



KnowPulse: A Web-Resource Focused on Diversity Data for Pulse Crop Improvement

Lacey-Anne Sanderson, Carolyn T. Caron, Reynold Tan, Yichao Shen, Ruobin Liu and Kirstin E. Bett*

Department of Plant Sciences, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:

Matthew Nicholas Nelson,
Agriculture and Food
(CSIRO), Australia

Reviewed by:

Ethalinda K. S. Cannon,
Iowa State University, United States
Elisa Bellucci,
Marche Polytechnic University, Italy

*Correspondence:

Kirstin E. Bett
k.bett@usask.ca

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 03 April 2019

Accepted: 10 July 2019

Published: 31 July 2019

Citation:

Sanderson L-A, Caron CT, Tan R,
Shen Y, Liu R and Bett KE (2019)
KnowPulse: A Web-Resource
Focused on Diversity Data for Pulse
Crop Improvement.
Front. Plant Sci. 10:965.
doi: 10.3389/fpls.2019.00965

KnowPulse (<https://knowpulse.usask.ca>) is a breeder-focused web portal for pulse breeders and geneticists. With a focus on diversity data, KnowPulse provides information on genetic markers, sequence variants, phenotypic traits and germplasm for chickpea, common bean, field pea, faba bean, and lentil. Genotypic data is accessible through the genotype matrix tool, displayed as a marker-by-germplasm table of genotype calls specific to germplasm chosen by the researcher. It is also summarized on genetic marker and sequence variant pages. Phenotypic data is visualized in trait distribution plots: violin plots for quantitative data and histograms for qualitative data. These plots are accessible through trait, germplasm, and experiment pages, as well as through a single page search tool. KnowPulse is built using the open-source Tripal toolkit and utilizes open-source tools including, but not limited to, species-specific JBrowse instances, a BLAST interface, and whole-genome CVITjs visualizations. KnowPulse is constantly evolving with data and tools added as they become available. Full integration of genetic maps and quantitative trait loci is imminent, and development of tools exploring structural variation is being explored.

Keywords: legumes, pulses, web resource, diversity, genotypic data, phenotypic data

INTRODUCTION

Legumes are immensely important in agricultural ecosystems with the legume family (Leguminosae) being second only to the grass family (Poaceae) in economic and nutritional value (Graham and Vance, 2003). Grain legumes, also known as “pulses,” are primarily marketed for human consumption and are a good source of dietary fiber, protein, slow-release carbohydrates, B vitamins, iron, copper, magnesium, manganese, zinc, and phosphorous (Tharanathan and Mahadevamma, 2003; Polak et al., 2015). They are also naturally low in fat, virtually free of saturated fat and cholesterol free (Polak et al., 2015). In recent years there has been an explosion of genome assemblies for legumes (Varshney et al., 2009, 2012, 2013; Schmutz et al., 2010, 2014; O’Rourke et al., 2014; Tang et al., 2014; Parween et al., 2015; Pandey et al., 2016). In addition, there has been a dramatic increase in sequence variation data (Kamfwa et al., 2015; Boutet et al., 2016; Moghaddam et al., 2016; Pandey et al., 2016; Gali et al., 2018; Ogutcen et al., 2018). In order to maximize the usefulness of this data, it should be curated with connections between phenotypic and genotypic data verified in a web resource which is friendly to both breeders and researchers.

Several legume-focused databases have been developed including Legume Information System (LIS; <https://legumeinfo.org>, Dash et al., 2015), Medicago truncatula Genome Database (<http://www.medicagogenome.org>, Krishnakumar et al., 2014), SoyBase (<https://www.soybase.org/>, Grant et al., 2009), PeanutBase (<https://peanutbase.org>; Dash et al., 2016), and Cool Season Food Legume Database (<https://www.coolseasonfoodlegume.org/>). While these resources are invaluable to their crop-specific and comparative communities, none provide the integration between germplasm, genotypic and phenotypic data to adequately develop the genetic markers useful in pulse breeding programs.

Over 100 plant and animal databases use Tripal (<https://www.drupal.org/project/tripal>; http://tripal.info/sites_using_tripal, Sanderson et al., 2013), an open-source, highly customizable toolkit providing efficient development of biological web portals. Tripal extends the popular Drupal content management system (CMS). Use of a CMS enables developers to focus on the specific needs of their community without the overhead of user and security management, or the database schema design frequently associated with web portal development. Tripal's use of the Generic Model Organism Database (GMOD) Chado schema (Mungall and Emmert, 2007) provides flexible support for biological data, while facilitating the exchange of data and expertise among Tripal sites through common infrastructure.

KnowPulse, a breeder-focused web portal, was first released in 2010 to serve the pulse breeders at the University of Saskatchewan. There is a focus on common bean, chickpea, field pea, lentil and faba bean, as these are the crops of interest in their program. KnowPulse is built using Tripal, with the purpose of serving as a reliable data storage solution with metadata preservation. It has since evolved into a public resource by housing a large number of continually expanding datasets focused on genetic variation. We describe the novel genetic variation display and tools of KnowPulse below to inform the greater legume community.

MATERIALS AND METHODS

Datasets

KnowPulse houses data for chickpea, dry bean, field pea, lentil, and faba bean. The magnitude of all data is summarized by type (e.g., germplasm, genotypes, phenotypes) on the home page. There is information on Genebank accessions and University of Saskatchewan cultivars. Users can access a number of genotypic (i.e., genetic markers, sequence variants, and genotypic calls) and phenotypic (i.e., traits, experiments, and measurements) datasets. Lastly, the pre-release genomic sequence information for *Lens culinaris* is available through the web portal by request. In an effort to provide researchers with data as soon as possible, KnowPulse houses unpublished data. However, all data is required to have a long-term data management plan ensuring integrity and availability.

Implementation

KnowPulse uses Drupal 7 (<https://www.drupal.org/>), an open-source enterprise-level content management system, and Tripal

3, which extends Drupal for biological data. The modular PHP framework provided by Drupal and Tripal allows KnowPulse to use community-contributed extensions and an advanced administrative interface to speed up development time and provide more functionality to users. The core Tripal modules power the ontology-driven content pages (e.g., genetic markers, germplasm accessions, research projects), content-type specific searches and semantic web-ready web services for all content. Customized displays were developed through extension modules. The entire technology stack is open-source and all extension modules are publicly available on GitHub and open to collaboration (<https://uofs-pulse-binfo.github.io/our-modules/>).

All data, excluding the BLAST databases, are stored in a single PostgreSQL instance using the Drupal schema and GMOD Chado schema (Mungall and Emmert, 2007) for web-related data and biological data, respectively. PostgreSQL constraints and data type checking ensure data integrity and standards compliance. For example, genotypic data must be linked to the germplasm assayed, the experiment, and the genetic marker including assay information. Well-chosen indices and materialized views mitigate any performance issues incurred by use of a relational database by speeding up queries. This combination allows us to meet the speed and data integrity needs of the user.

Permissions and Accessibility

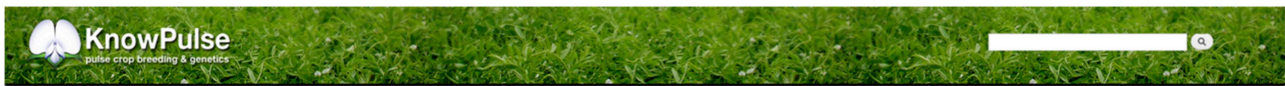
KnowPulse acts as both a public data portal and a private breeding program management system. All the functionality described herein is publicly available unless otherwise stated. Since KnowPulse provides access to pre-publication data, you may find restrictions on download for specific datasets and watermarked charts. Private data and tools can be accessed via a user account with specific permissions. If you need access to private data for your research, please contact Dr. Kirstin Bett, corresponding author, with an explanation and in most cases we will be happy to collaborate with you.

RESULTS AND DISCUSSION

Genomic Variation

In the genomic context, genotypic data are particularly important in KnowPulse. These data are used by researchers for marker development and association studies with the ultimate goal of facilitating pulse crop breeding. KnowPulse provides a germplasm-by-variant genotype matrix for researchers to explore genotypic data for their germplasm set (**Figure 1**). Since genotypic datasets are increasingly expanding, this tool provides filter options including experiment, variant list, genomic position, marker or variant type, and pairwise polymorphisms. Additionally, if the data is overwhelming to analyze within the browser, users can request permission to download it via KnowPulse in a variety of formats (e.g., comma-separated values, hapmap).

Sequence variants and genetic markers are each represented with their own pages in KnowPulse. Sequence variant pages list all the markers available for a given genomic position, whereas genetic marker pages provide details for a specific marker assay.



Home > Search Biological Data > Genotypes > Lens Genotypes > Search

Lens Genotypes

1 Choose germplasm you are interested in.
Simply enter the name of one germplasm (e.g. CDC Redberry) of interest below and then click the green plus (+) button. You can enter any number of germplasm you are interested in and each will be added to the matrix as they are entered.

Germplasm

CDC Rosetown	<input type="checkbox"/>	Lens culinaris	<input type="checkbox"/>	X
CDC Blaze	<input type="checkbox"/>	Lens culinaris	<input type="checkbox"/>	X
CDC Vantage	<input type="checkbox"/>	Lens culinaris	<input type="checkbox"/>	X
ILL 7502	<input type="checkbox"/>	Lens culinaris	<input type="checkbox"/>	X
Germplasm/Stock Name	<input type="text"/>	Lens culinaris	<input type="checkbox"/>	<input type="button" value="+"/>

Specify the stock (and species of the stock) you want to display the genotypes of.

Polymorphic Variants
Between and
Restrict the variants to those that have different allele calls for the selected germplasm.

2 Restrict 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 manageable.

Genome Range
From Start Position to
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.

3 Additional Filter criteria. (optional)

Bioinformatic Tools

- BLAST
- BLASTs
- JBrowse
- Chickpea (kabull, v1.0)
- Common Beans (Phytozome, v1.0)
- Lentil (LcUC, v1.2)
- Medicago (MGAG, v4.0)
- Soybean (Phytozome, v2.0)
- User-driven Charts
- CVTJs Whole-Genome Visualization
- Position Convert: Lc0.8 to Lc1.2
- VCF Bulk Export

Genomic Data

- Sequence Variants
- Markers
- Lentil Genotypes

Our Research

- Home
- Bioinformatics Resource
- Student Information
- Job Postings
- Field Lab
- Projects
- Publications
- Workshops
- Research
 - Breeding & Genetics
 - Pathology
 - Physiology
 - Tissue Culture
- Crops
 - Chickpea
 - Dry Bean
 - Faba Bean
 - Lentil
 - Pea
- All Species of Interest

