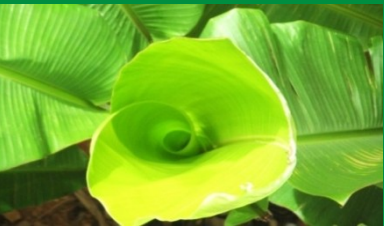


**Case studies in targeted resequencing and microarray
applications in
-Marker assisted Plant Breeding
-Veterinary Sciences**



Dr. Raja Mugasimangalam

2nd July 2010., University of Agricultural Sciences, GKVK, Bangalore India



Thanks to all of you – for registering online

Sequencing

PCR

qPCR

Microarrays



Agilent
Certified
Services Provider
Microarray-Based
Genomic Analysis



INFRASTRUCTURE & CAPABILITIES



Microarray Design team

Molecular biology lab

(RNA DNA extractions, protocol optimization qPCR)

Microarray Lab

Microarray analysis Team

NGS Lab

(NGS library preps, barcoding On Array Capture and InSolution Capture, QC)

Next Generation DNA sequence analysis team

Bridging the gap between Biology and Technology



Microarrays to NGS !

Home

About Us

Workshop

Contact Us

Careers

News

Clientele

Publications

Microarray Services

Genomics Solutions

Custom Microarrays

Data Analysis

Biointerpreter

Bio-IT

Promotions

Next Generation DNA Sequencing

**Parallel sequencing reactions on a
microarray slide!**

1 Million tiny wells (454 TECHNOLOGY)

**100 Million spots-- Solexa /Solid:
whole slide as a single well**

Illumina Sequencing Platform

1. Sequencing by synthesis
2. Reversible dye terminators



Solexa

Sequence reads with base quality, FASTQ format

```
*****
|SNAPSHOT OUTPUT
*****

*****
File Name: s_1_1_sequence.txt
File Size: 32635 MB
*****

*****
Printing TOP 20 lines
*****

@SLXA-EAS1_89:1:1:657:649/1
GCTACGGAATAAAAAC
+SLXA-EAS1_89:1:1:657:649/1
cccccccccccccccc
@SLXA-EAS1_89:1:1:657:649/1
GCAGAAAATGGGAGTGAAAATCTCCGATGAGCAGC
+SLXA-EAS1_89:1:1:657:649/1
ccccccccccbecbecb` `cecbcccZcc`^`bR^`
@SLXA-EAS1_89:1:1:708:653/1
GAGAGAGCAGTGGGCGAGGTTGGGACATGTCATGA
+SLXA-EAS1_89:1:1:708:653/1
cccccccccccccccccccccccccccccccc`cecb` `Y^
@SLXA-EAS1_89:1:1:675:644/1
GAACATTTATTATAATCCTATTCAATTATAATAAT
+SLXA-EAS1_89:1:1:675:644/1
cccccccccccccccccccccccccccccccc^bb^b
@SLXA-EAS1_89:1:1:721:668/1
GCTGTAGATCTGGAAATCGCAACGGAGGAAGAAAG
+SLXA-EAS1_89:1:1:721:668/1
cccccccccccccccccccccccccccccccccbbbb
```

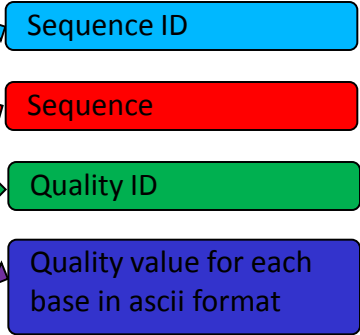


Fig1: Snapshot of the FASTQ file

Reference Sequence Alignment

Alignment Program for Short Read NGS Sequences:

MAQ (version 0.6.8)

Authors:

Heng Li and Richard Durbin

Publication:

Mapping short DNA sequencing reads and calling variants using mapping quality scores.

Heng Li, Jue Ruan and Richard Durbin

Genome Res. 2008 Nov;18(11):1851-8. Epub 2008 Aug 19.

