Comparative genomics in vertebrates: a role for the platypus

Stephen J. O'Brien

Laboratory of Genomic Diversity, National Cancer Institute, Frederick, MD 21702-1201, USA.
Email: steveobrien@mail.nih.gov

What an uncommon treat for a student of comparative genomics to witness the assembly of the Australian and International science community in Victor Harbour, South Australia, to celebrate the completion and assembly of the platypus whole genome sequence (WGS) (Warren *et al*. 2008; Warren and Grützner 2009). We've had a human genome sequence for 6 years, mouse WGS for a few, and dog and cat WGS only very recently. These and some 28 additional mammal species were nominated and funded for genome sequencing by the US National Human Genome Research Institute (NHGRI) as an adjunct to the human genome sequence primarily to provide an evolutionary context for understanding and interpreting the human genome organisation and all its extraordinary puzzlements (O'Brien *et al*. 2001; Green 2007; NHGRI 2009). The addition of the platypus adds enormous evolutionary depth because it is an outgroup to all these species.

Comparative inference in biology is not new: it began with comparative anatomy of cats and pigs to inform 17th century surgeons, and continued through comparative physiology, neurology, biochemistry and other biological disciplines. By the 21st century, the time has come for comparative genomics, as the technology and computer routines to sequence, assemble maps and align unimaginably large genome sequence sets of 3 billion nucleotide letters are now available.

When NHGRI staffed its 'Human Genome Annotation' Committee, chaired by Robert Waterston, in 2004, it nominated and funded whole genome assessment of selected mammals, anticipating four principal goals for this bold sequencing initiative: (1) to discover evolutionarily conserved sequence motifs that are responsible for regulatory and other critical genomic functions, particularly outside of protein coding genes; (2) to provide a framework for reconstruction of genome organisation, content and dynamics that have occurred during the mammalian radiations; (3) to empower new animal models of human disease and heritable phenotypes; and (4) to provide a starting point for assessment of the expansion, contraction and adaptation of gene families in different evolutionary lineages. All these goals are laudable and ongoing, and for reasons of evolutionary context the platypus genome sequence is a critical keystone of an amazing comparative genomics endeavour.

The platypus is a basal mammal, but only just. The paleontological and geologic records tells us that the earliest vertebrates (mostly jawless fishes) first appeared around 600 million years ago (MYA), morphing to primitive tetrapod amphibians and reptile ancestors 250 million years later (Hedges and Kumar 2009). The earliest mammals began their divergence from reptiles some 300 MYA when they relinquished eggs, scales, venom and feathers (modern birds are a polyphyletic inclusion of reptiles, notably dinosaurs; see Fig. 1) to warm-blooded animals with fur coats, giving birth to and nursing live young, having four-chambered hearts, and (in one species) remarkable cognitive neurodevelopment. Mammal proliferation fuelled a dynamic grand experiment of diverse ecological
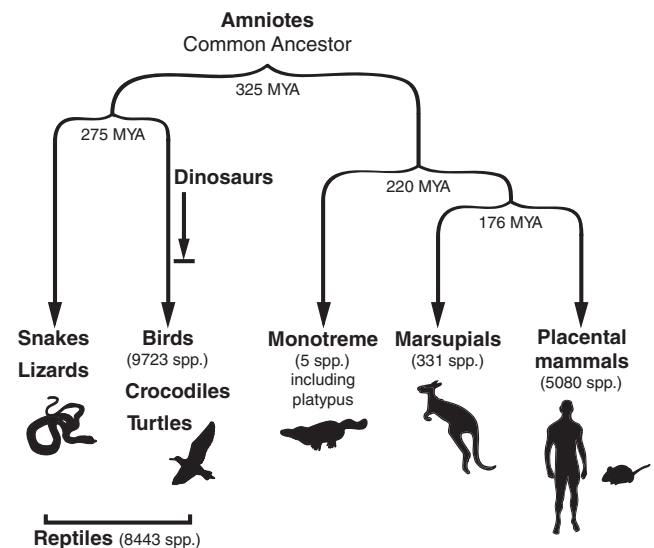
Fig. 1. A curious sort of mammal. Phylogenetic interpretation of the evolutionary relationship among mammals, reptile and birds with approximate dates for the divergence nodes among the different groups (see Hedges and Kumar 2009). The platypus remains a divergent outgroup in the monotreme lineage, which has only five living species (one platypus and four echidnas). The number of living species in other mammalian groups is listed in parentheses.

opportunity, mutational modification and accumulating adaptations.

