript (splice variant) 53 orthologues and 13	Z paralogues
 Gene families 	Transmitht	This gene has T u	anachpt (aprice variant), 55 orthologues and 17	paralogues.
Literature	Transcripts	Show transcript	table	
Genomic alignments				
- Gene tree	Gene tree Ø			
 Gene gain/loss tree 				
- Orthologues	GeneTree EPIGT0082000010	3607		
Paralogues	Number of genes	497		
- Gene Tree	Number of speciation nodes	361		
Ontologies	Number of duplication nodes	128		
- GO: Biological process	Number of ambiguous nodes	6		
- GO: Molecular function	Number of sens calls suggests	0		
- GO: Cellular component	Number of gene split events	1		
Genetic Variation	Highlight annotations	Hide annotations	table Clear highlighting	
- Structural variants				
└ Variant image				X
- Gene expression	Show 10 rentries			Filter
 Regulation External references 	highlight	Accession	Description	
 Supporting evidence 	479 members	IPR005333 @	Transcription factor, TCP	
ID History Gene bistory	• 477 members	IPR017887 @	Transcription factor TCP subgroup	
Aconfigure this page	141 members	IPR017888 密	CYC/TB1, R domain	
	4 members	IPR020467	Potassium channel, voltage dependent,	Kv1.4
者 Custom tracks	0 1 member	IPR001932	PPM-type phosphatase domain	
🛃 Export data	O 1 member	<u>IPR032675</u> 문	Leucine-rich repeat domain, L domain-lil	ke
Share this page	Showing 1 to 6 of 6 entries (filte	red from 297 total entries)		< 1 > >
+ Bookmark this page				
Gramene is produced in	SELMODRAFT 99227 Selaging	lla moellendorffi		
collaboration with Ensembl Plants	S SELMODRAFT 104521 Sebair	ella moellendorffi		
	D SELMODRAFT_104321, SERGI	ginella moellendorffi		
	SELMODRAFT_29102, Sela	ginella moellendorffi		
	o GLYMA13G36840, Glycin	e max		
	GIYMA12G 33640, Glycin	e max		
	TCP10. Prunus persica			

- 3) Customized data dump: Using the BioMart utility.
 - a) Go to http://ensembl.gramene.org/biomart/martview.
 - b) Select Database: "Plant Genes" and Dataset: "Zea mays genes".
 - c) Under "Protein Domains", select "Limit to genes with these family or (InterPro) domain IDs" and enter "IPR005333", "IPR017887" or "IPR017888".
 - d) Click on "Count". Alternatively, under "Attributes" select the associated data (e.g., gene or transcript ID, position, sequence, variants, GO terms, etc.) that you would like to download for these genes.
 - e) Answer: There are 46 maize genes with IPR005333, 45 with IPR017887 and 4 with IPR017888.

New Count Results	🔶 URL 🛛 🎝 XML	🔄 Peri 🕕 🚯 Help
Dataset 46 / 44300 Genes	Please restrict your	query using criteria below
Zea mays genes (AGPv4 (CampbellMaker2015Dec))	REGION:	
Filters	GENE:	
InterPro ID(s): [ID-list specified]	GENE ONTOLOGY:	
Attributes	PLANT ONTOLOGY:	
Transcript stable ID	■ ENVIRONMENT ONTOLOGY:	
	■ GRAMENE TAXONOMIC ONTOLOGY:	
Dataset	■ GROWTH STAGE ONTOLOGY:	
[None Selected]	TRAIT ONTOLOGY:	
	■ MULTI-SPECIES COMPARISONS:	
	PROTEIN DOMAINS:	
	□ Limit to genes	with coiled coils (Ncoils) Only Excluded
	Limit to genes with these family or domain IDs:	InterPro ID(s)
	Transmembrane domains	Browse No file selected.
		 Excluded
	□ Signal domains	 Only Excluded

- **c.** You have learned 3 ways to find orthologs for a given gene (via Search, Genome Browser and BioMart). Can you identify the (*Japonica*) rice ortholog of the maize *tb1* gene and highlight both genes in the TCP gene family tree?
- 1) Via Search:
 - a) From the "Homology" tab in Search results (see above), select *Zea mays* and *Oryza sativa japonica* from the drop-down menu on the top right of the Search page.
 - b) Click on "Show Orthologs".

