# Versatile, High-Throughput Sequencing Supports Australia Genomic Center Growth

NGS expands genomic projects for species ancient and new, large and small.

## Introduction

Two driving forces were at the heart of creating the Deakin Genomics Centre in Australia.[1] The first was to lower sequencing costs and provide a faster turnaround of sequencing data to Deakin researchers. The second was to create a genomics hub that could serve the university and greater community, including farmers, industry, and conservation scientists. It's delivering much more than that.

Chris Austin, PhD, Professor and Head of the School of Life and Environmental Sciences at Deakin University, established the Deakin Genomics Centre in late 2017. It's located at the University's main campus at Waurn Ponds Geelong, close to its Queenscliff Marine Research Station. In addition to a MiniSeq™ and two MiSeq™ Systems, the facility has a NovaSeq™ 6000 System to handle high-throughput, large-scale sequencing projects. Staffed with six people, the Deakin Genomics Centre is a vital teaching facility that supports the growth of the Deakin genomics research community and undergraduate and graduate teaching programs. It's also stimulated genomics collaborations with researchers in Australia, and well beyond its shores, who are studying a wide range of species and performing a variety of projects.

"We never know what plant or animal species we'll be working on next," Dr. Austin said. "We've sequenced everything from the ancient Wollemi pine to alpacas, clownfish, prawns, abalone, fungi, and bacteria. The Deakin Genomics Centre has become a hub for collaborations, rather than just a generator of fee-for-service agreements. We're meeting new researchers all the time because of the capabilities of our facility ."

iCommunity spoke with Dr. Austin and Larry Croft, Associate Professor of Genomics and Manager of the Deakin Genomics Centre, about its agriculture and aquaculture collaborations, biodiversity and conservation studies, and the value that next-generation sequencing (NGS) is bringing to these research efforts.

**Q: Why was the Deakin Genomics Centre established?**
**Chris Austin (CA):** When I arrived at the Geelong campus last year, there were several researchers undertaking genomic studies in different disciplines and Schools. However, they were isolated and independent, and most of the sequencing was outsourced at unnecessary expense, often with long delays and without any assistance on experimental design, sample preparation, or bioinformatics.

We realized that a genomics facility could provide a new, valuable research platform for Deakin University. By establishing the Deakin Genomics Centre, we could offer genomics services to researchers and support their studies from inception to data analysis. We'd also be able to introduce students to the latest genomic technologies and methods and give them the best chance of earning advance degrees or obtaining jobs in the commercial sector.

**Q: How has NGS advanced research studies that were once performed with older genomic methods?**
**Larry Croft (LC):** We've seen NGS transform research studies. For example, we identified a fascinating, bright red *Pseudomonas* colony more than 10 years ago. We knew that there was a point mutation somewhere in the genome, but we didn't know where, so we put the sample in the freezer. A decade later, we sequenced it with a MiSeq System and within three hours we identified the mutation that caused the red color. It was probably the fastest turnaround for a scientific paper that I've ever worked on.

**CA:** NGS added a significant new dimension to the research of Professor John Endler, a famous scientist at Deakin University and a member of the Australian Academy of Science. He's been maintaining populations of thousands of guppies under different light regimes for many years in the basement of our department. He's studying sexual selection and aspects of behavior under these environments and looking at the evolution of various phenotypic traits.

Chris Austin, PhD is Professor and Head of the School of Life and Environmental Sciences at Deakin University and Larry Croft is an Associate Professor of Genomics and Manager of the Deakin Genomics Centre.

A few years ago, we didn't have an NGS system that could enable us to sequence 80–100 individuals, let alone 1000s in a study. The NovaSeq 6000 System easily sequenced many guppy genomes cost-effectively. We've already identified specific genes related to the important phenotypic traits documented in these experimental populations. It's a classic case of how NGS can add another layer of value to existing research programs.

