What makes birds and bats the talk of the town

Comparative methods are used to study vocalization in songbirds and bats.

Vivien Marx

Songbirds and bats learn to vocalize from parents and peers, much as people do. For labs studying language and speech, especially vocal learning, "birdsong is a great model system," says Timothy Gentner, a computational neuroscientist at the University of California at San Diego (UCSD).

Stella and Disco are feathered, talking YouTube stars. "What do you think?" asks Stella, a starling, speaking directly into the camera. "I'm Disco and I know it," says Disco, a parakeet, in a separate video. Some songbirds have a knack for mimicry. Lyrebirds can imitate the songs of many bird species, the sound of car alarms, camera shutters and chainsaws. Disco sings his species-typical song as well as TV theme songs, and he can beatbox. "They'll pick up anything," says Gentner about songbirds.

Songbirds are social vocalizers. Gentner has sometimes whistled to the birds as he entered the colony to record birdsong from the several species there. One day, some whistled back. The same has happened to Rockefeller University neurogeneticist Erich Jarvis, who works with zebra finches and canaries, among others. "What they're thinking, I don't know," says Jarvis. Perhaps they perceive his sounds as similar to ones they use.

Barring a speech or other impediment, babbling human babies eventually turn into fluent adults. Songbirds, too. "Starlings will add material to their vocal repertoires for years and years and years," says Gentner. Starlings and canaries are open-ended learners, whereas zebra finches stop learning songs after a certain age. Studying that phase and vocal learning more generally can reveal details about development, changes to neural circuits during learning, sound production, perception or processing[1].

Among vertebrates, vocal learners are rare, says Jarvis. Among mammals the trait appears limited to people, whales, dolphins, elephants, seals and bats, and among birds, songbirds, hummingbirds and parrots. According to comparative neuroanatomy, songbirds and humans share analogous brain structures involved in vocal learning. Structures in songbirds are comparable to the cortical areas in mammalian brains involved in learning, sensory input processing and motor output, all of which play a role in speech.

Zebra finches have long been studied in the vocal communication field. As part of the Vertebrate Sequencing Project, the zebra finch genome is being sequenced again. (Credit: Sergio Mendoza Hochmann/Moment/Getty)

Sonja Vernes, at the Max Planck Institute for Psycholinguistics in the Netherlands, studies the genetics of vocalization in mammals, mainly people and bats. She and others have advanced bats as a model of vocal learning, and she cofounded Bat 1k, a project geared toward sequencing the genomes of all bat species. She and Ella Lattenkamp of Ludwig Maximilian University in Germany note that comparative approaches in a broader range of species are needed in vocal learning research, and also in understudied or neglected species[2]. The cross-species view will let labs learn more about the phylogenetic, ecological, morphological, developmental and social aspects and neural and genetic underpinnings of vocal learning.

Large-scale evolutionary and comparative genomics projects are under way toward this goal, while also advancing broader research questions[3]. The projects leverage cost drops in genome sequencing, assembly and RNA sequencing, and the greater availability of do-it-yourself options in optogenetics. The Vertebrate Genomes Project has set out to sequence and assemble reference genomes for all vertebrates and is about to release its first dataset of over a dozen complete genomes, including several vocal-learner species.

Song and speech

Vocal communication in people and birds involves much physiological machinery: the larynx, nerves, auditory processing, learning, memory. Human speech is made up of distinct sounds—phonemes—as is bird song. Both birds and people have evolved a sensitivity to the structure of vocal communication, says Gentner.

