Japanese Team Combines WGA With 454, Sanger to Sequence Uncultivable Bacteria

[April 15, 2008]

By Julia Karow

A team of researchers in Japan led by RIKEN have coupled whole-genome amplification with Sanger and 454 sequencing to decode the complete genome of a symbiotic bacterium from the termite gut that cannot be cultured.

The scientists, which published their study in the *Proceedings of the National Academy of Sciences* last week, showed that several hundred cells are sufficient to obtain a complete bacterial genome sequence, allowing them to study the bug's role in its natural environment.

Improvements in cell-sorting methods, they say, could make their approach applicable to a wider range of uncultivable bacteria.

To be sure, genomic information about such bacteria can be gleaned from metagenomic analyses, which provide "a comprehensive view on the diversity and functions of a complex microbial community," Yuichi Hongoh, an author of the *PNAS* study, told *In Sequence* by e-mail last week. However, "the taxonomic assignment of the numerously produced genome fragments is almost impossible in most cases."

Also, in order to study the functions and interactions of individual members in such communities, complete genome sequences must be determined, said Hongoh, a research scientist in the Environmental Molecular Biology Lab at the RIKEN Wako Institute.

Producing a genome sequence is especially important for bacteria in "deeply branching lineages" like the one the researchers chose for this project. "Once we [have] determined a complete genome sequence for a bacterium in a novel lineage, it becomes easier to bin the genome fragments by a combination of sequence similarity and other factors, such as GC content, in a metagenomic analysis, which may uncover genetic and functional diversity within the species or genus" he said.

In their project, the researchers sequenced the genome of the Rs-D17 bacterium, a member of the Termite Group 1 phylum of bacteria that lives inside the unicellular eukaryote *Trichonympha agilis*. About 4,000 Rs-D17 bacteria dwell in each *T. agilis*, which in turn lives in the gut of termites.

The researchers harvested several hundred Rs-D17 cells leaking out of the back of a ruptured *T. agilis* cell, yielding approximately 1 picogram of genomic DNA, according to Hongoh.

The reason they collected the bacteria from a single host cell is that "there were significant differences" in the non-coding regions between bacteria obtained from several different host cells, he said.

Using isothermal whole genome amplification, they multiplied the bacterial DNA more than 10 million times, obtaining approximately 50 micrograms of DNA, sufficient for Sanger sequencing of shotgun clones, two runs on the 454 GS 20, and PCR finishing and sequence confirmation of low quality regions.

> *"Reconstruction of complete or nearly complete genome sequences from single bacterial cells is still not practical."*

The researchers chose GE Healthcare's TempliPhi DNA-amplification kit for their whole genome-amplification process, which is based on the Phi29 DNA polymerase. That method, Hongoh said, shows less amplification bias than PCR-based methods, although the extent of bias depends on the sample.

"We actually experienced some difficult cases for other samples in which we could not reduce an extreme WGA bias by any modification of the procedure," Hongoh said. He added that "the mechanism of biased amplification by Phi29 DNA polymerase is unknown."

For Rs-D17, the scientists found less than 6-fold amplification bias, which was "not fatal for completing the genome sequence," Hongoh said.

To sequence the genome, they combined Sanger sequencing of a shotgun library, which covered the entire Rs-D17 genome, with two runs of pyrosequencing on the 454 GS 20, which covered about 98 percent of the genome with 42-fold redundancy.

That redundancy, Hongoh said, "was particularly important in our case" because WGA can generate chimeric sequences, "which can be identified and excluded by highly redundant sequences."

Another advantage of 454 sequencing, he pointed out, is that unlike Sanger sequencing, it is free from cloning bias. However, Sanger sequencing is still needed because homopolymers "occasionally cannot be correctly read by [the]

GS 20," he said.

Combining and analyzing 30 contigs from 454 data with Sanger data "provided the finished, complete circular chromosomal sequence, which contains no gaps," said Atsushi Toyoda, a senior scientist in the Genomic Sciences Center at RIKEN who was also involved in the project.

The method is limited by the large number of bacterial cells that must be available at the start. For example, Hongoh said that when he tried using just a few Rs-D17 cells as a template for the amplification, he observed too much bias to proceed. He estimated that between a hundred and a thousand cells are needed to obtain a complete, or at least near-complete, genome sequence.

Hongoh said that several research groups are currently trying to isolate single bacterial cells, and sequence their DNA following WGA, but "it is not easy to overcome the problems caused by contamination, biased amplification, and chimera formation. Thus, reconstruction of complete or nearly complete genome sequences from single bacterial cells is still not practical."

For example, last fall, researchers at Stanford University led by Steve Quake used a microfluidic chip to isolate individual *E. coli* cells, and amplified their DNA by multiple displacement amplification in small reaction volumes and suggested that it may be possible to sequence and assemble full genomes with this method (see *In Sequence* 9/25/2007).

The Japanese researchers' current approach could be applied to other symbiotic bacteria, Hongoh said, "as long as hundreds of nearly clonal cells can be collected by means of micromanipulation, microdissection, or centrifugation."

Termite guts, for example, harbor other "dominant bacterial species," and other insects contain "one or a few species-specific uncultivable endosymbionts" that could be analyzed using the same strategy, he said.

In order to make their approach applicable to uncultivable free-living bacteria, which are less easy to collect in the hundreds, Hongoh and his colleagues are planning to use improved flow cytometry methods to collect "nearly homogeneous cells" from environmental samples.

Hongoh said he also believes the method could be applied to generate draft genome sequences of eukaryotic microbes that cannot be cultivated.