



ARS-led team demonstrates efficiency of new genome assembly method in cattle and yaks

A multicenter collaboration led by scientists at the U.S. Department of Agriculture (USDA) Agricultural Research Service (ARS) has developed a new technique for assembling genomic data that is significantly more accurate than previous methods and can extrapolate parental genome assemblies from data for a single individual.

The original human genome assembly, first described in 2001, contained hundreds of thousands of gaps and tens of thousands of errors, despite investments of billions of dollars and the efforts of thousands of scientists. About eight years later, the cattle genome was first described with similar attributes of gaps and errors, at a cost of tens of millions of dollars and efforts of more than 100 scientists.

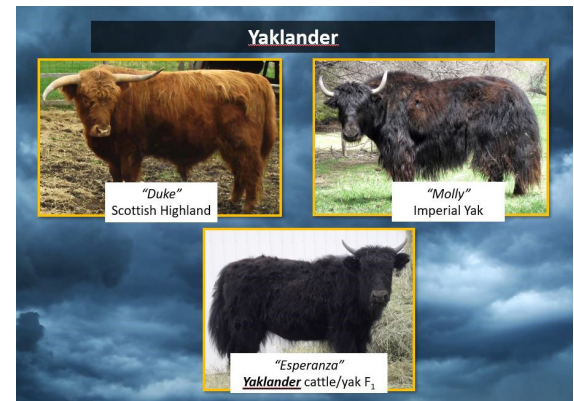
By comparison, the new work—led by the U.S. Meat Animal Research Center of the ARS—involved fewer than 40 total scientists and generated genome assemblies with only a few hundred gaps and a similarly low number of potential errors.

Early genome assemblies were based on inbred individuals, in whom differences between the maternally and paternally inherited sets of chromosomes are minimized, because these differences confused the computer algorithms used to create assemblies from short DNA sequences.

The new method uses long-read sequencing technology and highly heterozygous individuals (in whom the maternal and paternal chromosomes are very different) to generate two separate genome assemblies from that individual—each assembly representing either the maternal or paternal set of chromosomes.

This novel “trio binning” technique was first applied to crosses between diverse cattle breeds, producing reference-quality assemblies of the Angus and Brahman breeds. The technique was then applied to an interspecies hybrid, the offspring of a Highland bull and female yak, with maximal contrast between maternal and paternal chromosomes.

The quality of the resulting individual genome assemblies for both yak and cattle is equal to or better than any existing mammalian assembly, including those of humans or biomedically important species such as mice or rats. Approximately one third of the



Above: The interspecies cattle x yak trio used to create the first genome assemblies that spanned entire chromosomes without gaps in the sequence.

chromosomes for each of the two parental species were assembled with no gaps, achieving “finished” status.

The technology was transferred to the scientific and agricultural communities via the public repositories GitHub (for algorithms) and GenBank (for genome assemblies). All methods and software were described in detail in the journals *Nature Biotechnology* and *GigaScience*.

The effort required a broad range of expertise across multiple federal laboratories and universities. This included animal husbandry expertise from the USYAKS association of yak breeders; veterinary expertise from the University of Nebraska; DNA sequencing expertise and project leadership from the ARS; expertise in development and application of genome assembly algorithms from the National Human Genome Research Institute (NHGRI) and the ARS; and expertise in the development and application of genome scaffolding technology and algorithms from the University of California Santa Cruz and the NHGRI.

The efficiency of trio binning has inspired the creation of the International Bovine Pangenome Consortium, a collaboration to create high quality genome assemblies of all the cattle breeds in existence today. This consortium will work to identify the genetic basis for phenotypic differences between breeds, improve animal health and well-being, and increase sustainability of animal agriculture in the beef and dairy sectors. ☺