Assembling the Kiwi Genome to Understand Unusual Adaptations in Sight, Smell, and Flight

With the Nextera® Mate Pair kit, researchers generated up to 13 kb insert-size libraries for *de novo* sequencing of the kiwi genome on the HiSeq® System.

Introduction

The kiwi (*Apteryx mantelli*) is an evolutionary marvel. During the last 50 million years, it has evolved to become a flightless, nocturnal bird that forages for worms, seeds, and small ground dwelling creatures. It makes up for its poor eyesight and inability to fly with an exceptional sense of smell, thanks to nostrils on the end of its beak.

About the size of a chicken, the kiwi is the national bird of New Zealand and the subject of a recent sequencing study by Diana Le Duc, M.D., a Ph.D. student at the Max Planck Institute for Evolutionary Anthropology and part of the Medical Faculty in the Institute of Biochemistry at the University Leipzig. Working in collaboration with the Molecular Biochemistry Institute and The Minerva Research Group for Bioinformatics, her research isn't focused on the kiwi itself. She's more interested in what the bird's genome might tell us about the evolution of the genes responsible for the bird's unique features.

This perspective has been championed by the Head of the Institute of Biochemistry at the University Leipzig, Prof. Dr. Torsten Schöneberg, who uses evolution as a tool to understand the physiological relevance of individual G protein-coupled receptors (GPCRs) and to interpret the clinical significance of GPCR mutations. Using the Nextera Mate Pair Library Preparation Kit and next-generation sequencing (NGS) on the HiSeq System, Dr. Le Duc completed *de novo* sequencing of the kiwi genome. The data has provided information on the role of GPCR mutations in the bird's unusual traits, which may have implications into limb and vision development in humans. iCommunity spoke with Dr. Le Duc to learn more about the obstacles she faced in sequencing the kiwi genome and what she uncovered using Illumina technologies.

Kiwi are flightless birds native to New Zealand.

Diana Le Duc, M.D., is a Ph.D. student at the Max Planck Institute for Evolutionary Anthropology and part of the Medical Faculty in the Institute of Biochemistry at the University Leipzig.

Q: Why is the kiwi bird a good candidate for studying GPCR and other types of mutations?

Diana Le Duc (DLD): The kiwi is unique among birds. There are five kiwi species and all are nocturnal, lack wings and a tail, and have a very long bill with nostrils at the tip. We wanted to understand the genetic basis of these traits, including the role GPCRs play in the kiwi's sensorial adaptation and what variants are responsible for the lack of forelimb development. Assembling the kiwi genome enabled us to ask biological questions. We were lucky enough to find some changes in the genome that pointed to their unique biology.

Q: What size libraries did you produce initially as you began sequencing the kiwi genome?

DLD: When we started the project, we could only generate short insert-size libraries (240, 420, and 800 bp) and mate pair libraries with a range between 2 to 4 kb. We obtained coverage of about 50 fold, and we managed to get a good assembly. However, it was considered a draft genome because it was rather difficult to annotate the genes. The libraries required more than 10 µg of DNA, which became a bottleneck for us. The kiwi is an endangered species so it's difficult to access DNA. In fact, we had to obtain kiwi DNA from abandoned eggs that could not hatch through BNZ Operation Nest Egg, an organization dedicated to protecting the kiwi.

Q: Did the Nextera Mate Pair kit enable you to generate the libraries you needed?

DLD: We received the HiSeq System about the same time that the Nextera Mate Pair kit became available. With the Nextera Mate Pair kit, we could produce much larger libraries, from 7 to 13 kb, and generate the final assembly of the kiwi genome. Nextera Mate Pair also allowed us to reduce our DNA input down to 4 µg. We then had contiguous genes that we could annotate and draw biological conclusions. It helped move our draft genome to an assembled genome.

> With Nextera Mate Pair, you don't have to spend so much time in the lab. It gives you more time to spend on the analysis.

Q: How did the Nextera Mate Pair kit improve your workflow?

DLD: I like the fact that we don't need to shear DNA with the Nextera Mate Pair kit. Shearing requires more input DNA, because you lose so much material in the process. Also, the bigger the insert size the more material you lose in the library building process. With Nextera Mate Pair, the tagmentation step does the shearing for you. You just put the DNA on a gel and extract the sizes you're interested in. It's a great improvement, with a better workflow and less hands on time than other kits. With Nextera Mate Pair, you don't have to spend so much time in the lab. It gives you more time to spend on the analysis.

