

BARLEX – the Barley Draft Genome Explorer

Dear Editor,

Genome browsers visualize the end product of genome assembly, which is a highly contiguous sequence. However, how to visualize the intermediate products of genome sequencing? Next-generation sequencing has enabled genome sequencing in species with huge genomes, but most often the shotgun assemblies obtained from short-read sequence data do not meet the quality standards required for finished reference sequences. In particular, many plant genomes do not yet have finished reference sequences, but are instead represented by a developing genomic infrastructure built around incomplete draft assemblies and physical or genetic maps. For some applications, such as the comparison of genome maps with sequence resources, it may be desirable that these disparate resources are accessible in a single browser because generic genome browsers such as Gbrowse (Stein et al., 2002) can often not easily accommodate the multi-layered, incompletely ordered structure of plant draft genomes. To our knowledge, there is no single generic database system to integrate the heterogeneous data gathered in a plant genome sequencing project into one browsable interface.

We have developed BARLEX, a web-based application to access the developing genomic infrastructure of barley. Barley is a cereal grass of agronomical importance. The large genome size (5 Gb) and its high content of repetitive DNA have long obstructed the creation of a reference genome sequence of barley. In recent years, high-throughput technologies have expedited the construction of a physical map (Ariyadasa et al., 2014), a whole-genome shotgun assembly (International Barley Genome Sequencing Consortium, 2012), and an ultra-dense genetic map (Mascher et al., 2013a). BARLEX was developed in response to requests from members of the barley research community wishing to access and compare these comprehensive datasets (Supplemental Table 1).

BARLEX is centered on the genome-wide physical map of barley, which has been integrated with sequence assemblies and genetic maps (International Barley Genome Sequencing Consortium, 2012; Ariyadasa et al., 2014). The BAC identifiers and overlap data between BACs were imported into an Oracle relational database (Supplemental Methods and Supplemental Figures 1 and 2). User access to this database is provided through a web application developed with the Oracle Application Express (APEX) framework (<http://apex.oracle.com>). The BARLEX database is freely accessible at <http://barlex.barleysequence.org>. Moreover, we implemented a novel visualization strategy for interactive exploration of the information contained in the database. Each fingerprinted contig is represented as a graph structure, where BACs are the nodes of the graph and an edge between two BACs is present if they are overlapping within an FP contig (Figure 1A). The graph structure is drawn dynamically with Cytoscape Web (Lopes et al., 2010). Textual information for all clones in an FP contig can be accessed in a table below the graphs. On hovering with the mouse pointer over a node or an

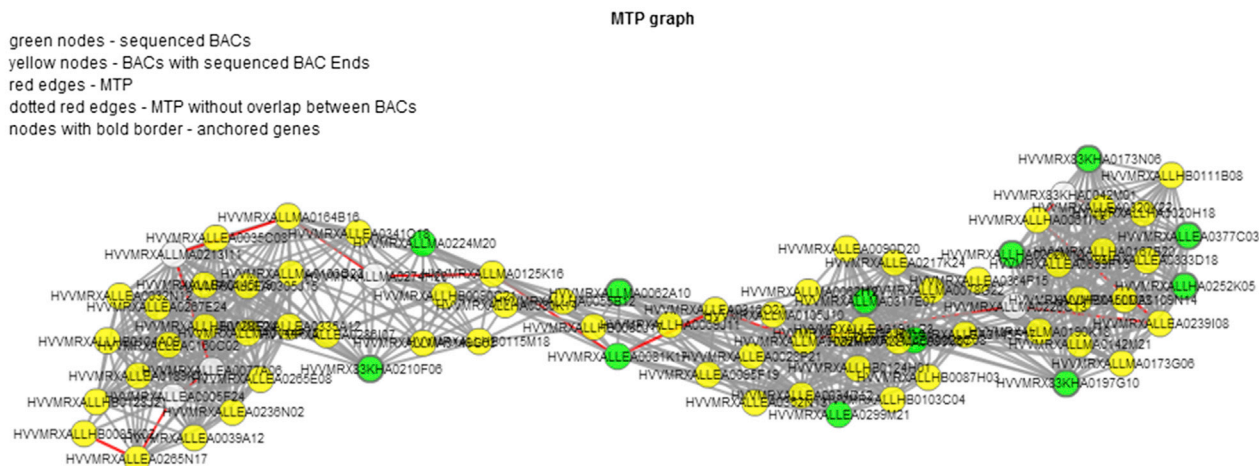
edge, summary information is shown as a tooltip. Left-clicks on nodes and edges open web pages with detailed information.

We used the BAC survey sequence information associated with the physical map to link the clones to the annotated whole-genome shotgun assembly of barley. For this purpose, BLAST searches of WGS contigs against sequence contigs of individual BAC or BAC end sequences were performed as described earlier (Ariyadasa et al., 2014). The visualization of the alignment results is centered on a single BAC. All BAC sequences, WGS contigs, and genes assigned to a clone are grouped around it (Figure 1B). The nodes of this graph are also clickable and direct the user to pages with information specific to the various elements. The pages of single-sequence BAC contigs contain a Kmasker plot (<http://webblast.ipk-gatersleben.de/kmasker/>) showing the distribution of k-mer frequencies to highlight low copy regions. The pages of WGS contigs (Figure 1C) contain genetic anchoring data as well as the positions of annotated genes (International Barley Genome Sequencing Consortium, 2012), exome capture targets (Mascher et al., 2013b), and markers on the 9K iSelect chip (Comadran et al., 2012). BARLEX also incorporates expression profiles of all genes across eight developmental stages (International Barley Genome Sequencing Consortium, 2012).

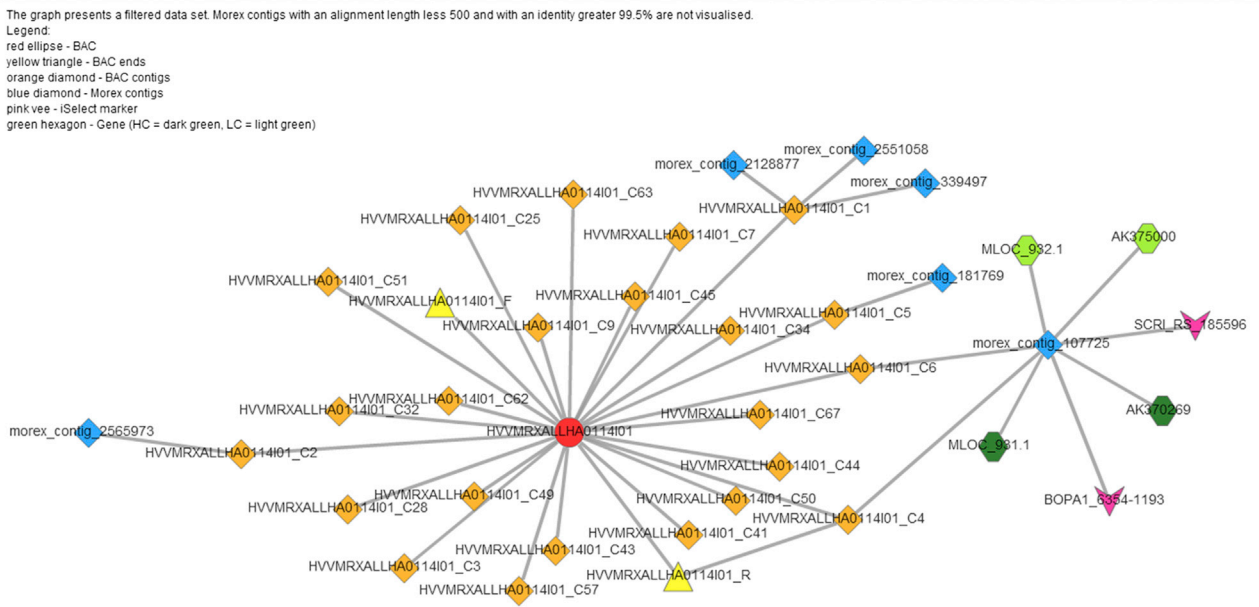
We have set up several entry points to BARLEX. FP contigs and BACs can be looked up with their names or their genetic positions. Likewise, a table of annotated genes can be searched with gene identifiers or the names of the WGS contig carrying them and is directly linked to the BAC information. All BLAST results underlying the links between BACs and contigs are also accessible as searchable tables with hyperlinks to information pages. All sequence data used to perform the alignments are publicly available (Supplemental Table 1). Entry points for sequences that are not immediately associated with the barley genome infrastructure can be found through BLAST searches submitted via a form available on the BARLEX homepage. Currently, the most widely used genotyping platform in barley is a 9K Illumina iSelect SNP chip (Comadran et al., 2012). We have linked all marker sequences of this platform to the BARLEX data. Finally, a portlet on the start page connects BARLEX to LAILAPS, a search engine engineered toward information retrieval in the life sciences (<http://lailaps.ipk-gatersleben.de/>).

So far, sequence information directly associated with the physical map of barley is scarce. While the entire MTP consists of 66 772 clones, currently full sequence information is publicly available for only 6278 BACs. As MTP sequencing comes along (International Barley Genome Sequencing Consortium, unpublished results), we will populate BARLEX with additional BAC sequence and overlap data. The new sequence data will be useful

A Network (MTP)



B Anchored Contigs/Genes



C Exome Capture Target

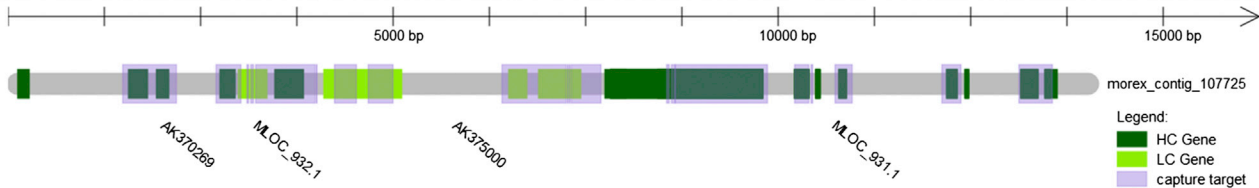


Figure 1. Selected Features of the BARLEX Web Interface.
(A) Graph-based visualization of the minimum tiling path of a fingerprinted contig.
(B) Overview of resources available for each sequenced BAC.
(C) Genes annotated on the WGS assembly of barley and exome capture target regions.

to validate the overlap information provided by the physical map. In analogy to the concept of fingerprinted contigs, we can define sequence overlap clusters. Two BACs located in the same cluster either directly overlap or there exists a path of overlapping BACs that connect the two BACs. Two previously disconnected FP contigs may then be joined if their clones at the ends of both contigs are

overlapping under stringent alignment criteria (99.8% identity, 5000 bp alignment length, see Supplemental Methods). To illustrate the detection of BAC overlaps on real data, along with the BAC data gathered in frame of the barley genome project, we also imported 17 BACs sequenced as part of a positional cloning project. Yang et al. (2014) found two FP contigs that harbored

genetic markers flanking a gene locus that controls susceptibility to barley yellow mosaic virus and sequenced a manually selected minimum tiling path between the flanking markers. Manual inspection of the assembled sequences of BACs confirmed the sequence overlap between the two contigs and enabled Yang et al. (2014) to pinpoint a candidate gene. Our automated overlap analysis was able to reproduce these results (Supplemental Figure 3 and Supplemental Text 1).

We will maintain BARLEX and increase its functionality as sequence assemblies of all clones in the MTP become available. In particular, the sequence-based validation of fingerprint-based clone overlaps will be an important tool. As the barley genome is highly repetitive, care needs to be taken to avoid erroneously joining BACs based on highly similar, but unrelated copies of transposable elements. Once the assemblies of all MTP clones have been completed, all individual BAC assemblies need to be integrated into a finished reference sequence. Constructing non-redundant pseudo molecules from more than 70 000 BAC assemblies is an endeavor that cannot be undertaken by manual curators. Automated procedures are needed to integrate all available sequence resources to create a contiguous sequence of the barley genome. The development of a customized pipeline for detecting and merging overlapping BACs, ordering merged BAC sequences to form larger supercontigs, and spotting potential mis-assembly or problematic clones can be assisted by manual inspection of exemplary cases in BARLEX. Clone overlaps detected in an automated manner by BLAST alignment can be examined in the BARLEX web interface.

