An ecological model organism flies into the genomics era

Anna W Santure[1]

[1]School of Biological Sciences, University of Auckland, Auckland 1010, New Zealand

Keywords: blue tit, genome assembly, transcriptome, genome browser, ecological genetics

Corresponding author: Anna W Santure, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland Mail Centre, Auckland 1142, New Zealand. Email: asanture@gmail.com

Running title: blue tit genome

Despite the very rapid 'genomicisation' of the field of Molecular Ecology in recent years, there have been relatively few annotated whole genome assemblies of non-model organisms published. Instead, molecular ecologists have more frequently utilised next generation sequencing technologies to develop large numbers of markers or to generate transcriptome data. Whole genome assemblies are more expensive, and require considerable computational resources and bioinformatic expertise. However, the availability of an annotated genome offers exciting opportunities to address fundamental questions in ecology and evolution that are difficult to address with moderate sets of markers or by transcriptome sequencing. Such questions include elucidating the roles of natural and sexual selection in shaping diversity, determining the roles of regulatory and protein-coding change in the evolution of traits, and determining the genomic architecture of sex-specific trait variation. Arguably, these questions are most tractable – and most interesting – in well characterised species for which there is already some knowledge of natural and sexual selection, and of the traits that are most likely to link to fitness. In this issue, Mueller *et al*. (2016) present the assembly and annotation of the genome of the blue tit (*Cyanistes caeruleus*), a model ecological species. In addition, by sequencing the transcriptome of male and female blue tits, the authors identify and annotate sex-biased gene expression, and conclude that noncoding RNA genes are likely to play a significant role in sex-biased expression. By making their assembly and annotation publically available and accessible via a genome browser, Mueller *et al*. (2016) offer exciting possibilities for further research into the genomic basis of adaptation, and investigation of the roles of natural and sexual selection, in this well-studied ecological model species.

The development of genomic resources offers exciting opportunity to better understand and characterise the ecology and evolution of non-model species (Ekblom and Galindo 2011, Stapley *et al*. 2010). A growing number of studies in the field of molecular ecology are employing next generation sequence data to generate large genome-wide marker sets, sequence targeted or

random regions of the genome at high sequence depth, or sequence transcripts across developmental stages, tissue types or sex. In many cases, such resources have enabled us to better answer questions that have engaged us for decades – for example, how are species related? How inbred is this endangered population? What has been the impact of climate and geography on species diversity? However, the opportunity to now generate, assemble and annotate whole-genome data offers the exciting possibility to answer many new questions key to our understanding of ecological and evolutionary processes. Such questions include: what are the loci of adaptation between species, and are the same regions of the genome responsible for within-species variation? How has the landscape of selection and recombination shaped diversity between and within species? What are the roles of regulatory and protein-coding change in the evolution of traits? Are the same genetic changes responsible for adaptation in replicate populations? What is the adaptive potential of the species in a changing world? How is sex-specific trait variation established and maintained, and what regions of the genome are responsible? How might genome features including rearrangements and transposable elements explain divergence and speciation? (Andrew *et al*. 2013; Ekblom and Galindo 2011, Ellegren 2013, Stapley *et al*. 2010).

Arguably, many of these questions are most meaningful in species where the ecology is well understood. A growing number of studies are beginning to demonstrate just how useful and informative it is to layer genomic information on top of a sound understanding of the environmental constraints and biological interactions that define the species ecology (for example, the recent demonstration that beak shape diversity across the iconic Darwin's finches is associated with a transcription factor affecting craniofacial development, Lamichhaney *et al*. 2015). One such well-studied, 'ecological model species' is the blue tit (*Cyanistes caeruleus*; Figure 1), a widespread and common Eurasian species that has been the subject of ecological study since the 1950s. While a number of groups have begun developing genomic resources for the blue tit, including a very recently published set of RAD-seq markers (Szulkin *et al*. 2016), the publication this month of an

annotated whole genome assembly of blue tit (Mueller *et al*. 2016) represents a valuable resource for the molecular ecology community that will facilitate further research into blue tit ecology, adaptation and evolutionary potential.

