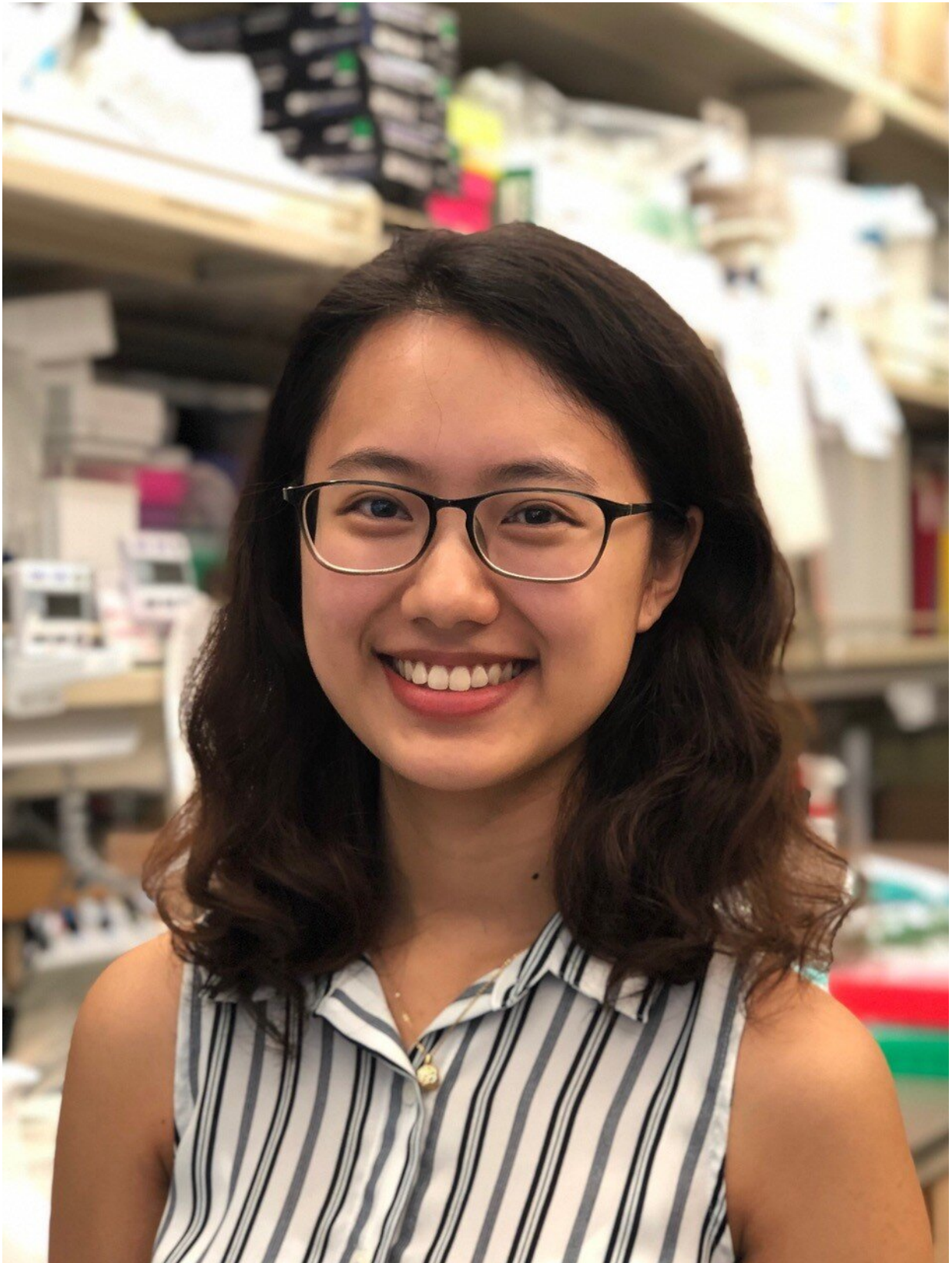


New human reference genome resources help capture global genetic diversity

October 16 2019



Karen Wong, BS, University of California, San Francisco. Credit: Karen Wong

Scientists have assembled a set of genetic sequences that enable the reference genome to better reflect global genetic diversity. The new sequences improve the utility of the human reference genome, a touchstone resource for modern genetics and genomics research, and were presented at the American Society of Human Genetics 2019 Annual Meeting in Houston, Texas

When the Human Genome Project was completed in 2003, its signature achievement was the human [reference genome](#), a set of DNA sequences that serves as a structure and representative example of the complete set of human genes. For areas of the genome where there is little variation among different people, the reference genome is an important resource that has helped move forward efforts in gene sequencing, [genome-wide association studies](#), and protein characterization.

Since almost all genetic sequencing experiments rely on the human reference genome, there is a pressing need to improve the reference to better capture the diversity found in different human populations, explained Karen Wong, BS, a graduate student in Professor Pui-Yan Kwok's laboratory at the University of California, San Francisco (UCSF), who presented the research. A more representative reference would benefit scientists using the millions of existing sequencing datasets, as well as future sequencing studies.

"As integral as it has been to the scientific community, the current reference genome does not represent the genetic diversity found in different human populations. It is limiting because the reference genome is constructed with DNA from a few people, and over 70% of its sequences comes from a single donor," Ms. Wong said. When studying

groups that may have more genetic differences from the reference genome, the reference genome is less useful and may even introduce error or bias to the results.

In particular, researchers are challenged when a particular sample includes insertion sequences—stretches of DNA that, relative to the reference, represent an addition. The length of these insertion sequences can range from a few hundred base pairs to an entire gene duplication, but because they do not exist in the reference genome, they cannot be mapped to the reference and so lack the necessary context to be properly studied. For this reason, such sequences are often discarded.

"The meaning of a sequence derives from knowing where it comes from," Ms. Wong said. "Knowing where a sequence fits in the genome allows researchers to interpret insertions based on sequences around them."

To accomplish this goal, Ms. Wong and colleagues at UCSF and the Institute of Biomedical Sciences at the Academia Sinica in Taiwan fully sequenced more than 300 genomes from around the world, taking care to include both male and female sequences from various subpopulations. Focusing on the areas of the genomes that did not map to the reference genome, they used a process called *de novo* assembly to identify new, unique insertion sequences and their locations relative to the reference. They were able to place a vast majority of the unique sequences missing from the reference, which enabled them to add detail to the reference genome, improving their and other researchers' ability to map future sequences and study them in context.

As next steps, Ms. Wong and colleagues plan to continue sequencing diverse global genomes, as well as explore how to best augment and organize the reference genome to make it most useful to researchers.

"The human genome has many complex regions, which no single reference structure can fully describe," Ms. Wong said. "It is important to consider what information should be represented in the human reference [genome](#), and what doesn't need to be included, to make sure it is helpful for a variety of uses but avoids overcomplexity.

More information: K Wong et al. (2019 Oct 16). Abstract: Constructing a reference genome that captures global genetic diversity for improved interpretation of whole genome sequencing data. Presented at the American Society of Human Genetics 2019 Annual Meeting. Houston, Texas.

Provided by American Society of Human Genetics

Citation: New human reference genome resources help capture global genetic diversity (2019, October 16) retrieved 10 April 2024 from <https://medicalxpress.com/news/2019-10-human-genome-resources-capture-global.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--