AMERICAN SHEEP INDUSTRY ASSOCIATION
YOUR INDUSTRY ADVOCATE SINCE 1865

# Implication of the Sheep Genome with Disease in Domestic and Wild Sheep

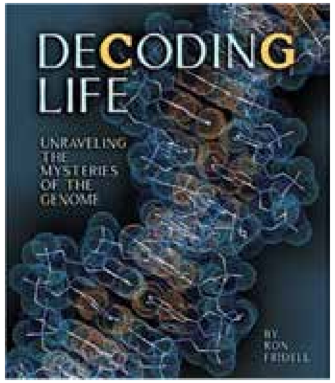
Brenda M. Murdoch, Temitayo Olagunju, Shangqian Xie, Emily Clark, Alan Archibald, Michelle Mousel, Stephanie McKay, Darren Hagen, Andrew Hess, Ted S. Kalbfleisch, Michael Heaton, Brad Freking, Tom Murphy, Derek Bickhart, Shannon Clarke, Rudiger Brauning, Tracy Hadfield, Noelle Cockett, Ben Rosen and Timothy P.L. Smith

USDA NIFA

American Sheep Industry Convention
Animal Health Committee
January 29, 2026

# Importance of Reference Genomes

- High quality reference genomes underpin all genomic research within a species.



- One of NSF 10 Big Ideas in 2019, “Understanding the Rules of Life” with the mission of “Elucidating the set of rules that predict an organism’s observable characteristics, its phenotype”.
- Vertebrate Genome Project (VGP) - generate reference genomes of species from all 260 vertebrate orders, with plans to sequence all 66,000 vertebrates.

VGP
VERTEBRATE GENOMES PROJECT
A PROJECT OF THE G10K CONSORTIUM

# Reference Genome Assemblies

| **Reference Genome Assembly** | **Species** | **Contig N50 (Mb)** | **Number of contigs** | **Assembly size (Gb)** | **Release date** |
|---|---|---|---|---|---|
| **GRCh38.p14** | Human | 57.9 | 996 | 3.1 | 2022 |
| Sscrofa11.1 | Pig | 48.2 | 1,117 | 2.5 | 2017 |
| ARS-UCD2.0 | Cattle | 25.9 | 2,343 | 2.8 | 2023 |
| ARS1 | Goat | 26.2 | 30,377 | 2.9 | 2016 |
| **Oar_rambouillet_v1.0** | Sheep | 2.6 | 7,486 | 2.9 | 2017 |
| **ARS-UI_Ramb_v2.0** | Sheep | 43.2 | 226 | 2.7 | 2021 |

❖ The *ARS-UI_Ramb_v2.0* ovine reference genome assembly has improved contiguity and error rates compared with other sheep genomes and is comparable to other species

# Reference Genome Assemblies

| **Reference Genome Assembly** | **Species** | **Contig N50 (Mb)** | **Number of contigs** | **Assembly size (Gb)** | **Release date** |
|---|---|---|---|---|---|
| **GRCh38.p14** | Human | 57.9 | 996 | 3.1 | 2022 |
| Sscrofa11.1 | Pig | 48.2 | 1,117 | 2.5 | 2017 |
| ARS-UCD2.0 | Cattle | 25.9 | 2,343 | 2.8 | 2023 |
| ARS1 | Goat | 26.2 | 30,377 | 2.9 | 2016 |
| **Oar_rambouillet_v1.0** | Sheep | 2.6 | 7,486 | 2.9 | 2017 |
| **ARS-UI_Ramb_v2.0** | Sheep | 43.2 | 226 | 2.7 | 2021 |
| **ARS-UI_Ramb_v3.0** | **Sheep** | **43.2** | **226** | **2.7** | **2023** |

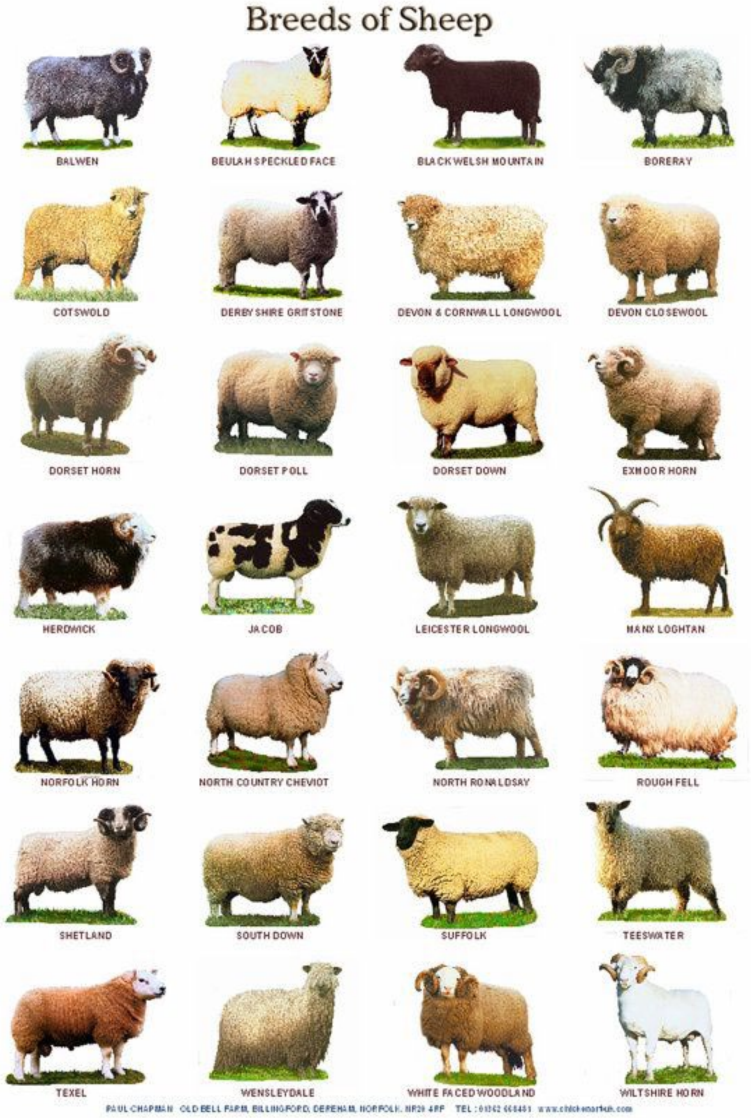
- The *ARS-UI_Ramb_v2.0* ovine reference genome assembly has improved contiguity and error rates compared with other sheep genomes and is comparable to other species

# Development of the Ovine Pangenome

A single representative reference genomes of a species are inadequate to describe the full genome variation across populations within that species.

**Objective to deliver high-quality genomes, for different breeds and species of sheep selected for their divergent traits.**

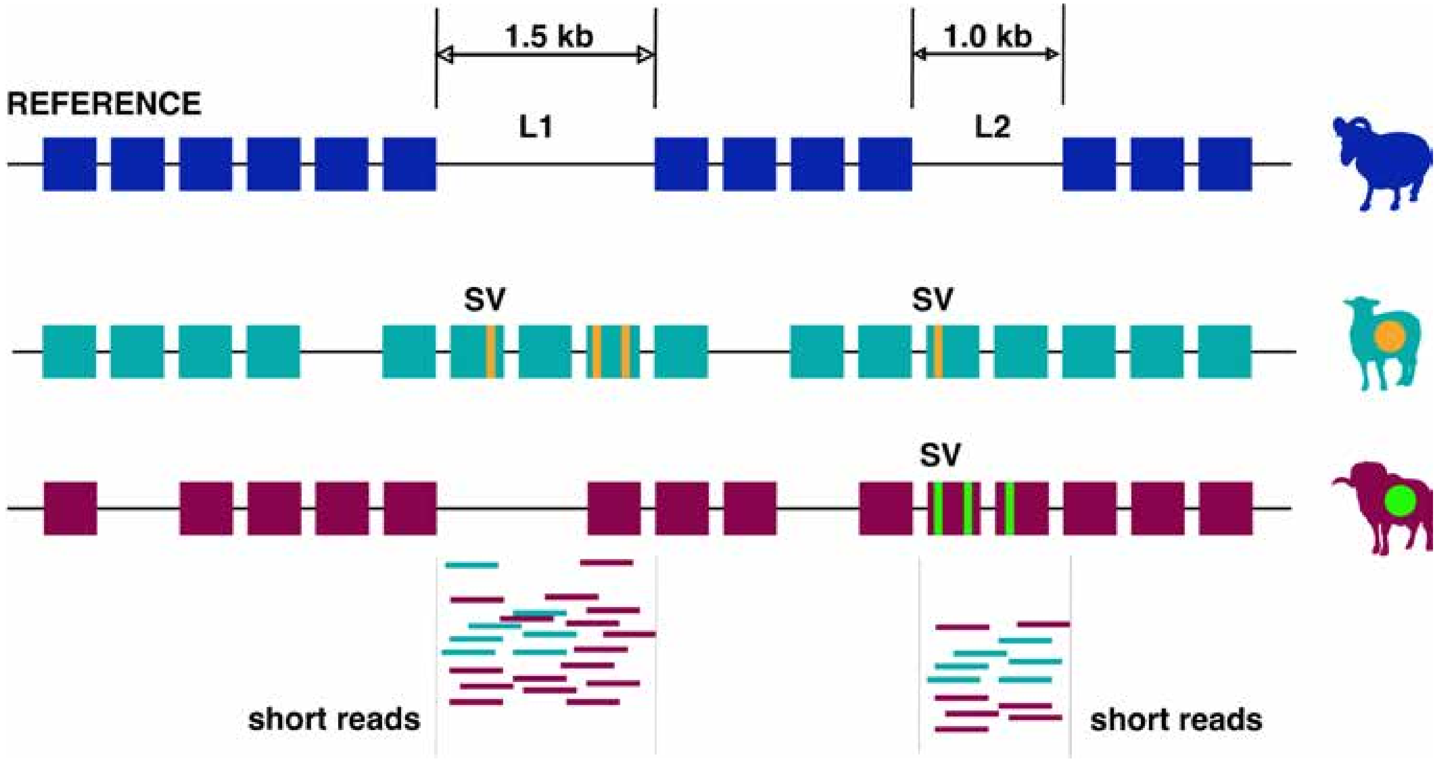
Characterize genomic regions of different ovine breeds and species, thereby identifying breed specific genomic regions associated with a range of phenotypically distinct important biological traits.



