



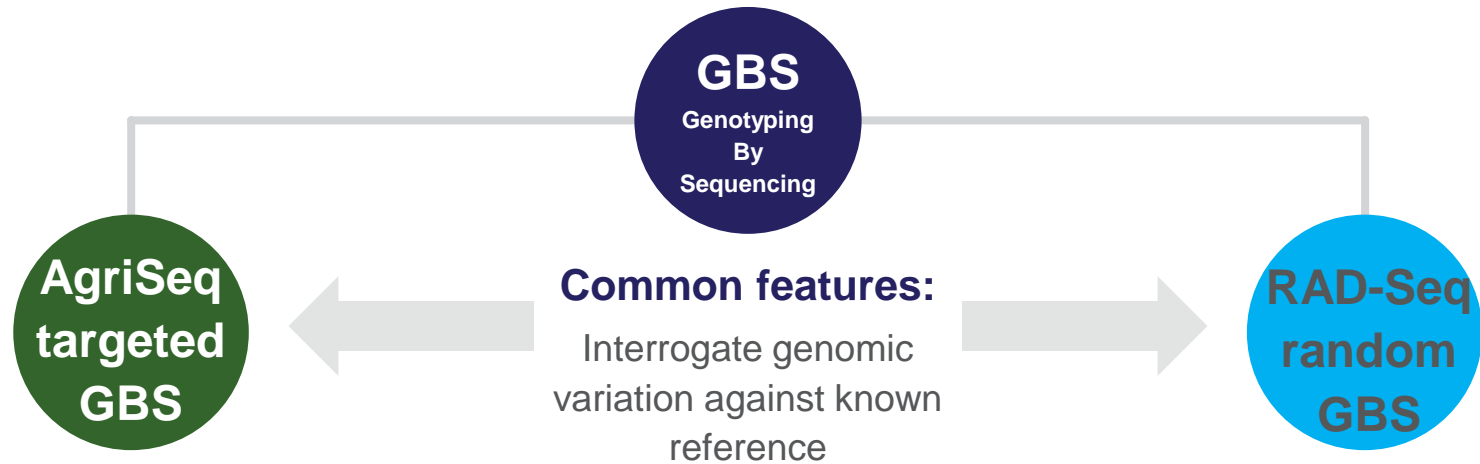
# End-To-End AgriSeq™ Targeted GBS Long INDELs Solution

Chris Willis

Haktan Suren

- AgriSeq™ targeted genotyping-by-sequencing (GBS) technology
- AgriSeq™ GBS throughput
- Long INDELS
- Long INDEL Design Strategy
- INVERSION Design Strategy
- Complete AgriSeq Sequencing Workflow
- Case Study

# AgriSeq vs. other GBS (Genotyping by Sequencing) Solutions



Mapping and/or screening	<b>Best fit</b>	Discovery
Amplicon	<b>Mediated by</b>	Restriction enzyme
Targeted—user selected	<b>Type of markers</b>	Random—Non reproducible
High >95%	<b>Consistency of markers called between samples</b>	Med (20–80%)
Up to 5,000	<b>Number of SNPs interrogated/sample</b>	Up to 100,000s

# AgriSeq™ Targeted GBS Solutions for Agriculture

AgriSeq is a powerful, customizable, flexible and cost-effective high throughput Targeted GBS workflow capable of rapidly genotyping 50 - 5000 markers across plant and animal species.