Q: How did you assess the quality of the kiwi genome assembly?

DLD: We assessed the quality of the assembly by comparing it to annotated genomes of the zebra finch and chicken. Comparatively, we did not see rearrangements and the genomes possessed common chromosome sequences, meaning synteny was kept. We did not have the entire chromosome, but our N50, which describes scaffold length, was good enough at about 4 Mb.

Q: What did the sequence tell you about the role of GPCR receptors in the kiwi genome?

DLD: To understand the genes involved in vision and olfaction, we annotated the olfactory receptor repertoire and all of the GPCRs. We found that olfactory receptors, classified as sensory GPCRs, have a higher diversification in these birds, providing them with a keen olfactory sense that they use in foraging. We also looked in genes we know are responsible for vision, like rhodopsin. We found that several opsins, belonging to the rhodopsin GPCR class that is involved in color vision are inactivated, most likely due to nocturnal adaptation.

Since the kiwi doesn't have problems with its lower limbs, we focused on genes known to be involved only in the development of the upper limbs. We annotated the Hox cluster and we looked in all of the fibroblast growth factors (FGF). One of the genes responsible for upper forelimb development, fibin, is most likely disrupted in the kiwi genome, and a potential cause for the absence of wings.

Q: How did you identify that the fibin gene may be disrupted in the kiwi genome?

DLD: We couldn't find this gene at all when we first annotated the genome. We used the Basic Local Alignment Search Tool (BLAST) to find regions of local similarity between a collection of sequences responsible for this gene. We took about 20 species of birds, mammals, and fish and blasted it against our entire assembly to see whether there is any remnant of it. We found what appears to be a gap close to the region where the fibin gene is supposed to be. We found the 3′ UTR, a sequence of about 500 bp, but we did not retrieve any significant hits for an adjacent 280 bp. The gap, of about 2000 bp, is towards the 5′ end. Given the fact that we did not see any mate pair support for that gap, we wondered why it appeared there.

> It's amazing that we are now at the point where technology, like Nextera Mate Pair and the HiSeq System, make it so easy to assemble genomes and come to biological conclusions based on the data.

The fibin gene is well-conserved gene among species, so we aligned short insert size libraries against all of the known sequences for these genes from other species. We couldn't align anything. We tried long-range PCR and designed primers according to the sequences from other species, and this also failed to retrieve anything. It could be a region that is difficult to sequence, but we had good coverage and the GC content was not that high.

Q: How can you prove the gene is no longer there?

DLD: A loss, the absence of something, is hard to prove. Fibin knockdown has only been biologically tested in zebrafish, not in birds. We know that the synteny is kept on the scaffold. The genes that are supposed to be to the right and to the left of the fibin are there on the scaffold where the 3′ UTR of fibin is located. The 3′ UTR is probably kept, but the rest of the gene is lost. I expect that there was probably a loss of constraint on this gene since the kiwi didn't need to fly anymore. It may be disrupted or completely missing. Or the gap may be filled with a sequence unrelated to the fibin. If that is the case, I would have to prove that the sequence doesn't align with and can't be PCR amplified with primers from sequences of other species. We need to research this further.

Q: What are the next steps in your research?

DLD: We will keep working on the fibin gene to determine if it's truly missing. Our lab is also conducting studies using transciptomics to characterize how a loss of a receptor impacts gene expression. I've been working on the GPCR34 knockout because this receptor is involved in immunity. The next step is to study the regulation of this receptor by using the HiSeq System to sequence transcriptomes of dendritic cells that are stimulated under different conditions.

Q: How have the HiSeq system and Nextera Mate Pair kit enabled your research?

DLD: It's amazing that we are now at the point where technology, like Nextera Mate Pair and the HiSeq System, make it so easy to assemble genomes and come to biological conclusions based on the data. It took a large team of specialized experts 10 years to work on the human genome. But now someone like me, an M.D. without any previous computational background, can sequence, assemble, and analyze a genome in three years. I believe that NGS is going to be used further in medicine. That's another reason why this project is so important to me. I expect that in the future, I'll be able to translate the results into the medical field, benefiting research into limb and vision development.

Illumina, Inc. • 1.800.809.4566 toll-free (U.S.) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

FOR RESEARCH USE ONLY

Pub. No. 770-2014-029 Current as of 03 September 2014.

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGAGC
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA