lobSTR algorithm rolls DNA fingerprinting into 21st century

April 27 2012, by Nicole Giese Rura

As any crime show buff can tell you, DNA evidence identifies a victim's remains, fingers the guilty, and sets the innocent free. But in reality, the processing of forensic DNA evidence takes much longer than a 60-minute primetime slot.

To create a victim or perpetrator's DNA profile, the U.S. Federal Bureau of Investigation (FBI) scans a DNA sample for at least 13 short tandem repeats (STRs). STRs are collections of repeated two to six nucleotide-long sequences, such as CTGCTGCTG, which are scattered around the genome. Because the number of repeats in STRs can mutate quickly, each person's set of these genetic markers is different from every other person's, making STRs ideal for creating a unique DNA fingerprint.

The FBI first introduced their STR identification system in 1998, when STRs were the darling of the genetics community. However, other identifying genomic markers were soon discovered and gained in popularity. Around the same time, high throughput sequencing allowed researchers to process vast amounts of DNA, but using methods that were ineffectual in repeated DNA, including STRs. STRs were mostly forgotten by geneticists, and innovations to study them stalled.

Now Whitehead Institute researchers have pulled STR identification into the 21st Century by creating lobSTR, a three-step system that accurately and simultaneously profiles more than100,000 STRs from a human genome sequence in one day—a feat that previous systems could never complete. The lobSTR algorithm is described in the May issue of

Genome Research.

"lobSTR found that in one human genome, 55% of the STRs are polymorphic, they showed some difference, which is very surprising," says Whitehead Fellow Yaniv Erlich. "Usually DNA's polymorphism rate is very low because most DNA is identical between two people. With this tool, we provide access to tens of thousands of quickly changing markers that you couldn't get before, and those can be used in medical genetics, population genetics, and forensics."

To create a DNA fingerprint, lobSTR first scans an entire genome to identify all STRs and what nucleotide pattern is repeated within those stretches of DNA. Then, lobSTR notes the non-repeating sequences flanking either end of the STRs. These sequences anchor each STR's location within the genome and determine the number of repeats at the STRs. Finally, lobSTR removes any "noise" to produce an accurate description of the STRs' configuration.

According to Melissa Gymrek, who is the first author of the Genome Research paper, lobSTR's ability to accurately and efficiently describe thousands of STRs in one genome has opened up many new research opportunities.

"The first and simple next step is to characterize the amount of STR variation in individuals and populations," says Gymrek, who was an undergraduate researcher in Erlich's lab when she worked on lobSTR. "This will provide knowledge of the normal range of STR alleles at each locus, which will be useful in medical genetics studies that would like to determine if a given allele is normal or likely to be pathogenic. Another direction we are looking at is to look at STRs in case/control studies to look for STRs associated with disease. The list goes on, but these are some of the first questions we're looking to tackle."

More information: "lobSTR: A short tandem repeat profiler for personal genomes" *Genome Research*, published in advance on April 20, 2012.

Provided by Whitehead Institute for Biomedical Research