# What is a Pangenome

## Reference bias

- Miss true variants in individuals due to absence of region on reference
- Erroneous variants due to reads aligning to wrong places



Single Reference

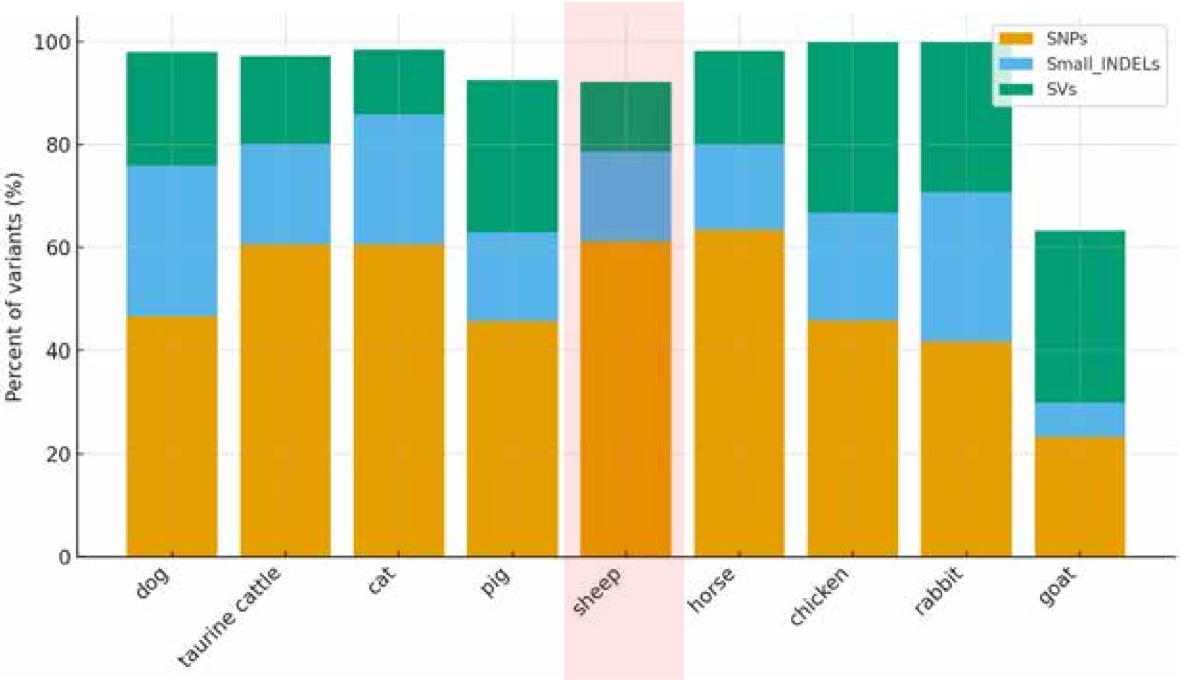
Variants absent in the reference

Short sequence reads aligned to the reference for variant discovery

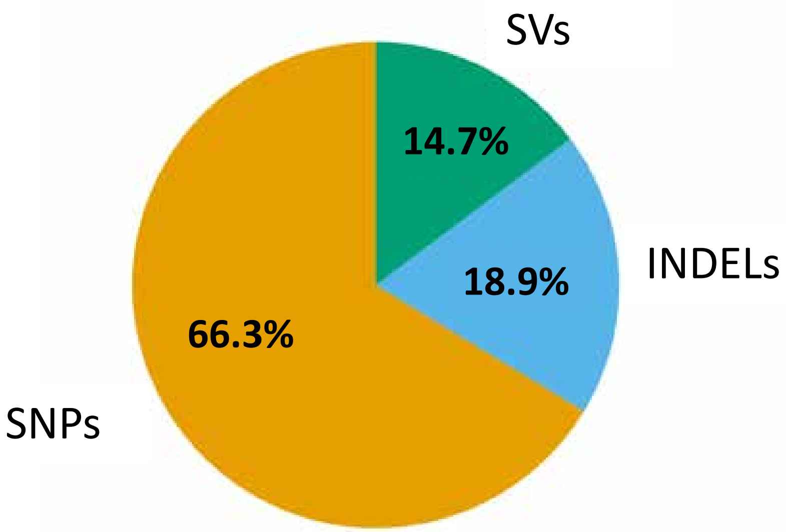
Variants = Single Nucleotide Polymorphism (SNP), Insertion/ Deletion (INDEL), Structural Variant (SV)

# Why a Pangenome

SVs = variants ≥ 20bp



Proportion of variant types by species on OMIA

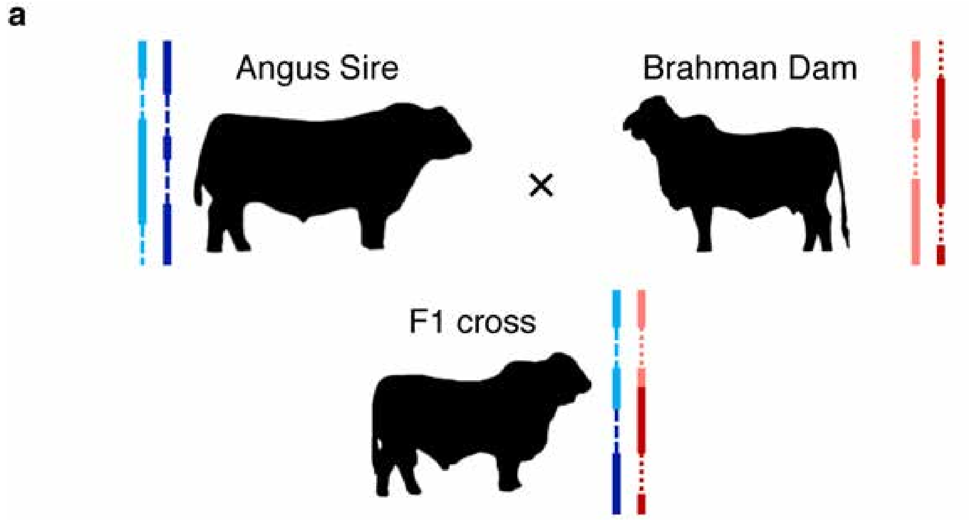


Proportion of **sheep** variants on OMIA

The low figures are mainly due to the shortcoming of many short read variant callers (Kosugi *et al*., 2019; Zook *et al*., 2020)

