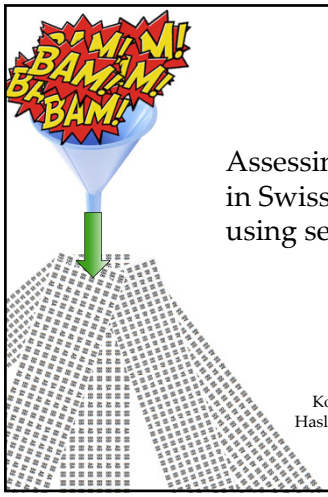
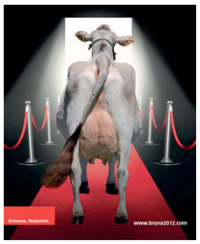


Assessing genomic variation in Swiss dairy cattle using sequence information

C.F. Baes, M.A. Dolezal, E. Fritz-Waters, J.E. Koltes, B. Bapst, S. Jansen, C. Flury, H. Signer-Hasler, C. Stricker, R. Fernando, F. Schmitz-Hsu, R. Fries, D.J. Garrick, J.M. Reecy, B. Gredler

Motivation



Identification of **high-quality** genetic variants from next-generation sequencing data is crucial for downstream analysis...

... but its **really** difficult!

Compare variant identification methods:

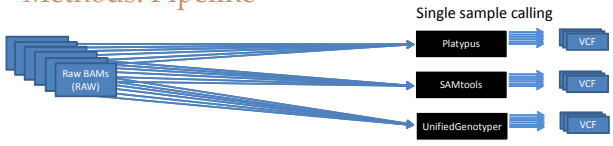
- Pre-calling methods
- Software
- Calling (variant identification) methods

Methods: Sequence Data

66 key ancestors:
- Brown Swiss, Braunvieh, Original Braunvieh
- (Red) Holstein, Swiss Fleckvieh, Simmental

- Illumina HiSeq2000 data
- Paired-end reads
- Duplicates marked
- Genome-wide filtered read depth = 12x

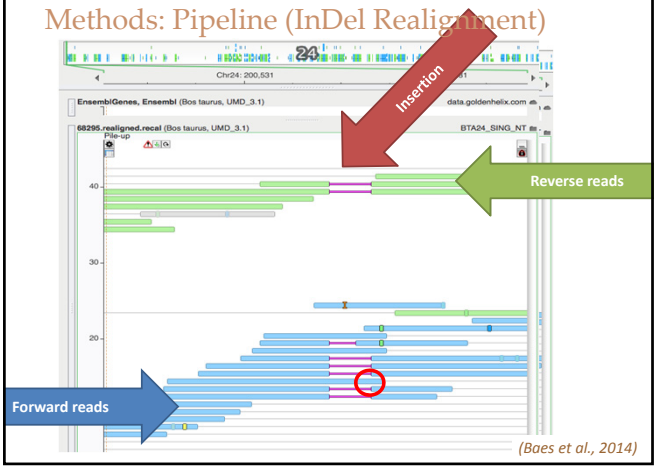
Methods: Pipeline



BAM = Binary alignment mapping file
VCF = Variant calling format file

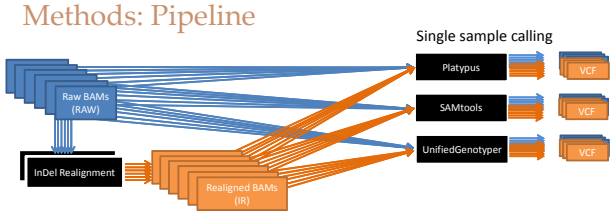
(Baes et al., 2014)

Methods: Pipeline (InDel Realignment)



(Baes et al., 2014)

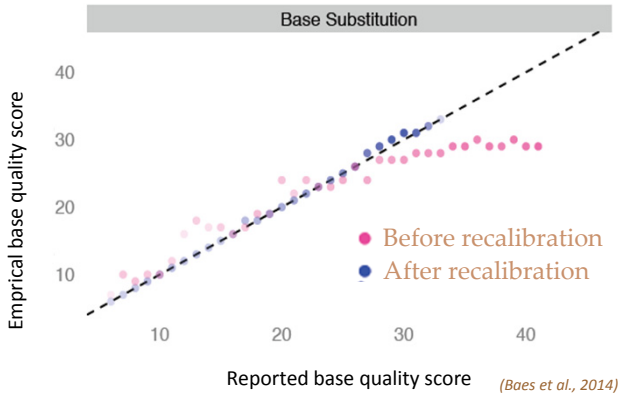
Methods: Pipeline



BAM = Binary alignment mapping file
VCF = Variant calling format file

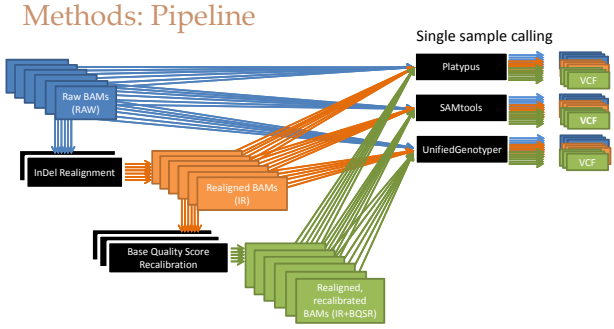
(Baes et al., 2014)