In studying the complex, varied song of the Bengalese finch, researchers at Kyoto University discovered that the birds can perceive 'grammar', a cognitive ability long thought to be uniquely human[4]. The scientists used strings of finch song syllables to create artificial Bengalese finch songs. Behavioral changes indicated the birds could discriminate between the two: in reaction to 'grammatical' song they gave short response calls, whereas artificial song led to increased calls. Neuroanatomic analysis revealed that this auditory processing in the finch occurs in brain areas such as the pallium, which shares properties
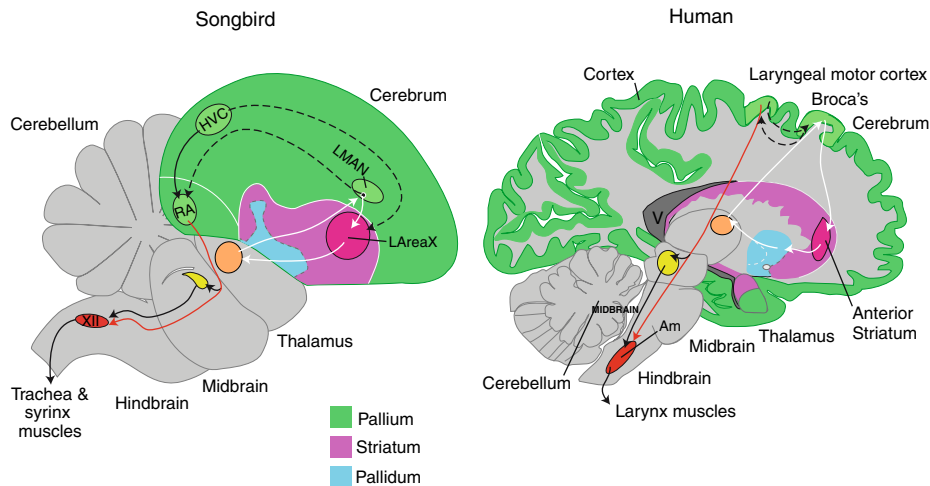
Areas of the songbird brain share properties with language-processing areas in the human brain. (Adapted with permission from ref. 6, AAAS; E. Dewalt, Springer Nature)

with the language-processing cortical region in the human brain.

Birds communicate with one another through song—"there's no doubt about that and they're using really acoustically rich and varied signals to do it," says Gentner. "But there are a lot of properties that language has that birdsong doesn't." We won't be speaking with birds meaningfully anytime soon, he says. Language and meaning sit atop a structured sequence of phonemes. "Language seems to have co-opted this structure in ways that songbirds don't seem to have co-opted for the modulation of meaning," he says. Given evolutionary conservation across the animal kingdom, the comparison of birds and humans can reveal details such as the similarity between neural circuits that control vocal communication, he says. Vocalization research needs the resolution to see, for example, how a subset of neurons generates vocal behavior or how plasticity lets us adapt to understand a speaker with a strong accent.

To parse birdsong, the Gentner lab collaborates with physicists at UCSD and the Salk Institute. They develop and apply computational tools to analyze sonograms, the plots of song frequencies and duration. They are working on high-throughput machine-learning-based methods for analyzing song from different bird species and comparing its sequential structure to that of human speech. It will be a more quantitative way to interpret birdsong, says Gentner.

Evolutionary radiation
Several bird traits, including vocalization, have resulted from convergent evolution, says Guojie Zhang, an evolutionary geneticist who splits his time between the University of Copenhagen and the China National GeneBank run by BGI Shenzen[5,6]. As Jarvis explains, a common ancestor may have hybridized with non-vocal learners who gave rise to parrots and songbirds, but he is "90% confident" that vocal learning in songbirds, hummingbirds and parrots evolved independently. Humans are 300 million years removed from our common ancestor with birds, while chimpanzees are at a 6-million-year distance. Yet with vocal learning, birds "are more similar to us than chimps," says Jarvis. The vocal nuclei in songbirds match regions in the human brain that are lacking in non-human primates.

Good reference genomes are needed to build a phylogenetic tree and for comparative genomics analysis, says Guojie Zhang.

Good reference genomes are needed for building a phylogenetic tree that lets labs reliably do comparative genomics analysis, says Zhang. Those genomes can help with the identification of protein-coding genes and regulatory non-coding regions. What makes phylogenetic analysis of birds challenging is rapid radiation: many species appear in a narrow period 60 or 70 million years ago, he says, near the mass extinction at the Cretaceous–Tertiary border.

Labs have been analyzing different subsets of genomic sequences and obtaining different phylogenetic trees, says Zhang. But the whole genome context is needed because genes never function in isolation. This motivated Zhang and others to launch the Avian Phylogenetics Project and then B10k, a project to sequence the genomes of all 10,500 known bird species, organized by institutions including the Kunming Institute of Zoology of the Chinese Academy of Sciences, the University of Copenhagen, Rockefeller University, the Smithsonian Institution, Imperial College and Curtin University in Australia. To date, teams have completed full genomes for 360 bird species.

For phylogenetic analysis, Zhang and colleagues have used software tools such as RAxML, and more tools are in development. Phylogenetic analysis with full genomes is a computational challenge. "This has become a bottleneck right now," says Zhang. Data visualization of many species at one glance is also on the to-do list. "We need to find smart ways to show the data," he says, and move beyond a static phylogenetic tree, which "you cannot really play with." It's a publishing challenge, too—a tree with 300 or more species is hard to display in a traditionally formatted paper.

Comparing genomes across species opens up research questions, says Zhang. "Birds have lost a lot of genes throughout evolutionary history," he says, which has made the bird genome quite small compared to those of other animals, such as reptiles. The bird genome is rather conserved: periods of gene loss were followed by low levels of gene change, but birds have gained many non-coding genes with regulatory function, which helps to explain bird diversity, he says. Once sequences are available, labs can explore changes through evolutionary time, take on validation of gene function and explore gene–behavior associations, and consider ways to apply gene-editing techniques to explore gene function.

Tuning in to bats
Bats echolocate and vocalize, too. To understand circuits in the brain of the Egyptian fruit bat that underlie vocal learning and production, neuroscientist and bioengineer Michael Yartsev and his team at the University of California, Berkeley, take a multidisciplinary approach. They analyze acoustic recordings, electrophysiology and optogenetic data, and carry out neuroanatomic tracing and behavioral tests. Promising anatomic parallels between the brains of bats, birds and people are beginning to emerge, says Yartsev.

Bats are good models for exploring vocal production, learning and development, and to study social aspects of vocalization, says Tobias Schmid, a graduate student in

To record vocalization from flying bats, wireless tools are used. (Credit: Steve Gettle/Minden Pictures/Getty)

the Yartsev lab. To better study and isolate electrophysiological signals in vocalizing Egyptian fruit bats, he does operant conditioning experiments. The bats learn to make calls in a new context. They love bananas and learn to make high-pitched trills and shorter, noisier screeches for fruit reward. The lab studies bats of various ages: some bats learn to vocalize only when young, others seem to be open-ended learners.

Scientists can order a transgenic mouse but few resources exist for non-standard model organisms, says Yartsev. That should not, in his view, deter labs from picking non-traditional model species and building what they need[7]. His lab team builds sound systems, setups for electrophysiology or optogenetics-based experiments. Optogenetics components for rodents work readily with bats, he says.

Consumer electronics have become so modular that labs can find microcontrollers and many other parts, says Schmid. He is building a bat-behavior box and calcium-imaging setup with open-source technologies such as 'miniscopes', miniaturized two-photon microscopes developed at UCLA. "You can build entire behavior and imaging or e-phys setups for relatively cheap," he says.

The lab needs wireless tools. "We cannot record from a flying bat if it's not going to be a wireless system," says Yartsev. Bats vocalize when they are close to one another, so a classic microphone won't reveal who is 'speaking'. The scientists developed a microphone, a wireless piezo-electronic device, as a bat necklace to detect with high accuracy who's vocalizing, he says.

Early results about the bat brain cortex indicate that regions active in vocalization could be analogous to human brain regions involved in speech, says Schmid. "We see analogous genes in the same place, and we see roughly analogous connections with our tracing—with our anatomical tracing," he says.

The Yartsev lab is not sequencing but contributes Egyptian fruit bat tissue to the Bat 1k sequencing project. Genetic resources for non-traditional model organisms will help the community compare birds, bats, people and other mammals, says Yartsev, and foster collaborative research. Labs might consider knockdown experiments to explore gene regulation and one day use CRISPR-based experiments. Bird vocalization and behavior are better understood than
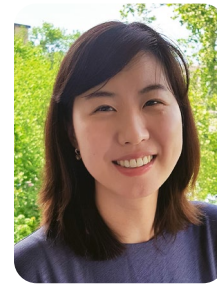
Phasing will help with the assembly of vertebrate genomes, says Arang Rhie.

bat vocalization, he says. By integrating techniques and a whole-organism approach, he hopes to tease out details of circuits and gene expression during vocal behavior.

Correcting genomes

The Sanger sequences of chicken, turkey and zebra finch DNA have long been available. But researchers have run into issues with the assembled genomes, says Jarvis. Scientists need genomes from many more species. Jarvis was frustrated with errors and gaps in bird genomes, including data he and colleagues had generated. "It was costing my students many, many months, sometimes a year's work, trying to fix gene structures, clone out the right gene," he says. These delays and headaches when working on individual genes motivated him to help build broader collaborative efforts and resources for the scientific community.

