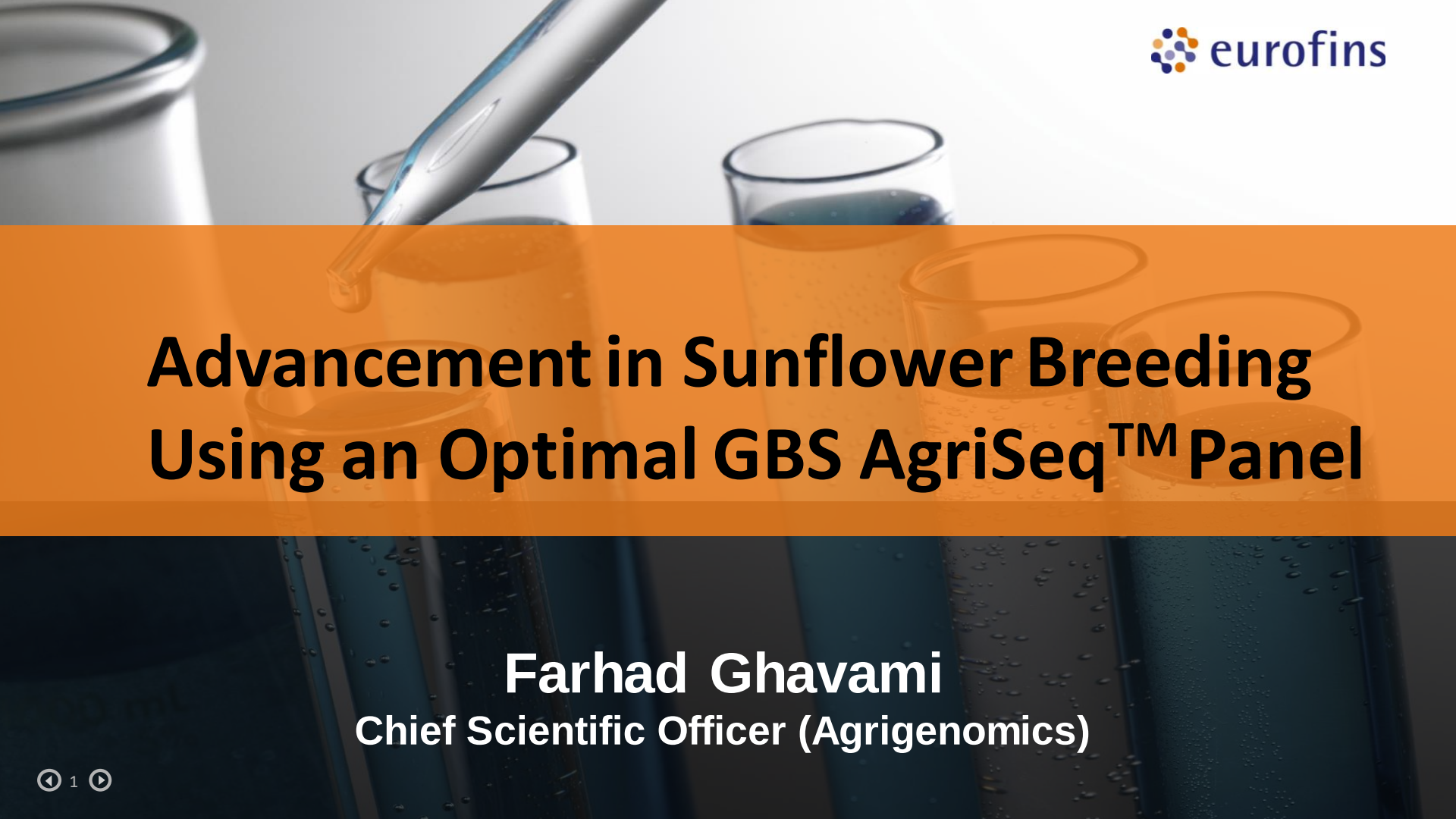


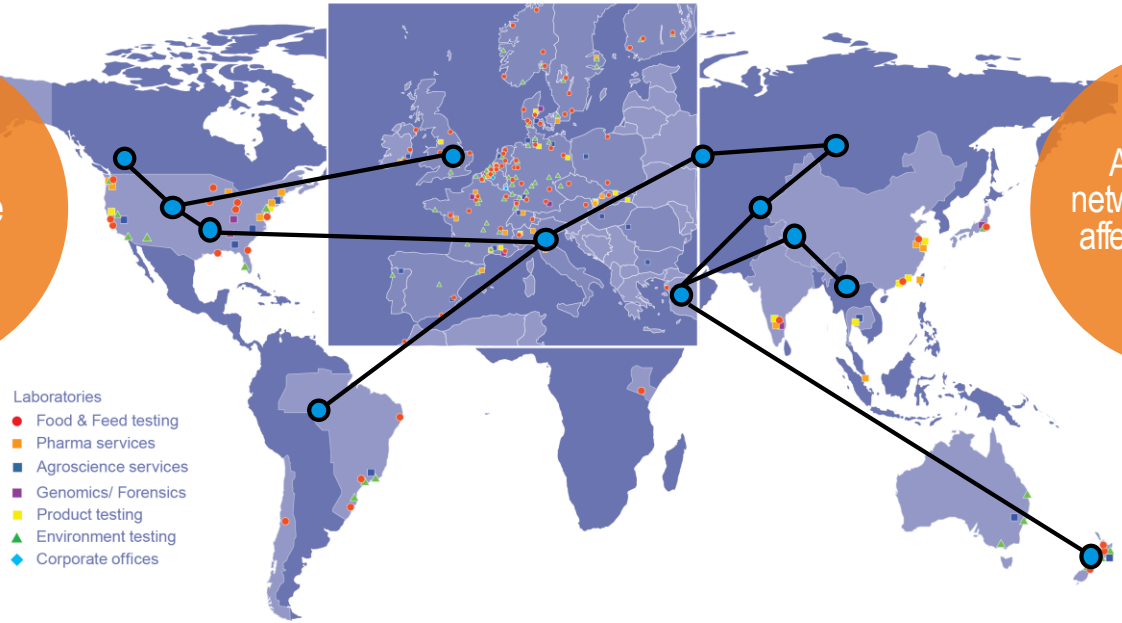
Advancement in Sunflower Breeding Using an Optimal GBS AgriSeq™ Panel

Farhad Ghavami
Chief Scientific Officer (Agrigenomics)

