KnowPulse Documentation

LA Sanderson, CT Caron, R Tan, Y Shen, KE Bett (University of Sa

May 26, 2021

Use Cases:

1	Get a list of SNPs/variants between 2+ varieties within a region of interest.	1
2	Developers	7

CHAPTER 1

Get a list of SNPs/variants between 2+ varieties within a region of interest.

Example:

- Region of interest: LcChr4:1875000-2075000
- Germplasm of Interest: ILL 8006, CDC Milestone

1.1 Step #1: Go to the Lentil Genotype Matrix Tool

- Go to https://knowpulse.usask.ca
- Click on "Genotypes" at the top of the page. This will scroll you down to the genotypes section.
- Click on "Lentil Genotypes" to access the Lentil Genotype Matrix Tool.



1.2 Step #2: Find genotypic data for your reference germplasm

- Enter the name of each germplasm you are interested in by typing it in the textfield labelled germplasm. Then check the correct species is selected to the right of the textbox. To add more then one germplasm click the green + button.
- For this example, Enter "ILL 8006" in the top germplasm textfield. You will notice that it autocom-

pleted as you type.

- Then click the green + button and enter "CDC Milestone" in the second textfield which appears.
- Each time you click the green + button or search, the genotypic data for the listed germplasm will be shown.

STAR AND DE STAR SERVICES

13 (15/10/17/17	000					Bioinformatic Tools		
le concept	C 5					▼ BI AST		
- Choose germ	plasm you are interested i	۱.				▼ Nucleotide Query		
Simply enter the green plus (+) bu entered.	name of one germplasm (e.g tton. You can enter any num	. "Eston AGL", "CDC I ber of germplasm you	Robin AGL", or "ILl 1 are interested in c	. 8007 AGL") of it and each will be a	nterest below and then click t dded to the matrix as they ar	he • BLASTn • BLASTx e • Protein Query		
Germplasm						• BLASTp • tBLASTn		
ILL 8006			O Len	s culinaris 🔹	×	▼ JBrowse		
CDC Milestone			O Len	s culinaris 🔹	×	• Lentil (UofS, v1.2)		
Germplasm/Stor	rk Name		0 I.an	s culinaris	<u> </u>	• Chickpea (Kabuli, v1.0)		
Specify the stock (an	ad species of the stock) you want to	display the genotypes of			•	Cnickpea (Desi, v2.0) Colletotrichum lentis		
						Common Bean (Phytozome, v2.1)		
Restrict to the	e region of the genome. (o recommend you filter to a g	ptional) ven region of the gend	ome to make the ge	notypes et more n	nanagable. For example, to s	v4.0) > User-driven Charts > CVTTjs Whole-Genome Visualizatio • Position Convert: Lco.8 to Lc1.2 • VCF Bulk Export		
of Lentil Chromo	some 4 you would enter From	n "LcChr4" to "LcChr4	f" leaving the start	and end position	blank.	VCF Position Search		
From Sequence	Name O Start Por	sition	to Sequence Name O End Position					
The range of the gen	ome you would like to display vari	ants from. If you enter jus	t the start or just the e	nd position then all v	ariants before or after that	Genomic Data		
location, respectively	location, respectively, will be displayed.							
						Markers		
Additional Filt	tor critoria (ontional)					 Lentil Genotypes 		
Additional Th	er chteria. (optional)							
Search						Phenotypic Data		
						✓ Phenotypes		
		Total Results?:	95,362			• Trait Distribution Plot		
		Unique Variants	: 39,436		Sort by Location, Varia	• Lentil Traits		
ad: Requires log in		emque variants						
ad: <i>Requires log in</i> nt Name	Backbone	Start 1	and ILL 800	6	CDC Milestone			
ad: <i>Requires log in</i> nt Name 7539578	Backbone	Start 1 5562 5	3nd ILL 800	6	CDC Milestone	Germplasm		
oad: <i>Requires log in</i> nt Name 7539578 7539537	Backbone LcChri LcChri	Start J 5562 5 5603 5	End ILL 800 562 603	б С	CDC Milestone A C	Germplasm		
nad: <i>Requires log in</i> nt Name 7539578 7539537 753960	Backbone LeChri LeChri LeChri	Start J 55562 5 5603 5 6081 6	3nd ILL 800 5562	6 C G	CDC Milestone A C C C C	Germplasm • Germplasm		

1.3 Step #3: Restrict the Sequence Variants to polymorphic between your germplasm

- Underneath germplasm, there is a filter to restrict to polymorphic variants. This filter compares two germplasm and only shows variants with different genotypic calls.
- For our example, we would select ILL 8006 in the first drop down and CDC Milestone in the second drop-down to see only sequence variants with differing genotypes (i.e polymorphic variants) between these two germplasm.
- Click Search to see the results.



Home

Lens Genotypes	S					Bioinformatic Tools
1 - Choose germplas	sm you are interested ir	ı.				▼ BLAST▼ Nucleotide Query
Simply enter the nam green plus (+) button entered.	ne of one germplasm (e.g. n. You can enter any num	 ● BLASTN ● BLASTX ▼ Protein Query ● BLASTp 				
Germplasm						• tBLASTn
ILL 8006				O Lens culinaris	<u> </u>	▼ JBrowse
CDC Milestone				O Lens culinaris	• ×	• Lentil (UofS, v1.2)
Germplasm/Stock N	Name			O Lens culinaris	• 🔂	 Chickpea (Kabuli, Vilo) Chickpea (Desi, v2.0)
Specify the stock (and sp	pecies of the stock) you want to	display the genotypes o	f.			Colletotrichum lentis
Restrict the variants to the restrict function of L ntil Chromosom Getome Range Front Sequence Na	hose that have different allele of sign of the genome. (or commend you filter to a gin the 4 you would enter From ame Start Pos	alls for the selected gern ptional) ven region of the gern "LeChr4" to "LeCh ition	nplasm. nome to mi r4" leaving to Seq	ake the genotype set mor the start and end position uence Name	e me tagable. For example, to see on b mk. En Position	all • VCF Position Search
The ange of the genome loca on, respectively, wi	e you would like to display varia ill be displayed. criteria. (optional)	ants from. If you enter ji	ust the start	or just the end position then a	Ill var ants before or after that	Genomic Data Sequence Variants • Markers • Lentil Genotypes
Search						Phenotypic Data
		Total Results?	24,704			Phenotypes Trait Distribution Plot Lentil Traits
Download: Requires log in	Berliner	Unique Variant	s•: 8,486	TTT Door	Sort by Location, Variant	Name
Variant Name	LaChar	Start	End	ILL 8006	CDC Milestone	
LeCo1753p60	LeChri	6081	6081	G		Germplasm
LcC01392p800	LeChri	363602	363602	C		- Complasm
LeC01392p854	Leunri	363757	303757	C		Germplasm Registered Varieties
LCC01392p920	LeChri	363823	303823	C	T	✓ Genetic Maps
LcC01392p932	LcChri	363835	363835	Т	C	MapViewer Tool
LcC01392p935	LeChri	363838	363838	G	T	
LcC01392p1135	LcChr1	364038	364038	С	Т	

1.4 Step #4: Restrict to you trait-implicated Region of the Genome.

- The second section of the filter criteria available for the genotype matrix allows you to enter the region of the genome you are interested in. Once you click search, the genotype matrix will only show sequence variants found in this region.
- In our example, the region of interest is LcChr4:1875000-2075000. To enter this we place LcChr4 for the Sequence Name, 1875000 for the start position and 2075000 for the end position.