Methods: Pipeline (Base Recalibration)



(Baes et al., 2014)

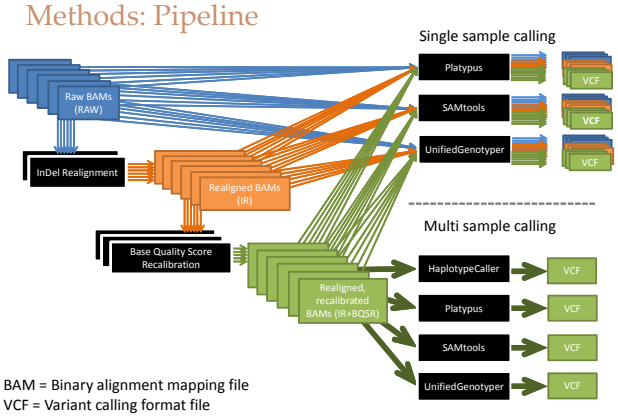
Methods: Pipeline



BAM = Binary alignment mapping file
VCF = Variant calling format file

(Baes et al., 2014)

Methods: Pipeline



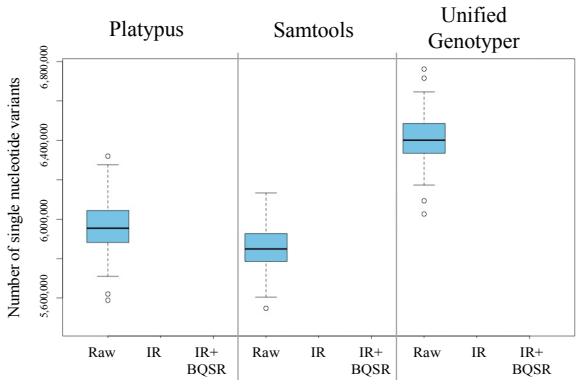
BAM = Binary alignment mapping file
VCF = Variant calling format file

(Baes et al., 2014)

Results:

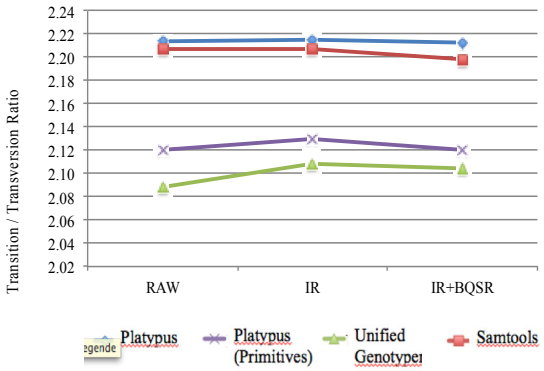
Pre-calling methods (single sample)

Results: Pre-calling methods



(Baes et al., 2014)

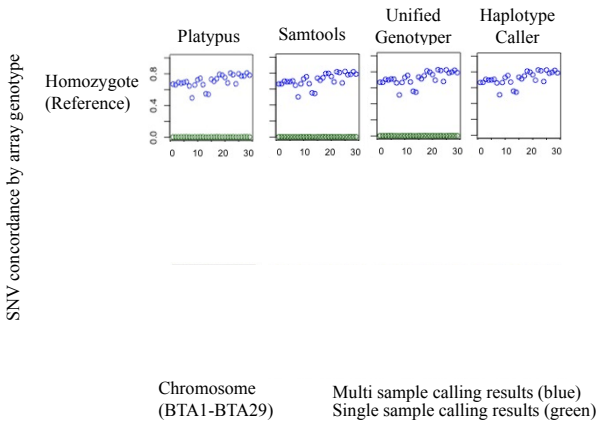
Results: Pre-calling methods



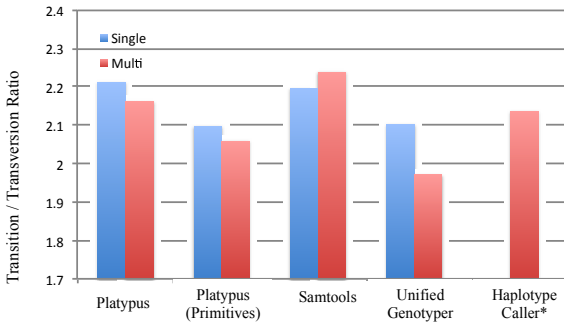
Results:

Single vs. multi sample variant identification

Results: Single vs. multi sample calling

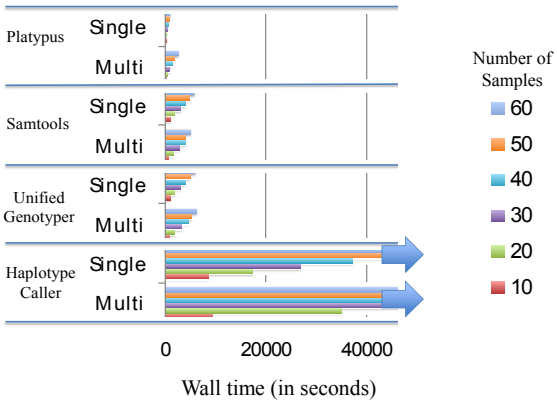


Results: Single vs. multi sample calling

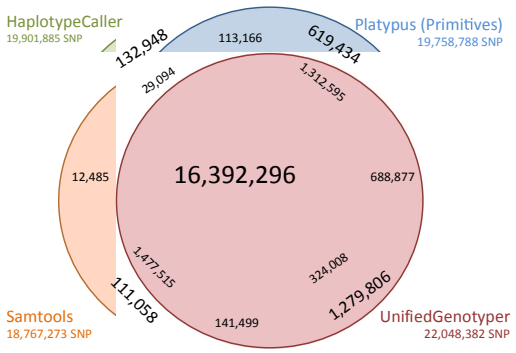


* No single sample results available
IR + BQSR was applied in all analyses

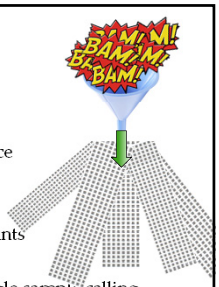
Results: Single vs. multi sample calling



Results: Consensus dataset



Summary



Pre-calling methods:

InDel realignment

- slightly improves variant quality
- likely to improve with improved reference

Base Quality Score Recalibration

- high computation time
- little benefit in number / quality of variants

Single vs. multi sample calling:

- no homozygous reference genotypes in single sample calling
- homozygous reference genotypes with multi sample calling, but depend on number of samples
- single sample calling convenient in terms of computation logistics
- slightly better concordance with SNP array for heterozygotes in multi sample calling

Consensus data set:

- a fantastic idea!

Thanks to Team Qualitas:
Madeleine Berweger, Jürg Moll, Franz Seefried, Urs Schuler

Financial support from the Association of Swiss Cattle Breeders and the Swiss Commission for Technology and Innovation is gratefully acknowledged

Objective

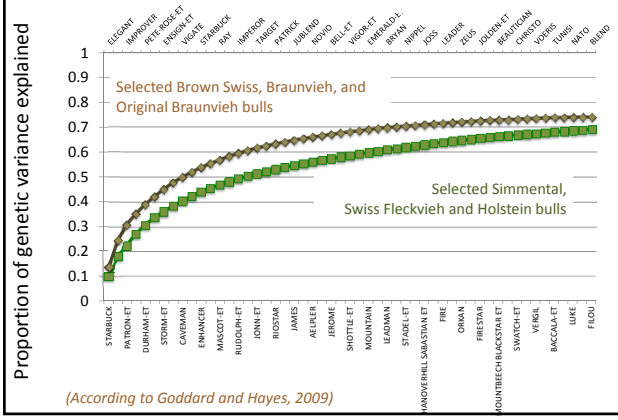
Compare variant identification methods:

Pre-calling methods
- InDel realignment *(dePristo et al., 2011)*
- Base quality score recalibration *(dePristo et al., 2011)*

Software
- Platypus *(Rimmer et al., 2014)*
- Samtools *(Li et al., 2009)*
- Unified Genotyper *(McKenna et al., 2010)*
- Haplotype Caller *(McKenna et al., 2010)*

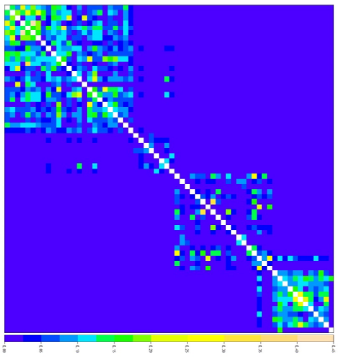
Calling methods:
- single sample variant identification
- multi sample variant identification

Methods: Bull selection



(According to Goddard and Hayes, 2009)

