Tilling and Eco-Tilling

TILLING (Targeting Induced Local Lesions in Genomes) is a method in molecular biology that allows directed identification of mutations in a specific gene. TILLING was introduced in 2000, as functional genomic tool for the model plant Arabidopsis thaliana. TILLING is a non-transgenic technology that utilizes a reverse genetics approach for the production and detection of mutation and has been successfully employed in zebrafish, maize, wheat, rice, soybean, tomato, lotus, mustard, and lettuce. It overcomes the limitation of knockout experiments. TILLING combines a standard and efficient technique of mutagenesis with a chemical mutagen such as ethyl methanesulfonate (EMS) with a sensitive DNA screening-technique that identifies single base mutations (also called point mutations) in a target gene. In contrast, Eco-TILLING is a variation of TILLING that investigates the natural variation among cultivar/inbred line/accession when aligned with a sequenced reference genome for the identification of SNPs.

Next-generation sequencing (NGS), with advantage of high throughput sequencing of the targets provides advantage over the traditional LICOR gel based system to identify mutations involving PCR amplification and digestion of mismatches using the restriction enzyme Cel-I. Drawbacks to the LICOR system is having to pour slab gels and long run times (~4 hours). The current NGS technologies are robust and convenient data analysis using bioinformatics tools to analyze thousands of samples at a time using different pooling schemes.

Xcelris Genomics offers the high-throughput of next-generation sequencing technology of HiSeq 2000/2500 and MiSeq along with resolving power of overlapping pool design for efficient and economical functional genomics analysis across thousands of TILLING population generated for different organisms.

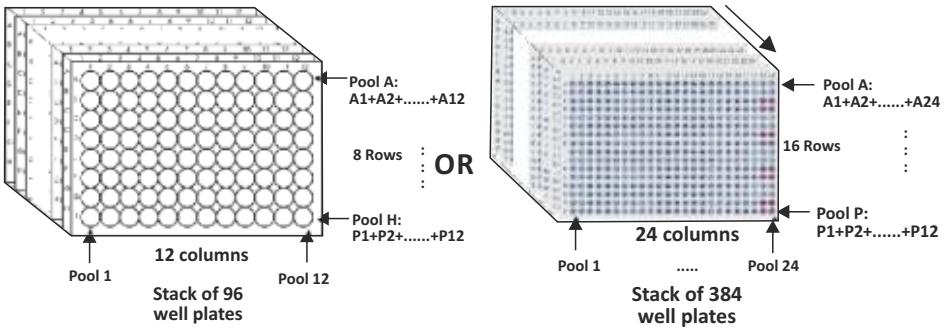
Sample Requirement:

1. Fresh Tissue Samples
   Total 2-5g of fresh tissue samples from which using high throughput (96 well format) technology will be utilized for isolation of gDNA of each sample.
2. DNA Samples
   High molecular weight intact double stranded gDNA samples (5-10 µg, conc. >100 ng/µl) dissolved in TE buffer, A260/280 ratio between 1.7 to 1.9 along with gene sequences in fasta format.
3. PCR products
   1 to 4 µg (conc. > 100 ng/µl) purified amplicon product.

Note: All type of samples should be transported in -20ºC to Xcelris Genomics, Ahmedabad, Gujarat, India.

Brief methodology for TILLING

- Isolation of gDNA samples
- Amplification of individual genes by PCR.
- Pooling of PCR products in equimolar ratio.
- Preparation of barcode indexed Illumina libraries
- Sequencing on HiSeq 2000/2500/MiSeq and filtration of high quality data
- Demultiplexing of the sequence reads based on barcode.
- Alignment of reads by gene wise and library wise using reference sequence.
- Variation calling (gene wise library wise).
- Variation intersection of row pools and column pools.



- Schematic pooling on 96/384 well format to generate 20/40 PCR pools
- Construction of bar-code indexed libraries(n=20/40)
- Paired end sequencing and mapping to reference sequences
- Mapping of mutations by bioinformatics analysis

Bioinformatics analysis/deliverables

The raw data generated from sequencer will be filtered using Trimmomatica and high quality reads will be mapped on reference gene sequence using MAQ for SNP calling. In-house perl scripts will be used for for SNP selection, filtration and matrix development **(Fig 1)**