Home

Choose purpues we can there ease the same of one germplane (e.g. "Eston AGL", "CDC Robin AGL", or "LL BOOT AGL", of interest below and then click the prevention of the germplane part of the same of one germplane part of the same of one germplane part of the same of	ens Genotype	S				Bioinformatic Tools				
Choose germplasm you are interested in. Subject sets the muse of one germplasm (g.g. "Eakon AGL", "CPC Relation AGL", or "LL Boos AGL" of justment below and then click the germ place () batton. You can enter any number of germplasm you are interested in and each will be added to the matrix as they are entered. Exercise the set of the set	51					▼ BLAST				
Single years the name of one germplane (a, 2. "CDX Cables AGL", "CTX Lasor AGL") of intervent below and then cick the green pire (>) (>) (>) (>) (>) (>) (>) (>) (>) (>)	1 - Choose germpla	asm you are interested ir	۱.			▼ Nucleotide Query				
Carmpian Company Construction Constructi	Simply enter the na green plus (+) butto entered.	rme of one germplasm (e.g. on. You can enter any num	 BLASTn BLASTx ▼ Protein Query 							
ILL 8000 Instantion Instantion Instantion ILL 8000 Instantion Instantion Instantion Comparison Instantion Instantion Instantion Primerical Content of the genome to make the genotypes set more managable. For example, to see all of the set of the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to inclust the endotion blank. Comme Range Instant to instant the region of the genome to make the genotype set more managable. For example, to see all of the region of the genome to inclust the endotion blank. Comme Range Instant to instant the region of the genome to inclust the endotion blank. Additional Filter criteria. (optional) Instant to in	Germplasm		BLASTp tBLASTp							
CDC Meeting Less calutaria * Current parks //Stock Name Exers calutaria * Specify that skal cala decide that obje variants datage the genotypes of. Exers calutaria * Parkers Mill 1 and Conditions * * Parkers Mill 1 and Conditions * * Parkers Mill 1 and Conditions * * * Parkers Mill 1 and Conditions *	ILL 8006		(Lens culinaris	- ×					
I we consider the stack and species of the anckly you want to display the genotypes at. Promyorphic Variants Retrict to the region of the genome. (optional) I argoinable to a given orgin of the genome to make the genotype set more managable. For example, to see all of Letter from "LCCrift" to "LCCrift" leaving the start and end position blank. Common Resen (Phytroame, Variants Resence) Prom	CDC Milestone	CDC Milestone		Lens culinaris		 Lentil (UofS, v1.2) 				
Chickge (Dely year) Chickge (Dely year) Chickge (Dely year) Syndry the stuck (and preder of the stuck) yea want to diaply the genotypes of. Polymorphic Variants Common Mean (Mytorome, v.a.) Retrict to the region of the genome. (optional) Retrict to the region of the genome. (optional) Polymorphic Variants Common Mean (Mytorome, v.a.) Retrict to the region of the genome. (optional) Retrict to the region of the genome. (optional) Polymorphic Variants Common Mean (Mytorome, V.a.) Retrict to the region of the genome. (optional) Retrict to the region of the genome. (optional) Polymorphic Variants Common Machinge (Mitchica) Retrict to the region of the genome. (optional) Is z55000 Carries VYTP Which Common Wualling Starch Stark Machine Locht 4 Zor Zoro Sympon Sympon Starch	CDC Milestone	CDC Milestone				 Chickpea (Kabuli, v1.0) Chickpea (Desi, v2.0) Colletotrichum lentis 				
Specify the stack (and species of the stack) you want during the genotypes of. Polymorphic Variants Retween 118 8000 and COC Missions Returns to those that have different allole calls for the selected germplane. * Retrict the variants to those that have different allole calls for the selected germplane. * Retrict the variants to those that have different allole calls for the selected germplane. * Retrict the variants to those that have different allole calls for the selected germplane. * Retrict the variants to those that have different allole calls for the selected germplane. * Retrict to the region of the genome, (optional) if applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of Left A 1 1875000 fo Cchr4 2075000 The range of the genome you would enter Prom "LcArdr4" to "LcArdr4" leaving the start and end position blank. Cenome Range From Cchr4 1875 * Addictional Filter criteria. (optional) Search * full Results': 15 Found: Requires log in the search or just the start or just the end position then all variants before or after that and have respectively, will be diaplayed: * Sequence Variants * Lattil Genotypes Lacha 194466 394466 TT 0 Search * Denotypes Lacha 194466 394466 TT 0 Search * Dial Results': 15 Search * Denotypes Lacha 194466 394466 TT 0 Correntee Calls * Denotypes Lacha 194466 394466 TT 0 Correntee Calls * Denotypes Lacha 194466 394466 TT 0 Correntee Calls * Denotypes Lacha 194466 394466 TT 0 Search * Denotypes Lacha 194466 394466 TT 0 Search * Denotypes Lacha 194466 394466 TT 0 Search * Denotypes Lacha 194466 394466 TT 0 Correntee Calls * Denotypes Lacha 194466 394466 TT 0 Search * Denotypes * Denotypes Lacha 194466 394466 TT 0 Searc	Germplasm/Stock	Name	(Lens culinaris	· 🖯					
Polymorphic Variants • Mediagn (MGAG, ve.0) Between Lubble and Co Messon • • Mediagn (MGAG, ve.0) • Mediagn (MGAG, ve.0) • Mediagn (MGAG, ve.0) • Mediagn (MGA	Specify the stock (and s	Specify the stock (and species of the stock) you want to display the genotypes of.								
Bettern LL 200 and CCC Milestone] Restrict to the tegion of the genome. (Optional) If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of Jentil Chromosome you would enter Prom "LChr4" to "LChr4" t	Polymorphic Varia	Polymorphic Variants								
Activity the variants to those that have different tailed calls for the sected germplaam. • Stybean (Phytocanon, Williams & a. v.o.) • Restrict to the region of the genome. (optional) • User-driven Charts of Lenti Chromosome 4 you would enter From "LcChr4" to "LcChr4" leaving the start and end position blank. • Ort Will Mike Stopt Cenome Range • Isoro to LcChr4 2075000 Prom LcChr4 1875000 to LcChr4 2075000 The range of the genome you would like to diplay variants from. If you enter just the start or just the end position then all variants before or after that location, respectively, will be diplayed. • Stephene (Phytocanon) Search • Jall Results"; 15 Sort by Locations, Variant Name and Xance Rakbone ILL 8006 CDC Milestone yoogsp3 LcChr4 194466 T G yoogsp3 LcChr4 194466 T <	Between ILL 8006	Between ILL 8006 • and CDC Milestone •								
Kestfort to the region of the genome. (optional) If explicable, we recommend you, filter to a given region of the genome to make the genotype set more managable. For example, to see all of Lettic Hornessome 4 you would enter Prom "LcChr4" to "LcChr4" leaving the start and end position blank. Comme Range Prom LcCh4 1875000 to LcChr4 10 LCCh4 1875000 to LcChr4 10 LCCh4 187500 to LcChr4 18750 to LcChr4 1875 to LcChr4 18750 to LcChr4 1875 to LcChr4 18750 to LcChr4 1875 to LcChr4 1875 to LcChr4 18750 to LcChr4 1875 to LcChr4 1875 to LcChr4 18750 to LcChr4 1875 to LcChr4 187	Restrict the variants to	those that have different allele of	calls for the selected germplasm.			 Soybean (Phytozome, Williams 82 v4.0) 				
• Restrict to the region of the genome. (optional) • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of LeCht4 • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of LeCht4 • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of LeCht4 • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of LeCht4 • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of LeCht4 • CVTTJ3 Whole-Genome Visualization If applicable, we recommend you filter to a given region of the genome you would like to display variants from. If you enter just the start or just the end position then all variants before or after that • Sequence Variants If applicable, we recommend you filter to display variants from. If you enter just the start or just the end position then all variants before or after that • Sequence Variants If applicable, in the you doel in the variants before or after that • Sequence Variants • Init Distribution Plot If approving LeCht4 104565						User-driven Charts				
J Graphicable, we recommend you filter to a given region of the genome to make the genotype set more managable. For example, to see all of Lentil Chromosome 4 you would enter From "LcChr4" leaving the start and end position blank. Position Convert: Loo 8 to Lc1. VCF Pulk Export VCF Pulk Export	Restrict to the re	egion of the genome. (or	otional)			► CViTjs Whole-Genome Visualization				
• (C) = prival (If applicable, we not	annund von Elton ta a ai	um magion of the generate make	the construct of	none managable. For example, to see all	 Position Convert: Lco.8 to Lc1.2 				
Genome Range 10 LChr4 2075000 The range of the genome you would like to display variants from. If you enter just the start or just the end position then all variants before or after that Sequence Variants • Additional Filter criteria. (optional) Sarch Sequence Variants • Sequence Variants • bring Kendbore Full Results*: 15 Sort by Location, Variant Name • Markers • Name Results*: 15 Sort by Location, Variant Name Nonofficial Lichr4 118.806 CDC Milestone % poops64 Lichr4 19466 T C % poops77 Lichr4 19466 T C % poops64 Lichr4 19466 T C % poops77 Lichr4 19466 T C % poops77 Lichr4 19470 A G % poops71 Lichr4 19466 T C % poops72 Lichr4 19466 T C % poops72 Lichr4 19466 T C % poops72 Lichr4 19466 A G % poops72 Lichr4	of Lentil Chromosor	me 4 you would enter From	n "LcChr4" to "LcChr4" leaving th	e start and end po	sition blank.	 VCF Bulk Export VCF Position Search 				
