Comparative genomics

Kangaroo genetics leaps ahead

JA Marshall Graves

Heredity (2002) 89, 406–407. doi:10.1038/sj.hdy.6800191

The Human Genome Project is more or less concluded. The Kangaroo Genome Project has made a small beginning with the first comprehensive linkage map of a marsupial (Zenger *et al.*, 2002).

To help interpret the massive amounts of data from the Human Genome Project, the genomes of other vertebrates are also being sequenced. Homologous regions can be lined up and genes and other important signals identified by their relative conservation — so-called phylogenetic footprinting. Mouse is nearly complete; rat, chimp, pets and farm animals are on the way, as are chicken and two fish species.

This array of genome projects spans 450 million years of evolutionary divergence from humans. Yet there is an awkward gap between other eutherian ('placental') mammals that radiated about 80 million years ago, and birds, which diverged from mammals about 310 MYA. Mouse is too close to guarantee that similarities are not just evolutionary hangovers, and birds are so far away that signal is lost, and homology may even be hard to spot.

Marsupials fill this gap. Like mice, marsupials are mammals, but they are much more distantly related to humans, having diverged about 130 MYA (Figure 1). Comparison of their genomes with those of humans and mice should afford plenty of variation in gene sequence, arrangement and function.

A marsupial genome sequence would be an important addition for a number of other reasons. Marsupials differ from eutherians in genetic control mechanisms, like X chromosome inactivation, that are highly conserved in eutherian mammals. Physiological processes like lactation and embryogenesis, are also done rather differently in marsupials. Having a marsupial genome sequence could be a crucial step towards dissecting the genetic basis of these processes in our favourite eutherian, *Homo sapiens*.

It might therefore be time to get serious about marsupial genomes — perhaps even choose one for full sequencing. Which marsupial? There are 269 species to choose from, 200 living in Australia, and the rest in (mostly South) America. Of these, three model species have been

chosen to represent distantly related branches of the infraclass; the tammar wallaby representing the classic Australian kangaroo family, the fat-tailed dunnart *Sminthopsis crassicaudata*, representing an Australian family of small insectivores/carnivores, and the Brazilian grey opossum *Monodelphis domestica*. Ideally we would like to map and sequence the genomes of all three, which are as different as man, mice and aardvarks.

One of these model species is the tammar wallaby, *Macropus eugenii*. Tammars are a small member of the kangaroo family, cheap to maintain and relatively tractable to scientific study. They have been studied intensively for many decades, so their physiology and reproduction are rather well known (Hinds *et al.*, 1990).

The chromosomes of tammars have also been well studied. Like other marsupials, the tammar has a genome much the size of the human, but divided up into only eight (rather than 23) chromosomes that are quite magnificent. The low diploid number is a cytologist's dream, so a lot of fundamental work on chromosome structure, function and evolution has been done on marsupials. Remarkably, the karyotypes of all marsupials are similar enough that an ances-

tral $2n = 14$ karyotype can be deduced from chromosome sizes and banding patterns (Rofe and Hayman, 1985). This has been confirmed by chromosome painting. The beauty of this is that when you've seen one marsupial genome, you've seen them all.

Therefore the tammar wallaby genome is an excellent candidate for full sequencing, and the information from its genome will be instantly transferable to other marsupial genomes. The first step toward some serious marsupial genomics is getting a linkage map that can be used to piece together genomic information.

Until now, only fragmentary information has been available for marsupials — a hotch-potch of gene homologues mapped by different methods in a variety of species (Samollow and Graves, 1998).

Professor Des Cooper's laboratory at Macquarie University in Sydney pioneered tammar genetics. By interbreeding two very divergent tammar populations, the Cooper lab solved the problem of finding markers in natural populations. Tammars from two islands off the coast of South Australia and Western Australia that have been isolated for about 100 000 years proved to have a wealth of fixed differences, both in microsatellites, and even within coding genes (McKenzie and Cooper, 1997). This promised to deliver a system as generally useful for mapping as the *Mus musculus* × *M. spretus* cross has been to mouse genetics. Unlike mouse, however, the tammar is not a laboratory mammal, and breeding and husbandry are not straightforward by any means.

Now Kyall Zenger, Mark Eldridge and

Figure 1 Relationships of kangaroos to other mammals and other vertebrates.

Des Cooper from the Macquarie group have published a paper in *Genetics*, detailing a comprehensive linkage map of the tammar wallaby. This is the first for any marsupial. Most of 64 markers used were microsatellites or anonymous DNA markers, but importantly, 14 are known coding genes. The coding markers allow some of the linkage groups to be anchored to chromosomes, and potentially will allow the map to be aligned with those of other species.

A collection of 353 informative meioses, largely from male hybrid animals, linked all but four markers to at least one other, suggesting that most of the genome is included in the map. The 60 linked markers fell into nine linkage groups. Based on the smattering of physical assignments available, three of these could be assigned to chromosomes 1 and 3, and one to the X chromosome (Graves, 1995).

True, the map is still sketchy. However, it contains many features of immediate interest. Chief among these is

the analysis of the sex differences between recombination rates.

Previous work on limited linkage data in the other two model marsupials, the dunnart and the Brazilian opossum, established that recombination in females was much lower than in males (Bennett *et al*, 1986; van Oorschot *et al*, 1992). This is just the opposite of eutherian mammals, in which recombination is much reduced in males. This exception strikes at the heart of Haldane's venerable hypothesis that recombination is lower in the heterogametic sex, perhaps due to reduced crossing over between differentiating sex chromosomes (Haldane, 1922). The tendency has been to ignore marsupials on the grounds that the data were too fragmentary.

The new and much more complete data from the tammar supports these earlier findings, and implies that the lower recombination in females is an ancient marsupial trait. The authors rightly insist that this reverse sex difference in recombination in marsupials can

no longer be ignored when assessing theories of sex and recombination.

Once again, therefore, marsupials are caught breaking the genetic rules. This is what makes them particularly valuable subjects for genetic study. Perhaps this map will herald the start of a much-needed onslaught on the Kangaroo genome (Wakefield and Graves, 2002). ■

Jennifer A. Marshall Graves is at the Research School of Biological Science, Australian National University, Canberra, ACT 2601, Australia.

e-mail: graves@rsbs.anu.edu.au

Zenger KR *et al* (2002). *Genetics* **161**: 321–330.
Hinds LA *et al*. (1990). *Aust J Zool* **37**: 223–234.
Rofe RH, Hayman DL (1985). *Cytogenet Cell Genet* **39**: 40–50.
Samollow P, Graves JAM (1998). *Inst Lab Animal Res J* **39**: 204–223.
McKenzie LM, Cooper DW (1997). *J Hered* **88**: 398–400.
Graves JAM (1995). *BioEssays* **17**: 311–320.
Bennett JH *et al*. (1986). *Nature* **323**: 59–60.
van Oorschot RA *et al*. (1992). *Cytogenet Cell Genet* **60**: 64–67.
Haldane JBS (1922). *J Genet* **12**: 101–109.
Wakefield MJ, Graves JAM (2002). *EMBO Rep* (in press).

Comparative Genomics

Islands of genetic novelty

JJ Ferretti and AN Suvorov

Heredity (2002) **89**, 407–408. doi:10.1038/sj.hdy.6800192

Two new reports of bacterial genome sequences provide revealing insights into the evolution of virulence mechanisms in an important group of human pathogens. Interestingly, the new work also demonstrates there is substantial heterogeneity among the genomes of this group, even among those of the same serotype that provoke a similar immune response in humans.

Streptococcus agalactiae, also known as group B streptococcus (GBS), is responsible for invasive, and sometimes life-threatening, infections in humans, particularly newborn babies. Neonatal infections of these bacteria have decreased since 1996 after the Centers for Disease Control issued guidelines recommending antibiotic treatment before birth for babies at high risk (Centers for Disease Control, 2002). However, GBS still remains a leading cause of newborn sepsis and meningitis, as well as of severe invasive diseases in adults.

GBS is traditionally classified into serotypes according to the degree to which its polysaccharide capsule stimulates an immune response (ie its antigenicity). Nine serotypes have been identified to date.

Now complete genome sequences for representatives of two of these serotypes, III and V, have been reported in September issues of, respectively, *Molecular Microbiology* and *Proceedings of the National Academy of Science*. Philippe Glaser and his co-authors report the sequence of the circular chromosome of serotype III GBS, which is highly associated with early onset disease in the newborn (Glaser *et al*, 2002). Hervé Tettelin and colleagues have sequenced serotype V GBS, associated with severe invasive diseases of nonpregnant adults (Tettelin *et al*, 2002).

In the *PNAS* study, Tettelin and colleagues used DNA microarrays to do comparative genomic hybridization between the serotype V strain that was sequenced and 19 other strains of various serotypes. Interestingly, at least 18% of the GBS-specific genes were either absent or divergent from these 19 other strains, including those of the same serotype. It seems there is an enormous amount of genetic heterogeneity within these bacterial serotypes.

One of the most intriguing findings of the two reports was the discovery that most genes apparently unique to specific

strains of the same serotype were clustered together in regions (islands). These islands consisted of 7 to 81 kb and encoded at least five contiguous genes. Of the 15 islands identified in the serotype V genome, 10 contained atypical nucleotide compositions differing from the 35.7% G + C content of the entire genome. One possibility is that these divergent islands correspond to horizontal gene transfer events.

The serotype III strain genome had 14 islands that contained the majority of known or putative GBS virulence factors. Importantly, all of these islands also contained sequences known to be associated with mobile genetic elements, eg insertion sequences, proteins of phages, plasmids and transposons. However, the role that these mobile genetic elements had played in the acquisition and spread of these islands remains uncertain because of substantial gene rearrangements among the islands. Such evidence of genetic deterioration may indicate that at least some of these elements may have been present in GBS for a long period of the organism's evolutionary history. Regardless, the strong association of genetic transfer elements and virulence factors in these chromosomal islands suggests that they may be pathogenicity islands and thus have an important role in virulence acquisition and genetic diversity.

Although a large number of phage and plasmid-related genes were found in the