Nature Reviews Genetics | Published online 14 Feb 2018; doi:10.1038/nrg.2018.5

John Cancalosi/naturepl.com/Alamy

GENOMICS

Next regeneration sequencing for reference genomes

> the 32 Gb axolotl genome is the largest genome assembled to date

Various species have remarkable abilities to regenerate body parts or entire organisms after injury, but a comprehensive understanding of the molecular basis of regeneration mechanisms will require detailed genomic resources. Two new studies report high-quality reference genomes for two classic regeneration model organisms with contrasting genome sizes: the axolotl salamander *Ambystoma mexicanum* and the planarium flatworm *Schmidtea mediterranea*.

Both studies centred on long-read sequencing using the Pacific Biosciences single-molecule real-time (SMRT) sequencing platform. The multi-kilobase reads enable various repetitive tracts to be spanned, which minimizes assembly breaks and increases assembly completeness relative to short-read sequences. To facilitate the bioinformatic assembly of repeat-rich long reads into contigs, the teams also devised a new *de novo* assembler called MARVEL.

The assembled contigs were then ordered onto larger scaffolds derived from long-range mapping information from Bionano Saphyr hybridization-based optical mapping (for the axolotl genome by Nowoshilow *et al.*) or from Chicago/HiRise sequencing of 3D-proximal locus pairs (for the planarian genome by Grohme *et al.*). From this multilayered pipeline, the resultant reference genomes are substantially more complete than previous draft genomes for these species. Moreover, the 32 Gb axolotl genome is the largest genome assembled to date, and is 29-fold more contiguous than the next-largest current genome assembly, the 22 Gb loblolly pine.

Despite being tenfold larger than the human genome, the axolotl genome has comparable numbers of genes to other vertebrates, but with larger intron sizes and intergenic distances, resulting in the overall large genome size. Interestingly, key developmental genes show more modest intron expansion (6–11-fold relative to human, mouse and frog) than non-developmental genes (13–25-fold), indicating that developmental genes are under size constraint, perhaps to retain tight regulatory control of their expression.

Although the assembled planarian genome is more modest in size at 782 Mb, both genomes were dominated by repetitive DNA and showed clear evidence of recent expansions of long terminal repeat (LTR) retroelements. The planarian genome contained a new class of particularly long >30 kb elements, which is a likely reason why previous short-read approaches failed to generate a high-quality reference genome. Even for these latest long-read reference genomes, LTR elements frequently occurred at the ends of contigs and caused breaks in the assembly, so they are likely to remain a hurdle that limits contig lengths even for long-read genome projects.

One of the main values of these reference genomes is as a resource to facilitate further investigation into the genomic underpinnings of regeneration phenomena. For example, one strategy involves comparing gene and transcript repertoires between species with and without these capabilities to identify putative regeneration genes, and then designing targeted reagents (such as RNA interference or CRISPR tools) to test the functional consequences of perturbation.

As one intriguing example, Nowoshilow *et al.* examined species-restricted transcripts in the limb blastema (from where axolotl limbs are regenerated) and identified a potential broad role for Ly6-family cell-surface proteins in regeneration. Other noteworthy biological inferences from the genomes and transcriptomes were that axolotls have lost the developmental gene *Pax3*, but the authors showed using gene editing that *Pax7* has evolved to carry out the functions of both *Pax3* and *Pax7*, and Grohme *et al.* demonstrated that planarians have lost the spindle assembly checkpoint genes *MAD1* and *MAD2*, which are almost universally essential in other species; checkpoint functions were instead provided by the ROD–ZW10–ZWILCH (RZZ) complex.

It will be fascinating to see what further functional insights into regeneration and development can be gained from these new genomic resources, including testing putative regenerative roles for species-restricted genes and repetitive elements.

Darren J. Burgess

ORIGINAL ARTICLES Nowoshilow, S. *et al.* The axolotl genome and the evolution of key tissue formation regulators. *Nature* https://doi.org/10.1038/nature25458 (2018) | Grohme, M. A. *et al.* The genome of *Schmidtea mediterranea* and the evolution of core cellular mechanisms. *Nature* https://doi.org/10.1038/nature25473 (2018)