

2010

The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps

Carson M. Andorf

United States Department of Agriculture, carson.andorf@gmail.com

Carolyn J. Lawrence

United States Department of Agriculture

Lisa C. Harper

United States Department of Agriculture

Mary L. Schaeffer

United States Department of Agriculture

Follow this and additional works at: http://lib.dr.iastate.edu/gdcb_las_pubs



Part of the [Agriculture Commons](#), [Bioinformatics Commons](#), [Computational Biology Commons](#), [United States Department of Agriculture Commons](#), [Genetics Commons](#), and the [Plant Breeding and Genetics Commons](#)

See next page for additional authors.

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/gdcb_las_pubs/31. For information on how to cite this item, please visit <http://lib.dr.iastate.edu/howtocite.html>.

The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps

Abstract

Methods to automatically integrate sequence information with physical and genetic maps are scarce. The Locus Lookup tool enables researchers to define windows of genomic sequence likely to contain loci of interest where only genetic or physical mapping associations are reported. Using the Locus Lookup tool, researchers will be able to locate specific genes more efficiently that will ultimately help them develop a better maize plant. With the availability of the well-documented source code, the tool can be easily adapted to other biological systems.

Disciplines

Agriculture | Bioinformatics | Computational Biology | Genetics | Plant Breeding and Genetics

Comments

This article is from *Bioinformatics* 26 (2010): 434–436, doi:[10.1093/bioinformatics/btp5](https://doi.org/10.1093/bioinformatics/btp5).

Rights

Works produced by employees of the U.S. Government as part of their official duties are not copyrighted within the U.S. The content of this document is not copyrighted.

Authors

Carson M. Andorf, Carolyn J. Lawrence, Lisa C. Harper, Mary L. Schaeffer, Darwin A. Campbell, and Taner Z. Sen

Databases and ontologies

The Locus Lookup tool at MaizeGDB: identification of genomic regions in maize by integrating sequence information with physical and genetic maps

Carson M. Andorf¹, Carolyn J. Lawrence^{1,2}, Lisa C. Harper^{3,4}, Mary L. Schaeffer^{5,6}, Darwin A. Campbell¹ and Taner Z. Sen^{1,2,*}

¹US Department of Agriculture – Agricultural Research Service, Corn Insects and Crop Genetics Research Unit,

²Department of Genetics, Development and Cell Biology; Bioinformatics and Computational Biology Program, Iowa

State University, Ames, IA 50011, ³US Department of Agriculture – Agricultural Research Service, Plant Gene

Expression Center, 800 Buchanan Street, Albany, CA 94710, ⁴Department of Molecular and Biology, University of

California Berkeley, Berkeley, CA 94720, ⁵US Department of Agriculture – Agricultural Research Service, Plant Genetics Research Unit and ⁶Division of Plant Sciences, University of Missouri Columbia, Columbia, MO 65211, USA

Received and revised on August 7, 2009; accepted on September 11, 2009

Associate Editor: Alex Bateman

ABSTRACT

Summary: Methods to automatically integrate sequence information with physical and genetic maps are scarce. The Locus Lookup tool enables researchers to define windows of genomic sequence likely to contain loci of interest where only genetic or physical mapping associations are reported. Using the Locus Lookup tool, researchers will be able to locate specific genes more efficiently that will ultimately help them develop a better maize plant. With the availability of the well-documented source code, the tool can be easily adapted to other biological systems.

Availability: The Locus Lookup tool is available on the web at http://maizegdb.org/cgi-bin/locus_lookup.cgi. It is implemented in PHP, Oracle and Apache, with all major browsers supported. Source code is freely available for download at http://ftp.maizegdb.org/open_source/locus_lookup/.

Contact: taner.sen@ars.usda.gov

Maize (*Zea mays* ssp. *mays*) has been an important model organism for nearly a century due to its large chromosomes, ease of making genetic crosses, and economic importance. Maize researchers have created many genetic recombination maps, and more than 1700 are available online via MaizeGDB, the maize model organism database (Lawrence *et al.*, 2006, 2007; Sen *et al.*, 2009). Other maize maps include physical, cytological [reviewed in Lawrence *et al.* (2006), and optical maps (Aston *et al.*, 1999; Zhou *et al.*, 2009)]. In addition to these maps, maize researchers now have access to genomic sequence of the B73 inbred line's genic regions (Schnable *et al.*, 2009) and pseudomolecules representing the sequenced regions of the genome (Wei *et al.*, 2009).

The availability of the B73 reference genome sequence opens up new research possibilities. Investigators can more easily narrow the genomic regions responsible for specific phenotypes with higher accuracy in a shorter time frame: 'walking to genes' has become

'running to genes'. Researchers can dissect gene structure and function faster, and more clearly define the relationship between genotype and phenotype.

Although sequence data are becoming increasingly available [especially with the advent of next-generation sequencing technologies (Simon *et al.*, 2009)], the bioinformatic tools needed to integrate sequence with existing map information are insufficient, not only for maize but for many other research model species. Although these datatypes reside side-by-side in databases, integrating the information in a meaningful way is not always straightforward. In the absence of expensive manual curation by database personnel, researchers often resort to copying and pasting rows of data for loci of interest into spreadsheets and analyzing and curating the data by hand.

To overcome some aspects of these challenges, the Locus Lookup tool was developed and deployed within the context of the GBrowse-based MaizeGDB Genome Browser (Sen *et al.*, 2009; Stein *et al.*, 2002). In overview, the Locus Lookup tool takes the name(s) of (i) a single locus or (ii) two loci that define a region and returns a snapshot representing the likely genomic region containing the locus of interest.

Here is how the Locus Lookup tool works. When a single search term is entered, the Locus Lookup tool (i) checks whether the locus/loci is/are already mapped to the B73 sequence, and if so, displays the genomic/nucleotide coordinates on the MaizeGDB Genome Browser. If not, then (ii) it checks whether a probe/molecular marker (e.g. a Bacterial Artificial Chromosome-BAC) that recognizes/contains that locus has known genomic coordinates. If no such probes/molecular markers are returned, then the Locus Lookup tool (iii) checks a user-specified genetic map for the nearest left and right features/markers neighboring the locus that are also placed on the B73 sequence in genomic coordinates. If such left and right features can be found, the Locus Lookup will return the coordinates between which the gene of interest is likely to reside. Note also that because the B73 sequence is BAC-based and BAC

*To whom correspondence should be addressed.

Chromosome coordinates based on MaizeSequence.org marker positions for: et1

The exact coordinates for *et1* were not found. The coordinates will be estimated based on the nearest flanking loci with known coordinates.

The Locus *et1* is flanked by the following loci with known chromosome coordinates: (*npi425a* and *bnlg1496*)

The Locus *et1* is **estimated** to be between position **214,541,600** and position **214,683,700** on **Chromosome 3** based on the map: **IBM2 2008 Neighbors 3**

This region is **142,100** base pairs long.

(Click [here](#) to go to the Genome Browser or scroll to the bottom of the page)

Locus	Chr	Map Position	Chr Start	Chr End
<i>npi425a</i>	3	759.9	214,541,600	214,546,500
<i>et1</i>	3	760.02	0	0
<i>bnlg1496</i>	3	760.9	214,678,800	214,683,700

Loci known to be between the flanking loci at map positions **759.9** and **760.9** on the **IBM2 2008 Neighbors map**:

▶ [Click to expand details](#)

The Locus Lookup tool works by (1) checking physical map coordinates to find out whether the locus is already placed. If so, your physically mapped locus is highlighted in red in the region returned. If not, the tool (2) checks the locus record at MaizeGDB to find out if any BACs are known to detect the locus and that BAC is returned within its genomic context. If not, (3) **genetically mapped probes that are nearest the input locus are identified, the tool checks whether those probes have known genomic coordinates (working outward until appropriate probes are identified) and finally the region of the genome contained by the identified probes is reported with bounding probes shown in red.**

[More information](#)

Your search for *et1* got a result with the logic described in (3) above.

The Locus *et1* is **estimated** between position **214,497,500** and position **214,884,600** on **Chromosome 3** located on one or more of the following BACs: [AC197543](#) and [AC208915](#)

This region is **387,100** base pairs long.

(Click on image to go to the Genome Browser)

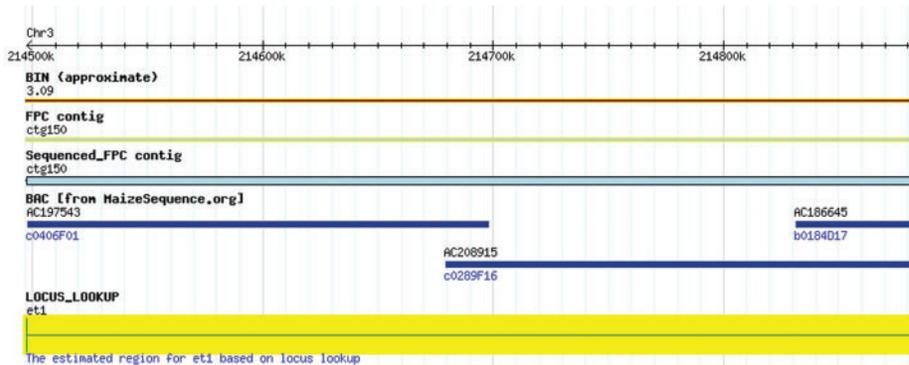


Fig. 1. The Locus Lookup view of the *etched 1* (*et1*) locus at MaizeGDB. At the top of the page, the estimated coordinates and the markers of known coordinates that are used for estimation are shown. In the middle of the page, a box with a bolded section specifies which method is used for estimation. At the bottom of the page, a view of the genomic region is shown. When the users click this view, they are lead to the MaizeGDB Genome Browser (with a custom Locus Lookup track), which enables them to harness the full capability of the MaizeGDB Genome Browser to address their research problems.

sequences consist of scaffolds as well as unordered and unoriented pieces, the returned snapshot contains the entire span of the left-most and right-most BACs in the region in case the locus of interest falls outside the window due to possible fragmentation of those BACs.