# Trio Binning Haplotype Assembly

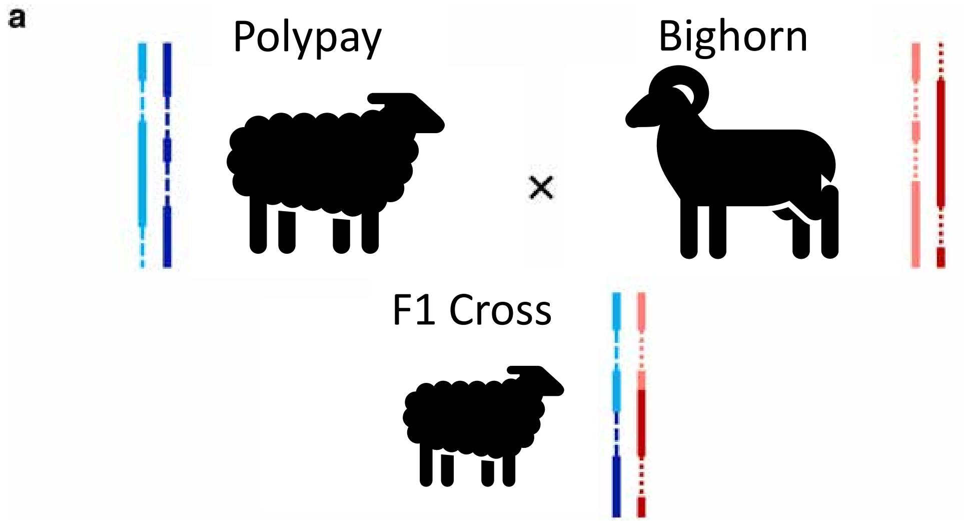
Generate high-quality (haplotype-resolved) references



(figure modified from Koren et al., 2018)

# Trio Binning Haplotype Assembly

Generate high-quality (haplotype-resolved) references



(figure modified from Koren et al., 2018)

# Genome Assemblies Included in Pangenome

- Generated F1 fetuses for the following crosses

| **Merino** | **Awassi** |
|---|---|
| ▪ Fine wool<br>▪ Semiarid adapt | ▪ Fat in tail<br>▪ Dessert adaptable |
| **St. Croix** | **Suffolk** |
| ▪ Parasite resistant<br>▪ Hair | ▪ Parasite susceptible<br>▪ Growth and meat |
| **East Friesian** | **Native Churro** |
| ▪ Dairy<br>▪ Multiple offspring | ▪ Polled – 4 horns<br>▪ Adapt to arid climate |
| **Polypay** | **Bighorn (*canadensis*)** |
| ▪ Fine wool<br>▪ Multiple offspring | ▪ Horn & hair<br>▪ Pneumonia susceptible |

# Genome Assemblies Included in Pangenome

- 8 sheep, includes 2 species promised in the grant (in bold)

| |
|---|
| **Merino** |
| **Awassi** |
| **St. Croix** |
| **Suffolk** |
| **East Friesian** |
| **Native Churro** |
| **Polypay** |
| **Bighorn (*canadensis*)** |

| |
|---|
| Dorper |
| Romanov |
| Shire |
| Romney |
| Wiltshire |
| Damara |

- 6 domestic breeds were added by Shannon Clarke and Tim Smith

- 8 high quality public sourced assemblies and 2 species were added

| |
|---|
| Chinese Merino |
| Kazak |
| Polled Dorset |
| Qiaoke |
| Ujumqin |
| Waggir |
| Yunnan |
| ***Rambouillet** |

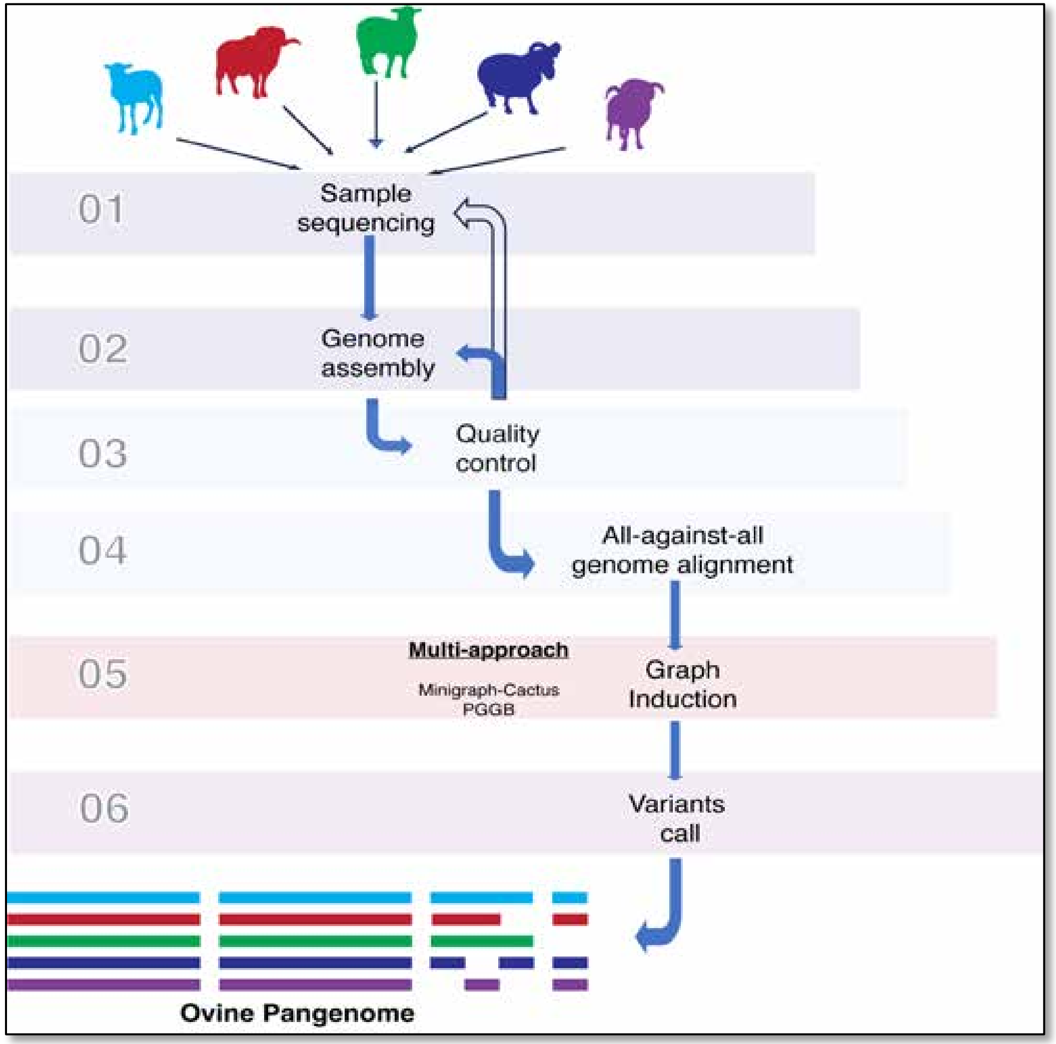
*(Ovis *aries*) reference

| |
|---|
| Mouflon (*orientalis*) |
| Argali (*ammon*) |

- Total of **24** sheep including **4** species

**Sergey Koren, Arang Rhie & Adam Phillippy**

# Pangenome Building Process



Step 01: Sequence: 75x HiFi, 100x & 30x UL ONT, 30 x Illumina (from parents)

Step 02: Assembly - Verkko v2.0

Step 03: Manual curation

Step 04: Alignments and Annotation

Step 05: Build graph

Step 06: Variants are being analyzed
(SNP, INDELs, SV)

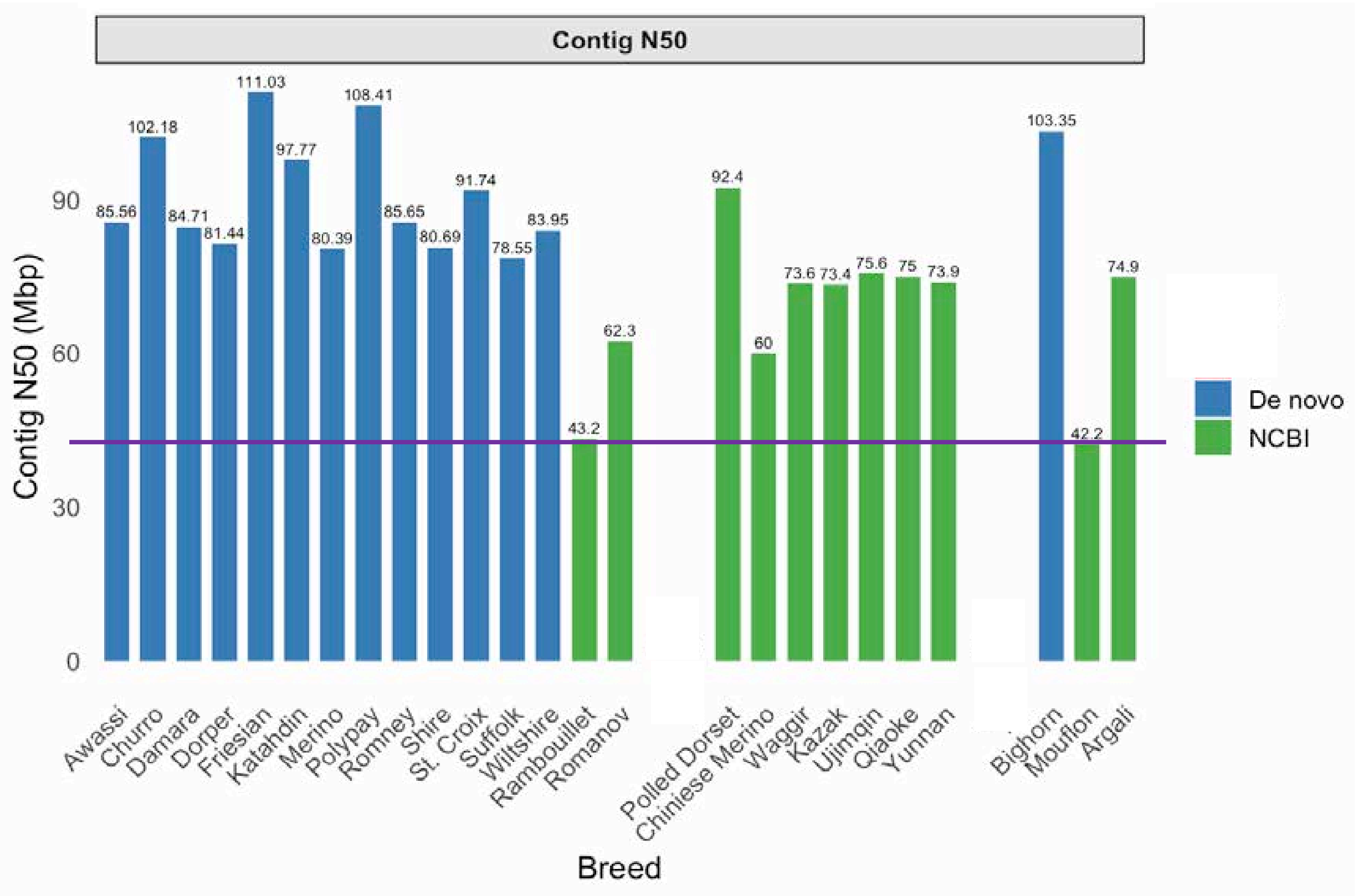
Multiple graph approaches are used – Minigraph-Cactus (MC) and PanGenome Graph Builder (PGGB)

# Pangenome Building Process

- PanGenome Graph Builder (**PGGB**)
- All-to-all alignment of the input genomes
- Reference-agnostic
- Retains more information but is harder to index and query against

- Minigraph-Cactus (**MC**)
- Selects a “reference” assembly as the backbone
- Progressively aligns other assemblies to the backbone
- Retains less information but easier to query

# Development of the Ovine Pangenome



**14 New assemblies**
13 domestic sheep
1 Bighorn

**10 Public assemblies**
Selected based on high contiguity (Contig N50 > 40Mb)

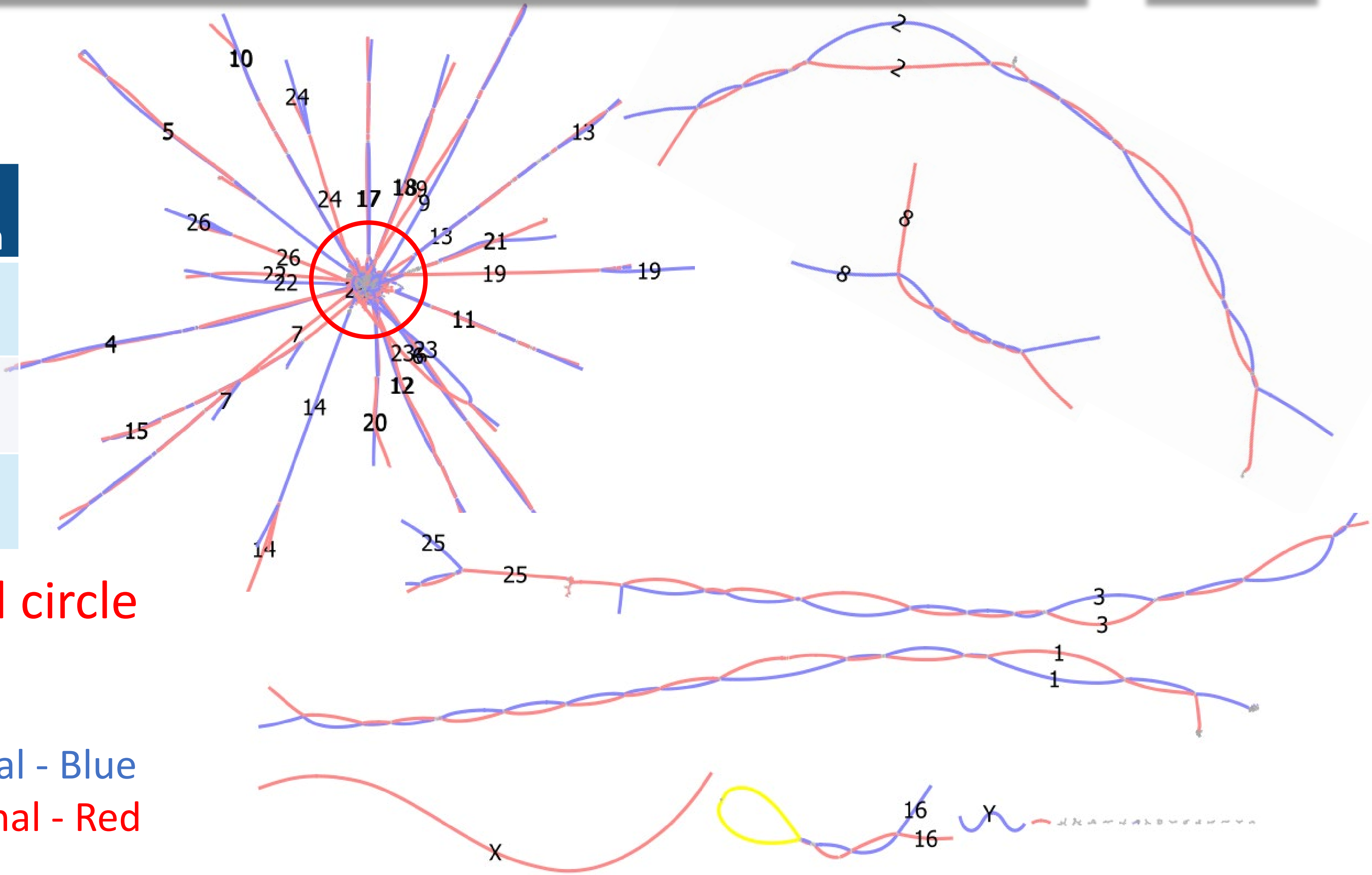
# Genome Assembly Graph

## Churro x Friesian

| | Native Churro | East Friesian |
|---|---|---|
| Contig/Scaff N50 Mb | 93/96 | 97/100 |
| Total gaps | 9 | 16 |
| near T2T* | 21/27 | 20/27 |