> "Whether its rice, cucumbers, silkworms, or alpacas, we're seeking to identify genetic markers to improve the species, while maintaining inherent diversity."

**Q: What types of collaborations is the Deakin Genomics Centre attracting?**
**CA:** Having the NovaSeq 6000 System is attracting interest from international collaborators. The university has strong ties in India and I have experience working with institutes in Vietnam and Malaysia. In addition to partnering with us on genomic projects in their countries, they are sending their students and staff to Deakin University for training. It's one more way we're benefiting from this new infrastructure and the presence of the NovaSeq 6000 System.

In addition, many of these international projects are for agrigenomics or aquagenomics studies of important commercial species, such as shrimp and prawns. That's very useful for Deakin University. We want to be viewed as a university performing high-quality research that has global impact and that matters to communities, such as improving efficiencies at the farm level through selective breeding.

Genomic selection is a genomics-powered extension of the traditional breeding that farmers have been performing for centuries. In developing countries, they often don't have access to genomics or an infrastructure to assist farmers in using genetic selection to improve their crops, herds, or aquaculture breeding. Improving a high-protein food staple, such as fish or shrimp, is a powerful strategy for countries to improve the well-being of its citizens.

**Q: What types of agriculture projects are you working on?**
**CA:** Our agriculture research tends to look at commercial crops and less conventional species. Whether its rice, cucumbers, silkworms, or alpacas, we're seeking to identify genetic markers to improve the species, while maintaining inherent diversity.

**LC:** The downside of traditional agriculture breeding is that it is essentially inbreeding. For example, alpacas have been bred for a desirable fleece. However, the genetic marker for that fleece is linked to a genetic disease that causes alpaca stillbirths. Sequencing enables us to identify and keep the traits we want, while removing the problems of inbreeding. By mixing the population, we're outbreeding and bringing diversity and wildness back into the population.

**Q: How are genetic selection projects for aquaculture different than those for agriculture?**
**CA:** Genetic improvement in aquacultures can be more challenging than in agriculture, because of the nature of the species. We have limited preexisting genomic knowledge of most fish and crustaceans. It reduces our ability to use molecular-based selective breeding as a tool to improve growth or disease resistance. Our first step in those cases is to sequence the genomes.

**LC:** Aquaculture genetic research often supports conservation studies. It brings together people breeding fish for specific traits and people focused on conserving the wild form of the species. It's important to bring wild fish into the breeding population to ensure diversity. As a result, breeders are invested in conservation of the wild species.

> "...we can perform large-scale population sequencing at great depth, inexpensively. That supports endangered species sequencing and new applications like eDNA sequencing for biodiversity studies."

**Q: How does NGS support conservation and environmental DNA (eDNA) biodiversity studies?**
**LC:** We've become very involved in sequencing endangered plants and animals. One of the first things to happen with endangered species is that they become isolated into fragmented population islands. Those fragmented populations start inbreeding, the genetic diversity of the population decreases and the species begins to die off.

**CA:** We can perform large-scale population sequencing with the NovaSeq 6000 system at great depth inexpensively. That supports endangered species sequencing and new applications like eDNA sequencing for biodiversity studies. In the old days, we had to grab a net and go out in the field to directly sample organisms that we were interested in sequencing. Today, we obtain a water sample, filter it, extract eDNA, and sequence it. It's a completely different perspective in biodiversity surveys.

**Q: What types of eDNA studies have you performed?**
**CA:** We are conducting eDNA studies of sharks and invasive marine species with one of our collaborators. We sequenced ocean water samples to detect their presence and can even identify genetic signatures of individuals.

We also sequenced a freshwater crayfish, known locally as the northern yabby or red claw crayfish (*Cherax quadricarinatus*), that has a large repetitive genome of over 200 chromosomes. Many of its populations are localized and threatened by environmental changes. The red claw crayfish is used for aquaculture and has been introduced widely outside their native range throughout the world. eDNA studies enable us to understand the extent and dynamics of its populations.