User login

Email or username *

 Password *

Create new account
 Request new password

Download: Requires log in

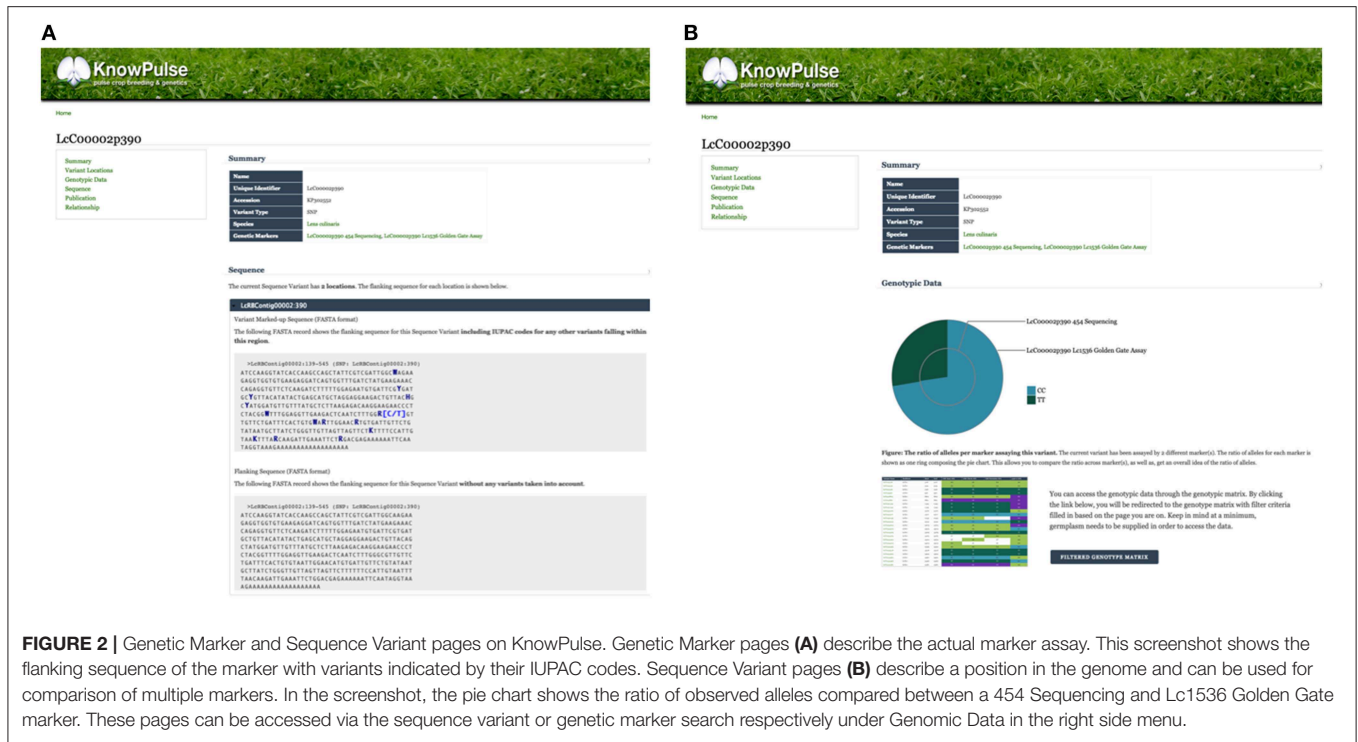
Total Results: 47
Unique Variants: 45

Sort by Location, Variant Name

Variant Name	Backbone	Start	End	CDC Rosetown	CDC Blaze	CDC Vantage	ILL 7502
LeC092692098	LeChr1	501711	501711	CC	CC	CC	GG
LeC00059312	LeChr1	5365369	5365369			AA	GG
LeC01355498	LeChr1	8246786	8246786	TT	TT	TT	
LeC21917231	LeChr1	14106568	14106568	TT	TT	TT	TT
LeC029992578	LeChr1	18771493	18771493	TT	TT	TT	
LeC23699475	LeChr1	20421764	20421764			TT	TT
LeC07589354	LeChr1	21288390	21288390	TT	TT	TT	CC
LeC04303409	LeChr1	22056682	22056682	TT	TT	TT	GG
LeC02136479	LeChr1	22067695	22067695	CC	CC	CC	TT
LeC21132691	LeChr1	22695065	22695065	AA	AA	AA	AA
LeC02136499	LeChr1	24841812	24841812	TT	TT	TT	TT
LeC029992758	LeChr1	25486920	25486920	TT	TT	TT	
LeC32069908	LeChr1	25541704	25541704	TT	GG	GG	GG
LeC02999232	LeChr1	25586239	25586239	TT	TT	TT	TT
LeC013782000	LeChr1	26495941	26495941	GG	AA	AA	AA
LeC02999459	LeChr1	26918317	26918317	GG	GG	AA	GG
LeC04997355	LeChr1	26995385	26995385	CC	CC	CC	CC
LeC00059312	LeChr1	33666446	33666446		AA		GG
LeC17038163	LeChr1	34227817	34227817	GG	GG	GG	GG
LeC07178591	LeChr1	34262344	34262344	CC	CC	CC	CC
LeC07042683	LeChr1	42842311	42842311		GG	GG	GG
LeC83949339	LeChr1	43437423	43437423	TT	CC	CC	TT
LeC06882495	LeChr1	44864335	44864335	CC	CC	CC	TT
LeC26599431	LeChr1	55631847	55631847	GG	AA	GG	GG
LeC01778299	LeChr1	55633165	55633165	AA	AA	AA	
LeC06463847	LeChr1	60786850	60786850	AA	AA	AA	AA
LeC0720966	LeChr1	60788855	60788855	GG	GG	GG	GG
LeC07591936	LeChr1	62429849	62429849	AA	AA	AA	AA
LeC10794249	LeChr1	68141831	68141831	CC	CC	CC	TT
LeC039976603	LeChr1	69780621	69780621	AA	AA	AA	GG
LeC14313328	LeChr1	69967105	69967105	AA	AA	TT	TT
LeC23246666	LeChr1	70013613	70013613	TT	TT	GG	
LeC23246666	LeChr1	70338647	70338647	CC	GG	CC	CC
LeC02912353	LeChr1	70349269	70349269	AA	AA		GG
LeC061287449	LeChr1	70444918	70444918	AA	AA		GG
LeC012229311	LeChr1	70999553	70999553	TT	TT	TT	TT
LeC14164965	LeChr1	71077304	71077304	TT	CC	TT	CC
LeC27249351	LeChr1	71339587	71339587	CC	AA	CC	CC
LeC170649236	LeChr1	71323483	71323483	GG	AA	GG	GG
LeC05328612	LeChr1	71342232	71342232	GG	AA	GG	GG
LeC022132006	LeChr1	71394063	71394063			AA	AA
LeC05509579	LeChr1	71503219	71503219	CC	AA	AA	AA
LeC25642009	LeChr1	71520591	71520591	AA	GG	AA	GG
LeC03907264	LeChr1	71529281	71529281		AA	GG	GG
LeC01042928	LeChr1	71650276	71650276		AA	CC	CC
LeC142482002	LeChr1	71667952	71667952	TT	TT	CC	CC
LeC00068665	LeChr1	71686610	71686610	AA	GG	AA	AA

= first 100 < previous LeChr1: 501710-71686610 next > last 100 =

FIGURE 1 | Germplasm by variant genotype matrix. This screenshot shows the genotype matrix for CDC Rosetown, CDC Blaze, CDC Vantage, and ILL 7502 restricted to the beginning of LcChr1. The form near the top provides additional filter options while the color-coded table below shows the allele calls for each known variant. Researchers can use this tool to inspect the genotypes of a region of interest (e.g., QTL region) for their germplasm set. This tool can be accessed in the right side menu under Genomic Data > Sequence Variants > Lentil Genotypes.



This distinction allows researchers to evaluate genotypes in context of the assay. Additionally, genetic marker pages pinpoint the location of the variant on each available genome assembly. More advanced features include: the flanking sequence with additional known variants indicated using their IUPAC codes, a pie chart summarizing the allele calls recorded, and a link to the genotype matrix to access specific calls for germplasm of interest (Figure 2A). Sequence variant pages reveal similar information with the context of all markers for that variant for comparison (Figure 2B).

A number of tools which provide further context to these genetic markers through whole-genome visualizations include CViTJs (<https://github.com/LegumeFederation/cvitjs>) and JBrowse (Buels et al., 2016). CViTJs provides whole-genome views of specific datasets such as gene and genetic marker distribution. These are available on KnowPulse for chickpea, common bean, lentil, soybean, and medicago (Figure 3A). CViTJs charts allow researchers to see broad trends across the genome; whereas, JBrowse instances are highly suitable for graphical browsing of a specific region of interest. KnowPulse has JBrowse instances for kabuli chickpea (v1.0, Varshney et al., 2013), common bean (v1.0, Schmutz et al., 2014), lentil (v1.2, Ramsay et al., 2014), soybean (v2.0, Schmutz et al., 2010), and medicago (v4.0, Tang et al., 2014) with tracks for gene sets, genetic markers, and putative orthologs from related species (Figure 3B).

Tripal BLAST (https://www.drupal.org/project/tripal_blast) provides sequence alignment searches for users with a region of interest but no prior information about its location in hosted genome assemblies. In KnowPulse, users can BLAST against pulse-specific datasets such as genome and transcript

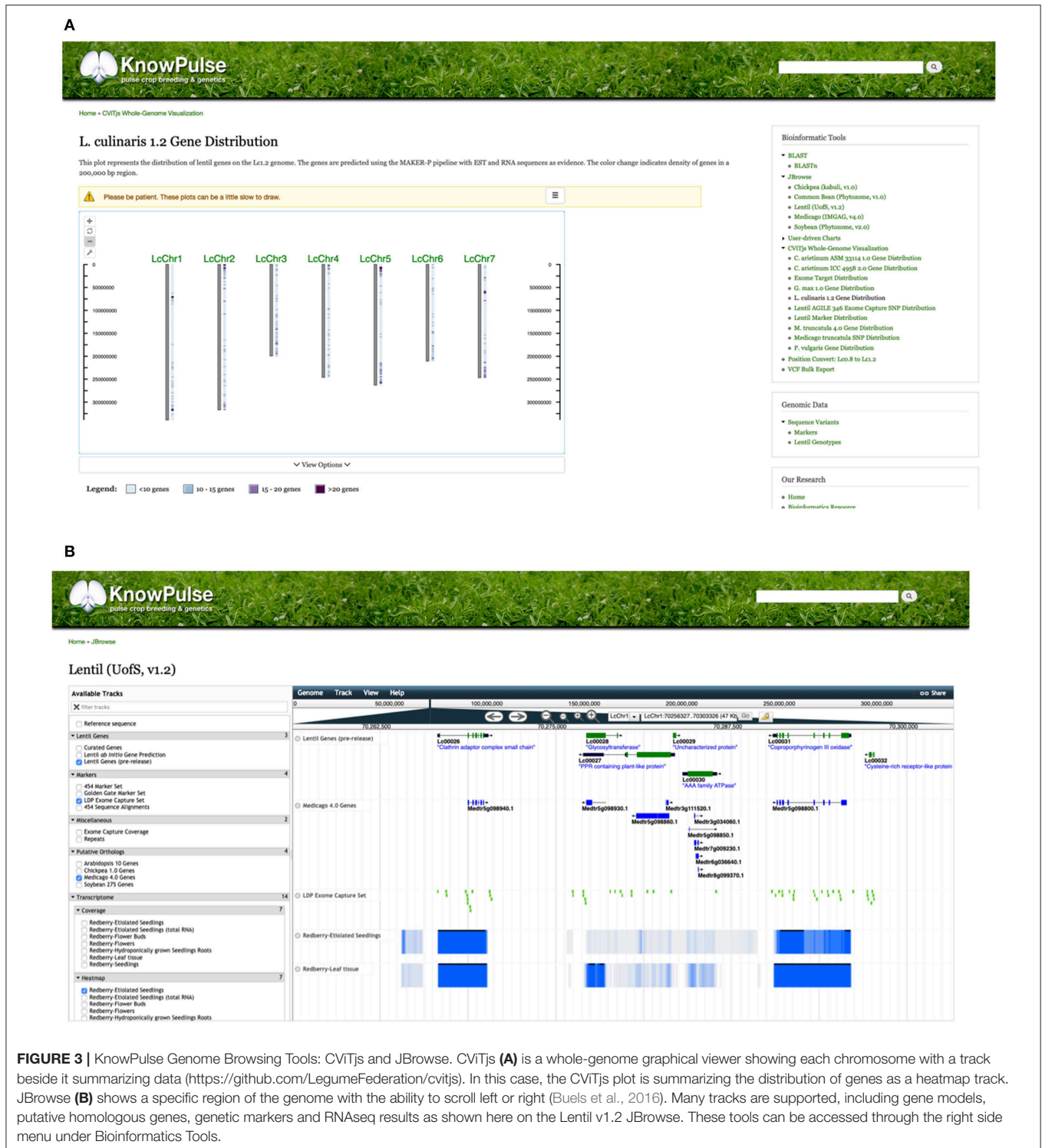
assemblies for crops (i.e., chickpea, common bean, field pea, and lentil), related wild species, and model legume species (i.e., soybean, lotus, medicago). The user simply enters their sequence in the search box, selects the dataset to BLAST against and clicks BLAST, which uses NCBI BLAST+ command-line tools (Camacho et al., 2009) to perform the search. The results are then displayed in a table with links to the appropriate JBrowse.

Phenomic Variation

With our focus on variation data, phenomics is a very important component of KnowPulse. Not only are phenotypic data used for association studies and marker discovery, they are also used for breeding activities such as germplasm selection and identification. As such, visualizations focus on the distribution of phenomic data, often in reference to specific germplasm and between site-years within an experiment.

KnowPulse provides trait distribution plots to summarize phenotypic data for a given experiment. Data from different site-years are stored separately but averaged across replicates. For quantitative data, violin plots are used to demonstrate data structure (i.e., median, interquartile range, and 95% confidence interval) and distribution. The x-axis labels each site-year, whereas the y-axis labels the observed values for the given trait (Figure 4A). Qualitative data is summarized with histograms which consist of a series for every site-year (Figure 4B). In both plot types, the phenotypic value for a given germplasm can be highlighted within the context of the larger dataset. This proves quite helpful in breeding programs to provide additional data for selections, highlight potential planting errors, and plan crosses.

Trait distribution plots can be accessed in a number of different ways. Plots are found on all associated trait, germplasm,



and experiment pages. There is also a tool which allows users to generate their own plots based on KnowPulse-housed data. This kind of integration ensures that the system is intuitive to all users. Context and summaries for the trait, experiment or germplasm being viewed is also provided.

Additionally, trait pages in KnowPulse contain an overview describing the trait, linking it to ontologies and describing the methodology used for data collection (Figure 5A). Experiments in which the traits were measured are listed, along with information on the number of associated site-years (Figure 5B).

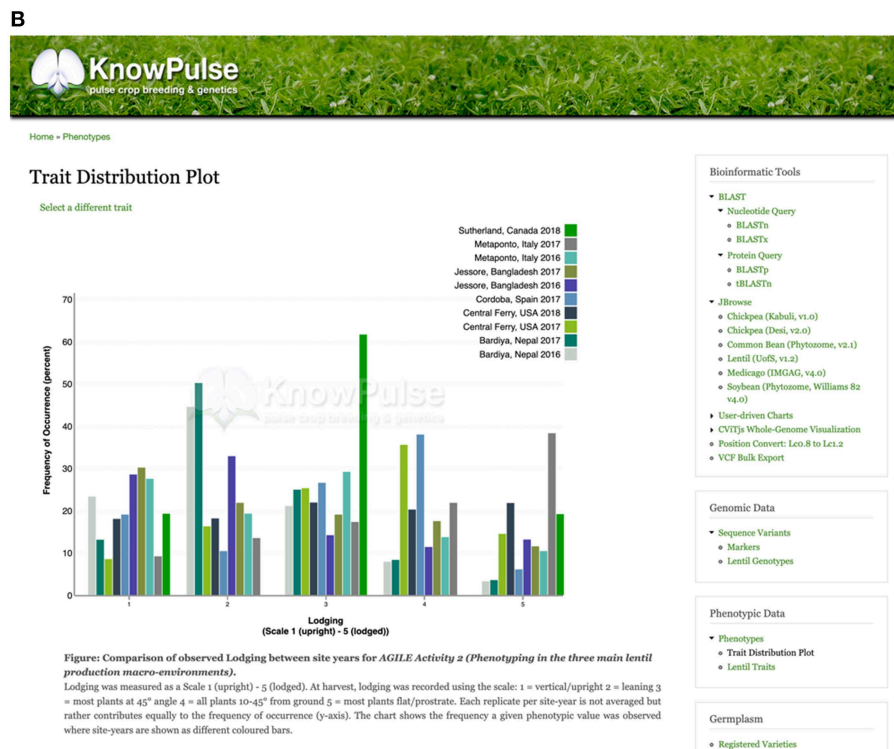
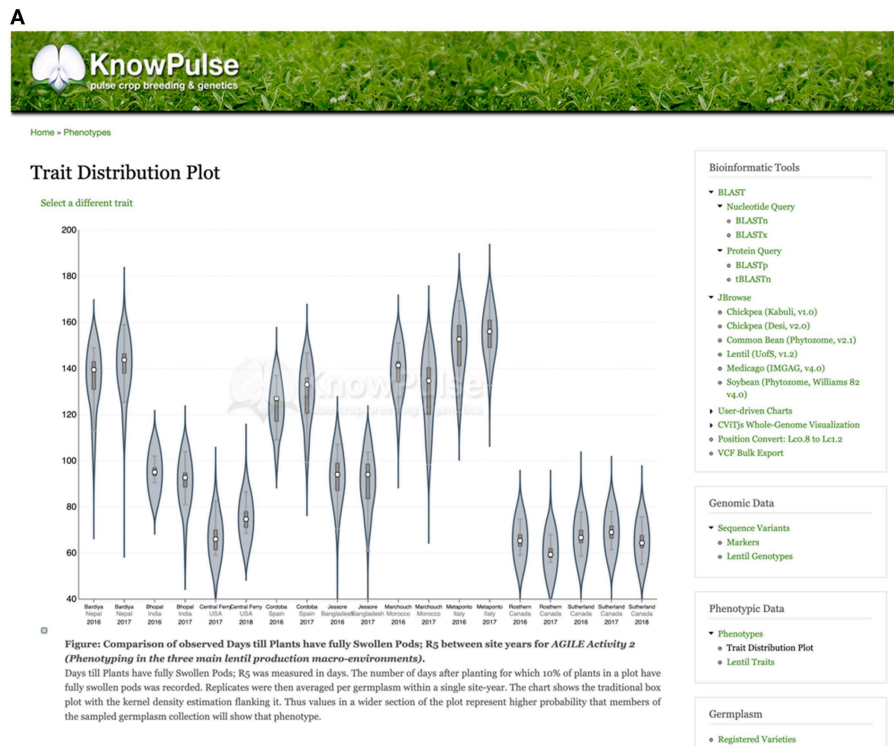
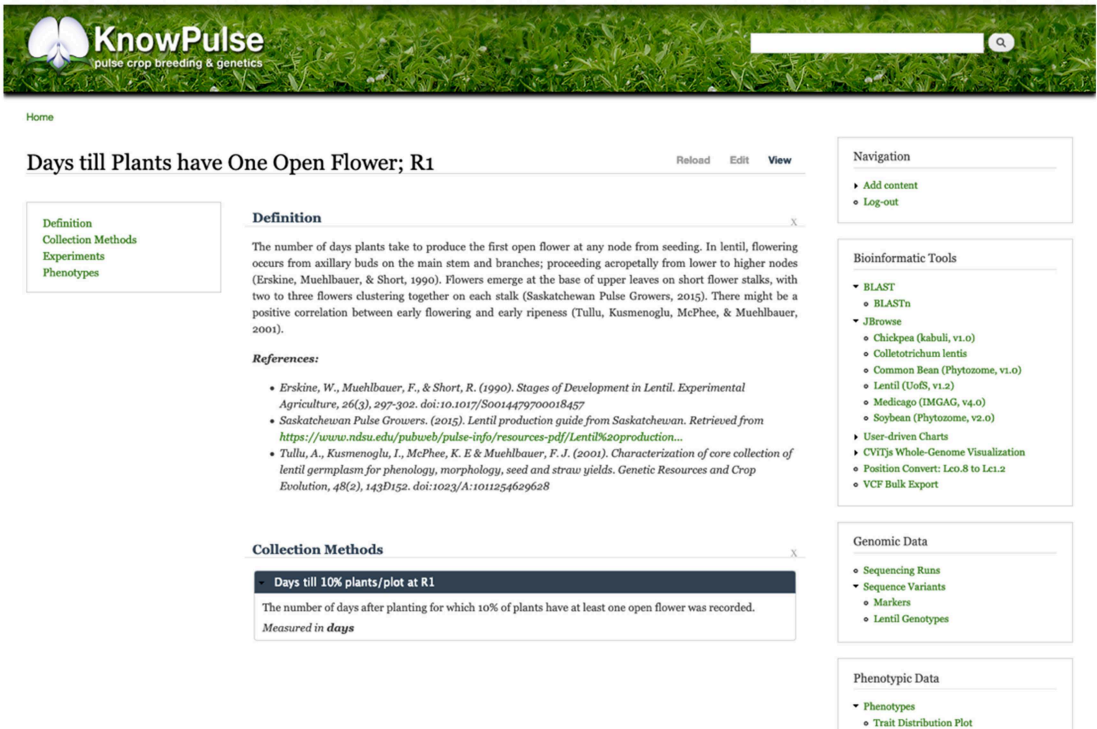


FIGURE 4 | Trait distribution plots summarizing phenotype data. Quantitative phenotypic data (A) is shown as a violin plot with site-year labeled by the x-axis and observed values labeled by the y-axis. This allows researchers to see the data structure (i.e., median, interquartile range, and 95% confidence interval) and distribution per site year. Qualitative phenotypic data (B) is shown as a multi-series histogram with each series representing a site-year and the observed phenotypes defined on the x-axis. The quantity of germplasm exhibiting each phenotype is shown on the y-axis allowing researchers to evaluate how prevalent a phenotype is in their population. These plots can be accessed via the trait distribution plot tool under Phenotypic data in the right side menu, as well as through trait, germplasm, and project pages with associated phenotypic data.

A



KnowPulse pulse crop breeding & genetics

Home

Days till Plants have One Open Flower; R1

Reload Edit View

Definition

Collection Methods

Experiments

Phenotypes

Definition

The number of days plants take to produce the first open flower at any node from seeding. In lentil, flowering occurs from axillary buds on the main stem and branches; proceeding acropetally from lower to higher nodes (Erskine, Muehlbauer, & Short, 1990). Flowers emerge at the base of upper leaves on short flower stalks, with two to three flowers clustering together on each stalk (Saskatchewan Pulse Growers, 2015). There might be a positive correlation between early flowering and early ripeness (Tullu, Kusmenoglu, McPhee, & Muehlbauer, 2001).

References:

- Erskine, W., Muehlbauer, F., & Short, R. (1990). *Stages of Development in Lentil. Experimental Agriculture*, 26(3), 297-302. doi:10.1017/S0014479700018457
- Saskatchewan Pulse Growers. (2015). *Lentil production guide from Saskatchewan*. Retrieved from <https://www.ndsu.edu/pubweb/pulse-info/resources-pdf/Lentil%20production...>
- Tullu, A., Kusmenoglu, I., McPhee, K. E & Muehlbauer, F. J. (2001). Characterization of core collection of lentil germplasm for phenology, morphology, seed and straw yields. *Genetic Resources and Crop Evolution*, 48(2), 143D152. doi:10.23/A:1011254629628

Collection Methods

Days till 10% plants/plot at R1

The number of days after planting for which 10% of plants have at least one open flower was recorded. Measured in *days*

Navigation

- Add content
- Log-out

Bioinformatic Tools

- BLAST
- BLASTn
- JBrowse
 - Chickpea (kabuli, v1.0)
 - Colletotrichum lentis
 - Common Bean (Phytozome, v1.0)
 - Lentil (UoS, v1.2)
 - Medicago (IMGAG, v4.0)
 - Soybean (Phytozome, v2.0)
- User-driven Charts
- CVTJs Whole-Genome Visualization
- Position Convert: Lc0.8 to Lc1.2
- VCF Bulk Export

