Advancing avian behavioral neuroendocrinology through genomics

David F. Clayton a,*, Sarah E. London b

a Biological & Experimental Psychology Division, School of Biological & Chemical Sciences, Queen Mary University of London, London E1 4NS, UK
b Department of Psychology, Institute for Mind and Biology, Committee on Neurobiology, University of Chicago, 940 E 57th Street, Chicago, IL, USA

Abstract

Genome technologies are transforming all areas of biology, including the study of hormones, brain and behavior. Annotated reference genome assemblies are rapidly being produced for many avian species. Here we briefly review the basic concepts and tools used in genomics. We then consider how these are informing the study of avian behavioral neuroendocrinology, focusing in particular on lessons from the study of songbirds. We discuss the impact of having a complete “parts list” for an organism; the transformational potential of studying large sets of genes at once instead one gene at a time; the growing recognition that environmental and behavioral signals trigger massive shifts in gene expression in the brain; and the prospects for using comparative genomics to uncover the genetic roots of behavioral variation. Throughout, we identify promising new directions for bolstering the application of genomic information to further advance the study of avian brain and behavior.

1. Introduction

The sequencing of the human genome (Lander et al., 2001; Venter et al., 2001) triggered a tidal wave of genomics which has spread across all of biology. The wave has now caught up with avian biology, affecting the study of model organisms important for brain, behavior and endocrinology. The resources and tools of genomics may seem alien to scientists schooled in the study of hormones, brain and behavior but are a valuable addition to existing research strategies. In this review, we summarize the current state of avian genomics, with a particular eye to the future: where will genomics have the greatest impact on the study and understanding of avian brain and behavior? Our focus will emphasize the zebra finch as a model organism for behavioral neuroendocrinology, but we also touch on relevant illustrative examples from other species and contexts.

We see three broad areas where the impact of genome sequencing is already being felt. First is the completion of the parts list for a specific organism – that is, the comprehensive description of all the molecular entities encoded by the genome. We show how an accurate accounting of all gene products can resolve long-standing questions in biology and guide future research. Second is the power of genomics to enable analyses of biological organization at a level beyond the reach of traditional methods focused on just one or a few genes or gene products. Studies of large-scale patterns of gene expression and gene networks are already providing unexpected insights into the molecular underpinnings of phenomena ranging from sex differences to vocal communication. Third is the way in which genomics facilitates the study of natural variation, both within and between species. Understanding variation is key to understanding biological systems at all levels, from the molecular (e.g., what is the mechanism of this mutation?) to the ecological (e.g., how has this species evolved to occupy this niche?).

2. What’s in a genome?

We begin by defining a few terms and surveying the current state of avian genomics. “Genomics” comprises a number of different technologies for determining the sequence and/or chromosomal location of a piece of DNA. The DNA sequence being examined may be the genomic DNA of an organism, or synthesized from its RNA (by enzymatically converting RNA to cDNA). A genome assembly refers to the sequence of all the genomic DNA obtained from an organism, after the sequence has been ordered with respect to position along chromosomes. A “transcriptome” refers to the collection of sequences determined from the RNA of a cell, tissue or organism, and thus represents the genetic elements that are expressed (synthesized from the genome) in that biological context. Genome sequencing technologies have been cleverly adapted to support not only the assembly of whole genomes and transcriptomes, but also to measure the relative copy numbers of
individual RNAs (thereby quantifying gene expression), and to compare large numbers of samples for sequence variation at single positions in the genome (Single Nucleotide Polymorphisms, SNPs).

Once a genome assembly is complete or “finished” (more on this below), the parts list for the organism is known, at least in principle. That is, with a complete genome, we have access to the genetic instructions needed to make all the macromolecular products synthesized by the cell (proteins and various non-coding RNAs, ncRNAs), along with the genetic control elements that may regulate their production. Reading this code is still a work in progress. The large ENCODE collaboration has famously estimated that 80.4% of the human genome is functional (I.T.E.P. Consortium, 2012) in that it has some influence on the information expressed by the genome (see Graur et al., 2013 for an interesting counterpoint to ENCODE’s conclusions regarding “function”). Only a few percent of a genome actually encodes protein, and “reading” that part (determining what proteins can be made) is relatively straightforward. Annotations of conserved protein-coding regions and other genome features can be accessed through various browsers on the internet such as those maintained by Ensembl (ensembl.org) and UCSC. The majority of the RNA synthesized from the genome, however, does not code for proteins and these RNAs are still poorly understood.

The first avian genome assembly (chicken) was annotated and published in 2004 (I.C.G.S. Consortium, 2004), followed 6 years later by the second (zebra finch) Warren et al., 2010 (Fig. 1). The pace has quickened since, with the publication of genome assemblies representing Turkey (Dallioul et al., 2010), Flycatcher (Ellegren et al., 2012), Budgerigar (Koren et al., 2012), Macaw (Seabury et al., 2013), Falcon (Zhan et al., 2013) and Rock pigeon (Shapero et al., 2013). At the present, Ensembl annotations are available only for chicken, turkey and zebra finch but more annotations are scheduled to come online in the near future. Even more avian genome assemblies are expected to emerge in 2013–2014 with the progress of the Genome 10K initiative and its affiliated avian component (progress can be monitored at URLs including http://songbirdscience.com/ and http://aviangenomes.org/). Additionally, transcriptomes are being produced for many species to provide a description of genes expressed without the cost and labor of a whole genome assembly. So far these include the great tit (Santure et al., 2011), quail (Wang and Chang, 2011), song sparrow (Srivastava et al., 2012), plover (Mohdad et al., 2013), house finch (Backström et al., 2013) and a number of other species (Künstner et al., 2010).

It is important to appreciate that not all genome assemblies are of equivalent quality – they can differ significantly in both accuracy and completeness. Strategies for both sequence generation and assembly are constantly being refined, and inevitably there are errors and omissions. Full consideration of sequencing methodologies is beyond the scope of this review (Schmutz et al., 2004; Mardis et al., 2002) but common problems in sequencing include repetitive regions that are hard to assemble and regions that are simply hard to sequence due to secondary structure. The gold standard technology at this point is still Sanger sequencing of cDNA inserts cloned into plasmids, which generates highly accurate reads of hundreds of nucleotides but at high cost. Next Generation Sequencing (NGS) has recently been added to the sequencing toolkit; it relies on much more rapid sequencing of much shorter fragments of genomic DNA (or, for transcriptome analysis, of cDNA derived from RNA). To help with ordering the plasmid or NGS reads along chromosomes and to resolve some of the issues that arise from repeated elements, even larger pieces of DNA of different sizes are cloned into vectors, often bacterial artificial chromosomes (BACs), whose ends can be readily sequenced. The majority of the BAC insert may remain unsequenced, but the size of the insert is known and this allows assembly of the BACs into a physical scaffold that can help locate the relative positions of plasmid and NGS sequencing reads within this scaffold. “Finishing” a genome assembly therefore requires the concerted application of a blend of sequencing approaches, and only a few animal genome sequences have reached a status approaching “finished” (Drosophila, C. elegans, mouse, and human). Even the human genome, the first one sequenced – has some remaining gaps, mostly in the centromere regions. A near-finished chicken genome assembly (galGal5, the fourth major publically-released version) on par with the human is anticipated within the next year, as is an improved zebra finch assembly (taeGut2; Warren and Mello, personal communication).

There are several clear technical limitations of current avian genome assemblies. The largest physically is that the W sex chromosome is not fully sequenced (Moghadam et al., 2012; Ayers et al., 2013), and given this precedent the initial assembly of the zebra finch genome focused solely on the Z chromosomes derived from a male; thus the W chromosome is not represented at all in the current zebra finch assembly. This is a significant limitation given that some genetic elements important for sex-typical behavior and neuroendocrine signaling are likely mapped to the W chromosome. Another limitation is that the assemblies tend to have gaps at the 5’ proximal regions of transcribed elements (Fig. 2). This is especially noticeable in the current (first) release of the zebra finch genome assembly. These regions tend to be enriched for regulatory

![Fig. 1. Schematic alignment of the 10 largest chromosomes in chicken (Gga) and zebra finch (Tgu), showing overall synteny. Rearrangements are common within but not between chromosomes. Only the macrochromosomes are shown. Birds also have numerous small chromosomes ("microchromosomes") which are gene-rich but difficult to karyotype. Reprinted from Stapley et al. (2010).](image-url)
elements but were technically difficult to sequence and assemble due to their high GC nucleotide content and possible repeated functional motifs. It will be essential to fill in these gaps to allow study of how genes involved in neuroendocrinology and behavior are regulated.

Despite these limitations, a draft genome sequence has immediate utility for the study of neuroendocrinology and behavior. It provides predicted protein sequences, useful for studies of molecular evolution and as a reference point for developing new antibodies or evaluating the reactivity of existing antibodies to specific protein sequences. It enables identification of predicted gene regulatory regions (e.g., transcriptional start sites, promoters, enhancers, introns, and ncRNA recognition sites) and supports the study of genetic variation associated with phenotypes of interest. These variations include not only single-nucleotide changes, but also the potential for exon-by-exon splice variation and large structural changes (chromosomal insertions, deletions, and rearrangements) that are a major factor in both evolution and disease (Weischenfeldt and Symmons, 2013).

3. Completing the parts list

At the most basic level, genomics provides a list of the molecular parts that comprise the organism. Several examples illustrate how this parts list can suggest new questions for neuroendocrinology and behavior, or answer old ones.

3.1. From parts to functions: olfaction

Genome sequencing has contributed directly to a fundamental insight about the sensory capacities of birds, and by extension, about social communication. Although some birds appeared to use olfactory cues for guidance (e.g., seabirds and vultures), most avian species were widely assumed to have minimal olfactory sensing capabilities, an inference supported in part by their relatively small olfactory bulbs (Roper, 1999; Hagelin and Jones, 2007). Examination of the chicken and zebra finch genomes, however, revealed hundreds of genes with the canonical structure of olfactory receptors (OR) (Steiger et al., 2009). Moreover, the predicted OR gene complement in birds includes a different clade of receptors compared to those in the lizard, suggesting evolution of distinct olfactory specializations in each group, and raising the possibility that examination of multiple bird genomes may reveal genome-encoded olfactory specializations (Steiger et al., 2009). A significant social function for olfactory discrimination in the zebra finch has now been revealed, as fledglings are able to distinguish between kin and non-kin based on olfactory cues alone (Krause and Caspers, 2012; Krause et al., 2012). Clearly, the effects of odor on behavior require much more study but we also note that only ~30–60% of the predicted OR genes are complete (Steiger et al., 2009). In this case, the genome unveiled the importance of smell but it also presented still-unanswered functional and evolutionary questions.

Fig. 2. Sequence-level data for the zebra finch ZENK (zif286, egr1, ngfi-a, krox24) gene, an IEG induced by hearing and producing song that may be regulated by steroids and that encodes a transcriptional regulatory protein (reviewed in Clayton (2013)), as displayed in the UCSC and Ensembl genome browsers. Graphical representation shows the position of ZENK on the minus strand of chromosome 13, with 5' end on the right side of the screen. Note an assembly gap near the 5' end of the predicted gene model (red circle). Also shown in these views are alignments of zebra finch cDNAs (red arrows) and cDNA sequences from other species (red asterisks).
3.2. From functions to parts: reproductive signaling

Gonadotropin releasing hormone (GnRH) is an essential peptide for reproductive behavior, but the ability to study it in model songbird species (e.g., starling and zebra finch) was hampered for at least 15 years by the inability to clone the gene; these failures had led to speculation that the peptide had diverged significantly in songbirds (reviewed in Stevenson et al. (2013)). GnRH was initially unannotated in the assembly, but access to the zebra finch genomic sequence was instrumental in isolating a clone that encodes the GnRH signaling decapptide (GnRH-I) and the associated transport protein (GAP). The clone revealed a highly conserved GnRH peptide sequence with an uncommonly high degree of nucleotide sequence divergence (Stevenson et al., 2009a). Examination of the entire prohormone cDNA suggested several potential sources of variation in GnRH signaling pathways (Stevenson et al., 2009b). These include polymorphisms within the zebra finch, and differences between songbirds (zebra finch and starlings) and non-songbird species that may affect GnRH signaling.

Reproductive signaling involves more than just GnRH-I (Bentley et al., 2003; Deviche et al., 2006; Small et al., 2008; Dawson et al., 2001; Pereyra et al., 2005; Tobari et al., 2010) One important peptide that interacts with GnRH-I is the gonadotropin inhibiting hormone (GnIH) peptide Tobari et al., 2010; Ubuka et al., 2009. GnIH is annotated in the current zebra finch genome assembly. On the other hand, the multiple forms of GnRH, each with a distinct gene, that have been described in other species have not yet been identified in the zebra finch genome (Okubo and Nagahama, 2008). There is immunohistochemical and behavioral evidence that songbirds have GnRH-II but the sequence and structural similarities between GnRH-I and -II require cautious interpretation until genes are defined (Stevenson et al., 2013, 2008; Maney et al., 1997; Bentley et al., 2004; Ubuka et al., 2008). The chicken GnRH-II gene is mapped to chromosome 4; the syntenic region in the zebra finch assembly has the surrounding genes annotated but there is no homologous sequence. It is possible that the gene sequence simply was not included in the assembly – there is an assembly gap in this region. However, homology searches of the zebra finch trace genome sequence archive using the chicken GnRH-II as seed sequence have yet to find sequence of significant similarity (Ikemoto and Park, 2006). It is difficult to demonstrate the absence of a gene, but not all GnRH peptides have been found in all species, thus it is possible that this is the case in the zebra finch, too.

3.3. When genes go missing

In some cases, there is strong evidence that birds lost genes encoded in the genomes of other organisms. Loss of a protein-coding gene can have practical implications for methodologies like Western blots that rely on epitope–antibody binding. But there are also theoretical implications to consider: is the function itself lost, or have other mechanisms evolved to perform the task in birds? Sometimes, gene loss does not have obvious detrimental consequences. For example, bird genomes do not have genes for casein milk or tooth enamel proteins (Warren et al., 2010). Surprisingly, however, we do find examples of “missing” genes that function centrally in the nervous and endocrine systems in other animals. We present here a few examples that do appear to be true losses and that pose enticing questions about the biological implications.

The Kisspeptin (KISS1) peptide modulates GnRH in mammals and fish, and is therefore central to reproduction in those animals. Sequence for the KISS1 peptide is not annotated and cannot be identified via homology searches in the current zebra finch or chicken assemblies. Also absent is a gene for the KISS1 receptor. Efforts to directly isolate the KISS1 peptide in birds have so far also been unsuccessful (Tobari et al., 2010, 2011). It is surprising that such a peptide would not exist in birds given its evolutionary conservation and role in reproduction; perhaps GnIH, either alone or in combination with other hormones, is sufficient to mediate equivalent regulation in birds.

Synapsins are localized to presynaptic terminals and regulate vesicle trafficking (Humeau et al., 2011; Shupliakov et al., 2011; Bogen et al., 2011; Porton et al., 2011). In mammals, there are three synapsin genes, 1–3 (SYN1–3). However, in zebra finch and chicken, only two genes, isoforms 2 and 3, are represented in the genome (Warren et al., 2010). In fact, a stretch of approximately 0.68 Mb on chromosome 1, corresponding to ~25 genes including SYN1 on the human X chromosome, appears to be absent in avian genomes. Evolutionarily, this is interesting because SYN1 was identified in animals that evolved before mammals diverged from birds (Warren et al., 2010; Humeau et al., 2011). This is more consistent with a loss of SYN1 in avian lineages than the duplication of genes in mammals. For those interested in neural function and behavior, the loss of SYN1 raises questions about synaptic transmission; it is reasonable to assume that loss of a protein that controls vesicle release would have profound consequences for brain and behavior. Indeed, synapsins are associated with neurological disorders and disease such as epilepsy and schizophrenia (Porton et al., 2011; Saviouk et al., 2007; García et al., 2004; Cavalleri et al., 2007; Lachman et al., 2005; Tsai et al., 2002). However, single, double, and triple SYN gene knock out mice survive into adulthood (Bogen et al., 2011; Gittler et al., 2004). Even triple SYN knock out mice do not show gross morphological disruptions, they have normally-sized brains with the same number of synapses, and behave largely as wild-types, though they are prone to seizures and display deficits in specific behavioral tasks, consistent with disregulation of synaptic function (Gittler et al., 2004). SYNA 2 and 3 are dynamically regulated in song control areas in zebra finches (Velho and Mello, 2008). It remains an open question whether or not these two proteins perform the functions that it takes three synapsins to do in mammals.

Steroid signaling is central in the development and function of multiple tissues, including the brain, where it directs global neuroendocrine processes and behavior. In the zebra finch, estradiol is the only factor known to functionally masculinize developmental organization of the song control system, and the brain has the capacity to synthesize steroids de novo as early as the day of hatching (Simpson and Vicario, 1991a,b; London et al., 2006, 2004, 2003) Steroids continue to signal in the song system in adulthood; estrogens are rapidly synthesized in response to experience and have potential for rapid and long-term effects on song processing (Re mage-Healey et al., 2008, 2011). Several key steroidogenic enzymes are members of gene families that demonstrate how even small modifications in sequence can alter characteristics such as substrate specificity to increase diversity of steroid signaling. Major enzyme types from two gene families (Abe et al., 2012, 2014) and 17α-hydroxysteroid dehydrogenase, are confirmed in zebra finch brain biochemically (Cam and Schlenger, 1998; London et al., 2010; Soma et al., 2004; Tam and Schlenger, 2007; Vanson et al., 1996). Examination of the genome, however, suggests that several members of both families that are present and active in mammals are absent in birds (London and Clayton, 2010). Within the subset of genes identified in the zebra finch genome, those experimentally confirmed in the zebra finch brain are among those highly conserved across clade. This suggests that the function of these steroidogenic enzymes is preserved; how the birds compensate for the loss of the other enzyme family members is yet unclear.

3.4. When genes proliferate

Genome parts lists also reveal cases where single genes are duplicated. Again, caution is required initially as automated
assembly of complex genomes can create spurious gene predictions (see below) but experimentation can confirm these predictions. Gene duplication raises the question of whether the second gene contributes new biological functions, or if the additional copy permits independent regulation (e.g., across different tissues or subregions of the same tissue) of the same function to control distinct process.

For example, consider growth hormone (GH), a highly conserved peptide in vertebrate evolution best understood as a hormone released from the pituitary to affect multiple somatic tissues (Etherton and Bauman, 1998). It is also produced in extra-pituitary brain regions, where its function is less clear (Harvey and Hull, 2003). The growth hormone gene has been duplicated in birds of the phylogenetic Order Passeriformes, which includes songbirds (genes mapped to Chromosomes 1 and 27) (Yuri et al., 2008). Both GH genes are transcribed in zebra finch brain (Replogle et al., 2008) though only one of the peptides has been directly measured (Xie et al., 2010). Sequence substitution measures show that both passerine GH genes have hallmarks of positive selection compared to nonpasserine genes and that the duplicate peptides are likely functionally divergent (Yuri et al., 2008). Data from a corvid passerine suggests that the two GH genes have different expression profiles and therefore could have independent functions (Arai and Igo, 2010). One intriguing speculation is that duplication of the GH gene could have led to the independent growth regulation of the passerine vocal control nuclei; these areas evolved only in songbirds and are distinctive in their growth rates relative both to each other and to the rest of the forebrain (Bottjer et al., 1985). An intriguing experiment would be to compare expression patterns of both GH genes in Passerine oscines (songbirds, with fully developed singing neural network) and Passerine suboscines (closely related non-songbirds, with rudimentary vocal areas) to assess this hypothesis.

Finding gene duplications and considering their functional ramifications can be exciting but caution is warranted as genome-wide gene prediction can “discover” new genes that do not exist. For example, the official Ensembl gene set (genebuild 71) contains three gene models that appear to be duplications of the key estrogen-synthetic enzyme aromatase. Even cursory examination of these models, however, demonstrates that only one (ENSTGUG00000006993) contains all 9 coding exons well-confirmed in the aromatase transcript whereas the others are clearly fragmentary (Fig. 3). It is possible that zebra finches have evolved some specialized truncated enzyme, but given the extensive characterization of aromatase gene and protein in multiple bird species, it is unlikely (Metzdorf et al., 1999; Saldanha et al., 1998, 2000; Schlinger and Arnold, 1993; Shen et al., 1994; Cornil et al., 2011; Dickens et al., 2013; Ramachandran et al., 1999). The single nucleotide polymorphism rate in zebra finches is high, predicted to be one per every ~700 bp, thus it is likely that the donor bird for the genome sequence was heterozygous at the aromatase locus, and the two alleles were different enough to be treated as independent genes during the genome assembly process.

Other cases of predicted gene duplications are less clearly diagnosed as totally spurious. When the zebra finch genome was first assembled, 31 copies of the p21-activated kinase (PAK3) gene were bioinformatically predicted (Warren et al., 2010; Kong et al., 2010). Fourteen genes were mapped to 5 different chromosomes, including the Z sex chromosome, and the other 17 were mapped to an unincorporated supercontig, a large stretch of genomic sequence that was not incorporated into the final genome (Kong et al., 2010; Itoh et al., 2011). It is not surprising that birds would have multiple PAK3 genes; 6 PAK3 genes have been described in mammals (Boda et al., 2006; Jaffer and Chernoff, 2002; Kreis and Banner, 2009; Ramakers, 2002). At least a subset of the predicted genes are transcribed. In zebra finches, PAK3 is expressed widely within the brain and in the testes (Kong et al., 2010; Itoh et al., 2011) but not all predicted PAK3 genes may be functional. Although multiple functions have been reported for the different mammalian PAK3 genes, many of the avian gene models are incomplete and the assignment of some models to non-chromosomal contigs is highly suggestive of allelic sequence divergence that inflates gene model predictions (Kong et al., 2010). Evidence is that multiple PAK 3 genes do exist; the challenge is to experimentally confirm which models accurately represent genes and to ascribe functions to them.

4. Gene networks, landscapes and terrains

As we have just seen, a sequenced genome can be invaluable in studies of individual gene products or gene families. But it also enables the use of new methods for measuring the expression of essentially all gene simultaneously in a tissue sample (i.e., “genome-wide expression analysis” using microarrays or RNAseq). The shift from single-gene to genome-wide analysis is both a technical and a philosophical one.

4.1. The genome as an agent of physiological signaling

The genome itself is a central agent in physiological signaling – an idea rooted in the earliest investigations of molecular biology and applied specifically to steroid signaling since the early 1960s (Hechter and Halkerston, 1965). During the 1970s and 1980s, the concept of signal transduction was widely extended to include pathways by which diverse extracellular signaling molecules lead to genome responses in the cell nucleus. The concept of “Immediate Early Gene” (IEG) was borrowed from virology to describe the genes that are expressed first, such as c-fos and egr-1, with the assumption that these first responses would lead to an ongoing cascade of “late” gene responses (Cole et al., 1989; Morgan and Curran, 1989; Sheng and Greenberg, 1990; Clayton, 2000). IEGs became increasingly used as indicators of neuronal (neurogenomic) response to explicit behavioral or environmental stimuli; in birds such paradigms include the sound of birdsong (Mello et al., 1992), sexual activity (Meddle et al., 1997), changes in photoperiod (Ginty et al., 1993) or exposure to a novel challenge requiring learning (Javars et al., 1995; Beck and Fibiger, 1995; Guszok et al., 2001; Jones et al., 2001).

In the last 10 years, however, researchers have moved beyond the simple “signal-response” paradigm of IEGs, and recognized that behavioral change can involve large-scale shifts in entire landscapes of gene expression as revealed using DNA microarrays and more recently, direct quantitative sequencing of RNA populations (i.e., RNAseq) (Mortazavi et al., 2008; Hitzemann et al., 2013). Two notable examples from the zebra finch literature were extensions of earlier research that began as studies the IEG response to the production and perception of birdsong. Wada et al. (2006) estimated that more than 100 genes are regulated within the first few hours after a bird begins to sing, in at least four different anatomical loci and six temporal expression patterns. Dong et al. (2009) observed hundreds of gene expression changes within the first 30 min after an isolated bird hears a song playback – but found a different and even larger set of genes changed their expression as the behavioral response to the song became habituated. These results were focal points for further analysis in the zebra finch genome sequencing project (Warren et al., 2010), which confirmed that a large number of both protein coding mRNAs and ncRNAs respond to song production or perception, and began to apply transcriptional network modeling to analyze the temporal dynamics of gene responses (Fig. 4).
The involvement of ncRNAs adds a new wrinkle to the study of behavioral and physiological signaling. The sound of song playback, for example, results in a rapid (within 30 min) decrease of hundreds of transcripts in the auditory forebrain, a third of which do not encode proteins (Warren et al., 2010; Dong et al., 2009). Song-responsive ncRNAs include both long non-coding RNAs (lncRNAs), as well as microRNAs and other very short RNAs species (e.g., piRNAs) Gunaratne et al., 2011. Many ncRNAs are believed to function by regulating the transcription, stability or translation of many target mRNAs at once (Wang and Chang, 2011; He and Han-non, 2004). This represents an additional level of coordinate gene regulation that can be superimposed on top of primary control systems mediated by transcription factor proteins and could contribute to functional modulation in response to physiological signals. For example, one song-regulated microRNA, miR-2954, is expressed at higher levels in all tissues in males than in females (Luo et al., 2012) and declines even further in the brain of females – but not males – when they hear song playback (Gunaratne et al., 2011) (Fig. 4). miR-2954 is therefore the first genomic element discovered that responds to song playbacks differently in males and females. Females show different behavioral responses to male courtship songs with distinct acoustic properties (Balzer and Williams, 1998; Vyas et al., 2009; Woolley and Doupe, 2008). It is intriguing to speculate that this sex difference in microRNA regulation could contribute to female-typical behavioral responses to song.

4.2. Emergent properties revealed

By analyzing the expression of many genes at once instead of just one or a few, new principles of biological function may be revealed. In each of the following three examples, study of large gene sets was essential to the discoveries made.
The first example concerns the study of fundamental sex differences. In most species, various mechanisms of dosage compensation result in equivalent expression of Z- (or X-) linked genes in both sexes, despite the two-fold difference in chromosomal dosage between them. Dosage compensation was assumed to be essential in all species, mainly to avoid developmental catastrophes. However, genome-enabled microarray analysis of all Z-linked vs. autosomal genes in chickens and zebra finches revealed that male birds generally express Z genes at a higher level than females do, in direct contradiction to the previous dogma (Warren et al., 2010; Itoh et al., 2007, 2010; Melamed and Arnold, 2007). The ramifications of this are still being worked out (Arnold and Itoh, 2011; Itoh et al., 2011; Mank, 2013).

Another example concerns the concept of “habituation”, often thought of as a general loss of responsiveness to a particular stimulus (Dong and Clayton, 2009). Early studies of the genomic response to song playbacks demonstrated habituation of the ZENK response to a song after that song had been presented repeatedly to a bird – reexposure to the habituated song resulted in no change in ZENK expression compared to birds hearing only silence. Naively, if focusing on just the ZENK gene, one might expect that the brain gene expression profile of birds in the habituated state would in general resemble that of birds hearing only silence. But microarray analysis revealed an unexpected picture, with habituated birds showing a distinct pattern of gene expression from birds left in silence in the auditory forebrain, suggesting habituation is an actively acquired state much more complicated than a simple loss of the ZENK response (Dong et al., 2009).

A third example comes from the recent study of Hilliard et al. (2012), who examined how differences in gene expression might be correlated with various behavioral measures linked to song production. They combined microarray analysis with a statistical procedure (Weighted Gene Coexpression Network Analysis, WGCNA) that clusters genes when they show similar variations in expression across a large set of samples – in this case, 26 different birds. Using a microarray to probe essentially all genes expressed in the brain, they identified 21 modules (coexpressed gene sets) in song control nucleus Area X, five of which are correlated with various measures of singing behavior in their study population (Fig. 5). Importantly, the modules differ in how they map to the different behavioral measures, suggesting that a modular organization of gene expression could underlay a modular organization of behavior.