In what might be considered very good luck, given molecular ecologists' proclivity to study birds, avian genomes are small (typically between 1.2-1.4 Gb) and are remarkably stable, with high levels of synteny and relatively few large-scale genome rearrangements (Ellegren 2013). The recent release of a 'flock' of 48 well-annotated avian genomes (Zhang *et al*. 2014; with more to come) has certainly been facilitated by the excellent assemblies and annotations of the chicken (ICGSC 2004) and zebra finch genomes (Warren *et al*. 2010). This remarkable avian synteny has enabled Mueller *et al*. (2016) to use the closely related Tibetan ground tit genome (Cai *et al*. 2013) to order *de novo*-assembled blue tit scaffolds into superscaffolds across the genome, and has aided the identification and removal of likely contig and scaffold misassembles. By comparing to all publically available avian protein sequences, Mueller *et al*. (2016) have annotated >21,000 coding genes in the genome, with transcriptome sequencing of males and females making possible complete gene annotations, including the identification of untranslated regions, for >17,000 of these genes. Mueller *et al*. (2016) aligned transcripts, including non-coding RNA sequences, to the genome, and demonstrated a significant role for non-coding regulatory RNA elements in sex-biased expression. In addition, the genome has been annotated with over 500,000 SNPs identified from the transcriptome sequencing of 10 individuals. The single male bird that had been genome sequenced was also transcriptome sequenced, which has enabled the validation of SNPs in the transcriptome; it is somewhat remarkable to note that SNP genotypes were identical in 99.99% of the SNPs detected in both transcriptome and genome sequenced regions, and indicates the quality of both assemblies. Interestingly, indels between datasets were much less concordant, indicating that best practice guidelines for RNA-based indel calling still struggle with differentiating true indels from sequencing or assembly error.

A significant challenge remaining for avian genome assembly is the identification, assembly and annotation of the smallest of the ~30 microchromosomes, including the W sex chromosome. It is thought that there are around 40 chromosomes in most avian genomes. In the best annotated of these, the chicken genome (2n = 78), only 32 chromosomes plus 3 linkage groups are described (ICGSC 2004), which is a major limitation when chicken is used as a template to name and identify chromosomes in other avian genomes (including the blue tit). In addition, it is common, as in Mueller *et al*. (2016), to sequence the homogametic sex – male ZZ – in order to accurately assemble the Z chromosome, but this comes at the expense the W chromosome not being sequenced. A full understanding of the nature of natural and sexual selection and trait architecture at a genome scale will require that we better characterise these elusive microchromosomes. Further, while large genome structure is remarkably stable in avian genomes, the detection of local rearrangements of gene order in blue tit and many other avian genomes will require the construction of a genetic linkage map or confirmation of scaffold order and orientation by, for example, long read sequencing.

The assembled blue tit genome sequence and annotation is publically available in an easy to navigate genome browser (http://public-genomes-ngs.molgen.mpg.de), with SNPs and male and female transcripts annotated, along with homology to the Tibetan ground tit, medium ground finch and zebra finch genomes, and superscaffolds assigned to zebra finch chromosomes. Chromosomes are named according to synteny with chicken chromosomes; this rather unusual convention (rather than naming chromosomes by size) has developed since the publication of the zebra finch genome (Warren *et al*. 2010) and means that, for example, in passerines, including blue tit, chromosome 2 is the largest chromosome, rather than chromosome 1, which has undergone a fission in the passerine lineage (Ellegren 2013). However, the consistency of naming chromosomes based on the chicken genome, helped by the relatively stable karyotype in all birds, makes comparisons with other avian genomes straightforward. It is also noteworthy that the publication in a genome browser format is

116 likely to encourage the use of genome information in the molecular ecology community that will find
117 it most useful. Presenting data in a visual format, with clear organisation and the ability to download
118 sequences easily, is to be encouraged for other genome sequencing projects.

References

Andrew RL, Bernatchez L, Bonin A, *et al*. (2013) A road map for molecular ecology. *Molecular Ecology* **22**, 2605-2626.

Cai Q, Qian X, Lang Y, *et al*. (2013) Genome sequence of ground tit *Pseudopodoces humilis* and its adaptation to high altitude. *Genome Biology* **14**, R29.

Ekblom R, Galindo J (2011) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1-15.

Ellegren H (2013) The evolutionary genomics of birds. *Annual Reviews of Ecology, Evolution and Systematics* **44**, 239-259.

ICGSC (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716.

Mueller J *et al*. (2016) Characterisation of the genome and transcriptome of the blue tit *Cyanistes caeruleus*: polymorphisms, sex-biased expression and selection signals. *Molecular Ecology Resources* **16**, XXX-XXX.

Lamichhaney S *et al*. (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**, 371-375.

Stapley *et al*. (2010) Adaptation genomics: the next generation. *Trends in Ecology and Evolution* **25**, 705-712.

Szulkin M, Gagnaire P-A, Bierne N, Charmantier A (2016). Population genomic footprints of fine-scale differentiation between habitats in Mediterranean blue tits. *Molecular Ecology* **16**, XXX-XXX.

Warren *et al*. (2010) The genome of a songbird. *Nature* **464**, 757-762.

Zhang G, Jarvis ED, Gilbert MTP (2014) A flock of genomes. *Science* **346**, 1308-1309.

Figure legend

Figure 1: Adult blue tit, copyright Ueli Raz