Gramene	Search for genes, species, pa	hways, ontology terr	ns, don	nains		0	2 genes in 2 genomes -
- 1	Gene Tree Orthologs of TB1 🗙						
o o o o o o <u>o</u> <u>o</u>	o e e o O. sativa Japonica e emo Zea mays	1	в				
OS03G07065 unknown	500 Oryza sativa Japonica	Group er	Oryza su	stiva Japonica Group –	TCP18 / Transcription	Arabidopsis th factor TCP18	Model Episcies Homolog
TB1 Zm00001d03 Transcription factor TE Location Homole	33673 Zea mays COSINTE BRANCHED 1 297 X-refs		eace on the	CB0300705500 CB0300705500 Cx0300705500 protein; TCP tamby transcription factor containing protein, expressed; Teoloitie timulichest protein; CDNA clone.002;121:007, Juli risent sequence	TCP18 / Transcription	Arabidopsis th factor TCP18	Model Species Homolog Italiana
Compara Gene This phylogram sho	Tree ws the relationships between this	genes ind others si	milar to	it, as determined by Ensem	ıbl Compara.		
Magnolic Magnolic	AT3G18550 2m00014033673 2m00014033677 Pacese: 6 genes, 4 para Pacese: 5 genes, 3 paralogs	logs gs					
Search Gramene			Links to	other resources			
Show All Homolog Show Orthologs Show Paralogs	95 407 50		Ensemb	al Gene Tree view			

- 2) Via Genome Browser:
 - a) From the left side menu of the Gene Summary page or the Plant Compara Gene Tree view (see above), select the (Plant Compara) "Orthologues" option
 - b) Type "japonica" on the "Filter" box to select to show only rice orthologues in the results table.

Gramene BLAST BioMart	Tools Downloads Help Feedba	ack UploadData			Search Zea mays	Login/Register
Zea mays (AGPv4) V Location: 1:	270,553,676-270,554,776 Gene: Zm	00001d033673 Tr	ans: Zm00001d033673_T0	01 Jobs 🔻		
Gene-based displays	Gene: Zm00001d033673					
 Splice variants 						
- Transcript comparison	Description	Zm00001d033673				
E Sequence	Location	Chromosome 1: 2	70,553,676-270,554,776 for	ward strand.		
Secondary Structure	About this gene	This gene has 1 tr	anscript (splice variant), 53 d	orthologues and 17 paralog	Jes.	
 Gene families Literature Plant Compara 	Transcripts	Show transcript t	able			
Genomic alignments Gene tree Gene gain/loss tree	Orthologues @					
Paralogues Pan-taxonomic Compara	Download orthologues					
- Gene Tree	Summary of orthologues of this	gene				
Orthologies Orthologies GO: Biological process	Click on 'Show details' to display the or	thologues for one or n	nore groups of species. Alter	rnatively, click on 'Configure	this page' to choose a	custom list of species.
GO: Molecular function GO: Cellular component	Species set	Show details	With 1:1 orthologues	With 1:many orthologues	With many:many orthologues	Without orthologues
Genetic Variation	All (48 species)		0	28	8	<u>12</u>
 Variant table Structural variants 	Amborellales (1 species)		0	0	0	1
- Variant image	Bryophyta (1 species)		0	0	1	0
- Gene expression - Regulation	Chlorophyta (2 species)		0	0	0	2
- External references	Liliopsida (20 species)		0	19	1	0
ID History Gene history	Lycopodiophyta (1 species)		0	0	1	0
	Rhodophyta (3 species)		0	0	0	3
Configure this page	Eudicotyledons (15 species)		0	9	5	1
Custom tracks	Selected orthologues					
🛃 Export data						X
< Share this page	Show All - entries	S	now/hide columns		japor	.ica
Bookmark this page	Species Type dN/	gene name	entifier & Compare	e Location		%id %id
Gramene is produced in collaboration with Ensembl Plants	Oryza sativa <u>1-to-many</u> 0.1; Japonica	2778 OS03G0706 Os03g0706500 transcription fac protein, express branched1 prot clone:002-121-1 sequence [Sour	1500 - Region Alignme protein; TCP family tor containing sed; Toosinte ein; cDNA 301, full issert rce:	Comparison 3:28428504-28- ant (protein) ant (cDNA) ree (image)	430438: <u>1</u>	57.99 % 61.48 %

c) Click on the "Gene Tree (image) link" for the rice orthologue.



- 3) Via BioMart:
 - a) From the "*Zea Mays* genes" data set in BioMart (see above), under the "Gene" filter, select "ID list limit".
 - b) Enter "Zm00001d033673" as the "Gene stable ID" for maize *tb1*.
 - c) Under "Attributes", select "Homologs".
 - d) From the "Homologs" attributes form, under "Gene Attributes" select "Gene stable ID", and under "Orthologs" select "Oryza sativa Japonica gene stable ID" and any additional data desired (e.g., % identity).
 - e) Click on "Results". Customize how to view and export your results.