Nineteenth century zoologists described one group of bizarre Australian species, the monotremes (the platypus and two echidna genera) that had attributes of mammals, birds and reptiles all wrapped together (Moyal 2001). Living monotremes are considered to be the descendents of a 220 million-year-old lineage of the early mammals (Madsen 2009). Given the amalgam of mammalian, avian and reptilian biological features, the platypus and echidnas are considered to be living 'missing links' between the two epochs of reptilian and mammalian pre-eminence (Fig. 1). When specimens of the platypus were first examined by British naturalists they suspected a taxidermy hoax or a species created by a government committee. Neither postulate proved true; rather, the platypus simply evolved in its own way and demonstrated its ability to survive so many millions of years, preserving a marvellous link between two major dominations of the planet, reptiles and mammals.

The platypus (*Ornithorhynchus anatinus*) lives today in freshwater streams and lakes in Eastern Australia and Tasmania, feeding on worms, insect larvae, frogs, shrimp and crayfish. Like other mammals, this curious species is warm-blooded (albeit at a low average temperature of 30–32°C) with a four-chambered heart, produces milk for nursing through abdominal glandular patches (not nipples), and has a furry coat. But, like reptiles, the platypus has a single cloaca vestibule that opens to the intestine, exocrine ducts and genital tubules (hence its ordinal name Monotremata). It also has an elaborate venom production mechanism on its hind limbs that delivers potent snake-like venom to intruders through a bony spur. Like birds, the platypus has webbed feet and a bill superficially like a duck, lacks adult teeth, lay eggs and incubates them to hatching.

Similarities to birds and reptiles extend to the sex chromosomes. Nearly all mammals have two sex chromosomes (X and Y) that are monophyletic, whereas the platypus has 5 X and 5 Y chromosomes that during meiosis assort in a chain of interleaving translocation chromosomes (Grutzner *et al*. 2004), reminiscent of some termite species (Syren and Luykx 1977). The platypus X chromosomes share no homology with the sex chromosomes of therian mammals, but extensive homology with the chicken/bird sex chromosome (Z), indicate that the platypus, and probably its cousin monotreme species the echidna, follow the avian Z/W sex determination and not the more recently derived X/Y strategy of the therian mammals (Veyrunes *et al*. 2008).

The many fascinating features of the platypus genome are described by Warren *et al*. (2008) previously and here (Warren and Grutzner 2009), alongside a treasure trove of derivative studies about genome aspects of platypus biology from venom, reproduction, digestion, egg laying, cytology, molecular ecology, disease immune defences and sex determination dynamics in mammals. There is yet more to mine, including the genetic circuits that drive electrosensory physiology adapted for underwater foraging, endogenous retroviral families, nuclear mitochondrial DNA (numt) loci, pseudogenes, single nucleotide polymorphisms (SNPs) and several paralogous gene families that compare to the mammalian and reptilian radiations.

An interesting lesson from the first attempt to annotate the platypus genome assembly is the special difficulty for a species that is so distant from its closest comparative genome sequence. The platypus diverged ~220 MYA from all other mammals, and its only close relatives are two echidnas. Mostly because of these distant evolutionary roots, the sequence assembly and comparative analysis of the platypus genome presented a huge challenge. Although sequenced to ~6-fold coverage using a combination of whole genome shotgun (WGS) and BAC sequencing, only 409.4 Mbp (22.3%) out of 1.84 Gbp of the sequence assembly (or 17.4% of 2.35 Gbp whole genome) could be ordered on the 21 platypus autosomes chromosomes and five X chromosomes, and then only with the assistance of 279 markers physically mapped by fluorescence *in situ* hybridisation. Between the FISH-mapped anchor markers, scaffolds had therefore to be ordered by comparative inference with eutherian mammalian genome assemblies (Warren *et al*. 2008). Such an approach is reasonable, but likely to introduce mapping errors due to any chromosome rearrangements that occurred during the monotreme lineage. This limitation also precludes a confident high resolution analysis of conserved syntenic blocks (stretches of chromosome containing homologous DNA sequences in the in two compared mammal species) in the platypus relative to other mammals (O'Brien *et al*. 1993; Murphy *et al*. 2005; Pontius *et al*. 2007). These deficiencies can be easily remedied by construction of a denser physical map, either FISH, radiation hybrid, or fingerprint based (Lewin *et al*. 2009).

Future refinement of the platypus genome and comparative genomic analysis would be an invaluable contribution to reconstructing the genome organisation of the ancestors of all mammals. Further, as sequencing costs plummet to US $1000–10 000 per genome, we can envisage the extension of the 32 mammals nominated by NHGRI for WGS to the Genome 10K era, when the WGS of 10 000 vertebrate species, including most available reptiles, birds and mammals, will also be sequenced (G10KCOS 2009). The platypus and echidna genome sequences will certainly be important to anchor the outgroup for this massive analytical undertaking.

Today's geneticists can now dig very deeply into the code of life prescribed by genome sequences, in some cases in such depth as to lose sight of the biological conundrum they intend to study. The platypus genome provides an antidote to such diversion, the unique biology of this species is sufficiently amazing to prevent such distraction. The papers in this volume really do extend 'beyond the platypus genome', but together the genome and biology provide that for which all students of biology seek, an understanding of the development, survival and strategy of evolution's wondrous creations.

References

G10KCOS (2009). Genome 10K: a proposal to obtain whole genome sequence for 10 000 vertebrate species. *Journal of Heredity*, in press.

Green, P. (2007). 2x genomes – does depth matter? *Genome Research* **17**, 1547–1549. doi: 10.1101/gr.7050807

Grützner, F., Rens, W., Tsend-Ayush, E., El-Mogharbel, N., O'Brien, P. C., Jones, R. C., Ferguson-Smith, M. A., and Graves, J. A. M. (2004). In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* **432**, 913–917. doi: 10.1038/nature03021

Hedges, S. B., and Kumar, S. (Eds) (2009). 'The Timetree of Life.' (Oxford University Press: Oxford.)

Lewin, H. A., Larkin, D. M., Pontius, J., and O'Brien, S. J. (2009). Every genome sequence needs a good map. *Genome Research*, in press. doi: 10.1101/gr.094557.109

Madsen, O. (2009). Mammmals (Mammalia). In 'The Timetree of Life'. (Eds S. B. Hedges and S. Kumar.) pp. 459–461. (Oxford University Press: Oxford.)

Moyal, A. (2001). 'Platypus.' (Allen & Unwin: Sydney.)

Murphy, W. J., Larkin, D. M., Everts-van der Wind, A., Bourque, G., Tesler, G., *et al*. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science* **309**, 613–617. doi: 10.1126/science.1111387

NHGRI (2009). Multiple mammalian genomes for comparative annotation. Available at http://www.genome.gov/25521745 [Verified 98 October 2009]

O'Brien, S. J., Womack, J. E., Lyons, L. A., Moore, K. J., Jenkins, N. A., and Copeland, N. G. (1993). Anchored reference loci for comparative genome mapping in mammals. *Nature Genetics* **3**, 103–112. doi: 10.1038/ng0293-103

O'Brien, S. J., Eizirik, E., and Murphy, W. J. (2001). Genomics: on choosing mammalian genomes for sequencing. *Science* **292**, 2264–2266. doi: 10.1126/science.1059393

Pontius, J. U., Mullikin, J. C., Smith, D. R., Agencourt Sequencing Team, Lindblad-Toh, K., *et al*. (2007). Initial sequence and comparative analysis of the cat genome. *Genome Research* **17**, 1675–1689. doi: 10.1101/gr.6380007

Syren, R. M., and Luykx, P. (1977). Permanent segmental interchange 1300 complex in the termite Incisitermes schwarzi. *Nature* **266**, 167–168. doi: 10.1038/266167a0

Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., *et al*. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. *Genome Research* **18**, 965–973. doi: 10.1101/gr.7101908

Warren, W. C., and Grützner, F. (2009). The enigma of the platypus genome. *Australian Journal of Zoology* **57**, 157–165. doi: 10.1071/ZO09051

Warren, W. C., Hillier, L. W., Marshall Graves, J. A., Birney, E., Ponting, C. P., *et al*. (2008). Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183. doi: 10.1038/nature06936