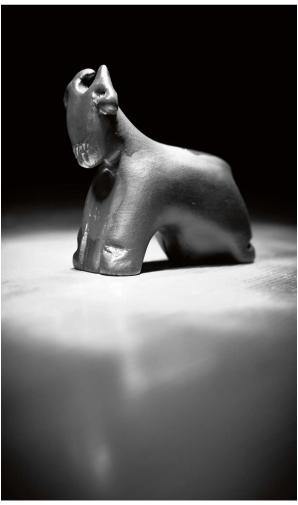# Genomes go platinum

Sequencing a vertebrate genome had almost become routine by 2017, but, with very few exceptions, assemblies of most diploid genomes remained highly fragmented and incomplete. The domestic goat genome ARS1 created a new standard for de novo assemblies of complex genomes.

This was not the first de novo goat genome assembly, but it was by far the most complete. The number of contigs, sequences without gaps, was reduced from more than 330,000 in its predecessor CHIR_1.0 to fewer than 31,000 in ARS1; the number of scaffolds, sequences with gaps, was reduced from 77,431 to approximately 30,000, including all autosomes and the X chromosome. The success of the ARS1 goat genome project hinged on the development

> " The domestic goat genome … created a new standard for de novo assemblies of complex genomes "

Credit: Ian M Butterfield (Concepts) / Alamy Stock Photo

and improvement of multiple technologies that had been used in recent genome assemblies, albeit never at once: high-throughput short read DNA sequencing (MILESTONE 5), PacBio long-read sequencing (MILESTONE 8), optical mapping, and Hi-C chromatin interaction data (MILESTONE 10).

Although low-cost high-throughput sequencing provided a way to quickly generate huge amounts of sequence data, these were not sufficient for assembling genomes from scratch. Long-read sequencing promised great advancements in genome completeness but came with the cost of high error rates. In 2012, Koren and colleagues took advantage of the complementary properties of these approaches to develop a strategy for polishing out the errors in the long reads using highly accurate short reads. Around the same time, scaffolding methods were hitting their stride. Optical mapping, a restriction enzyme-based method originally developed in 1993, was merged with microfluidics to produce high-resolution physical maps that could be used to guide genome assembly. This technology was used by Dong et al. to generate the first domestic goat genome (CHIR_1.0) in 2013, as well as by Seo et al. in 2016 to obtain the most contiguous diploid human genome at the time. In parallel, Shendure and colleagues repurposed the intrachromosomal contacts generated in Hi-C experiments to inform the order and orientation of reads to generate chromosome-length scaffolds.

Bickhart et al. showed that these techniques could be combined to capitalize on each of their respective strengths. Optical mapping can correct orientation errors introduced by Hi-C mapping and assembly errors in PacBio contigs, which in turn are consolidated into longer scaffolds by Hi-C. Both scaffolding methods benefit from very long contigs, made possible by the PacBio reads (with sequencing errors corrected by short reads). Together, these four

ingredients comprise the recipe for a platinum genome.

This approach was noteworthy for another reason: its price tag. Compared with shotgun-assembled genomes, the goat genome was ~3 times more expensive, but compared with existing reference genomes the savings were substantial. This also opened the door for using a fairly small number of reference taxa as anchors for genome assemblies of closely related species, which could then be sequenced using the more affordable shotgun method.

Today, a true platinum genome requires the ingredients listed above, with the potential addition of haplotype phasing and removal of associated false duplications. Trio-binning, that is the use of parental genomes to resolve haplotypes in diploid individuals, has been proposed as the most effective phasing approach, and it is already being implemented in some of the output of the Vertebrate Genomes Project — an effort aimed at producing complete, error-free and phased diploid genomes for every known vertebrate species on Earth.

The drive towards telomere-to-telomere genomes (MILESTONE 17) is taking place at breakneck speed, but the humble goat serves as an important signpost marking the start of the platinum era.

Brooke LaFlamme,
Communications Biology

ORIGINAL ARTICLE Bickhart, D. M. et al. Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nat. Genet.* **49**, 643–650 (2017)
FURTHER READING Koren, S. et al. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nat. Biotechnol.* **30**, 693–700 (2012) | Dong, Y. et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat. Biotechnol.* **31**, 135–141 (2013) | Seo, J.-S. et al. *De novo* assembly and phasing of a Korean human genome. *Nature* **538**, 243–247 (2016) | Burton, J. N. et al. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119–1125 (2013) | Koren, S. et al. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nat. Biotechnol.* **36**, 1174–1182 (2018) | Rhie, A. et al. Towards complete and error-free genome assemblies of all vertebrate species. Preprint at *bioRxiv* https://doi.org/10.1101/2020.05.22.110833 (2020) | Hastie, A. R. et al. Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. *PLoS ONE* **8**, e55864 (2013)