**>100 custom panels  
for over 30  
species and counting**

## Animals

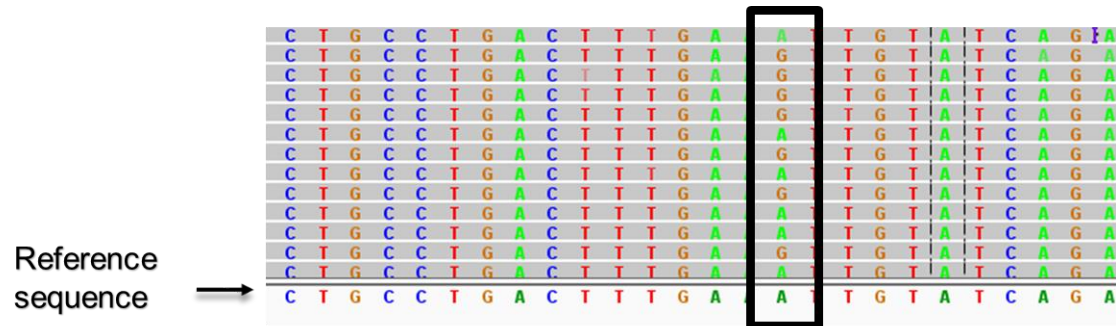
- Bovine
- Porcine
- Equine
- Ovine
- Canine
- Feline
- Deer
- Chicken
- Tuna
- Trout
- Turbot
- Salmon
- Black Soldier fly

## Plants

- Barley
- Cacao
- Canola
- Corn
- Cotton
- Cucumber
- Eucalyptus
- Lettuce
- Oats
- Onion
- Pine
- Potato
- Rice
- Sorghum
- Soybean
- Spinach
- Spruce
- Sunflower
- Tomato
- Wheat
- Watermelon

# AgriSeq™ targeted genotyping-by-sequencing (GBS) technology

- High throughput, Customizable, Cost effective
- Support different types of markers
  - single nucleotide polymorphisms (SNPs),
  - multiple nucleotide polymorphisms (MNPs),
  - structural variants (e.g. inversions, duplications),
  - insertions and deletions (INDELs) up to 50nt in length.



- However, genotyping long INDELs present unique challenges when using NGS technology

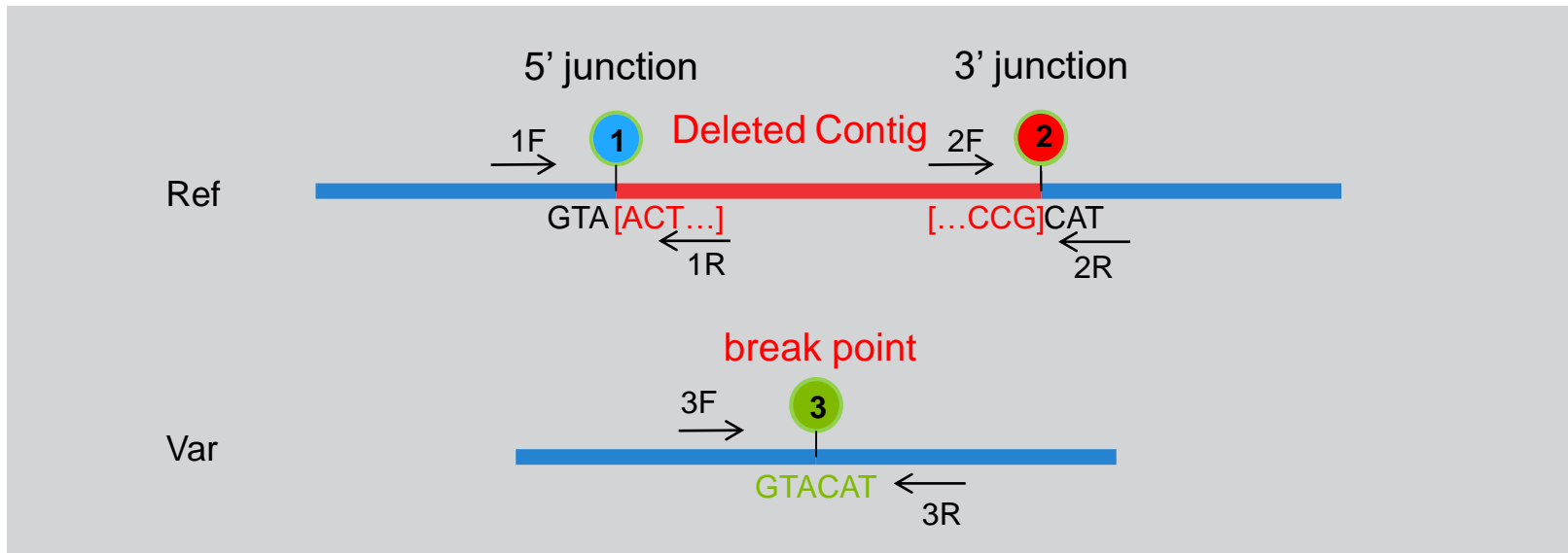
# What are Long INDELS?

