

The UniformMu Resource

Overview:

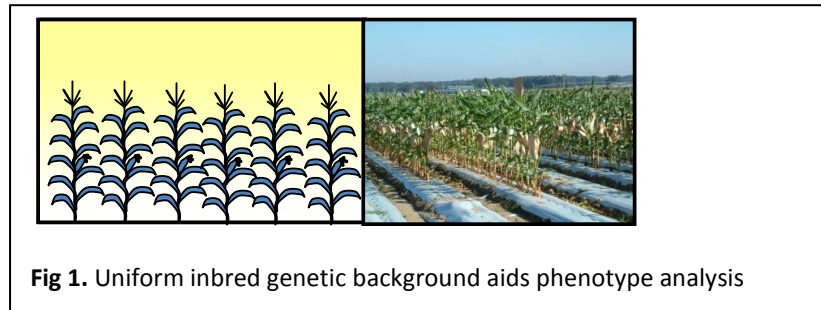
- I. Key features and current status of the UniformMu resource
- II. Finding insertions at MaizeGDB.org
- III. Recommended care and handling of UniformMu lines
- IV. UniformMu Methods: Genetics, sequencing and bioinformatics

I. Key features and current status of the UniformMu resource (March, 2011):

Definition: UniformMu is a unique maize population developed for genetics research (McCarty et al., 2005). This material is widely used for experimental analysis of maize gene function, because 1) the uniform plants and seeds provide an excellent basis for comparing unaltered, wild-type controls to detectable changes in new plant phenotypes, 2) the UniformMu transposable elements actively generate new mutations when they insert into individual genes, and 3) a searchable database (at MaizeGDB.org) matches gene sequences to specific mutants with Mu insertions in genes of interest (with seed stocks immediately available from the Maize Genetics Cooperation Stock Center via MaizeGDB.org). Gene roles can then be studied through loss-of-function effects at the biochemical, metabolic, and whole-plant levels. The population was developed by introgression of an active Mu transposon system from Robertson's Mutator into the genetically uniform W22 maize inbred. Mutations in genes are identified by sequence-based mapping of germinal Mu insertions in the maize genome.

Key features:

1. Uniform background: W22



2. Genetic stability: Mu-activity is monitored with a bronze color marker in kernels.

Conversion of lines from mutagenic to stable is achieved by using the color marker to select those with and without active transposase, respectively. The active, mutagenic, Mu-on lines are indicated by purple speckles on bronze kernels (sites where Mu excision has restored wildtype color), whereas stable, Mu-off lines have bronze kernels. For more, see UniformMu Methods.

3. Mapped, heritable insertions:

~ 10 new insertions per line if a visible phenotype is present.

~ 5 new insertions per line if visible phenotype not present.

Ancestral insertions are defined: There are ~ 40 canonical Mu elements in the W22 inbred.

4. Sustainable seed resource:

Seed stocks from sib-pollinated F3 plants are available immediately via MaizeGDB

Sib-pollination preserves deleterious mutations as heterozygotes.

Current status of UniformMu (as of March, 2011):

The latest UniformMu data release at MaizeGDB (March 2011) includes 26,211 germinal insertions mapped in 5,952 seed stocks, each available immediately and free of charge from the Maize Genetics Cooperation Stock Center. By July, 2011, we will have mapped up to 38,000 germinal, heritable insertions in 8,256 seed stocks. The collection currently contains insertions in an estimated 31% of maize genes based on mapped Mu sites in at least 12,127 high-confidence maize genes in the Filtered Gene Set (v 5b.60, MaizeSequence.org). These genes also include 5,329 that have multiple insertion alleles invaluable for confirming genotype-phenotype relationships (Table 1).

