The Neurobiology of Vocal Communication

Researchers at Duke University Medical Center are studying songbird genomes for clues about human vocal learning.

Introduction

Vocal learning, the ability to hear a sound and repeat it, is a complex behavior that few animals have evolved to perform. Studying songbirds and other animals that can imitate sounds is providing insight into molecular mechanisms underlying the capacity for spoken language, one of the crucial traits that differentiates humans from other animals.

Erich Jarvis, PhD, a Professor of Neurobiology at Duke University Medical Center and Investigator of the Howard Hughes Medical Institute, leads a team of researchers who study songbirds to understand the neurobiology of vocal learning. They’re using songbirds as a model for how the human brain generates, perceives, and learns spoken language.

iCommunity spoke with Dr. Jarvis about his research and how Illumina sequencing systems are advancing knowledge into the genetic basis of vocal learning and the mechanisms of brain function.

Q: Why are songbirds a good model for studying vocal learning?
Erich Jarvis (EJ): Songbirds, parrots, hummingbirds, and a handful of other animals have the ability to imitate sounds. These include some mammals such as dolphins, whales, sea lions, bats, elephants, and, of course, humans.

In contrast, monkeys and great apes cannot imitate sounds using their larynx. Many people find this quite surprising because they are our closest relatives among primates. Monkeys and great apes can only produce innate sounds that they learn how to create in different contexts. However, the actual cryptic structure and the syntax are innate.

So although humans and song-learning birds are distantly related to each other, we share a similar organization of brain pathways for vocal communication. These include a premotor pathway necessary for vocal learning and a motor pathway necessary for production of learned vocalizations.

Q: What is the Bird 10,000 Genomes (B10K) project?
EJ: I’m part of the B10K initiative, which was officially launched in 2015 by BGI and other collaborators to generate genome sequences from all extant bird species within the next 5 years. Currently, we are using Illumina technology for sequencing. The B10K Project builds upon the success of the Avian Phylogenomics Project, which provided the first proof of concept for carrying out large-scale sequencing of all representative species of orders across a vertebrate class, and a window into the types of discoveries that can be made with such genomes. The Avian Phylogenomics Project was lead by Guojie Zhang of BGI, Tom Gilbert of Copenhagen University, and myself.

The B10K project will allow the completion of a genomic level “tree of life” of the entire living avian class, enabling us to identify the links between genetic and phenotypic variation. In the process, we hope to uncover the correlation of genetic evolutionary, biogeographical, and biodiversity patterns across a wide-range of species, evaluate the impact of various ecological factors and human influence on species evolution, and unveil the demographic history of an entire class of organisms.

My focus is studying genetics of vocal learning. My investigations require that we compare the genes, vocal behavior, and associated brain pathways of the few rare groups that have vocal learning, with most species that do not.

Q: How many bird genomes have been sequenced as part of the B10K project and what sequencing technologies were used?
EJ: At last count, we’ve sequenced more than 200 bird genomes representing one species per family, most done at BGI. Forty-eight of these genomes have already been published. For several reasons, we’re attempting to sequence in stages of deeper taxonomic levels of this project. First, it gives us phylogenetic breadth at each successive taxonomic level. Second, we could use the genomes at each level to understand the phylogeny of birds, which has been a contentious issue for decades.

Previously, the chicken and the zebra finch were sequenced using Sanger sequencing. Other bird genomes were sequenced using next-generation sequencing (NGS) technologies. Most were sequenced using Illumina sequencing systems, while a few others were sequenced with either Roche 454 or PacBio sequencing.

Q: Why did you choose NGS and the HiSeq System for your studies?
EJ: Repetitive regions are hard to assemble using shot gun sequencing. We began using NGS to generate jumping libraries that provided better assembly quality through repetitive regions, and TruSeq® DNA Library Prep (version 3 chemistry) that provides better sequencing through GC-rich regions. For the bird genomes

Erich Jarvis, PhD is an Professor of Neurobiology at Duke University Medical Center.
that we published at the end of 2014, about half of them were performed using multiple jumping libraries, the other half with just 2 jumping libraries.5,7

The primary reasons that we chose the HiSeq System were cost, coverage, speed, and flexibility. We need a certain amount of coverage to accurately call a base pair to generate base-pair accuracy and to obtain reads through hard to sequence regions. We also need sequencing at a certain price point to cost-effectively generate the 20–30× coverage needed, or up to 100× sometimes for generating higher-quality assemblies. The HiSeq System is also fast and enables us to sequence multiple species in one lane, in one run. That quickens the pace of our studies by increasing the volume of species we can sequence at one time. Finally, it’s flexible, with pools, libraries, and algorithms that can take Illumina-based data and perform many things with it in different ways.