4.3. Genomic integration of developmental and environmental influences

Ordinary experience is integrative and ongoing, and many behaviors are shaped gradually over time. Following Waddington (1957), the image of a changing landscape may provide the best metaphor for this, and genome technology now provides tools to help define landscapes in molecular terms. Three examples come from the Songbird Neurogenomics Initiative, a recently completed collaborative project to analyze samples collected from multiple laboratories and behavioral paradigms with a common microarray analysis pipeline (Reploge et al., 2008; Drnevich et al., 2012). Mukai et al. (2009) compared hypothalamic gene expression patterns in the breeding (spring) and nonbreeding (autumn) seasons of territorial song sparrows as they experienced a simulated territorial intrusion. Each of the two factors (shift in seasons and behavioral challenge) alone triggered different changes in gene expression. Moreover, 88 cDNAs showed significant interactions between season and the simulated intrusion indicating that the environmental context modulated the genomic response to a particular experience (Fig. 6). Stevenson et al. (2012) measured gene expression in the telencephalic song control nuclei of starlings as they underwent changes in photoperiod. They

![Fig. 5. Diverse relationships between gene expression network modules and specific behavioral measures. Using WGCNA, 21 modules were constructed from gene expression data. Reprinted from Mukai et al. (2009).](image)

![Fig. 6. Gene regulatory network displays functional connections between multiple genes regulated by season. This network includes multiple peptides such as Vasopressin Inhibitory Peptide (VIP), Somatostatin (SST), Follicular Stimulating Hormone (FSH), Adenylate Cyclase Activating Polypeptide 1 (ADCYAP1), secretogranin (SCG2) and transthyretin (TTR1). Red gene symbols indicate higher expression in autumn compared to spring; blue is opposite. Darker color intensity indicates large fold change of that gene between seasons. Open symbols are genes without data. Reprinted from Mukai et al. (2009).](image)
observed a “categorical switch” in gene expression that was both dependent on reproductive state and different from the waves of gene expression observed in the preoptic area of the hypothalamus. Thompson et al. (2012) studied gene expression in different hormonal and photoperiodic conditions in seasonally-breeding Gambel’s white-crowned sparrows, and in two different telencephalic song control nuclei (HVC and RA). They observed evidence for different molecular programs in HVC and RA in response to seasonal change. All three studies observed changes in gene expression for a broad range of functional categories, including genes that regulate thyroid hormone action and neuroplasticity (Mukai et al., 2009; Stevenson et al., 2012); neurogenesis and angiogenesis (Stevenson et al., 2012; Thompson et al., 2012), electrophysiology (Thompson et al., 2012) and epigenetic processes (Stevenson et al., 2012).

These microarray studies highlight how gene expression profiles can be molded and reshaped over time, storing information about past history and environmental context and modulating the response to immediate behavioral or physiological signals. This raises two questions about the links between the neurogenome, physiological signaling and behavior. How does the same underlying genome give rise to different responses to a specific cue when it is encountered in different environments, developmental conditions or cell types? And how can we trace from such complex shifts in gene expression to coherent functional consequences? These questions are largely unanswered although progress is being made at both technical and theoretical levels.

The first question can be appreciated as the central question in the modern study of epigenetics (where the word is used to mean stable differences in gene expression despite the same underlying gene sequence) (Jaenisch and Bird, 2003; Weaver et al., 2004). The concept has its roots in developmental biology (Waddington, 1942) and has become linked to the concept of “phenotypic plasticity” (more than one phenotype from the same genotype) in ecology and evolution (Renn and Schumer, 2013; Kültz et al., 2013). In other study organisms, great advances have been made in defining biochemical mechanisms that give rise to persistent changes in gene expression (e.g., DNA methylation and chemical modification of histone proteins). Studies of epigenetic mechanisms in avian biology are very limited at this writing, however. A few reports have considered the possible role of DNA methylation in long-term evolutionary processes (Liebl et al., 2013; Hu et al., 2013) or transgenerational inheritance (Natt et al., 2012) but there is only a single recent report of an epigenetic measurement related to neuro-behavioral signaling: a difference in DNA methylation around the androgen receptor gene correlated with measures of singing behavior in wild and domesticated strains of Lonchura (Wada et al., 2013). This contrasts with the situation in mammals where epigenetic mechanisms have been directly implicated in behavioral processes ranging from partner preference formation (Wang et al., 2013) to learning and memory (Day and Sweatt, 2011, 2010) and cognitive aging (Oliveira et al., 2012). Clearly this is an area ripe for further study in avian behavioral models.

The second question, how coherent organismal function emerges from complex shifts in gene networks, can be appreciated as the central question of modern systems biology (Medina, 2013). Systems biology is focused on the analysis of network topology and the identification of key nodes in the network that represent critical control points for the overall network function, e.g., (Zeidán-Chulúa et al., 2013; Ausländer and Fussenegger, 2013; Wang et al., 2012). The application of WGCNA described above (Hilliard et al., 2012) represents an early illustration of the potential for developing a systems biology approach to avian behavioral models.

5. Leveraging natural variation

One of the most powerful aspects of the new genomics is the access and precision it gives to measurements of genetic variation. When genetic variation can be correlated with functional (phenotypic) variation, this can lead directly to inferences about the mechanisms underlying the function. Genome technologies provide the tools for analyzing genetic variation efficiently in large populations of organisms both at the whole-genome level (e.g., using SNPs and microsatellite markers) and at the precise nucleotide sequence level in targeted regions of specific genes of interest. These tools are being applied now to study genotype–phenotype correlations both within and between avian species.

5.1. Nature’s experiments

In some cases, serendipitous natural structural variation within a single species may be exploited. For example, it has been known for 50 years that an inversion within chromosome 2 distinguishes two morphs of the white-throated sparrow, Zonotrichia albicollis (Thornycroft, 1966). The morphs differ in plumage and behavior, with the “tan” morph (TS) investing more in parental care and the “white” morph (WS) more in territorial and sexual aggression (Tuttle, 2003; Horton et al., 2012, 2013). The polymorphism is maintained in the population by balancing selection through negative assortative mating (each morph prefers to breed with the opposite morph) and through recombination suppression. A rare homozygous individual for the “white” inversion showed an extreme phenotype supporting the hypothesis that alleles within the inverted segment confer high aggression (Horton et al., 2013). To define the structure of the inversion at high resolution, BAC clones and other materials derived from the zebra finch genome sequencing project were first used (Thomas et al., 2008), followed by targeted genomic sequencing around the inversion breakpoints (Davis et al., 2011; Huynh et al., 2011) (Fig. 7). Clearly this chromosomal polymorphism must have a strong and direct effect on social behavior, but the inversion is large enough (~100 Mb) to contain hundreds of genes. One next step is to identify genes in or near the inversion that show alternate expression patterns that sort with the behavioral phenotype. This experiment, technically unfathomable even 10 years ago, is now straightforward using high-throughput sequencing-based approaches.

Chromosomal variation has also been observed within and between colonies of the zebra finch. These include inversions in chromosome 6 in domesticated birds (Itoh and Arnold, 2005) and an inversion within the Z sex chromosome present in domesticated and wild zebra finches of both the Australian (T. guttata castanotis) and Timor (T. guttata guttata) subspecies (Itoh et al., 2011, 2006). The Z inversion breakpoint appears to involve one of the numerous PAK3-like genes which expanded in number in the zebra finch after divergence from the chicken (Kong et al., 2010; Itoh et al., 2011). So far, no behavioral or functional phenotypes have been connected to these variations, though generally breakpoints are located in highly repetitive regions associated with transposable elements.

**Fig. 7**. Fluorescent In Situ Hybridization (FISH) demonstrates chromosome 2 inversion that sorts with color and behavioral phenotypes in white-throated sparrows. Metaphase spreads of chromosome 2, hybridized with three genes, here labeled by BAC clone number, show inverted order and arm displacement of three genes in the TS (left) compared to the WS (right), and illustrate shift in centromere position. Reprinted from Thomas et al. (2008).
elements that can lead to gene expansions; as discussed above, it is possible for these expansions to have functional consequences but each needs to be validated.

5.2. Breeding for behavioral variation

As the zebra finch breeds readily in captivity, classical genetic strategies to parse out associations between specific genes and behavioral traits may be feasible. These strategies include Quantitative Trait Locus (QTL) mapping (though see Slate (2013) for a sobering assessment) and selective breeding for lines that differ in physiological or behavioral properties. The growing arsenal of sequence-based genetic markers in the zebra finch should greatly enhance these strategies (Stapley et al., 2010; Dawson et al., 2013; Forstmeier et al., 2012; Backstrom et al., 2010). One example highlights both the promise and the challenge of the genetic approach in zebra finches. Artificial selection for responses to a mild stressor, following a paradigm used extensively in past studies of galliformes, was successful over four generations in raising the peak corticosterone response in two zebra finch lines (Evans et al., 2006). Unfortunately, resources were not in place to maintain these zebra finch lines and they are no longer available for study.

5.3. Comparative genomics

A strength of avian biology is the great diversity of accessible species, each with a unique set of evolutionary and ecological adaptations. Comparative approaches can help identify underlying functional mechanisms responsible for specific traits that are segregated among otherwise closely related species, e.g., (Devoogd, 2004; Brenowitz and Beecher, 2005; Beecher and Brenowitz, 2005; Denver et al., 2009). In principle, genome sequencing is a powerful new tool for comparative studies. For example, vocal learning ability is shared among the thousands of different oscine songbirds, but for the most part is absent among the hundreds of subspecies which are also members of the Order Passeriformes. Among the oscines, song learning occurs in some species only during a limited juvenile period (e.g., zebra finch) Immelmann and Hinde, 1969; Slater et al., 1991; Jones et al., 1996, whereas other species can continue to copy and incorporate new song elements throughout their lives (e.g., starling) Bohner et al., 1990; Chaikin, 1994. Other behavioral traits that clearly differ systematically across songbird species include social group size and territoriality, mating styles (ranging from monogamy to polygamy and polyandry) and parental care (e.g., biparental vs. maternal-only) (Cockburn, 2006; Goodwin et al., 1982; Goodson et al., 2005). Each of these stable species differences must have a root in the genome.

Might it be possible to identify genetic elements responsible for characteristic behaviors, simply by directly comparing genome sequences of different species that do and do not display that behavior? The challenge to be faced is the background of enormous sequence variation unrelated to the trait of interest, even among individuals of the same species. For example, the paired diploid chromosomes in a single zebra finch differ at more than 1,740,000 sites (SNPs) – one every ∼700 bp (Warren et al., 2010). In practice this means that comparisons using even thousands of individuals may still be statistically under-powered, a disappointing conclusion of many genome-wide association studies (GWAS) in humans Gandhi and Wood, 2010; Rakyan et al., 2011; Hakonarson and Grant, 2011. To interpret all of the genomic changes, it is necessary to tap into the strengths of the bird system: comparative phylogeny and existing paradigms for the study of endocrinology and complex natural behavior.

For example, taxonomy and comparative sequencing can be used to pinpoint chromosomal regions responsible for a trait, by studying the pattern of trait loss across a broad phylogenetic tree. A recent proof-of-concept for this approach focused on the ability to synthesize vitamin C, an ancestral vertebrate trait that was lost in at least four independent mammalian lineages. A large phylogenetic tree aligning 27 sequenced mammalian genomes identified only a single gene that was lost in all of these and only these four lineages; this gene is indeed central to vitamin C synthesis (Hiller et al., 2012). It seems doubtful that such a straightforward strategy will be feasible for more complex multigenic traits, ones that emerged more recently in evolution, or ones for which large phylogenetic trees of gain and loss are not available.

