

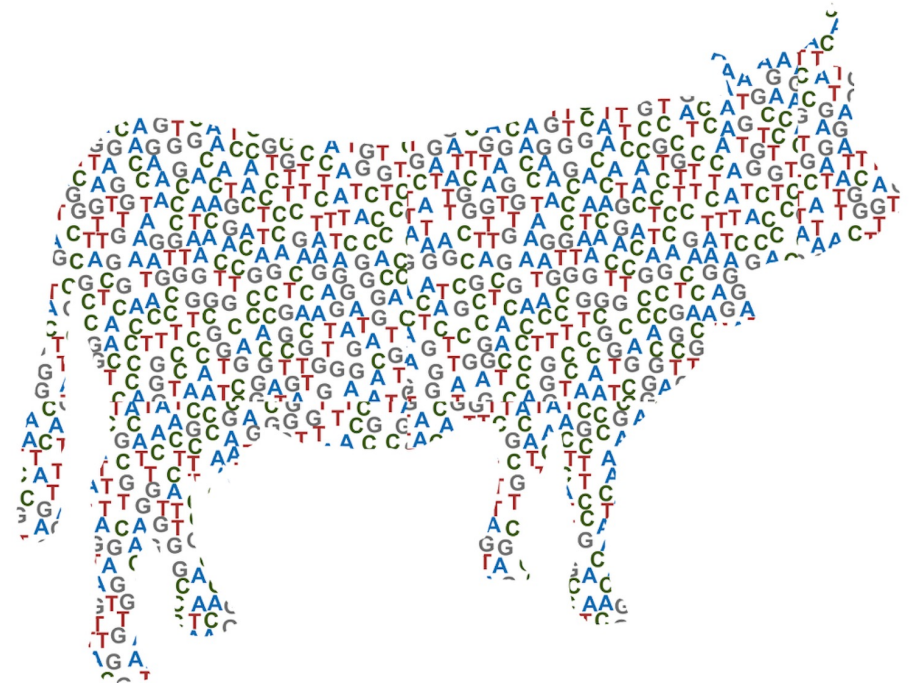


Current projects at the ETH Animal Genomics group

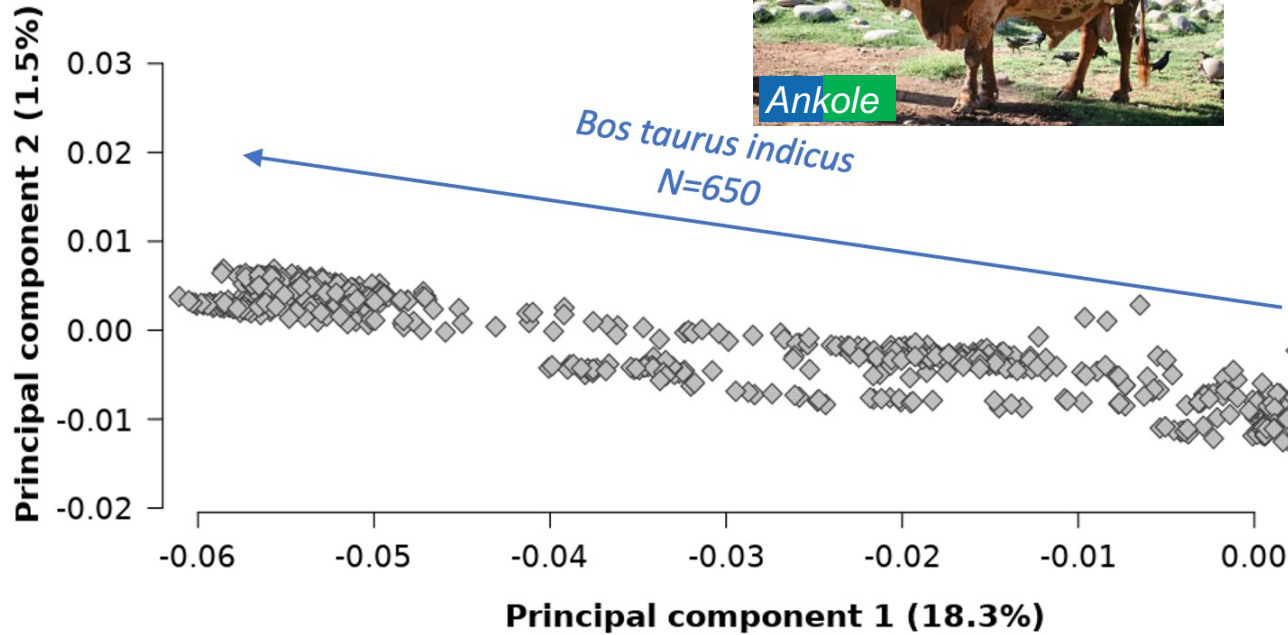
Hubert Pausch
Animal Genomics
10 November 2021

Topics covered today

- Genome assembly and pangenome integration
- Genome and transcriptome analyses to pinpoint male fertility-associated loci



Reference-guided variant discovery revealed >140 million polymorphic sites in >4700 *taurine* and *indicine* cattle genomes¹



¹ Run 8 of the 1000 Bull Genomes Project (<http://www.1000bullgenomes.com/>)

dairy

beef

The linear reference sequence has *limitations*.