Genome Length	% of reference sequence mapped	Reads mapped	Reads unmapped	Avg. read depth	Avg. length of gaps	No. of Gaps	Length of Gaps
14 Mb	98.92	5.3 Million	7.1 Lakh	20	24	6431	154.5 Kb

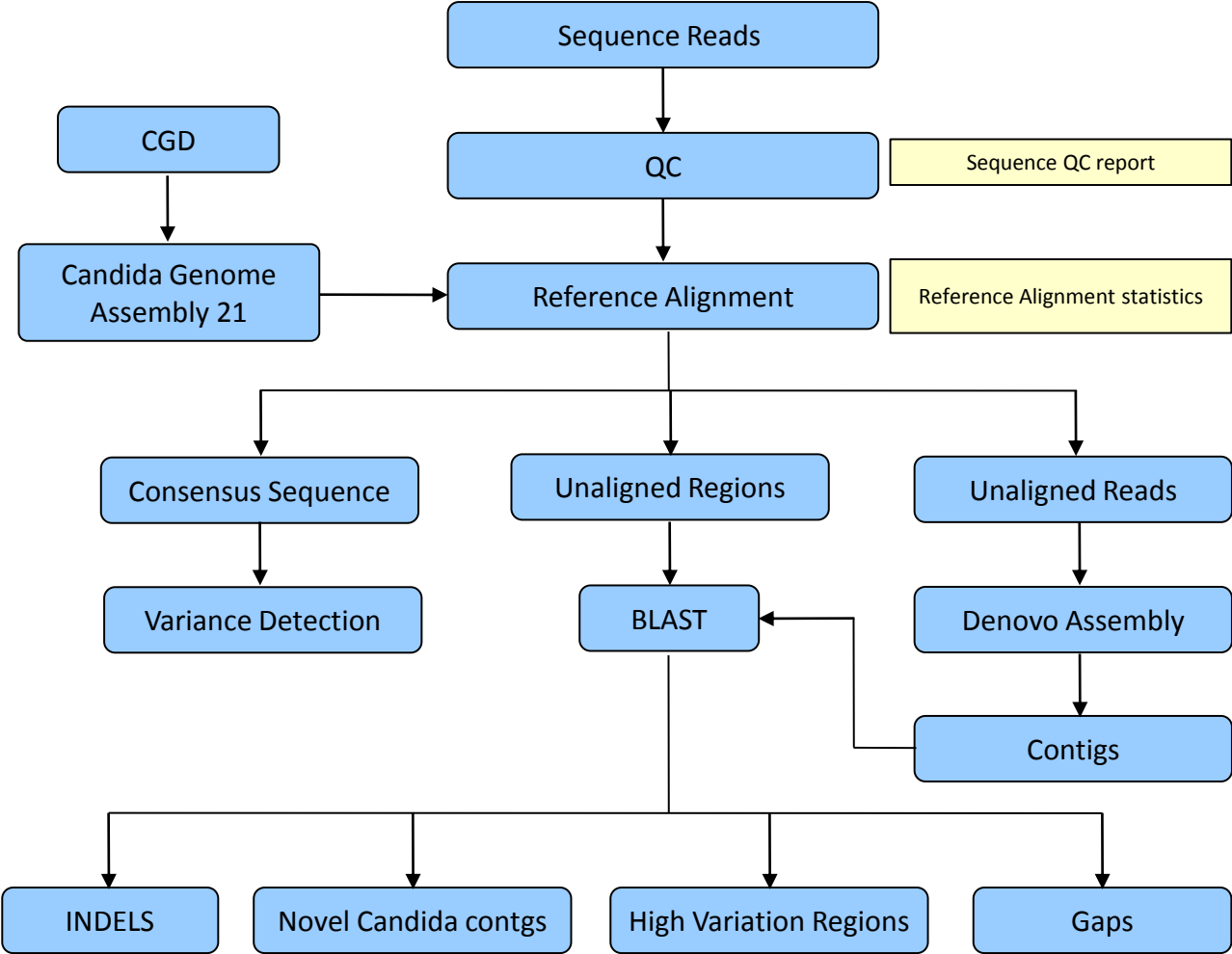
***Gaps might be deletions or regions of high variance.**

Single Base Variance Categorization:

Single Base Change	Type of Change	All Chr.	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	ChrR	Mitochondria
A-G	Transition	4682	1028	650	415	585	469	360	356	819	0
G-A	Transition	4247	934	573	398	527	446	335	303	731	0
T-C	Transition	4643	992	649	450	591	432	413	341	773	2
C-T	Transition	4153	888	575	379	531	457	311	302	709	1
A-T	Transversion	1166	256	140	107	154	140	102	77	190	0
A-C	Transversion	663	137	84	73	81	64	43	61	120	0
G-C	Transversion	473	101	78	31	64	53	45	32	69	0
G-T	Transversion	740	171	106	66	93	85	52	45	122	0
T-A	Transversion	1207	265	150	109	154	125	99	90	214	1
T-G	Transversion	722	165	103	59	84	68	51	69	123	0
C-A	Transversion	779	187	99	65	103	80	56	62	126	1
C-G	Transversion	513	118	73	33	64	64	33	54	74	0
N-A	Unknown-Known	6	1	1	0	0	0	0	0	4	0
N-T	Unknown-Known	49	11	6	8	7	2	0	0	15	0
N-G	Unknown-Known	55	7	13	8	11	1	0	0	15	0
N-C	Unknown-Known	29	8	5	2	5	1	0	0	8	0

Analysis Workflow

Deep Sequencing of *Candida albicans* Indian clinical isolate



Sequence Files



Ca21_
chromosomes.fasta

Description: Candida albicans genome, assembly 21

File Size: 14 MB

No. of Sequences: 9

Total Sequence Size: 14 Mb



s_7_sequence
_GT492L.txt

Description: Candida albicans Indian Clinical Isolate Illumina Sequenced Reads

File Size: 1 GB

No. of Sequenced Reads: 6 Millon

Total Sequence Size: 300 Mb

20X coverage

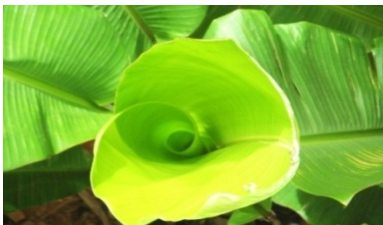
Anything that is living, lived are being sequenced worldwide

What are we in India doing??

Moringa oleifera (drum stick)

Turmeric

Bitterguard



Next Generation Sequencing

- **1 lane provides 2GB of DATA !!**
- **Is it enough??**

- **For a 3GB genome at 2GB output per lane-**
- **To sequence a whole genome 20X, we need minimum 30 lanes!!**

WHY Targeted ReSequencing??

- **Sequence Exons (whole Exome)**
- **Sequence Exons in a chromosome**
- **Sequence a Genomic Region (s)**
- **Sequence Chloroplast and mitochondria**
- **Sequence REGIONS THAT MATTER**

➤ **Whole Genome sequencing at >50X is still very expensive (INR 25 lakhs OR USD 50K per sample)**

➤ **Sequence focused regions at a much lower cost - Multiple samples**

INR 1.5 lakhs (USD 3000) for 1 sample

On-Array DNA Capture Workflow

Provide Genotypic list of Genes / Exons / Genomic regions of Interest or Choose from catalog capture designs

