

QUALITAS⁺

Swiss Low Input Genetics (SLIG): Breeding functional dairy cows for low-input production systems using whole genome sequence data

SABRE Treffen

04. Juni 2013

QUALITAS⁺

Übersicht

- Das Projekt
 - Sequenzierung
 - Ziel
 - Umfang
 - Partner
 - Zeitplan
- Der aktueller Stand

Gründe für die Sequenzierung



Ziel

Ziel ist es, eine genomische ZWS zu entwickeln basierend auf Sequenzdaten



Projektumfang

- Dauer
 - 3 Jahre
- Kosten
 - 1.2 Millionen CHF
- Partner
 - 5 Partner aus Forschung und Industrie

Partner im „1000 Bull Genomes Project“

- 492 Tiere im aktuellen Run
- > 20 Millionen Varianten (SNP, InDels, CNVs...)



Projektpartner



- Christian Stricker



- Christine Flury, Fritz Schneider, Heidi Signer-Hasler

IOWA STATE
UNIVERSITY

- Rohan Fernando, Dorian Garrick

QUALITAS⁺

- Beat Bapst, Birgit Gredler, Jürg Moll, Urs Schnyder

swissgenetics 

- Stefan Felder-Reiche, Fritz Schmitz-Hsu

1000 bull genomes project

- Hans Daetwyler, Ben Hayes

TUM

- Ruedi Fries, Sandra Jansen



UNIVERSITÀ
DEGLI STUDI
DI MILANO

vetmeduni
vienna 

- Marlies Dolezal

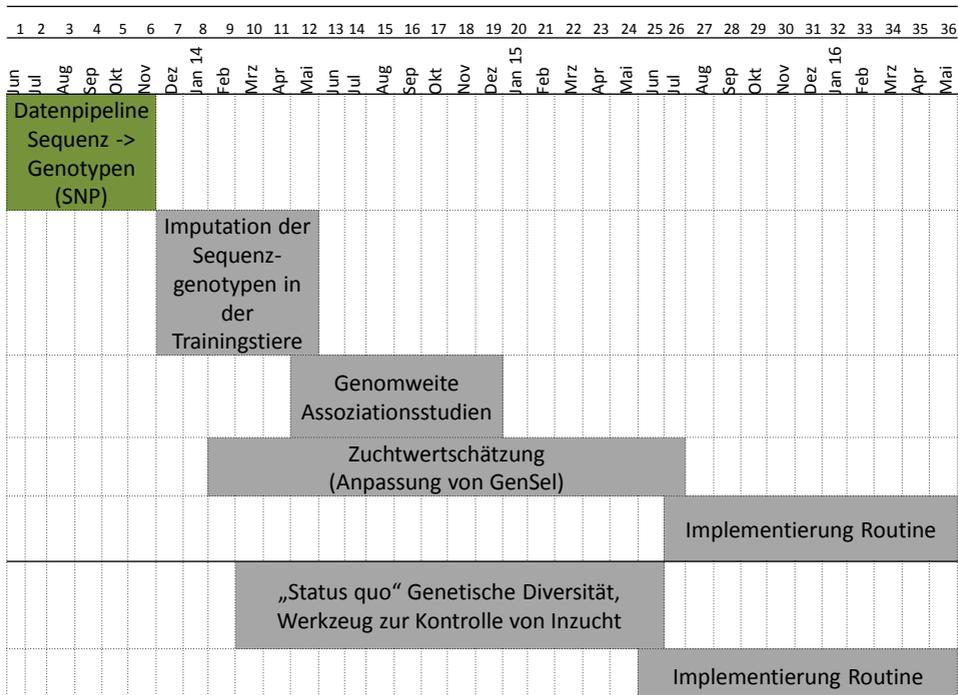
Projektumfang

Subprojekt 1

- (Stierauswahl)
- Datenpipeline: Sequenzinfo → Genotypen (SNP)
- Imputation der Sequenz-Genotypen in Trainingstiere
- Genomweite Assoziationsstudien
- Anpassung der Effektschätzungs-Methoden (GenSel)
- Schätzung von genomische Zuchtwerte anhand von Sequenzinformation
- Implementierung in der Routine-Zuchtwertschätzung