Because the precision of Locus Lookup tool lies in the accuracy of both genetic coordinates on the selected genetic map and genomic coordinates on the B73 sequence, the imperfections in both are inherited. For the maize genetic maps, in particular consensus maps

such as Neighbors and the Genetic 2008 that are composites of many different maps and experiments, genetic order and distance are only approximated. In addition, among different maize cultivars or lines, the order of certain genes along a chromosome varies in some cases (Fu and Dooner, 2002). Some of the current limitations (as of July, 2009) of the maize B73 sequence are: (i) it is not yet known how much sequence exists between physically mapped BAC contigs; (ii) the sequence of each BAC consists of scaffolds, as well as unordered

and unoriented pieces separated by strings of 'N's; (iii) the order and/or orientation of BACs within a contig on the physical map may be wrong; and (iv) the order or orientation of sequence contigs may be wrong. In addition, as of this writing, the current version of the genome assembly (MaizeSequence.org's 3b.50 release) consists of 16 581 sequenced BACs, 1024 of which (i.e. 6%) are not assigned to a linkage group (chromosome). Sequences not assigned to a chromosome will be missed by the Locus Lookup tool.

The Locus Lookup tool at MaizeGDB can be reached a few different ways: (i) from the MaizeGDB homepage (<http://www.maizegdb.org>) in the left green margin, (ii) via selecting 'genome browser' in the dropdown to the left of the search box displayed at the top and bottom of every MaizeGDB page, and (iii) from the search box on the MaizeGDB Genome Browser (<http://gbrowse.maizegdb.org>). To do a Locus Lookup search, researchers enter the search term (locus name) and click on the search button. In the results page either the genomic coordinates are shown as a snapshot of the MaizeGDB Genome Browser or, if the position is based on flanking features, a genomic region will be specified and presented [see the example of *etched1* (*et1*) locus in Fig. 1]. Clicking the snapshot adds a custom track to the Genome Browser allowing the region to be viewed in its genomic context. In addition, if a genomic region bounded by two loci is desired, researchers can type 'locus1..locus2' into the search box to define the region bounded by the two loci using the same logic as above.

The logic and implementation for the Maize Locus Lookup tool are not specific to maize and generally applicable to any research model organism and other model organism database groups. For example, SoyBase (Grant *et al.*, 2009) and SGN (Mueller *et al.*, 2005) are already planning to implement the Locus Lookup tool's logic for the soybean and Solanaceae research communities, respectively.

ACKNOWLEDGEMENTS

The authors thank the maize community for their enthusiastic support and feedback, the MaizeGDB Working Group for their guidance, and the Maize Genome Sequencing Consortium for releasing the B73 genome as it is being sequenced.

Funding: US Department of Agriculture, Agricultural Research Service.

Conflict of Interest: none declared.

REFERENCES

- Aston,C. *et al.* (1999) Optical mapping and its potential for large-scale sequencing projects. *Trends Biotechnol.*, **17**, 297–302.
- Fu,H. and Dooner,H.K. (2002) Intraspecific violation of genetic colinearity and its implications in maize. *Proc. Natl Acad. Sci. USA*, **99**, 9573–9578.
- Grant,D. *et al.* (2009) SoyBase, The USDA-ARS Soybean Genome Database. Available at <http://soybase.org> (last accessed date November 19, 2009).
- Lawrence,C.J. *et al.* (2006) Predicting chromosomal locations of genetically mapped loci in maize using the Morgan2McClintock Translator. *Genetics*, **172**, 2007–2009.
- Lawrence,C.J. *et al.* (2007) MaizeGDB's new data types, resources and activities. *Nucleic Acids Res.*, **35**, D895–D900.
- Mueller,L.A. *et al.* (2005) The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. *Plant Physiol.*, **138**, 1310–1317.
- Schnable,P.S. *et al.* (2009) The B73 maize genome: complexity, diversity and dynamics. *Science*, **326**, 1112–1115.
- Sen,T.Z. *et al.* (2009) MaizeGDB becomes 'sequence-centric'. *Database* [Epub ahead of print, doi:10.1093/database/bap020]
- Simon,S.A. *et al.* (2009) Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.*, **60**, 305–333.
- Stein,L.D. *et al.* (2002) The generic genome browser: a building block for a model organism system database. *Genome Res.*, **12**, 1599–1610.
- Wei,F. *et al.* (2009) The physical and genetic framework of the maize B73 genome. *PLoS Genet.*, **5**, e1000715.
- Zhou,S. *et al.* (2009) B73 Optical map: a single molecule map of the maize genome, Maize Genetics Coop. *Newsletter*, **83**, 106–107.