Reference-guided variant discovery is prone to «reference allele bias»

- «*soft* reference bias»
 - Reads that contain only alleles that match corresponding reference nucleotides are more likely to align correctly than reads that also contain non-reference alleles
- «*hard* reference bias»
 - Reads originating from DNA fragments that are highly diverged from corresponding reference nucleotides will obtain low alignment scores, or align at incorrect locations, or remain un-mapped
 - Reads originating from DNA fragments that are missing in the reference genome remain unmapped. Reference-guided variant discovery is blind to these sites

Graph-based references mitigate reference allele biases



Danang Crysanto, ETH

- ARS-UCD1.2 used as backbone
- Augmented with SNPs and Indels prioritized based on allele frequency using *VG* (Garrison et al., 2018)
 - *Breed-specific genome graphs*
Variants were prioritized within four breeds
 - *Pangenome graphs*
Variants were prioritized across all breeds
- **Conclusion:** A graph-based reference *improves* read mapping, variant genotyping, and allelic balance at tractable complexity, **but** this approach is largely blind to more complex variant.

Crysanto and Pausch *Genome Biology* (2020) 21:184
<https://doi.org/10.1186/s13059-020-02105-0> Genome Biology

RESEARCH Open Access

Bovine breed-specific augmented reference graphs facilitate accurate sequence read mapping and unbiased variant discovery

Danang Crysanto and Hubert Pausch

Linear reference Variation-aware reference Consensus linear reference

ATGGCTTGAAGTTCGTAGGGTGCC ATGGCTT AAGTTCG AGGGTGCC ATGGCTT AAGTTCGTAGGGTGCC

Frequency in the population **Filtered based on breeds**

Decreasing minimum AF

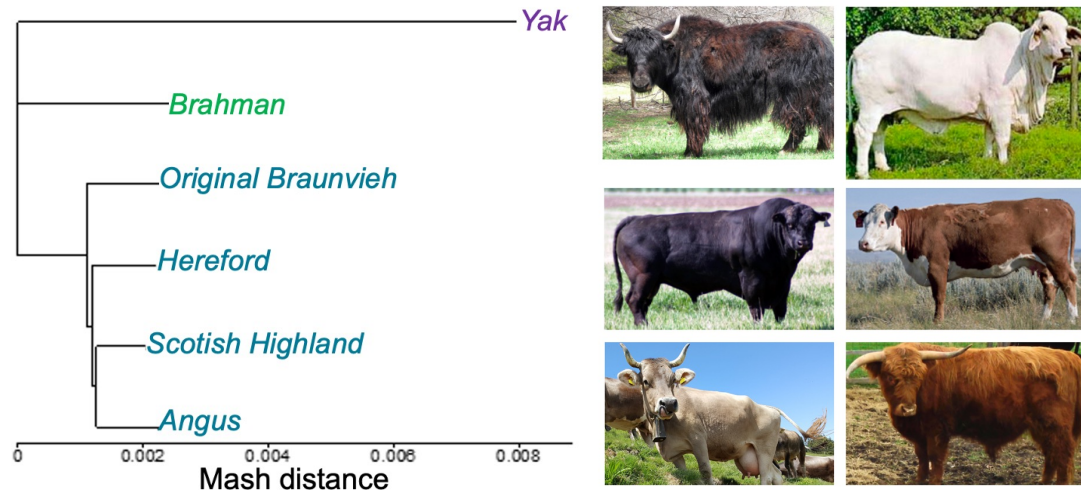
Breed-specific variants Multi-breed variants Random variants

A prototype of a bovine multi-assembly graph

(*Bos taurus*, *Bos indicus*, *Bos grunniens*)



Danang Crysanto, ETH



RESEARCH ARTICLE



Novel functional sequences uncovered through a bovine multiassembly graph

Danang Crysanto, Alexander S. Leonard, Zih-Hua Fang, and Hubert Pausch

[+ See all authors and affiliations](#)

PNAS May 18, 2021 118 (20) e2101056118; <https://doi.org/10.1073/pnas.2101056118>

Edited by Harris A. Lewin, University of California, Davis, CA, and approved April 2, 2021 (received for review January 18, 2021)

