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

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

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. 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
The Locus Lookup tool at MaizeGDB

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 SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol., 138, 1310–1317. 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