Our History: Rooted in Tradition and Science

Our Global Footprint

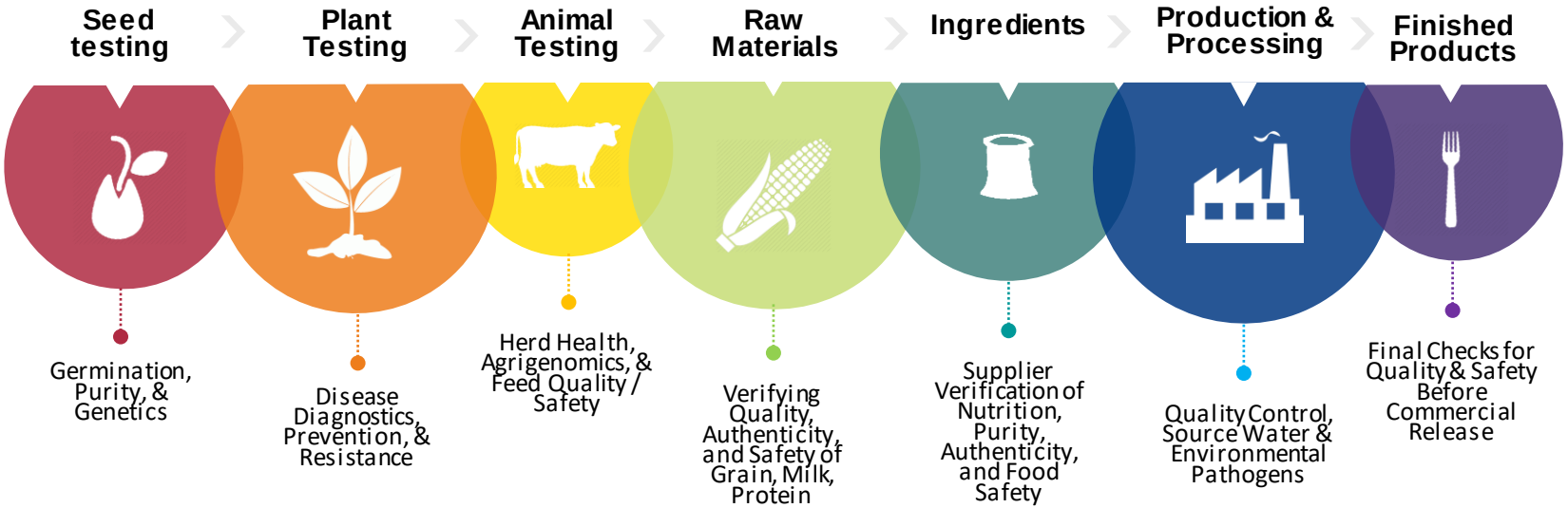
Local competence centers



A connected network for issues affecting multiple markets

Quality Analysis at Every Stage

To fulfill our mission, Eurofins is committed to providing value from farm to fork.



Eurofins BioDiagnostics

Through our portfolio, personalized service, quality systems, and industry expertise, Eurofins offers our clients a unique advantage that other testing companies cannot provide.

Eurofins BioDiagnostics
Service Areas

Seed Analysis
- Germination
- Vigor
- Physical Purity
- Seed treatment

Seed Health
- Viruses
- Bacteria
- Bacterial Fruit Blotch
- Mycology

Seed Genetics
- DNA Fingerprinting
- Trait Mapping
- Marker-Assisted Breeding & Backcrossing
- Genomic Selection

Plant Health
- Grapevine Diseases
- Custom Diagnosis
- Resistance Screening
- Bioreba Kits and Reagents

Agrigenomics
- Genetic Purity and Authenticity
- Variety ID
- Genotyping
- Genetic based ID

Other Services
- Microscopy
- Fatty Acid Profiling
- Adventitious Presence
- Consulting

Agrigenomics services
Genotyping Solutions

EBDI provides solutions for low, medium, high and ultra high density Genotyping for any plant breeding needs

High Density SNP Genotyping

Applied Biosystems ™ GeneTitan for Axiom array

illumina iScan for infinium array

Low Density SNP Genotyping

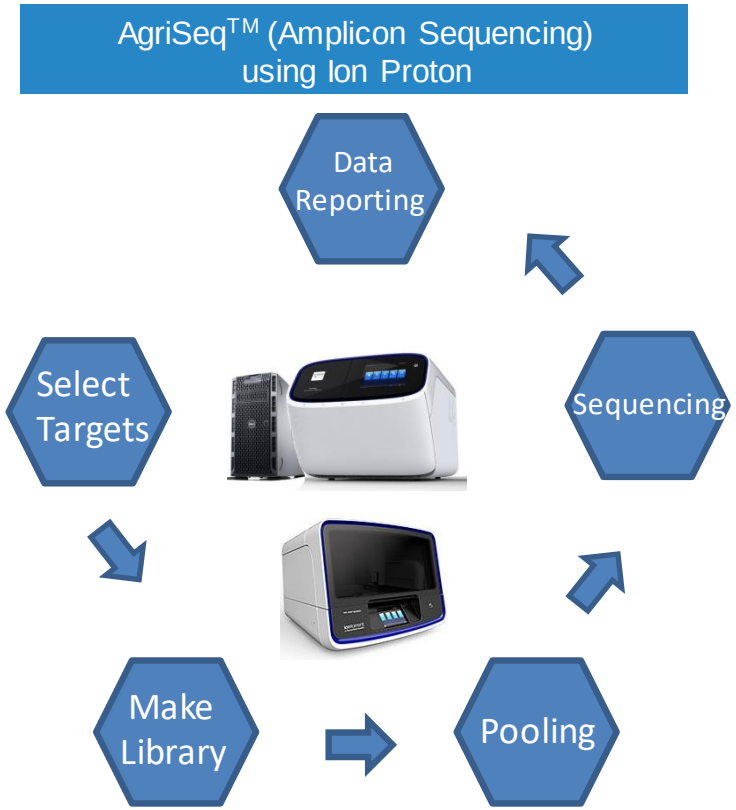
TaqMan™ and Applied Biosystems ™

Array Tape Platform

AgriSeq™: Medium Density Genotyping

- Genotype samples with little input DNA
- Target 100 to 3000 of genes or markers in a single run (Medium Density Genotyping)
- Use Pre-Designed Panels or create customized panel using Ion AmpliSeq Designer

  - **Fast TAT**
  - **Easy custom panel design**
  - **Very cost effective**



Our Exprience with AgriSeq™ panels

Bovine ISAG 200 SNP panel

Porcine 1500 SNP panel

Other species are in the pipeline

Conifer Custom SNP panel

Sunflower 700 SNP panel

Why Sunflower?

