**Incorporating Next-Gen Sequencing Data and Automation into JCVI’s Viral Finishing Pipeline**

**Nadia Fedorova, Danny Katzel, Tim Stockwell, Jessica Hostetler, Peter Edworthy Rebecca Halpin, David Spiro, and David Wentworth**

The J Craig Venter Institute, Rockville, MD, U.S.A.

JCVI is now in its sixth year of high throughput viral genomics projects. Over ten different viral projects are in progress supported by the NIAID Genomic Sequencing Center for Infectious Disease (GSCID). These projects represent many viruses including Influenza, Coronavirus, Rotavirus, Paramyxovirus, Adenovirus, Arbovirus, Measels, Mumps, Rubella, and Norovirus. The viral sequencing and finishing pipeline at JCVI employs both amplicon-based Sanger sequencing and next generation sequencing technologies. These approaches, combined increasingly with automated data processing, have allowed us to complete over X viral genomes in the last 12 months, and X genomes since 2006.

A highly automated and optimized Sanger sequencing and finishing pipeline, initially developed for influenza viruses, is now integrated into all viral projects. The primary software component, **VAPOR**, is a suite of tools that loads sequencing reads into a project database, assembles, and validates viral samples. Vapor’s companion tool, **autoTasker**, examines samples for low coverage and low quality areas, suggests PCR tasks needed to finish samples, and produces NCBI submission files. Ultimately we want to combine these programs into a single, integrated software suite for rapid and efficient viral genome sequencing.

Our next-gen sequencing and finishing pipeline utilizes SISPA-generated genomic libraries with 454 and Illumina sequencing technologies. Thus, we are able to efficiently and completely sequence the genomes of traditionally challenging samples (e.g., avian influenza, and previously unknown viruses). The automated next-gen assembly pipeline employs **CLC** command-line tools and a cas to ace conversion tool, called **cas2consed.** Finishing work is performed using the widely available **Consed** editor assited by the integration of **autoTasker** quality control scripts**.**

To streamline our viral pipelines, we are adapting **JIRA** for sample tracking and our goal is to create a semi-automated tracking interface that follows the progress of viral samples from acquisition through to NCBI submission. The combination of highly optimized sequencing technologies and automated software tools allows for large volumes of sample processing with limited manual interaction. Although these new developments will reduce labor and costs, they will increase speed and production capacity.

The advantages of these innovative technologies, reduced labor per sample combined with the ability to manage a higher volume of samples, will increase the number of viable applications for genomic sequencing and provide the scaling needed to support this growth.