Advancing Agrigenomic Discoveries with Sequencing & GWAS Research

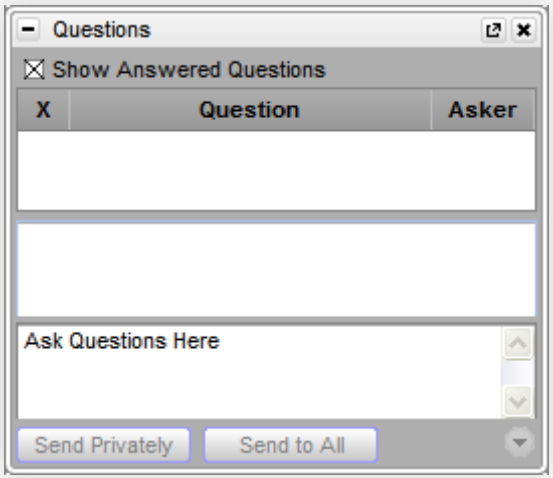
January 8, 2014

Greta Linse Peterson
Director of Product Management and Quality

GOLDEN HELIX
Accelerating the Quest for Significance™

Questions during the presentation

Use the Questions pane in your GoToWebinar window



Agenda

1. Overview
2. GWAS for Agrigenomics
3. DNA-Seq for Agrigenomics
4. Conclusion

Agrigenomic Strategic Relationships

USDA

UC DAVIS
UNIVERSITY OF CALIFORNIA

Cobb®

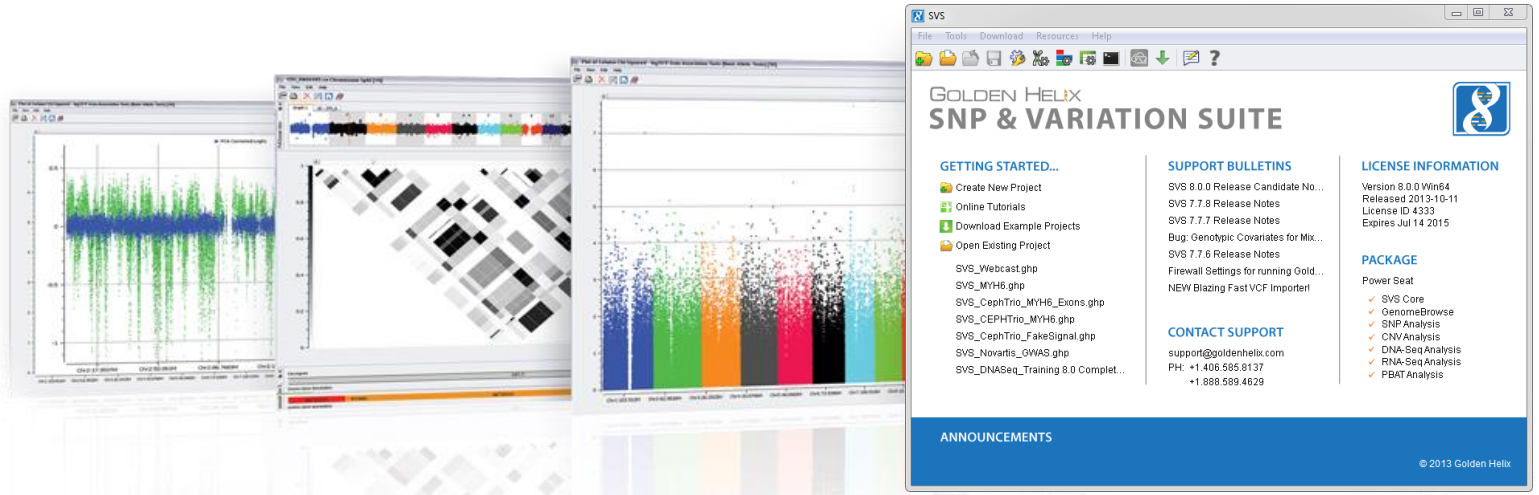
U.K.®
UNIVERSITY OF
KENTUCKY®

CIMMYT MR
International Maize and Wheat Improvement Center

unesp

UNIVERSITY OF
ILLINOIS

SNP & Variation Suite (SVS)



Core Features

- Powerful Data Management
- Rich Visualizations
- Robust Statistics
- Flexible
- Easy-to-use