Subprojekt 2

- Werkzeug zur Kontrolle von Inzucht
- Implementierung in der Praxis

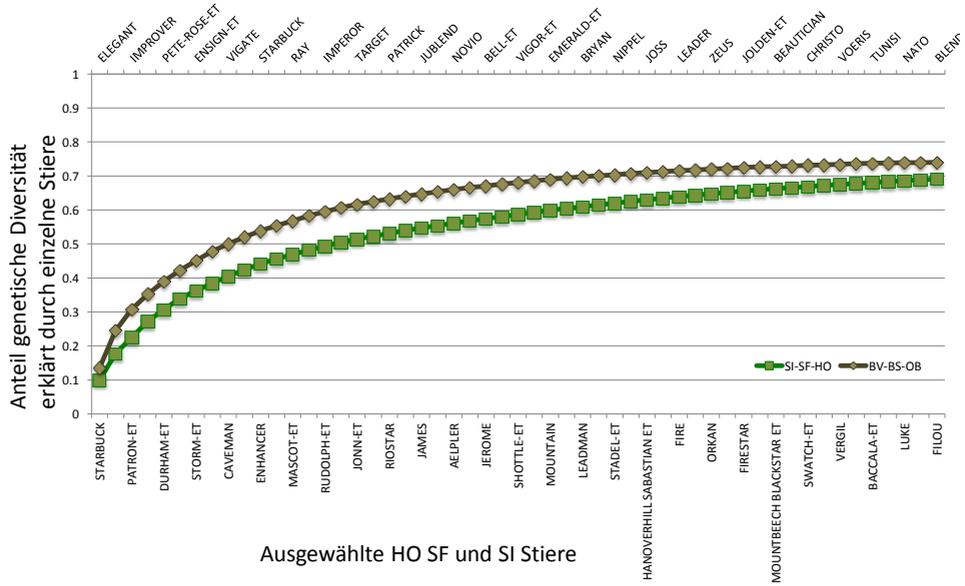




Der aktueller Stand

Stierauswahl

Ausgewählte BS, BV und OB Stiere



Aktueller Stand: Sequenzierung

TUM / Helmholtz Zentrum

- Illumina HiSeq2000, Paired-end reads (101bp) → FASTA
- Alignment zu UMD 3.1 mit Burrows-Wheeler-Aligner
→ SAM
- Indizieren und Sortieren mit Samtools → BAM
- PCR Duplikate markieren mit Picardtools

- Sehr grosse Datenmengen
 - ca. 28 GB / Datei * 60 Stiere = 1.68 TB
- Datentransfer aufwendig (mehrere Tage, MD5 Prüfung)

Grundgerüst der aktuellen Pipeline

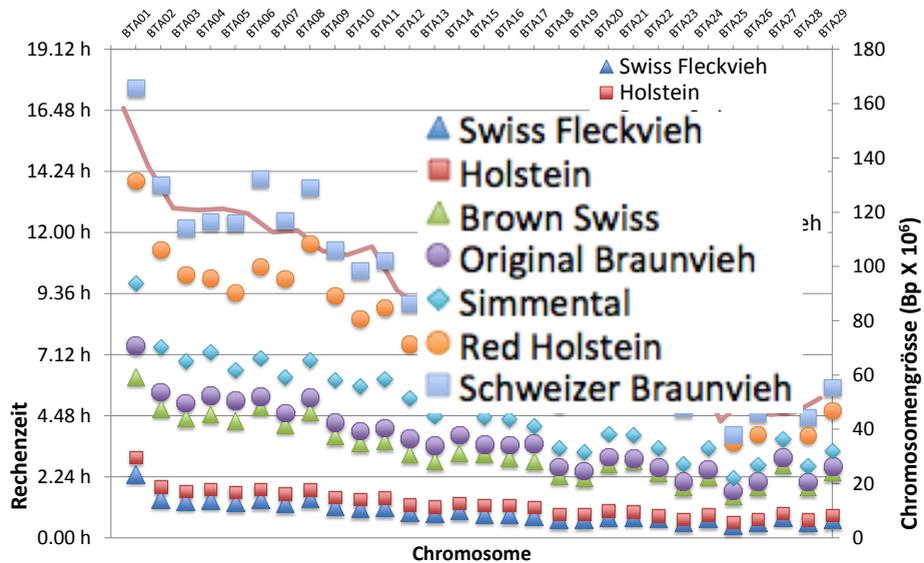
Variant Calling (Single / Multisample):

- Samtools + BCFTools (*Li et al., 2009*)

Software	Aufgabe
Samtools	Sort and index BAM files
Samtools + BCFTools	Variant calling
BCFTools	Filter variants
Samtools	Index Fasta (reference)
Picardtools	Create dictionary for GATK
GATK	Create input file for Beagle
Beagle	Phase haplotypes
	(unzip files)
GATK	Convert Beagle output file to VCF

- GATK (Genome Analysis Toolkit) (*McKenna et al., 2010*)

Rechenzeit: Variant Calling



Variant Calling: Beispiel BTA24

	Original Braunvieh		Average	
	(n=7)		(n=31)	
Variant Calling (Single* or Multi)	Single	Multi	Single	Multi
Total number of variants	164'060	346'462	162'398	456'316
Mean variant read depth (DP Tag)	11	75	11	337
Mean quality score (QUAL)	112	635	114	806
* Single Sample Calling berechnet als Durchschnitt				

QUALITAS⁺

Ausblick

1. Single-sample vs. Multi-sample variant calling
 - Anzahl / Qualität von Varianten
 - Schnittmengen von Varianten (Überlappungen)
 - Allelfrequenzspektrum

2. Samtools vs. GATK
 - Anzahl / Qualität von Varianten
 - Schnittmengen von Varianten (Überlappungen)
 - Rechenzeit

3. Rassenunterschiede

QUALITAS⁺

Danke an Team Qualitas...

Beat Bapst
 Madeleine Berweger
 Andreas Bigler
 Birgit Gredler
 Jürg Moll
 Franz Seefried
 Urs Schuler
 Urs Schnyder



Bild: simplyfantasticbooks.com

... und Danke Ihnen für Ihre
 Aufmerksamkeit

KTI/CTI

ASR
 Arbeitsgemeinschaft Schweizerischer Rindzüchter
 Communauté de travail des éleveurs bovins suisses

Dieses Projekt wird mitfinanziert von ASR und der
 Kommission für Technologie und Innovation