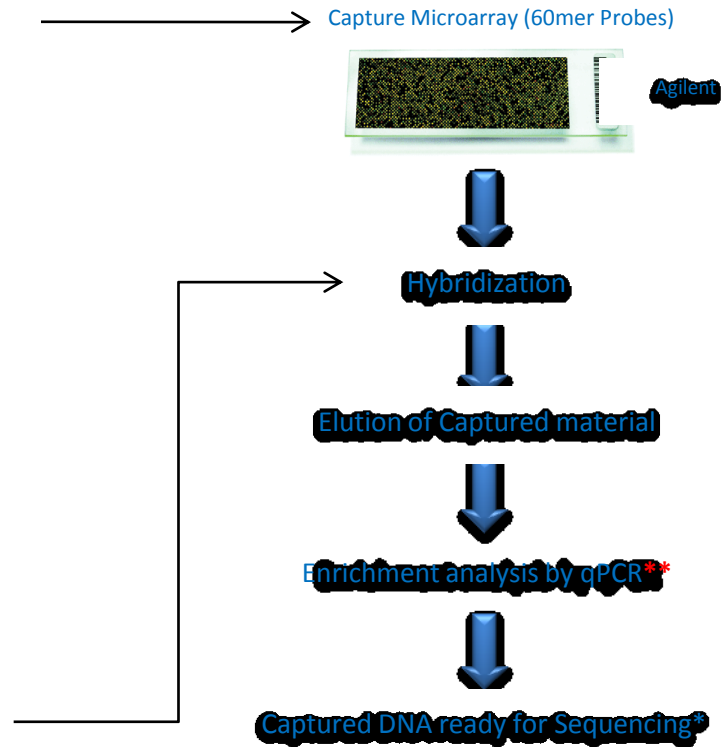
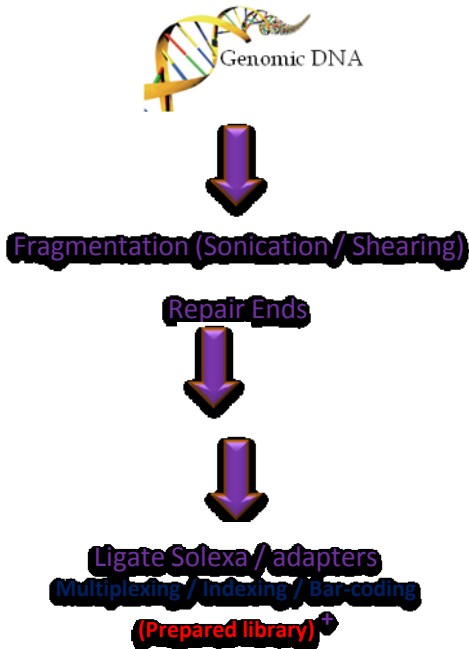


Fig 7. Sheared DNA

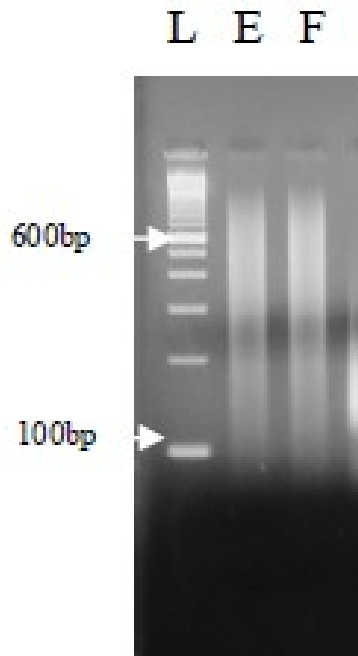
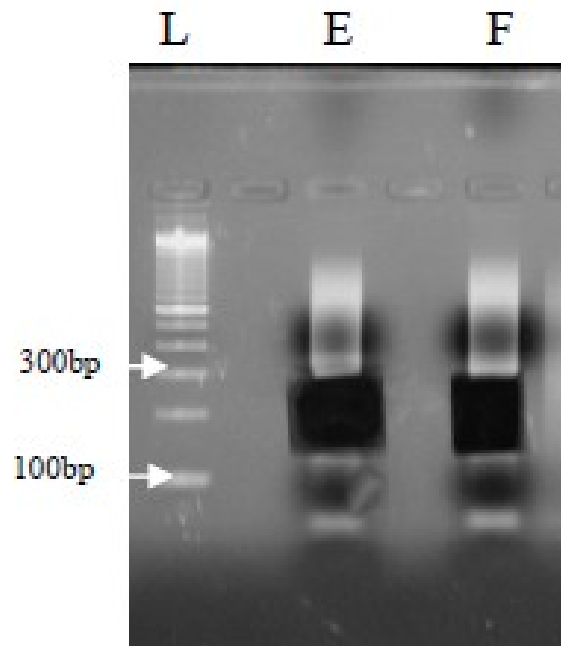
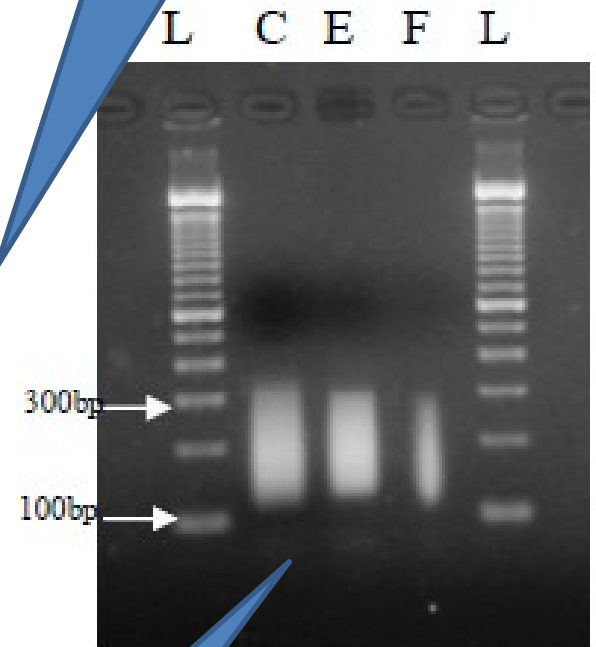