New Count Results	🔶 URL 🗛	NML 🚽 Perl 💿 Help	
Dataset 1 / 44300 Genes	Please select columns to be i	ncluded in the output and hit 'Results' when ready	
Zea mays genes (AGPv4 (CampbellMaker2015Dec)) Filters	Features Homologs Structures Sequences Variation		
Gene stable ID(s): [ID-list specified]	GENE:		
Attributes Gene stable ID Oryza sativa Japonica gene stable ID % identity	Gene Attributes © Gene stable ID Protein stable ID Protein stable ID Chromosome/scaffold name Gene start (bp)	☐ Band ☐ Gene name ☐ Source of gene name ☐ Gene description ☐ Gene biotype	
Dataset	Strand	Transcript count	
[None Selected]	ORTHOLOGS: Aegilops tauschii gene stable ID Aegilops tauschii protein stable ID Aegilops tauschii protein stable ID Aegilops tauschii art (bp) Aegilops tauschii art (bp) Representative protein or transcript ID Ancestor	☐ Homology type ☐ % identity ☐ Aegilops tauschii % identity ☐ dN ☐ dS ☐ Orthology confidence [0 low, 1 high]	
	Amborella trichopoda Orthologs Amborella trichopoda gene stable ID Amborella trichopoda protein stable ID Amborella trichopoda chromosome/scaffold Amborella trichopoda start (bp) Amborella trichopoda end (bp) Representative protein or transcript ID Ancestor	□ Homology type □ % identity □ Amborella trichopoda % identity □ dN □ dS □ Orthology confidence [0 low, 1 high]	
	Arabidopsis lyrata Orthologs	n Homoloav tvoe	

Answer: <u>OS03G0706500</u> (IRGSP1) or <u>LOC_Os03g49880</u> (MSU6).

d. Identify genes in the tree that have been associated with flower development.

Hint: GO:0009908 is the GO term identifier for "flower development".

- 1. From the Plant Compara Gene Tree view (see above), enter the term "flower" in the Filter box to identify GO or InterPro term(s) for flower development.
- 2. Select GO:0009908



Exercise 2. Identify tomato transcription factors within the TCP gene family with a SNP that results in a truncated peptide.

Note: The SNP will introduce a stop codon(*) resulting in a truncated protein product.

Hints:

- 1) Via Search and/or Genome Browser:
 - a) From the Transcript page of Solyc06g069240.1, the tomato ortholog of maize *tb1* (proceed as above for the rice ortholog of TB1)
 - b) Select "Domains & features" from the left side menu.

- c) Find the IRP017887 domain and click on "Display all genes with this domain".
- d) Copy the resulting gene list and use it as input to mine for tomato variants with a "stop_gained" as functional consequence.

Solanum lycopersicum (SL2.50) V	Location: 6:43,006,78	31-43,008,127	Gene: Soly	c06g069240.1 Tra	ns: Solyc06g069240	.1.1 Jobs V		
Transcript-based displays								
E Sequence	Transcript: S	olyc06g06	9240.1.1					
- Exons	Location		Chromo	some 6: 43,006,781-4	3.008,127 forward stra	and.		
- cDNA	About this transcrip	ot	This tra	ascrint has 2 exons is	annotated with 8 dom	ains and features and is	associated with 116 variation	8
Protein Information	Cono		The da		annotatoa man <u>o aom</u>	and reating and it		
Protein summary	Gene		This trai	nscript is a product of	gene Solyc06g069240	.1 Show transcript tab	le	
 Domains & features 								
└─ Variants	Domains & fe	eatures @						
Genetic Variation Variant table								
- Variant image	Domains							
 Population comparison 								_
 Comparison image External References 	Show/hide colum	nns					Filter	
- General identifiers	Domain type	Start 🔶	End 🔶	Description	Accession	InterPro		÷.
- Supporting evidence	PANTHER	17	173	-	PTHR31072	-		
E ID History	PANTHER	17	173	-	PTHR31072:SF6@			
 Transcript history 	PANTHER	191	297		PTHR31072 g			
Protein history	PANTHER	191	297	-	PTHR31072:SF6@	-		
Configure this page	PROSITE profiles	255	272	CYC/TB1, R domain	n PS51370@	IPR017888 @ [Display	all genes with this domain]	
者 Custom tracks	PROSITE profiles	115	173	Transcription factor TCP subgroup	PS51369@	IPR017887 @ [Display	all genes with this domain]	
🛃 Export data	Pfam	114	275	Transcription factor, TCP	PF03634	IPR005333 @ [Display	all genes with this domain]	
Share this page	Other features							
+ Bookmark this page								-
Gramene is produced in	Show/hide colum	nns					Filter	i i
collaboration with Ensembl Plants	Feature type			Start		End		÷.
	Low complexity (See	g)		185		204		

2) Via BioMart: First use IPR017887 (TCP domain) to Filter the tomato genes data set. Then select tomato variations as a second data set and under Filters, use "stop gained" as "Consequence type".