Satellite repeats - Red circle

*missing satellite annotations



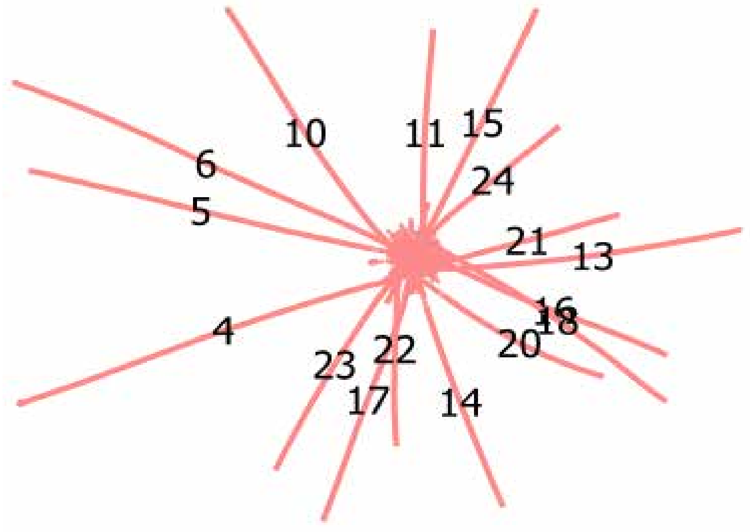
Paternal - Blue

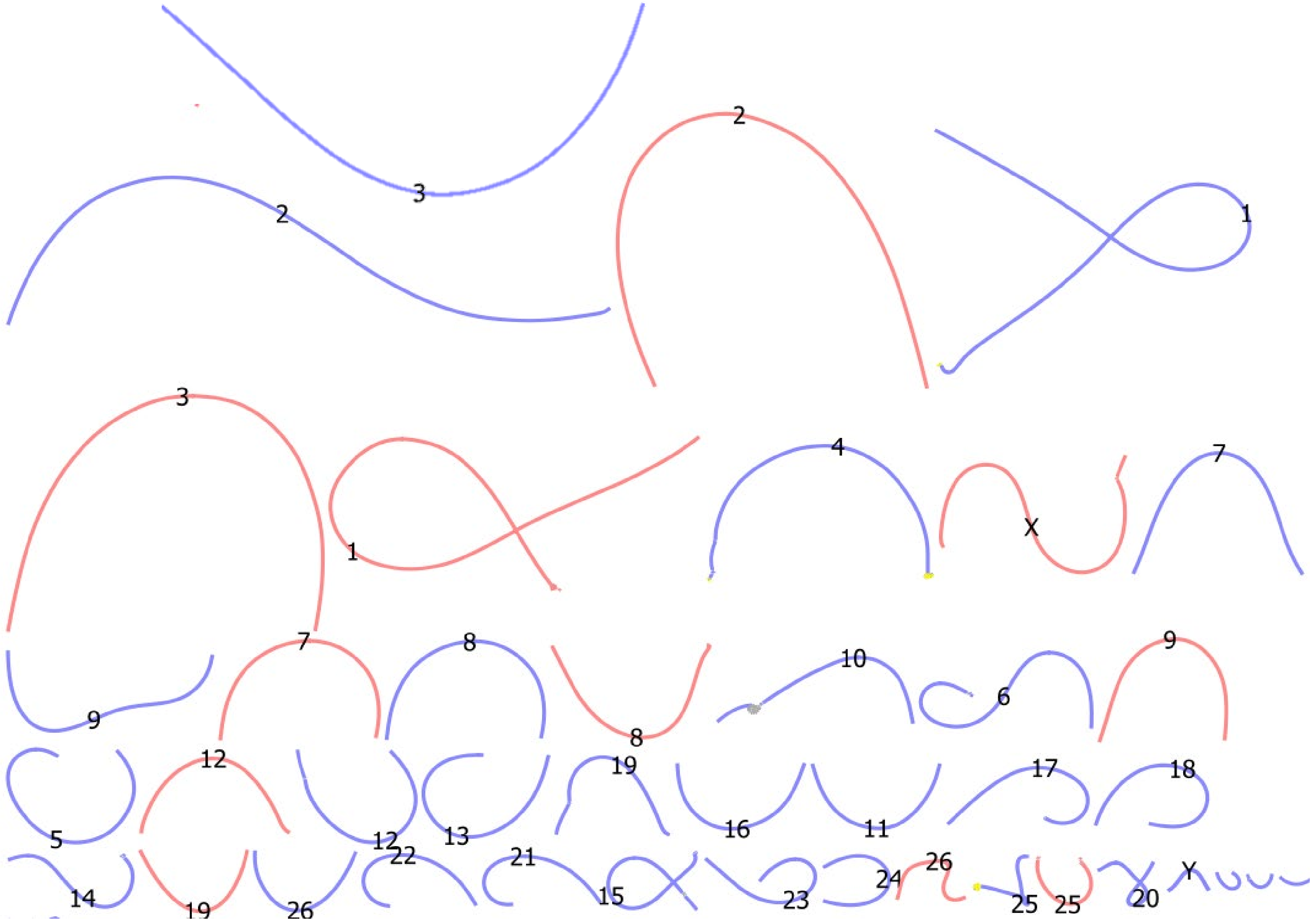
Maternal - Red

# Genome Assembly Graph: Interspecies Cross

## Bighorn x Polypay

| | Bighorn | Polypay |
|---|---|---|
| Contig/Scaff N50 Mb | 103/103 | 95/95 |
| Total gaps | 3 | 16 |
| near T2T* | 24/27 | 23/27 |



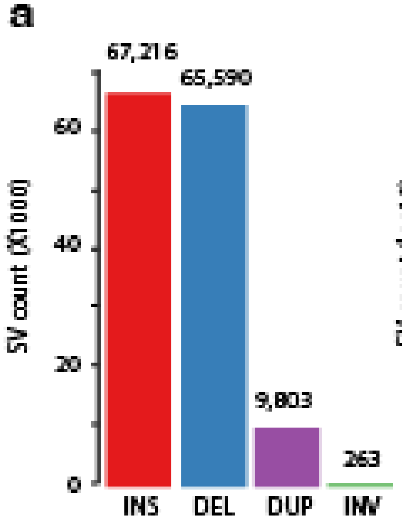


# Bighorn Telomere-to-Telomere (T2T)

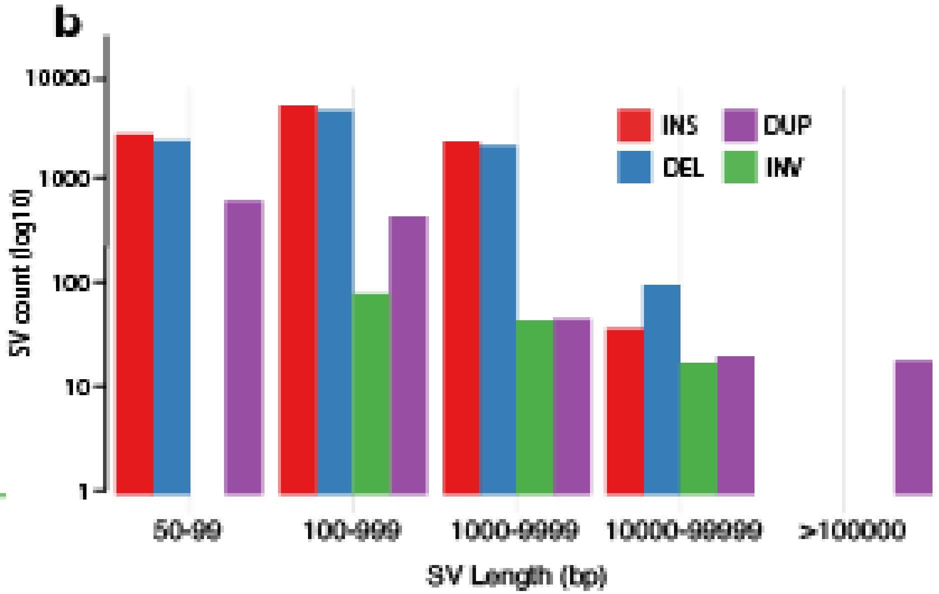
## Bighorn sheep T2T genome assembly reveals differences in immune genes: a potential cause of high morbidity due to respiratory pathogens

Temitayo A. Olagunju, Mariia Pospelova, John C. Schwartz, Michelle R. Mousel, Lindsay M.W. Piel, Paige C. Grossman, Kathryn P. Huyvaert, Kristen L. Kuhn, Tajbir Raihan, Morgan R. Stegemiller, Sarem F. Khilji, Gordon K. Murdoch, Ahmed Tibary, Lisette P. Waits, Arang Rhie, Sergey Koren, Adam M. Phillipi, Stephanie D. McKay, Shannon M. Clarke, Emily L. Clark, Rudiger Brauning, Noelle E. Cockett, John A. Hammond, Margaret Highland, Yana Safonova, Timothy P.L. Smith, Benjamin D. Rosen and Brenda M. Murdoch

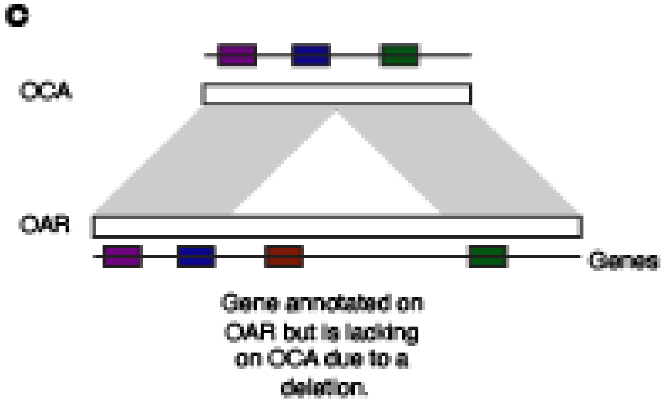
Number of SVs bighorn



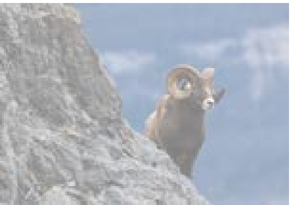
The size distribution of the SVs



Bighorn sheep (OCA)
domestic sheep genome (OAR)



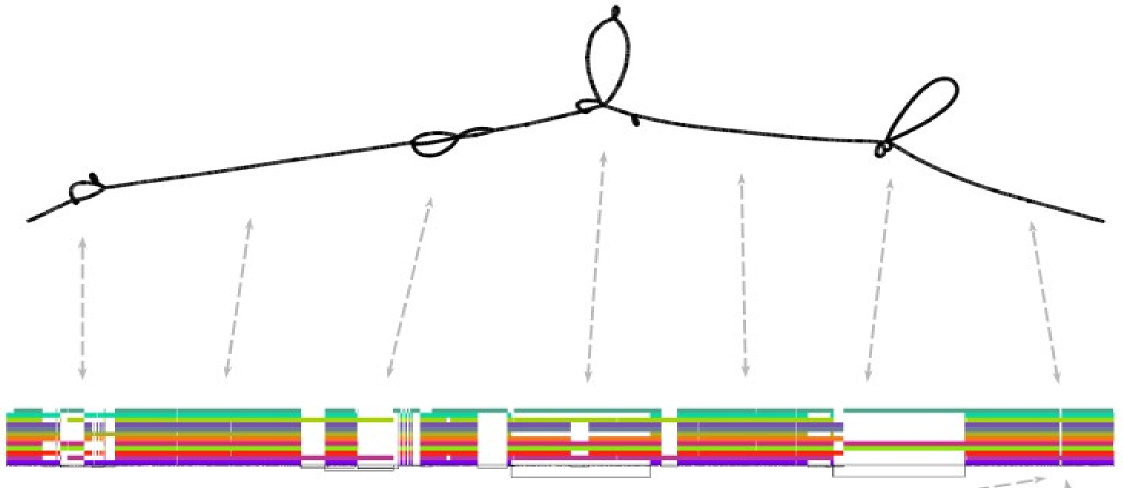
**Two genes lacking on Bighorn-T2T**
***GBP5 & ZNF501***

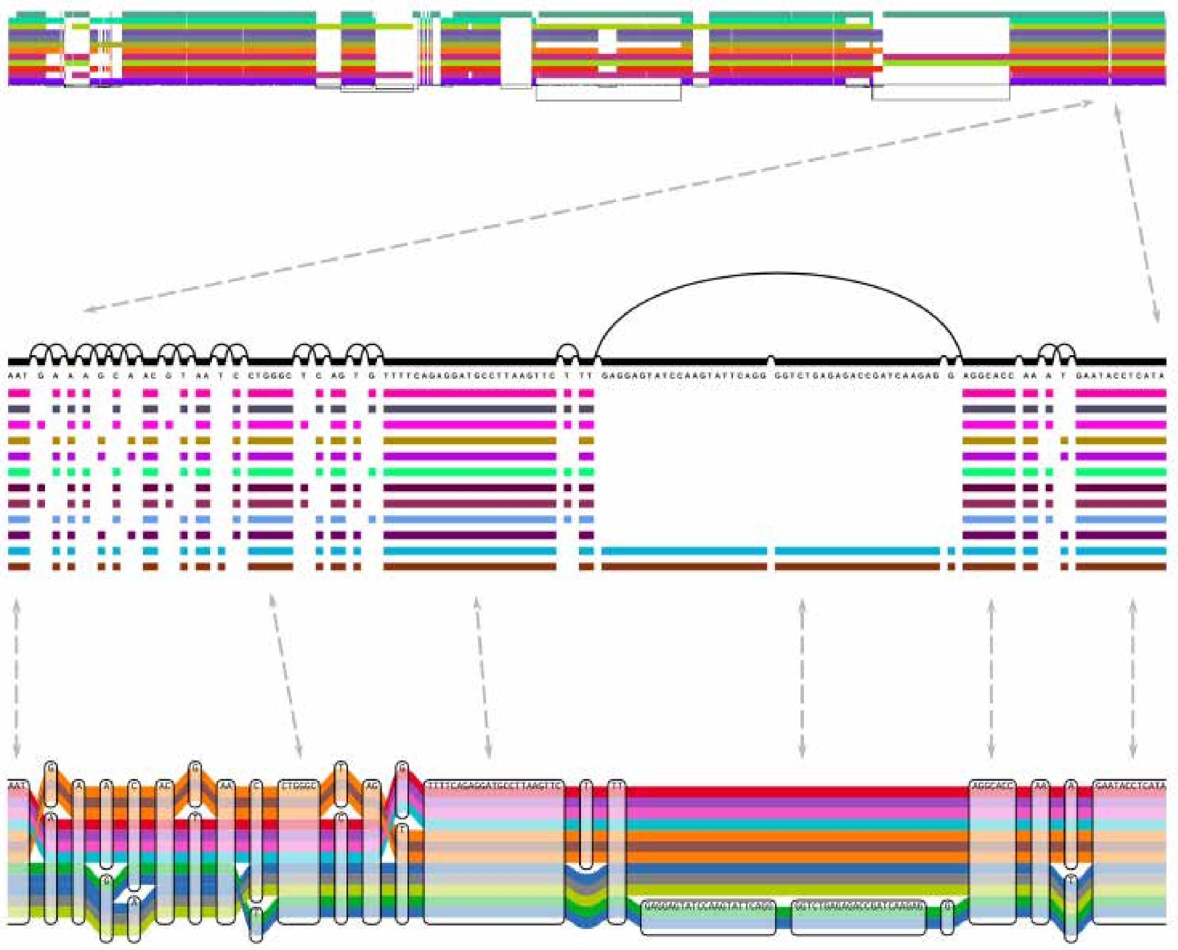


# Pangenome Graph

**Objective to deliver genetic resources as a tool for studying the ovine genome through online databases.**

- Genome graphs are a very flexible way of truthfully representing the structure and sequence of genomes as paths.



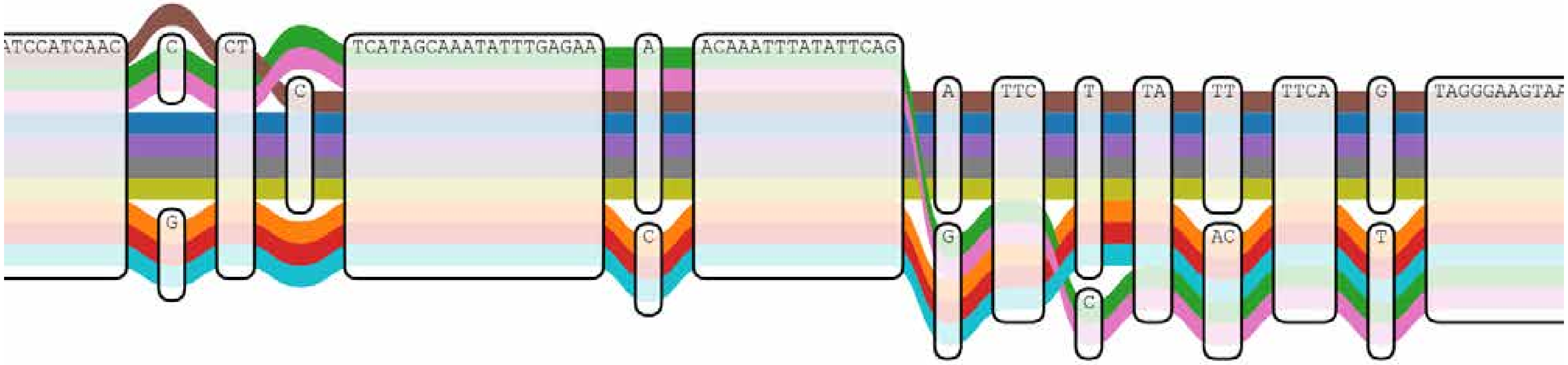


https://pangenome.github.io/

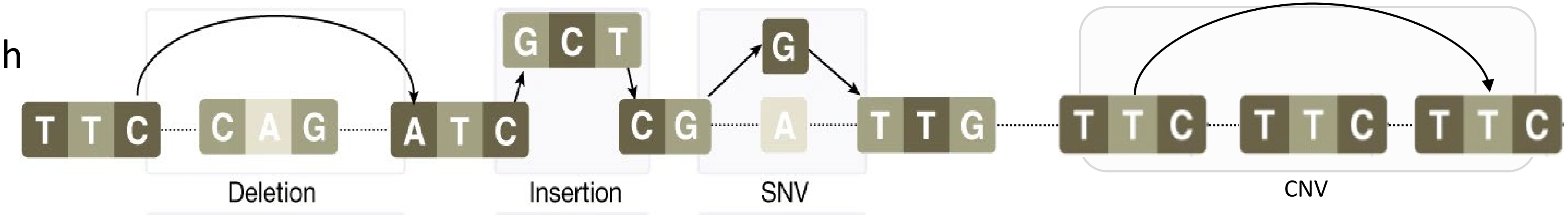
# Pangenome Graph

- Representing a pangenome
  - Variation graph (VG)

Representation



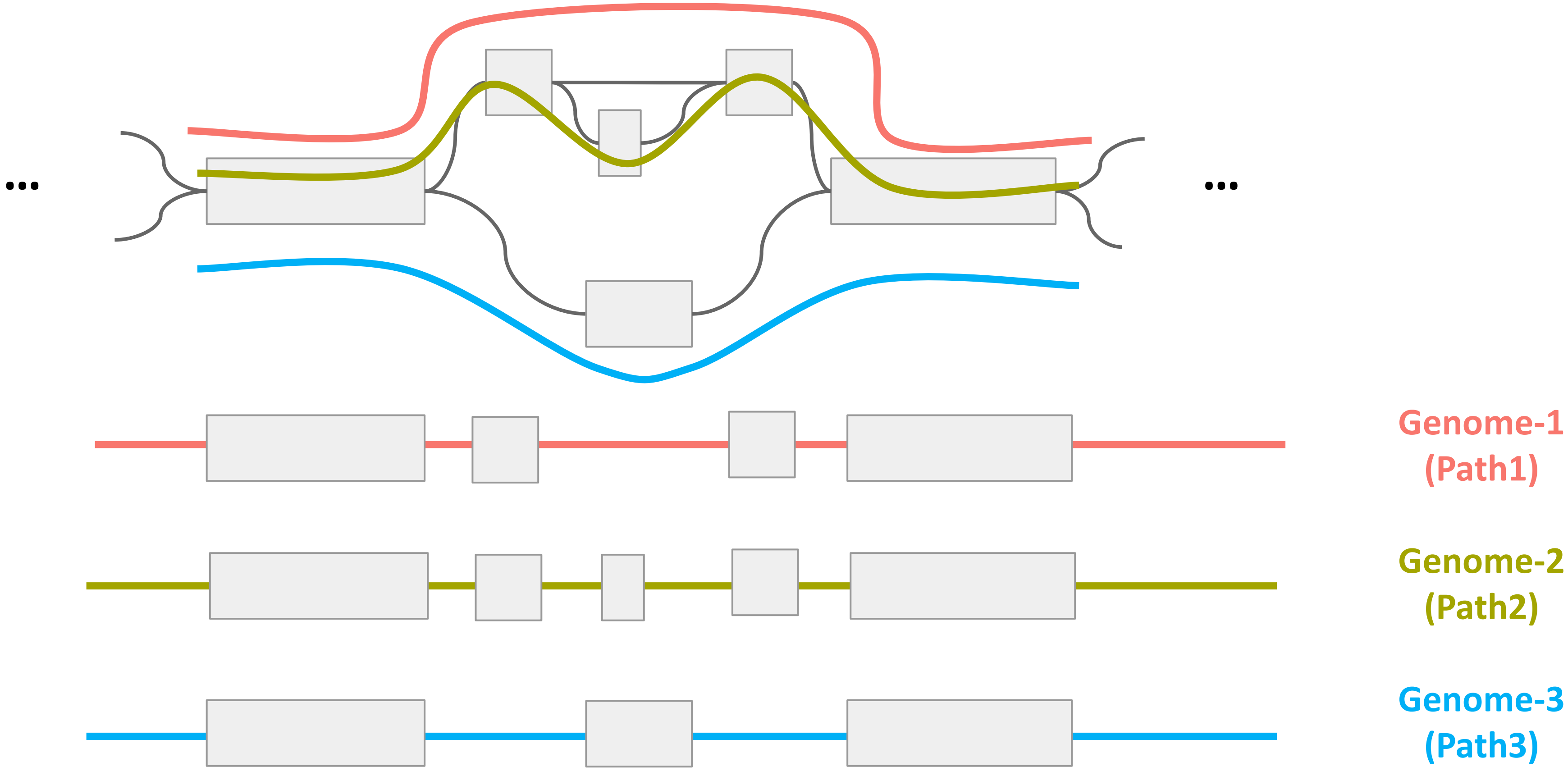
SVs on a Variation Graph



Adapted from Erik Garrison

# Pangenome Graph

- Representing a pangenome
  - Variation graph (VG)