- There is no high or medium density SNP panel available for sunflower since the one that was created in 2010
- The number of markers and the cost of the genotyping using the arrays were prohibitive for the day to day breeding selections
- Targeted Genotyping by sequencing (TBGS) with medium size panels of 500 to 1000 SNPs has been used extensively in other crops like corn and soybean not much in sunflower.

"Eurofins BioDiagnostics partnered with Applied Biosystems (Thermo Fisher) to develop AgriSeq™, a genotyping by sequencing (GBS) panel for sunflower which can help the sunflower breeders with their QTL mapping, marker assisted backcrossing and genomic selection projects"

SeedWORLD

History: EBDI collaborations with NSA

- 6 diverse lines was used
- 105 M variants found
- 16467 variants fit illumina bead array criteria

De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach

Venkatramana Pegadaraju, Rick Nipper, Brent Hulke, Lili Qi and Quentin Schultz

BMC Genomics 2013 **14**:556

Categories	Idenifyed SNPs/InDels	# Single Bead Assay
Fixed variants	8313	6323
> 1SNP/contigs	5361	2072
RAD clustering to common EST	1167	430
Het variants	1557	1175
Total	16398	10000
Synthesis failure		-1277
Final set		8723

History: EBDI collaborations with NSA

- 1291 samples were genotyped
- A consensus map of 5019 SNP markers was created based on 3 populations
- A number of disease resistance genes were mapped in sunflower
- A number of PCR based SNP markers developed for marker assisted selection

OPEN ACCESS Freely available online

PLOS ONE

A High-Density SNP Map of Sunflower Derived from RAD-Sequencing Facilitating Fine-Mapping of the Rust Resistance Gene R_{12}

Zahirul I. Talukder[1], Li Gong[2], Brent S. Hulke[3], Venkatramana Pegadaraju[4], Qijian Song[5], Quentin Schultz[4], Lili Qi[3]*

Theoretical and Applied Genetics
April 2016, Volume 129, Issue 4, pp 741–752 | Cite as

Genetics and mapping of a novel downy mildew resistance gene, *Pl* $_{18}$, introgressed from wild *Helianthus argophyllus* into cultivated sunflower (*Helianthus annuus* L.)

Molecular Breeding
October 2015, 35:196 | Cite as

Map saturation and SNP marker development for the rust resistance genes (R_4, R_5, R_{13a}, and R_{13b}) in sunflower (*Helianthus annuus* L.)

History: Sunflower SNP Consensus Map

Table 1. SNP and SSR marker distributions in the three component maps and the consensus map of sunflower.

	Pop1 (HA 89×RHA 464) F_2					Pop2 (B-line×RHA 464) F_2			Pop3 (CR29×RHA 468) F_2			Consensus map				
Linkage groups	No. of markers			Map length cM	Density cM/marker	No. of SNP markers	Map length cM	Density cM/marker	No. of SNP markers	Map length cM	Density cM/marker	No. of markers			Map length cM	Density cM/marker
	SSR	SNP	Total									SNP	SSR	Total		
LG1	9	285	294	88.56	0.30	337	73.82	0.22	33	54.97	1.67	384	9	393	76.09	0.19
LG2	9	82	91	80.22	0.88	172	73.92	0.43	34	40.59	1.19	214	9	223	81.99	0.37
LG3	6	122	128	105.89	0.83	185	88.99	0.48	145	88.72	0.61	327	6	333	95.33	0.29
LG4	6	142	148	57.45	0.39	94	100.30	1.07	156	108.94	0.70	273	6	279	102.45	0.37
LG5	10	179	189	100.96	0.53	241	91.77	0.38	146	82.09	0.56	374	10	384	91.87	0.24
LG6	2	51	53	48.03	0.91	117	58.80	0.50	67	56.58	0.84	168	2	170	62.99	0.37
LG7	8	62	70	67.28	0.96	72	66.09	0.92	60	55.18	0.92	140	8	148	68.31	0.46
LG8	8	214	222	67.66	0.30	166	62.97	0.38	172	81.40	0.47	320	8	328	75.42	0.23
LG9	10	108	118	106.79	0.91	179	86.50	0.48	228	108.83	0.48	352	10	362	104.60	0.29
LG10	13	386	399	94.84	0.24	437	94.48	0.22	95	76.61	0.81	503	13	516	90.89	0.18
LG11	10	142	152	76.32	0.50	103	88.37	0.86	117	95.86	0.82	246	10	256	99.82	0.39
LG12	8	141	149	19.84	0.13	142	62.96	0.44	98	66.12	0.67	255	8	263	67.00	0.25
LG13	4	44	48	31.19	0.65	248	69.45	0.28	136	77.70	0.57	296	4	300	72.93	0.24
LG14	1	43	44	38.56	0.88	160	78.56	0.49	191	73.63	0.39	285	1	286	76.47	0.27
LG15	5	54	59	53.95	0.91	158	80.28	0.51	95	84.78	0.89	225	5	230	85.46	0.37
LG16	5	77	82	105.80	1.29	223	95.77	0.43	144	107.67	0.75	333	5	338	101.28	0.30
LG17	4	36	40	21.37	0.53	202	97.94	0.48	206	57.52	0.28	324	4	328	90.94	0.28
Total	118	2168	2286	1164.71	0.51	3236	1370.97	0.42	2123	1317.19	0.62	5019	118	5137	1443.84	0.28