- Chromosomal structural changes can cause problems with growth, development, and function of the body's systems. The effects of structural changes depend on their size, location and the dosage (allele frequency).
- Structural variants and Insertion-deletion mutations longer than 50 bp are treated as long variants.
- Long INDELS have unique design and analysis Challenges
  - Variable lengths of the INDELS
  - Long INDEL positions are not fixed
  - Standard variant caller tools are not suitable for the long INDEL detection

# Long INDEL Design Strategy & Analysis

- Three amplicon strategy facilitates genotype calls from both alleles.

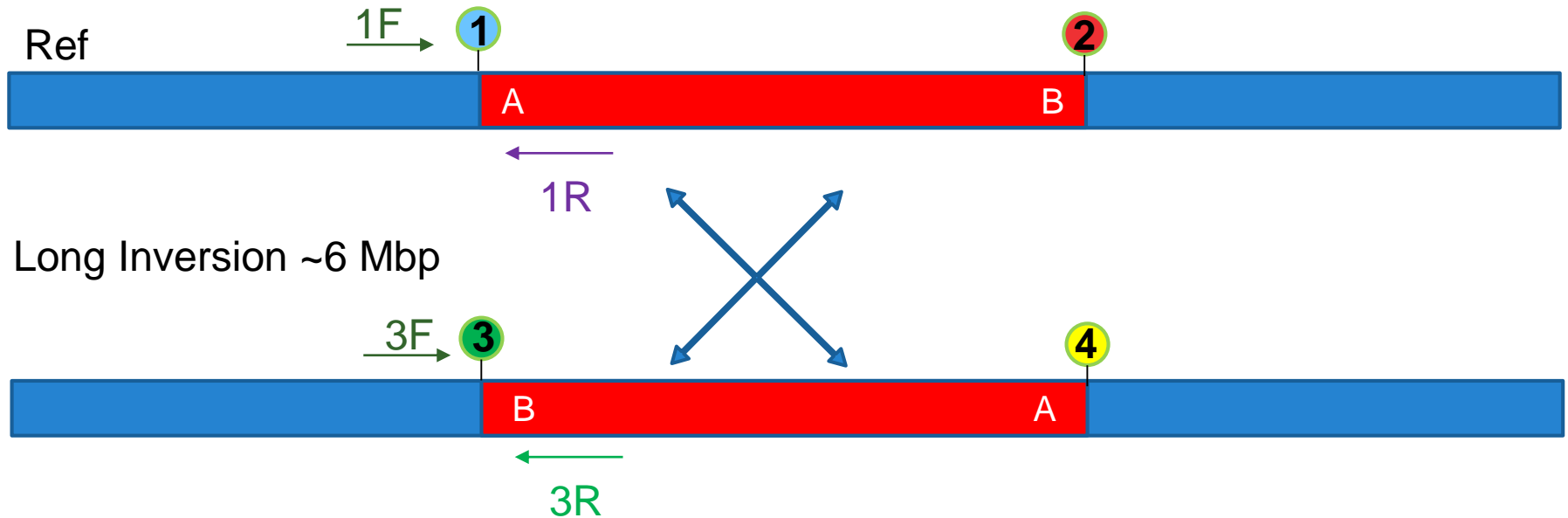
Deletion      Var\_01    GCCAACAGTA[ACTAGC.../-]CATAGCTCTG



- New pipeline to automate the generation of present/absent calls

# INVERSION Design Strategy

- 4 possible targets.



