

Pacific Biosciences Releases Highest-Quality, Most Contiguous Individual Human Genome Assembly to Date

October 8, 2018

Diploid Assembly of a Puerto Rican Female Adds a Rich Resource to Population-Specific Reference Genomes

MENLO PARK, Calif., Oct. 08, 2018 (GLOBE NEWSWIRE) -- Pacific Biosciences of California, Inc. (Nasdaq:PACB), the leading provider of high-quality sequencing of genomes, transcriptomes and epigenomes, today announced it has produced the most contiguous diploid human genome assembly of a single individual to date, representing the nearly complete DNA sequence from all 46 chromosomes inherited from both parents. The sample used was derived from a Puerto Rican female who was a donor to previous population genetics studies such as the [1000 Genomes Project](#). The new assembly adds to a growing list of high-quality, population-specific human genome reference assemblies generated using PacBio® long-read, Single Molecule, Real-Time (SMRT®) Sequencing.

The new publicly available assembly (PacBio [HG00733](#)) has the fewest gaps of any human genome assembly, with more than half of the genome contained in gapless sequence at least 27 Mb long. The primary contig assembly is 2.89 Gb long and consists of 865 contigs that were assembled with PacBio data generated with the company's Sequel® System. Using the [FALCON-Unzip assembler](#), maternal and paternal haplotypes were resolved over more than 80% of the genome. Maternal and paternal haplotype blocks were then further phased using Hi-C technology and the [FALCON-Phase method](#) developed in collaboration with Phase Genomics. The genome was then *de novo* scaffolded using Phase Genomics' [Proximo Hi-C platform](#), resulting in the first chromosome-scale diploid assembly of a single individual accomplished with only two technologies. More specific details about the assembly are included on the PacBio blog.

"This level of human genome resolution was not possible until now and is uniquely enabled by PacBio sequencing technology," said Michael Hunkapiller, Ph.D., Chairman and CEO of Pacific Biosciences. "Previous sequencing methods were unable to separate the sequence from the 23 chromosome pairs, resulting in human genome assemblies that were half the size and composited in a haphazard manner from DNA sequences inherited from the maternal genome and paternal genome, respectively. It is now possible to resolve haplotype sequences from each parent, resulting in the most complete view of a diploid human genome and the full range of an individual's unique genetic diversity."

"The Genome Reference Consortium has been using data generated with a number of new technologies including PacBio sequencing over the past 5 years as part of our efforts to improve the human genome reference assembly. This new achievement is a prime example of the level of quality that can now be achieved on a single human genome," said Valerie Schneider, member of the Genome Reference Consortium. "We look forward to the availability of this and other highly contiguous, phased assemblies for the opportunities they offer in better understanding and representing human genomic diversity."

The current version of the human reference genome assembly released by the Genome Reference Consortium ([GRCh38](#)) represents the chromosomes of the human genome as a mosaic haploid sequence that was derived from sequencing the DNA of more than 50 individuals and combining the data. By contrast, the PacBio HG00733 diploid genome assembly separately resolves the maternal and paternal chromosome sequences and includes diversity specific to the Puerto Rican population, which is rich itself in ethnic diversity.

SMRT Sequencing has demonstrated that any individual diploid human genome contains more than 20,000 unique structural variants (defined as ≥ 50 bp in length) and another $\sim 400,000$ insertions or deletion variants (ranging in length from 1 bp to 49 bp). Importantly, more than 80% of these variants are not currently accessible using short-read whole genome sequencing methods due to coverage bias, ambiguity in read mapping and inability to span large variants. In contrast, sensitive detection of these larger variants in human genome studies has been widely demonstrated using PacBio long-read sequencing. More than 40 [global initiatives](#) are currently underway to apply these *de novo* assembly methods to individuals representing multiple ethnic populations, thereby extending the diversity of available human reference genomes.

"In order to enable precision medicine for all populations, it is crucial to achieve high-quality DNA sequencing and to better represent true ethnic diversity within genomic databases," said Jonas Korlach, PhD, Chief Scientific Officer of Pacific Biosciences. "In addition to the many human population sequencing projects underway, PacBio is enabling major initiatives to sequence every type of life form on earth, from the smallest bacteria to the most complex plants and animals."

The data are available using NCBI accession IDs: BioProject: ([PRJNA483067](#)), assembly: [[RBJD00000000](#)] and sequence data ([SRP155659](#)).

Additional Resources

- [Interactive map](#) showcasing global initiatives underway to generate reference-quality human genome assemblies for diverse populations
- [BioReport Podcast](#) on the value of ethnic-specific reference genomes
- *Nature Reviews Genetics* paper from NHGRI: [Prioritizing diversity in human genomics research](#)
- Article in *The Journal of Precision Medicine*: "[Minority Report – Ethnic Diversity and the Real Promise for Precision Medicine](#)"
- Article in *Bio-IT World*: "[Genomic Data Standards Are a Necessity](#)"
- NHGRI Project Award: [High Quality Human and Non-Human Primate Genome Assemblies](#)

More details are available on the PacBio website:

- Blog post: [Data Release: Highest-Quality, Most Contiguous Individual Human Genome Assembly to Date](#)

- Blog post: [For Reference-Grade Human Genome Assemblies, SMRT Sequencing Yields Optimal Results](#)
- Webinar: [Assembling High-Quality Human Reference Genomes for Global Populations](#)
- FALCON-Phase [press release](#) and article [preprint](#)
- PacBio research focus webpage about [Human Population Genetics](#)

About Pacific Biosciences

Pacific Biosciences of California, Inc. (NASDAQ:PACB) offers sequencing systems to help scientists resolve genetically complex problems. Based on its novel Single Molecule, Real-Time (SMRT[®]) technology, Pacific Biosciences' products enable: *de novo* genome assembly to finish genomes in order to more fully identify, annotate and decipher genomic structures; full-length transcript analysis to improve annotations in reference genomes, characterize alternatively spliced isoforms in important gene families, and find novel genes; targeted sequencing to more comprehensively characterize genetic variations; and real-time kinetic information for epigenome characterization. Pacific Biosciences' technology provides high accuracy, ultra-long reads, uniform coverage, and the ability to simultaneously detect epigenetic changes. PacBio[®] sequencing systems, including consumables and software, provide a simple, fast, end-to-end workflow for SMRT Sequencing. More information is available at www.pacb.com.

Forward-Looking Statements

All statements in this press release that are not historical are forward-looking statements, including, among other things, statements relating to future availability, uses, accuracy, quality or performance of, or benefits of using, products or technologies, the suitability or utility of methods, products or technologies for particular applications, studies or projects, the expected benefits of sequencing projects, and other future events. You should not place undue reliance on forward-looking statements because they involve known and unknown risks, uncertainties, changes in circumstances and other factors that are, in some cases, beyond Pacific Biosciences' control and could cause actual results to differ materially from the information expressed or implied by forward-looking statements made in this press release. Factors that could materially affect actual results can be found in Pacific Biosciences' most recent filings with the Securities and Exchange Commission, including Pacific Biosciences' most recent reports on Forms 8-K, 10-K and 10-Q, and include those listed under the caption "Risk Factors."

Pacific Biosciences undertakes no obligation to revise or update information in this press release to reflect events or circumstances in the future, even if new information becomes available.

Contacts

Media:

Nicole Litchfield

415.793.6468

nicole@bioscribe.com

Investors:

Trevin Rard

650.521.8450

ir@pacificbiosciences.com



Source: Pacific Biosciences of California, Inc.