BARLEX is currently specific to the genomic infrastructure of barley. However, none of the data sets incorporated into BARLEX are specific to barley. Similar sequence and mapping resources have recently become available for hexaploid bread wheat (Choulet et al., 2014) and the wild relatives of rice (Wang et al., 2014). A central repository very much like BARLEX could provide an easy-to-use presentation and interactive resource to facilitate genomic research in the grasses.

SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

FUNDING

Barley genome sequencing was supported by the German Ministry of Education and Research (BMBF) in frame of the grants BARLEX (FKZ0314000) and TRITEX (0315954A) to N.S. and U.S. and by a grant from the Leibniz Association (Pakt für Forschung und Innovation) to N.S. and U.S.

ACKNOWLEDGMENTS

We thank Steffen Flemming, Markus Oppermann, and Matthias Lange for maintaining the database as well as the application server. Thomas Münch and Heiko Mieke are acknowledged for administration of our UNIX servers. Finally, we thank Danuta Schüler and Doreen Stengel for managing sequence data in the IPK laboratory information management system. No conflict of interest declared.

Received: November 6, 2014
Revised: February 11, 2015
Accepted: March 18, 2015
Published: March 24, 2015

*Christian Colmsee, Sebastian Beier,
Axel Himmelbach, Thomas Schmutzer,
Nils Stein, Uwe Scholz and Martin Mascher**

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben,
Corrensstraße 3, D-06466 Stadt Seeland, Germany

*Correspondence: Martin Mascher (mascher@ipk-gatersleben.de)
<http://dx.doi.org/10.1016/j.molp.2015.03.009>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

REFERENCES

- Ariyadasa, R., Mascher, M., Nussbaumer, T., Schulte, D., Frenkel, Z., Poursarebani, N., Zhou, R., Steuernagel, B., Gundlach, H., Taudien, S., et al. (2014). A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. *Plant Physiol.* **164**:412–423.
- Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., et al. (2014). Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345**:1249721.
- Comadran, J., Kilian, B., Russell, J., Ramsay, L., Stein, N., Ganai, M., Shaw, P., Bayer, M., Thomas, W., Marshall, D., et al. (2012). Natural variation in a homolog of *Antirrhinum* *CENTRORADIALIS* contributed to spring growth habit and environmental adaptation in cultivated barley. *Nat. Genet.* **44**:1388–1392.
- International Barley Genome Sequencing Consortium (2012). A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**:711–716.
- Lopes, C.T., Franz, M., Kazi, F., Donaldson, S.L., Morris, Q., and Bader, G.D. (2010). Cytoscape web: an interactive web-based network browser. *Bioinformatics* **26**:2347–2348.
- Mascher, M., Muehlbauer, G.J., Rokhsar, D.S., Chapman, J., Schmutz, J., Barry, K., Muñoz-Amatriáin, M., Close, T.J., Wise, R.P., Schulman, A.H., et al. (2013a). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* **76**:718–727.
- Mascher, M., Richmond, T.A., Gerhardt, D.J., Himmelbach, A., Clissold, L., Sampath, D., Ayling, S., Steuernagel, B., Pfeifer, M., D'Ascenzo, M., et al. (2013b). Barley whole exome capture: a tool for genomic research in the genus *Hordeum* and beyond. *Plant J.* **76**:494–505.
- Stein, L.D., Mungall, C., Shu, S., Caudy, M., Mangone, M., Day, A., Nickerson, E., Stajich, J.E., Harris, T.W., Arva, A., et al. (2002). The generic genome browser: a building block for a model organism system database. *Genome Res.* **12**:1599–1610.
- Wang, X., Kudrna, D.A., Pan, Y., Wang, H., Liu, L., Lin, H., Zhang, J., Song, X., Goicoechea, J.L., Wing, R.A., et al. (2014). Global genomic diversity of *Oryza sativa* varieties revealed by comparative physical mapping. *Genetics* **196**:937–949.
- Yang, P., Lüpken, T., Habekuss, A., Hensel, G., Steuernagel, B., Kilian, B., Ariyadasa, R., Himmelbach, A., Kumlehn, J., Scholz, U., et al. (2014). PROTEIN DISULFIDE ISOMERASE LIKE 5-1 is a susceptibility factor to plant viruses. *Proc. Natl. Acad. Sci. USA* **111**:2104–2109.