Title

A Metagenomic Analysis of Environmental and Clinical Samples Using a Secure Hybrid Cloud Solution.

Journal

Journal of Biomolecular Techniques, 30(Suppl)

ISSN

1524-0215

Authors

Ounit, Rachid
Mason, Chris
Lonardi, Stefano
et al.

Publication Date

2019-12-01

Peer reviewed

The libraries were run on the Sequel System with chemistry v3.0 and software v6.0, generating ~11 Gb of sequence per SMRT Cell with 10 hour movies, and followed by de novo genome assembly with FALCON. The resulting assemblies had high contiguity (contig N50s over 1 Mb) and completeness (as determined by conserved BUSCO gene analysis) when at least 30-fold unique molecular coverage is obtained.

This new low-input approach now puts PacBio-based assemblies in reach for small highly heterozygous organisms that comprise much of the diversity of life. The method presented here is scalable and can be applied to samples with starting DNA amounts of 150 ng per 500 Mb genome size.

4

A Metagenomic Analysis of Environmental and Clinical Samples Using a Secure Hybrid Cloud Solution

Rachid Ounit[1], Chris Mason[2], Stefano Lonardi[1], Niamh O'Hara[3]

[1]Department of Computer Science & Engineering, University of California, Riverside, [2]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, [3]Jacobs Technion-Cornell Institute, Cornell Tech, New York

The number and types of studies about the human microbiome, metagenomics and personalized medicine, and clinical genomics are increasing at an unprecedented rate, leading to computational challenges. For example, the analysis of patient/clinical samples requires methods capable of (i) accurately detecting pathogenic organisms, (ii) running with high speed to allow short response-time and diagnosis, and (iii) scaling to ever growing databases of reference genomes. While cloud-computing has the potential to offer low-cost solutions to these needs, serious concerns regarding the protection of genomic data exist due to the lack of control and security in remote genomic databases.

We present a novel metagenomic analysis system called "Virgile" that is capable of performing privacy-preserving queries on databases hosted in outsourced servers (e.g., public or cloud-based). This method takes as input the sequenced data produced by any modern sequencing instruments (e.g., Illumina, Pacbio, Oxford Nanopore) and outputs the microbial profile using a database of whole genome sequences (e.g., the RefSeq database from NCBI). The algorithm for the microbial profile aims to estimate without bias the abundance of microorganisms present using a genome-centric approach.

Result: Using an extensive set of 65 simulated datasets, negative and positive controls, real clinical samples, and mock communities, we show that Virgile identifies and estimates the abundance of organisms present in environmental or clinical samples with high accuracy compared to state-of-the-art and popular methods available, including MetaPhlAn2 and KrakenUniq. Running at high speed, Virgile can also be run on a standard 8 GB RAM laptop.

Virgile is a novel privacy-preserving abundance estimation algorithm called Virgile that can efficiently and rapidly discern the abundance and taxonomic identification of organisms present in a metagenomic sample, including those from medical environments. To the best of our knowledge, Virgile is the only metagenome analysis system leveraging cloud computing in a secure manner.

5

A Robust, Streamlined, Enzyme-based DNA Library Preparation Method Amenable to a Wide Range of DNA Inputs

Amanda Hoppers[1], Lynne Apone[1], Vaishnavi Panchapakesa[1], Karen McKay[1], Pingfang Liu[1], Chen Song[1], Christine Sumner[1], Christine Rozzi[1], Melissa Arn[1], Jonathan Sanford[1], Bradley W. Langhorst[1], Timur Shtatland[1], Fiona J. Stewart[1], Daniela Munafo[1], Eileen T. Dimalanta[1], Theodore B. Davis[1]

[1]New England Biolabs

Advances in next generation sequencing (NGS) platforms have outpaced those in library preparation. While hundreds to thousands of NGS libraries can be sequenced in a single run of an Illumina instrument, a single library is constructed using a multi-step procedure requiring expensive consumables and specialized equipment. To overcome these limitations and increase the ease and throughput of library construction, we have developed a robust, streamlined library construction method that integrates enzyme-based DNA fragmentation with end repair and dA-tailing. The NEBNext Ultra II FS DNA Library Prep Kit eliminates the need for specialized equipment to fragment DNA and reduces the number of steps in the library construction protocol. This method produces high quality libraries from gDNA isolated from organisms whose genomes vary widely in GC content, as well as amplicons and DNA purified from blood. In addition, the FS kit generates libraries with substantially higher yields than those using mechanically sheared DNA, enabling greatly reduced input requirements. Libraries constructed using the FS kit from inputs as low as 100 pg of human gDNA for amplified libraries and 100 ng for PCR free, show similar coverage uniformity and GC bias compared to libraries constructed with 100 ng of mechanically sheared DNA.

The ability to generate high quality libraries from low amounts of starting material and a broad range of inputs will help advance the widespread implementation of NGS in both basic science and the clinic.

6

ABRF-sPRG 2018-2019: Development and Characterization of a Stable-Isotope Labeled Phosphopeptide Standard

Anthony Herren[1], Brian Searle[2], David Hawke[3], Kimberly Lee[4], Antonius Koller[5], Ryan Leib[6], Brett Phinney[1], Allis Chien[6], Bhavin Patel[7]

[1]University of California, Davis, [2]University of Washington, Proteome Software, [3]UT MD Anderson Cancer Center, [4]Cell Signaling Technology, [5]Northeastern University, [6]Stanford University, [7]Thermo Fisher Scientific