

An update from the International Sheep Genomics Consortium

S. Clarke¹, S. McWilliam², H. Daetwyler³, A. McCulloch¹, N. Cockett⁴ and R. Brauning¹
on behalf of the members of the consortium

¹AgResearch, New Zealand, ²CSIRO, Australia, ³Agriculture Victoria, Australia, ⁴Utah State University, USA

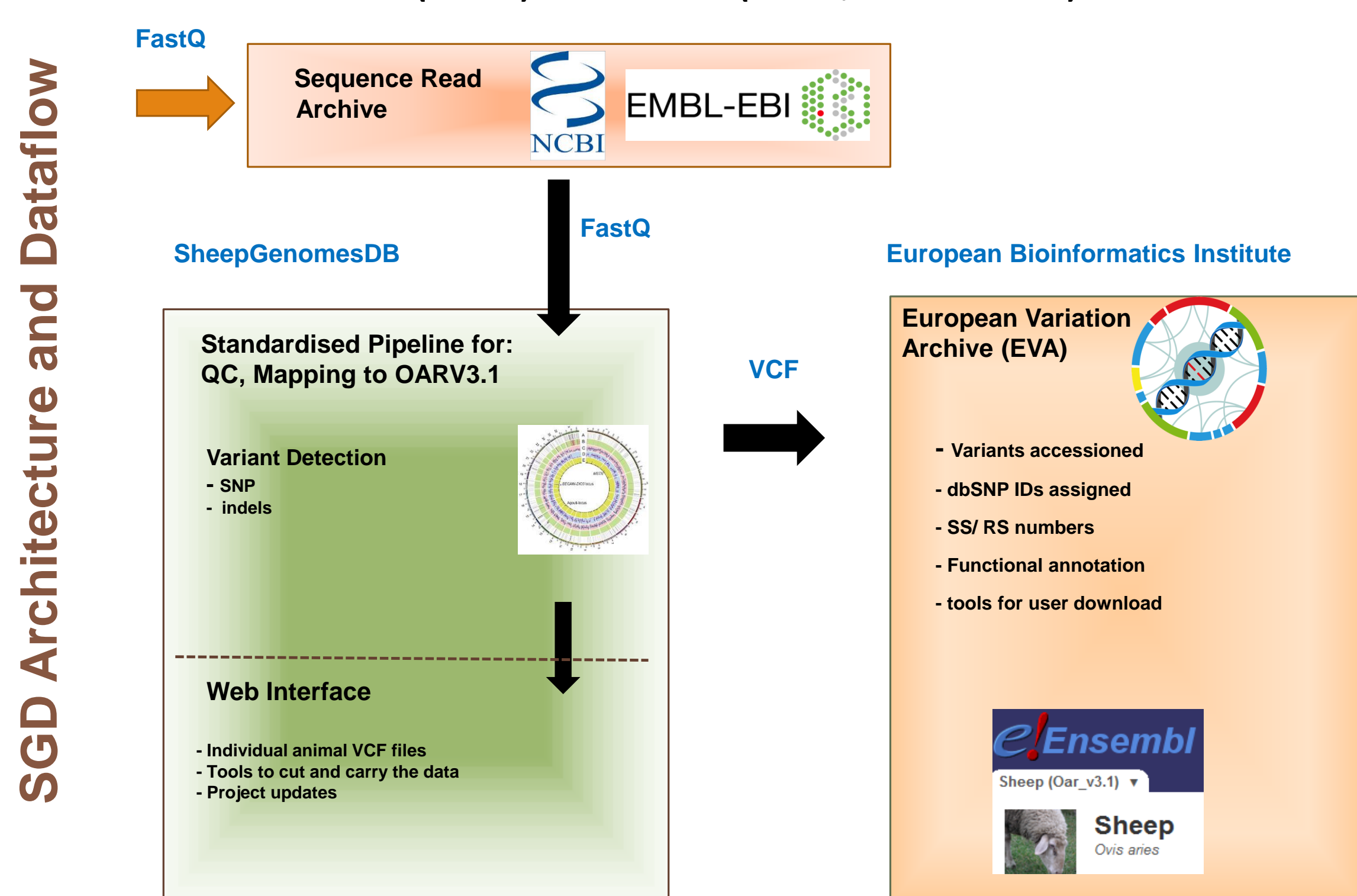


The long-term goal of the International Sheep Genomics Consortium (ISGC) to develop underpinning resources for the sheep research community resulted in continued improvement of the sheep genome assembly, and the development of low, medium and high-density Illumina SNP chips.

The ISGC members have continued to make available whole genome sequence data to the community that has been captured via the Sheep Genomes Database, an initiative of the ISGC that extends the consortium's recent achievements.

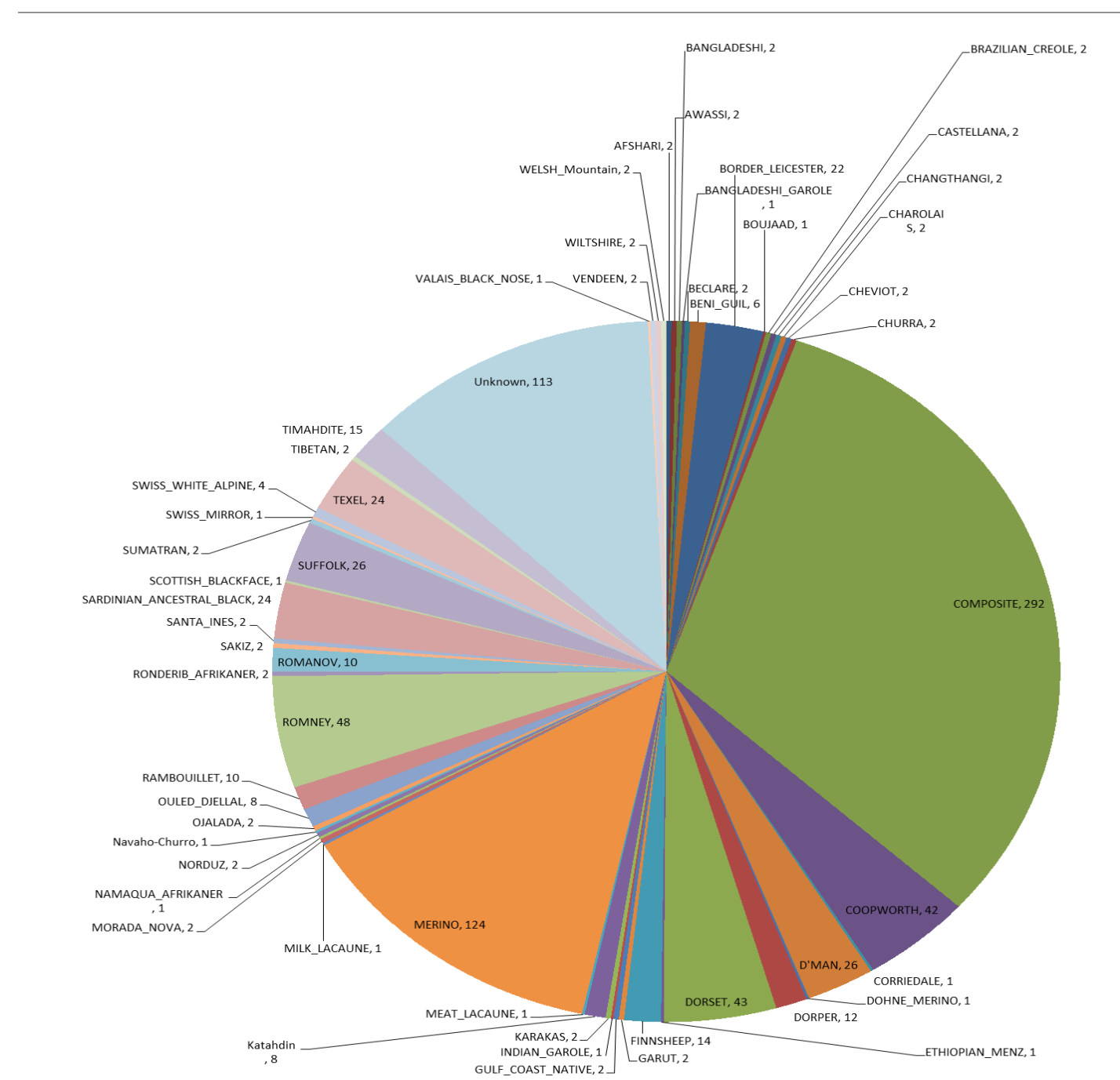
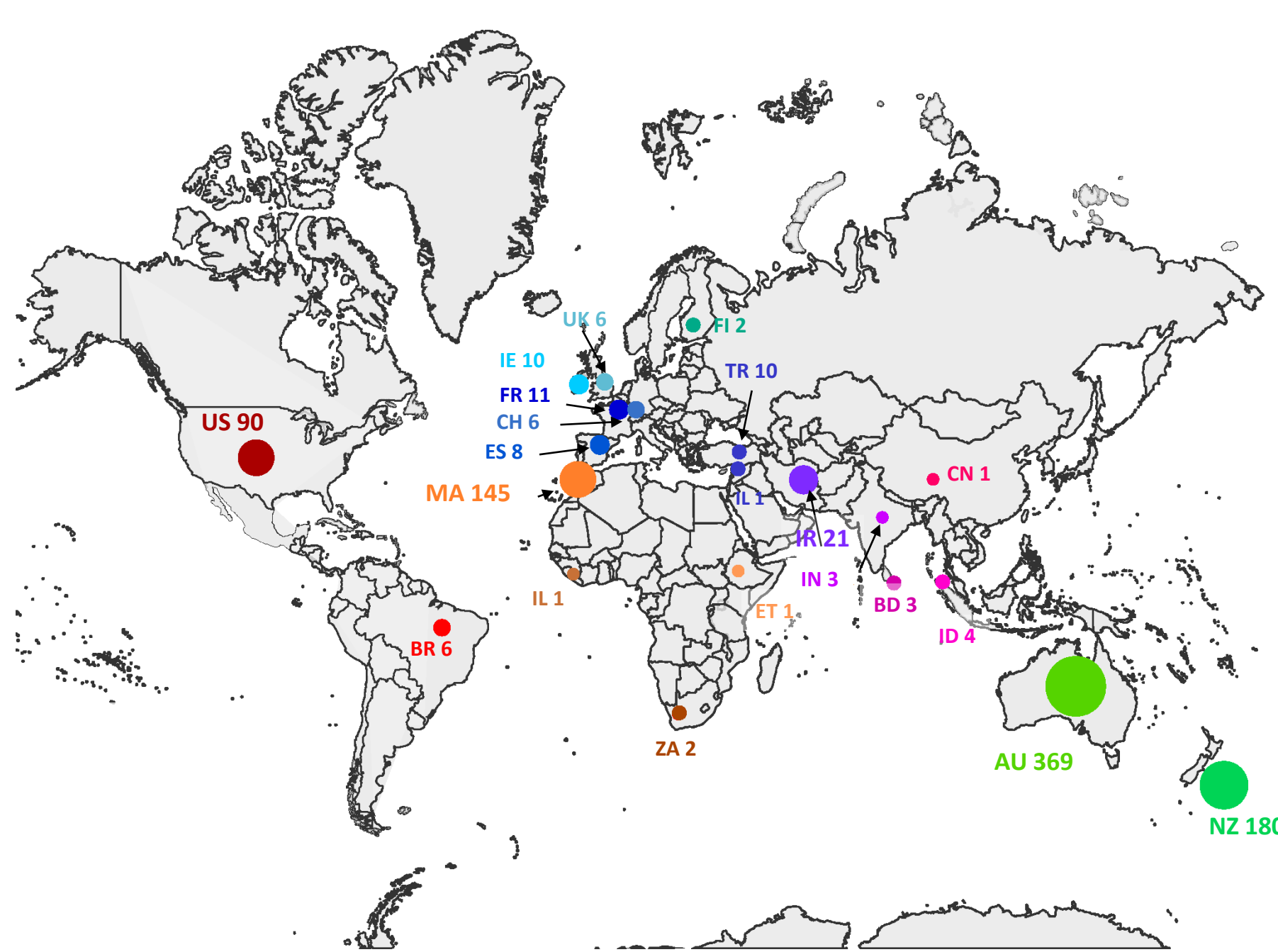
SheepGenomesDB (SGD)

- An electronic warehouse containing sequence variants called from the expanding collection of sheep genomes
- Application of a single harmonised pipeline for read QC, mapping, variant detection and annotation
- Built around NCBI (SRA) and EBI (EVA, Ensembl) resources



Run 2: 935 samples, 21 countries, 69 breeds

- ~50 million filtered variants
- Utilised GATK and Samtools with OAR v3.1 genome assembly
- Completed with results at EVA (<https://www.ebi.ac.uk/eva/?eva-study=PRJEB31241>)

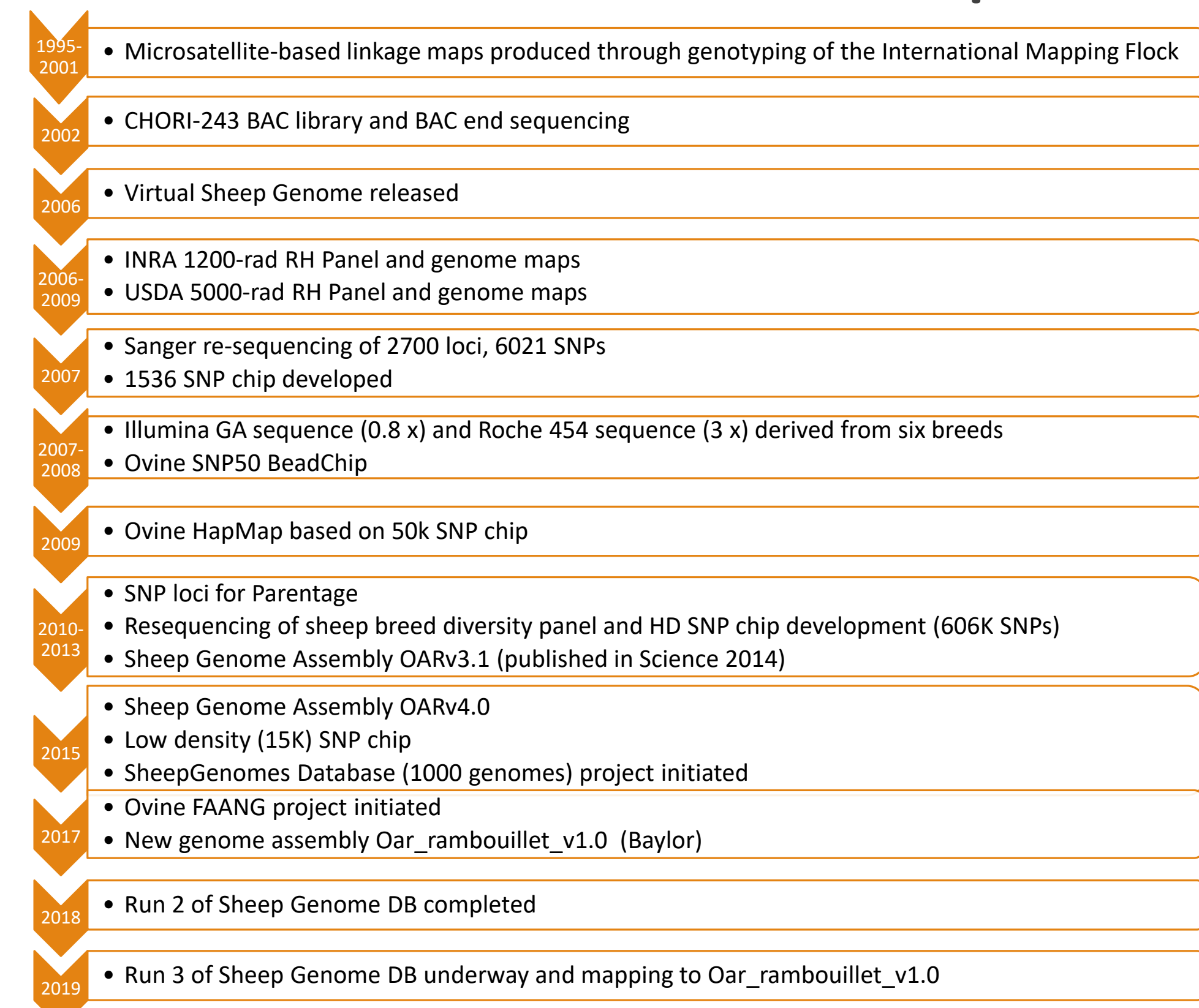


Run 3 - ongoing