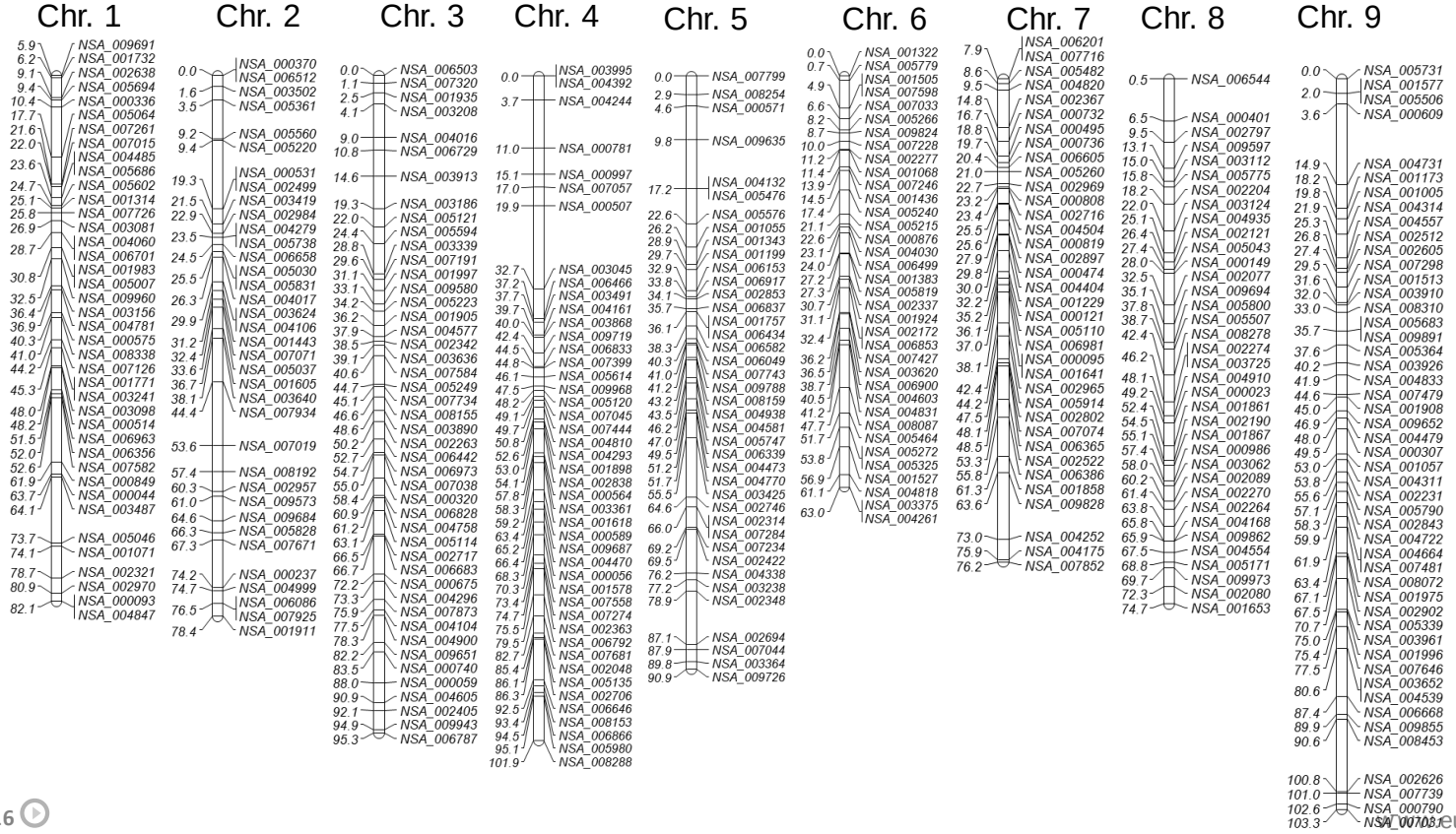
doi:10.1371/journal.pone.0098628.t001

The EBDI Sunflower 700 SNP Panel

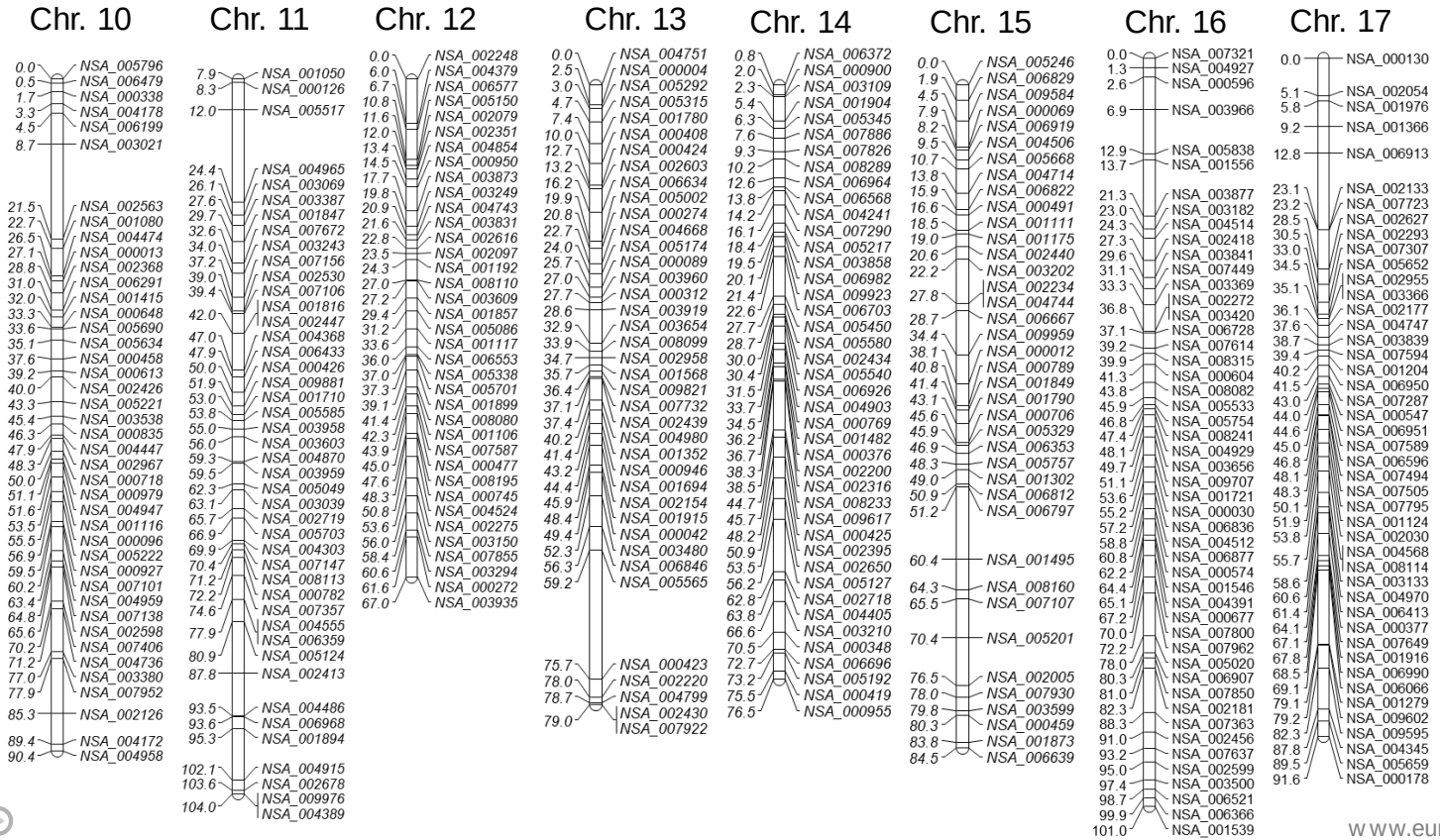
Chromosome	Markers	Chr. Length (CM)	Marker's Distance	PIC
1	40	76.09	1.90	0.30
2	36	81.99	2.28	0.34
3	46	95.33	2.07	0.33
4	45	102.45	2.28	0.33
5	40	91.87	2.30	0.33
6	36	62.99	1.75	0.34
7	36	68.31	1.90	0.33
8	36	75.42	2.10	0.32
9	49	104.60	2.13	0.33
10	42	90.98	2.17	0.32
11	44	99.82	2.27	0.34
12	37	67.00	1.81	0.31
13	39	72.93	1.87	0.33
14	42	76.47	1.82	0.31
15	39	85.46	2.19	0.30
16	49	101.28	2.07	0.34
17	45	90.94	2.02	0.34
Total	701	1443.93		
Average	41.23529412	84.94	2.05	0.33