,	1_5				1 7	•
New Count Results			🔶 URL 🕴	🛚 XML 🛛 🛃 Perl 🛛 🌘	Help	
	-	2000				
Dataset 34 / 38735 Genes	Export all results to	File		ा 😒	SV 😒 🛛 Unique results	only 🥝 Go
Solanum lycopersicum genes (SL2.50 (2014-10- EnsemblPlants))	Email notification to					
Filters	View	50 😒 rov	s as HTML	🗿 🗹 Unique re	sults only	
Gene type : protein_coding InterPro ID(s); [ID-list specified]	InterPro ID InterPro short description	Gene stable ID	Variation ID	Chromosome name	Position on chromosome (bp)	Consequence to transcript
Attributoe	IPR017887 TF_TCP_subgr	Solyc02q094290.1	vcZ14MA8R	2	54869495	stop gained
Attributes	IPR017887 TF TCP subgr	Solvc01a103780.2	vcZ11HNA9	1	92305148	stop gained
InterPro ID	IPR017887 TF_TCP_subgr	Solyc02g077250.2	vcZ13UI6K	2	42244608	stop_gained
InterPro short description	IPR017887 TF TCP subgr	Solvc02a077250.2	vcZ13UI6W	2	42244773	stop gained
	IPR017887 TF_TCP_subgr	Solyc02q068200.1	vcZ13LECQ	2	38218628	stop gained
	IPR017887 TF_TCP_subgr	Solyc02g068200.1	vcZ13LECR	2	38218656	stop_gained
Dataset 28964 / 71156260	IPR017887 TF_TCP_subgr	Solyc09g008030.1	vcZ1PBLRI	9	1505249	stop gained
SNPs	IPR017887 TF_TCP_subgr	Solyc09g008030.1	vcZ1PBLSE	9	1505460	stop_gained
Solanum lyconersicum variations	IPR017887 TF_TCP_subgr	Solyc08q048390.1	vcZ1NTIMO	8	13139333	stop gained
(SI 2 50 (2014-10-	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTINJ	8	13139565	stop_gained
EnsemblPlants))	IPR017887 TF_TCP_subgr	Solyc08q048390.1	vcZ1NTIOE	8	13139834	stop gained
Filtere	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIOL	8	13139867	stop_gained
Filters	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIP7	8	13140047	stop_gained
Consequence type :	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIP9	8	13140068	stop_gained
stop_gained	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIPJ	8	13140137	stop_gained
Attributes	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIPL	8	13140160	stop_gained
Gono stable ID	IPR017887 TF_TCP_subgr	Solyc08g048390.1	vcZ1NTIPS	8	13140221	stop_gained
Variation ID	IPR017887 TF_TCP_subgr	Solyc06g069240.1	vcZ1HQBH8	6	43008047	stop_gained
Observation	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDAV	5	5987540	stop_gained
Chromosome name	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDAY	5	5987573	stop_gained
Position on chromosome (bp)	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDBG	5	5987705	stop_gained
Consequence to transcript	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDBU	5	5987804	stop_gained
	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDC2	5	5987870	stop_gained
	IPR017887 TF_TCP_subgr	Solyc05g012840.1	vcZ1CCDCG	5	5988050	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT61M	8	13379406	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT62A	8	13379763	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT62P	8	13379946	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT638	8	13380111	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT63M	8	13380265	stop_gained
	IPR017887 TF_TCP_subgr	Solyc08g048370.2	vcZ1NT63T	8	13380326	stop_gained
	IPR017887 TF_TCP_subgr	Solyc05g032780.1	vcZ1EDEOS	5	44839431	stop_gained
	IPR017887 TF_TCP_subgr	Solyc05g032780.1	vcZ1EDEPQ	5	44839867	stop_gained

Exercise 3. Explore the genetic variation associated with a gene

We will now explore genetic variants along the Arabidopsis *cle18* gene to find 2 SNPs reported to have drastic functional consequences for the CLE18 peptide. CLE18 is a CLAVATA3/ESR-related (CLE) peptide with diverse roles in plant growth and development. Two functional consequences (Cao *et al*, 2011)[Nature Genetics].

- **a.** Visualize the genetic variants for this gene
- **b.** Are there any new stop codons introduced (nonsense variants) in this gene? Compare your findings with Supplementary Table 3
- c. Are there any transcript-specific variants for this gene?
- **d.** Download a subset of the variants (e.g., those that introduce an amino acid change in the protein)

Note: In addition to the Ensembl "Tools" for genomic analysis, other genetic analysis (e.g., Simple Sequence Repeat Identification Tool or SRIT) tools can be accessed through Gramene's archival Diversity pages at http://archive.gramene.org/diversity/tools.html

Exercise 4. Explore a metabolic pathway and compare it in two species

Let's now look at the carotenoid biosynthesis pathway in Gramene. You will notice that we currently offer two pathway platforms: Plant Reactome (rice reference pathways & orthology-based projections to 64 species) and BioCyc-based (10 plant species now served via Cyverse @ http://pathway.iplantcollaborative.org).

