⇥ GENOMICS

Building a giant in tiny steps

A milestone in genome sequencing has been achieved with the publication of a draft sequence of the giant panda genome: the first large eukaryotic genome to be assembled *de novo* using next-generation sequencing technology alone. This paper shows the way to more rapid and cost-effective *de novo* genome assembly and offers numerous insights for evolutionary and conservation biology.

The throughput and efficiency of next-generation sequencing platforms is substantially greater than the capillary-based Sanger sequencing that has been used hitherto for large genome projects, but reads are often <100 bp. Li and colleagues overcame the challenge that these short reads present for assembling long contigs by taking a step-wise approach to genome assembly. From the DNA of a female giant panda they constructed paired-end sequencing libraries with insert sizes ranging from 150 bp to 10 kb. They used the Illumina Genome Analyzer to generate a massive quantity of sequence reads: equivalent to 73× genome coverage with an average length of 52 bp. For the assembly, they first used overlapping reads from libraries with small inserts to assemble short contigs (~1.5 kb), then the authors used paired-end information, step-by-step from the small insert to the large insert libraries, to join the contigs. Their published assembly covers 2.25 Gb, which is ~94% of the whole genome, and has an average contig size of 40 kb.

The authors tested the accuracy of their method at single base, local and large-scale levels and found that it compared favourably to the traditional approach. Also, the contigs and scaffolds were sufficiently long for gene prediction and comparative analyses, which suggests that this assembly has similar utility to Sanger-based genome assemblies.

Can the panda genome help us to understand the biology and evolution of this highly endangered animal? The panda is of the order Carnivora, but is famous for its largely herbivorous diet of bamboo. Li *et al.* found that the panda seems to have retained the genes of a carnivorous digestive system, but they did not find genes that are predicted to be needed for digesting bamboo. The authors suggest that a specialized gut microbiome might facilitate bamboo digestion and that loss of function in a gene involved in tasting protein-rich food might partly explain the diet switch.

The authors also present analyses that indicate the panda genome will prove useful for understanding phylogenetic relationships between carnivores and for determining whether the small remaining panda population has sufficient genetic variability to be sustainable. Therefore, the panda's genome represents a methodological advance and might help safeguard the future of this animal.

Mary Muers

DIGITAL VISION

ORIGINAL RESEARCH PAPER Li, R. *et al.* The sequence and *de novo* assembly of the giant panda genome. *Nature* 13 Dec 2009 (doi:10.1038/nature08696)
FURTHER READING Metzker, M. L. Sequencing technologies — the next generation. *Nature Rev. Genet.* **11**, 31–46 (2010)