- 768 Markers were selected from the 5019 markers with consensus map information
- Markers with best performance selected
- Markers with High MAF >0.25 selected
- Markers with high PIC scores >0.25 selected

The Sunflower GBS Panel Consensus Map



The Sunflower GBS Panel Consensus Map

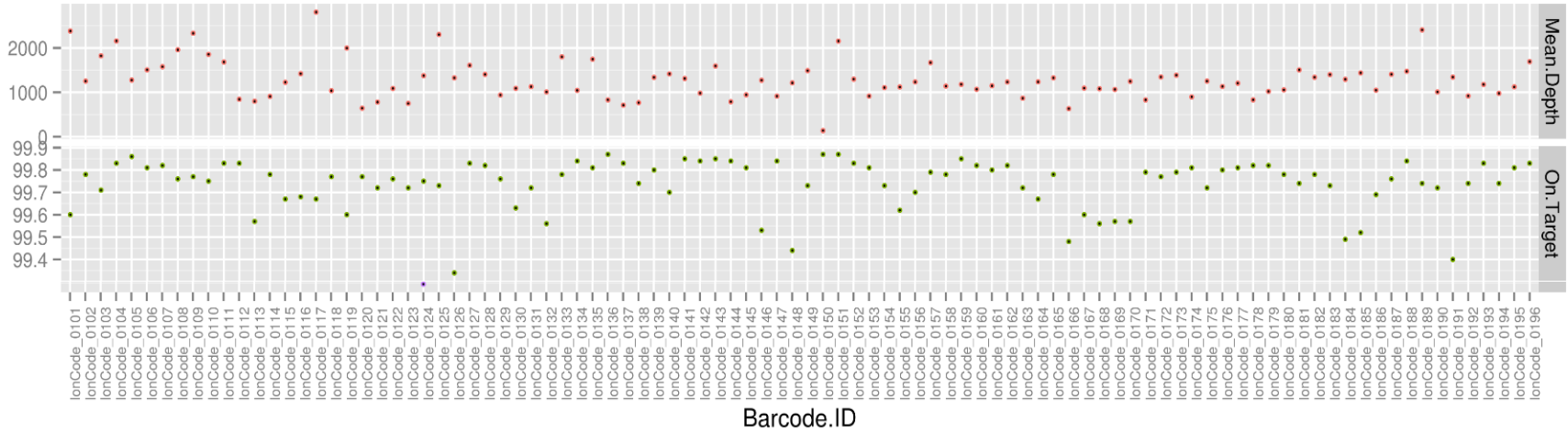


Validation of Sunflower 700 SNP Panel

Study Plan:

- 192 samples were acquired from Germplasm Resources Information Network (GRIN) Thanks to Lisa Burke, Seed storage manager, USDA-ARS-PIRU
- 96 samples were selected for this study (illumina 10K data was available for these samples)
- DNA were extracted from pool of 5 seeds
- A plate of 96 samples were genotyped at EBDI in 2 replications and in ThermoFisher research lab.
- The quality of the panel was analyzed
- The reproducibility of the data was studied
- The concordance of the TBGS data was compared between the two labs
- The concordance of the data was calculated between TGBS and the 2010 illumina array data

Quality Metrics for Sunflower GBS Panel



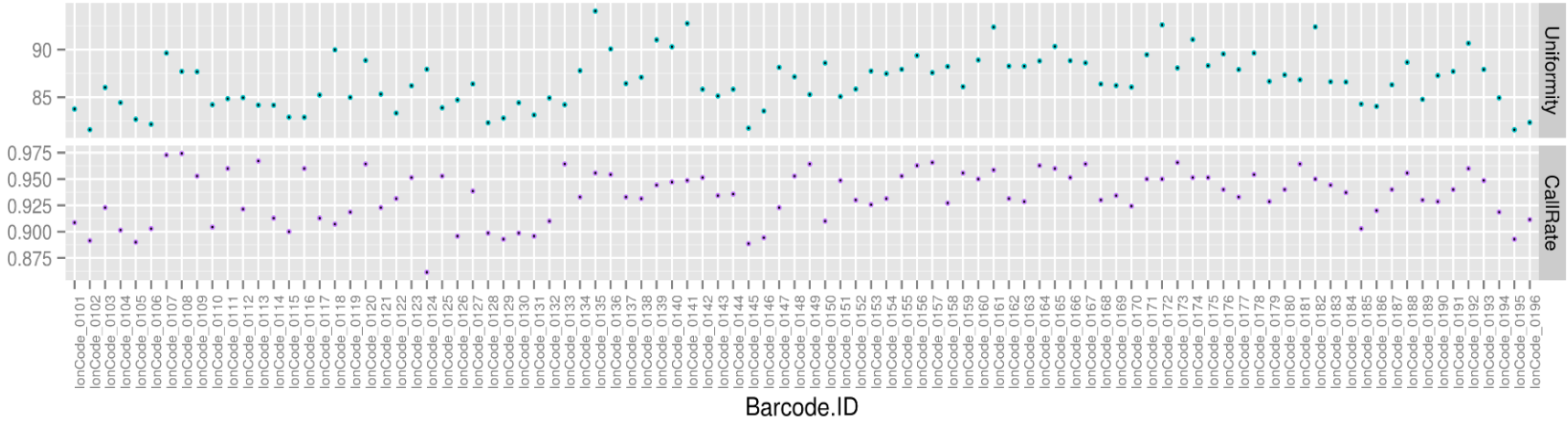
1279X

99.7%

Mean Depth is also known as Coverage. It is a measure of how many reads per amplicon were attained during sequencing.

On Target Reads is a percentage of the mapped reads that were aligned correctly over a target region. This metric infers off-target alignments, which can be indicative of issues such as sample contamination and uncharacterized genetic variation.

Quality Metrics for Sunflower GBS Panel



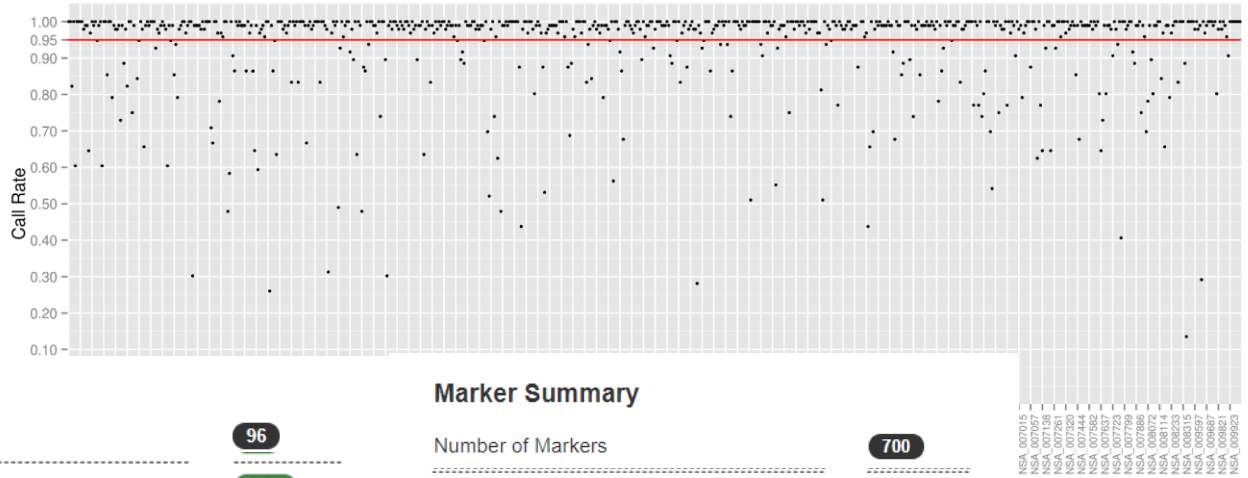
86.7%

93.4%

Uniformity is represented as a percentage of the target bases that had at least 20X read depth. It is a measure of end-to-end coverage over a target region. In our experience, the average value should be greater than 90%.

Sample Call Rate is the percentage of markers generating a genotype call for a specific sample. This metric allows one to see the impact of sample-to-sample variation as a factor in determining marker call rates.

Quality Metrics for Sunflower GBS Panel



93.4%

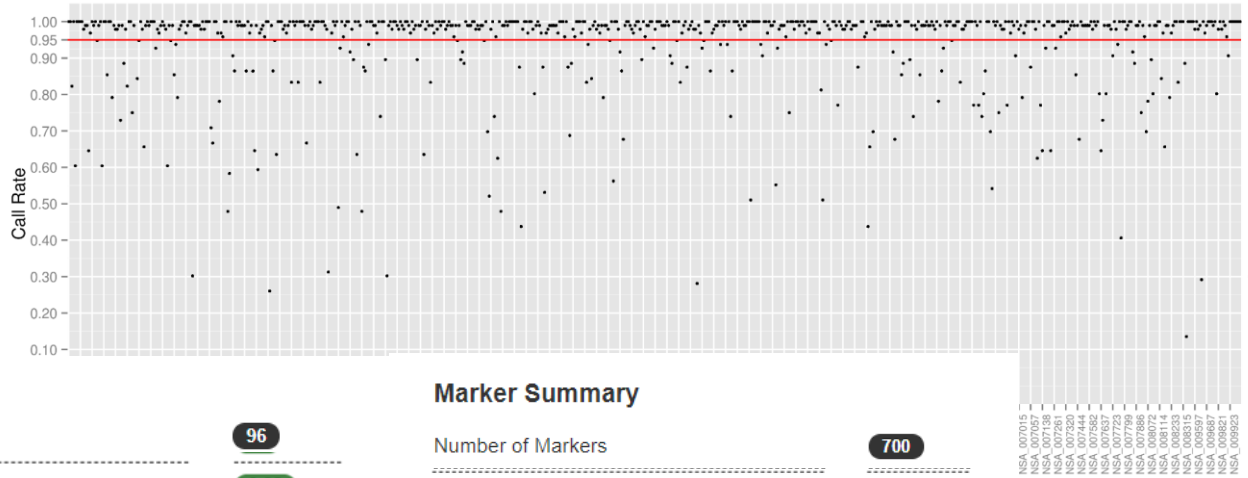
Sample Summary

Samples Run	96
Samples Call Rate	0.93
# Samples = 100% CR	0
# Samples > 90% CR	84
# Samples > 50% CR	96
# Samples = 0% CR	0

Marker Summary

Number of Markers	700
Marker Call Rate	0.93
# Markers = 100% CR	252
# Markers > 90% CR	567
# Markers > 50% CR	683
# Markers = 0% CR	1

Quality Metrics for Sunflower GBS Panel



93.4%

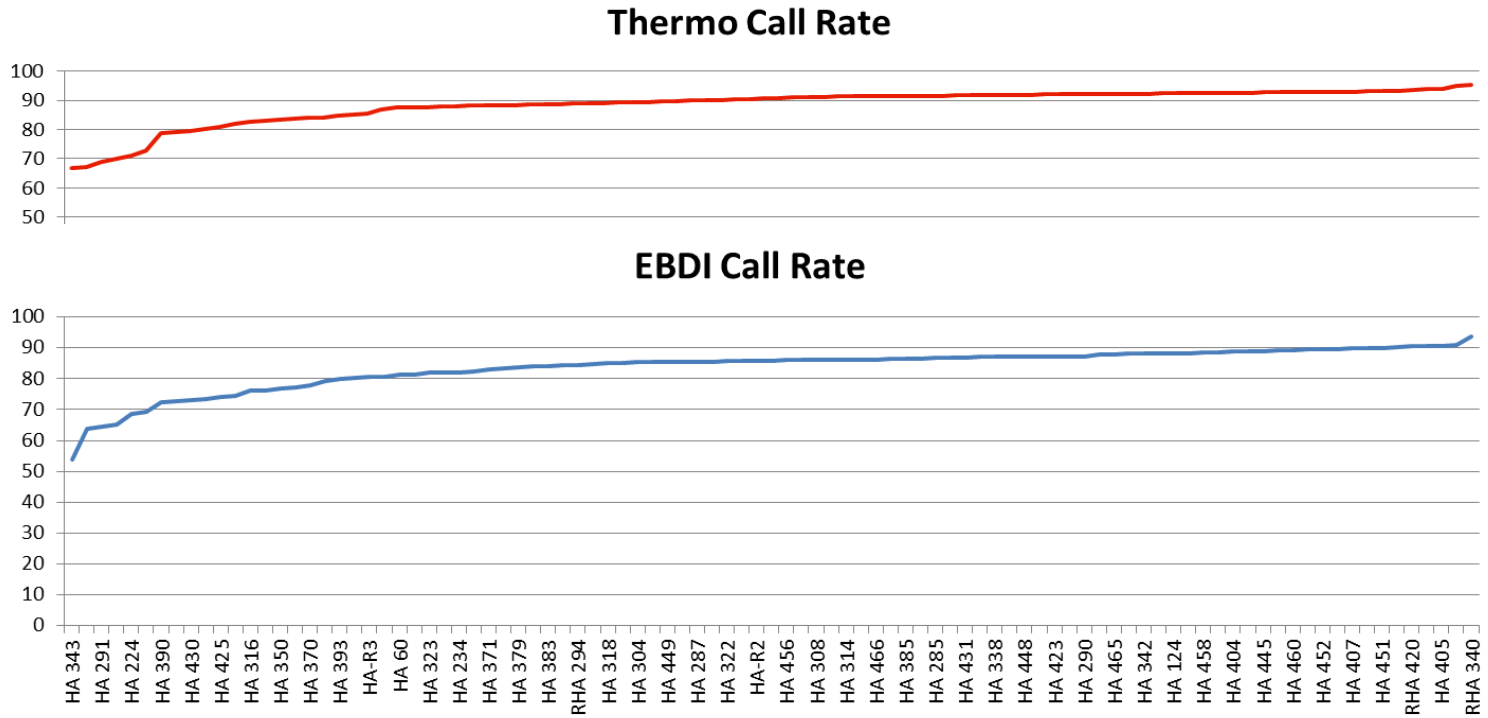
Sample Summary

Samples Run	96
Samples Call Rate	0.93
# Samples = 100% CR	0
# Samples > 90% CR	84
# Samples > 50% CR	96
# Samples = 0% CR	0

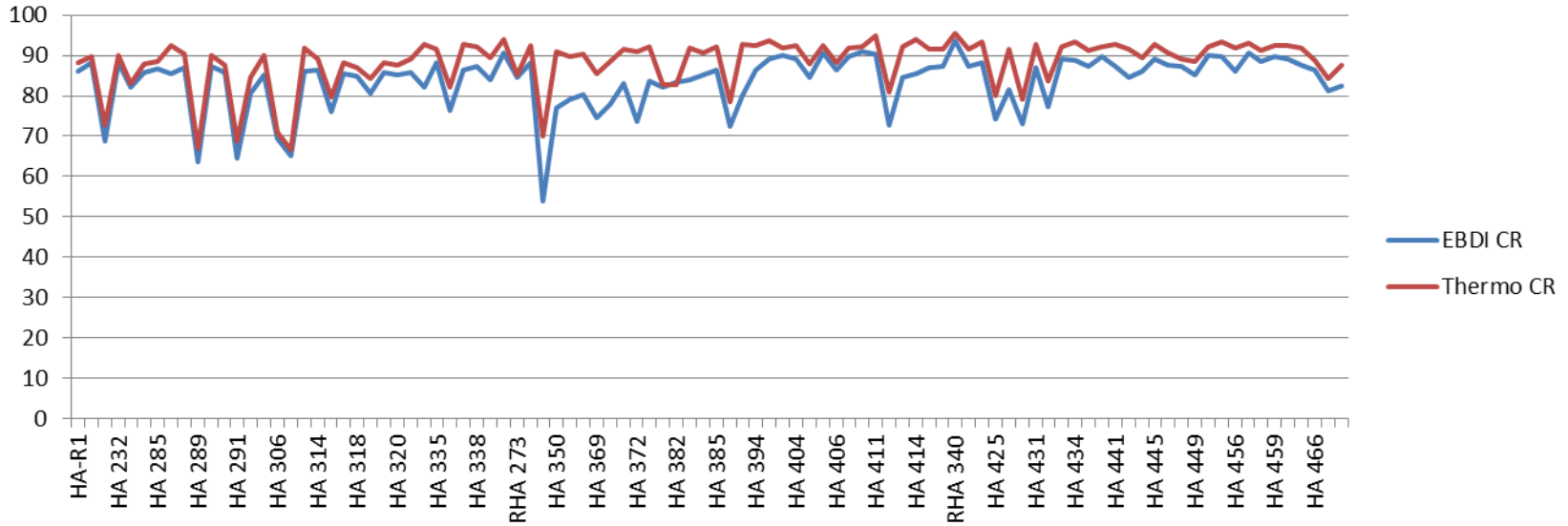
Marker Summary

Number of Markers	700
Marker Call Rate	0.93
# Markers = 100% CR	252
# Markers > 90% CR	567
# Markers > 50% CR	683
# Markers = 0% CR	1

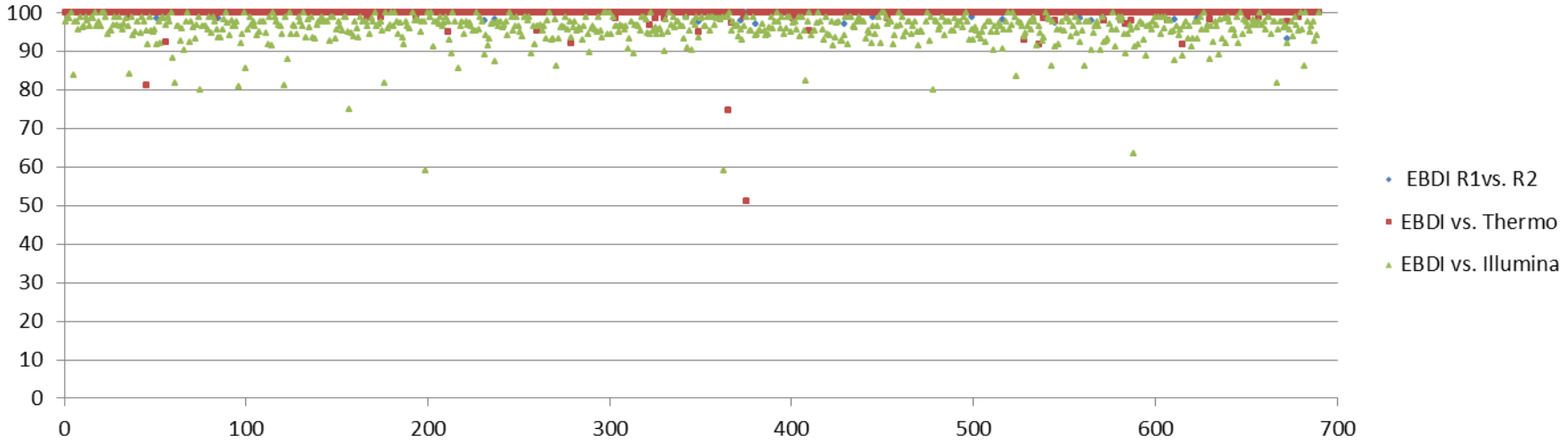
Summary of Sunflower Panel Call Rates



Call Rates/Sample: Indicator of Sample Type/Quality



Concordance (different runs and illumina 10K array)



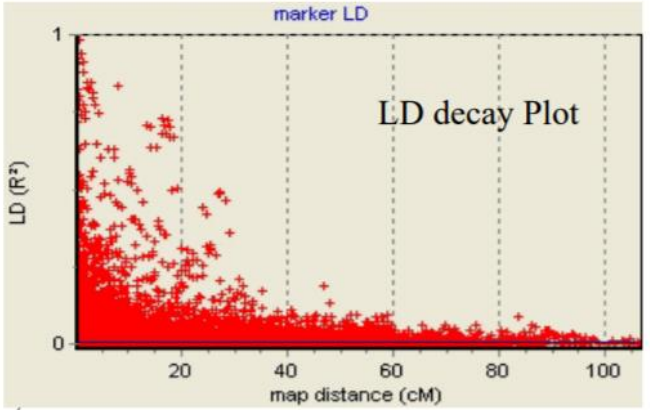
EBDI R1 vs R2 Overall Genotype Concordance = 99.95%

EBDI vs. Thermo TGBS run Overall Genotype Concordance = 99.72%

EBDI vs. illumina 10K array Overall Genotype Concordance = 96.18%

LD in Sunflower

LD decay is highly variable in different species depending to the mating systems, breeding history, genomic rearrangements, and historical recombination happened.

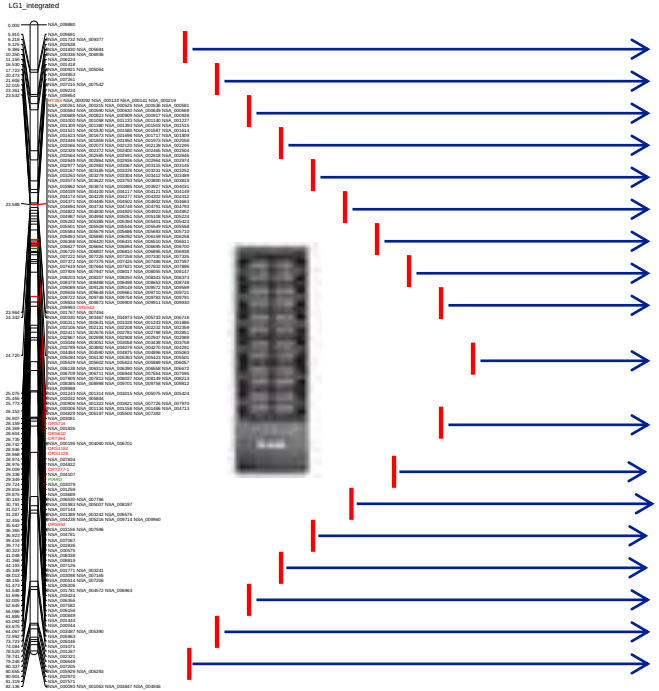


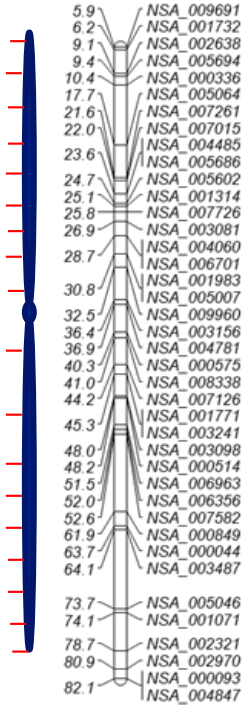
Genotyping 271 diverse samples using illumina 10K array the analysis of LD revealed that in most regions, LD declined quite rapidly as a function of genetic distance. The correlation between most pairs of SNPs dropped rapidly below $r^2 \leq 0.10$ within 3 cM. Mandel J.R. et al. 2013 PLoS Genetics 9:3

In another study using 9480 SNP markers (5788 mapped) the LD decay was observed to be averaged at 2.5 cM window.

Nambessan S. U. 2015 BMC Plant Biol. 15: 84

LD in decay and optimal number of markers





LD=3cM

LG1= 76.9 cM

Total LD blocks= 26

Total Markers = 40

Total Genome Length = 1443.9 cM

Minimum number of Markers = 482

Conclusion

- To design an optimal panel for sunflower breeding 700 markers were selected that covers all sunflower genome with average distance of 2cM.
- A consensus map of the 700 SNP markers is available from their previous location published in 2014.
- A comprehensive study has been done to validate the quality of the panel. The data shows the consistency and reproducibility of the data is very high.
- The genotyping results of AgriSeq™ TGBS panel was 96% in concordance with illumina bead array genotyping data previously produced for the same lines.
- Considering the average LD block size in sunflower that spans around 3cM, the panel will be optimal for any genomic selection or QTL mapping studies in sunflower.

Thank you !

Questions?