- a. Search for "carotenoid biosynthesis"
- **b.** Browse through results categories
- **c.** Select annotations in the Plant Reactome (rice)
- d. Download a list of proteins associated with this rice pathway in Reactome
- e. Go back to the search results and look up for what species have annotated pathways in the BioCyc platform
- **f.** Select rice and compare with maize
- **g.** Download a list of all the genes associated with a carotenoid biosynthesis pathway in each, rice and maize
- **h.** Check out other resources for maize pathways in MaizeGDB.Org

Exercise 5. Upload, visualize and share your own data into a new genome browser track

The Ensembl genome browser allows users to upload their own data and visualize it on a custom track. Data may be formatted in various file formats including GFF, GTF, BED, BAM, VCF, bedGraph, gbrowse, PSL, WIG, BigBed, BigWig, and TrackHub. Some data like GFF annotations may be directly uploaded from a local machine. Large data files like BED/BAM alignments or BigWig graphic display configurations need to be uploaded onto a local server that is accessible to the browser via an URL. Another way to share third-party data is via a DAS (Distributed Annotation System) registry, which would need to be set up by a software engineer.

The test data sets that we will upload and visualize for this exercise have been preloaded onto a local server that is publicly accessible: http://data.gramene.org/public/Zea_mays4m/methylome/. The data consists of BAM alignments and CpG methylation for B73 & Mo17 maize lines used in the study by Regulski *et al* (2013) [Genome Research 23:1651] and were used to create expression tracks in Gramene build 45.

Copy/paste the following sample VCF:

```
##fileformat=VCFv4.0
##fileDate=20161018
##source=MaizeHapMapMockUp
##reference=RefGenv4
##phasing=no
##INFO=<ID=MQ,Number=.,Type=Float,Description="RMS mapping quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
#CHROM POS
             ID
                   REF
                        ALT
                               QUAL FILTER INFO FORMAT
B73:MZ M97:MZ MKN009:MZ
                            MKN010:MZ
                                          MKN011:MZ
    46100
            PZE010000071 T
                                  . PASS MQ=92 GT
1
                              С
                                                            0/0
0/0
    0/0
         0/0
              0/0
    46232
           PZE010000203 G A .
                                        PASS MQ=91 GT
1
                                                            0/0
0/0
    0/0 0/0
             0/0
  Zea mays (AGPv4) VEP V
```