Fig 8. Gel Elution elution



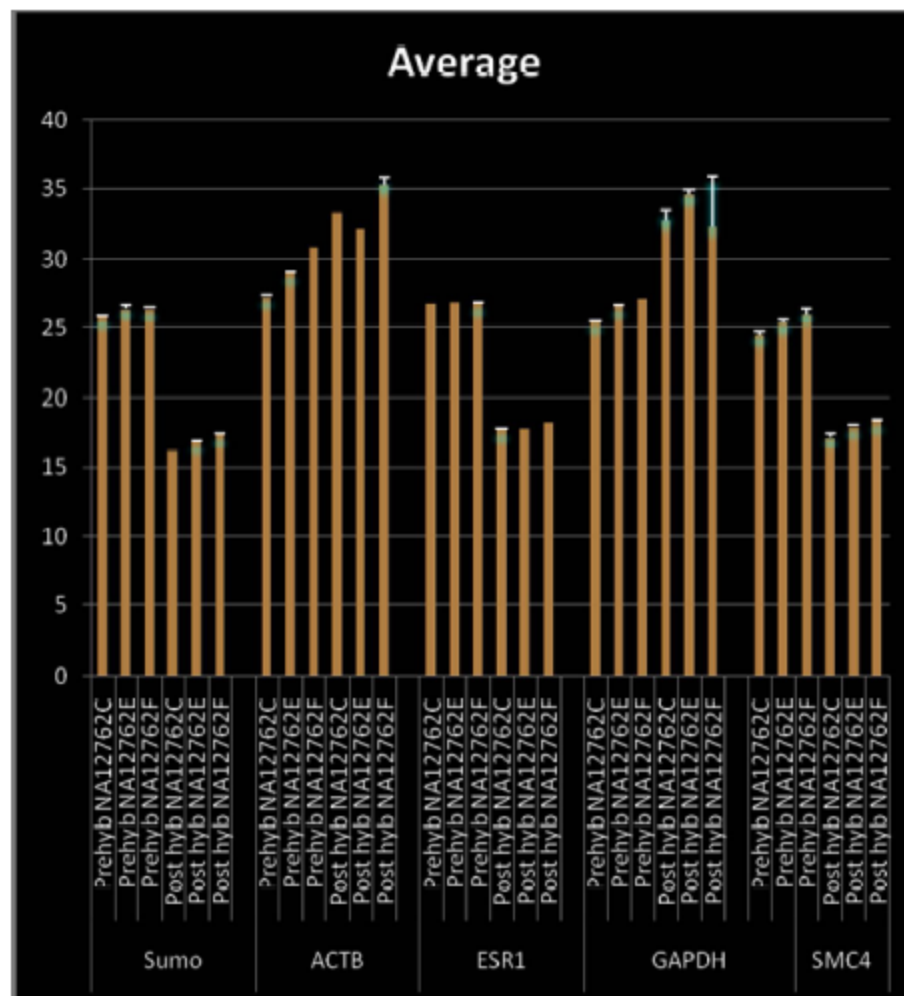
Ligate to
sequencing
adapters

Agarose Gel Electrophoresis (18 cycles)