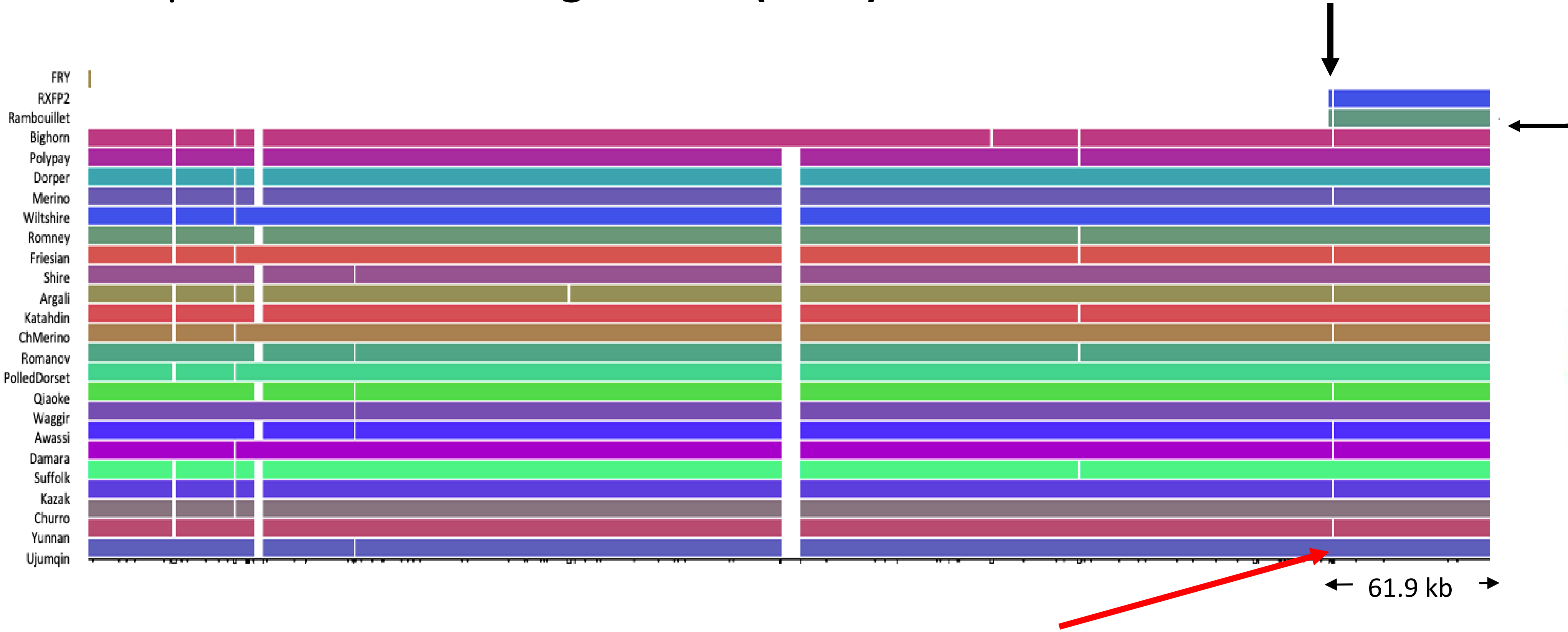


Temitayo
Olagunju

Adapted from Erik Garrison

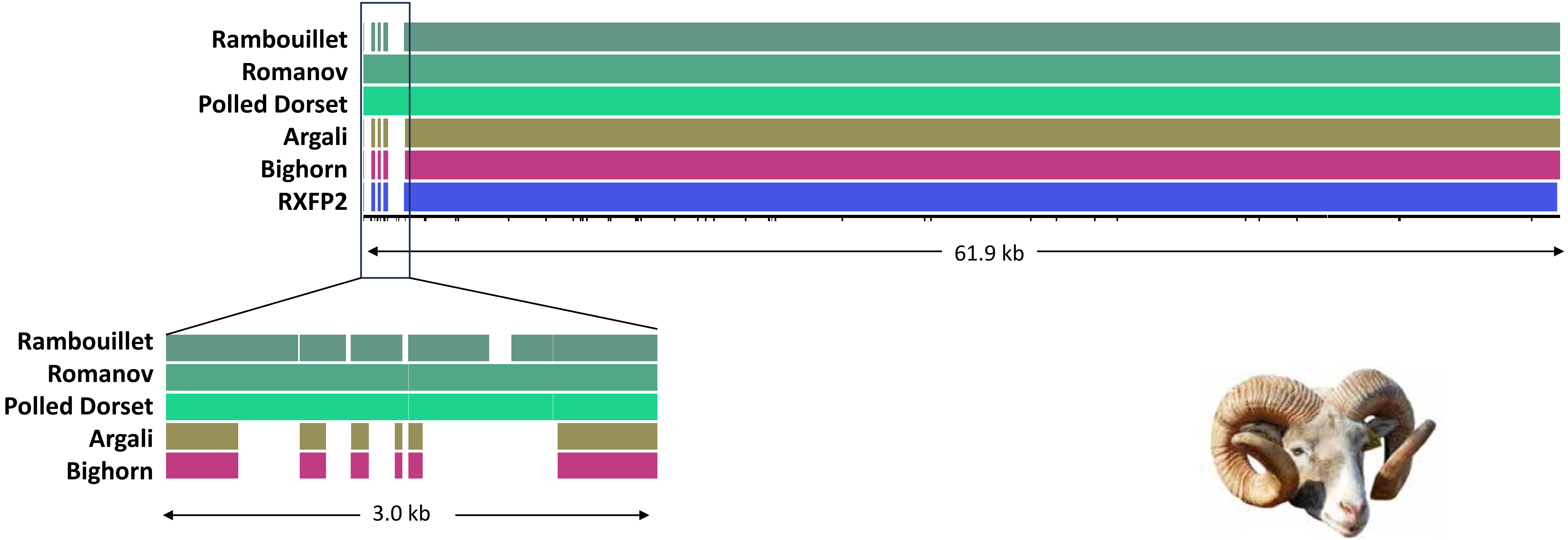
# Relaxin Family Peptide Receptor 2 (RXFP2)

- A 1.8-kb insertion in the 3'-UTR of RXFP2 is associated with polledness in sheep. **Wiedemar & Drögemüller (2015)**



1.6 kb insertion

# Relaxin Family Peptide Receptor 2 (RXFP2)



# Summary

- We have presented a sheep pangenome comprising 26 breeds which were selected based on diverse traits
- T2T Bighorn genome assembly identified differences in immune genes
- The pangenome was built using PGGB and Minigraph-Cactus

# Impact

High-quality pangenomes with well-annotated regulatory elements are critical to facilitate our abilities to link genomes to phenomes in all species and will advance scientific discovery and benefit the US animal agriculture.

This pangenome project will increase our understanding of how various genomes result in different phenomes and adaptation in different environments.

**Pangenomes will empower scientists to more effectively predict the physiological outcomes from genomic variation towards the improvement of breeding strategies in different environments.**

# Acknowledgements

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, award number USDA-NIFA-2021-67016-33416.

USDA

NIFA

University of Idaho
Department of Animal, Veterinary and Food Sciences

UtahState University

ROSLIN

USDA ars

M U

OSU

T2T

T2T
TELOMERE-TO-TELOMERE CONSORTIUM

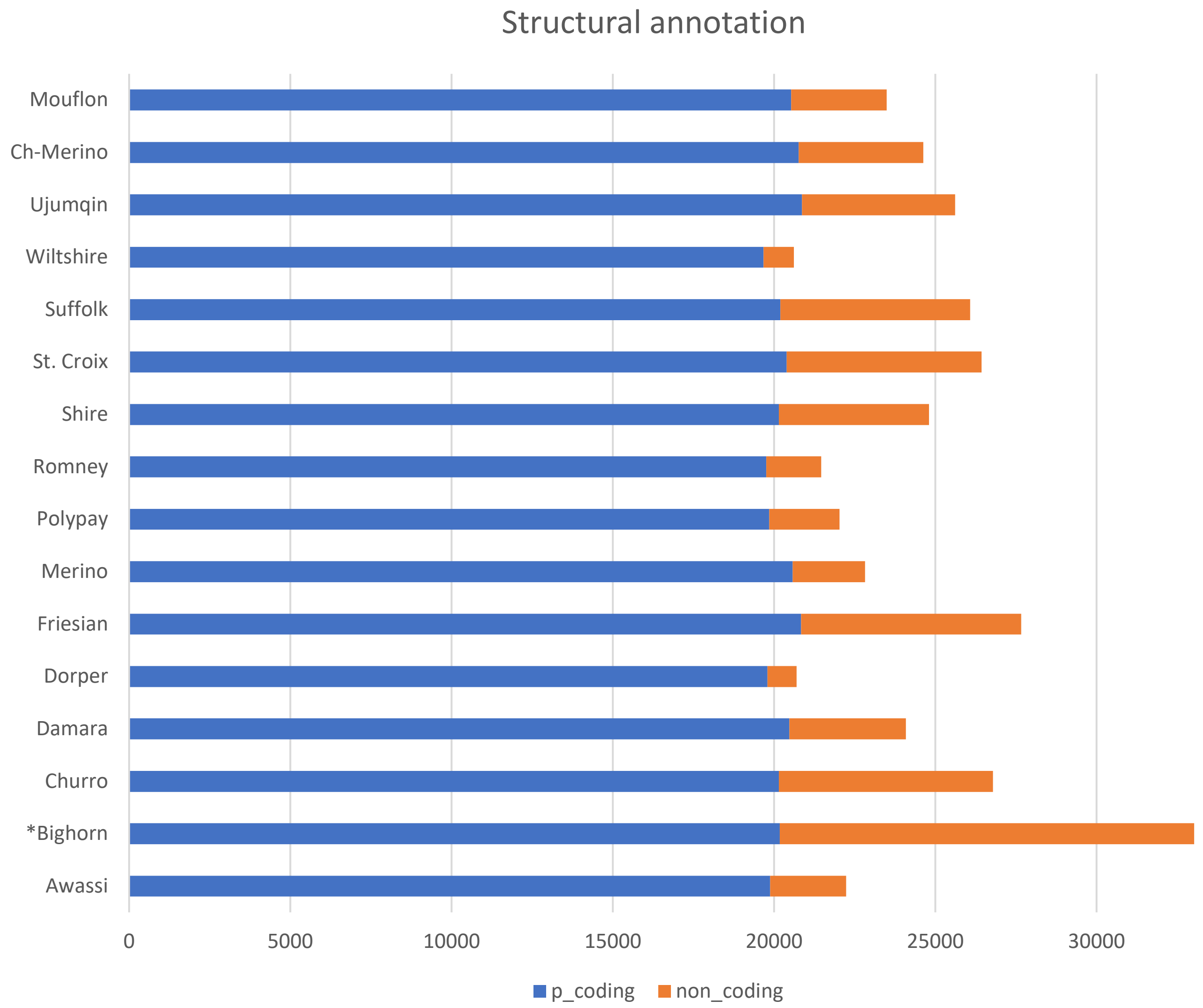
FAANG
Functional Annotation of Animal Genomes

agresearch

ISGC
International Sheep Genomics Consortium

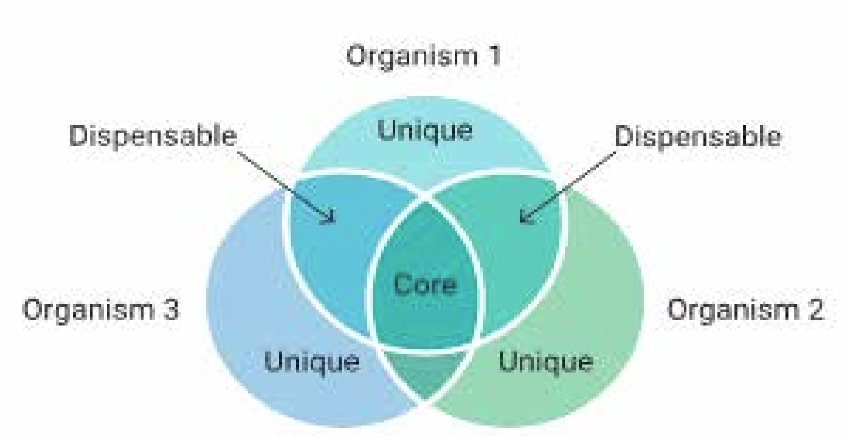
# Questions ?

# Development of the Ovine Pangenome



**Gene annotation sources**

- *NCBI RefSeq
- NCBI Egapx
- Liftoff



Source: PacBio

| Core | Dispensable | Unique |
|---|---|---|
| 13,860 | 3,291 | 155 |