Methods: Bull selection

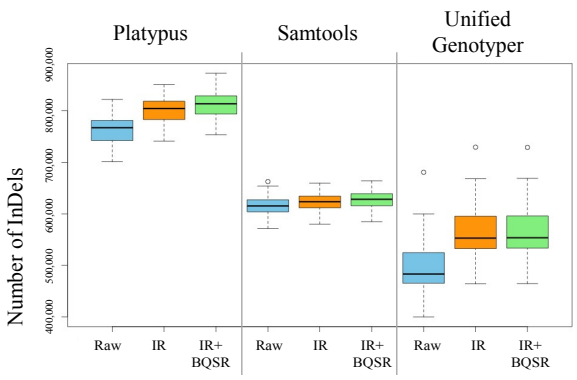


Cluster 1
- Brown Swiss
- Braunvieh
- Original Braunvieh

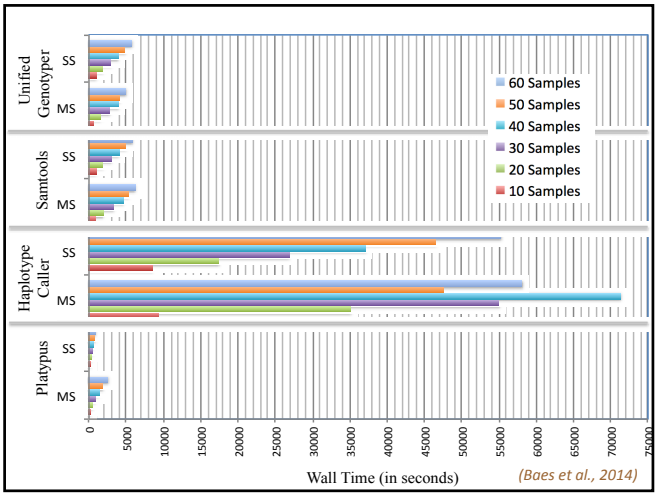
Cluster 2
- Simmental
- Swiss Fleckvieh
- (Red) Holstein

(Baes et al., 2014)

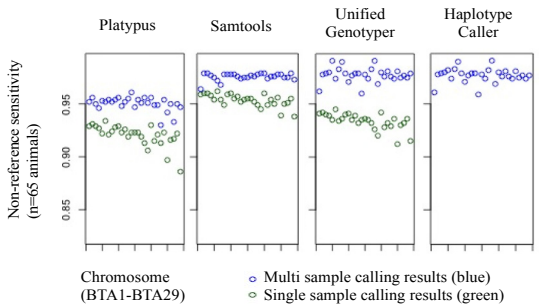
Results: Pre-calling methods



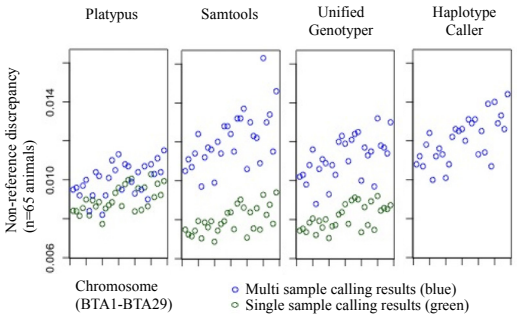
(Baes et al., 2014)



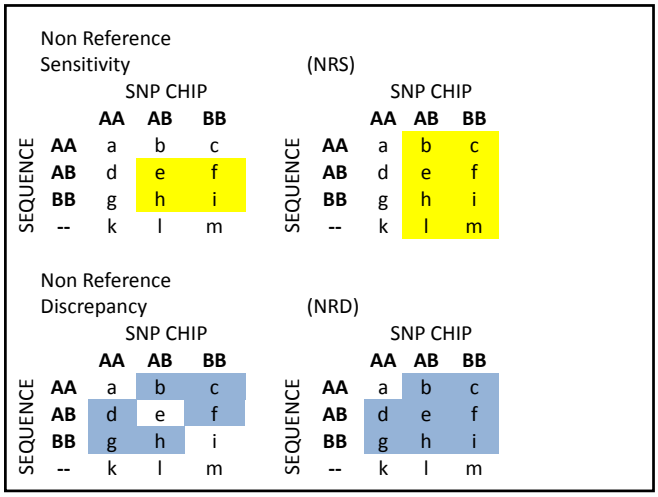
(Baes et al., 2014)



(Baes et al., 2014)



(Baes et al., 2014)


Non Reference
Sensitivity
(NRS)
SNP CHIP
AA AB BB
SEQUENCE
AA a b c
AB d e f
BB g h i
-- k l m
SNP CHIP
AA AB BB
SEQUENCE
AA a b c
AB d e f
BB g h i
-- k l m
Non Reference
Discrepancy
(NRD)
SNP CHIP
AA AB BB
SEQUENCE
AA a b c
AB d e f
BB g h i
-- k l m
SNP CHIP
AA AB BB
SEQUENCE
AA a b c
AB d e f
BB g h i
-- k l m