Table 1. Summary statistics for UniformMu March 2011

12,127	genes hit out of 39,475 total maize genes (based on the Filtered Gene Set v 5b.60, MaizeSequence.org)
31%	of maize genes have UniformMu insertions
5,329	genes with 2 or more mutant alleles
26,211	heritable Mu insertions mapped in
5,952	lines at the Maize Genetics Cooperation Stock Center

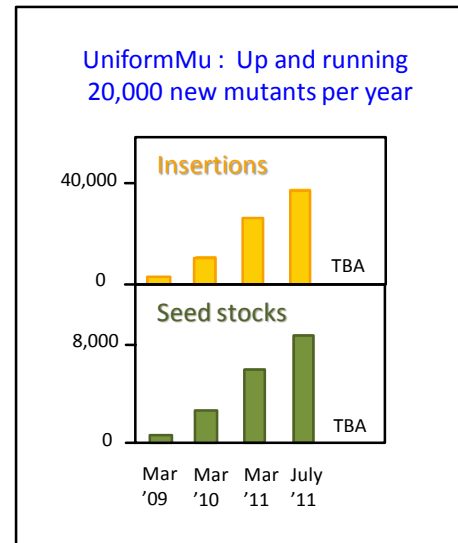


Fig. 2. Project statistics for UniformMu: Germinal insertions mapped (gold), and Seed stocks produced for the Maize Genetics Cooperation Stock Center (green).

UniformMu methods overview:

(More detail at end of document):

- Development of seed stocks:
 - Uniform background, heritable insertions, stable Mu-off, defined set of mapped Mu insertions in each line, visible ancestral mutants removed, lines immediately available, and free of charge.
- Mapping of Mu insertion sites:
 - High-stringency selection, validated
- Illumina sequencing protocols
 - Multiplexed, DNA-barcoded, grid-based

II. Finding insertions at MaizeGDB.org:

UniformMu insertions are displayed as track in the MaizeGDB genome browser (Fig 3). To search for insertions in a gene of interest (e.g. using a GRMZM identifier), paste the gene name in the MaizeGDB search window and select "genome browser - B73 RefGen_v2" from the drop down list. When the genome browser display for the gene appears, scroll down and locate the UniformMu track. Any insertions located in the displayed region will be displayed. Each unique insertion locus in the maize genome is given a unique identifier (e.g. mu10128088). Clicking on the insertion will bring up a locus page containing additional information about the insertion. To locate the seed stock containing the insertion, scroll down to the section labeled "Variations/Polymorphisms" and click on the link. The variation page will include a list of available stocks for the insertion (e.g. UFMu-01828). Clicking on the

stock link will open a stock page. The stock page includes a list of all insertions that have been mapped in the selected stock (typically 5-10 insertions, sometimes more). To request seed from the Maize Genetics Cooperation Stock Center, use the "Order this stock.." link located in the right-hand side panel.

Navigating MaizeGDB.org :

Data type:

Locus: [mu1014568](#)

Variation: [mu1014568::mu](#)

Stock: [UFMu-01040](#)

Each unique insertion site is a locus.
 Each locus has one variation – the Mu insertion.
 Each variation is assigned to one or more seed stocks.

UniformMu insertions are displayed as a track in the genome browser.

Browser view for GRMZM5G882075

UniformMu track

This gene has three insertions

Click an insertion site to open a locus page. Then follow the links to request seed.
 Locus page -> variations page -> stock page -> Stock Center Request Form (in right side panel).
 If more than one stock is listed for your insertion, we recommend getting multiple stocks.
 We also recommend including our color converted W22 wild type (X17EA) with your request (see text).

Fig 3. Navigation of MaizeGDB and the genome browser

II. Recommended care and handling of UniformMu insertion lines:

Planting UFMu (UniformMu) and W22 wild type seed stocks:

At least 15 seeds will typically be provided for each request of UFMu seed from the Maize Genetics Cooperation Stock Center. Because the seed supply is limited, it is important to plan carefully

before planting to ensure recovery of seed carrying the insertion of interest from as many plants as possible. If you have any problems, please contact us (DRM@UFL.edu or KEKoch@UFL.edu). The envelope will contain pooled seed from sib-pollinated F3 plants (see UniformMu Methods). The F3 lines are sibbed to minimize selection against insertions that may have deleterious phenotypes (e.g. lethal's, male steriles, etc). Many interesting, but deleterious mutations are thus preserved as heterozygous insertions; whereas, roughly half of the Mu insertions in UFMu lines are homozygous and will be present in all plants. Even if the Mu-insert of interest is deleterious, it should still be carried by at least some heterozygous F3 seed in the sample provided.

-- We recommend planting at least eight seeds in the field or greenhouse such that all plants can be sampled for DNA and pollinated.

-- We also recommend concurrently planting W22 wild type seeds (available from the Stock Center – catalog # X17EA), so that these can be used as additional mothers for pollen from mutant plants (Additional W22 can also be planted about 10 days before and after the mutant seed to increase potential for silk availability at the time of mutant anthesis).

This is important for three reasons. First, although UniformMu lines generally have good fertility, some may carry mutations that reduce vigor and fecundity (e.g. male sterility). Having wild type pollen and silks available helps ensure recovery of seed from the selected UFMu line. Second, backcrosses to wild type are often a necessary step towards developing a family of segregating progeny, where wildtype and mutant siblings can be tested for co-segregation of any phenotypic features with presence of the Mu-insert of interest. This is especially important if the insertion is homozygous in the UFMu line (about 50% of UniformMu insertions in F3 stocks are homozygous), in which case this generation will be needed to generate a suitable segregating family. Third, the UFMu lines typically carry 5 to 10 novel insertions (sometimes more), so a backcross will aid separation and independent assortment of these in subsequent generations (A given phenotype in the first generation may or may not be due to the Mu-insertion of interest). Recommended backcrosses can thus speed analyses.

PCR confirmation of Mu insertions:

1. To confirm presence of the Mu insertions in a gene of interest, genomic DNA must be extracted from individual plants and analyzed by PCR using gene-specific primers.

2. At least two gene specific primers will need to be designed for each insertion (one upstream and one downstream of the insert site), and these will need to be compatible with a Mu TIR specific primer. See Table 1 for a list of TIR primers we use (Latshaw and McCarty, unpublished, and Settles et al., 2004). We recommend TIR6 or TIR8 for routine confirmation and genotyping. Note that TIR8 is a mixture of four primers that are nested with respect to TIR6. This allows their use in two stages for enhanced specificity. TIR8 amplicons contain sufficient TIR sequence downstream of the primer to authenticate the Mu insertion. For sequence validation of insertion sites, we typically do an initial PCR reaction with TIR6. Product is gel purified and sequenced using the nested TIR8 primer and/or gene-specific primer.

3. Online tools such as Primer3 may aid design gene-specific primers. We typically use pairs of 21-27bp gene specific primers, with primers that anneal upstream and downstream of the predicted insert site. Where possible, we order three such pairs, so that they can be tested in alternate combinations (see more below on this). Ideally, the upstream and downstream gene specific primers should be separated by less than 500 bp in the wild type sequence to maximize amplification efficiency. BLASTN searches of the maize genome should be used to avoid primers that are present in multiple copies. Because there is currently limited genome sequence available for the W22 inbred background of UniformMu, we typically rely on the B73 reference genome for primer design. Sequence polymorphisms in the target region can affect primer efficiency and predicted product sizes. Where possible, targeting

exon sequences can reduce likelihood of sequence polymorphisms that will compromise optimal primer annealing. Table 2 contains our standard two-step, touchdown PCR protocol. We have found that including 5% DMSO in the reaction or addition of commercial PCR “adjuvant” cocktails (such as Invitrogen PCRx Enhancer System) can be effective for improving PCR success with GC rich maize genome sequences. We add fresh DMSO from aliquots stored at -20°C.

4. Gene-specific primers that flank the insertion site are first PCR-tested in pairs with wild type W22 DNA. This confirms capacity of the selected primers to amplify the expected wild type fragment.

Table 2. Two-step PCR protocol:

1. 94°C 1 min
2. 94°C 25 sec
3. 62°C 30 sec
4. 72°C 1 min (depends on expected size)
5. 8-10 cycles of 2 to 4.
6. 94°C 25 sec
7. 56°C 30 sec
8. 72°C 1 min (depends on expected size)
9. ~27 cycles of 6 to 8.
10. 72°C 5min
11. 4°C pause

The advantage of designing and ordering three pairs of primers, is that individual upstream primers can be combined with other downstream primers if the first pairs tested fail to optimally amplify the wild type sequence. All possible forward-reverse combinations can be tested with W22 DNA. W22-B73 sequence polymorphisms and multiple copy sequences are common causes of primer failure. The former are more likely to be encountered where one or both primers are located in the promoter or 5'-UTR. While placing primers in those regions is often unavoidable due to a strong 5' bias for Mu insertions, testing multiple primers will usually produce a successful pair.

Table 3. TIR primers

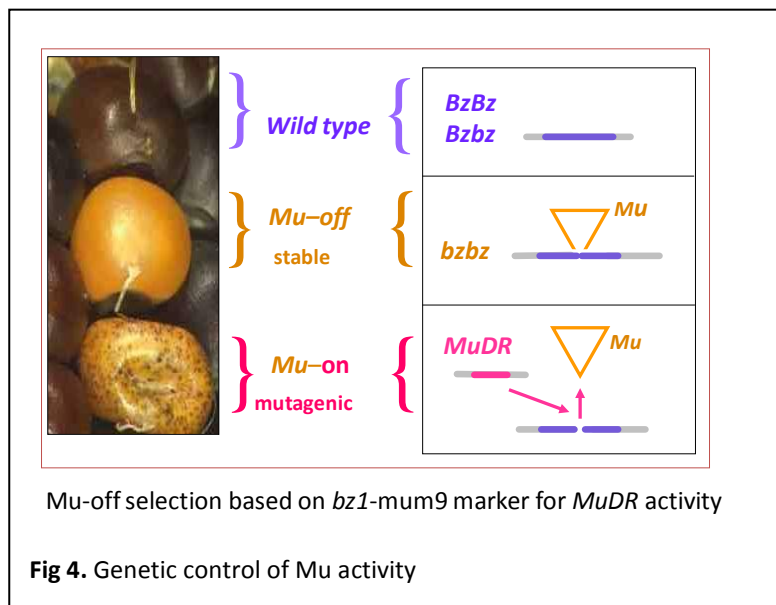
Sequence	Primer	Tm	Mu classes
AGAGAAGCCAACGCCAWGCCTCYATTCGTC	TIR6	71.7	
CGCCTCCATTTTCGTCGAATCCCCTS	TIR8.1	69.5	Mu7, Mu5
CGCCTCCATTTTCGTCGAATCCSCTT	TIR8.2	67.9	Mu1.7, Mu1, Mu1.4, Mu3, Mu4
SGCCTCCATTTTCGTCGAATCCCKT	TIR8.3	68.7	MuA2, MuDR, Mu8
CGCCTCCATTTTCGTCGAATCACCTC	TIR8.4	67.9	Mu7
GCTCTTCKTCYATAATGRCAATT	TIR5	58.3	

5. Once the best primer pairs have been identified for the gene of interest in W22 control DNA, we test **BOTH** forward and reverse primers in combination with the TIR6 primer in the presence of DNA from individual UFMu plants. We have seen a limited number of cases where only one of the primers produced PCR bands with TIR6 even though together they amplified wild type alleles. We subsequently confirm the location of Mu insertions by sequencing PCR products. Parallel PCR reactions performed with the pair of flanking gene specific primers can be used to test for presence of the wild type allele and heterozygosity. However, amplification of a wild-type-sized fragment can sometimes occur even in situations where an insertion is homozygous, if the gene is duplicated in the maize genome. If no products can be obtained from any plants using TIR6, try repeating the PCR using TIR8. If that also fails, contact us and we may be able to assist (e-addresses above).

6. The sequence analyses described in #3, may still result in primers that are not strictly gene specific. You may have to compromise and use them. If so, they may amplify two or more genomic segments that are not be resolved by standard gel electrophoresis. This situation could arise where a homozygous Mu insertion is present in a gene of interest that has paralogs or duplicated segments in maize genome. Such homozygous plants may be incorrectly genotyped as heterozygous because the wild type primer pairs are likely to generate PCR products visible as a single band in the agarose gel. There are two independent approaches to resolve and confirm such insertions. First, the PCR products can be directly sequenced with the gene specific primers that were used to amplify, or with additional internal sequencing primers. If primers amplify one or more additional loci, polymorphisms should be detectable with the PCR products. Second, a genetic test can be done by backcrossing the lines to the W22 inbred to produce multiple F1 ears. If parental plants are homozygous for the Mu insertions, all the F1 plants will typically carry the insertion.

IV. UniformMu Methods: genetics, sequencing and bioinformatics

Field genetics: Production of F3, UniformMu seed stocks for sequence-indexing and public distribution.



To generate and isolate new mutations, plants heterozygous for a highly active MuDR (i.e. self-pollinated ears that segregate 3:1 for densely spotted, *bz1-mum9* and stable bronze kernels) are backcrossed to our standard W22 ACR inbred strain (McCarty et al., 2005). The resulting F1 plants are grown, self-pollinated and screened for stable *bz1* seed (Fig 4; McCarty et al., 2005). Because the F1 plants are heterozygous for MuDR, the self-pollinated ears segregate Mu-inactive seed that have a stable bronze (non-spotted) kernel phenotype. The stable bronze F2 seed are grown and self-pollinated to establish confirmed, homozygous MuDR(-), Mu-inactive derivative lines. Twenty, spot-free F3 kernels are selected from each line, grown and the plants sib-pollinated.

To create a seed stock suitable for sequence analysis and public distribution through the Stock Center, a balanced pool of 200-300 seeds is prepared from 2-5 ears (depending on ear size, and quality) from each family. To ensure highest possible seed quality, all seed stocks are screened manually by experienced personnel (including students, mentors, and technical staff) to eliminate residual Mu-activity, cracking, mold or other defects that could impact seed viability.

UniformMu maize plants are grown and pollinated at the UF PSREU experimental farm at Citra, FL and at the Gray Research facility located in Askum, Ill. The Florida location allows two generations per year while the relatively mild central Illinois summer season is optimal for production of high quality seed stocks.

Sequence-indexing of Mu insertions by Illumina sequencing:

We have developed a highly efficient protocol for sequencing Mu flanking sequences using the Illumina platform (Illumina.com). Plants are grown in 24 X 24 grids (576 lines per grid) creating a 2D array of 48 pooled DNA samples (24 for each axis). The 48 DNA samples are sequenced in a single multiplexed library using one lane of an Illumina flow cell. Multiplexing is implemented using a custom set of Illumina Flow Cell A sequencing adaptor oligos each incorporating one of 64 different 4-base key codes into the 5'-end of the sequence. The fourth base of the code is a check sum that allows detection of sequencing errors in the multiplex key. To eliminate chimera PCR artifacts that can result in "key swapping" among the multiplex samples, all steps of library construction are performed separately for each of the 48 DNA samples and sub-libraries are pooled prior to the cluster generation stage in the Illumina sequencing protocol.

