

## NEWS AND COMMENTARY

Zebra finch genome sequenced

# The singing genome

H Ellegren

*Heredity* (2011) 106, 533–534; doi:10.1038/hdy.2010.109; published online 8 September 2010

Adding to a rapidly growing list of organisms that have had their genomes decoded, the genome sequence of zebra finch (*Taeniopygia guttata*) was recently reported (Warren *et al.*, 2010). Being the second bird to be genome sequenced (chicken was the first; ICGSC (International Chicken Genome Sequencing Consortium), 2004), it opens up new possibilities for comparative genomic approaches to studies of evolutionary processes in birds. It also increases the power in using birds as an outgroup in analyses of mammalian molecular evolution.

The zebra finch is a major avian model organism for studies of behaviour, endocrinology and neurobiology. Neuroscientists have been attracted by zebra finches for many years, as songbirds (oscine birds of the order Passeriformes) represent perhaps the most vital model for understanding the cognitive processes behind the trait of vocal learning. This is a complex trait—which includes learning, memory and vocalization—that songbirds only share with humans and a few other avian lineages (parrots and hummingbirds). This made a strong case for zebra finch genome sequencing (Clayton *et al.*, 2009), which was approved and funded by the National Human Genome Research Institute (NHGRI) in 2005. The sequencing work has been completed at the Washington University Genome Sequencing Centre.

The assembled zebra finch genome is 1.2Gb, of which 1.0Gb has been assigned to chromosomes. The genome size in birds is typically 30–50% of that of mammals and, similar to what was observed in chicken genome sequencing, in zebra finch this can largely be explained by the low frequency of mobile repetitive elements (7.7% of the genome). In all 17475 genes were predicted, which is slightly less than what is often considered for mammalian genomes. A set of these genes is differentially expressed in the auditory forebrain in song experiments (Dong *et al.*, 2009) and are thus candidates for involvement in vocalization and learning functions. Moreover, differentially expressed transcripts include a large number of

non-coding RNAs, indicating that other types of sequences than protein-coding genes might mediate complex cognitive behaviours, for example, microRNAs. The same was found in analyses of patterns of gene expression in song control loci of the zebra finch brain, yielding a catalogue of about 800 genes that alter their expression in response to a song. As vocal learning is a derived trait in songbirds, one would expect that some of these genes have been subject to positive selection in the evolutionary lineage leading to songbirds. An analysis of the evolutionary rates of genes shared between zebra finch, chicken and mammals revealed 49 that have both been subject to adaptive evolution in the zebra finch lineage and are suppressed in expression upon song exposure. Interestingly, there is a significant overrepresentation among these of genes with the functional annotation 'ion channel activity' and, indeed, ion channels are critical components of the nervous system. In a more detailed study, Nam *et al.* (2010) demonstrated an overrepresentation of glutamate receptors among positively selected and differentially expressed genes in zebra finch song control loci. Incidentally, the gene *ASPM*, which is associated with microcephaly in humans and has been implicated in the emergence of modern human cognition (a conclusion that has been questioned), is positively selected both in the zebra finch (Nam *et al.*, 2010) and in the human lineage (Evans *et al.*, 2004).

A relatively large number of papers covering particular aspects of zebra finch genomics accompanied the main genome paper. These include topics such as the peptidome (Xie *et al.*, 2010), the degradome (Quesada *et al.*, 2010) and immunogenetics (Balakrishnan *et al.*, 2010), to name but a few. Genome sequencing was also accompanied by the development of a zebra finch linkage map (Stapley *et al.*, 2009; Backström *et al.*, 2010). By integrating the data from the assembly and the genetic map, it is possible to estimate the rate of recombination in different parts of the genome. This revealed a highly heterogeneous recombination landscape with one of the most pronounced

'telomere effects' seen in any species (Backström *et al.*, 2010; Stapley *et al.*, 2010). Up to 90% of recombination events apparently occur in the terminal 10% of the larger chromosomes; the central parts of chromosomes thereby essentially form recombination deserts. One consequence of this is that the degree of linkage disequilibrium is also highly heterogeneous in the zebra finch genome. This means that association mapping should be expected to more easily find correlations between markers and trait loci located in the central parts of chromosomes. On the other hand, however, the low recombination rate in these regions will make it more difficult to fine-map such loci by pedigree approaches.

