Sequencing the "Tree of Life"

By sequencing the baobab tree, Monsanto is providing scientists with the genomic resources to protect and sustain the baobab, while informing our understanding of tree evolution.



Recognizing that its technology could play a critical role in alleviating global hunger, malnutrition, and poverty, Illumina created the Agricultural Greater Good initiative. Each year, Illumina awards Greater Good grants to agricultural research organizations that are focused

on identifying and breeding plants and animals that will increase the sustainability, productivity, and nutritional density of crop and livestock species. Under the grants, Illumina sequencing and genotyping reagents are provided free of charge.



Monsanto 2012 Illumina Greater Good Initiative Award Winner

Introduction

For people familiar with Disney's *The Lion King*, the African baobab tree is the "Tree of Life." The depiction in the animated movie isn't far from the truth. Capable of living thousands of years, it provides shelter, food, and water for animals and humans living on the African savannah.

In a departure from its usual focus on crop species, Monsanto is using its expertise and resources to better understand the baobab tree and what makes it so resilient. "We don't often have the opportunity to work on species outside of our portfolio," said Todd Michael, Ph.D., former Head of the Genome Analysis Center at Monsanto and now Senior Principal Research Scientist, Ibis Biosciences, Abbott Laboratories. "In fact, we couldn't have initiated this project without receiving the Greater Good Initiative award. We decided to sequence the African baobab species, *Adanonia digitata*, providing scientists in Africa with the genomic resources to protect and sustain the baobab, while informing our understanding of tree evolution."

Baobab trees can reach over 80 feet (25 meters) in height, with trunk diameters that range from 20–40 feet (6–12 meters). With large branches and leaves creating huge shade canopies, the trees rely on animals and humans for seed dispersal. "The baobab generates huge pods containing 50–100 seeds," said Dr. Michael. "While the pod is used for bowls and decorative items, the pulp is ground into a drink that's high in vitamin C, similar in taste to lemonade. The seeds are about the size of a pea and are disbursed by people and animals. People grow them and eat the seedlings, which are tasty and very nutritious. The bark is used for medicinal purposes."

Dr. Michael's team is using the Illumina HiSeq[®] 2000 system to sequence the *A. digitata* genome. "As far as I know, we're the first team to sequence the Baobab tree," said Dr. Michael. "Our first goal was to make a reference genome of *A. digitata* in order to access the variation of the organism itself, and then sequence the seven other baobab species at a lower level. One of the original scientific papers about this species reported that it is heterozygous, polyploidy, and has between 50 and 150 chromosomes¹. Chromosome number and the ploidy level will all play into how we develop markers for breeders. There are several groups in West Africa that want to use molecular markers to optimize baobab cultivation."

The team obtained seeds from four different *A. digitata* seed pods, taken from a cluster of baobab trees that were presumably cross



Todd Michael, Ph.D., was formerly the Head of the Genome Analysis Center at Monsanto and is now a Senior Principal Research Scientist at Ibis Biosciences, Abbott Laboratories.



Baobab trees can liver for thousands of years, reaching over 80 feet (25 meters) in height, with trunk diameters that range from 20–40 feet (6–12 meters).

October 2013

pollinated. They sequenced five seeds at 100× and one seed at 30×. "The amount of variation between the different seeds in that small population was very large," said Dr. Michael. "The interesting thing about a long-lived species is that typically they hold a lot of heterozygosity. That dovetails with being polyploidy and having lots of different chromosomes."

Scientists had estimated that the baobab genome was over a gigabyte. "We found it to be smaller than predicted, about 700 MB," said Dr. Michael. "That could mean several things. Since we don't have physical resources, such as BAC sequencing or physical maps, there's a chance that we're still missing part of the genome."