“At last count, we’ve sequenced more than 200 bird genomes representing one species per family.”

Q: Outside of the B10K project, are you participating in any other sequencing initiatives?
EJ: I’m working with leaders of the G10K project, which is focused on sequencing 10,000 vertebrate genomes.5 These leaders include Drs. Steven O’Brien of St. Petersburg State University, David Haussler of University of California, Santa Cruz (UCSC), Oliver Ryder of the San Diego Zoo, Klaus-Peter Koepfli of The Smithsonian, Beth Shapiro of UCSC, and myself, among others. Depending on whose number you trust, there’s between 60,000–66,000 vertebrate species. We hope to sequence one genera per vertebrate species, which adds up to about 9500 species.

I’ll be using the bird and other vertebrate genomes to answer questions on the evolution and mechanisms of vocal learning. I’m specifically interested in studying the genomes of the different vocal learners to understand why, even within each lineage, different bird species to imitate and learn new songs while others cannot. Then I’d like to manipulate these genetic differences into mouse and vocal non-learning bird brains. Some of these genes do control connectivity, and unpublished findings suggest that they might control some vocal behavior. However, we’ve not yet identified a master regulating gene that’s changing the whole network of genes responsible for creating these brain circuits.

“We’re now repeating those array experiments with RNA-Seq using Illumina sequencing systems to identify genes that were missed.”

Q: Are you also performing follow-on RNA or mRNA studies?
EJ: In 2014, we published RNA expression analysis using microarrays, where we found convergent gene expression changes in humans and vocal learning birds.6 The convergent gene expression differences occurred in human speech brain areas and in songbird song brain areas. We’re now repeating those array experiments with RNA-Seq using Illumina sequencing systems to identify genes that were missed. The genes identified in the microarray RNA expression studies will serve as positive controls. The advantage of RNA-Seq is that it can detect all genes expressed in the brain regions and their cells. RNA expression analysis using microarrays can detect only the limited set of genes that are placed on the array. Because of this, we feel very confident that we missed potentially important genes, such as a master regulator that causes differences in a large gene network in speech areas.

Beyond that, we’re trying to find out why these brain regions have convergent changes in tens of genes—up to 50–70 genes per brain region—between birds and humans. We’re using ChIP-Seq to see if there are enhancer differences in motor regions in vocal learning birds and in the speech areas of the human brain. If we didn’t have the ChIP-Seq approach, it would take decades, if not hundreds of years, to identify the enhancer region differences one gene at a time per brain region.

“We’re now repeating those array experiments with RNA-Seq using Illumina sequencing systems to identify genes that were missed.”

Q: Does your songbird research have implications for understanding human speech?
EJ: For a long time, I thought we would make all these discoveries in birds and then other scientists would take those discoveries and apply them to humans. I soon discovered that even though people were excited by our work, they weren’t translating our findings into humans. I decided a few years ago to perform those studies myself.
This includes the study that compared the genomes and brain regions of birds and humans and how vocal learning genes are expressed. We were helped by colleagues with other institutions involved in analyzing human and non-human primate brains, such as the Allen Institute for Brain Science and RIKEN Brain Sciences Institute. There could be clinical applications that result from this research. After we figure out how to induce or modify communication circuits in the mouse, we want to see if we can figure out how to repair them when they’re damaged. Ultimately, that could enable us to repair speech brain circuits after a stroke or other type of trauma, or develop drugs designed to modulate specific genes in the brain circuits of autistic children.

If we didn’t have the ChIP-Seq approach, it would take decades, if not hundreds of years, to identify the enhancer region differences one gene at a time per brain region.

Q: What are some of the next steps in your research?
EJ: I’d like to work on a project that will take the genomes from the B10K group and redefine the species concept. In the past, the way species identification and distinction were determined was mainly through morphology. For example, a bird’s wing color or shape, or the size of a mammal’s paw would be used to determine which species are related to each other and who gave rise to whom. In studying the genomes in B10K, we’re finding that many parts of these morphology-based phylogenies of species are wrong, because many morphological features are convergent. If we look at the underlying genome, we find there are differences in who is related to whom. You also find that sometimes what we considered to be a single species, is really 2 different species, and what we considered 2 species, is actually one.

I am also working with my colleagues Beth Shapiro and Ed Green at UCSC, who are interested in trying to reconstruct the common ancestor of a set of species. If you have the genome sequences of all species of a particular class and there’s enough diversity represented, you have algorithms that can theoretically reconstruct the common ancestry of every base in the genome. We could then synthesize those chromosomes, put them into cells, fertilize an embryo—let’s say a chicken or some mammal—and give birth to this common ancestor genome.

References

Learn more about the Illumina systems and products mentioned in this article: