A single gene knock-out resource for maize

Charles Du[1], Wenwei Xiong[1], Yubin Li[2], Harrison Wang[2], Mithu Chatterjee[2], Limei He[2], Qinghua Wang[2], Jun Huang[2], Gregorio Segal[2], and Hugo K Dooner[2]

[1]Department of Biology, Montclair State University, Montclair, NJ 07043
[2]Waksman Institute of Microbiology, Rutgers University, Piscataway, NJ, 08854

The availability of a mutant line in which a single gene has been disrupted gives biologists a powerful tool in understanding that gene's action. Thus, sequence-indexed collections of single insertions are critical resources for elucidating gene function in organisms with a sequenced genome. Our NSF-PGRP-funded project to generate such a resource (Li et al., MNL 87: 20-21, 2014) has delivered three substantial outcomes.

Launching platforms

Almost 200 lines carrying a transpositionally active, marked *Ds-GFP (Ds*)* element have been generated and the location of 86 launching platforms has been mapped to all 20 chromosome arms of the maize genome (Figure 1). The red marks on the chromosomes identify the location of the mapped T-DNA platforms; the white marks, the location of the bin core markers. The space between two core markers corresponds to roughly 20 cM in the genetic map. Over 600 *Ds* transpositions were produced from each of several platforms scattered across the genome. Around 90% were confirmed as concordant germinal events by testcrosses, and mapped relative to the original platform.

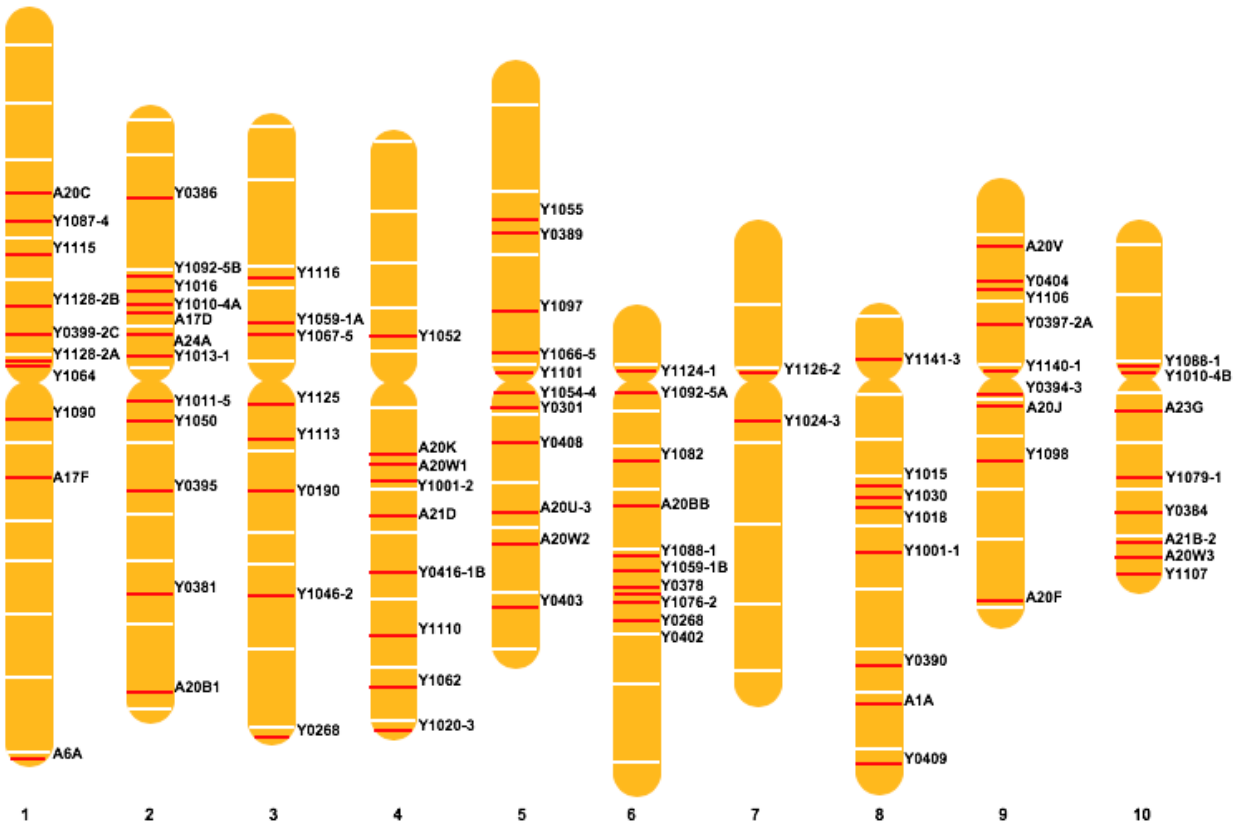

Figure 1. Location of 86 T-DNA platforms currently mapped in the genome. All 20 maize chromosome arms are represented among the 86 mapped platforms.

