BCVISS: a web application for analyzing mixed 16S rRNA gene chromatograms

A.V. Popova, L. S. Karan, S. N. Gavrilov, A. D. Neverov

Central Research Institute of Epidemiology, Moscow, Russia, weidewind@gmail.com

Direct sequencing of 16S rRNA gene is commonly used in diagnostics of septic diseases for identification of bacterial pathogens. In contrast to bacteriological methods, sequencing is faster; it can detect bacteria that are unculturable or difficult to culture. NGS platforms are currently limited in usage in clinical practice, mainly owing to a need for complete device load for cost effectiveness. In many cases Sanger sequencing is still less expensive and faster than high-throughput methods, and it is broadly available in clinics. A substantial proportion of clinical samples contain mixtures of different bacterial species, and their sequencing results are difficult to interpret with standard methods. Base Caller with Vocabulary (BCV) is a tool for deciphering of mixed chromatograms, which uses a vocabulary of sequences similar (but not essentially identical) to the target DNA [1].

Now we present BCVISS (Base Calling with Vocabulary and Identification of pathogenic bacteria taxonomy based on the Small Subunit rRNA gene sequence**)**, a web application specially designed for deciphering direct 16S rRNA sequencing chromatograms in ABIF format obtained for human clinical samples. BCVISS identifies sequences comprising the data and assigns them to taxonomic categories. Taxonomic identification is based on phylogenetic analysis, performed with the help of modified STAP method [2]. Unlike the original version, which doesn't provide confidence of a taxonomy assignment, modified method assigns to a query sequence a taxonomy category that corresponds to the nearest ancestral node with confidence value above the predefined significance threshold.]. This version leverages GreenGenes taxonomic database [3] (2013 release), stored on the server. Besides the list of taxons and their confidence values, BCVISS outputs names of the closest relatives present in the aforementioned node, the list of sequences predicted by BCV (.fasta), their expectation, the best hit found in GreenGenes by BLASTN and its identity and coverage. At the moment BCVISS is the only publicly available free application capable of deciphering mixed chromatograms. The only other application at the time being that could

process mixed chromatograms is a commercial online service, RipSeq [4]. Ripseq exclusively accepts chromatograms that are read from a restricted set of sequencing primers, while BCVISS can process chromatograms corresponding to any part of 16S gene.

We show utility of BCVISS on several samples. For 8 chromatograms obtained from RipSeq website (https://www.ripseq.com/login/login.aspx) RipSeq annotates totally 18 taxa whereas BCVISS recovers 16 taxa that match on species- or genus-level (13 and 3 respectively). We also reanalyzed two gastric mucosa samples from our previous work [1] that were characterized by cloning and sequencing of PCR amplified 16S gene. Due to the last version of GreenGenes database used by BCVISS taxonomy assignments of identified species are improved. In addition we successfully applied BCVISS to mixed chromatograms obtained for two clinical samples, for which bacteriological results were available: a case of fulminant septicemia caused by *Capnocytophaga canimorsus*, rarely associated with human infections, and a case of septicemia caused by *Streptococcus* of *mitis* group.

BCVISS is available at http://bioinf.fbb.msu.ru:8080/bcviss/bcvjob/form.gsp

1. Y.S.Fantin et al. (2013) Base-Calling Algorithm with Vocabulary (BCV): Method for Analyzing Population Sequencing Chromatograms, *PLoS ONE,* **8(1):** e54835.

2. D. Wu et al. (2008) An Automated Phylogenetic Tree-Based Small Subunit rRNA Taxonomy and Alignment Pipeline (STAP), *PLoS ONE, **3*****(7):** e2566.

3. T. Z. DeSantis et al. (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB, *Applied and Environmental Microbiology*, **72:**5069-72.

4. Ø. Kommedal, et al. (2009) Direct 16S rRNA Gene Sequencing from Clinical Specimens, with Special Focus on Polybacterial Samples and Interpretation of Mixed DNA Chromatograms, *Journal of Clinical Microbiology*, **47(11):**3562-3568.