Taxonomic Unit Identification Tool: a bioinformatical approach for the improvement of human cornea microbiome classification using 16S rRNA gene sequencing data

Alexander I. Tuzhikov

Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami, School of Medicine, Miami, FL, United States, alexander.tuzhikov@gmail.com

Alexander Y. Panchin

Institute for Information Transmission Problems, Russian Academy of Science, alexpanchin@yahoo.com

Qunfeng Dong

Departments of Biological Sciences, Computer Science and Engineering, University of North Texas, Texas, TX, United States, qunfeng.dong@unt.edu

David Nelson

Department of Biology, Indiana University, Bloomington, IN, United States nelson@indiana.edu

Terrence O'Brien

Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami, School of Medicine, Miami, FL, United States, tpob@hotmail.com

Valery I. Shestopalov

Department of Ophthalmology, Bascom Palmer Eye Institute, University of Miami, School of Medicine, Miami, FL, United States, VShestopalov@med.miami.edu

The healthy cornea is typically culture-negative and is considered nearly free of microbiota due to the bactericidal action of tear film. However, our new experimental data obtained using next generation DNA sequencing of 16S RNA genes indicates the presence of a resident microbial community on the healthy ocular surface (OS). When compared to culture-based methods, deep sequencing of the 16S rRNA gene libraries of total DNA from healthy conjunctiva and cornea has detected a much more diverse bacterial community. However, some bacterial genera are consistently "overlooked" by mainstream sequence classification algorithms such as the RDP-II Classifier and many 16S rRNA gene sequences identified in the human ocular surface remain unclassified if only available bioinformatical approaches are used. To become

clinically feasible, pathogen identification based on next generation sequencing has to deliver the species or even the strain levels of identification for as many microbes as possible. Due to the redundant sequence similarity between closely related species, a clear discern remains a challenging task. A plain homology-based search constantly provides a vast amount of highly similar sequences from different taxonomic subgroups, and thereby it is imperative that a robust algorithm would explicitly identify a pivotal taxonomic subunit within the output. The purpose of this study was to improve microbial classification using a combination of the BlastN algorithm against the most current GenBank database and a data filtering algorithm for the analysis of microbiota at the human cornea.

In order to address the problem of micriobial classification based on 16S rRNA gene sequences, we developed a new algorithm TUIT (Taxonomic Unit Identification Tool) and applied it to annotate (classify) 16S reads rejected (unclassified beyond Bacteria) by RDP-II Classification tool. TUIT utilizes BlastN–based search against the latest sequencing data available with the most current sequencing GenBank databases. For 1300 16S RNA gene sequences obtained from the human cornea that remained "unclassified beyond Bacteria" with RDP-II Classification our algorithm was able to 1) classify 29% of the sequences at the genus level, 2) classify 1% of the sequences at the species level. Importantly, out of 165 bacterial genera identified on the human ocular surface, 105 were only detected by TUIT and not by RDP-II Classifier. This includes some bacterial genera that appeared to be well-represented in the human ocular surface such as *Rickettsia*, *Holospora*, *Chitinophaga*, *Geobacter*, *Spirillum* and *Cardiobacterium*

Our approach to combine RDP-II Classifier and TUIT significantly improved the characterization of the ocular microbiome diversity by reducing the fraction of unclassified phylotypes and allowed us to achieve species-level characterization for some representatives of the ocular surface microbiota. The TUIT program may be applied for the study of other microbiomal data and might be of interest for other researchers in the field of microbiome analysis.