Web Tools But Tools But Tools But Tools Ver analysis of pasted data Assembly Converter ID History Converter	Variant Effe	tics 🖯	or result	ts Ø														
Configure this page	Category	(ount	Consequence	is (all)			Coding cons	equence	s								
* ****	Variants processe	rd á																
Custom tracks	Variants remainin	g after filtering 2																
els Export data	Novel / existing vi	ariants C	(0.0%) / 2 (100.0%)														
	Overlapped gene	s 2			• d	winstream_gene_var	iant: 87%			e sync	nymous_vari	int: 50%						
Share this page	Overlapped trans	cripts 8			• sj	issense_variant: 7%	70			 miss 	ense_variant	50%						
- Bookmark this pape	Overlapped regul	atory features -																
Gramene is produced in collaboration with Ensembl Plants																		
	Results preview	1																
	- Nevigetion			Q Filters				네 Do	wnload									
	Onungation							(approximation) and a		MOT NO	TO THE							
	Page: et al 1 of	1 🕞 👘 丨 Shor	v: <u>1 All</u> varia	unts Uploaded variant	🔁 is	defined		Add All: BioMa	art: Varia	ants g G	enes@							
	Page: (a) a 1 of Show/hide colu	1 🕞 🍋 I Shor	v: <u>1 Al</u> l varia	unts Uploaded variant	is	C defined		Add All: BioMa	art: <u>Vari</u> a	ants g G	enes:9							
	Page: e a 1 of Show/hide colu Uploaded variant	1 ())) I Short mms (13 hidden) Location	v: <u>1 Al</u> l varia	unts Uploaded variant	is Impact	C defined	Feature type	Feature	art: Varia	Exon	cDNA position	CDS position	Protein position	Amino acids	Codons	Existing variant	Distance to transcript	Feature strand
	Page: e a 1 of Show/hide colu Uploaded variant P2E0100000071	1 ()))) Short mms (13 hidden) Location 1:46100-46100	v: <u>1 Al</u> l varia Allele C C I	Ints Uploaded variant	is Impact	Gene	Feature type Transcript	Feature	art: Varia	Exon	cDNA position	CDS position 948	Protein position 316	Amino acids N	Codons AAT/AAC	Existing variant PZE0100000071	Distance to transcript	Feature strand
	Page: et al of Show/hide colu Uploaded variant PZE010000071 PZE010000071	1 (F)	v: 1 All varia Allele C C C	Donsequence synonymous_variant downstream_gene_variant	is Impact	defined Gene Zm00001d027230 Zm00001d027231	Feature type Transcript Transcript	Add All: BioMa Feature Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T001	Exon	cDNA position	CDS position 948	Protein position 316	Amino acids N	Codons AAT/AAC	Existing variant PZE0100000071 PZE0100000071	Distance to transcript 4777	Feature strand 1 -1
	Page: 1 of Show/hide colu Uploaded Variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	1 ► ► I Short mns (13 hidden) Location 1:46100-46100 1:46100-46100 1:46100-46100	x: 1 All varia	Deploaded variant	is Impact	defined Gene Zm00001d027230 Zm00001d027231 Zm00001d02723 Zm00001d027 Zm00001d027 Zm00001d027 Zm00001d027 Zm00001d027 Zm00001d027 Zm00001d027 Zm00001 Zm00001 Zm00001 Zm0000 Zm000 Zm000 Zm00 Zm00 Zm00 Zm00 Zm0 Zm00 Zm0 Zm0 Zm0 Zm0 Zm0 Zm Zm0 Zm Zm0 Zm Zm0 Zm Z	Feature type Transcript Transcript Transcript	Add All: BioMa Feature Zm00001d02723 Zm00001d02723 Zm00001d02723	art: <u>Varia</u> 10_T001 11_T001 11_T002	Exon 3/9	cDNA position : 1010	CDS position 948 -	Protein position 316 -	Amino acids N	Codons AAT/AAC	Existing variant PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777	Feature strand 1 -1 -1
	Page: ≪ < 1 of	1 (► (►) 1 Short mns (13 hidden) Location 1:46100-46100 1:46100-46100 1:46100-46100	x: 1 All varia	Ints Uploaded variant Consequence synonymous, variant downstream, gene, variant downstream, gene, variant	Impact	Gene Zm00001d027230 Zm00001d027231 Zm00001d027231 Zm00001d027231	Feature type Transcript Transcript Transcript Transcript	Add All: BioMa Feature 2m00001d02723 2m00001d02723 2m00001d02723 2m00001d02723	ent: Varia 80_T001 81_T001 81_T002 81_T003	Exon 3/9 -	cDNA position 1010 -	CDS position 948 - -	Protein position 316 - -	Amino acids N - -	Codons AAT/AAC - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777 4785	Feature strand 1 -1 -1 -1 -1
	Page: (a) (b) Show/hide colu Uploaded (c) Uploaded (c) (c) Variant (c) (c) PZE010000071 (c) (c) PZE010000071 (c) (c) PZE010000071 (c) (c) PZE010000071 (c) (c)	1 ► ► I Show mns (13 hidden) Location 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100	Allele C C I C I C I C I C I C I	Inte Uploaded variant Consequence synonymous, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant	LOW MODIFIER MODIFIER MODIFIER MODIFIER	defined Gene Zm00001 d027230 Zm00001 d027231 Zm00001 d027231 Zm00001 d027231 Zm00001 d027231	Feature type Transcript Transcript Transcript Transcript Transcript	Feature Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T001 11_T002 11_T003 11_T004	2007 VE ants (2) G Exon 3/9 - - -	cDNA position 1010 -	CDS position 948 - -	Protein position 316 - -	Amino acids N - -	Codons AAT/AAC - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777 4785 4787	Feature strand 1 -1 -1 -1 -1 -1
	Page: <1 of	1 > > > > > > Show mms (13 hidden) Location 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100	Allele C C I C I C I C I C I C I C I C I	Ints Uploaded variant Consequence synonymous, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant	LOW MODIFIER MODIFIER MODIFIER MODIFIER MODIFIER	defined Gene Zm00001d027230 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231	Feature type Transcript Transcript Transcript Transcript Transcript	East Air BioMa Feature 2m00001d02723 2m0001d02723 Zm00001d02723 2m00001d02723 2m00001d02723 Zm00001d02723 Zm00001d02723 2m00001d02723	art: Varia 10_T001 11_T001 11_T002 11_T003 11_T004 11_T004 11_T005	Exon 3/9 - -	cDNA position 1 1010 - - -	CDS position 948 - - - -	Protein position 316 - - - -	Amino acids N - -	Codons AAT/AAC - - - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777 4785 4787 4827	Feature strand 1 -1 -1 -1 -1 -1 -1 -1
	Page: ● < 1 of	Location 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100	Allele C C I C I C I C I C I C I C I C I C I C	Inte Uploeded variant Consequence synonymous, variant downstream_gene, variant downstream_gene, variant downstream_gene, variant downstream_gene, variant downstream_gene, variant	LOW MODIFIER MODIFIER MODIFIER MODIFIER MODIFIER	Gene Zm00001d027230 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript	Easture Easture Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	0_T001 1_T001 1_T001 1_T002 1_T003 1_T004 1_T005 1_T005	2007 VE ants 22 G Exon 3/9 - - - - -	cDNA position : 1010 - - - -	CDS position 948 - - - - -	Protein position 316 - - - -	Amino acids N - - - -	Codons AAT/AAC - - - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777 4787 4827 4827 4835	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1
	Page: I of Show/hide colu Uploaded variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071	Location 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46232-46232	Allele C C I C I C I C I C I C I C I C I C I C	Uploaded variant Consequence synonymous, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant	Impact	defined Gene Zm00001d027230 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript	Easture Zm00001d02723	0_T001 1_T001 1_T001 1_T002 1_T003 1_T004 1_T005 1_T005 1_T006 10_T001	Exon 3/9 - - - - 4/9	cDNA position 1 1010 - - - - 1047	CDS position 9 948 - - - - 985	Protein position 316 - - - - 329	Amino acids N - - - - - - - - - - - - - - - - - -	Codons AAT/AAC - - - - - - - - - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4775 4785 4787 4825 4825 -	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -
	Page: I of Showhide colu Upleaded variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000073 PZE010000071 PZE010000073 PZE010000071	Location Location Location Li46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46102-46120 1:46232-46232 1:46232-46232	X:1 All varia	Uptoceded variant Uptoceded variant Donssequence synonymous, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant downstream, gene, variant missense, variant	Impact Impact LOW MODIFIER MODIFIER MODIFIER MODIFIER MODIFIER MODIFIER	Gene 2m00001d027230 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	Eeature BioMa Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T002 11_T003 11_T003 11_T005	Exon 3/9 - - - - 4/9 -	cDNA position (1010 - - - - 1047	CDS position () 948 - - - - - 985 -	Protein position 316 - - - - 329 -	Amino acids N - - - - G/S -	Codons AAT/AAC - - - - - - - - - - - - - - - - - -	Existing PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071	Distance to transcript - 4777 4777 4777 4785 4787 4827 4827 4835 - 4845	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -
	Page: I of Show/hide colu Uploaded Variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000073 PZE010000071 PZE010000070 PZE010000071 PZE010000070 PZE010000070	1 E E I Shor mns (13 hidden) Location 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46100-46100 1:46102-46102 1:46232-46232 1:46232-46232	X:1 All varia	this Uploaded variant consequence synonymous, variant downsteam, gene, variant	Impact	Cene Zm00001d027230 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027231 Zm00001d027230 Zm00001d027230	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	Easture BioMar Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T001 11_T002 11_T003 11_T004 11_T005 10_T001 11_T005 10_T001 11_T002	Exon 3/9 - - - - 4/9 -	cDNA position (1010 - - - 1047 -	CDS position § 948 - - - 985 - -	Protein position 316 - - - 329 - -	Amino acids N - - - - G/S - -	Codons AAT/AAC - - - - - - - - - - -	Existing variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000073 PZE010000023 PZE010000023	Distance to transcript - 4777 4787 4787 4827 4827 4825 - 4825 - 4645	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 1 1 -1 -1 -1 -1
	Page: 1 of Showhide colu Uploaded Variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000073 PZE0100000071 PZE0100000073 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203	1 > > > > > > > > > > > > > > > > > > >	X:1 All varia	Uploaded varient ansequence synonymous, varient downsteam, gore, varient	Impact	Cene Zm000016027230 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	All: BioMa Feature 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T001 11_T002 11_T003 11_T004 11_T005 11_T005 11_T005 11_T001 11_T002 11_T002 11_T002 11_T002 11_T003	2007 VE ants (2) G Exon 3/9 - - - - - - - - - - - - - - - - - - -	cDNA position (1010 - - - 1047 - -	CDS position 2 948 - - - 985 - - 985 - -	Protein position 316 - - - 329 - - -	Amino acids N - - - - G/S - -	Codons AAT/AAC - - - - - - - -	Existing variant PZE010000071 PZE010000071 PZE0100000071 PZE0100000071 PZE0100000073 PZE0100000203 PZE0100000203 PZE0100000203	Distance to transcript - 4777 4777 4787 48827 48827 4885 - 4845 4645 4645	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -
	Page: 1 of Showhide colu Uploaded variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000070 PZE010000070 PZE0100000203 PZE0100000203 PZE0100000203	1 • • • • • • • • • • • • • • • • • • •	Allete C C I C I C I C I C I C I C I C I C I C	Uploaded variant apromymous, xariant apromymous, xariant downsteam, gene, variant downsteam, gene, variant	Impact	Cene 2m00001d027230 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231 2m00001d027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	All: BioMa Feature 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 2m00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723 Zm00001d02723	art: Varia 10_T001 11_T001 11_T002 11_T002 11_T003 11_T004 11_T005 11_T005 11_T005 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T002 11_T005	2007 92 ants (2) G Exon 3/9 - - - - - - - - - - - - - - - - - - -	cDNA position 1010 - - - 1047 - - - - -	CDS position \$ 948 - - - - 985 - - - 985 -	Protein position 316 - - - 329 - - - - -	Amino acids N - - - - - - - - - - - - - - - - - -	Codons AAT/AAC - - - - - - - - - - - - - -	Existing variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE0100000071 PZE0100000071 PZE0100000007 PZE0100000003 PZE0100000003	Distance to transcript - 4777 4775 4787 4785 4787 4835 4835 - 4653 4655	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -
	Page: 1 of Showhide colu Uploaded variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000073 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203	1 • • • • • • • • • • • • • • • • • • •	Allele C C I C I C I C I C I C I C I A I A I A I A I A I A I A I A I	Uploaded varient ansequence ansequence ansequence, varient downsteam, gene, varient	Impact	Cene Zm000016027230 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231 Zm000016027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	Local All: Biodes Biodes Peature Zm0000102723 Zm0000102723 Zm0000102723 Zm00010102723 Zm000102723 Zm00010102723 Zm00010102723 Zm00010102723 Zm00010102723 Zm00010102723 Zm00010102723 Zm00010102723 Zm00010102723	art: Varia 0_T001 11_T001 11_T002 11_T002 11_T004 11_T005 11_T006 11_T005 11_T002 11_T002 11_T003 11_T003 11_T003 11_T003 11_T005	Exon 3/9 - - - - - - - - - - - - - - - - - - -	cDNA position 1 1010 - - - - 1047 - - - -	CDS position • 948 - - - - - - 985 - - - - - - - -	Protein position (316 - - - - - - - - - - - - - - - - - - -	Amino acids N - - - - G/S - - - - - - - - - - - - - - - - - - -	Codons AAT/AAC - - - - GGC/AGC - - -	Existing variant PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000071 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203	Distance to transcript - 4777 4775 4785 4787 4785 4787 4785 - 4827 4827 4827 4825 4645 4645 4645 4655	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -
	Page: 1 of Showhide colu Uploaded Variant PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203 PZE0100000203	Location 1.46100-46100 1.4632-46232 1.4632-46232 1.4632-46232 1.4632-46232 1.4632-46232 1.4632-46232 1.4632-46232 1.4632-4632 1.4632	Allele C C C C C C C C C C C C C C C C C C C	Uploaded varient asynonymbus, xariant asynonymbus, xariant downsteam, gene, variant downsteam, gene, variant	Impact	Cene Zm000014027230 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231 Zm000014027231	Feature type Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript Transcript	Lett All: Biodat Biodat Feature Zm00001002723 Zm00001002723 Zm00001002733 Zm00001002732 Zm00001002733	art: Varia 0_T001 11_T002 11_T002 11_T002 11_T003 11_T005 11_T006 10_T001 11_T005 11_T004 11_T005 11_T004 11_T005 1	2017 YE was a set of the set of t	CDNA position (1010 - - - 1047 - - - - - - - - - - - - - - - - - - -	CDS position \$ 948 - - - 985 - - - - - - - - -	Protein position 316 - - - - 329 - - - - - - - -	Amino acids N - - - - - - - - - - - - - - - - - -	Codons AAT/AAC - - - - - - - - - - - - - - - -	Existing variant PZE110000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000071 PZE010000072 PZE010000072 PZE010000072 PZE010000073 PZE010000073 PZE010000073 PZE010000073 PZE010000073 PZE010000073 PZE010000073 PZE010000074 PZE010000075 PZE010000075 PZE01000075	Distance to transcript - 4777 4785 4787 4785 4787 4827 4827 4827 4825 - 4845 4645 4645 4645 4655 4695 4695	Feature strand 1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -

Exercise 6. BLAST a sequence. Determine synteny for a genomic region. Convert coordinates between different genome assemblies.

In this exercise, we will identify orthologues of a species whose reference genome is not available in Gramene via BLASTX and find corresponding synteny blocks in other species. **a.** Use the nucleotide sequence of the *Sorghum virgatum* Sh1 gene taken from Lin *et al* (2012) [Nature Genetics 44:720] to identify orthologous genes in maize, rice, and *Arabidopsis thaliana*.

Note: The corresponding gene in S. bicolor appears to be missing two exons.

- **b.** Highlight the orthologs in two of those species in the tree as you learned in Exercise 1.
- **c.** Download the genetic variation for each of the maize Sh1 orthologs as you learned in Exercise 2. How many nonsense substitutions can you find in each of these genes?
- d. Lin et al (2012) also provide RefGen_v2 coordinates for maize shattering QTLs in Supplementary Table 5. Identify synteny blocks for the intervals at maize chromosomal regions (RefGen_v2) chr1: 259,223,260 261,622,457 and chr5: 15,806,322 16,428,681 in rice and sorghum. Download the synteny images that you generate.

Note: You will need to first use the Assembly converter tool to map the QTL intervals to RefGen_v3 and subsequently to RefGen_v4 coordinates. This will be publicly available in the upcoming Gramene build 52.

- **e.** Download all the genes for a given synteny block. Can you identify a *Sh1* orthologous (YABBY-like) gene in it?
- **f.** Compare your results with those in *Lin et al* (2012) [Nature Genetics 44:720]

>S. variegatum Sh1 CDS