**LC:** We were surprised by the size of the red claw crayfish genome. There seems to be a correlation between large genome size and low population diversity. These crayfish live in isolated streams, so are often breeding with their relatives. That limits population mixing and genetic diversity.

We sequenced the mitochondrial DNA of the red claw crayfish and placed them into taxonomic trees. We then took water samples from various streams and performed eDNA sequencing to determine the crayfish population structure. It enabled us to identify streams where there is significant inbreeding or where there are endangered crayfish populations.

**CA:** We have another project in Vietnam to sample shrimp ponds and use eDNA sequencing to identify the presence of harmful viruses and bacteria. We look for a microbial signature that indicates if the ponds are healthy or might be susceptible to disease. The data might help us develop probiotics to improve the health of aquaculture species.

**Q: Have you performed *de novo* sequencing on any unusual species?**
**CA:** The first fish that my research group sequenced was the arowana, a freshwater fish found in Malaysia. After moving to Deakin University, we continued to study this species. It's an ancient fish that's at the base of the fish family tree. Its genome is of considerable interest to evolutionary biologists. The species has several different color forms and occurs naturally in different countries in southeast Asia, including Malaysia and Indonesia.

The arowana is one of the most economically valuable fish in the world. The Deakin Genomics Centre team created the first arowana reference genome.

The arowana is one of the most economically valuable fish in the world. In addition to having commercial value, it's known as the dragonfish and considered a lucky fish by people of Chinese heritage. Natural arowana populations are bright red or gold in color and there is a significant aquaculture industry. However, there are major challenges in breeding them because it's difficult to determine the sex of the fish. We have not yet nailed down the mechanism of sex determinism.

When we decided to sequence the arowana, we attracted significant interest from the fish farming community and people interested in using biotechnology to support breeders and suppliers. The project has been beneficial for the Deakin Genomics Centre because it has opened doors to working with companies to develop a test for fish gender easily and identify markers for the different color forms. As a result, creating the first arowana reference genome has led to various other projects, extending the project beyond its initial scope.

We also sequenced the Murray cod, which is Australia's largest freshwater fish. It's popular with anglers and is stocked in many different water ways. Our research generated interest from aquaculture researchers and even conservationists. There are some populations that are of special conservation significance, which need to be managed carefully. Once again, having a reference genome has provided a platform that enabled us to expand our original studies.

**Q: Are there any unique species that you've sequenced?**
**CA:** Australia has been isolated geographically for millions of years, so we have many plants and animals that are found only here. Everyone is familiar with our kangaroos, koalas, and parrots, yet our plants and invertebrates are equally distinctive and interesting biologically.

As environmental biologists we're always interested in studying unique flora and fauna. Fairly recently, the Wollemi pine was discovered in a canyon about 150 kilometers outside of Sydney. It's basically a living fossil.

**LC:** Finding a living Wollemi pine is the botanical equivalent of finding a dinosaur in the wild. Millions of years ago, these trees once covered Australia, New Zealand, and Antarctica. They've identified about 60 other Wollemi pine trees in the wild, making the species critically endangered. It's now been propagated, with progeny growing all over the world. These pines grow easily, even at high latitude in the Northern Hemisphere.

**CA:** The NovaSeq 6000 System enabled us to sequence the Wollemi pine genome. It turns out that it has a large, repetitive genome, with virtually no genetic variation. It's remarkable.

**LC:** We're using a super computer to assemble the Wollemi genome, which is more than twice the size of a human genome. It's going to take several months to complete the assembly.

**Q: Have you sequenced any other interesting species?**
**CA:** I lived in Darwin and northern Australia and worked in the museum there with some colleagues who were studying clownfish taxonomy and evolution. I realized that no one had sequenced the clownfish, so we went down to the local reef and collected some. We selected a species that is brown in color (*Amphiprion ocellaris*) rather than the usual orange and white striped version that people know from the movie *Finding Nemo*.[2]

The clownfish reference genome will enable studies to answer ecological and biological questions about why there are so many color forms and how these fish tolerate the neurotoxins produced by the anemones that are their hosts.