Collection of Sequence-indexed Ds transposon insertions*

A set of 16,000 *transposed Ds** (*trDs**) elements was selected for mapping to the reference genome using a strategy that takes advantage of unique sequences in the *Ds** element to specifically amplify maize sequences adjacent to the *trDs** element (*dsg* sites), thereby avoiding amplification of junctions from endogenous elements. The amplified *Ds** junctions were sequenced in 3-D pools or "cubes" of 960 individuals (arranged in ten 96-well plates, i.e., 96 samples per plate pool x 120 samples per row pool x 80 samples per column pool) in an Illumina MiSeq high-throughput machine in the Waksman Institute sequencing facility and mapped to the B73 reference genome using software, InsertionMapper, developed specifically for the project (Figure 2). Of the first 16000 *dsg* sequences, 14184 (~90%) have been unequivocally indexed to a specific cell in the 3-D pool and mapped to the maize genome, illustrating the efficiency of NGS and 3-D pooling for sequence-indexing a large collection of *Ds** insertions.

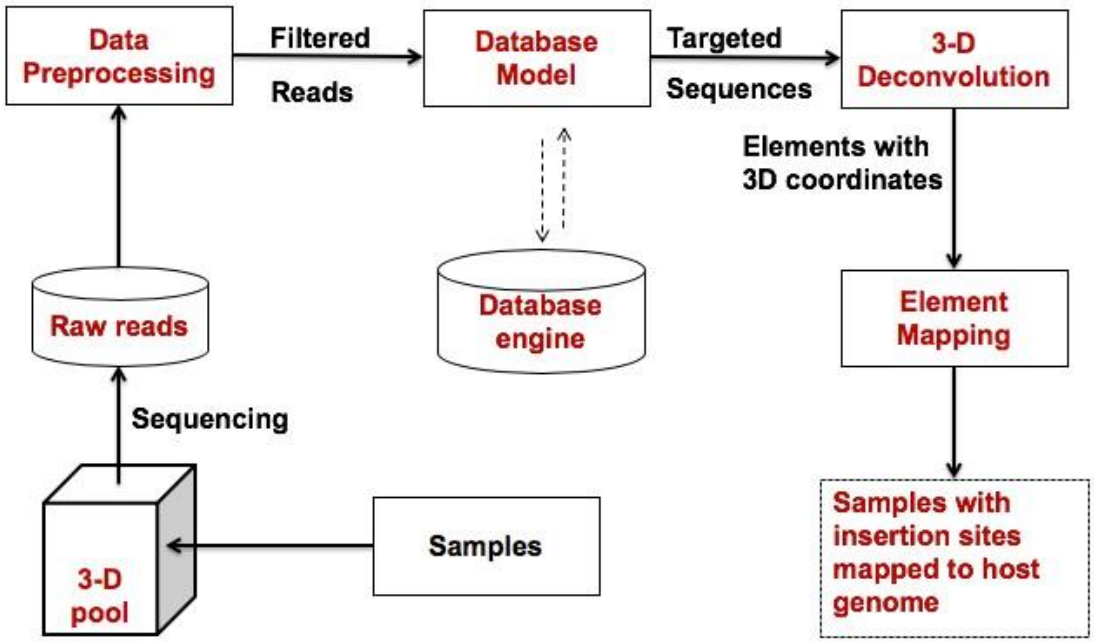

Figure 2. Schematic representation of InsertionMapper modules. InsertionMapper consists of four modules, each of which can be configured and run independently without affecting the others. This modular design of InsertionMapper ensures optimal output through fine-tuning parameters of each module.

The *trDs** elements were generated in a hybrid background containing B73 and A188 sequences from the Hi-II parent used in transformation and W22 sequences from the recurrent parent used to identify transpositions. Among the total of 14,184 *dsg* insertion lines, 11,335 (80%) were mapped to the same relative locations in B73 and W22, 1075 (7.6%) to different locations in B73 and W22, representing apparently translocated sequences between the two inbreds, 742 (5.2%) to unique locations in W22, and 409 (2.9%) to uniqe locations in B73 (Figure 3). Only 4.4% (625) dsgs insertions could not be mapped to either W22 or B73 and may represent sequences present in A188, but not in B73 or W22.

