

Battelle Team Says Next-Gen Sequencing Shows Promise for Forensics, Criminal Investigations

October 05, 2011

Battelle Team Says Next-Gen Sequencing Shows Promise for Forensics, Criminal Investigations

By [Molika Ashford](#)

Researchers from the Battelle Memorial Institute have found that sequencing using the Illumina Genome Analyzer platform not only compares favorably with forensic identification using standard PCR methods but provides additional information that could be useful for criminal investigations.

The group, led by Seth Faith, has been working on adapting sequencing with the anticipation that forensic laboratories are likely to pick up the technology within a few years.

In a presentation at the International Symposium on Human Identification this week, the team reported that its sequencing method was able to identify short tandem repeats as accurately as standard methods. In addition, they found that whole-genome and microbiome sequencing identified SNPs and other variants suggestive of medical, dietary, and geographic details, which could play an important role in police investigations as an adjunct to forensic identification.

Faith told *Clinical Sequencing News* that while information about diet and microbes doesn't improve forensic identification itself, these extra factors could be important for other aspects of criminal investigation.

Sequencing's broader output, he said, "increases the spectrum of alleles that are available for determination on an individual . . . and will help law enforcement, in the future, create data that will lead them in the right direction based on the genetic code."

For example, the researchers reported at the meeting that they were able to find genetic evidence of what one volunteer subject may have eaten — yogurt or cheese — as well as microbial DNA linked to the Ohio region where the subject's cohort was located.

"If someone is trying to build up a case or find a suspect or missing person, there could be some information they could glean off of some of this extra data," he said.

TR Massey, a Battelle spokesperson, explained that current forensics methods include several different techniques. For example, PCR and capillary electrophoresis are used to identify short tandem repeat markers while PCR and mass spectrometry are used for mitochondrial DNA analysis and some labs have adopted next-gen sequencing for SNP-based identification.

Next-gen sequencing could conceivably measure "all those markers," he said, "but what's lacking are protocols that can successfully do that for all of them and that will be useful in a forensic environment."

Faith's work, he said, "is demonstrating that, 'Yes indeed you can determine STR alleles from next-gen sequencing and here's one way to do it.'"

In the study discussed in the ISHI presentation, the Battelle team sequenced DNA from saliva samples from a small group of 35 anonymous donors who provided written descriptions of their ancestry and "some physical features," according to Faith.

The researchers used the Illumina GA to sequence the samples and developed two bioinformatics approaches to impute STR alleles from the raw short-read sequence data. Using this approach, they were able to gather "very solid" data for 12 of the 13 STRs in the Federal Bureau of Investigation's Combined DNA Index System panel.

One bioinformatics pipeline, which Faith discussed in his presentation, is "reference alignment based," he said. The other, developed in collaboration with researchers at Ohio State University, uses a hidden Markov model to evaluate the raw data. OSU researchers presented information on that approach separately at the conference.

Overall, the hidden Markov model was quicker because it evaluated raw data, "however, the reference alignment method was more robust and found more sequence variation such as SNPs," Faith said.

Comparable Results

At the ISHI presentation, Faith discussed the sequencing of six of the 35 samples for STRs as well as one whole-genome analysis. In the abstract for the presentation, the Battelle researchers reported that their sequencing approach matched the accuracy of a standard STR forensic assay. STR loci identified through sequencing "accurately matched allele calls from a [Promega] PowerPlex 16 analysis of the same sample," they wrote.

Additionally, the group was able to interrogate several more layers of genetic information, according to Faith. "We looked at SNPs, did some ancestry determination. We also then looked at physical features such as eye color, hair color, freckling, and baldness," he said.

The researchers also performed microbiome sequencing of the subjects' saliva. "We found one unique [organism] called *Streptococcus thermophilus*, which is used in industrial processes to make yogurt and cheese. So it's likely this person had ingested yogurt or cheese," Faith said. "Interestingly, anecdotally from the genome, this person is also lactose intolerant," he added.

The group also found *Histoplasma capsulatum*, a fungus endemic to the environment of the Columbus, Ohio, area where the study's donor cohort lives, Faith said.

Because of the fungus's geographic specificity, it could be used as a locator in a criminal investigation, Faith suggested. And there are many other microorganisms that could potentially play the same role, according to Massey.

"The thought is that every region of the world has microorganisms that are distributed in that part and not other parts. So what [Faith is] saying is that by examining the microbiome, you can get additional hints as to where a person is from or has been," he said.

Alongside the Illumina work, the researchers also evaluated Roche's 454, and have begun looking at Life Technologies' Ion Torrent PGM as another potential platform. Faith said that they chose to highlight their Illumina results at the conference because the technology has been regarded as the least ideal for forensic sequencing due to its relatively short reads.

"We thought for this presentation it would be more pertinent to show the Illumina data, not only for the depth, but also because the size of the fragments is somewhat smaller than the 454," Faith said. "One of the things with STRs, when using next-generation sequencing, is you need a very long read to cover these repeat regions."

"So 454, because it would read out to 400 base pairs, is certainly amenable to that and there are groups working on that angle. But no one has yet shown data with the Illumina technology ... and here we'll be showing data on the GA, showing it actually can be used for finding STR alleles."

Illumina has also reported interest in applying its sequencing platform to forensic applications. This week, the company announced that it is collaborating with the Institute of Applied Genetics and the Department of Forensic and Investigative Genetics at the University of North Texas Health Science Center to develop forensics protocols using its sequencing technologies ([see related story, this issue](#)).

The Battelle project was conceived as an attempt to work out methods for forensic sequencing in anticipation that other labs will begin adopting sequencing in the near future.

"We've been somewhat forecasting into the future that it's going to become more efficient for laboratories to operate these machines, more data can be produced, and possibly at a cheaper cost per sample," said Faith.

However, while sequencing has shown itself to be comparable to other forensic methods, "it will still have to go through the same rigors in validation that the current methods went through" before it can be adopted in forensic labs, he noted.

Massey added that researchers are "going to have to deal with the error rates on sequencing," which are still high. "That will be problematic when they go to validate this for court acceptance," he said. "

Another issue is that the cost of sequencing is prohibitive compared to standard assays. Nevertheless, Massey said Battelle anticipates that the shift to sequencing for forensic identification may be "a year or two away."

In the meantime, the group plans to continue "working on the methods now that will make that technology successful when the price point gets to where the labs can afford it."

Have topics you'd like to see covered in *Clinical Sequencing News*? Contact the editor at [mashford \[at\] genomeweb \[.\] com](mailto:mashford@genomeweb.com).