Recombination rate variation also has implications for genome evolution. Using a chicken tiling path microarray, Völker *et al.* (2010) identified 32 genomic regions that show copy number variation (CNV) distinguishing between chicken and zebra finch. As a clue to the mechanistic basis for structural variation in the genome, it was found that both the incidence of CNVs and chromosomal rearrangements were associated with regional variation in the recombination rate.

Zebra finch genome sequencing was done in the 'standard way', using Sanger technology and by integrating sequence assembly data and information on the physical map via BAC clones. Genome sequencing is now gradually turning to assemblies based on next-generation sequencing (NGS) technology. For example, the 2.25-Gb giant panda genome was recently *de novo* assembled at 56× coverage using short Illumina reads (Li *et al.*, 2010). It is interesting to note that the zebra finch and panda assemblies achieved approximately the same mean length of continuous stretches of assembled sequence (contigs) and of assembled contigs with gaps (scaffolds/supercontigs). This would suggest that about 10 times more sequence data are currently needed to obtain the same *de novo* assembly quality with short reads compared with Sanger sequencing coupled with physical mapping. Clearly, this will come to change with increasing NGS read lengths.

## Conflict of interest

The author declares no conflict of interest. Professor H Ellegren is at the Department of Evolutionary Biology, Uppsala University, Norbyvägen 18D, Uppsala SE-75236, Sweden.

e-mail: Hans.Ellegren@ebc.uu.se

- Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E *et al.* (2010). The recombination landscape of the zebra finch *Taeniopygia guttata* genome. *Genome Res* **20**: 485–495.
- Balakrishnan CN, Ekblom R, Völker M, Westerdahl H, Godinez R, Kotkiewicz H *et al.* (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol* **8**: 29.
- Clayton DF, Balakrishnan CN, London SE (2009). Integrating genomes, brain and behaviour in the study of songbirds. *Curr Biol* **19**: R865–R873.
- Dong S, Replogle KL, Hasadsri L, Imai BS, Yau PM, Rodriguez-Zas S *et al.* (2009). Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc Natl Acad Sci USA* **106**: 11364–11369.
- Evans PD, Anderson JR, Vallender EJ, Gilbert SL, Malcolm CM, Dorus S *et al.* (2004). Adaptive evolution of ASPM, a major determinant of cerebral cortical size in humans. *Hum Mol Genet* **13**: 489–494.
- ICGSC (International Chicken Genome Sequencing Consortium) (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**: 695–716.
- Li R, Fan W, Tian G, Zhu H, He L, Cai J *et al.* (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* **463**: 311–317.
- Nam K, Mugal C, Nabholz B, Schielzeth H, Wolf JB, Backstrom N *et al.* (2010). Molecular evolution of genes in avian genomes. *Genome Biol* **11**: R68.
- Quesada V, Velasco G, Puente XS, Warren WC, López-Otín C (2010). Comparative genomic analysis of the zebra finch degradome provides new insights into evolution of proteases in birds and mammals. *BMC Genomics* **11**: 220.
- Stapley J, Birkhead TR, Burke T, Slate J (2009). A linkage map of the zebra finch *Taeniopygia guttata* provides new insight into avian genome evolution. *Genetics* **179**: 651–667.
- Stapley J, Birkhead TR, Burke T, Slate J (2010). Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Res* **20**: 496–502.
- Völker M, Backström N, Skinner BM, Langley EJ, Bunzey SK, Ellegren H *et al.* (2010). Copy number variation, chromosome rearrangement, and their association with recombination during avian evolution. *Genome Res* **20**: 503–511.
- Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A *et al.* (2010). The genome of a songbird. *Nature* **464**: 757–762.
- Xie F, London SE, Southey BR, Annangudi SP, Amare A, Rodriguez-Zas SL *et al.* (2010). The zebra finch neuropeptidome: prediction, detection and expression. *BMC Biol* **8**: 28.

#### Editor's suggested reading:

- Hansson B, Ljungqvist M, Dawson DA, Mueller JC, Olano-Marin J, Ellegren H *et al.* (2010). Avian genome evolution: insights from a linkage map of the blue tit (*Cyanistes caeruleus*). *Heredity* **104**: 67–78.
- Li MH, Merilä J (2010). Extensive linkage disequilibrium in a wild bird population. *Heredity* **104**: 600–610.
- Hauber ME, Sewell MA, Zuk M (2008). Sequencing one sex or the other has to be justified: gender genomics and equality. *Heredity* **101**: 395.