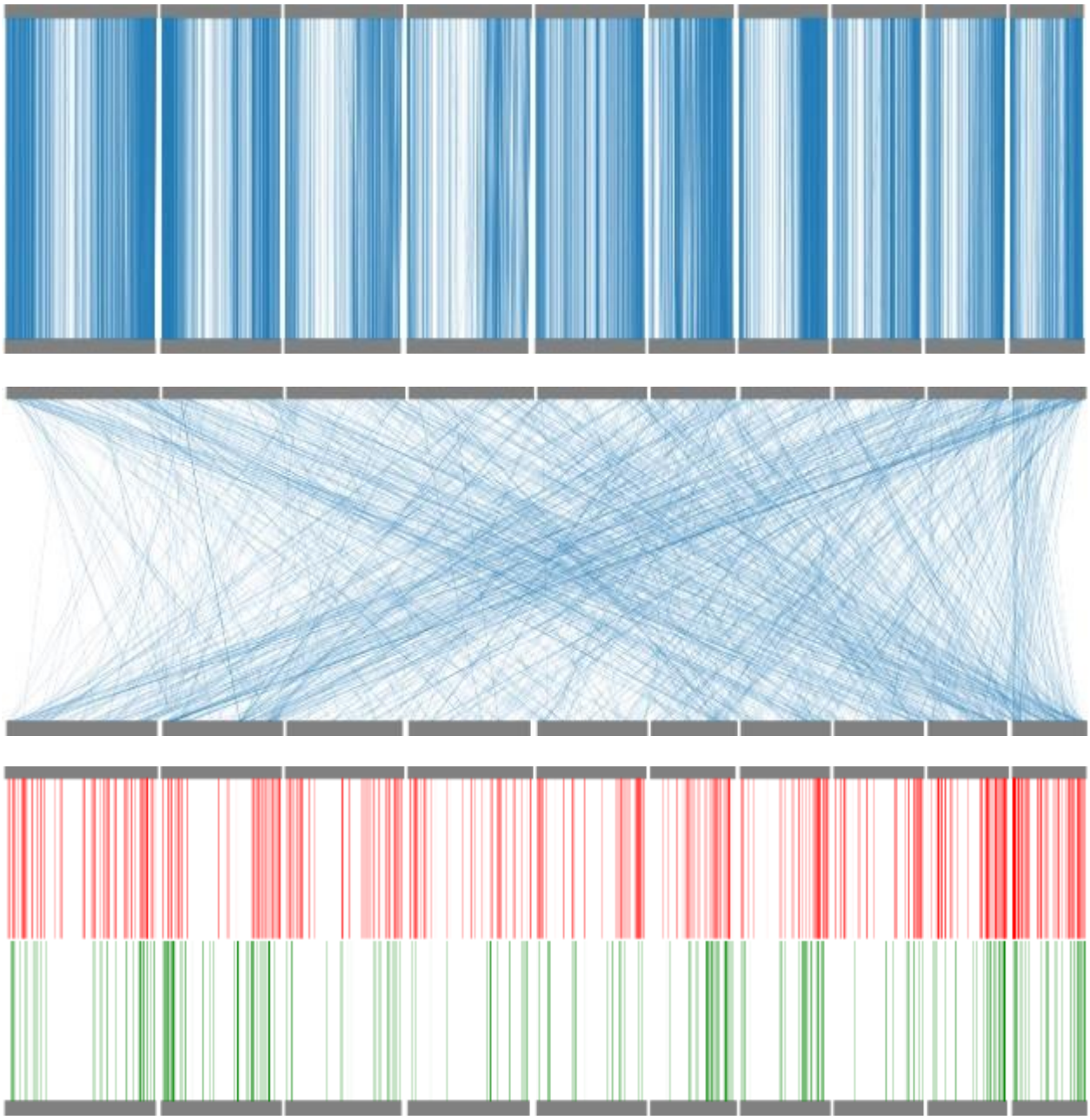

Figure 3. Top panel, *dsg*s shared between B73 and W22; middle panel, *dsg*s apparently translocated between B73 and W22; bottom panel, *dsg*s exclusive to each inbred line (red for W22 and green for B73). For all the three panels the top bars represent W22 and the bottom bars, B73. Chromosomes are arranged in order from 1 to 10, starting from the left.

Web-searchable database of insertion site sequences

All the information on sequence-indexed *dsg* sites, including their matching transposant lines, are shared via a web browser hosted at Montclair State University (http://acdsinsertions.org). We have set up a MaizeGDB-compatible relational database for the sequence-indexed transposant lines by using the freely available MySQL software. The user interface includes web searching forms written in Java and BLAST search tools. So far, we have sent 10,155 *dsgs* insertion lines to the Maize Genetics Stock Center, where they are available for distribution.