FINDING AND FIXING ANNOTATION ERRORS THROUGH COMMUNITY CURATION

Marcela K. Tello-Ruiz, PhD



PB20 PLANT BIOLOGY JULY 27 - 31 | WORLDWIDE SUMMIT





Outline



Cristina F. Marco

Marcela K. Tello-Ruiz

- Gene Structural Annotation & Community Curation What?
- Maize Annotation Jamborees How?
- <u>Course-based</u> <u>Undergraduate</u> <u>Research</u> <u>Experience</u> Why?





GENE STRUCTURAL ANNOTATION





GENE STRUCTURAL ANNOTATION

"The value of a genome is only as good as its annotation"

• Genome sequencing revolution \rightarrow 1,000s new genomes





50 Sorghum

- Annotation errors are prevalent and very \$\$\$
 - For example, CRISPR experiments
- Manual annotation is time-intensive (\$\$\$)





Developing & Customizing Annotation Tools

> Gramene Tree Visualizer: Flag suspicious gene models



http://gramene.org





Andrew Olson

> Apollo Browser: Evaluate and curate flagged genes



http://genomearchitect.org





Kapeel Chougule



Scaling up to Community-Level



Cold Spring Harbor (Dec 2017)



St Louis (March 2019)

PLANT BIOLOGY

- 4 workshops
 (2017-2020)
- 51 participants;
 38 institutions
- 12 PUI faculty, 8 researchers & 26 grad students
- ~250 genes manually curated



San Diego (Jan 2019)



Virtual (March 2020)





Community Annotation – Pilot Study



PLOS ONE

PLANT BIOLOGY

 \mathbf{P}

PUBLISH ABOUT BROWSE

Double triage to identify poorly annotated genes in maize: The missing link in community curation

Marcela K. Tello-Ruiz 🚾, Cristina F. Marco 🔤 🔄, Fei-Man Hsu, Rajdeep S. Khangura, Pengfei Qiao, Sirjan Sapkota, Michelle C. Stitzer, Rachael Wasikowski, Hao Wu, Junpeng Zhan, Kapeel Chougule, Lindsay C. Barone, Cornel Ghiban, Demitri Muna, Andrew C. Olson, Liva Wang, Doreen Ware, David A. Micklos [view less]

Published: October 28, 2019 • https://doi.org/10.1371/journal.pone.0224086

Tello-Ruiz et al (2019)





109 Manually Curated Maize V4 genes

http://www.gramene.org/curated_maize_v4_gene_models



Tello-Ruiz *et al* (2019)

ramene



2019 <u>Course-based Undergraduate Research Experiences</u>









CUREs



Rebecca

Seipelt

REVISING GENES IN THE NITROGEN USE EFFICIENCY PATHWAY OF ZEA MAYS FOR SUSTAINABLE AGRICULTURE Jewel Larkins, Rija Asim, Rebecca L. Seipelt-Thiemann Honors College, Biology Department, Middle Tennessee State University

Results

Table 2 Reference Sequence Domain Evidenc



Introduction

·Maize (Zea mays) is one of the highest produced and most important grain crops worldwide. •Nitrogen fertilizer application has been used to high-yield agriculture as plant growth depends of acquisition of nitrogen, an essential element defi

most soils (Ladha et al. 2016). ·Of the total amount of nitrogen fertilizer applie less than 40% is taken up by the plant (Zuluaga . The remaining nitrogen is lost to the atmosphere greenhouse gases or leached into the soil, polluti and terrestrial ecosystems (Hirel et al. 2011). •Nitrogen Use Efficiency (NUE) comprises a pla to take up nitrogen from the soil and convert it ir vield (Kumar et al. 2018).

·Ultimately, NUE is a trait influenced by various Knowing where in the plant these genes are expr when they are expressed, as well as in what form important in making genetic improvements aime expanding crop yield while being environmental sustainable

Hypothesis

Computationally predicted gene models fo in the nutrient use efficiency pathway of n show both accuracies and inaccuracies that our understanding of maize's role in agriculation sustainability; inaccuracies will be revised generate true, evidence-based gene models

Materials & Methods

·Selection of the Nutrient Use Efficiency in Zea mays and identification of the Refer Sequences (REFSEO) for each gene using ·Identification of differences in conservation amino acids and protein domains in Zea ma related plant species . Corn leaf RNA extraction and reverse tran ·Resuspension of forward and reverse tubu ·PCR and gel electrophoresis on corn leaf cDNA ·Determined splicing patterns observed for gene Zm00001d052233 using Apollo and designed

·Found appropriate annealing temperature for primers through gradient PCR ·PCR and gel electrophoresis on leaf, root, and shoot cDNA for gene Zm00001d052233 ·Analysis of all observed, expected and unexpected isoforms

		distance."	Marine Streement &		descent sector descent sector se	And an average of the local data	Resident discount in the billion of	Openain accus
sustain		General A	2=0000400400	ringer and a second	Overserent 24,04(32)			PANTHER
n the		General 2	2=0000402-07	riser reserved	Dramanert 26,003D			Gene3D Superfamily
in the		term 2	1-000000147	reporters street response	Donated 1843/2			SMART
icient m				2.8	116, CF, 28D for an internet.			Ptam
					10,30,00			CDD
d to crops,		benu's	1-00000-00	transporter ferrife 2	13 412 Differentiations	,		Plan
et al. 2019).		ince i	2+3000940732	Press Ally Property of	Instantic In some		•	DANTHER
e as potent		time 7	2-000040528	Press NEW YORK OF	Destaural 20802			HAMAP
ing aquatic		Generit	2-000000885	PassionElg/EliOastaria	Overseend 10,07,070			PROSITE per
0 1		0wm7	1-000140253	Prose-1475/7717481233	Oversored LLMD-00-		,	
ant's ability		Darie 30		nuu-me;mean.co	Conservation Party		,	< 2x00001405
nto grain		Dana 11		**************************************	THE REPORT OF THE PARTY OF	,		
nto gram			-		10,26,16,10 to a line of			protein coding
					13C06/Millionaniment			< 2m00003.005
s genes.		Garra 1.1	1-10009-0841	annanium transporte 2	Depresent 28,00,00.			protein coaing
ressed and		Garm 3.4	1-000094346	and a standard and a standard as 1	DUH/Minariani			
ns, will be		Oare 13	1-0001-0189	Protection of a particular state	Oversonal MIRAD MUMIPERAWARKS		•	22.93 kg
ed at		Own 30	2-0003424*	Podedkipteptpd Tangora genderhendigt	Overson of Malling		2	
llv		Darie 27		Turnet strappers	2101001001 10.00.00	,		Figure 1. a
		Dana 30		phoset service services	Promotion - Challen	,	,	and includ
		terrs 12	1-000040128	Participation and Participation	Dominant IN ROR	,		
		inter M	In the second se	territorian .	Doctored Distances			
				deread.	100,480,712 (completent)			
		1.070.00		-	11000	_	11 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	
r genes		Letine	1.1.4	- Comment		_		-
a genes							Company of the	- 1
aize will		1-04	a.longf	-	-	_		- 1
t affect				1762				- 1
ultural				Armi			- 0 - 0	- 1
to				42913				-
								199.100
		(migro	a young which	Tab. (1988)		_		199 QP
1.1.1						\$553		
		Fie	ure 2. S	creen Shot of	Apollo Tran	scrint Ex	idence for Zm0	0001052233
1000			Λ		A A	۸	Λ	٨
athway		1	2		4 - 5 -	6	7 8	2
ence		-	∇			$\neg /$		
Gramone				v 1	· ·	~		V
Gramene				10			7	
on or		Th Th	gure 5. P	roposed Gen	e Structure I	or zea m	ays Zm00001a	oszzos.
ays and	and a second	bo	e olack c	sent alternatio	a regions. Int	rons of Z	m00001d05223	3 are
		rer	resented	by horizonta	lines. Angled	lines ab	ove and below th	he image
scription		rep	resent sp	licing pattern	s. The horizo	ntal line b	below the image	indicates I
lin		the	chosen	region of inter	est. Feature s	izes are r	elative to their a	ctual size
mm		in	base pair	s (bp). This st	ructure was d	esigned u	ising RNA evide	nce 1
		pro	vided vi	a the Apollo i	nstance for Ze	ia mays g	enome, version	5

Table 1. Gene Summary Table



	80	265	Histidine phosphatase superfamily, clade-1
	81	155	Histidine phosphatase superfamily, clade-1
	81	295	Histidine phosphatase superfamily, clade-1
	176	284	Histidine phosphatase superfamily, clade-1
R	1	331	Phosphopycerate mutase 1
	79	305	Phosphoglycerate mutase 1
E patterns	83	92	Phosphoglycerate/bisphosphoglycerate mutase, active site
11 (5223), Teo 11 (5223), Teo 11 (5223), Tio 11 (5223), Tio 11 (5252), Tio 11 (5252), Tio 11 (5252), Tio 11 (525), Tio	barneling, barneling,	forence Sequence 7 which are locates	and Domain Evidence. The region near the end of Zm00001d052233
Tabul			
	-		
8	di ute		
shoot	ked namber		

Results

cvision to Observed?

Table 3. Isoform Comparison Table RT-PCR size Transcrip Provides Expected evidence for and/or

						1 Star
237	lseform 1	80	E, O	80	yes	
207	lsoform 2	80	E	yei	80	
315	Isedorm 3	yes	0	unknown	unknown	
400	lseform 4	yes	0	unknown	unknown	12
500	Isedorm 5	yes	0	uakaowa	unknown	
750	lseform 6	yes	0	uakaowa	unkaowa	

Conclusions and Future Directions

·Leaf RNA extraction and reverse transcription was unsuccessful as bands were not present on gel electrophoresis. Jori's leaf RNA was used for the duration of the research. Zm00001d052233 primers were visible at all annealing temperatures

57C was used. ·Reverse transcription on corn leaf, root, and shoot tissues using Zm00001d052233 primers was successful.

 There were two expected isoforms with only one observed after RT-PCR. Additionally, four unexpected isoforms were observed. The expected and observed isoform would likely have no domains missing and therefore be functional. However, the isoform which was not observed would be nonfunctional due to missing domains and alterna splicing. The nonfunctional proteins include those in the histidine phosphatase superfamily and phosphoglycerate mutases. Proteins in the histidine phosphatase family are involved in metabolic regulations intermediary metabolism and developmental processes. Moreover, phosphoglycerate mutase is an important enzyme in the pathway of glycolysis. No functionality in each of these proteins could influence nitrogen use efficiency in Zea mays. These results indicate that the original gene model is not entirely correct and should be revised. •Further research is needed to determine whether the observed isoform which were not expected would have missing domains or he functional

References







Spring 2020:

1 male)

PB20 PLANT BIOLOGY

Regional University

Fall 2019: Honors

Genetics, Sophomores:

18 students (17 female,

Undergrad Research, Soph-Junior (2 male, 1 female)

primers primers for region of interest







Rebecca Seipelt

Outcomes: The Stats (F19)

- ➤ 207 genes examined
- > 9+ genes evaluated in high detail
- > 18 regions tested within 9 genes

Medium term	Student Quote	Long term	Student Quote			
Increased self-efficacy	"Personally, I learned more about my strengths and weaknesses , time management, flexibility, accommodation, trial and error and the importance of learning from mistakes. "	Enhanced science identity	"Not only did this research give me new skills and knowledge, but it helped me boost my confidence of being a science major. It showed me that I have what it takes to be a female student in the STEM field, and			
Increased motivation in	"There is so much that goes on in the background of science that does not get appreciated and I found a		that I can do much more."			
science External validation from scientific community Increased	new passion for science through this experience." "Although many sleepless nights were spent on this research, the realness of it all was the most rewarding aspect. I felt like what I was doing in lab actually mattered beyond a simple grade." "sometimes results don't come in the first try.	Career clarification	"This project personally has meant a great deal to me, because it was the first time I was actually able to experience real lab conditions. It reassured me that I was in the right field of study, and I whole-heartedly enjoyed every minute of the research (minus deadline stress)."			
tolerance for obstacles	But continuing to try is what is important, you can't just give up if something doesn't turn out the way you had hoped."	Persistence in science	"this lab experience because it has given me the confidence to do other research at MTSU, which I was unsure if I wanted to do previously."			







2020 Virtual Jamboree



- 20 awardees (28 participants) from 17 institutions across US
- 5 teaching faculty (3 @ PUIs), 11 graduate students, 1 postdoc, 2 staff scientists
- 4 x 3-hour sessions (March 10-20)
- About 700 genes evaluated with the Gramene Gene Tree Visualizer and 42 genes curated in Apollo













@marcelakarey

Teaching Opportunities

- Example tests:
 - Refined algorithm to select a canonical transcript
 - Previously curated V4 models improved in V5
 - Compared *ab initio* gene predictions (benchmark: 44 genes on chr1)
- Example errors:
 - Internal sequence gain/loss
 - Missing START Met
 - Multiple internal STOP codons





http://curate.gramene.org/nam?gene=Zm00001e018087





@marcelakarey



Zm00001e017759 - Small internal deletion?

ay mode 🗸	Color Scheme -	Multiple Sequence Alignment	nent: Amino acid MSA. Drag slide	to reposition.		K		Curate: click to flag
,								okay Of flag
	Zea mays b73: Zm00001e017759 P001	FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSG PCRL	PFYGG	AAY DY HI YAGORS	FAPLE	flag G,L,_
	Zea mays ki11: Zm00030a018309 P001	MSYSHWAENVVLS	IPALAGPVLVPCHVTTQWLWFSI	NTHSGYHEPESPERL	PFYGG	AAYHDYHHYAGGRSQSN	FAPLE	
Г	 Zea mays p39; Zm00040a027243_P003 		IPALAGPVLVPCHVTTQWLWFSI	NTHSGY HEPESPER	PFYGG	AAYHDYHHYAGGRSQSN	FAPLE	
t .	Zea mays ky21; Zm00031a027169_P003		IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSGY HEPESPER	PFYGG	AAYHDYHHYA <mark>ggr</mark> sQSN	FAPLF	
	2ea mays b97; Zm00018a026085_P004		IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSGYHFPFSPCR	PFYGG	AAY DY HHYAGGRS	FAPLF	
	Contraction 2005 2000 2000 2000 2000 2000 2000 200	FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTOWLWFSICLIEG	NTHSGYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	FAPLF	
	Zea mays cml228; Zm00022a026178_P002		IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSGYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	FAPLF	
	Zea mays m162w; Zm00033a027502_P004	F T AP TG FAM <mark>S</mark> Y <mark>SH</mark> WA <mark>BN</mark> VVL <mark>S</mark>	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSGY HFPFSPCRL	PFYGG	AAYHDYHHYA <mark>GGR</mark> S <mark>OSN</mark>	<mark>FA</mark> PLF	
	Zea mays cml333; Zm00026a024986_	PO(FTAPTGFAMSYSHWANNVLS	IPALAGPVLVPCHVTTQWLWPSICLIEG	NTHSCYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	<mark>FA</mark> PLF	
	Zea mays cml322; Zm00025a024816_	PO(FTAPTCFAMSYSHWARNVVLS	IPALAGPVLVPCHVTTQWLWPSI	NTHSCYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>gors</mark> OSN	<mark>FA</mark> PLF	
	Cea mays nc358; Zm00037a024301_P0	001						
	Zea mays cml69; Zm00020a025040_P00	05-FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTQWLWFSICLIEG	NTHSGYHFPFSPCR1	PFYGG	-AAYHDYHHYA <mark>ggrs</mark> <mark>osn</mark>	FAPLF	
	Cea mays m37w; Zm00032a026069_P004	4F T A <mark>PTG</mark> FAM <mark>S</mark> Y <mark>SHWAENVVLS</mark>	IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSGYHFPFSPCR	PFYGG	AAY DY HHYAGGRSQSN	FA <mark>P</mark> LF	
	Zea mays b73_Ab10; Zm00043a025235_	PO(FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSGYHFPFSPCRL	PFYGG	AAYBDYHHYA <mark>ccr</mark> sQSN	FAPLF	
	Zea maysb73v4; Zm00001d040979_P001	F TA<mark>PTG</mark>FAMSY<mark>SHWA@NVVLS</mark>	IPALAGPVLVPCHVTTOWLWFSICLIEG	NTHSG				
	Zea mays mo18w; Zm00034a027031_	PO(FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSGYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>ccr</mark> sQSN	FAPLF	
	Cea mays cml103; Zm00021a026404_	PO(FTAPTGFAMSYSHWANNVVLS	IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSGYHFPFSPCRL	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	<mark>FAP</mark> LF	
	Zea mays cml277; Zm00024a024811_F	000;FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSCYHFPFSPCRL	PFYGG	AAYHDYHHYA <mark>ccr</mark> s	<mark>FAP</mark> LF	
	Cea mays ki3; Zm00029a026056_P004	FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSCYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>ggr</mark> sQSN	<mark>FA</mark> PLF	
	Cea mays cml247; Zm00023a026259_P0	03-F TAPTG FAMSY <mark>SHWAZNVVLS</mark>	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSGY HFPFSPCR	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	FAPLF	
	C Zea mays oh7b; Zm00038a026081_P004	F TAPTG FAMSY <mark>SHWAZNVVLS</mark>	IPALAGPVLVPCHVTTOWLWFSIRLIEG	NTHSGYHFPFSPCRL	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	FA <mark>PLF</mark>	
	Zea maysb73v3; GRMZM2G062531_P01	F TAPTC FAMSY <mark>SHWAEN</mark> VVLS	IPALAGPVLVPCHVTTQWLWFSIRLIEG	NTHSCYHFPFSPCR	PFYGG	AAYHDYHHYA <mark>GGR</mark> SQSN	FAPLF	
	Zea mays cml52; Zm00019a023352_P005	FTAPTGFAMSYSHWADNVVLS	IPALAGPVLVPCHVTTOWLWFSI	NTHSGYHFPFSPCR	PFYGG	AAYHDYHHYAGGRS	FAPLF	
	Cea mays nc350; Zm00036a017658_P003	FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTOWLWFSICLIEG	NTHSGYHEPESPER	PFYGG	AAYHDYHHYAGGRS	FAPLF	
	Zea mays tzi8; Zm00042a026577_P003	FTAPTGFAMSYSHWANVVLS	IPALAGPVLVPCHVTTQWLWFSI	NTHSGYHFPFSPCRL	PFYGG	AAYHDYHHYAGGIS	FAPLF	
4	Zea mays ms71; Zm00035a025530_P003	FTAPTGFAMSYSHWADNVVLS	IPALAGPVLVPCHVTTQWLWFSI	NTHSGYHFPFSPCRL	PFYGG	AAYHDYHHYA <mark>GOR</mark> SQSN	FAPLF	
¥	Zea mays il14h; Zm00028a025051_P004	FTAPTGFAMSYSHWAENVVLS	IPALAGPVLVPCHVTTQWLWFSIELIEG	NTHSGYHFPFSPCRL	PFYGG	AAYHDYHHYAGGES	FAPLF	
	Sorghum bicolor; OQU86919		IPALAGPVLVPCHVTTQWLWFSIELIEG	NTHSCYHFPFSPSRL	PRYGG	AAYHDYHHYAGG:SQSN	FAPLE	

NOT a deletion \rightarrow RNA evidence for alternative transcript in Apollo







Zm00001e002409 – Multiple STOP codons

Display mode -	Color Scheme -	Multiple Sequence Al	gnment: Amino acid MSA. Dra	ac slider to reposition.	Curate: click to flag genes
					okay Or flag
	Zea mays b73: Zm0000	1e002409 P002TV	P*GFDWE*G*WGGLKGLESLTILF*I	V VSSPLDLLHSIYSOTSP	fiag,G
гţ	Zea maysb73v3: GRMZ	M2G077937 P01	(P		
rt -	Zea mays cml322;	Zm00025a003452 P002 TV	P*GFDWE*G*WGGLKGLESLTILF*I	W <mark>RNSSPLDLLHS</mark> IY <mark>SQTSP</mark>	
	Zea mays cml333;	Zm00026a003353_P001	P*GFDWE*G*WGGLKGLESLTILF*I	W <mark>RNSSPLDLLHS</mark> IY <mark>SQTSP</mark>	
	Zea mays cml52; Zr	m00019a003203_P002	(P*GFDWE*G*WGGLKGL <mark>ESLT</mark> ILF*I	W <mark>RNSSPLDLLHS</mark> IY <mark>SQTSP</mark>	
	Cea mays ki11; Zm00	0030a002495_P001	(P*GFDWE*G*WGGL <mark>KGLESLT</mark> ILF*I	WRNSSPLDLLHSI	-
	🔽 Zea mays m162w; Zm	100033a003891_P001¶V	P*GFDWE*G*WGGLKGLESLTILF*I	W <mark>RNSSPLDLLHS</mark> IY <mark>SQTSP</mark> Y	
٦II	[😳 Zea mays ky21; Zm00	0031a003750_P002¶V	P*GFDWE*G*WGGLKGLESLTILF*I	W <mark>RSSPLDLLHS</mark> IYSQTSPY	-
	🔽 Zea maysb73v4; Zm0	0001d030016_P001VA	P		
11	ൂ Zea mays ki3; Zm	n00029a003647_P001	P*GFDWE*G*WGGLKGLESLTILF*I	W <mark>RNSSPLDLLHS</mark> IY <mark>SQTSP</mark> Y	
	Zea mays nc350;	; Zm00036a002390_P001	P*GFDWE*G*WGGLKGLESLTILF*I	W RNSSPLDLLHS IY SQTSPY	
	Zea mays mo18w;	; Zm00034a003677_P002	P*GFDWE*G*WGGLKGLDSLTILF*I	W RNSSPLDLLHS IYSQTSP	-
	Zea mays cml228	3; Zm00022a003601_P001	P*GFDWE*G*WGGLKGLDSLTILF*I	W RNSSPLDLLHS IYSQTSP	-
	🛛 🕹 Zea mays tzi8; Zr	m00042a003768_P001	P*GFDWE*G*WGGLKGLDSLTILF*I	W RNSSPLDLLHS IYSQTSPYSQTSP	-
	Zea mays cml103;	Zm00021a003626_P002	P*GFDWE*G*WGGLKGLDSLTILF*I	V VNSSPLDLLHSIYSQTSP	
	🛛 🤄 Zea mays cml247; Zr	m00023a003755_P002	P*GFDWB*G*WGGLKGLDSLTILF*I	N NSSPLDLLHSIYSQTSP	
	└○ Zea mays m37w; Zm0	00032a003523_P002	P*GFDWB*G*WGGLKGLBSLTILF*I	VNSSPLDLLHSI	
	Zea mays b73_Ab10; Z	m00043a003391_P002	P*GFDW3*G*WGGLKGLSSLTILF*I	NNNSSPLDLLHSI	
L[°	Zea mays nc358; Zm0003	37a003386_P001	• • • • • • • • • • • • • • • • • • •	VASSPLULLHSI	
	Zea mays cmi277; Zm000	J24a003531_P001		VECTOTION VECTO	
	Zea mays p39; Zm00040a				-
142	Zea mays 11411; Zm0002c	412002625 0001			
	Zea mays 0303, 211000			V NSSDIDITEST VSOTSD	
4	Zea mays cmios, 2mood Zea mays ms71: 7m000	035a003445 P002	P*GFDWX*G*WGGLKGLPSLTLF*T	V NSSPLDLIRS I	
	2ea mays oh7h: 7m000	00000000000000000000000000000000000000	P*GFDWX*G*WGGLKGLPSLTLF*T	V INSSPLDLIRST YSOTSP	
4	Zea mays ob43: Zm0003	39a003608 P004TV	P*GFDWK*G*WGGLKGLPSLTILF*I	WRNSSPLDLLRSIYSOTSP	
∫¶ ro z	ea mays hp301: Zm00027a	a003494 P001	(P		
5 z	ea mays b97; Zm00018a00	03617_P002TV	(P		
C Sor	ghum bicolor; EER94459	_	(P		
Cryz	a sativa Japonica Group; (Os10t0162832-01			-





Gramene Zm00001e002261 – Missing Starting Met

NAM genomes

lay mode -Zea mays b73; Zm00001e002261 P001 Zea maysb73v4; Zm00001d029810_P001 Zea mays ky21; Zm00031a003521_P001 Zea mays cml103; Zm00021a003404 P001 Zea mays ki3; Zm00029a003411_P001 Zea mays cml228: Zm00022a003339 P0 Zea mays tzi8; Zm00042a003517 P001 Zea mays il14h; Zm00028a003203_P001 Zea mays nc358; Zm00037a003179_P00 Zea mays cml52; Zm00019a003001_P0 Zea mays cml333; Zm00026a003131_P Zea mays mo18w; Zm00034a003440_P00 Zea mays m162w; Zm00033a003632 P001 Zea mays b73_Ab10; Zm00043a003170_F Zea mays oh43; Zm00039a003379_P001 Zea mays ms71; Zm00035a003203 P001 Zea mays p39; Zm00040a003540_P001 Zea mays hp301; Zm00027a003269 P002 Zea mays nc350; Zm00036a002237 P00 Zea mays tx303; Zm00041a003382_P00 Zea mays ki11; Zm00030a002325_P001 Zea mays cml247; Zm00023a003514 P0 Zea mays cml277; Zm00024a003325_P001 Zea mays m37w; Zm00032a003303_P001 Zea mays oh7b; Zm00038a003434_P001 Contraction 2 Co Zea maysb73v3: GRMZM2G127067 P01 Zea mays cml69; Zm00020a003227 P001 Zea mays cml322; Zm00025a003228_P001 Sorahum bicolor: EER94557 Brachypodium distachyon; KQJ97662 Oryza sativa Japonica Group; Os10t0545200-01 Orvza sativa Japonica Group: Os03t0108500-01 Brachypodium distachyon: PNT76894 Brachypodium distachyon; KQK20571 Brachypodium distachyon; KQK17305 Sorghum bicolor; OQU77931 Brachypodium distachyon; PNT64046 Sorahum bicolor: OQU80062 Sorahum bicolor: 2 genes Zea maysb73v3: 16 genes

Color Scheme -

Γ	Multiple	Sequence Alignment: Amino acid MSA. Drag slider to repo	osition.		<u>г</u>			Cı	irate: click to flag	genes	
									okay Or flag		
		RERVHTP T ACREKLWGLCSS	-FSPPA-RSLSLVSRLLHSKRTGA	GERRE <mark>OR</mark> AAHSPPPDFLGVAD		SPL <mark>GS</mark> P-	MLPYAT	[flag G,_,_	\$	
	-MPTRAP	RVRL <mark>RROIHPTVRVTPROGRTHRGRVHTP</mark> B T BAGRHKLWGLG <mark>SS</mark>	F <mark>SPPA-<mark>R</mark>SLSLVS<mark>R</mark>LLHS<mark>KR</mark>TGA</mark>	GERRE <mark>OR</mark> AA <mark>HS</mark> PPPDFL <mark>G</mark> VAD	RSPGORGT AG	SPL <mark>GS</mark> P-1	MLPYAT				
L							MLPYAT				
L	MOTOAD	RVRERROTHPTVRVTPROGRTHRGRVHTP-T-AGREREWGLGSS	FODDA DOLOUNOPLINCE	CEPPEORATES DODI PLOVAD	RSPGORGTAG	SPLGSP	MLPIAT				
01							MLPYAT				
Ľ							MLPYAT				
L		TVRVTPROGRTHRGRVHTPBTBAGRHKLWGLGSS	FSPPA- <mark>RSLS</mark> LVS <mark>R</mark> LLHS <mark>KR</mark> TGA	GE <mark>RREOR</mark> AA <mark>HS</mark> PPPDFLGVAD	RSPGORGT AG	SPL <mark>GS</mark> P-	MLPYAT-				
		HRGRVHTP TBAGR KLWGLGSS	F <mark>SPP</mark> A- <mark>RSLS</mark> LV <mark>SR</mark> LLHS <mark>KR</mark> TGA	GE <mark>RREOR</mark> AA <mark>HS</mark> PPPDFLGVAD	RSPGORGT AG	SPL <mark>G</mark> SP-1	MLPYAT				
01							MLPYAT				
001							MLPYAT				
		HRGRVHTP T AGREELWGLGSS	FSPPA-RSLSLVSRLLHSKRTGA	GERREORAAHSPPPDFLGVAD		SPLGSP-I	MLPYAT				
200	1	QIHPTVRVTPROGRIHRGRVHTPETEAGRHKLWGLGSS	FSPPA- <mark>RSLSLVSR</mark> LLHS <mark>KR</mark> TGA	GERREORAAHSPPPDFLGVAD	RSPGORGT AG	SPL <mark>GS</mark> P-	MLPYAT				
L	-MPTRAP	RVRL <mark>RRQIHPTVRVTPROGRTHRGRVHTPETEAGRHK</mark> LWGL <mark>GSS</mark>	F <mark>SPP</mark> A- <mark>RSLSLVSR</mark> LLHS <mark>KR</mark> TGA	GERREORAAHSPPPDFLGVAD	RSPGORGT AG	SPL <mark>GS</mark> P-1	MLPYAT				
L							MLPYAT				
L	2MPTRAP	RVRLRROIHPTVRVTPROGRTHRGRVHTP=T=AGRHKLWGLGSS	FSPPA-RSLSLVSRLLHSKRTGA	GERREORAAHSPPPDFLGVAD		SPLGSP-	MLPYAT				
			FSPPA RSLSLVSRLLHSURTGA	GERRECRATHSPPPIFLEVAN	RSPGORGTAG	SPLGSP	MLPYAT				
							MLPYAT				
1	TRAP	RVRLRRQIHPTVRVTPROGRTHRGRVHTPSTBAGRHKLWGLGSS	FSPPA- <mark>RSLSLVSR</mark> LLHS <mark>KR</mark> TGA	GERREORATHSPPPDFLGVAD	RSPGORGT AC	SPLGSP-	MLPYAT				
0 1							MLPYAT				
L	APRPRAP	HVRVRGQIHPTVRVTPROGRSHRGRVHTP=T=AGRHKLWGLGSS	FSPPA- <mark>RSLSLVSR</mark> LLHSO <mark>R</mark> TGA	GERRE <mark>OR</mark> ATHSPPPDFLGVAD	RSPGORGT AC	SPL <mark>GS</mark> P-1	MLPYAT				
L	-MPTRAP	RVRL <mark>RROIHPTVRVTPROGRTHRGRVHTPST</mark> SAGRHKLWGLGSS	FSPPA- <mark>RSLSLVSR</mark> LLHS <mark>KR</mark> TGA	GERREORATHSPPPDFLGVAD	RSPGORGT AG	SPLG <mark>S</mark> P-1	MLPYAT				
L	-MPTRAP	KVRDROIHPTVRVTPROGRTHRGRVHTP-T-AGREKLWGLGSS	FSPPA- <mark>RSLSLVSR</mark> LLHS <mark>KRTG</mark> A	GORROORAAHSPPPDFLGVAR	RSPGORGTAG	SPLGSP-1	MLPYAT-				
L							MLPYAT				
L	2MPTRAP	RVRLRRQIHPTVRVTPRQGRTHRGRVHTP=T=AGRHKLWGLGSS		CERRE <mark>OR</mark> AAHSPPPDFLGVAD		SPL <mark>GSP-</mark>	MLPYAT				
							MLPYAT				
L							MLPYAT-				
L							MLPWAT				
							MLPYAT				
L							MLPYAT				
1							MLPYAT				
L							MLPYAT				
1							MIRYAT				
1		MLLIKFGORKEE OO BESKHYKYGLHATTPLTPSRP	VELIYIH-TL <mark>CK</mark> SEELSPONYSOOK	KEKIP <mark>RTWKR</mark> SERIALDNIPINPPA/	AAN <mark>RRSK</mark> HRAY	<mark>(</mark>	MLPYAT				
1							MIPYGT				
L							MMPYGT				
1							rull ^y I GI				





CUREs derived from the 2020 VMAJ

- 2020 Spring CUREs:
 - Middle Tennessee State University Rebecca Seipelt
 - University of Central Missouri Selene Nikaido
 - Cornell University Kevin Ahern
 - University of Toledo Ankita Abnave & John Gray
- 2020 Fall CUREs under development at Whitman College and College of Southern Nevada











Interested in Joining?

- PUI faculty network
- Biweekly meetings, slack channel, office hours
- Educational resources: lesson plans, videotutorials, slides
- Resources to curate Maize and Sorghum genes
- Contact us: <u>telloruiz@cshl.edu</u>







Cristina Fernández-Marco Kapeel Chougule Andrew Olson David Micklos Doreen Ware

CUREs Faculty & Students

Rebecca Siepelt (MTSU) Christos Noutsos (SUNY OW) Selene Mikaido (UCM) Kevin Ahern (Cornell U) Ankita Abnave/John Gray (U Toledo)







The project is supported by NSF IOS-1127112, MCB-1744001, IOS-1445025, and USDA ARS 8062-21000-041

Outreach Funds