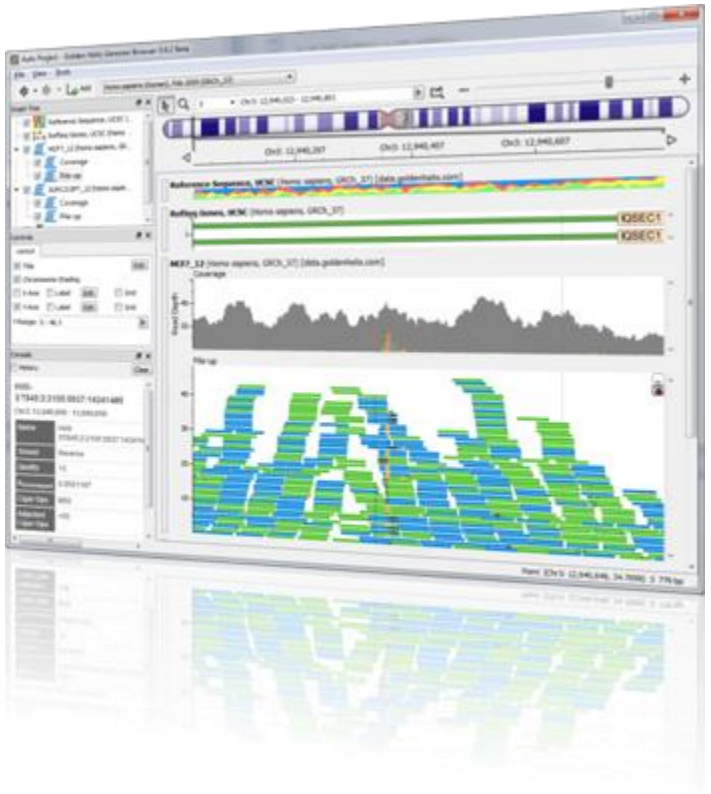
Applications

- Genotype Analysis
- DNA sequence analysis
- CNV Analysis
- RNA-seq differential expression
- Family Based Association

GenomeBrowse

- Free sequencing visualization tool
- Launched in 2011
- Makes the process of exploring DNA-seq and RNA-seq pile-up and coverage data intuitive and powerful
- Stream public annotations via the cloud
- Use it to validate variant calls, trio exploration, de Novo discovery, and more

Golden Helix GenomeBrowse™



Merging of Two Great Products

Species Supported

- Bovine
- Canine
- Equine
- Gallus gallus
- Mus musculus
- Ovine
- Porcine
- Rhesus macaque
- And over 16 more plant and animal genomes








Agenda

1. Overview
2. **GWAS for Agrigenomics**
3. DNA-Seq for Agrigenomics
4. Conclusion

First a little about the dataset…

- Sheep HapMap SNP50_Breedv1 dataset
- Illumina 50k SNP array
- 49,034 markers were left after filtering by the consortium
- 110 unmapped markers
- Only 1 marker in Chr Y

GWAS Workflow

- Phenotype Distribution
- Sample Quality
- SNP Quality
- Population Structure

- Correlation/Trend Test
- Linear Regression with PCA Correction
- EMMAX (Mixed Linear Model Analysis)

- QQ Plots
- Manhattan Plots and Annotation Sources in GenomeBrowse

GOLDEN HELIX

SNP & VARIATION SUITE

[Demonstration]

[Poll]

Agenda

1. Overview
2. GWAS for Agrigenomics
3. **DNA-Seq for Agrigenomics**
4. Conclusion

First a little about the dataset…

- 2 Bovine Partial Exomes + 1 Bison Partial Exome
- All female
- From Sequence Read Archive (SRP007095)
- Illumina Genome Analyzers for sequencing
- GATK for alignment and variant calling
- UMD_3.1 Reference Sequence

Cosart et al. BMC Genomics 2011, 12:347
http://www.biomedcentral.com/1471-2164/12/347

BMC Genomics

METHODOLOGY ARTICLE Open Access

Exome-wide DNA capture and next generation sequencing in domestic and wild species

Ted Cosart[1,2,3*], Albano Beja-Pereira[3*], Shanyuan Chen[3], Sarah B Ng[4], Jay Shendure[4] and Gordon Luikart[3,5]

Abstract

Background: Gene-targeted and genome-wide markers are crucial to advance evolutionary biology, agriculture, and biodiversity conservation by improving our understanding of genetic processes underlying adaptation and speciation. Unfortunately, for eukaryotic species with large genomes it remains costly to obtain genome sequences and to develop genome resources such as genome-wide SNPs. A method is needed to allow gene-targeted, next-generation sequencing that is flexible enough to include any gene or number of genes, unlike transcriptome sequencing. Such a method would allow sequencing of many individuals, avoiding ascertainment bias in subsequent population genetic analyses.

We demonstrate the usefulness of a recent technology, exon capture, for genome-wide, gene-targeted marker discovery in species with no genome resources. We use coding gene sequences from the domestic cow genome sequence (*Bos taurus*) to capture (enrich for), and subsequently sequence, thousands of exons of *B. taurus*, *B. indicus*, and *Bison bison* (wild bison). Our capture array has probes for 16,131 exons in 2,570 genes, including 203 candidate genes with known function and of interest for their association with disease and other fitness traits.

Results: We successfully sequenced and mapped exon sequences from across the 29 autosomes and X chromosome in the *B. taurus* genome sequence. Exon capture and high-throughput sequencing identified thousands of putative SNPs spread evenly across all reference chromosomes, in all three individuals, including hundreds of SNPs in our targeted candidate genes.

Conclusions: This study shows exon capture can be customized for SNP discovery in many individuals and for non-model species without genomic resources. Our captured exome subset was small enough for affordable next-generation sequencing, and successfully captured exons from a divergent wild species using the domestic cow genome as reference.

DNA-Seq Workflow – Unrelated Samples

- Alternate Read Ratio
- Read Depth
- Genotype Quality

- Exon Region Filter
- Examine Variant Sharing Between Samples
- Sample Pattern Filtering
- Variant Classification

- Candidate Variants
- VCFs
- BAMs
- Annotation Sources

GOLDEN HELIX

SNP & VARIATION SUITE

[Demonstration]

Agenda

1. Overview
2. GWAS for Agrigenomics
3. DNA-Seq for Agrigenomics
4. **Conclusion**

Questions or more info:

- Email info@goldenhelix.com
- Request an evaluation of the software at www.goldenhelix.com

Questions?

Use the Questions pane in your GoToWebinar window