Chrine Kange IB75000 fo LChr4 2075000 The range of the genome you would like to display variants from. If you enter just the start or just the end position then all variants before or after that location, respectively, will be displayed. Sequence Variants • Markers • Additional Filter criteria. (optional) Image: Comparison of the provide Variants from the pr	Company Borner									
International prime production would like to display variants from. If you enter just the end position then all variants before or after that location, respectively, will be displayed. Genomic Data • Additional Filter criteria. (optional) • Additional Filter criteria. (optional) International prime variants from. If you enter just the end position then all variants before or after that location, respectively, will be displayed variants from. For any of the prime variants from. If you enter just the end position then all variants before or after that location, respectively, will be displayed variants from. Genomic Data • Additional Filter criteria. (optional) State End Ist Book Optional • Phenotypic Data • Phenotypic Data search • Dime Variants*: 10 Sort by Location, Variant Name To Optional • Thit Distribution Plot • Thit Distribution Plot • Thit Distribution Plot • Thit Distribution Plot yoops64 LaChr4 194464 194465 T Germplasm • Registered Varieties yoops70 LaChr4 194464 194465 T Genomic Data • Genomic Data yoops20 LaChr4 194465 19466 T Genomic Data • Genomic Data yoops20 LaChr4 19469 19469 T Genomic Data	From LaChat	Genome Range								
Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position them at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position in the at variable before of after that Intrage of the genome you would use to usaging variable from 1 you enter just the start of just the end position 1 you enter just the start of just the end position 1 you enter just the start of just the end position 1 you enter just the end	The second active and the second	LcC.hr4 O 1875000 ⁴⁰ LcChr4 O 2075000								
• Additional Filter criteria. (optional) • Sequence Variants Search • Makers load: Requires log in • Unique Variants ² : 15 national Filter criteria • Unique Variants ² : 15 roopop64 LeChr4 194665 114.8006 CDC Milestone ryoopp53 LeChr4 194652 194652 T C ryoopp577 LeChr4 194652 194652 T C ryoopp570 LeChr4 194652 194652 T C ryoopp570 LeChr4 194652 194652 T C ryoopp571 LeChr4 194652 194652 T C ryoopp570 LeChr4 194652 194652 G C ryoopp571 LeChr4 194652 194652 G C ryoopp570 LeChr4 195010 AA G C ryoopp571 LeChr4 195010 AA G AA G ryoopp57 LeChr4 195010 AA G AA G AA G ryopp50<	location, respectively, w	The range of the genome you would like to display variants from. If you enter just the start or just the end position then all variants before or after that location, respectively, will be displayed.								
• Additional Filter criteria. (optional) • Markets • Latridi Genotypes • Search • Latridi Genotypes • Latridi Genotypes • Search • Latridi Genotypes • Theorypic Data • Search • Dime Variants*: 15 • Sort by Location, Variant Name • Markets • Dime Variants*: 16 • Oct Milestone • Name • Backbone • Li Booé • Doct Milestone • Poopo50 LcChr4 191465 11 • Coc • Soorb by Location, Variant Name • Chr4 191465 1 • Coc • Soorb optional LcChr4 191465 1 • Coc • Cermplasm • Registered Varieties • Soorpopa LcChr4 191470 AA 6 • Markets • Markets 0120450 LcChr4 191470 AA 6 • Cermplasm • Registered Varieties • Germplasm 0120450 LcChr4 194370 AA 6 • Markets • Markets 0120450 LcChr4 194370 AA 6 • Markets • Markets 0120450 LcChr4 204384 204390 T<		 Y M supervised and the supervised and								
Additional Filter criteria. (optional) Phenotype Search Phenotype Search Phenotype Additional Filter criteria. (optional) Notal Results*: 1: 5: 5: 5: 5: 5: 5: 5: 5: 5: 5: 5: 5: 5:										
Seach Definition Sort by Location, Variant Name Phenotype at Xance Backbone End IL 3006 OC Milestone 9 100 1000 9 100000 9 100000 9 100000 9 100000 9 100000 9 100000 9 100000 9 100000 9 100000 9 1000000 9 1000000 9 1000000 9 1000000 9 1000000 9 1000000 9 1000000 9 1000000 9 1000000 9 10000000 9 1000000000000 9 1000000000000000000000000000000000000	Additional Filter	r criteria. (optional)								
Search Plenotypic Data Sortby Location, Variant Nome Ant Name Sortby Location, Variant Nome Ant Name Rekobone Sortby Location, Variant Nome Torus Variant Nome Ant Name Rekobone Sortby Location, Variant Nome Torus Variant Nome Ant Name Rekobone Sortby Location, Variant Nome Torus Variant Nome Sortby Location, Variant Nome Sortby Location, Variant Nome Sortby Location, Variant Nome Planotypic Data Ant Nome Sortby Location, Variant Nome Planotypic Data Ant Nome Colspan="2">Planotypic Data Planotypic Data Planotypic Data Sortby Location, Variant Nome Sortby Location, Variant Nome Planotypic Data Planotypic Data Planotypic Data Oplanotypic Data										
India: Requires log in India: Resultors log in India: Resultors log in Sort by Location, Variant Name and Name Sackbone State India I	Search					Phenotypic Data				
Talk Best basesNade: Requires log inSort by Location, Variant Nameand NameBackboneStatEndILL 8006CDC Milestone7900p664LcChr4191456191456TG7900p577LcChr4191466191466TC7900p577LcChr4191470191470TC7900p577LcChr4191470191470AG900p577LcChr4191470191470AG900p577LcChr4196110196110AAG9017031LcChr41985063195063TC0132p152LcChr4204371204371TA0132p256LcChr4204386204386GA0132p266LcChr4204386204386GA0132p206LcChr4204380204380CTLcChr4204380204380CTC0132p205LcChr4204380204380CT0132p205LcChr4204380204380CT0132p205LcChr4204380204380CT015010frmatics Resource30100TC015010frmatics Resource3010010000100000015010frmatics Resource30100010000001000000015010frmatics Resource301000000000001000000000000000000000000000000000000						- Dhanatanaa				
Ant Service V intre Variant Value Sort by Location, Variant Name • Lentil Traits ant Name Backbone Stat End ILI Soo6 CDC Milestone 7900p664 LcChr4 191456 191456 T G 7900p573 LcChr4 191456 191456 T C 7900p573 LcChr4 191456 191457 T C 9900p577 LcChr4 191456 191458 T C 9900p573 LcChr4 191459 191459 AA G 9900p57 LcChr4 191459 19459 T C 9900p57 LcChr4 19459 19459 T C 9900p57 LcChr4 19459 19459 T C 917331 LcChr4 204371 204371 T A 9132925 LcChr4 204386 204386 T C Our Research 9132925 LcChr4 204386 2043802 C T			tal Regults?			Trait Distribution Plot				
Backbone Start End IL 8006 CDC Milestone 7900p664 LcChr4 19466 19466 T G 7900p57 LcChr4 19466 19466 T G 7900p57 LcChr4 19469 19469 T G 7900p57 LcChr4 19469 19469 T C 990p570 LcChr4 19469 19469 T C 990p570 LcChr4 19469 19469 T C 990p570 LcChr4 19469 19469 A G 990p570 LcChr4 19469 19469 A G 917931 LcChr4 198503 195063 T C 0132p450 LcChr4 204384 204384 T C Our Research 0132p251 LcChr4 204384 204386 G A Home 0132p206 LcChr4 204300 2045002 C T Our Resea	load: Requires log in		Loique Variants?: 10		Sort by Location Variant Name	Lentil Traits				
yoop664 LcChr4 194565 194565 T G yoop653 LcChr4 194666 194666 T C yoop677 LcChr4 194656 194656 T C yoop677 LcChr4 194652 194652 T C - Germplasm - Registered Varieties yoop570 LcChr4 194690 194790 19470 A G - Germplasm - Registered Varieties - Germplasm - Registered Varieties - Registered Varieties - Genetic Maps - MapViewer Tool - MapViewer Tool 1917031 LcChr4 1985063 1985063 T C - MapViewer Tool - MapViewer Tool 0132p450 LcChr4 204371 204371 T A -	ant Name	Backbone	Start End IL	L 8006	CDC Milestone					
Interfact Interfactory Interfactory <td>79000664</td> <td>LcChr4</td> <td>1914565 1914565</td> <td>т</td> <td>G</td> <td></td>	79000664	LcChr4	1914565 1914565	т	G					
Markadow	79000583	LcChr4	1914646 1914646	т	.C	Germplasm				
yoopy/// kc/ki yoopy/// yoopy/// kc/ki yoopy/// yoopy//// </td <td>9000577</td> <td>LcChr4</td> <td>1914652 1914652</td> <td>т</td> <td>C</td> <td>▼ Germplasm</td>	9000577	LcChr4	1914652 1914652	т	C	▼ Germplasm				
Spapping LcKrd Spin Spin AA G Spapping LcKrd 196100 196100 AA G Spapping LcKrd 1985063 T C Otagaptson LcKrd 204371 204371 T A Otagaptson LcKrd 2043849 2043849 T C Our Research Otagaptson LcKrd 2043866 2043866 G A Our Research Otagaptson LcKrd 2043866 2043866 G A Our Research Otagaptson LcKrd 2043866 2043866 G A Our Research Bioinformatics Resource Bioinformatics Re	79000520	LeChr4	1914709 1914709	Α	G	Registered Varieties				
Arror V Arror V Arror V Arror V Arror V • MapViewer Tool 39179331 LcChr4 1985063 1985063 T C 0132p4450 LcChr4 2043711 T A 0132p295 LcChr4 2043849 T C 0132p206 LcChr4 2043066 2043866 G A • Home • Bioinformatics Resource • Student Information	5929069	LcChr4	1961110 1961110	AA	G	▼ Genetic Maps				
Jackard Jogdorg Jogdorg <t< td=""><td>20170221</td><td>LeChr4</td><td>1085062 1085062</td><td>т</td><td>C</td><td> MapViewer Tool </td></t<>	20170221	LeChr4	1085062 1085062	т	C	 MapViewer Tool 				
Arrow Arrow Arrow 013229312 LcChr4 2043849 T C 013229295 LcChr4 2043866 C Arrow 013229206 LcChr4 2045002 C T «first 100 « previous LcChr4: 1914564-2045002 next) last 100 »	77*/F031	LeChr4	2042711 2042711		A					
USAPPORE Declina 2043049 2043049 Previous C Our Research 01322295 LcChr4 2043866 2043866 G A • Home • Bioinformatics Resource 01322206 LcChr4 2045002 2045002 next, last 100 Student Information	01329450	LeCh-t	2043/11 2043/11		A					
U12g2p255 LcCur4 2043866 2043866 C A 0132p206 LcChr4 2045002 2045002 C T • Home « first 100 « previous LcChr4: 1914564-2045002 next , last 100 »	0132p312	LcUnr4	2043849 2043849		C	Our Research				
V0132p200 LcChr4 2045002 2045002 C Image: Constraint of the second	0132p295	LeChr4	2043866 2043866	G	A	• Home				
« first 100 « previous LcChr4: 1914564-2045002 next , last 100 » • Student Information	0132p206	LcChr4	2045002 2045002	С	Т	Bioinformatics Resource				
		« first 100	« previous LcChr4: 1914564-2045	002 next last	100 »	Student Information				