- Only two of them, 1+3 or 2+4 or 1+4 or 2+3, are needed to detect the inversion.

① + ③ → Inversion Present      ② + ④ → Inversion Present  
① + ④ → Inversion Present      ② + ③ → Inversion Present



# Long INDEL Variant Calling Pipeline

Calculate  
Target  
Coverage






Calculate  
Allele  
Frequency

Genotype  
Calling

# Long INDEL Solutions: Case Study

- Custom Canine SNP Genotyping panel
  - Automated design process
  - Optimized for GC content, melting temperature ( $T_m$ ), secondary structure, uniqueness etc.
- 22 Long INDEL markers with varying insert sizes (62 to 6.5Mb) included
  - Size range: 62bp to >6 Mbp
- 384 Samples tested in the standard Agriseq Workflow
- Results compared to Orthogonal data where possible.

# AgriSeq Targeted GBS Sequencing Workflow

Customize targets	Construct library	Prepare template	Run sequence	Analyze data
<b>Sample-to-results in 2–3 days*</b>				
 <p>Custom bioinformatics design service</p>	 <p>AgriSeq library kits Ion Torrent™ IonCode™ Barcode Adapters</p>	 <p>Ion Chef System</p>	 <p>Ion GeneStudio sequencer with Ion Chip kit</p>	 <p>Torrent Suite™ Software</p>
<b>Hands-on time</b>	<3 hr (manual) <1 hr (automated)	<15 min	<15 min	<15 min
<b>Total time</b>	6–7 hr	Overnight	2.5 hr	6–24 hr

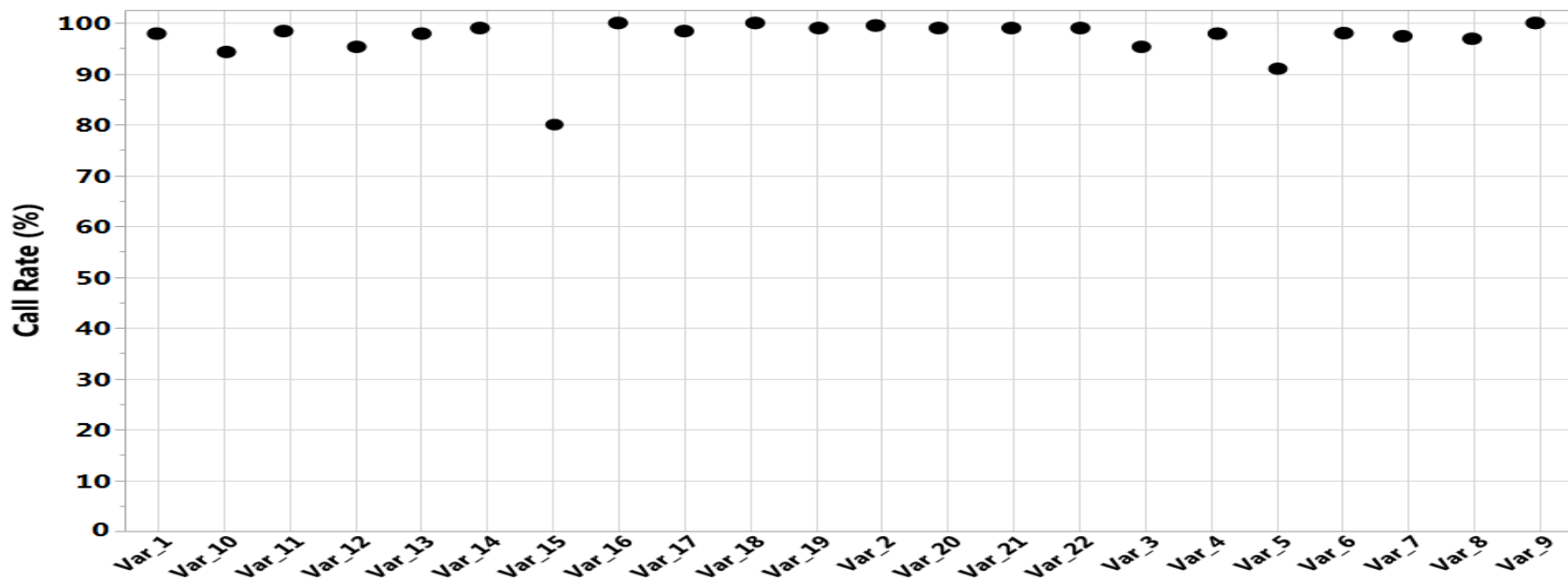
\* Depends on customer implementation or varies depending upon system used and samples/chip.