Table 1. Details of six bovine genome assemblies

Assembly (Species)	Sex*	Primary data used for the assembly [†]	Type of assembly	Assembler	Contig N50 (Mb)	Scaffold N50 (Mb)	Length of the autosomes
Hereford (<i>Bos taurus taurus</i>)	F	PacBio (80-fold CLR)	Primary	Falcon	21	108	2,489,385,779
Angus (<i>Bos taurus taurus</i>)	M	PacBio (136-fold CLR)	Haplotype-resolved	TrioCanu	29.4	102.8	2,468,157,877
Highland (<i>Bos taurus taurus</i>)	F	PacBio (125-fold CLR)	Haplotype-resolved	TrioCanu	71.7	86.2	2,483,452,092
Original Braunvieh (<i>Bos taurus taurus</i>)	F	PacBio (28-fold HiFi)	Primary	Hifiasm	86.0	96.3	2,607,746,442
Brahman (<i>Bos taurus indicus</i>)	F	PacBio (136-fold CLR)	Haplotype-resolved	TrioCanu	23.4	104.5	2,478,073,158
Yak (<i>Bos grunniens</i>)	F	PacBio (125-fold CLR)	Haplotype-resolved	TrioCanu	70.9	94.7	2,478,308,164

A bo
(*Bos ta*

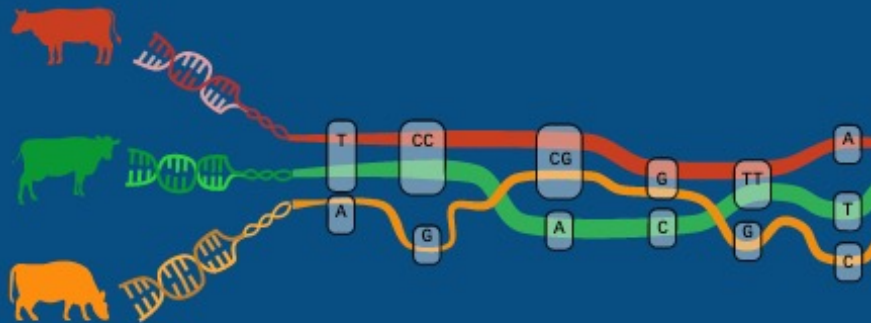
uilt with *minigraph*



Danang Crysanto, ETH

Establishing Bovine Pangenome Graphs

Danang Crysanto
DISS. ETH NO. 27672



RESEARCH ARTICLE

ETH zürich

Login Help Deutsch

Research Collection

Search

Home → Theses → Doctoral Thesis → View Item

Browse

Organisational Units

Publication Types

Authors

Publish

New Submission

Statistics

Downloads by Country

Most Popular Items

Most Popular Authors

Establishing Bovine Pangenome Graphs



Open access

Author

Crysanto, Danang

Date

2021

Type

Doctoral Thesis

ETH Bibliography

yes

Download

Full text (PDF, 11.32Mb)

Rights / license

Creative Commons Attribution 4.0 International

Abstract

The assembly of the draft *Bos taurus* reference genome was a milestone for genetics- and genomics-oriented research in cattle. The reference genome of domestic cattle was built from a single animal from the Hereford breed. However, the linear reference sequence does not represent the genetic diversity. [Show more](#)

Permanent link

<https://doi.org/10.3929/ethz-b-000510180>

Publication status

published

External links

[Search print copy at ETH Library](#)

Contributors

Examiner: Pausch, Hubert

Examiner: Guldbrandtsen, Bernt

Examiner: MacHugh, David

Haplotype-resolved	TrioCanu	23.4	104.5	2,478,073,158
Haplotype-resolved	TrioCanu	70.9	94.7	2,478,308,164

How to best integrate *haplotype-resolved* assemblies?

All reported findings are in our preprint



Alexander Leonard




bioRxiv
THE PREPRINT SERVER FOR BIOLOGY

New Results

 [Follow this preprint](#)

Bovine pangenome reveals trait-associated structural variation from diverse assembly inputs

 Alexander S. Leonard, Danang Crysanto, Zih-Hua Fang, Michael P Heaton, Brian L. Vander Ley, Carolina Herrera, Heinrich Bollwein, Derek M. Bickhart, Kristen L. Kuhn, Timothy PL. Smith, Benjamin D. Rosen, Hubert Pausch

doi: <https://doi.org/10.1101/2021.11.02.466900>

This article is a preprint and has not been certified by peer review [what does this mean?].

Male fertility.