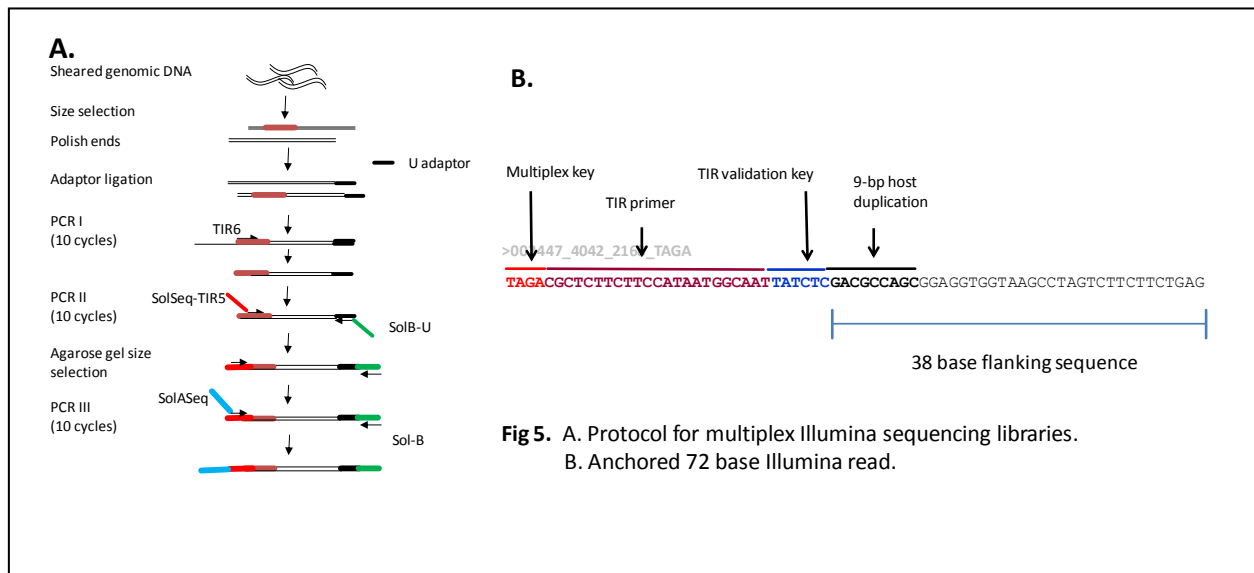


Fig 5. A. Protocol for multiplex Illumina sequencing libraries. B. Anchored 72 base Illumina read.

In brief, to obtain unbiased sampling of the genome, each DNA sample is randomly sheared using a BioRuptor sonication instrument, end-polished and ligated to a universal adaptor oligonucleotide (Fig 5). PCR I is performed using a Mu TIR specific primer and the universal adaptor primer. A nested PCR II reaction is then performed to incorporate the Illumina A and B flow-cell adaptor primers into Mu specific amplicons. The 3' half of the long Illumina A-sequencing adaptor is fused to a nested Mu TIR specific oligo that is anchored six bases from the 3'-end of the TIR (Fig 5a), and the B-adaptor sequence is incorporated by fusion to the universal adaptor primer. The six bases of TIR in the genomic sequence downstream of the A-primer are used to confirm authentic Mu flanking PCR products. Less than 1% of reads are rejected for mispriming. The PCR II products are size-selected by excision from agarose gels where each sample is run in a separate gel to eliminate the possibility of cross-contamination. Incorporation of the 5-prime portions of the A-sequencing adaptor is completed in a third PCR reaction. All PCR steps use a minimum number of cycles to mitigate PCR bias.

Bioinformatics analysis and mapping of Mu flanking sequences:

In a typical 72 base run performed at NCGR (Santa Fe, NM), our custom multiplex libraries yield 25-32 million reads per lane providing greater than 500,000 quality trimmed reads per multiplex DNA sample. With current Illumina technology, 72-base runs return high yields of reads conducive to statistically robust axis assignments while allowing relatively fast turn-around due to availability of 72-base lanes at sequencing centers (e.g. NCGR). Longer reads may be obtained from the same libraries without modification. After parsing of the key code, trimming of the Mu TIR sequence, reads retain 38 bases of precisely anchored genomic DNA sequence flanking the Mu insertion (Fig 5 b). The trimmed reads are mapped to the B73 maize genome (AGP v2 pseudomolecules) using BLASTN run in parallel on the UF HPC computing cluster (www.hpc.ufl.edu). Using 64 processors, mapping of 32 M reads is completed in about 2 hours. The BLAST output is parsed into a database of mapped Mu insertion sites using a cutoff BLAST expectation score of 10^{-7} . Flanking sequences that map within 3 bp in the maize genome are assigned to the same insertion locus.

For each insertion locus, read counts in each grid axis are used to locate maize lines in grid that contain that insertion (Fig 6). Insertion loci that map to low copy locations in the genome (maximum of 4 unresolved copies; >95% are single copy) and that have unambiguous, statistically robust assignments (axis counts with a X^2 score less than 10^{-5}) to individual lines in the grid are assigned a unique six digit identifier (e.g. mu103480) and submitted to MaizeGDB.org. In parallel, seed stocks (minimum of 200 high-quality seed per line) for each grid are deposited in the Stock Center. Consensus flanking sequences for each insertion are assembled from the Illumina reads using the AMOS simple consensus program (Meyers et al., 2000) and paired flanking sequences are deposited in Genbank as gapped 76 bp sequences (or longer depending on Illumina read length).

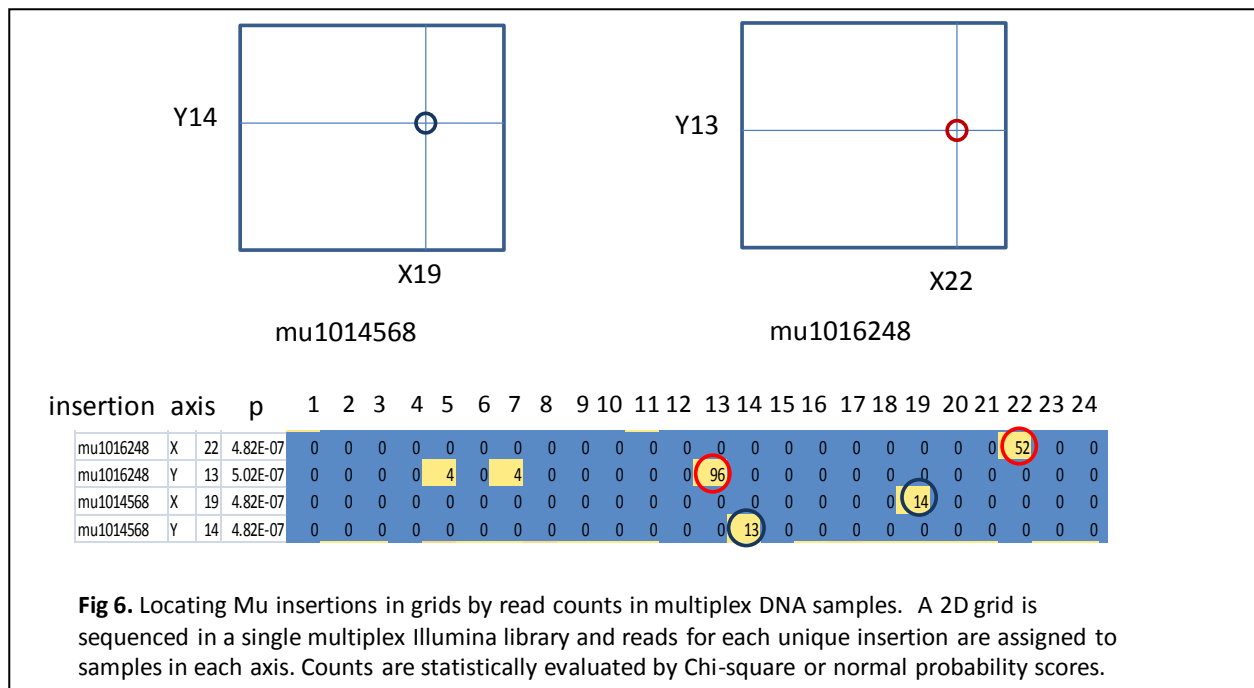


Fig 6. Locating Mu insertions in grids by read counts in multiplex DNA samples. A 2D grid is sequenced in a single multiplex Illumina library and reads for each unique insertion are assigned to samples in each axis. Counts are statistically evaluated by Chi-square or normal probability scores.

References:

- McCarty DR, Meeley RB. (2009) Transposon resources for forward and reverse genetics in maize. In. Handbook of Maize: genetics and genomics. (Ed. JL Bennetzen, S. Hake). Springer Science Berlin. pp 561-584.
- McCarty DR, Settles AM, Suzuki M, Tan BC, Latshaw S, Porch T, Robin K, Baier J, Avigne W, Lai J, Messing J, Koch KE, Hannah LC. (2005) Steady-state transposon mutagenesis in inbred maize. *Plant J.* 44:52-61.
- Myers E, Sutton G, et. al. (2000) A Whole-Genome Assembly of *Drosophila*. *Science* 287:2196-204
- Schnable PS et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112-5
- Settles AM, Holding DR, Tan BC, Latshaw SP, Liu J, Suzuki M, Li L, O'Brien BA, Fajardo DS, Wroclawska E, Tseung CW, Lai J, Hunter CT 3rd, Avigne WT, Baier J, Messing J, Hannah LC, Koch KE, Becraft PW, Larkins BA, McCarty DR. (2007) Sequence-indexed mutations in maize using the UniformMu transposon-tagging population. *BMC Genomics.* 8:116