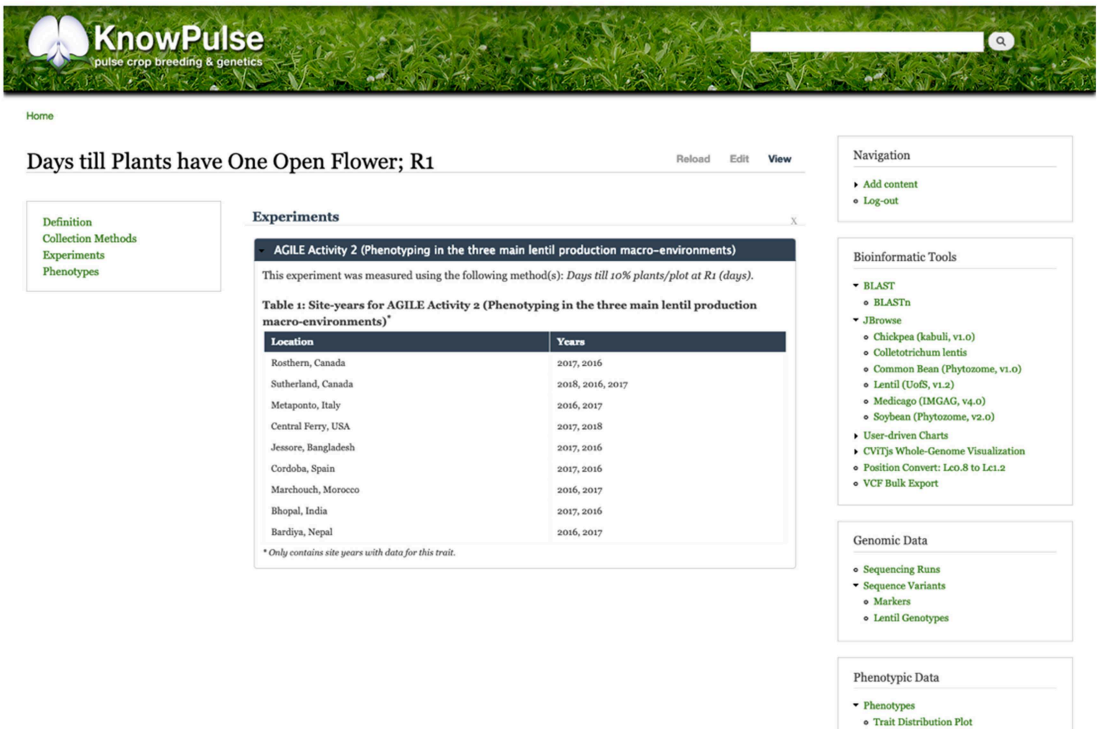
Genomic Data

- Sequencing Runs
- Sequence Variants
- Markers
- Lentil Genotypes

Phenotypic Data

- Phenotypes
- Trait Distribution Plot

B



KnowPulse pulse crop breeding & genetics

Home

Days till Plants have One Open Flower; R1

Reload Edit View

Definition

Collection Methods

Experiments

Phenotypes

Experiments

AGILE Activity 2 (Phenotyping in the three main lentil production macro-environments)

This experiment was measured using the following method(s): *Days till 10% plants/plot at R1 (days)*.

Table 1: Site-years for AGILE Activity 2 (Phenotyping in the three main lentil production macro-environments)*

Location	Years
Rosthern, Canada	2017, 2016
Sutherland, Canada	2018, 2016, 2017
Metaponto, Italy	2016, 2017
Central Ferry, USA	2017, 2018
Jessore, Bangladesh	2017, 2016
Cordoba, Spain	2017, 2016
Marchouch, Morocco	2016, 2017
Bhopal, India	2017, 2016
Bardiya, Nepal	2016, 2017

* Only contains site years with data for this trait.

Navigation

- Add content
- Log-out

Bioinformatic Tools

- BLAST
- BLASTn
- JBrowse
 - Chickpea (kabuli, v1.0)
 - Colletotrichum lentis
 - Common Bean (Phytozome, v1.0)
 - Lentil (UoS, v1.2)
 - Medicago (IMGAG, v4.0)
 - Soybean (Phytozome, v2.0)
- User-driven Charts
- CVTJs Whole-Genome Visualization
- Position Convert: Lc0.8 to Lc1.2
- VCF Bulk Export

Genomic Data

- Sequencing Runs
- Sequence Variants
- Markers
- Lentil Genotypes

Phenotypic Data

- Phenotypes
- Trait Distribution Plot

FIGURE 5 | KnowPulse trait pages. Each trait page describes the methods and units used for collecting the data (A) and lists all experiments the trait was measured in (B). This information provides context to help researchers better interpret the data. The trait distribution plot shown in Figure 4 is also available on each trait page. Trait pages can be accessed via the crop-specific trait search in the right side menu under phenotypic data.

Genome Canada [grant: 8302], Government of Saskatchewan [grant: 20150331], and the University of Saskatchewan.

ACKNOWLEDGMENTS

KnowPulse would not be where it is today without the continued development and maintenance of Tripal, and thus we extend a big thank you to the entire Tripal community and Dr.

Stephen Ficklin. KnowPulse is a member of the Legume Federation and, as such, we would like to thank them for their guidance and their collaborative efforts. We are also grateful for the support, guidance, and feedback from our community and colleagues in the Pulse group at the University of Saskatchewan. Some of the development work was done under AGILE, a Genome Canada funded project managed by Genome Prairie.

REFERENCES

- Boutet, G., Alves Carvalho, S., Falque, M., Peterlongo, P., Lhuillier, E., Bouchez, O., et al. (2016). SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genom.* 17:121. doi: 10.1186/s12864-016-2447-2
- Buels, R., Yao, E., Diesh, C. M., Hayes, R. D., Munoz-Torres, M., Helt, G., et al. (2016). JBrowse: a dynamic web platform for genome visualization and analysis. *Genome Biol.* 17:66. doi: 10.1186/s13059-016-0924-1
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+:architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Dash, S., Campbell, J. D., Cannon, E. K. S., Cleary, A. M., Huang, W., Kalberer, S. R., et al. (2015). Legume information system (LegumeInfo.org): a key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* 44, D1181–D1188. doi: 10.1093/nar/gkv1159
- Dash, S., Cannon, E. K. S., Kalberer, S. R., Farmer, A. D., and Cannon, S. B. (2016). “Chapter 8, PeanutBase and other bioinformatic resources for peanut,” in *Peanut Genetics, Processing, and Utilization*, eds H. T. Stalker and R. F. Wilson (Urbana, IL: AOCS Press), 241–252. doi: 10.1016/B978-1-63067-038-2.0008-3
- Gali, K. K., Liu, Y., Sindhu, A., Diapari, M., Shunmugam, A. S. K., Arganosa, G., et al. (2018). Construction of high-density linkage maps for mapping quantitative trait loci for multiple traits in field pea (*Pisum sativum* L.). *BMC Genomics.* 18, 1–25. doi: 10.1186/s12870-018-1368-4
- Graham, P. H., and Vance, C. P. (2003). Legumes: importance and constraints to greater use. *Plant Physiol.* 131, 872–877. doi: 10.1104/pp.017004
- Grant, D., Nelson, R. T., Cannon, S. B., and Shoemaker, R. C. (2009). SoyBase, the USDA-ARS soybean genetics and genomics database. *Nucleic Acid Res.* 38, D843–D846. doi: 10.1093/nar/gkp798
- Kamfwa, K., Cichy, K. A., and Kelly, J. D. (2015). Genome-wide association study of agronomic traits in common bean. *Plant Genome* 8, 1–12. doi: 10.3835/plantgenome2014.09.0059
- Krishnakumar, V., Kim, M., Rosen, B. D., Karamycheva, S., Bidwell, S. L., Tang, H., et al. (2014). MTGD: the medicago truncatula genome database. *Plant Cell Physiol.* 56:e1. doi: 10.1093/pcp/pcu179
- Moghaddam, S. M., Mamidi, S., Osorno, J. M., Lee, R., Brick, M., Kelly, J., et al. (2016). Genome-wide association study identifies candidate loci underlying agronomic traits in a middle American diversity panel of common bean. *Plant Genome* 9, 1–21. doi: 10.3835/plantgenome2016.02.0012
- Mungall, C. J., and Emmert, D. B. (2007). A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics* 23, i337–i346. doi: 10.1093/bioinformatics/btm189
- Ogutcen, E., Ramsay, L., von Wettberg, E. B., and Bett, K. E. (2018). Capturing variation in *Lens* (Fabaceae): development and utility of an exome capture array for lentil. *Appl. Plant Sci.* 6:e1165. doi: 10.1002/aps3.1165
- O'Rourke, J. A., Bolon, Y. T., Bucciarelli, B., and Vance, C. P. (2014). Legume genomics: understanding biology through DNA and RNA sequencing. *Ann. Bot.* 113, 1107–1120. doi: 10.1093/aob/mcu072
- Pandey, M. K., Roorkiwal, M., Singh, V. K., Ramalingam, A., Kudapa, H., Thudi, M., et al. (2016). Emerging genomic tools for legume breeding: current status and future prospects. *Front. Plant Sci.* 7:455. doi: 10.3389/fpls.2016.00455
- Parween, S., Nawaz, K., Roy, R., Pole, A. K., Venkata, S. B., Misra, G., et al. (2015). An advanced draft genome assembly of a desi type chickpea (*Cicer arietinum* L.). *Sci Rep.* 5:12806. doi: 10.1038/srep12806
- Polak, R., Phillips, E. M., and Campbell, A. (2015). Legumes: health benefits and culinary approaches to increase intake. *Clin Diabetes.* 33, 198–205. doi: 10.2337/diaclin.33.4.198
- Ramsay, L., Sharpe, A. G., Cook, D. R., Penmetsa, R. V., Gujaria-Verma, N., Vandenberg, A., et al. (2014). “Draft genome assembly and survey of genetic diversity within *Lens culinaris*,” in *International Plant and Animal Genome Conference XXII* (San Diego, CA).
- Sanderson, L. A., Ficklin, S. P., Cheng, C. H., Jung, S., Feltus, F. A., Bett, K. E., et al. (2013). Tripal v1.1: a standards-based toolkit for construction of online genetic and genomic databases. *Database* 2013:bat075. doi: 10.1093/database/bat075
- Schmutz, J., Cannon, S. B., Schlueter, J., Ma, J., Mitros, T., Nelson, W., et al. (2010). Genome sequence of the palaeopolyploid soybean. *Nature* 463, 178–183. doi: 10.1038/nature08670
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., et al. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* 46, 707–713. doi: 10.1038/ng.3008
- Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., et al. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genom.* 15:312. doi: 10.1186/1471-2164-15-312
- Tharanathan, R. N., and Mahadevamma, S. (2003). Grain legumes—a boon to human nutrition. *Trends Food Sci. Technol.* 14, 507–518. doi: 10.1016/j.tifs.2003.07.002
- Varshney, R. K., Chen, W., Li, Y., Bharti, A. K., Saxena, R. K., Schlueter, J. A., et al. (2012). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* 30, 83–89. doi: 10.1038/nbt.2022
- Varshney, R. K., Close, T. J., Singh, N. K., Hoisington, D. A., and Cook, D. R. (2009). Orphan legume crops enter the genomics era! *Curr. Opin. Plant Biol.* 12, 202–210. doi: 10.1016/j.pbi.2008.12.004
- Varshney, R. K., Song, C., Saxena, R. K., Azam, S., Yu, S., Sharpe, A. G., et al. (2013). Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. *Nat. Biotechnol.* 31, 240–246. doi: 10.1038/nbt.2491

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Sanderson, Caron, Tan, Shen, Liu and Bett. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.