Maya Hiltbold

A 1-bp deletion in *QRICH2* is associated with sperm defects in Brown Swiss bulls

- Identified through a forward-genetic screen in one bull (born in 2003, twelve ejaculates collected, all discarded)
- Assigned onto a 675 kb haplotype (181 BovineHD SNPs) that segregates at a frequency of 5%
- Found **one** live male animal (6 months old) in Germany
- Bull was housed at our research station for approx. 1 year
- Seven ejaculates collected, defect verified



An eQTL cohort to map molecular phenotypes that are associated with male reproductive performance



Xena Marie Mapel

- eQTL cohort
 - Testis tissue sampled from 76 mature bulls at a commercial slaughterhouse
 - DNA sequencing to an average of 12.6-fold depth
 - Deep total paired-end RNA sequencing (283.6 million reads per sample)
 - Used for a genome-wide sQTL study

communications biology

ARTICLE



<https://doi.org/10.1038/s42003-021-02725-7>

OPEN

The intronic branch point sequence is under strong evolutionary constraint in the bovine and human genome

Naveen Kumar Kadri ^{1✉}, Xena Marie Mapel¹ & Hubert Pausch ¹



Naveen's talk in the afternoon



Contributors & funding



- Ben Rosen, Derek Bickhart, Tim Smith, USDA, USA, from the *Bovine Pangenome Consortium*



Horizon 2020
European Union funding
for Research & Innovation



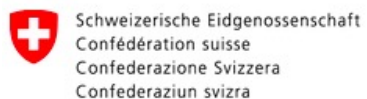
Eine Kooperation in Bildung und Forschung:
strickhof ETH zürich University of Zurich



Functional Genomics Center Zurich



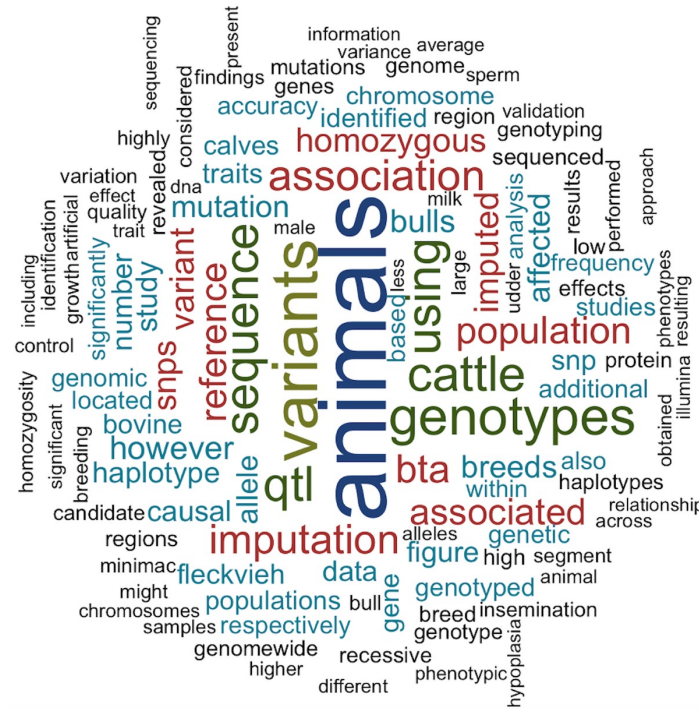
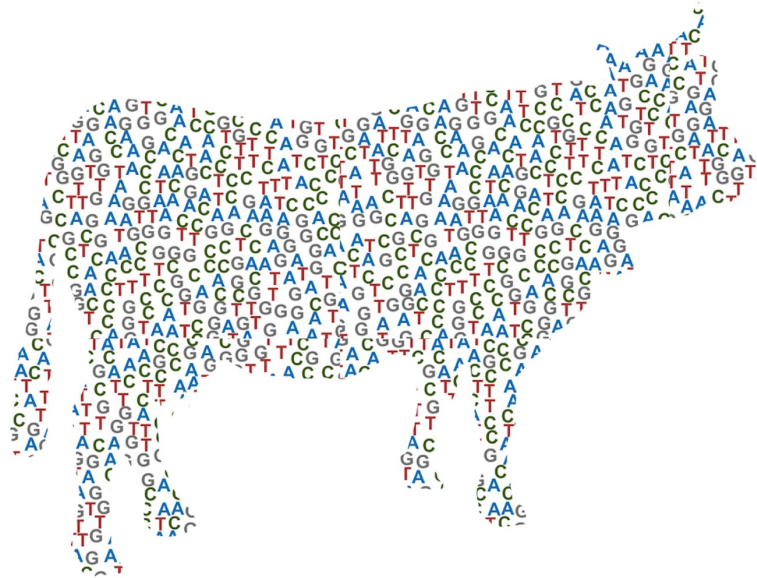
FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFONDS
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION



Federal Office for Agriculture FOAG



Thank you for listening!



ETH Zürich
Hubert Pausch
Animal Genomics
Universitätstrasse 2 | LFW B 58.2
8092 Zürich

hubert.pausch@usys.ethz.ch
www.ag.ethz.ch