**Workflow is achievable with 1 FTE**

**Highly scalable, with automation to help reduce hands-on time**

# 100% concordance comparing to truth data, 97% avg call rate

Approximate Call Rate					Approximate Call Rate				
Name	Type	Length	(%)	Concordance	Name	Type	Length	(%)	Concordance
Var_01	DEL	8 Kb	97.9	29/29 (100%)	Var_12	INS	230 bp	95.3	NA
Var_02	DEL	15 Kb	99.5	7/7 (100%)	Var_13	DEL	300 bp	97.9	NA
Var_03	DEL	130 Kb	95.3	4/4 (100%)	Var_14	INS	3 Kb	99	NA
Var_04	DEL	400 Kb	97.9	5/5 (100%)	Var_15	DEL	3.5 Kb	80	NA
Var_05	DEL	6 Mb	91	1/1 (100%)	Var_16	DEL	4 Kb	100	NA
Var_06	INS	60 bp	98	NA	Var_17	INS	4 Kb	98.4	NA
Var_07	INS	80 bp	97.4	NA	Var_18	DEL	10 Kb	100	NA
Var_08	INS	150 bp	96.9	NA	Var_19	DEL	15 Kb	99	NA
Var_09	INS	170 bp	100	NA	Var_20	DEL	18 Kb	99	NA
Var_10	INS	180 bp	94.3	NA	Var_21	DEL	40 Kb	99	NA
Var_11	DEL	180 bp	98.4	NA	Var_22	INV	5 Mb	99	NA



# Conclusions

- AgriSeq™ Genotyping By Sequencing enables the NGS detection of large INDELs (>50bp) marker type
- AgriSeq™ Genotyping By Sequencing design framework works for design detection of large INDELs (>50bp) and Inversions
- Long INDEL variant calling pipeline works for wide range of variant length (62 to 6.5Mb).
- High Call Rate across multiple samples indicates the reproducibility and flexibility of the method.
- High concordance with 'truth' data.

# Acknowledgements

- Thermo Fisher Scientific AgriGenomics Team
  - Haktan Suren (Bioinformatics)
  - Prasad Siddavatam (Bioinformatics)
  - Krishna Reddy Gujjula1 (Bioinformatics)
  - Jason Wall (R&D)
  - Claudio Carrasco (Business)
  - Jeanette Schmidt (Bioinformatics)
  - Rick Conrad (R&D)

For Research Use Only. Not for use in diagnostic procedures. AgriSeq is restricted for use with plants, agricultural animals or companion animals only. This product is not for use with human samples and/or in commercial applications. © 2019 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified.

# Experience the power of AgriSeq with 2 Enabling Options

Stop by Thermo Fisher Scientific (booth #1) to learn about:

## GBS EASY Jump Start bundles

appliedbiosystems



Agrigenomics Ion GeneStudio instrument  
and starter kit promotion

## See more - from STRs to SNPs



Free genotyping of your sample using  
AgriSeq GBS Panels