Post Sequence
capture

Fig 10-QPCR QC of captured selected Exons from samples C,E and F



Gene	Sample	dCT	Fold Enrichment
Sumo	NA12762C	9.58	765
	NA12762E	9.5	724
	NA12762F	8.96	498
ACTB	NA12762C	-6.1	0.01
	NA12762E	-3.19	0.11
	NA12762F	-4.4	0.05
ESR1	NA12762C	9	512
	NA12762E	9	512
	NA12762F	8.48	357
GAPDH	NA12762C	-7.3	0.01
	NA12762E	-8	0.00
	NA12762F	-5.2	0.03
SMC4	NA12762C	7.3	157
	NA12762E	7.3	157

Sequencing by Illumina GAI

S.No	1	2
Fastq file name	s_1_seq_915.txt	s_1_seq_986.txt
Fastq file size	1.59 GB	1.53 GB
Time taken for Analysis	17.31 minutes	18.95 minutes
Maximum Read Length	50	50
Minimum Read Length	50	50
Median Read Length	50	50
Total Number of Reads	9370291	8990756
Total Number of HQ Reads 1*	8591825	8098708
Percentage of HQ Reads	91.692 %	90.078 %
Total Number of Bases	468514550	449537800
Total Number of HQ Bases 2*	430768489	407440589
Percentage of HQ Bases	91.943 %	90.635 %
Total Number of Non-ATGC Characters	154664	147377
Percentage of Non-ATGC Characters	0.033 %	0.033 %
Number of Reads with Non-ATGC Characters	86327	82839
Percentage of Reads with Non-ATGC Characters	0.921 %	0.921 %

Alignment to Reference Genome

Statistics	Sample 1 (s_1_seq_915.txt)	Sample 2 (s_1_seq_986.txt)
Reference Sequence	Chr11-17713870- 33798836.txt	Chr11-17713870- 33798836.txt
Total Reads	9370291	8990756
Aligned Reads	2427518	2276330
Unaligned Reads	6942773	6714426
Capture Region Length	702099	702099
% of Reference Sequence covered	97.19	96.99
Average Read Depth	138 X	120 X
Length of Gaps	19762	21136

In Solution Capture

RNA “Baits”

High concentration of Baits over template

- **Bait:**

- A single oligo sequence of pre-determined length (120 bp) that complements a targeted region of the genome



- **Bait Group:**

- Consists of a group of Baits designed to complement a single or set of targeted intervals
- May be formed from baits generated within eArray, baits uploaded into eArray, or bait search results within eArray



- **Library:**

- Consists of one or more Bait Groups
- Represents the set of oligos that will be produced for the kit



Ideal for large number of samples

On Array

vs

InSolution

DNA Oligos on Array

RNA “Baits”

High Template conc

High Bait Conc

Highly flexible (one array per design)

Minimum 10 reactions

Difficult work Flow

Easy work flow

Suitable for small studies

Large studies

Multiplex capture

Not possible

Genotypic and Targeted Reseq pipeline

Send Genomic DNA

Choose from Catalog (whole exomes or X chromosome exomes)

Discuss and Design capture Array or Baits (insolution capture)

Get back Captured ready to sequence library (with enrichment QC- 3 targeted and 2 untargeted regions)

Get it sequenced by Genotypic / at your facility

Analysis at Genotypic** / at your facility

SNP /Mutation calls and confirmation

**** HPC at Genotypic and Super computing facility of CRL Pune**

The **largest linkedin Group for NGS** is run by Genotypic & *Yahoo group Nextgenseq*



Genomics: Next Generation DNA Sequencing (NGS) and Microarray

Already a member

To discuss novel Genomics applications of NGS Next Generation Sequencing and Microarray and combinations including Targeted resequencing. SUBGROUPS: NGS Platforms, Sequence Analysis, Business, Jobs, Services, R&D, Resequencing, Lab protocols

Yesterday's Activity: [Discussions \(1\)](#) [News \(1\)](#) [Jobs \(2\)](#)