**LC:** The Deakin Genomics Centre is assisting with Tasmanian devil studies. The Tasmanian devil population is in decline because of a unique transmittable cancer that has decreased its numbers. We're using genome sequencing to support selective breeding of the remaining animals to diversify the population and increase its genetic vigor.

We're also performing similar studies with quoll, an endangered marsupial. Deakin University is involved in several quoll breeding programs and we're supporting those efforts by sequencing the individuals.

**Q: What advice would you give someone who is considering making the transition to NGS?**
**CA:** For someone considering NGS, I'd recommend that they start slowly. We developed our knowledge and skills over time.

For researchers who can't afford their own NGS systems, I'd recommend collaborating with service providers. That will allow them to send in some samples and understand the benefits of the technology.

The Wollemi pine once covered Australia, New Zealand, and Antarctica millions of years ago. A living Wollemi pine was recently identified outside of Sydney.

We have a program where people send in their samples and we provide them with premium data, review the results with them, and then assist with the next steps. That's the difference between us and other service providers. We can assist with experimental design and library preparation and, if needed, data analysis and preparation of manuscripts for publication.

**Q: What are the emerging applications for NGS in agrigenomics?**
**CA:** I think NGS with the NovaSeq 6000 System will be a fundamental tool for advancing selective breeding of conventional and unusual species. We have sheep, wheat, and cotton conventional breeding work in Australia that will now also benefit from high-level genomics. There's also a range of species that we don't know much about genetically. To understand these species, we don't need a long history of genomic and genetic information anymore. We can pick a species and now have the capacity to sequence it and generate large amounts of data quickly.

I believe that whole-genome sequencing (WGS) will enable us to perform selective breeding more effectively. WGS increases the rate of genetic improvement more than 20% over traditional breeding. Microarrays do not have the density of markers for genetic improvement in most species. WGS gives us the breadth of genetic information to understand the links between desirable and deleterious traits, enabling us to maintain genetic diversity while still selecting for desired traits. This approach is going to be incredibly powerful.

**Q: What types of studies will you be using the NovaSeq 6000 System for in the future?**
**CA:** We're discussing a project to sequence all of Australia's freshwater fish species. It would benefit our understanding of the evolution of these species and contribute to their conservation. We're also working with the New South Wales Herbarium on a project to perform genome skimming of every plant species in New South Wales. These projects are perfect examples of the scale and nature of the projects that we hope to initiate in the future with the NovaSeq 6000 System.

**CA:** There's an incredible diversity of applications where sequencing could be useful. We're just beginning to address the possibilities to answer interesting environmental questions.

## See how the Deakin Genomics Centre is sequencing new species:

Sequencing Newly Discovered Species in Australia, www.illumina.com/company/video-hub/0QCP7wHJH6o.html

## Read more about sequencing at the Deakin Genomics Centre:

Deakin Genomics Centre Case Study: www.illumina.com/science/customer-stories/icommunity-customer-interviews-case-studies/austin-croft-deakin-case-study-ag-novaseq.html

## Learn more about the systems mentioned in this article:

NovaSeq 6000 System, www.illumina.com/systems/sequencing-platforms/novaseq.html

MiniSeq System, www.illumina.com/systems/sequencing-platforms/miniseq.html

MiSeq System, www.illumina.com/systems/sequencing-platforms/miseq.html

## References

1. Deakin Genomics Centre. Deakin University. www.deakin.edu.au/life-environmental-sciences/facilities/deakin-genomics-centre. Accessed April 2, 2019.
2. Finding Nemo: Scientists sequence the genome of iconic clownfish. Deakin University. www.deakin.edu.au/about-deakin/media-releases/articles/finding-nemo-scientists-sequence-the-genome-of-iconic-clownfish. Accessed April 2, 2019

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