- Utilising one pipeline (BWA alignment; GATK haplotype caller for SNPs and indels) with Rambouillet v1 genome assembly
- Whole genome sequence data had to be deposited at SRA and project number sent to Rudiger Brauning (rudiger.brauning@agresearch.co.nz) at AgResearch for inclusion in Run 3
- ~300 new animals to be included
- Aim is to have completed by August
- Updates and the pipeline utilised will be shared via email and sheepgenomesdb.org

<https://www.sheepmap.org/> <https://sheepgenomesdb.org/>

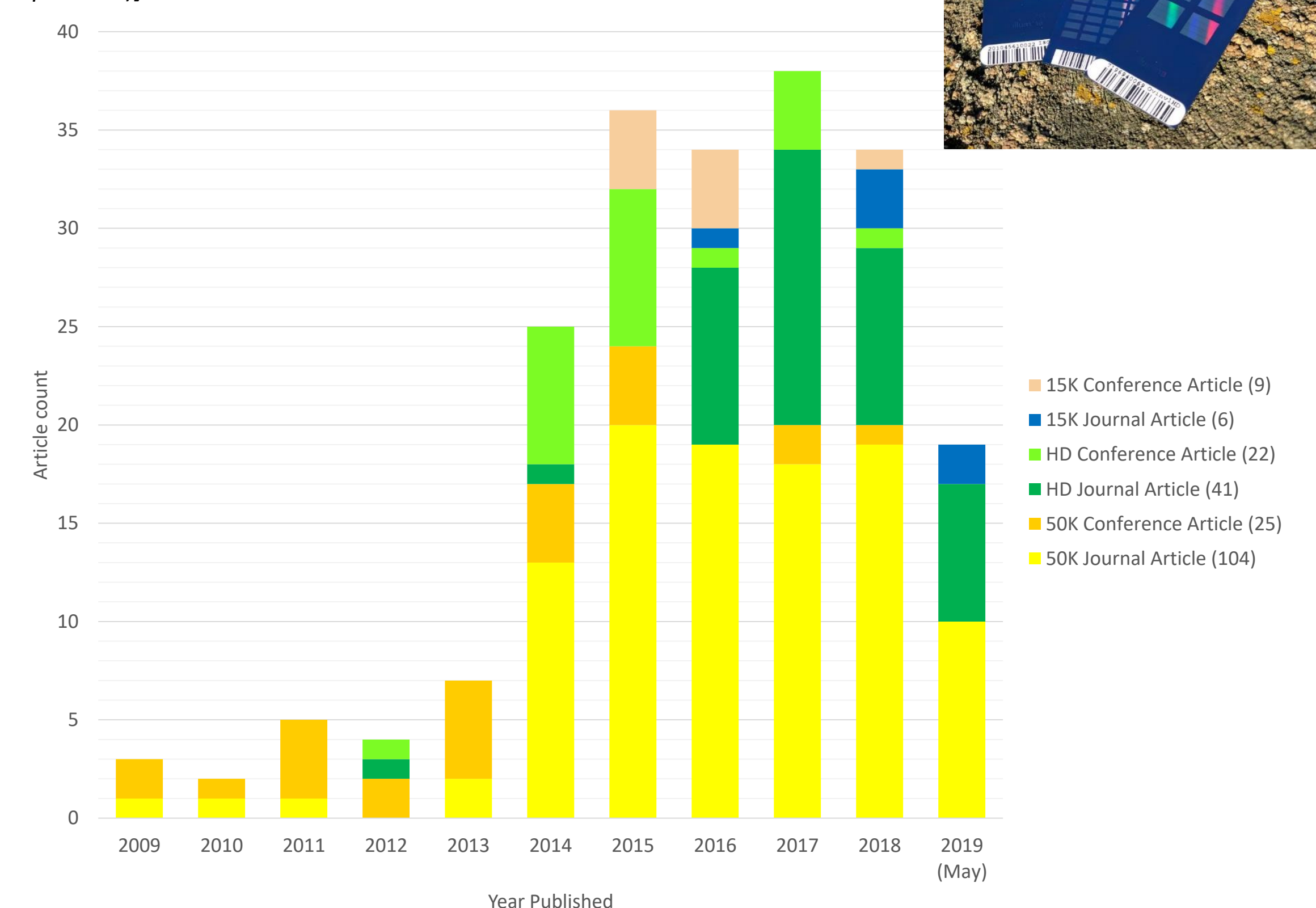
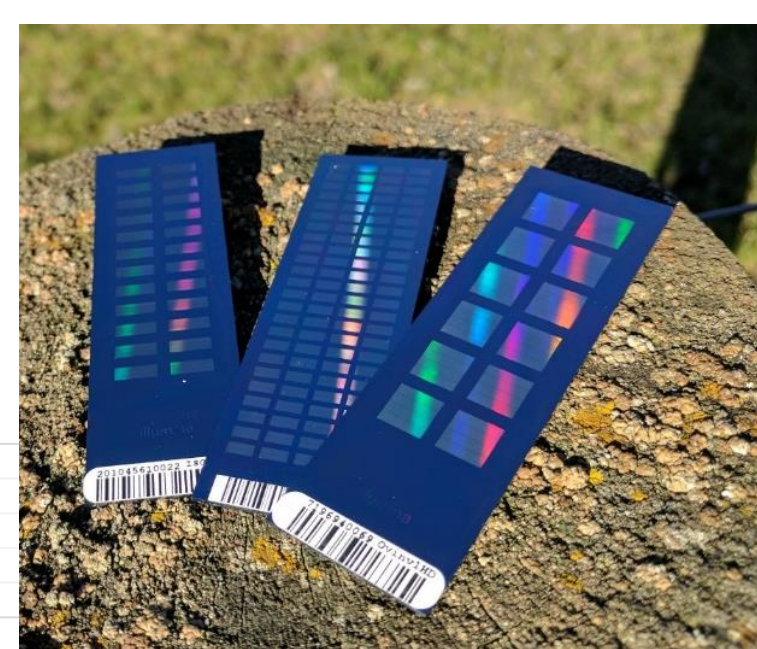
ISGC Genomic tool development



ISGC Illumina SNP chips

- Both the 50k and the 15k SNP chips are publicly available
- The HD (606k) chip currently remains an ISGC consortium chip

[Various low density ovine chips derived from the ISGC chips by individual groups (AgResearch, CSIRO, Teagasc...) to aid genotyping efficacy for genomic selection (S; imputation)]



Mapping of ISGC SNP chip probes to Rambouillet v1 (15k, 50k and HD chips)

- Probe sequences were taken from the Illumina manifests and mapped onto the Rambouillet genome (GCA_002742125.1) using bwa mem v0.7.17-r1188 with default settings (Indels were ignored)
- For each SNP a probe pair was constructed by using AlleleA_ProbeSeq and appending either the reference or the alternative allele.
- Only probe pairs were accepted that passed following filters.
 - both probes are mapped
 - one probe is mapped with 0 mismatches
 - both probes are not multi-mapped
 - no indels were allowed
 - both probes had to map in the same orientation
 - both probes had to map to the same position
 - mapped probes had between them exactly 1 mismatch, the SNP.
- Also mapped to Oar_v3.1 and Oar_v4 to enable comparison of mapping approach to NCBI and Ensembl.
- SNP name, position and allele from the consortium arrays available at <https://doi.org/10.6084/m9.figshare.8424935.v2>