Owner: [Raja Mugasimangalam](#) | [2,032 members](#) | [Share](#)

www.linkedin.com/groups?about=&gid=1907871&trk=anet_ug_grppro



Plant breeding & Genetics AND Agri Genomics

Already a member

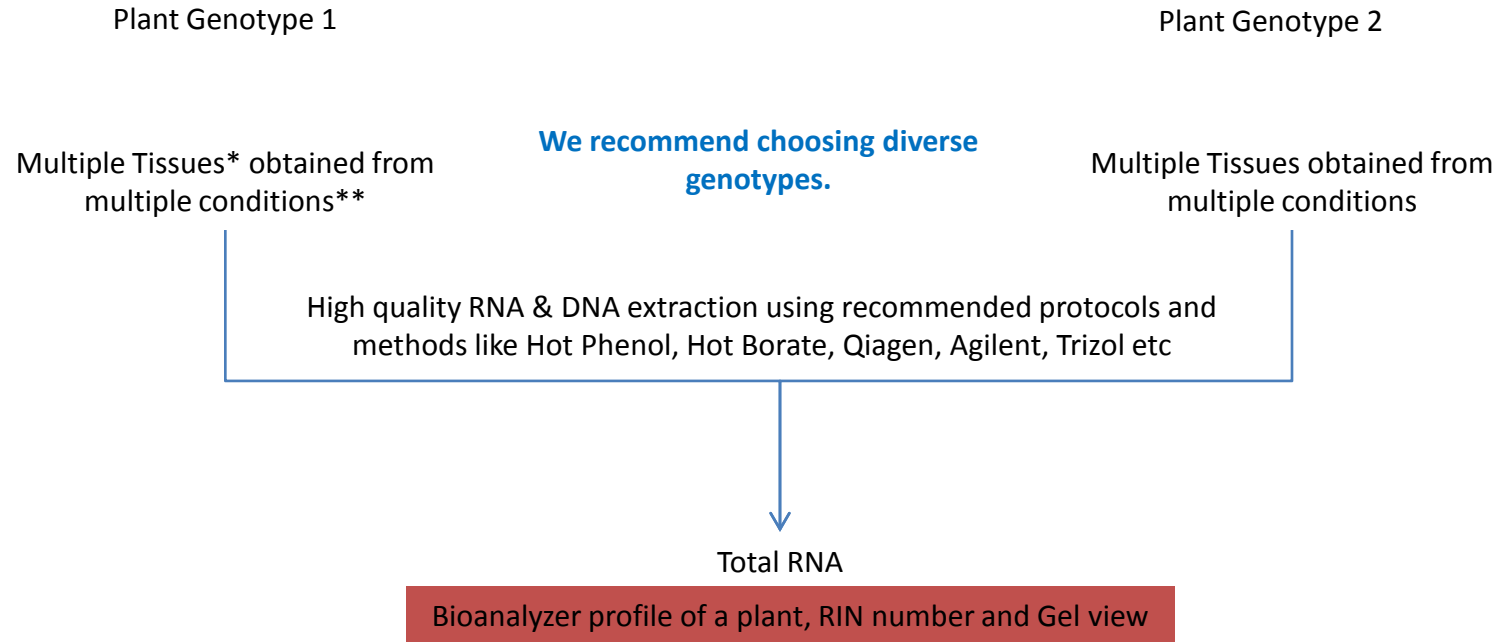
A group to discuss technologies available for Marker assisted plant breeding.

Owner: [Raja Mugasimangalam](#) | [16 members](#) | [Share](#)

http://www.linkedin.com/groups?about=&gid=3186656&trk=anet_ug_grppro

Step 1 – High quality DNA & RNA extraction and pooling strategy

Genotypic Recommended Workflow



* *Young leaves, Senescence leaves, Root, Fruit, Flower, Seed, Shoot etc.*

** *Heat shock, cold shock, wound, hormone treatment, light exposure etc.,*

Step 1a – High quality DNA & RNA extraction and pooling strategy

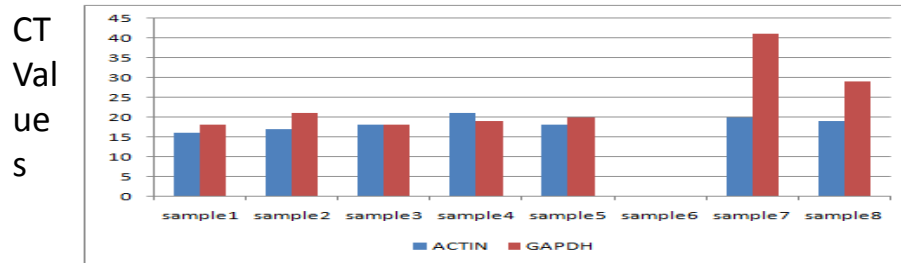
Genotypic Recommended Workflow

Bioanalyzer QC Passed RNAs



qRT-PCR for GAPDH and ACTIN for validation of RNA samples.

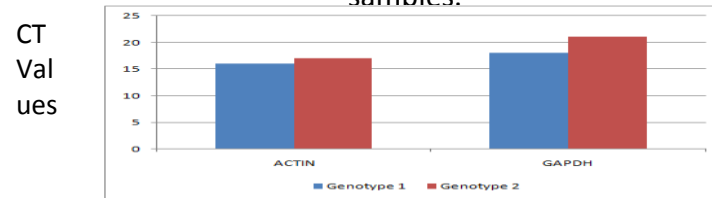
Pool RNAs which pass qRT-PCR tests.



Pool RNAs from Genotype 1 which pass qRT-PCR tests.

Pool RNAs from Genotype 2 which pass qRT-PCR tests.

qRT-PCR for GAPDH and ACTIN for validation of RNA samples.



Organism	Technology	Read Length (b)	Application	???
Arabidopsis	Illumina	50	Targeted Re-sequencing and variation analysis	Re-Sequencing
Bacillus indicus	Illumina	55	WGS	Re-Sequencing
Bovine	Illumina	75	Targeted Re-sequencing and variation analysis	Re-Sequencing
Brassica	454	>500	Transcriptome Sequencing	Denovo
Candida albicans	Illumina	55	WGS, variation analysis	Re-Sequencing
Chickpea	Illumina	72	Transcriptome Sequencing	Denovo
Cucurbitacea family	Illumina	35	small RNA	Re-Sequencing
Escherichia coli	Illumina	35	Variation Analysis	Re-Sequencing
Homo sapiens	Illumina	50	Targeted Re-sequencing and variation analysis	Re-Sequencing
Homo sapiens	Illumina	54	Targeted Re-sequencing and variation analysis	Re-Sequencing
Homo sapiens	Illumina	36	ChIP-Seq	Re-Sequencing
Labeo rohita	Illumina	54	WGS	Denovo
Lentil	Illumina	72	Transcriptome Sequencing	Denovo
Leptospira	Illumina	57	WGS	Re-Sequencing
Mus musculus	Illumina	36	ChIP-Seq	Re-sequencing
Mycobacterium tuberculosis	454	->500	WGS	Re-Sequencing
Oryza sativa	Illumina	72	Transcriptome Sequencing	Denovo
Pichia pastoris	Illumina	75	WGS	Re-Sequencing
Plasmodium vivax	Illumina	50	WGS	Re-Sequencing
Rhodospiridium toruloides	Illumina	72	WGS	Denovo
Sorghum bicolor	454	>500	Target Enrichment analysis	Re-Sequencing
Spirulina	Illumina	50	WGS	Re-Sequencing
Staphylococcus aureus	Illumina	54	Variation Analysis	Re-Sequencing
Tobacco	Illumina	57	Transcriptome Sequencing	Denovo (Semi)
Melon	Illumina	36	Small RNA - Virus induced	Re-Sequencing

How dynamic is the NGS field?

- **Roche, IBM Co-Developing Nanopore-based DNA Sequencing Technology**
- July 01, 2010

No Gold standard for analysis

No Gold standard for de novo whole genome sequencing

Reach us for any NGS project

Library prep

Targeted sequence Capture

Sequencing

NGS Analysis

Thank you

www.genotypic.co.in

www.NextGenSeq.com