1.5 Step #5: (Optionally): Restrict to specific variants.

- Say you are interested in a specific set of variants and would like to see that subset.
- You can enter the specific variant names by expanding the Additional Filter criteria section then clicking Search.



Home

Lens Genotypes					Bioinformatic Tools
1					- BLAST
1 Choose germplasm yo Simply enter the name of green plus (+) button. You entered. Germplasm ILL 8006 CDC Milestone Germplasm/Stock Name Specify the stock (and species of Polymorphic Variants Between ILL 8006 Restrict to the region If applicable, we recomm	pu are interested in. one germplasm (e.g., "Est i can enter any number of the stock) you want to disp of the stock) you want to disp and CDC Milestone * hat have different allele calls of the genome. (optic end you filter to a given	iton AGL", "CDC Robin AGL", or "IL of germplasm you are interested in a left of the genotypes of. for the selected germplasm.	L 8007 AGL") of inter and each will be adde is culinaris • 3 is culinaris • 3 is culinaris • 4 enotype set more man	est below and then click the d to the matrix as they are	 Nucleotide Query BLASTn BLASTx Protein Query BLASTp BLASTp tBLASTn JBrowse Lentil (UofS, v1.2) Chickpea (Kabuli, v1.0) Chickpea (Kabuli, v1.0) Chickpea (Cosi, v2.0) Colletotrichum lentis Common Bean (Phytozome, v2.1) Medicago (IMGAG, v4.0) Pea (Caméor; v1a) Soybean (Phytozome, Williams 82 v4.0) User-driven Charts CVTTjs Whole-Genome Visualization Position Convert: Lco.8 to Lc1.2 VCF Bulk Export
Genome Range	ou would enter From "L	cCnr4 to LCCnr4 leaving the start	ana ena position bla	пк.	VCF Position Search
From LcChr4	0 1875000	to LcChr4	0 207	5000	
The range of the genome you we location, respectively, will be d	ould like to display variants isplayed.	from. If you enter just the start or just the e	and position then all varia	nts before or after that	Genomic Data Sequence Variants Markers Lentil Genotypes
3 Additional Filter criter	ia. (optional)				
We recommend you fill ou in.	it as many of the followi	ing optional filters as possible to nar	row the genotype set	to those you are most interested	
Variant Name(s) LcC00132p450 LcC00132p312 LcC00132p295 LcC00132p206					 Phenotypes Trait Distribution Plot Lentil Traits
				li.	
A list of variant names you wis	h to see genotypes for with o	ne variant per line.			Germplasm
Proje t Name The name of the project you wa	ant to restrict genotypes to.			0	 Germplasm Registered Varieties Genetic Maps MapViewer Tool
- Choose One to Filter - The types of variants you woul	d like to see genotypes for (e.	.g. indels only).			
Marl r Type - Ch vse One to Filter - The ty s of markers you woul	d like to see genotypes for (e	.g. exome capture).			Our Research Home Bioinformatics Resource Student Information Job Postings
Searc		Total Results?: 7			 Field Lab Projects Publications Workshops Research
ownload: Requires log in		Unique Variants?: 4		Sort by Location, Variant Name	Breeding & Genetics
Variant Name	Backbone	Start End ILL 800	6 CI	OC Milestone	Pathology Physiology
LcC00132p206	LcChr4	2045002 2045002	С	т	Thysiology Tissue Culture
LcC00132p295	LcChr4	2043866 2043866	G	А	 Crops
LcC00132p312	LcChr4	2043849 2043849	т	С	Chickpea
LcCoo132p450	LcChr4 « first 100 « pr	2043711 2043711 evious LcChr4: 2045001-2043711	T next > last 100 »	A	• Dry Bean • Faba Bean • Lentil

CHAPTER 2

Developers

Warning: This guide is meant for the KnowPulse Development Team.

This guide is meant for the KnowPulse team to share information and document procedures. That said, it may be helpful for anyone developing a Tripal site -we are welcoming to all developers and welcome suggestions and fixes to our documentation.

2.1 Developer Workflow

This document describes the workflow that the KnowPulse team follows for development, which depends on the use of GitHub features and the [ZenHub Agile Project Management Extension for GitHub](https://www.zenhub.com/).

