ISGC March 2023 meeting call

Present: Shannon Clarke, John McEwan, Michelle Mousel, Brenda Murdoch, Emily Clark, Tim Smith, Rudiger Brauning, Lingzhao Fang, George Liu, Susan Johnston, Andrew Hess, Iona MacLeod, Gwenola TOSSE-KLOPP, Josephine Pemberton

Agenda:
1. “1000” Sheep Genome Run 3 update
2. Ovine HD SNP chip - where to next
3. AgR Sheep Genomics 60K plus chip
4. Farm Gtex-George Liu
5. Other business

Meeting notes:

1. “1000” Sheep Genome Run 3 update
   Run3 >3500 animals
   - Metadata for all 3522 animals used in run3 is available on figshare https://doi.org/10.6084/m9.figshare.21775100.v1
   - Animals are aligned against the Rambouillet_v2 (ARS-UI_Ramb_v2.0) reference genome
   - Variants are detected using GATK best practice guidelines as implemented in https://github.com/ESR-NZ/human_genomics_pipeline

   Reference genome
   - ARS-UI_Ramb_v2.0+ mitochondrial genome + chrUn
   - adding a Y chromosomal sequence to the Rambouillet ewe
     - this new Y assembly will be ready soon (upload to NCBI to ensure coordinates are fixed)

   Method (rudiger.brauning@agresearch.co.nz) Utilising one pipeline (BWA; GATK haplotypeCaller for SNPs and indels).
   Aim to provide users with variants defined by chromosomal location, functional annotation, animals of interest, breeds of interest.
   - BioSample IDs.
   - Include all samples from genus Ovis if
     - present in run2
     - WGS & Illumina & >= 14.5 Gbp & >50% Ovis content

Note:
- limited metadata for many of the samples submitted (e.g. breed, country of origin, sex)
- a breed prediction will be carried out
- Once completed, Run 3 including breed prediction & data will be made available at sheepmap.org.
-variants positioned on ARS-UI_Ramb_v2.0 functionally annotated genome made available at ensembl.
-collaborate with T2T-share vcf when complete
-update from Emily Clark and ARS-UI__Ramb_v2.0 genome onto Ensembl as the main annotated reference genome for the community:
  * the Ensembl Genebuild pipeline is currently over estimating the number of protein coding genes. Could be a repeats related issue and they see the same thing with other genomes including water buffalo, so it's not specific to sheep or the technologies used so far as they can tell.
  * re-run the Genebuild pipeline for Rambouillet v2 with the latest code, but the issue they were seeing with the inflated count for protein-coding genes is even more pronounced. They now believe it's related to transposable elements with protein-coding like ORFs, rather than repeats, but both could be an issue. The manual annotation team, HAVANA, are going to looking into it to put an actual solution in place. It is not only a problem with sheep, they also saw it in water buffalo, but some of the sheep assemblies seem to have the most pronounced issues with it.

2. **Ovine HD SNP chip—where to next**

- Illumina are phasing out the current HD format and this affects the Ovine HD chip. Illumina do have a new chip format that we could transition too, but this would require a large volume consortium order to get competitive pricing.
- AgResearch has a limited number of HD chips in stock for next 12 months—contact shannon.clarke@agresearch.co.nz if you required HD genotyping
- AgResearch have investigated an alternative to the HD Chip with a pilot project with Gencove. This uses low-pass sequencing and imputation.
  * Ref genome: ARS-UI__Ramb_v2.0
  * Ref panel: ~1000 animals -predominantly the animals present in Run 2
  
  https://figshare.com/articles/dataset/SheepGenomesDB_run2__list_of_animals/11530665
  * 164 samples (with HD genotypes) sequenced to ~0.5x and compared imputed vcf with genotype HD array data
  * ~96% of sites in array were called in imputed vcf and mean of 99.2% concordance (leave one out) for 540K PASS sites (see PAG2023 ISGC presentation).
  * Emily Clark is also trialling with African breeds.
  * contact shannon.clarke@agresearch.co.nz for a quote for low-pass sequencing and return of vcf and HD array equivalent genotypes. If previously genotyped with an Illumina array, concordance can be reported.

3. **AgR Sheep Genomics 60K plus chip**

The AgResearch Ovine Genomic 60Kplus XT chip contains ~50K +15K +other
-about to do another update with addon content finalised ~ November 2023.
  * happy to include your SNPs (contact shannon.clarke@agresearch.co.nz)
  * current genotyping cost $26 NZ (~16 USD) (inclusive of DNA extraction, parentage assignment and causative/single gene SNPs reported. Note those genotyping for genomic selection, upload to your preferred provider to estimate gBVS is also available.

4. **Farm Gtex—George Liu**

1) The Sheep&GoatGTEx project are open for RNA-Seq data contribution (discovery or validation populations) by 31th May 2023.

2) The molQTL reference of Sheep&GoatGTEx will be ready for the community by 31th Oct 2023.
3) The Sheep&GoatGTEx is looking for functional genomics data in sheep and goat for understanding the action mechanisms underlying regulatory variants. The deadline of functional data contribution will be 31th Dec 2023.

4) The Sheep&GoatGTEx is looking for large-scale GWAS of many complex traits in sheep and goat to demonstrate the importance of regulatory variants in prioritizing causal variants and genes of complex traits. The deadline of GWAS data contribution will be 31th Dec 2023.

5) The public release of results (including the publications) from the pilot phase of Sheep&GoatGTEx will approximately on 1th April 2024.

6) The Sheep&GoatGTEx is open for collaborations. So if you are interested in, please feel free to contact lingzhao.fang@qgg.au.ak or george.liu@usda.gov.