For neuroendocrinology and behavior, comparative genomics may be most effective when it can be combined with deep knowledge of a particular phenotype and its physiology, especially when the phenotype has diverged relatively recently in evolution. The classic example of a genetic difference underlying a species difference in behavior is the discovery of a polymorphism in the V1a receptor gene (avpr1a), which encodes a receptor for the nonapeptides vasopressin and oxytocin. Variations in a simple repeat “microsatellite” in the avpr1a promoter gene contribute to polygamous/monogamous pair bonding behaviors in two species of vole (reviewed in Donaldson and Young (2008)). This discovery came only after a century of fundamental biology had established oxytocin and vasopressin as key hormones in this behavior and defined species-dimorphic expression of the receptor (Insel and Shapiro, 1992).

A key question, unanswered at this point, is how many (if any) aspects of behavior can ultimately be reduced to such single-gene/single-protein effector traits that are also broadly conserved in evolution. Even the case of the V1a receptor polymorphism is more complex than initial reports implied, as a phylogenetic analysis of 25 rodent species failed to confirm the association of this particular polymorphism with social monogamy more broadly (Fink, 2006). There may be many different functional pathways to achieve a particular behavioral output. Consider the case of critical period learning in songbirds. A multitude of different factors could account for a restriction of song learning to a discrete period in different species. These may include social factors that drive attention to a specific tutor at a particular age; neurodevelopmental factors in any of the various circuit components needed for song copying (i.e., song perception, motor control, integration of perceptual and motor systems); factors that govern the amount of singing or the distribution of song practice across the day; even factors that engage the system during sleep, which has been increasingly appreciated as a significant factor in the biology of song (Shank and Margoliash, 2009). While it seems certain that genetic factors underlie species differences in the life trajectory of song learning behavior, it also seems likely that different specific mechanisms (and genes) may have evolved in different species to result in what is overtly a similar behavioral phenotype. This reality reflects the utility of pushing forward with genome-scale tools; it will be essential to measure many genomic elements and understand the relationships between them to parse the connections between genome dynamics and systems-level traits.

6. Moving forward

The field of behavioral neuroendocrinology has firmly established the fundamental importance of hormones in brain function and behavior. One of its strengths has been the integration of multiple levels of investigation, from small molecules, across tissues, into the whole animal, and in complex environments. Genomes represent a great new opportunity to incorporate another level of
analysis and thereby deepen the understanding of complex biological systems. In addition to ongoing improvements in sequencing technologies and statistical tools, anticipated advances in two other directions will help to further integrate genome science into the study of avian behavior and neuroendocrinology.

6.1. Data integration

The breadth and accuracy of genome data is only going to increase: its digital nature facilitates the accumulation, refinement and analysis of data. This is one of the strengths of genomics, but it also necessitates a strategy for dealing with expanding data set sizes. As more genomic and RNA-based sequences are generated, it will be essential to collect and share these data in a way that supports efficient access and analysis by the appropriate research communities – in this case, including both neuroscience and avian biology communities. At the present, more and more laboratories are collecting genome-based data, but there is as yet no efficient mechanism for sharing and comparing data within the avian community. Efficient data integration could lead to improved gene annotations, genome assemblies, comprehensive gene expression maps, and so forth. Nor is there good integration of genome resources with phenotypic data, as seen for example in the research communities studying plants (www.iplantcollaborative.org), humans (dbGAP and OMIM at www.ncbi.nlm.nih.gov/), mice (www.informatics.jax.org/) and zebra fish (http://zfin.org/) (see Kültz et al., 2013 for further discussion). In the avian community, important first steps have been taken in this direction with the development of scattered web-based resources for various measures of brain gene expression in various avian species (e.g., www.zebrafinchlatlas.org, geisha.arizona.edu, songbirdtranscriptome.net, avianigenomes.org, songbirdgenome.org). A fledgling resource that begins to link across levels and methods of analysis is seen in http://songbirdscience.com/, but much more needs to be done. The Songbird Neurogenomics Initiative (Replège et al., 2008) is a recent attempt to link experimenters working on diverse songbird species and research questions, and permitted a meta-analysis that revealed broader principles of neural gene expression (Drnevich et al., 2012). The opportunity for such collaborations will only increase, and bird researchers studying behavioral neuroendocrinology would benefit from their creation.

6.2. Efficient gene manipulation in avian brain

The impact of genomics on avian neuroendocrinology (and avian biology in general) will be much greater once key improvements are made to methods for in vivo gene manipulation. Several approaches have already been demonstrated in birds, including production of transgenics (Sato and Lansford, 2013; Seidl et al., 2013; Agate et al., 2009; Mozdziak and Petitte, 2004; Poynter et al., 2009; Poynter and Lansford, 2008; Scott and Lois, 2005, 2006), use of viruses as vectors for delivering and expressing specific sequences injected directly into the brain (Wada et al., 2006; Haesler et al., 2007; Roberts et al., 2008; Bauer et al., 2008; Schulz et al., 2010), and optogenetic manipulation of targeted neural circuits (Roberts et al., 2012). These strategies have enabled visualization of individual cells across development and within brain circuits, identified new anatomical connections, and tested functional contributions of genes and brain regions for behavior, demonstrating their great potential for discovery. Methodological improvements are still needed, however, to make these as routine techniques in the toolkit for avian behavior and neuroendocrinology.

7. Conclusion

Study of avian genomes has revealed unexpected behavioral potentials, variations in the machinery of neuroendocrine signaling, and a probable role for epigenetic processes in the integration of signals and behavior across changing environments. We have shown how even the “finished genome” is inevitably a work in progress, in part for the challenges of assembling a complete and accurate genome, and in part for the reality that the genome is different in every individual (with millions of potential single-nucleotide variations plus the occasional large restructuring of chromosomal pieces). Both genomics and neuroendocrinology will advance by better integration of the tools and datasets emerging from these different approaches and traditions.

Acknowledgments

We are grateful to the anonymous reviewers for useful organizational and textual suggestions.

References


Shupliakov, O., Haucke, V., Pechstein, A., 2011. How synapsin I may cluster synaptic
Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H.,
Schulz, S.B., Haesler, S., Scharff, C., Rochefort, C., 2010. Knockdown of FoxP2 alters