For two years, the scientists worked on technology development, says Jarvis, also with companies such as Pacific Biosciences (PacBio), which makes long-read sequencers[8]. The goal was a sequencing and assembly pipeline that delivers reliable data so researchers don't need to reclone or resequence genes of interest.

Adam Phillippy, a researcher at the National Institutes of Health's National Human Genome Research Institute (NIH/NHGRI), and Arang Rhie, a postdoctoral fellow in his lab, are in the assembly working group of the Vertebrate Genomes Project (see Box 1, "Some vertebrate genome projects"). The team has built and tested the pipeline now in place. Sequences are uploaded to GenomeArk, a dedicated spot on an Amazon web services cloud with hosting free of charge. Assembly happens on the cloud because it's easier "to take analysis to the data" than to download data to a lab's compute cluster, says Phillippy. The datasets are too big to adequately store in National Center for Biotechnology Information databases.

The desire for data breadth from many species is, says Phillippy, representative of labs "wanting to move away from a

---

Box 1 | Some vertebrate genome projects

In 2009, Genome10k was launched by David Haussler of the University of California, Santa Cruz; Stephen O'Brien, formerly of NCI-Frederick, now at T. Dobszhansky Center for Genome Bioinformatics at St. Petersburg State University; and Oliver Ryder from the San Diego Institute for Conservation Research.

The Avian Phylogenetics Project began in 2010, spearheaded by Guojie Zhang, from the University of Copenhagen and China National GeneBank, BGI Shenzen; Rockefeller University researcher Erich Jarvis; and Tom Gilbert of the University of Copenhagen. This project led to B10k, which aims to sequence all 10,500 known bird species. Then came the idea for the Vertebrate Genomes Project. "Instead of trying to pick and choose, I just said 'let's do 'em all,' " says Jarvis. There are an estimated 66,000 vertebrate species.

The Vertebrate Genomes Project started sequencing genomes in the spring of 2017 and is about to release its first dataset of more than a dozen assembled genomes, among them those of several vocal learners. Sequencing takes place at Rockefeller University, the Wellcome Sanger Institute and the Max Planck Institute (MPI) of Molecular Cell Biology and Genetics.

The data are going to GenomeArk, on an Amazon cloud, an arrangement facilitated by the genome-analysis company DNANexus.

Bat 1k aims to sequence all approximately 1,300 species of bats. The initiative involves researchers at Max Planck Institute for Psycholinguistics and the MPI of Molecular Cell Biology and Genetics, Stony Brook University, Texas Tech University, University College Dublin, University of Copenhagen and others.

(Sources: E. Jarvis, Rockefeller Univ.; A. Phillippy, A. Rhie, NIH/NHGRI; Springer Nature)

---

single-genome way of thinking to a more population way of thinking in comparative genomics." Data will be released in batches. "When people see this huge dataset that's so rich, it'll be irresistible to go develop new methods for comparative genomics, assembly, annotation."

In phase

Previously, Rhie developed methods to explore details of the human Korean genome[9]. Her phasing methods for separating haplotypes revealed genomic regions with insertions and copy-number variation particular to the Asian population. Some tools have emerged since then, but at the time there were none for phasing PacBio sequence reads. "I had to write my own code," she says. In her view, phasing will help with the assembly of vertebrate genomes.

The assembly pipeline now in place for the Vertebrate Genomes Project involves sequencing with PacBio sequencers, scaffolding and assembly steps with commercial platforms and services from companies such as Arima, Bionano Genomics, Dovetail Genomics, Phase Genomics and 10X Genomics. Hi-C assays are helpful for scaffolding, too. When two contiguous genome chunks, contigs, are found to interact heavily in Hi-C analysis, says Phillippy, they are more likely to be next to one another.

"I tried all kinds of combinations," says Rhie: sequencers such as Oxford Nanopore platforms, many software tool combinations. The assembly pipeline will keep evolving. "Six months from now we might adapt to a new pipeline and want to use that for everything," says Phillippy.

Genome assembly with phased long reads can help fix misassembled genes such as heterozygous *DUSP1* in the zebra finch. (Credit: Adapted from ref. 8., OUP; A. Rhie, NIH; E. Dewalt, Springer Nature)

When choosing assembly tools, says Rhie, one consideration is the number of repeat sequences in a genome. They have to be overcome with adequately sized contigs. That's challenging to estimate when many of these species are being sequenced de novo. The human genome has a highly repetitive six-kilobase (kb) sequence called LINE, for long interspersed element. PacBio reads are generally around 10 kb long, which gives adequate coverage for at least one copy of LINE. But when repeats are tandemly laid out throughout the genome, it gets harder to have contigs of the right length. Even though they test all approaches on the human reference genome, says Phillippy, each genome presents its unique challenges.

Some reference genomes contain errors that have become apparent during alignment of gene expression data. The zebra finch sequence was completed with Sanger sequencing, but the haplotypes were combined, which disturbs the gene structure, says Phillippy. These genomes will be sequenced again and phased, as will the human reference genome, he says. He, Rhie and others have developed 'trio binning'[10], a phasing method using parental and offspring genomes. Phased genomic information shows haplotype differences, which means teams can reconstruct structural variants at the sequence level across species and work at higher resolution.

"We want the same quality that we have for human for all of these vertebrates," says Phillippy. As a computer scientist who has long worked in genome assembly of many organisms, he knows vertebrate genomes will present many puzzles. "We're problem solvers," he says of the assembly team. "This is more exciting to me than working at Facebook."

Comparative genomics at scale

Jarvis says the genomic data will help scientists explore differences between known vocal learners and non-learners, as well as with tracking of genetic changes that occurred in vocal learners. Comparative approaches can shed light on a possible behavioral continuum between open-ended and closed-ended vocal learners, an idea advanced by his former student, Gustavo Arriago. The sequences will offer ways to study the genetic underpinnings of song diversity: among the 4–5,000 songbird species are those with simpler and more complex song.

The information reaped from scanning the vertebrate tree will deliver insights for much neurobiological and behavior analysis and for larger questions, says Jarvis. "Once we have all these genomes done, I think we'll be able to redefine the concept of what is a

Labs are moving toward more population-based thinking in comparative genomics, says Adam Phillippy.

living species," he says. Researchers can look at genomic distances and redefine relations across the entire vertebrate tree.

A pervasive assumption in some scientific circles is that it's best to use closely related species for studying biology and behavior, says Gentner. But that is not always true: for the study of vocal learning, to also understand what makes human language unique, mice are inferior to songbirds, he says. The natural world does not offer infinite options for comparison. Technology needs to catch up, so labs studying a variety of animal species can draw on the options currently available for mice. "If we focus all our resources on a very small number of model organisms, then we run the strong risk of letting our methods define our questions, rather than the other way around," he says. Large genomic reference datasets will help with analysis of a cross-species library of birdsong. Using these data to make genetic modifications, teams can explore cell-specific mechanisms and neural circuit specificity in the songbird in ways currently possible only in the mouse. The genomic data, he says, are "a gateway to a whole new set of techniques that we can apply to a whole old set of questions." ❐

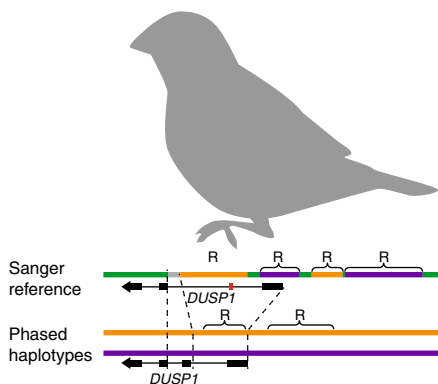Vivien Marx
Technology editor for Nature Methods.
e-mail: v.marx@us.nature.com

References
1. Kiggins, J. T., Comins, J. A. & Gentner, T. Q. *Front. Evol. Neurosci.* **4**, 6 (2012).
2. Lattenkamp, E. Z. & Vernes, S. C. *Curr. Opin. Behav. Sci.* https://doi.org/10.1016/j.cobeha.2018.04.007 (2018).
3. Jarvis, E. D. *Annu. Rev. Anim. Biosci.* **4**, 45–59 (2016).
4. Abe, K. & Watanabe, D. *Nat. Neurosci.* **14**, 1067–1074 (2011).
5. Zhang, G. et al. *Science* **346**, 1311–1320 (2014).
6. Pfenning, A. R. et al. *Science* **346**, 1256846 (2014).
7. Yartsev, M. M. *Science* **358**, 466–469 (2017).
8. Korlach, J. et al. *Gigascience* **6**, 1–16 (2017).
9. Seo, J. S. et al. *Nature* **538**, 243–247 (2016).
10. Koren, S. et al. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2018/02/26/271486 (2018).