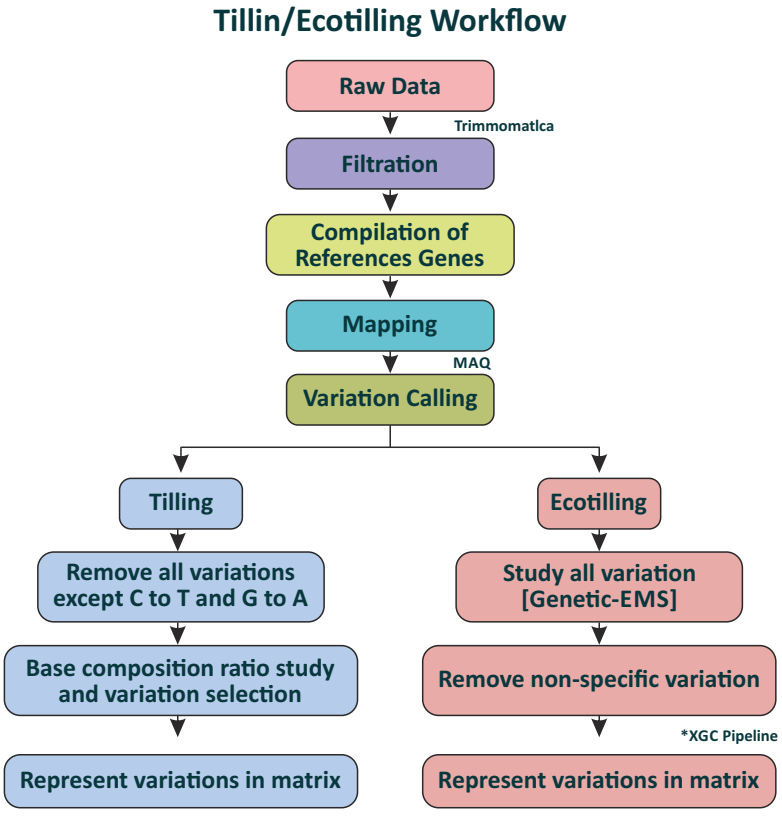


Fig 1. Schematic workflow for bioinformatics analysis for mutation detection in TILLING population and EcoTILLING analysis.

Deliverables:

- High quality data in fastq format for all the libraries
- Mapped data in fastq format for all the libraries
- Alignment files (*.map) for all the libraries
- All the variations in the libraries for Tilling and Eco-Tilling in .xls, with gene ID, SNP position, read depth, variation count and percentage with respect to reference.
- Matrix for all the libraries for Tilling and Eco-Tilling in .xls with gene ID, SNP position, read depth, variation count and percentage with respect to reference.

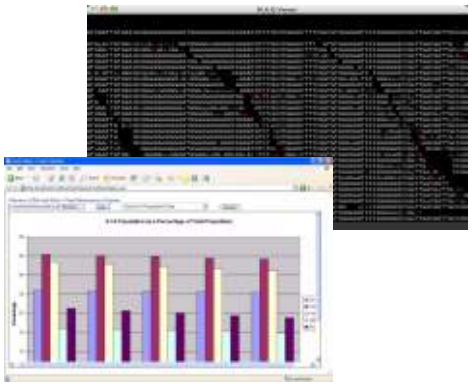


*Fig 2.SNP calling representation (graphical and terminal)

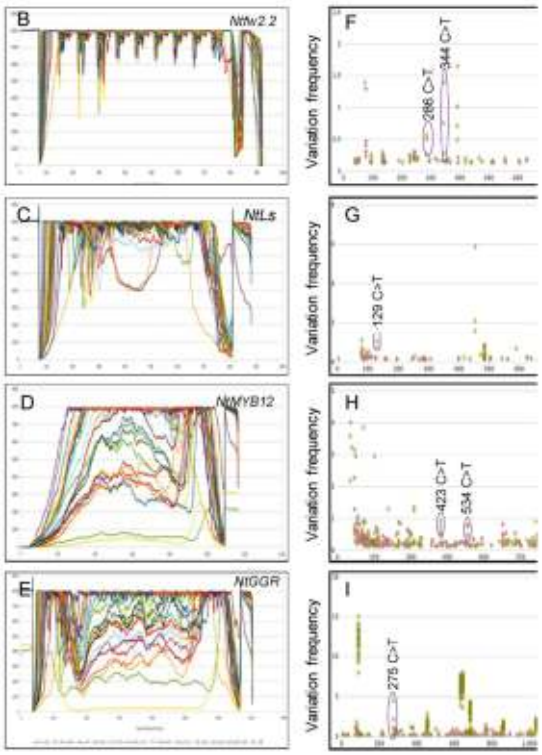


*Fig 3. Gene wise sequence coverage and variability across pools

Time line: It depends upon the type of sample, number of genes, number of samples, complexity and coverage required. The generalized time line is four to six months.

*TV Reddy et al., Industrial Crops and Products, Volume 40, Issue null, Pages 324-335