All Tasks, Bugs, Features, etc. should be GitHub Issues

- Create an issue on github: Be Descriptive!
- Label the issue (bug? enhancement?)
- Using Boards, indicate if this is a long term goal (icebox) or if you intend to do it right away (Up Next)
- If you know who will be doing work on this issue then assign it to them.
- Keep issues to reasonable amounts of work. A good rule of thumb is that if you think the issue would take more than a week of development time to complete, then the scope of work is not reasonable for a single issue ;-).
- Also keep in mind, the smaller the issue the easier it is to review and the faster it will be released.

2.1.1 Workflow

- 1. Choose an issue to work on. Bugs should take priority. Look first at issues assigned to you.
- 2. Once you've chosen a task, assign it to yourself and move it to "In Progress" using Boards.

- 3. Create a local branch and push it to github _before_ making a commit. Use the format *[issue number]-[short description]* when you name the branch.
 - Create local branch: *git checkout -b [issue number]-[short description]*
 - Push it to github: git push origin [issue number]-[short description]
 - Any changes you will make in completing this issue should now be done within this branch.
- 4. Work on the issue.
 - Commit often to your local branch using descriptive commit messages.
 - Push to github at least twice a week!
 - Also, try to to keep a record of design decisions and approach to solving the issue by commenting on the issue itself. The more context that you provide, the less you will have to explain to your co-workers during review ;-)
- 5. Once work is complete, signal a Review Create a pull request between your branch and the master branch
 - 1. On the GitHub repository page, click on the "Pull Requests" tab and click the "New Pull Request" button.
 - 2. Leave *base:master*; change *compare:master* to point to your branch created in step 3.
 - 3. The page immediatly shows the changes between your branch and master. Create the pull request by clicking the "Create Pull Request" button.
 - 4. Link the pull request to the issue
 - 5. Describe the original issue, how you fixed it and how to test your code: what pages to look at, etc.
 - Request all team members review the pull request
 - In Boards, move the issue to the "Review" pane
- 6. Review
 - Give your team members a day or two to look over your code
 - Answer any questions they have
 - Wait to implement suggestions until the review process is completed by everyone
 - Reviewers should indicate whether they approve your changes or not.
- 7. All reviewers approved your changes?
- If no, then go back to step 4.
 - Make sure to address all concerns.
 - If you are uncertain their approach is the best one, then **defend yours**. Ask for clarification on any suggestions if necessary.
 - Keep in mind, you do not need to create a new pull request after addressing concerns. The pull request
 will automatically update as you make changes to the branch created in step 3. That is why you are doing
 this work in a issue-specific branch.
- If yes, then move onto step 8
- 8. Close the issue
 - Done by lead Reviewer (Once all issues/suggestions are addressed)
 - Lead Reviewer is (1) Lacey or (2) the person who did the review if Lacey is gone.
 - Pull request is merged into the master branch

- 1. Click "Conversation" tab at the top of the pull request
- 2. Scroll down, click "Merge pull request" assuming there are no conflicts with the base branch. If there are conflicts then those should be fixed by the person who submitted the pull request.
- 3. After you will see "Pull request successfully merged and closed" with a "Delete Branch" button. Click "Delete Branch".
- Issue is closed
 - 1. The issue should be referenced in the original description of the pull request -go to it.
 - 2. Click "Close Issue".
- · New issue is created to update module on KnowPulse
 - 1. Go to https://github.com/UofS-Pulse-Binfo/KnowPulse
 - 2. Create an issue specifing that module XXX has received either new features or bug fixes and needs to be updated. Reference the issue linked to the pull request by entering [organization]/[repo]#[issue number] (e.g. UofS-Pulse-Binfo/tripal_daemon#3)
 - 3. Assign this issue to whomever is in charge of updating KnowPulse (e.g. Lacey).

2.1.2 Review

- When reviewing, look at the code changes directly _and_ test through the interface.
 - 1. To see all the changes, click on "Files Changed" at the top of the pull request.
 - 2. To start a code review, hover in front of the first line you would like to comment on then click the blue + icon that shows up. This opens a comment box where you can enter your suggestion/critique. Then click the green "Start review" button.
 - 3. Continue adding comments throughout the changes in this manner.
- If you don't actually have code review comments to make, you can submit your review by:
 - 1. Click on "Files Changed" at the top of the pull request.
 - 2. Click on "Review Changes" button at the top right of the page.
 - 3. In the resulting popup, enter your comments and select whether you approve or request changes.
- If you simply need clarification, just comment on the pull request with your questions and hold off reviewing.
- For interface functionality or design critique/suggestions, enter a comment (preferrably with screenshot) in the "Conversation" tab of the pull request.
- You may need to checkout the branch on a development site to adequetaly test the code.
 - 1. git fetch origin will fetch all branches from that repository
 - 2. git checkout -b [branch] origin/[branch] will pull the code from the branch which you want to test
 - 3. Once you've finished testing, you can revert to the master branch with git checkout master

2.2 KnowPulse Content Types

The following breaks down our content types based on the primary Chado table storing their data.

• chado.organism table (full table)

- Species (OBI:organism)
- chado.project table (seperated using projectprop)
 - Research Project (semantic science ontology; study; SIO:001066)
 - Genome Project (local:genome)
- chado.publication (full table)
 - Publication (TPUB; Publication)
- chado.feature table (seperated using feature.type_id)
 - Variant (Sequence Ontology; sequence_variant)
 - Genetic Marker (Sequence Ontology; genetic_marker)
 - QTL (Sequence Ontology; QTL; SO:0000771)
- chado.stock (seperated using stock.type_id)
 - Germplasm Accession (Crop ontology; accession; CO_010:0000044)
 - Breeding Cross (local:F1)
 - Registered Varieties (Sequence Ontology; cultivar; CO_010:0000029)
 - Recombinant Inbred Line (local:Recombinant Inbred Line)
 - Biological Sample (biological sample; sep:00195)
- chado.cvterm (specific cv_id)
 - Chickpea Trait (CO_338:ROOT)
 - Common Bean Trait (CO_335:ROOT)
 - Lentil Trait (CO_339:ROOT)
- chado.featuremap (full table)
 - Genetic Map (data onotology; Map)
- chado.stock_collection (full table)
 - Stock Collections (don't know what term to use yet)

2.2.1 Species (Organism)

The organism content type will include each of our crops and wild relatives. Stored in the Chado Organism table. Design-wise we are thinking of these as launch pads which both summarize the data per crop and link to pre-filtered searches.

Genome Pane

- @todo Current Assembly: This field should provide information on the most recent genome assembly or assemblies for this species.
- @todo Genome Browser: This field should summarize the data in the current JBrowse and have an image linking to the JBrowse instance for the current assembly
- @todo Whole-Genome Diagram: This field should show an embedded CViTjs diagram summarizing the current assembly's annotation and markers.

• @todo Past Assemblies: This field should list the past assemblies with links to archived information.

Genetic Markers Pane

- Types: This field summarizes the types of markers available for this organism.
- Quick Search: This field provides a simple keyword search with limited advanced search collapsed. When the user clicks search, they are redirected to the marker search page with their results pre-filtered.

Genetic Maps & QTL Pane

- @todo Tiled display of available maps with each tile containing map metadata such as population, date generated, number of markers, QTL list
- @todo QTL Quick Search: simple keyword search for QTLs.

Germplasm Pane

- Types: This field summarizes the types of germplasm available for this organism. Should this be subject to permissions?
- Quick Search: This field provides a simple keyword search. No need for advanced search or breakdown by type.
- @todo Recent UofS Varieties: a list of the most recent varieties.
- @todo Variety Spotlight: Highlight a single variety
- @todo RIL Populations: This could be a cool search that lets them select things like interspecific, species, phenotype of interest. Essentially allow people to search for a RIL without knowing it's number
- @todo Germplasm Diversity: Show a map with each of our germplasm indicated.
- @todo Wild Relatives: summarize the gene pools and/or phylogeny

Phenotypes

- @todo Quick Search: This field provides a simple keyword search. No need for advanced search? It would be helpful this field searched both names and linked cvterms.
- @todo Phenotype Projects: Highlight our biggest and/or most recent phenotyping projects. Should show metadata for each such as population phenotyped, purpose, siteyears w/ data, funders logos if applicable. Links to project pages.

2.2.2 Research Project

Currently just a regular old project page. Stored in the Chado Project table. Need a way to break out? What about datasets which are currently projects with data? These pages should include all metadata about the project and summarize all the data generated by it.

Warning: Not yet available as a Tripal3 content type. Currently in progress.

Overview Pane

- Project meta-data (stored as project properties)
 - Years
 - Collaborators
 - Funders
 - Locations
 - Measurements/Traits being collected
 - Genotyping being done?
 - Germplasm collection
- Abstract
- Methods

Genomic Data

- Barchart or infographic summarizing the types and amount of genomic data (sequence, genes, etc.) generated as part of this project.
- Assembly information if generated as part of this project?

Genotypic Data

- Barchart or infographic summarizing the types and amount of genotypic data (markers, variants, germplasm assayed, genotype calls, etc.) for this project.
- Genotypes Search: Redirects to the genotype matrix pre-filtered for the current project. Provides form elements to further restrict the matrix.

Phenotype Data

• Barchart or infographic summarizing the siteyears, traits and germplasm phenotyped for this project.

Germplasm

- List the germplasm associated with this project with download. It might be a long list so make it searchable. Also, we should support stock collections which would then be listed at the top. For example, AGILE is using the Lentil Diversity Panel so it would be nice to say that at the top so users know it's not a random list ;-).
- We also need a way to add germplasm to a project through the form so keep in mind this one will need a form element.

2.2.3 Publication

Stored in the Chado Publication table. These are really basic pages meant to provide attribution to data in KnowPulse. As such **All panes will be open by default**.

Warning: Not yet available as a Tripal3 content type. Currently in progress.

Overview

- citation
 - this should be really obvious. Currently highlighted in green.
- meta-data (stored as project properties)
 - date published
 - journal
 - authors
- abstract
- obvious link to full text

Data

- associated data?
- maybe in a bubble/circle chart?

2.2.4 Genome Project

The plan is to store the genome assembly as collection of analysis records (one for each program ran to create the assembly) grouped together by this project.

Warning: Not yet available as a Tripal3 content type.

Overview

Open by Default

- Publication (prominent) with link
- Meta-data about the assembly
 - Scaffolds: size + number
 - Contigs: size + number
 - Scaffold N50/L50
 - Contig N50/L50
- Meta-data about the annotation

Assembly Details

- How was the assembly generated?
 - paragraph description
 - a list of all the analysis' records with versions, etc.
- How was it annotated?
 - paragraph description
 - a list of all the analysis' records with versions, etc.

Contacts

• List of contacts for who will answer questions

Whole Genome Browsing

Open by Default

- Genome Browser
 - summarize the data in the current JBrowse
 - have an image linking to the JBrowse instance for this assembly
- Whole-Genome Diagram
 - show an embedded CViTjs diagram summarizing this assembly's annotation and markers.

BLAST Search

· BLAST quick search where database is this assembly

Bulk Data

· Files for download associated with this genome

Restrictions on dataset usage

Open by Default

• any agreements, etc.

2.2.5 Variant

This page indicates a point of known variation in the genome. This single point may be assayed by many genetic markers. Stored in the Chado Feature table broken out by type_id.

Overview

- Meta-data
 - Name
 - Туре
 - Species
 - Gene name and location in gene (if within gene)

Variant Locations

· variant locations by assembly

Genetic Markers

• List all the genetic markers assaying this variant

Genotypic Data

- pie chart summarizing the known alleles
- Genotype look-up field: mini genotype matrix for a single variant. Germplasm are the rows.
- · Genotypes Matrix Link: Redirects to the genotype matrix pre-filtered for the current variant.

Sequence

• Variant Marked-up Sequence (FASTA format) -state assembly? for all locations

Genome View

• JBrowse field located on this marker

2.2.6 Genetic Marker

The genetic marker page is currently a customized feature page. Stored in the Chado Feature table broken out by type_id.

Overview

- Meta-data
 - Name
 - Marker technology
 - Species
 - @todo Gene name and location in gene (if within gene)
 - @todo any additional meta-data helpful for running the marker (e.g. primers)

Marker Locations

• marker locations by assembly (@todo not yet by assembly)

Genotypic Data

- pie chart summarizing the known alleles
- @todo Genotype look-up field: mini genotype matrix for a single variant. Germplasm are the rows.
- @todo Genotypes Matrix Link: Redirects to the genotype matrix pre-filtered for the current marker.

Sequence

• Variant Marked-up Sequence (FASTA format) (@todo -state assembly?) for all locations

Genome View

• @todo JBrowse field located on this marker

Genetic Maps

• @todo List all the genetic maps containing this marker

QTL

• @todo List of all QTL this marker falls within?

2.2.7 Germplasm Accession

Stored in the Chado Stock table broken out by type_id.

Warning: Not yet available as a Tripal3 content type.

Overview

- Basic information such as name, accession, what-not. This will remain a default Tripal pane
- It would be great to finally have the breeder short-hand for the pedigree! This could be a separate field from the pedigree diagram so I can put it in the overview.

Parental Pedigree

- Shows the parentage pedigree diagram with all but the first 3 levels collapsed.
- All diagrams should now include a figure legend.

Genotypes Pane

- Types: This field summarizes the number of markers per type with calls for the current germplasm.
- Genotype Look-up field: a mini-genotype matrix filtered to this germplasm only. User enters a marker or region to display.
- Genotype Matrix Link: Redirects to the genotype matrix pre-filtered for the current germplasm.

Breeding Markers

• List the calls for a set of CORE breeding markers. Make the set configurable at the field level.

Phenotypes Pane

• List the value for a set of CORE phenotypes (e.g. seed coat, cotyledon color). Make the set configurable at the field level. Should it be entereable on the form? How do we handle multiple experiments, siteyears for a single trait/variety combo.

Relationships

• list any non-parental/progeny relationships.

2.2.8 Cultivated Variety

Stored in the Chado Stock table broken out by type_id. See https://www.crop-diversity.org/mgis/accession/ 01BEL084245

Warning: Not yet available as a Tripal3 content type.

Overview

- Basic information such as name, accession, what-not. This will remain a default Tripal pane
- It would be great to finally have the breeder short-hand for the pedigree! This could be a separate field from the pedigree diagram so I can put it in the overview.
- specific variety information
- link to registration pub
- abstract from pub

Parental Pedigree

- Shows the parentage pedigree diagram with all but the first 3 levels collapsed.
- All diagrams should now include a figure legend.

Genotypes Pane

- Types: This field summarizes the number of markers per type with calls for the current germplasm.
- Genotype Look-up field: a mini-genotype matrix filtered to this germplasm only. User enters a marker or region to display.
- Genotype Matrix Link: Redirects to the genotype matrix pre-filtered for the current germplasm.

Breeding Markers

• List the calls for a set of CORE breeding markers. Make the set configurable at the field level.

Phenotypes Pane

- List the value for a set of CORE phenotypes (e.g. seed coat, cotyledon color). Make the set configurable at the field level. Should it be entereable on the form? How do we handle multiple experiments, siteyears for a single trait/variety combo.
- Phenotype pane should highlight phenotypes specifically of interest for this cultivar.

Relationships

• list any non-parental/progeny relationships.

2.2.9 Breeding Cross

Stored in the Chado Stock table broken out by type_id.

Warning: Not yet available as a Tripal3 content type.

Overview

- Basic information such as name, accession, what-not. This will remain a default Tripal pane
- Overview will include information about the crossing block
- It would be great to finally have the breeder short-hand for the pedigree! This could be a separate field from the pedigree diagram so I can put it in the overview.

Crossing Block

- year, season,
- number of crosses made in the same crossing block

Parental Pedigree

- Shows the parentage pedigree diagram with all but the first 3 levels collapsed.
- All diagrams should now include a figure legend.

Progeny

• list of progeny the current cross was a parent for (includes crossing block info).

Genotypes Pane

- Types: This field summarizes the number of markers per type with calls for the current germplasm.
- Genotype Look-up field: a mini-genotype matrix filtered to this germplasm only. User enters a marker or region to display.
- Genotype Matrix Link: Redirects to the genotype matrix pre-filtered for the current germplasm.

Breeding Markers

• List the calls for a set of CORE breeding markers. Make the set configurable at the field level.

Phenotypes Pane

• List the value for a set of CORE phenotypes (e.g. seed coat, cotyledon color). Make the set configurable at the field level. Should it be entereable on the form? How do we handle multiple experiments, siteyears for a single trait/variety combo.

Relationships

• list any non-parental/progeny relationships.

2.2.10 Recombinant Imbred Line

Stored in the Chado Stock table broken out by type_id.

Warning: Not yet available as a Tripal3 content type.

Overview

- Basic information such as name, accession, what-not. This will remain a default Tripal pane
- RIL Specific information
 - the original cross
 - parents,
 - latest generation,
 - status
 - number of individuals.
- It would be great to finally have the breeder short-hand for the pedigree! This could be a separate field from the pedigree diagram so I can put it in the overview.

RIL Development Pane

- Reiterate the type of cross (Interspecific, intraspecific)
- Indicates the parents of the RIL including their species.
- Development timeline including which years & how many individuals
- If the RIL is complete, indicate that and the final number of individuals.

Parental Pedigree

- Shows the parentage pedigree diagram with all but the first 3 levels collapsed.
- All diagrams should now include a figure legend.

Genotypes Pane

- Types: This field summarizes the number of markers per type with calls for the current germplasm.
- Genotype Look-up field: a mini-genotype matrix filtered to parents of this RIL. User enters a marker or region to display.
- Genotype Matrix Link: Redirects to the genotype matrix pre-filtered for the parents of this RIL.
- Links to the VCF Filter form for the current RIL to facillitate downloading of the data.

Breeding Markers

• List the calls for a set of CORE breeding markers. Make the set configurable at the field level.

Phenotypes Pane

- List the value for a set of CORE phenotypes (e.g. seed coat, cotyledon color). Make the set configurable at the field level. Should it be entereable on the form? How do we handle multiple experiments, siteyears for a single trait/variety combo.
- Phenotype pane should highlight phenotypes specifically of interest for this RIL.

Relationships

• list any non-parental/progeny relationships.

2.2.11 Stock Collection

Warning: Not yet available as a Tripal3 content type.

Overview Pane

- paragraph describing the population
- · development/source information

Genomic Data

• Barchart or infographic summarizing the types and amount of genomic data (sequence, genes, etc.) collected for this collection.

Genotypic Data

- Barchart or infographic summarizing the types and amount of genotypic data (markers, variants, germplasm assayed, genotype calls, etc.) for this collection.
- Genotypes Search: Redirects to the genotype matrix. Provides form elements to choose germplasm from this collection.

Phenotype Data

• Barchart or infographic summarizing the siteyears, traits and germplasm phenotyped for this collection.

Germplasm

• List the germplasm associated with this collection with download. It might be a long list so make it searchable.

2.3 KnowPulse Field Types

The following sections described common field types.

2.3.1 Quick Search (Should be in Core?)

- This field type allows a user to search for additional content quickly and easily.
- It consists of a single keyword search box and optional, configurable advanced search form.
- Configuration: Search URL this field redirects to what key to use in the query string for entered keywords -Drupal Form API definition of advanced search elements where the keys of the elements are used in the query string - Description and/or help text to display to the user - Quick Search Title

2.3.2 Summary Field (Should be in Core?)

- This field displays a graphical summary based on a materialized view.
- Configuration: Select the Materialized View to get the summary from Select which mview column maps to the legend (e.g. type name) Optional filter by field: select field and mview column mapping Select chart type: Pie, bar, etc. Figure Title Figure Description Chart type specific configuration

2.3.3 Embedded JBrowse by organism

- Embeds a JBrowse instance
- Provided by the JBrowse integration module
- Contains linkout to full JBrowse page for more intensive browsing

• Configuration: - Select Organism Field for the current content type - Show/Hide Menubar, search, track selector - For each organism: select jbrowse, tracks, and location to show

2.3.4 Whole-Genome View by organism

- Embeds a CViTjs diagram
- Provided by the CViTjs integration module
- Contains linkout to full diagram
- Configuration: Select Organism Field for the current content type Show/Hide Controls, track selector For each organism: select diagram and tracks to show

2.4 Task-based Tutorials

2.4.1 Update KnowPulse

The following documentation outlines the steps and testing that should be done when updating KnowPulse, whether it be simply updating a custom module or updating Drupal itself. Updating between major versions (e.g. Drupal $7 \Rightarrow 8$ or Tripal $2 \Rightarrow 3$) is beyond the scope of this document.

The following are the basic steps to be taken. There is more information below if needed.

- 1. Create a fresh clone (Dev Fresh) of the production site
- 2. Upgrade the clone.
 - In the directory for the site, execute *drush up*. This will tell you what modules need to be updated and prompt you to do so. Remember, if you don't have permission on the directory you will need to use sudo.
- 3. Test the upgrade by checking the pages listed under "To be Checked" below.
- 4. If there are no errors (check the current page, the recent log messages and the apache error log) then upgrade the production site using the same steps.
 - Recent Log Messages: http://knowpulse.usask.ca/portal/admin/reports/dblog
 - Apache Error Log: /var/log/apache2/error.log (requires sudo).
- 5. Check the pages listed under "To be Checked" on the production site as well.

Pages To Be Checked

This list should be ever expanding. Feel free to document the specifics under each page.

- Home Page: http://knowpulse.usask.ca/
- BLAST: http://knowpulse.usask.ca/portal/blast/nucleotide/nucleotide
 - You should actually execute an example blast to ensure that the functionality works.
- JBrowse: http://knowpulse.usask.ca/portal/jbrowse/Lentil
 - Ensure you have your cache killer turned on!
 - Make sure to turn on/off tracks, navigate side to side and search.
- Views Search: http://knowpulse.usask.ca/portal/search/variants

- Check multiple combinations of filters (e.g. genus:Lens, type:SNP, name:p390).
- If you get no results and see a green message as follows, you likely need to re-apply the Tripal Views patch.

Warning: > Exception in KP Variant Search[kp_variant_search]: SQLSTATE[42P01]: Undefined table: 7 ERROR: relation "feature" does not exist > LINE 5: feature feature

- To apply views patch, navigate to views module directory and execute *git apply* .../tripal/tripal_views/views-sql-compliant-three-tier-naming-1971160-30.patch
- Feature Page: http://knowpulse.usask.ca/portal/feature/lens-culinaris/SNP/LcC00002p390
 - Check the genotypes & sequence panes especially.
- · Raw Phenotypes
 - Chart: http://knowpulse.usask.ca/portal/phenotypes/raw
 - Backup: http://knowpulse.usask.ca/portal/phenotypes/raw/backup
- Germplasm: http://knowpulse.usask.ca/portal/stock/Lens/culinaris/Variety/KP%3AGERM58
 - Check that you can collapse nodes & follow links on the pedigree
- Research Category: http://knowpulse.usask.ca/portal/node/21
 - Check that there are projects listed & pager works.

Create a Clone of KP

Note: Replace * * * with the name of the database user in all the commands below.

- 1. Backup the production site to your home directory
 - pg_dump kp3_production --user *** --host thunder > ~/ kp3_production.2017Aug18.sql
- 2. Remove the existing Clone. This should be safe since no one is supposed to develop on the clone. That said, it wouldn't hu
 - sudo rm -R dev/fresh
 - psql --user *** --host thunder --command "DROP DATABASE kp3_fresh"
- 3. Re-create the clone database and restore the production site backup to it.
 - psql --user *** --host thunder --command "CREATE DATABASE kp3_fresh WITH OWNER ***"
 - psql kp3_fresh --user *** --host thunder < ~/kp3_production.
 2017Aug18.sgl
- 4. Copy the production site files to dev/fresh.
 - sudo cp -R ../portal fresh
 - sudo chown -R lacey-local:www-dev fresh/
- 5. Change the database in the dev/fresh/sites/default/settings.php from kp3_production to kp3_fresh.
 - You need to use sudo to change this file.

6. Change the write permission of dev/fresh/sites/default/files to be writable by all.

```
• sudo chmod 777 files/
```

2.4.2 Add new Genome Build to KP

Overview

The following list is a general overview. More details will be added below as needed.

- 1. Create a blast database and register it with the Tripal Blast UI module.
- 2. Create a JBrowse instance and register it with the Tripal JBrowse module.
- 3. Setup bulk download, if desired.