Low coverage sequencing was performed on seeds from the seven other baobab species, five of which are native to Madagascar (*A. grandidieri, A. suarezensis, A. rubrostipa, A. perrieri, A. madagascariensis,* and *A. za*) and one to Western Australia (*A. gibbosa*). "Our initial look at the data shows that the variation between the species found on different continents is as great as the variation within any given baobab species," Dr. Michael said. "Interestingly, we performed our initial sequencing on seeds, and many of the reports about genome size and characteristics were done on the leaf tissue of mature trees. We're wondering if there's a dramatic difference between the two, implying that the genome is going through lots of changes over time."

The team will be performing deeper sequencing at higher coverage of some of the species samples to determine the degree of variation. "With that information, we'll work with the Donald Danforth Science Center to develop markers," said Dr. Michael. "Those will be used to perform a larger population study to determine the extent of variation in places where there's been human selection versus places where there hasn't been any selective breeding."

The final goal is to understand how these trees live so long. "Clearly, they've been saved by the fact that they have no value from a wood standpoint," Dr. Michael said. "It's a succulent. You can remove the bark all the way around the tree and it won't kill it. There have been reports that when these trees die, they turn to mush."

The tree's longevity may also coincide with being high in nutrients or other beneficial characteristics. Understanding the genetic architecture of the baobab genome might help to explain the baobab's longevity. "A recent study showed there's definitely a positive selection or a selective force resulting from cultivation in West Africa². We want to determine whether cultivation is driving the baobab away from being a long-lived organism."

To study baobab longevity, the team would like to obtain leaf samples from several of the very old trees living in Africa. "We'd like to take a leaf from one side of the tree and compare it to a leaf from the other side of the tree, said Dr. Michael. "Those samples represent different points in the lifetime of the tree, possibly separated by thousands of years. By sequencing the methylome, and maybe even parts of the epigenome, such as histones, we could see the impact time has had on the genome."

According to Dr. Michael, the opportunity to explore genomes is really unprecedented. "A year or two ago, you wouldn't have thought about doing this type of project, because it would be too expensive to even think of trying. But the amount of sequence the HiSeq 2000 produces for its price point changes the questions we're able to ask and explore."

Scientists can't learn a lot from just one genome, preferring to view them in the context of larger populations. "We've spent the last 10 to 12 years in plant genomics learning a lot about a few genomes, *Arabidopsis*, rice, maize, etc.," said Dr. Michael. "There's a real opportunity to leverage the information we're getting from non-model, non-crop genomes like the baobab. We can start making observations about a greater number of genes and what they control."

Only several tree species have been sequenced to date. "The more genomes we sequence that represent different parts of the evolutionary tree, the more data we have to make deductions about the evolution of gene families," Dr. Michael said. "It turns out that the baobab is in the same family (*Malvaceae*) as cotton. It would be interesting to look at the differences between the diploid cotton putative D-genome parent (*Gossypium raimondii*), which is a small tree found in Mexico, and baobab genomes to identify the genes involved in driving the baobab to be long lived and those that confer medicinal characteristics. As we sequence more genomes, we can use them as positive and negative controls in studies. Marrying genomics, proteomics, and metabolomics will speed up the process of finding these beneficial characteristics."

References

- Baum, DA and Oginuma, K (1994) A review of chromosome numbers in Bombacaceae with new counts for Adansonia. Taxon 43:11–20.
- Assogbadjo AE, Kyndt T, Sinsin B, Gheysen G, and van Damme P (2006) Patterns of genetic and morphometric diversity in baobab (*Adansonia digitata*) populations across different climatic zones of Benin (West Africa). Ann Bot. 97: 819–30.

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

© 2013 Illumina, Inc. All rights reserved. Illumina, illumina, Dx, BaseSpace, BeadArray, BeadXpress, cBot, CSPro, DASL, DesignStudio, Eco, GAItx, Genetic Energy, Genome Analyzer, GenomeStudio, GoldenGate, HiScan, HiSeq, Infinium, iSelect, MiSeq, Nextera, NuPCR, SeqMonitor, Solexa, TruSeq, TruSight, VeraCode, the pumpkin orange color, and the Genetic Energy streaming bases design are trademarks or registered trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 1370-2013-003 Current as of 18 October 2013

illumina