Blast

1. Prefix the record names in your FASTA file using this GIST. The naming convention is The first letter of the Genus and two letters from the species followed by the version and an underscore. For example, Lens culinaris 1.0 would have Lcu1.0_ added to the beginning of each record name.

> perl prefix_fasta.pl [original fasta] [new fasta] [prefix]

2. Create the BLAST indicies.

> formatdb -p F -t Genus_species_version_genome -i Genus_species_version_genome.fasta_ --n Genus_species_version_genome -o T

3. Create BLAST database on KnowPulse. Go to Admin Toolbar > Content > Add Content > Blast Database. Convention for naming is Genus species: Genome version.

Note: File Location: Both the prefixed FASTA and the associated indices should be located in the blast_dbs folder with all the other KnowPulse blast databases. If you are uncertain where that is, simply check the page for an existing blast database to see it's path.

Note: Restrict Access: You can restrict the users which can blast against a specific blast database by editing the page created in step #3 above. Near the bottom there is a section for Access in the horizontal tabs. Simple select the group you would like to give access (usually University of Saskatchewan) and all users not in that group will not see that blast database as an option.

Warning: More information on Link-outs to come.

JBrowse

1. Setup a new JBrowse following the instructions here: https://jbrowse.org/install/. The folder containing the JBrowse files should be named [random 5 letters]-[genome prefix].

- Use the prefixed fasta file created for the BLAST as the backbone.
- Make sure to update any annotation so the backbone/chromosome feature names match.
- 2. Add KnowPulse themeing to the new JBrowse by copying the CSS files from an existing JBrowse. This needs to be improved. Ideally all KnowPulse specific changes would be in a single CSS file added to the index.
- 3. Edit jbrowse.conf to change the Share link to point to the KnowPulse path by adding the following to the [General] section.

Note: Notice the [Instance Path] in the URL above; this should be changed to the path selected in step 4. Use existing jbrowse.conf as an example for other configuration changes.

- 4. Register the JBrowse with KnowPulse by creating a JBrowse Instance. Go to Admin Toolbar > Content > Add Content > JBrowse Instance and fill in the requested information. Also, change the URL path settings to provide an alias for the JBrowse page. Conventions are as follows:
 - Cultivated crops, current version: [Common Name]
 - Cultivated crops, archived version: [Common Name]/[version]
 - Wild Species, current version: [Genus species]
 - Wild Species, archived version: [Genus species]/[version]

2.5 Tool Customization

The following documents discuss how we customize third-party tools for better integration with KnowPulse in an effort to provide a consistent user experience.

2.5.1 JBrowse

We have changed the colours of our JBrowse to make it match KnowPulse for more seamless integration. The following shows our customizations:

```
/* Change the color of the upper trapezoid */
#GenomeBrowser.jbrowse div.locationTrap {
    border-bottom-color: #17313E
}
#GenomeBrowser.jbrowse div.locationThumb {
    border: 2px solid #17313E;
}
/* Change the color/Background of the top menu bar */
#GenomeBrowser.jbrowse .menuBar {
    background: url('http://knowpulse.usask.ca/portal/sites/all/themes/kptheme/images/
    ·mainbarbg.jpg') repeat-x;
    text-align: right;
}
```

(continues on next page)

(continued from previous page)

```
/** Customize HTML Features */
#GenomeBrowser.jbrowse .feature2,
    #GenomeBrowser.jbrowse .plus-feature2,
    #GenomeBrowser.jbrowse .minus-feature2 {
        height: 7px;
        background-color: #62d335;
    #GenomeBrowser.jbrowse .match_part,
    #GenomeBrowser.jbrowse .plus-match_part,
    #GenomeBrowser.jbrowse .minus-match_part {
        height: 10px;
        background-color: #00FF00;
}
```

2.6 ReadtheDocs Guide

Documentation for our generic modules should be written in **Restructured Text**, compiled with Sphinx, and built/hosted with ReadTheDocs. This directory, when compiled, is hosted at https://[yourmodule]. readthedocs.io/en/latest/

For minor changes, you can simply Edit the file using the Github editor, which will allow you to make a Pull Request. Once approved, your changes will be reflected in the documentation automatically!

2.6.1 Install Sphinx

For minor changes, you don't need to build the documentation! If you want to see how your changes will look on the built site, however, you will need Sphinx installed.Sphinx packages are published on the Python Package Index. The preferred tool for installing packages from PyPI is pip. This tool is provided with all modern versions of Python.

On Linux or MacOS, you should open your terminal and run the following command.

\$ pip install -U sphinx

After installation, type sphinx-build --version on the command prompt. If everything worked fine, you will see the version number for the Sphinx package you just installed.

For more information, please see the Sphinx setup guide: http://www.sphinx-doc.org/en/master/usage/quickstart.html

2.6.2 Building your changes

For more extensive edits, or when contributing new guides, you should build the documentation locally. From the docs root (e.g. [your module]/docs/, execute make html. The built site will be in docs/_build/html/ index.html.

2.6.3 Guidelines

Please follow these guidelines when updating our docs. Let us know if you have any questions or something isn't clear.

Please place images in the same folder as the guide text file, following the convention [file_name]. [n].[optional description].[extension]. For example, configuring_page_display.3. rearrange.png or configuring_page_display.1.png are both located in docs/user_guide/ and are part of the configuring_page_display.rst guide.

We currently use the following syntax:

```
Title of File (using title case)
_____
Introduction text.
.. note::
 This would show up as a boxed note to the reader. It's good for ensuring important
-points/hints are seen but should be used sparingly. It's also a good way to note_
→that "This guide is under developement."
Section Title
 _____
We use double backticks to indicate ``inline-code`` including file names, function_
→and method names, paths, etc.
Longer code-blocks should begin with the ``.. code-block:: [type]`` directive and
\hookrightarrow should be indented at least one
level. There should also be a blank line before and after it as shown below.
.. code-block:: bash
 if ($needs_documentation) {
     use $these_guidelines;
      $contribute_docs = $appreciated;
Section 1.1 Title
~~~~~~
The use of appropriate sections makes reading documentation and later specific,
→details easier. Sub sections such
as this one will be hidden unless the main section is already selected.
The following toctree specifies that there are 3 files with additional content for_
\rightarrow the current page.
The order is as it will appear in the sidebar and the link titles will be the "Title_
\rightarrow of File" for each file.
.. toctree::
  :maxdepth: 2
  :caption: Contents:
  dir/file1
  dir/file2
  dir/file3
```

For more information on syntax, see this RestructuredText Primer.

2.6.4 Starting Docs for a new Module

- 1. Install sphinx using pip and navigate yourself into git directory.
- 2. Create a docs folder there and go inside. Run sphinx-quickstart and set to your needs. Simply accept the defaults except for:

```
The project name will occur in several places in the built documentation.
> Project name: Genotypes Loader
> Author name(s): Carolyn T Caron et al., University of Saskatchewan, Pulse_
→Bioinformatics
> Project release []:
```

- 3. Create one .rst document per planned section and populate it accoding to the guidelines above. If a given section is long enough to be broken up into sub-sections then create a folder with the same name as [your section].rst and place all .rst files for that section within it.
- 4. For each .rst file, ensure it is added to the toctree for that section. The base sections should be added to the index.rst
- 5. Edit the index.rst: change the title to [My Module]: Available Documentation and remove the Indices and Tables section.
- 6. Ensure your conf.py matches the one provided here. by comparing each section. Do not change your project information and pay special attention to PHP Highlighting and Options for HTML output.
- 7. Copy_static from this repo to your docs folder. This contains the UofS-Pulse-Binfo theme alterations.
- 8. Build your documentation using make html and check _build/html/index.html to ensure your documentation looks as it should.
- 9. Add _build/ to your .gitignore file and commit your changes pushing them out to github.
- 10. Go to https://readthedocs.org/dashboard/ and add this module as a project.

Note: If Theme error: no theme named 'sphinx_rtd_theme' found (missing theme. conf?) is encountered in step 8. This error indicates sphinx theme package sphinx-rtd-theme is not available in your system and it can be installed by executing following command on Linux or MacOS.

sudo pip install sphinx-rtd-theme